

RNA-Chromatin Interactome Reveals Enhancer-Promoter Connectivity in 3D Genome

Xiao Li^{*1}, Bing Zhou^{*1}, Liang Chen¹, Hairi Li¹, and Xiang-Dong Fu^{1,2}

¹Department of Cellular and Molecular Medicine, ²Institute of Genomic Medicine, University of California, San Diego, La Jolla, CA 92093-0651, USA

*co-first authors

Corresponding author:

Xiang-Dong Fu

Email: xd.fu@ucsd.edu

Phone: 858-534-4937

Summary

High eukaryotic genomes are populated with enhancers, but it has been a major challenge in defining specific enhancer-promoter relationship. Enhancers can also be divided into typical and super-enhancers, yet their functional distinctions remain to be understood. Here, we report a strategy to capture *in situ* Global RNA Interactions with DNA by deep sequencing (GRID-seq). By deducing general RNA background on chromatin, we unexpectedly detect a highly selective set of RNAs (including both lncRNAs and protein-coding pre-mRNAs) decorated on enhancers, particularly super-enhancers. Based on the origins of those RNAs and functional perturbation of enhancer activities, we are able to infer global enhancer-promoter connectivity, which is significantly beyond the traditional framework. These findings provide a functional RNA-chromatin interactome in 3D genome.

Introduction

Recent decades of genomic research reveal that mammalian genomes are more prevalently transcribed than previously anticipated¹. It is now quite clear that mammalian genomes express not only protein-coding RNAs but also a large repertoire of non-coding RNAs that have regulatory functions in different layers of gene expression. Many of those regulatory RNAs appear to directly act on chromatin, as exemplified by various lncRNAs^{2,3}. Some of those regulatory RNAs mediate genomic interactions only in *cis*, while others, such as *MALAT1* and *NEAT1*, are capable of acting in *trans*⁴. These findings suggest an emerging paradigm in regulated gene expression via specific RNA-chromatin interactions.

Various techniques have been developed to localize specific RNAs on chromatin. These methods, such as Chromatin Isolation by RNA Purification or ChIRP⁵, Capture Hybridization

Analysis of RNA Targets or CHART⁶, and RNA Affinity Purification or RAP⁷, all rely on using complementary sequences to capture a specific RNA followed by deep sequencing to identify targets on chromatin. Importantly, all of these methods only allow analysis of one known RNA at a time, and up to date, a global view is lacking on all RNA-chromatin interactions, which is critical to address a wide range of functional genomics questions.

One of the most pressing questions is which enhancers interact with individual gene promoters to regulate gene expression. Current approaches to this problem rely on long distance DNA-DNA interactions detected by Hi-C or ChIA-PET coupled with immunoprecipitation of RNA polymerase II (RNA Pol II) and DNA-DNA interaction mediators⁸⁻¹⁰. However, such interactions are often embedded in numerous static physical interactions confined within largely cell type-independent topologically associating domains (TADs), which would thus impede accurate assignment¹¹. Another pressing question concerns the newly introduced concept of super-enhancers^{12,13}. As super-enhancers may be argued as a set of stitched typical enhancers¹⁴, it has been unclear whether super-enhancers are indeed superior over typical enhancers in action range, potency or both in activating promoters.

To fill these critical gaps, we sought to develop a general approach for systematic localization of all potential chromatin-interacting RNAs in an unbiased fashion. Here we present such an approach for mapping Global RNA Interactions with DNA by deep sequencing (GRID-seq) via using a bivalent linker to ligate RNA to DNA *in situ* and present a global picture on RNA-chromatin interactions, exposing distinct classes of *cis*- and *trans*-chromosomal interacting RNAs in both human and *Drosophila* genomes. Our analysis reveals a large set of both coding and non-coding RNAs that are prevalently associated with enhancers, particularly super-enhancers, which enable us to infer enhancer-promoter connectivity in 3D genomes.

Results

Ligating RNA to proximal DNA *in situ*

We chose a triple negative breast cancer MDA-MB-231 cell line to develop a global and unbiased strategy to map RNA-chromatin interactions. To stabilize RNAs on chromatin, we double fixed cells with disuccinimidyl glutarate (DSG) and formaldehyde, isolated nuclei, and digested DNA *in situ* with a frequent cutter (AluI). We designed a biotin-labeled bivalent linker consisting of an ssRNA portion for ligation only to RNA and a dsDNA portion for ligation only to DNA ([Extended Data Fig. 1a](#)). As diagrammed in [Fig. 1a](#), part 1, we first performed RNA ligation *in situ*, washed away excess free linkers, and extended the DNA primer in the linker into ligated RNA with reverse transcriptase. We next ligated the linker to local genomic DNA followed by affinity purification with streptavidin beads. In [Fig. 1a](#), part 2, we released ssDNA from beads, converted to dsDNA, and used the type II restriction enzyme MmeI to cut DNA ~20 nt upstream and downstream from the two built-in recognition sites in the linker. The products were resolved in native gel to detect two defined DNA fragments with the larger one (~85 bp) corresponding to linker ligation to both RNA and DNA and the smaller one (~65 bp) to linker ligation to either RNA or DNA. We isolated the larger fragment for amplification and adapter ligation for deep sequencing, typically generating >200 million 100 nt raw reads (~40 million uniquely mapped read mates) per library on human cells (see [Extended Methods](#)), which showed high concordance between replicates ($R^2 > 0.95$, [Extended Data Fig. 1b,c](#)).

We developed a strategy for controlling non-specific interactions by mixing isolated nuclei from MDA-MB-231 cells and *Drosophila* S2 cells ([Fig. 1b](#)). Because the human genome is much larger than the *Drosophila* genome, we mixed human and *Drosophila* cells in roughly equal molar genome. By using uniquely and unambiguously mapped RNA and DNA read mates

to human or *Drosophila* genome (see [Extended Methods](#)), we estimated 6.8% human RNA linked to *Drosophila* DNA and 8.4% *Drosophila* RNA linked to human DNA ([Fig. 1b](#)). While these cross-species mates were likely resulted from ligation of fragmented RNAs in solution, which randomly landed on exposed chromatin during the GRID-seq reaction, the majority of RNA-chromatin interactions we detected reflected *in situ* interactions in both human and *Drosophila* genomes. Such non-specific cross-genome interactions in fact enabled us to construct a general background, which is essential to evaluate specific RNA-chromatin interactions within each genome (see below).

Identifying hit RNAs on chromatin

We rank-ordered genes based on RNA reads on chromatin, observing a set of RNAs, including both known lncRNAs and various pre-mRNAs, that prevalently interacted with chromatin in both human and *Drosophila* genomes (red-labeled in [Fig. 1c](#), [Extended Data Fig. 2a](#)). These chromatin-interacting RNAs are expressed from a small fraction of active genes with RNA reads evenly derived from their gene bodies, yet account for ~90% of all detected RNAs on chromatin ([Extended Data Fig. 2b](#)). Close examination of individual examples revealed that many RNAs showed discrete binding while others exhibited relatively scattered distribution on chromatin. We therefore plotted length-normalized abundance of RNAs against their highest binding density on chromatin to identify RNAs with sufficient densities in both RNA reads and interactions on chromatin, leading to 973 “hit” RNAs in MDA-MB-231 cells ([Supplementary Table 1](#)). These hit RNAs consist of both small RNAs (2.1%, which predominantly correspond to snRNAs and snoRNAs, orange-labeled in [Fig. 1d](#), [Extended Data Fig. 2c](#)) and long RNAs (including 6.7% lncRNAs and 87.1% protein-coding pre-mRNAs, red-labeled in [Fig. 1d](#),

[Extended Data Fig. 2c](#)). When compared with gene expression, it became evident that hit RNAs are not necessarily from highly expressed genes in both human and *Drosophila* cells ([Fig. 1e](#), [Extended Data Fig. 2d,e](#)). These data revealed a set of RNAs with propensity to frequently and specifically interact with chromatin in human and *Drosophila* genomes.

Taking advantage of the small *Drosophila* genome, thus having sufficient density of human RNAs linked to its genome, we constructed a background for non-specific RNA-chromatin interactions. Besides randomly fragmented RNAs, expressed RNAs are also known to extensively explore nuclear space after being released from their sites of transcription ¹⁵. We thus reasoned that *trans*-chromosomal signals from pre-mRNAs of protein-coding genes might be used to construct a background. Indeed, such deduced background with endogenous RNAs in *Drosophila* S2 cells is highly concordant with that built from *trans*-species RNA-chromatin interactions ([Extended Data Fig. 3a,b](#)). This suggests that we may use endogenous RNAs to build a similar background for human cells, which is reproducible based on replicated GRID-seq datasets ([Extended Data Fig. 3c](#)). GRID-seq signals significantly above the background may therefore reflect RNAs that either traveled extensively to engage in specific *trans*-chromosomal interactions or interacted with chromatin in spatial proximity to their sites of transcription.

Global view of RNA-chromatin interactions

When displaying all hit RNAs on chromosomes in the human genome after normalization against the background ([Extended Methods](#)), it was clear that the majority of RNAs interacted with chromatin near their sites of transcription, and surprisingly, a limited number of RNAs were extensively engaged in *trans*-chromosomal interactions, such as U2 snRNA and two pseudo U2 snRNAs, *MALAT1* and *NEAT1* ([Fig. 2a](#)). Both *MALAT1* and *NEAT1* reside in Chromosome 11

and an enlarged view of Chromosomes 11 and 17 showed their prevalent interactions with similar efficiency on all other chromosomes in the human genome. Further analysis revealed their general preference for active chromatin regions, such as promoters marked by H3K4me3 and active enhancers decorated with H3K27ac ([Extended Data Fig. 4a](#)), which were concordant with previous reports ^{6,16}. We also detected abundant 7SK RNA on chromatin ([Extended Data Fig. 4a](#)), similar to the signals detected by ChIRP ¹⁷, but its interactions with chromatin in most locations were similar to background ([Fig. 2a](#)), as background RNA-chromatin interactions also tend to associate with open chromatin ([Extended Data Fig. 4b](#)). With the expended chromosomal view, we noted multiple coding (pc) and non-coding (nc) RNAs that are able to interact with numerous loci in the same chromosomes ([Fig. 2b](#)), suggesting that these RNAs may help define (as either cause or consequence) nuclear territories in the cell. Such extensive RNA interactions with chromatin are highly reproducible based on replicated GRID-seq experiments, even with increasing resolutions ([Extended Data Fig. 5a,b](#)).

In *Drosophila* S2 cells, we also detected a large number of RNAs on chromatin, and enlarged chromosomal view showed *roX2*, a lncRNA known to be involved in dosage compensation in *Drosophila* ¹⁸, decorated only on Chromosome X ([Extended Data Fig. 6a](#)). We further compared our *roX2*-chromatin interaction results with the published *roX2* ChIRP and ChART data ^{5,6} as well as the ChIP-seq data on MSL3, a known *roX2*-interacting factor ¹⁹ and observed a high degree of concordance among all mapping results on Chromosome X ([Extended Data Fig. 6b,c](#)). These data unambiguously demonstrated the ability of unbiased GRID-seq to recapitulate known specific RNA-chromatin interactions, thus empowering discovery and characterization of new RNA-chromatin interactions.

RNA-chromatin interaction ranges

Some RNAs appeared to interact with chromatin within rather confined regions while others showed more broad chromatin interactions within the same chromosomes as well as specific interactions across chromosomes. We therefore characterized hit RNAs based on their chromatin-interacting ranges in local (± 10 Kb from their genes), *cis* (beyond local regions) or *trans* (across chromosomes). We first analyzed such behaviors of lncRNAs, observing that the majority of lncRNAs showing all three modes of chromatin interactions with a few exceptions (Fig. 2c). A set of lncRNAs, as exemplified by *HCG18*, were rarely engaged in *trans*-chromosomal actions, while *U2 snRNA* showed no local action, likely because it interacts with chromatin during co-transcriptional splicing only after being assembled into small nuclear ribonucleoprotein particle (snRNP). Circos plots²⁰ further illustrated the ability of *MALAT1* to interact with chromatin in all three modes and with similar efficiency, while *PVT1* was predominantly engaged in local and *cis*-chromosomal interactions (Fig. 2d).

Interestingly, when the same analysis was applied to chromatin-interacting pre-mRNAs, the majority of them were able to participate in all three modes of interactions, but without any exclusively acting in *trans* (Fig. 2e). Circos plots again showed *HMG42* pre-mRNA interacted with chromatin in all three modes, although with much reduced frequencies in *trans*-chromosomal interactions, while *SMAD5* pre-mRNA interactions with chromatin were largely confined in local and *cis*-chromosomal interactions (Fig. 2f). These findings suggest that many pre-mRNAs may behave like lncRNAs in the nucleus. Additionally, while *Drosophila* S2 cells showed the same trend as human cells, a much larger number of small RNAs, predominantly snoRNAs, were involved in chromatin interactions in all three modes (Extended Data Fig. 7, Supplementary Table 1).

Cell type-specific interactions

We next wondered whether specific RNA-chromatin interactions we discovered reflect cell type-specific activities in mammals, and therefore extended GRID-seq analysis to another well-characterized human multiple myeloma cell line MM.1S, which would enable us to take advantage of previously generated functional data on this cell type²¹. As on MDA-MB-231 cells, we again detected a set of *trans*-acting RNAs, including *MALAT1* and *NEAT1* ([Extended Data Fig. 8a](#)), as well as both coding and non-coding RNAs that broadly interacted with numerous loci near their sites of transcription. Interestingly, we detected *XIST* (expressed in MM.1S cells, but not in MDA-MB-231 cells), which extensively decorated on Chromosome X ([Extended Data Fig. 8b](#)), further demonstrating the power and accuracy of GRID-seq in identifying specific RNA-chromatin interactions.

Interestingly, cross analysis between MDA-MD-231 and MM.1S cells revealed cell type-specific RNA-chromatin interactions ([Fig. 3a,b](#)), as exemplified by different RNA signals on Chromosome 4 ([Fig. 3c](#)) and by differential binding of a selective set of hit RNAs on Chromosome 6 ([Fig. 3d](#)). In contrast, background RNA-chromatin interactions were similar between MDA-MB-231 and MM.1S cells ([Extended Data Fig. 9a,b](#)). This is reminiscent of enhancers, the majority of which are cell type and tissue-specific, as shown previously based on H3K4me1/2 and H3K27ac²². Indeed, we noted that nearly all chromatin-interacting RNAs were linked to both shared and cell type-specific enhancers ([Fig. 3e](#), [Extended Data Fig. 9c,d](#)) and in a quantitative fashion ([Extended Data Fig. 9e,f](#)) in both MDA-MB-231 and MM.1S cells. This can be directly visualized on specific examples, showing not only cell type-specific interactions of RNAs with their own genes but also with nearby enhancers ([Fig. 3f,g](#)). Surprisingly, even with

the same RNA detected in both cells, it reached out to distinct enhancers (Fig. 3h), suggesting cell type-specific rewiring of transcription program. Although we did not have sufficient read density for eRNAs, which are believed to link enhancers to promoters²³, our data indicated that pre-mRNAs from actively transcribed genes were also associated with their enhancers, perhaps reflecting putative hubs for enhancer-promoter interactions in the nucleus.

Prevalent RNAs on super-enhancers

Recent studies suggested that enhancers can be segregated into typical and super-enhancers, the latter of which were defined based on much higher density of enhancer marks, such as MED1 and BRD4 ChIP-seq signals, which generally tracked H3K27ac, and importantly, such “stitched” enhancers appeared to be more potent in activating nearby genes^{13,21}. Given most enhancers were associated with GRID-seq signals, we sought to determine whether such signals could also be used to define the strength of typical versus super-enhancers. By sorting enhancers based on RNA coverage, we found that enhancers highly decorated with RNAs mostly corresponded to super-enhancers in both MDA-MB-231 and MM.1S cells (Fig. 4a, Extended Data Fig. 9g), which was also evidenced by much higher RNA coverage on super-enhancers relative to typical enhancers in both cell types (Fig. 4b, Extended Data Fig. 9h). Therefore, chromatin-associated RNAs may provide yet another independent measure of enhancer activities.

As super-enhancers are more potent than typical enhancers in activating nearby genes in MM.1S cells²¹, we sorted the GRID-seq signals based on RNA coverage and compared the expression of neighboring genes from 50 Kb flanking enhancers by using the published gene expression data on MM.1S cells. We found that genes adjacent to top 10% RNA-decorated enhancers were more active than those adjacent to bottom 10% (Fig. 4c,d), and similarly, the

genes associated with top 10% RNA-decorated enhancers were more responsive than those with bottom 10% to functional perturbation with the BRD4 inhibitor JQ1 (Fig. 4e), thus confirming the previous conclusions^{13,21}. We performed the same set of experiments on MDA-MB-231 cells by using GRO-seq to score nascent RNA production and transcriptional response to JQ1, and reached the same conclusions (Extended Data Fig. 9i,j,k). Combined, these data suggest that the levels of chromatin-associated RNAs reflect enhancer activities in activating gene expression, which can be used to differentiate super from typical enhancers.

Inferring enhancer-promoter connectivity

One of the fundamental problems in regulated gene expression is to pair enhancers and promoters. In fact, published ChIA-PET and Hi-C experiments have already indicated that enhancers may reach out to promoters that are far away in terms of linear DNA distance^{24,25}. However, numerous static long-distance DNA-DNA interactions tend to obscure true enhancer-promoter interactions in those experiments, as TADs are largely cell type invariant¹¹. We therefore sought to use chromatin-interacting RNAs to infer potential enhancer-promoter relationships. Interestingly, as exemplified on MM.1S cells, RNAs tend to reach out from their sites of transcription to a medium distance of ~1 Mb (Fig. 5a).

As enhancers tend to attract RNA in general, we used *trans*-chromosomal interaction signals of protein-coding RNAs to deduce a statistical model for *trans*-chromosomal interacting RNAs on enhancers, and by using a highly stringent criterion of $z \geq 3$, we identified RNA-chromatin interactions that likely reflect the physical proximity between gene loci and enhancers (Extended Data Fig. 10a). This analysis clearly suggests that enhancers may reach out to

promoters significantly beyond the traditional confinement of 50 Kb, but interestingly, typical and super-enhancers seem to have a similar action range ([Extended Data Fig. 10b,c](#)).

We next visualized GRID-seq defined enhancer-promoter network with Cytoscape by using self-organized layout ²⁶, as illustrated on Chromosome 1 and with several enlarged views in MM.1S cells ([Fig. 5b, Supplementary Table 2](#)). Based on this network, we found that each enhancer, whether typical or super, seemed to control 1 or 2 genes in most cases ([Fig. 5c](#)). In contrast, each gene seemed to be regulated by multiple typical enhancers, but a given gene is only associated with 1 or 2 super-enhancers ([Fig. 5d](#)). Because these observations may be interpreted to indicate that super-enhancers are many individual enhancers stitched together, our finding at least suggests that super-enhancers do not control more genes than typical enhancers.

Functional evidence for connectivity

Because chromatin-interacting RNAs enabled the estimation of long-distance enhancer-promoter connectivity significantly beyond the traditional framework, we next wished to seek functional evidence for newly predicated enhancer-promoter partnerships. On a specific example in MM.1S cells, RNAs from two transcribing genes (*SNX5* and *RPBPI*) were interacting with one super-enhancer and six typical enhancers ([Extended data Fig. 10d](#)). In response to JQ1 treatment, both genes were down regulated ([Extended data Fig. 10e](#)) and the super-enhancer showed more reduced BRD4 binding than all other typical enhancers ([Extended Data Fig. 10f](#)). We next extended the analysis to all RNA connected enhancers and promoters in MM.1S cells, asking whether genes associated with at least one super-enhancers (plus typical enhancers) might be more sensitive to perturbation of enhancer activities by JQ1 than those only linked to typical enhancers. We found that genes associated with at least one super-enhancer

were indeed more responsive to JQ1 treatment than those only linked to typical enhancers based on GRID-seq signals within the traditional 50 Kb range ([Fig. 5e](#)), and importantly, we made the same observation based on the connectivity without setting any action range ([Fig. 5f](#)). These data suggest that chromatin-interacting RNAs may indeed be used to deduce long-distance enhancer-promoter interactions and that super-enhancers are superior in activating both local and distal genes.

Our analysis has been focused on *cis*-chromosomal interactions, but numerous RNAs were still able to reach out to loci in other chromosomes with sufficient significance based on our *trans*-chromosomal interaction model, implying the formation of enhancer-promoter hubs even between chromosomes in 3D genome. To visualize such interactions, we then displayed the network of all deduced enhancer-promoter interactions in MM.1S cells with Cytoscape by using self-organized layout. Strikingly, the resulting global network revealed that individual chromosome tended to cluster into multiple intra-chromosomal hubs, each centered by gene(s) expressing chromatin-associated RNAs, many of which were also connected with other chromosomes ([Fig. 5g](#)). These findings suggest that chromatin-interacting RNAs may be used to construct a functional 3D map that resembles the chromosomal organization in the nucleus.

Discussion

We present a method for unbiased global detection and analysis of RNA-chromatin interactions, which will be a powerful tool for studying regulatory RNAs on chromatin. To infer specific RNA-chromatin interactions, we established a strategy to set the background for non-specific RNA-chromatin interactions by mixing human and *Drosophila* nuclei. Interestingly, such true background is comparable to that built with endogenous RNAs, thus permitting the use

of endogenous RNAs to construct background in human cells. Because those “non-specific” RNAs tend to search for open chromatin to interact, it is important to keep in mind that such background may obscure some specific RNA-chromatin interactions. In any case, a major surprise is relatively few RNAs capable of engaging in *trans*-chromosomal interactions in the human genome, *MALAT1* and *NEAT1* being the major lncRNAs as reported in literature. Interestingly, unlike human cells, a large number of snoRNAs appear to participate in chromatin interactions in *Drosophila* S2 cells, raising an intriguing possibility that various snoRNAs may have important roles at the chromatin levels in *Drosophila*. The identification of many unannotated chromatin-interacting transcripts provides rich resources for future functional studies.

Besides *trans*-chromosomal chromatin-interacting RNAs, we detected a large set of chromatin-interacting RNAs that are not necessarily from the most highly expressed genes in the cell. Strikingly, many of these RNAs are able to reach out to chromatin that are megabases away in linear DNA distance, and in some extreme cases, specific RNAs can decorate the entire chromosome arm or the full chromosome, which has only one precedent *XIST* in human and one precedent *roX1/2* in *Drosophila* cells, both involved in X-chromosome dosage compensation^{18,27}. These findings beg yet another intriguing question on whether some RNAs are involved in dosage compensation or other large-scale regulatory activities in various autosomes.

Strikingly, the majority of specific chromatin-interacting RNAs are associated with enhancers marked by H3K4me1/2 and H3K27ac. Recent studies suggest that a subset of enhancers may be considered super-enhancers because of their clustered distribution in mammalian genomes^{12,13,21}. Interestingly, about half of all chromatin-interacting RNAs detected by GRID-seq in two mammalian cell types decorated on those super-enhancers,

therefore providing an independent measure for enhancer activities. Based on functional perturbation of enhancer activities, previous studies also indicate that super-enhancers appear to be more potent than typical enhancers in enhancing gene expression and more sensitive to inhibition of BRD4²¹. Our analysis based on GRID-seq signals confirmed the previous conclusions and further revealed that super-enhancers are similar to typical enhancers in action range.

We took advantage of chromatin-interacting RNAs to address an outstanding question on enhancer-promoter connectivity, which has been inferred in various cases by long-distance DNA-DNA interaction data, especially those enriched with critical transcription factors, such as RNA Pol II^{10,28}, or indirectly deduced via enhancer-promoter units (EPUs) based on their co-regulation patterns in different tissues²⁹. Functional studies also validated some of those inferred relationships in various biological contexts^{30,31}. Importantly, our data enable us to infer global enhancer-promoter connectivity without constraint of the traditional boundary of ~50 Kb, which is in fact in line with examples based on physical DNA-DNA interactions^{25,32}. We envision that such “long-distance” interactions are actually quite local in the 3D space of the nucleus, and therefore, enhancers, particularly super-enhancers, may share hubs with gene promoters they regulate.

Thus far, we have been only using chromatin-interacting RNAs to relate to various gene activities, but an important question to be addressed in future studies is whether those RNAs may actually play more active roles in nucleating enhancer-promoter hubs. Furthermore, while many of those chromatin-interacting RNAs are well-characterized lncRNAs, the majority are actually pre-mRNAs transcribed from typical protein-coding genes. This implies that many such pre-mRNAs may function as lncRNAs in nucleus. In fact, increasing evidence suggests the

functional importance of nascent RNAs from both pre-mRNAs and lncRNAs in mediating a range of regulatory activities on chromatin, as exemplified by the recruitment of a *de novo* DNA methyltransferase³³, transcriptional activators^{34,35}, or repressors³⁶⁻³⁸. The GRID-seq technology is therefore expected to expedite the discovery of a variety of RNA-mediated regulatory activities on chromatin.

Acknowledgements

We wish to express our gratitude to Xiong Ji and Richard Young for sending us MM.1S cells; Cheng-Ming Chiang for JQ1 inhibitor; Yuanchao Xue for advice during technology development, and members of the Fu lab for stimulating discussion 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-seq experiments; X.L., B.Z., and X.-D.F. wrote the manuscript.

Author Information:

High-throughput data are deposited in Gene Expression Omnibus under accession number GSEXXX for all GRID-seq and GRO-seq experiments.

Reference:

- 1 Djebali, S. *et al.* Landscape of transcription in human cells. *Nature* **489**, 101-108, doi:10.1038/nature11233 (2012).
- 2 Rinn, J. L. & Chang, H. Y. Genome regulation by long noncoding RNAs. *Annu Rev Biochem* **81**, 145-166, doi:10.1146/annurev-biochem-051410-092902 (2012).
- 3 Fu, X. D. Non-coding RNA: a new frontier in regulatory biology. *Natl Sci Rev* **1**, 190-204, doi:10.1093/nsr/nwu008 (2014).
- 4 West, J. A. *et al.* The long noncoding RNAs NEAT1 and MALAT1 bind active chromatin sites. *Mol Cell* **55**, 791-802, doi:10.1016/j.molcel.2014.07.012 (2014).
- 5 Chu, C., Qu, K., Zhong, F. L., Artandi, S. E. & Chang, H. Y. Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. *Mol Cell* **44**, 667-678, doi:10.1016/j.molcel.2011.08.027 (2011).
- 6 Simon, M. D. *et al.* The genomic binding sites of a noncoding RNA. *Proc Natl Acad Sci U S A* **108**, 20497-20502, doi:10.1073/pnas.1113536108 (2011).
- 7 Engreitz, J. M. *et al.* The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science* **341**, 1237973, doi:10.1126/science.1237973 (2013).
- 8 Chepelev, I., Wei, G., Wangsa, D., Tang, Q. & Zhao, K. Characterization of genome-wide enhancer-promoter interactions reveals co-expression of interacting genes and modes of higher order chromatin organization. *Cell Res* **22**, 490-503, doi:10.1038/cr.2012.15 (2012).
- 9 Zhang, Y. *et al.* Chromatin connectivity maps reveal dynamic promoter-enhancer long-range associations. *Nature* **504**, 306-310, doi:10.1038/nature12716 (2013).
- 10 Li, G. *et al.* Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell* **148**, 84-98, doi:10.1016/j.cell.2011.12.014 (2012).
- 11 Jin, F. *et al.* A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature* **503**, 290-294, doi:10.1038/nature12644 (2013).
- 12 Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934-947, doi:10.1016/j.cell.2013.09.053 (2013).
- 13 Whyte, W. A. *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307-319, doi:10.1016/j.cell.2013.03.035 (2013).
- 14 Pott, S. & Lieb, J. D. What are super-enhancers? *Nat Genet* **47**, 8-12, doi:10.1038/ng.3167 (2015).
- 15 Shav-Tal, Y. *et al.* Dynamics of single mRNPs in nuclei of living cells. *Science* **304**, 1797-1800, doi:10.1126/science.1099754 (2004).
- 16 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).

- 17 Flynn, R. A. *et al.* 7SK-BAF axis controls pervasive transcription at enhancers. *Nat Struct Mol Biol* **23**, 231-238, doi:10.1038/nsmb.3176 (2016).
- 18 Gelbart, M. E. & Kuroda, M. I. Drosophila dosage compensation: a complex voyage to the X chromosome. *Development* **136**, 1399-1410, doi:10.1242/dev.029645 (2009).
- 19 Alekseyenko, A. A. *et al.* A sequence motif within chromatin entry sites directs MSL establishment on the Drosophila X chromosome. *Cell* **134**, 599-609, doi:10.1016/j.cell.2008.06.033 (2008).
- 20 Krzywinski, M. *et al.* Circos: an information aesthetic for comparative genomics. *Genome Res* **19**, 1639-1645, doi:10.1101/gr.092759.109 (2009).
- 21 Loven, J. *et al.* Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* **153**, 320-334, doi:10.1016/j.cell.2013.03.036 (2013).
- 22 Heintzman, N. D. *et al.* Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**, 108-112, doi:10.1038/nature07829 (2009).
- 23 Wang, D. *et al.* Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. *Nature* **474**, 390-394, doi:10.1038/nature10006 (2011).
- 24 Fullwood, M. J. *et al.* An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature* **462**, 58-64, doi:10.1038/nature08497 (2009).
- 25 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).
- 26 Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498-2504, doi:10.1101/gr.1239303 (2003).
- 27 Payer, B. & Lee, J. T. X chromosome dosage compensation: how mammals keep the balance. *Annu Rev Genet* **42**, 733-772, doi:10.1146/annurev.genet.42.110807.091711 (2008).
- 28 Mercer, T. R. *et al.* DNase I-hypersensitive exons colocalize with promoters and distal regulatory elements. *Nat Genet* **45**, 852-859, doi:10.1038/ng.2677 (2013).
- 29 Shen, Y. *et al.* A map of the cis-regulatory sequences in the mouse genome. *Nature* **488**, 116-120, doi:10.1038/nature11243 (2012).
- 30 Ji, X. *et al.* 3D Chromosome Regulatory Landscape of Human Pluripotent Cells. *Cell Stem Cell* **18**, 262-275, doi:10.1016/j.stem.2015.11.007 (2016).
- 31 Lupianez, D. G. *et al.* Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* **161**, 1012-1025, doi:10.1016/j.cell.2015.04.004 (2015).
- 32 Tang, Z. *et al.* CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription. *Cell* **163**, 1611-1627, doi:10.1016/j.cell.2015.11.024 (2015).
- 33 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, doi:10.1101/gad.590910 (2010).

- 34 Sigova, A. A. *et al.* Transcription factor trapping by RNA in gene regulatory elements. *Science* **350**, 978-981, doi:10.1126/science.aad3346 (2015).
- 35 Lai, F. *et al.* Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. *Nature* **494**, 497-501, doi:10.1038/nature11884 (2013).
- 36 McHugh, C. A. *et al.* The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature* **521**, 232-236, doi:10.1038/nature14443 (2015).
- 37 Yin, Y. *et al.* Opposing Roles for the lncRNA Haunt and Its Genomic Locus in Regulating HOXA Gene Activation during Embryonic Stem Cell Differentiation. *Cell Stem Cell* **16**, 504-516, doi:10.1016/j.stem.2015.03.007 (2015).
- 38 Wei, C. *et al.* RBFox2 Binds Nascent RNA to Globally Regulate Polycomb Complex 2 Targeting in Mammalian Genomes. *Mol Cell*, doi:10.1016/j.molcel.2016.04.013 (2016).
- 39 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, doi:10.1016/j.cell.2015.01.015 (2015).

Extended Methods

Cell culture

MDA-MB-231 breast cancer cells (HTB-26 ATCC) and MM.1S multiple myeloma cells were grown at 37°C and 5% CO₂. MDA-MB-231 cells were cultured in Dulbecco's Modified Eagle Medium (Thermo Fisher Sci.) supplemented with 10% fetal bovine serum. MM.1S cells were cultured in RPMI-1640 supplemented with 1% GlutaMAX (Thermo Fisher Sci.) and 10% fetal bovine serum. For JQ1 treatment, MDA-MB-231 cells were resuspended in fresh media containing 500 nM JQ1 (a gift from Cheng-Ming Chiang, UT Southwestern) or 0.05% DMSO as vehicle for a duration of 6 h. *Drosophila* S2 cells were cultured in Schneider's *Drosophila* Medium (Thermo Fisher Sci.) supplemented with 10% fetal bovine serum and 2 mM L-glutamine (Thermo Fisher Sci.), at room temperature in ambient CO₂.

Construction of GRID-seq library

A bivalent linker was chemically synthesized (IDT), as illustrated in Extended Data Fig.1. The DNA strand consists of 5'-/5Phos/GTTGGAGTTCGGTGTGAGCTGTGTC -3', and the DNA/RNA hybrid strand contains 5'-/5Phos/rGrUrUrGrGrArUrUrCrNrNrNrGrACACAGC/iBiodT/CACTCCCCACACACCGAACTC CAAC -3' (rN: Random ribonucleotide; iBiodT: biotin-conjugated T). The DNA/RNA hybrid stand was pre-adenylated by using the DNA 5' Adenylation Kit (NEB) following manufacturer instruction, and was purified by Phenol:Chloroform:Isoamyl Alcohol (pH 8.0, Thermo Fisher Sci.) followed by ethanol precipitation. Equal molar quantity of the two stands were mixed, incubated at 80°C for 5 min, and annealed after slow cooling to room temperature at

approximately 0.1°C per sec. The annealed linker was adjusted to the final concentration of 8 pmol/μl.

Approximately 2 million mammalian cells or 10 million *Drosophila* cells were used for each GRID-seq library construction. Cells were washed twice with PBS and crosslinked for 45 min at room temperature with 2 mM PBS-diluted DSG solution. Cells were washed and further crosslinked for 10 min at room temperature with 3% PBS-diluted formaldehyde solution 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% NP-40, 1 U/μL RiboLock (Thermo Fisher Sci.), 1× Protease inhibitor (Sigma-Aldrich.)) for 15 min on ice. To prepare nuclei, fixed cells were washed in 200 μl of 1× Tango Buffer (Thermo Fisher Sci.) and then incubated in 320 μl of Buffer B (1× Tango Buffer, 0.2% SDS) for 10 min at 62°C. SDS was 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× Tango Buffer, resuspended in 500 μl of AluI solution (1× Tango Buffer, 1 U/μl RiboLock, 1× Protease inhibitor, 1% Triton X-100, 0.5 U/μl AluI (Thermo Fisher Sci.)), and incubated at 37°C for 2 h with agitation. Nuclei were collected, resuspended in 400 μl of PNK solution (1× Tango Buffer, 1 U/μl RiboLock, 1× Protease inhibitor, 1 mM ATP, 0.35 U/μl T4 PNK (Thermo Fisher Sci.)), 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× RNA Ligase Buffer (NEB), resuspended in 500 μl of RNA ligation solution (1× 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 H₂O, 36 μl of 1 M KCl, 32 μl of 10 mM dNTP mix, 28 μl of 5× RT First Strand Buffer (Thermo

Fisher Sci.), 28 μ l of 100 mM DTT and 5 μ l of SuperScript III Reverse Transcriptase were mixed directly into the suspension, and the reaction was incubated at 50°C for 45 min. For *in situ* linker ligation to AluI-cut genomic DNA, nuclei were collected, washed twice with 200 μ l of 1 \times DNA Ligase Buffer (NEB) to remove free linker, resuspended in 1.2 ml of DNA Ligation Solution (0.2 U/ μ l RiboLock, 1 \times DNA Ligase Buffer, 1 mg/ml BSA, 1% Triton X-100, 1 U/ μ l T4 DNA Ligase (Thermo Fisher Sci.)) and incubated overnight at 16°C with rotation.

Nuclei were collected, washed with PBS, resuspended in 266 μ l 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 Sci.) and incubated at 65°C for 30 min. After adding 20 μ l of 5 M NaCl, protease-treated nuclei were incubated for another 1.5 h. Total DNA was extracted, dissolved in B&W Buffer (5 mM Tris-Cl pH 7.5, 1 M NaCl, 0.5 mM EDTA, 0.02% Tween-20). 300 μ g of Streptavidin-conjugated magnetic beads were washed with B&W Buffer and mixed with purified DNA for affinity purification of biotinylated linker. After incubation at 37°C for 30 min, beads were extensively washed for 5 times with B&W Buffer, and incubated in 100 μ l of 150 mM NaOH at room temperature for 10 min. Cleared supernatant was collected, neutralized with 6.5 μ l of 1.25 M Acetic Acid, and diluted with 11 μ l of 10 \times TE Buffer (100 mM Tris-Cl pH 7.5, 10 mM EDTA). Released ssDNA was precipitated by isopropanol from the supernatant and dissolved in 30 μ l of H₂O. Second strand synthesis was performed by mixing ssDNA with 250 ng Random Hexamer Primers and 5 μ l of 10 \times NEB Buffer CutSmart, incubating at 98°C for 5 min, adding 8.5 μ l of H₂O, 5 pmol dNTP and 5 U Klenow Fragment (3' to 5' exo-) enzyme (NEB) and incubating at 37°C for 1 h. After heat inactivation at 70°C for 10 min, 5 pmol Sadenosylmethionine (NEB) and 1 U MmeI enzyme (NEB) was added and incubated at 37°C for 30 min followed by addition of another 3 U MmeI and incubation for 30 min. The reaction was

treated with 40 µg Proteinase K at 65°C for 20 min. Digested DNA was extracted and purified before loading to 12% native PAGE gel for size-selection. The desired band at 84 bp was excised and DNA was extracted.

Adapters were prepared by annealing the following two oligonucleotides (IDT) in 1× NEB Buffer 2 to a final concentration of 25 mM: 5'-/5Phos/AGATCGGAAGAGCACACGTCT-3' and 5'-ACACTCTTCCCTACACGACGCTCTCCGATCTNN-3'. Purified DNA was dissolved in 10 µl of 1× NEB Buffer CutSmart and 0.5 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 36 µl H₂O, mixed with 10 µl of 10× T4 DNA Ligase Buffer (NEB), 32 µl of PEG-6000, 200 pmol of 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 Polynucleotide Kinase (NEB) supplemented by 100 pmol ATP and incubated at 37°C for 30 min. Nick was then ligated by addition of 1 µl of 10× T4 DNA Ligase Buffer, 100 pmol ATP and 1,600 U T4 DNA Ligase (NEB), and incubation at room temperature for 30 min. DNA along with excessive Adapters were extracted and purified before loading to 10% native PAGE gel for size-selection. The desired band at approximately 185 bp was excised. DNA was extracted and dissolved in 20 µl of H₂O. 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 Sci.)) was prepared. PCR primers consist of Primer#1 (5'-AATGATAACGGCGACCACCGAGATCTACACBBBBBACACTCTTCCCTACACGACGCTCTCCGATCT -3' (BBBBB: 5 nt barcode for multiplexing libraries)) and Primer#2 (5'-CAAGCAGAAGACGGCATACGAGACGTGTGCTCTCCGATCT -3'). PCR was performed

with an initial 30 sec denaturation at 98°C, followed by 16 cycles of 10 sec denaturation at 98°C, 30 sec annealing at 65°C and 15 sec extension at 72°C. The PCR product was separated on a native 10% PAGE gel and the band at 188 bp was recovered. DNA was subsequently subjected to Illumina HiSeq 2500 for single-end 100 bp sequencing, with the sequencing primer (5'-ACACTTTCCCTACACGACGCTTCCGATCT -3').

Parallel analysis of Human-*Drosophila* mixed nuclei

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. 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. The construction of the mix library was performed in parallel with 2 million MDA-MB-231 cell nuclei and 10 million S2 cell nuclei.

GRID-seq raw data processing and mapping

Upon sequencing, individual libraries were segregated according to multiplexing barcodes and then both barcode and residual adapter sequences were removed from each tag to produce tags with non-uniform length, the majority of which ranged from 84 bp to 87 bp in each library. To precisely remove linker sequence from the DNA and cDNA tags, MmeI motifs were used for defining linker boundaries. Linker orientation also dictated whether the tags at each side were originated from genomic DNA or RNA. To minimize the loss of tags due to sequencing errors, tags were first filtered based on the presence of two opposite-orientated MmeI motifs, then aligned to the linker sequence from both directions to determine its orientation. DNA and

corresponding RNA tags, most of which ranged from 18 bp to 23 bp, were extracted for alignment.

All processed tags were aligned to their indicated genome build using Bowtie2 with parameter of --local⁴⁰. Human samples were aligned to genome build hg38 and *Drosophila* samples to genome build dm3 with parameters -q 2. To estimate the numbers of cross-species-ligated RNA and DNA tags in the mix of MDA-MB-231 and S2 cell nuclei, RNA tags were first aligned independently to the transcriptome builds of hg38 and dm3, with the most stringent parameter of -q 44. Linked DNA tags of those RNA tags that were unambiguously aligned to human transcriptome were then aligned to human genome with the parameter of -q 2. These DNA tags failed to align to the human genome were then aligned to the *Drosophila* genome, with the parameter of -q 2. Conversely, DNA tags linked to the RNA tags that were unambiguously aligned (-q 44) to the *Drosophila* transcriptome were first aligned to the *Drosophila* genome (-q 2), and those unaligned DNA tags were then aligned to the human genome (-q 2)

Seven GRID-seq libraries were generated from the current study:

Species	Samples	Raw tags	Linker-clipped tags	Uniquely mapped tag pairs (% of alignment filtering)
<i>Drosophila</i>	S2 Rep1	113.7 M	95.7 M	4.5 M (4.7%)
<i>Drosophila</i>	S2 Rep2	165.7 M	74.0 M	2.0 M (2.7%)
Mixed	MDA-MB-231 + S2	46.2 M	38.7 M	--
Human	MDA-MB-231 Rep1	213.0 M	168.0 M	44.0 M (26.2%)
Human	MDA-MB-231 Rep2	294.9 M	144.8 M	40.3 M (27.8%)
Human	MM.1S Rep1	300.8 M	135.0 M	37.7 M (27.9%)
Human	MM.1S Rep2	363.6 M	149.6 M	39.2 M (26.2%)

Identification of hit RNAs

Genomic regions with enriched GRID-seq RNA reads were detected by MACS2⁴¹ using the model for broad-peak detection. These mapped regions with significant enrichment ($p < 0.001$) and overlapping with known-gene annotation (Ensemble genes GRCh38.83 for human and BDGP5.78 for *Drosophila*) were assigned to their respective largest annotated genes. Enriched regions that did not overlap with any known gene were assigned as “unannotated transcripts”. The RNA read-coverage of genomic regions was calculated by BEDtools⁴² and SAMtools⁴³. The genes with the tag-coverage above the sliding-window threshold [$(N_{i+n} - N_i) \geq n$, where i was the rank of given RNA, N was the read-counts of this RNA and n was the 1/100 of the total number of ranked RNAs] were selected as abundant chromatin-interacting RNAs. A subset of abundant RNAs with sufficient RNA read-density [RPK (reads per Kb) ≥ 100] or with significant DNA read-densities (RPK ≥ 10) associated at any genomic region was identified as hit RNAs for further analysis (Supplementary Table 1).

Construction of background for non-specific RNA-chromatin interactions

To determine the specific RNA-chromatin interaction pattern of each hit RNA, we developed a genome-wide background for non-specific interactions. Ideally, for each library, we could setup mixed nuclei from different species in order to evaluate cross-species RNA-chromatin interactions, which would represent the true background. In practice, it is feasible to generate such background with sufficient density on the *Drosophila* genome by using RNA reads from human cells because of the much smaller *Drosophila* genome, but the RNA density on the human genome based on the RNA reads from *Drosophila* is often too scattered to provide a reliable background. We thus sought to deduce the background in the *Drosophila* genome based on endogenous RNA reads in comparison with the true background based on cross-species RNA-

chromatin interactions. Toward this goal, we selected RNAs from protein-coding genes engaged in *trans*-chromosomal interactions, and calculated their density on the 1 Kb-binned genome. Such deduced background read-density at each 1 Kb bin was then smoothed by a moving widow of 100 Kb and normalized by the total read number and chromosome size. The resulting background in the *Drosophila* genome was highly correlated with the cross-species background. This strategy enabled us to deduce the background on any human cells by using endogenous RNA reads.

Normalization of RNA-chromatin interactions and construction of genomic binding matrix

To evaluate specific RNA-chromatin interactions of each hit RNA at each genomic bin, we first summarized the DNA reads for each gene at each genomic bin, and then normalized them to 1 million reads followed by further normalization of total number of bins of each chromosome. A ratio was next calculated by dividing the normalized DNA read density with background read density, which represents specific binding of this RNA at the genomic bin.

To construct a robust global binding matrix for all hit RNAs, genomic bins with significant binding level (at least 3 bins with binding level ≥ 2 in every 10 bin-window) were preserved and further smoothed by a moving-widow of 10 bins. All subsequent analyses involving binding levels and patterns were based on such genomic binding matrix.

Identification of active enhancers and inference of enhancer–promoter connectivity

Active enhancers were identified by using published H3K27ac ChIP-seq data²¹. Briefly, enriched peaks of H3K27ac were first detected by MACS2; the peaks within 2.5 Kb around

known promoters were removed; and the qualified peaks were stitched together if they were within a 12.5 Kb region.

As specific binding at enhancers indicated specific footprints of hit RNA on the chromatin, we built a statistical model to identify significant long-distance RNA-chromatin interactions between enhancer and promoters based on *trans*-chromosomal interactions from hit RNAs of protein-coding genes. We used this model and selected a stringent confidence threshold (Z -score ≥ 3) to segregate the interaction levels of travelling RNAs across the nuclear space from those likely due to the spatial proximity in 3D genome.

Construction of enhancer-promoter network

The enhancer-promoter network based on significant intra- and inter-chromosomal interactions was constructed with inferred enhancer-promoter pairs at different significance levels ([Supplementary Table 2](#)). The network of the Chromosome 1 in MM.1S cell was built based on RNA-chromatin interactions of all the hit genes in the chromosome at a very stringent significance ($z \geq 3$), and the network of the whole genome was based on significant RNA-chromatin interactions from protein-coding genes with slightly reduced stringency ($z \geq 2$). The network was then imported into Cytoscape (version 3.3), a versatile software for network visualization and analysis, and displayed with a self-organized layout algorithm, Edge-Repulsive Spring-Electric Layout, supported by the third-party app of AllegroLayout ^{[26](#)}.

Functional perturbation of general enhancer activities

MDA-MB-231 cells were treated with the BRD4 inhibitor JQ1 or DMSO for 6 h, and immediately harvested for global nuclear run-on ^{[23,44](#)}. To quantify transcription activities in an

unbiased manner, GRO-seq read densities were initially normalized using total uniquely-mapped read numbers to remove variations between libraries. To minimize the bias introduced by promoter pausing or gene length, only reads that were aligned within the 2 Kb region 1 Kb downstream from the TSS were selected to calculate the transcription activity. For multi-isoform genes, the transcript with highest read counts within the 2 Kb window was selected to represent the gene's transcription activity.

Public datasets analyzed

The data of RNA Pol II ChIP-seq for MDA-MB-231 cells were obtained from European Genome-phenome Archive (<https://www.ebi.ac.uk/ega>). All other datasets were obtained from NCBI Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo/>).

Data type	Cell type	Accession ID
roX2 ChIRP-seq	<i>Drosophila</i> S2	GSM820427, GSM820428
roX2 ChART-seq	<i>Drosophila</i> S2	GSM833475, GSM833476
MLS3-TAP ChIP-seq	<i>Drosophila</i> Clone 8	GSM296247
Total RNA-seq	<i>Drosophila</i> S2	GSM480160
Total RNA-seq	Human MDA-MB-231	GSM929913
H3K27ac ChIP-seq	Human MDA-MB-231	GSM1204474, GSM1204475
H3K4me3 ChIP-seq	Human MDA-MB-231	GSM1204472, GSM1204473
RNA Pol II ChIP-seq	Human MDA-MB-231	EGAN00001343502, EGAN00001343503
H3K27me3 ChIP-seq	Human MDA-MB-231	GSM2058911, GSM2058912
RNA expression array	Human MM.1S (DMSO, 6 h)	GSM1094100, GSM1094101
RNA expression array	Human MM.1S (JQ1, 0.5 mM, 6 h)	GSM1094092, GSM1094093
H3K27ac ChIP-seq	Human MM.1S	GSM894083
RNA Pol II ChIP-seq	Human MM.1S	GSM894086
H3K4me3 ChIP-seq	Human MM.1S	GSM894084

H3K27me3 ChIP-seq	Human MM.1S	GSM1252088
Brd4 ChIP-seq	Human MM.1S (DMSO, 6 h)	GSM1038275
Brd4 ChIP-seq	Human MM.1S (JQ1, 0.5 mM, 6 h)	GSM1038271

Extended references

- 40 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359, doi:10.1038/nmeth.1923 (2012).
- 41 Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137, doi:10.1186/gb-2008-9-9-r137 (2008).
- 42 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842, doi:10.1093/bioinformatics/btq033 (2010).
- 43 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).
- 44 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, doi:10.1126/science.1162228 (2008).

Fig. 1

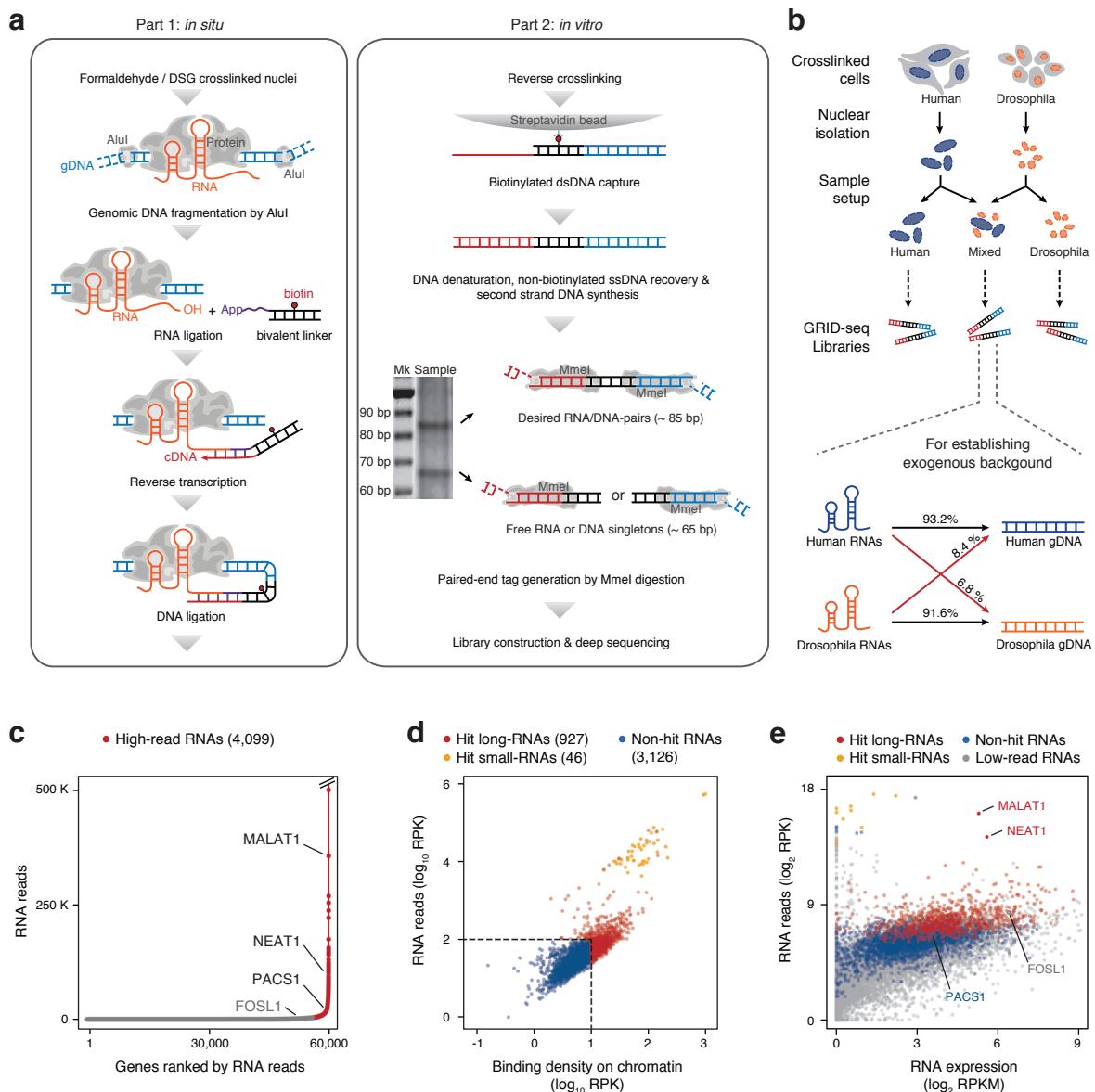


Fig. 1 | Chromatin-interacting RNAs revealed by GRID-seq.

a, Schematic presentation of the GRID-seq technology. Left: steps performed *in situ* on isolated nuclei; Right: steps 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 (upper band) or to either DNA or RNA (lower band). After native gel resolution, the excised upper band was subjected to library construction by adapter ligation followed by deep sequencing. **b**, Top: Scheme for using human MDA-MB-231 cells, *Drosophila* S2 cells, or the mix of both for GRID-seq library construction. Bottom: The percentages of human RNAs ligated to human DNAs or *Drosophila* DNAs and the percentages of *Drosophila* RNAs ligated to *Drosophila* DNAs or human DNAs. **c**, Genes rank-ordered by GRID-seq RNA reads in MDA-MB-231 cells. Red-labeled are genes that passed the cutoff based on the amounts of their chromatin-interacting RNAs. **d**, Scatterplot of length-normalized RNA reads from annotated gene (y-axis) and the read density of the largest peak on DNA (x-axis) in MDA-MB-231 cells. RPK: reads per Kb. Dashed lines are the set thresholds for specific hits with red dots for long RNAs and orange dots for small RNAs. Blue dots are chromatin-interacting RNAs without sufficient density. **e**, Comparison between gene expression detected by RNA-seq (based on the data from GSM929913) and chromatin-interacting RNAs detected by GRID-seq for all human genes in MDA-MB-231 cells. Grey dots are genes not showing frequent interactions with DNA as in **c** and colored genes correspond to those in **d**. Highlighted are one or two representative genes in each class. RPKM: reads per Kb per million mapped reads.

Fig. 2

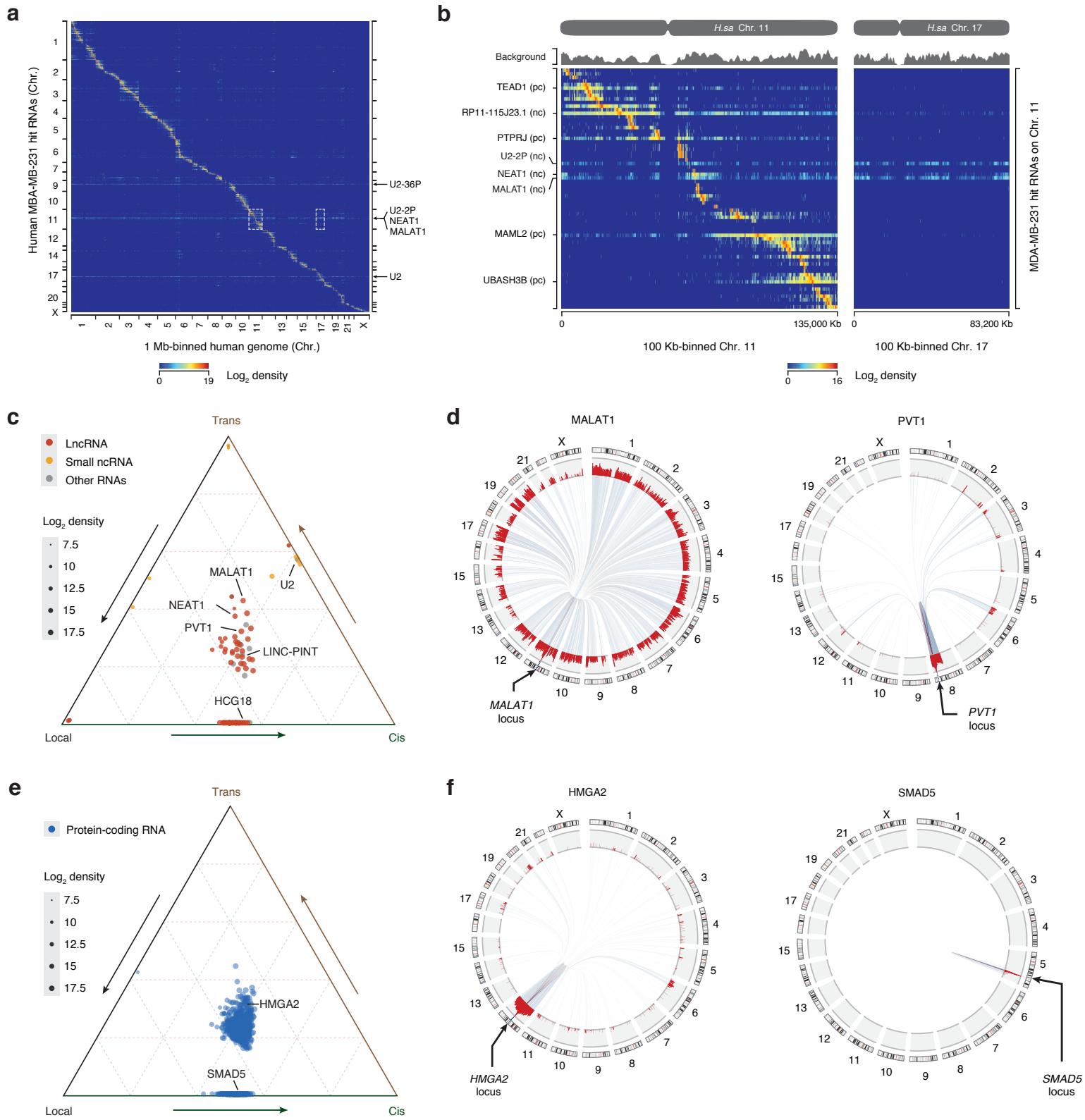


Fig. 2 | Global view of RNA-chromatin interactions in human MDA-MB-231 cells

a, A heatmap showing the interaction of all hit RNAs across the whole human genome in MDA-MB-231 cells. Row: hit RNAs from their origins of transcription. Column: hit RNAs linked to DNA in the 1 Mb-binned human genome. Representative *trans*-chromosomal interacting hit RNAs are labeled on the right. U2-36P and U2-2P are transcripts from pseudo U2 snRNA genes.

b, Two representative regions boxed in **a** are enlarged, showing detailed interaction patterns of hit RNAs from Chromosome 11 on 100 Kb-binned Chromosome 11 (left) and Chromosome 17 (right). Representative hit RNAs are labeled on the left (pc: protein-coding RNAs, nc: non-coding RNAs). Top: The background deduced from endogenous *trans*-chromosomal interacting RNAs from all protein-coding genes. **c**, Ternary plot of non-coding hit RNAs based on their relative chromatin-interaction levels in local (± 1 Kb from their genes), *cis* (the same chromosome the gene resides except local), and *trans* (all other chromosomes except its own chromosome) modes. Colors of dots represent different types of RNAs and sizes represent the levels of chromatin-interacting RNAs. **d**, Circos plots of the interactions of two non-coding hit RNAs *MALAT1* (left) and *PVT1* (right) with chromatin in the human genome. **e**, Ternary plot of protein-coding hit RNAs shown similarly as in **c**. **f**, Circos plots of the interactions of two protein-coding hit RNAs *HMG A2* (left) and *SMAD5* (right) with chromatin in the human genome.

Fig. 3

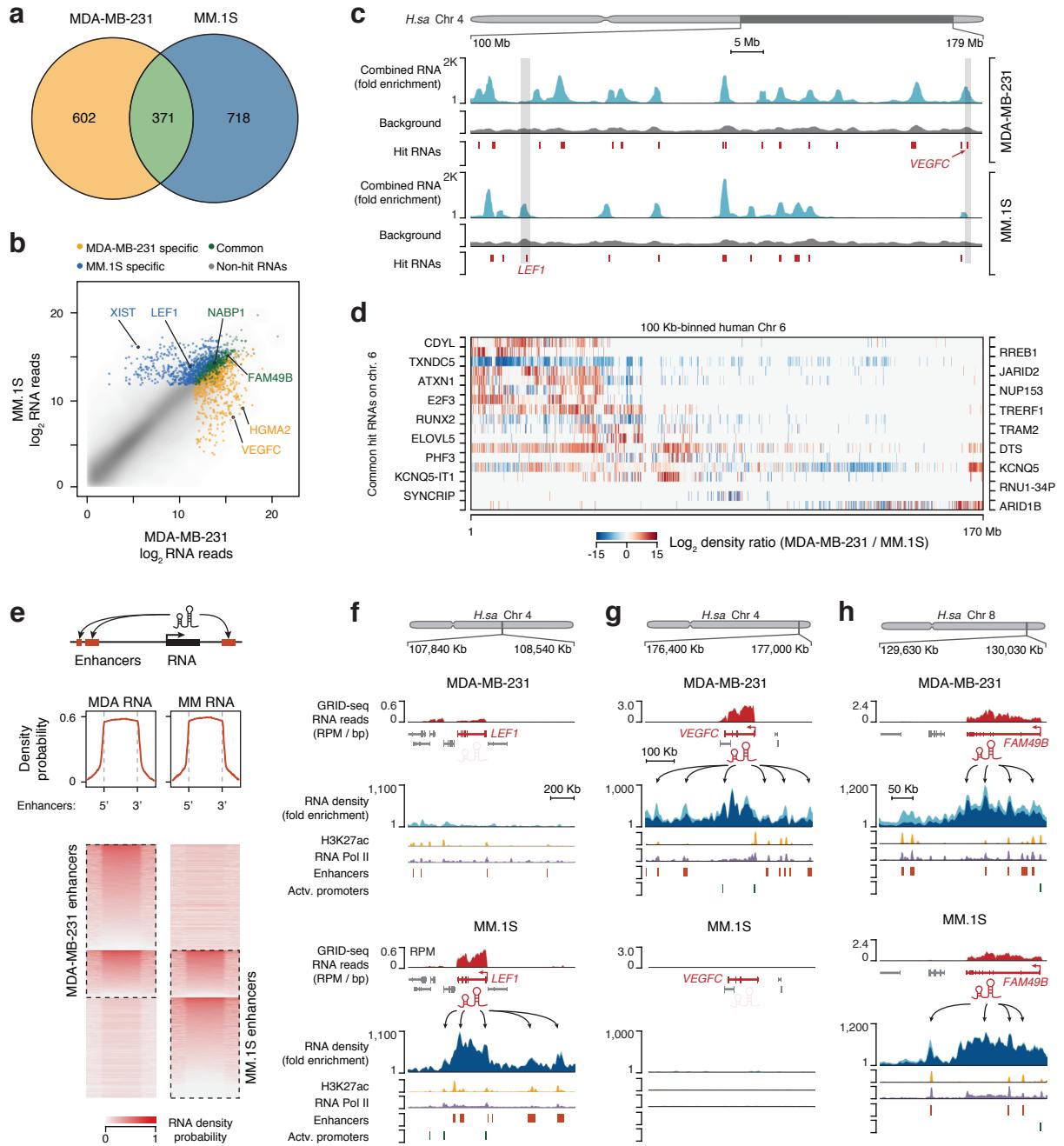


Fig. 3 | Cell type-specific RNA-chromatin interactions in mammalian cells

a, A Venn diagram showing numbers of cell type-specific and common hit RNAs between MDA-MB-231 and MM.1S cells. **b**, Comparison of individual GRID-seq RNA reads between the two cell types. Smoothed scatter plot (grey) represents non-hit RNAs and colored dots represent hit RNAs, a few of which are highlighted. **c**, Comparison of hit RNAs and their chromatin interactions on Chromosome 4 between MDA-MB-231 and MM.1S cells. Two representative cell type-specific regions are shaded and shown with further details in panels **f** and **g**. **d**, A heatmap illustrating differential chromatin interactions of common hit RNAs between the two cell types on Chromosome 6. Blue and Red respectively indicate lower and higher levels of chromatin interactions in MDA-MB-231 cells relative to MM.1S cells. **e**, Top: Meta-analysis demonstrating enrichment of hit RNAs on enhancers in MDA-MB-231 and MM.1S cells. Bottom: heatmaps of enhancers ranked by normalized GRID-seq RNA signals in the two cell types, showing both cell type-specific and common RNA-chromatin interactions on enhancers. **f**, **g**, **h**, Examples showing broad chromatin interactions of hit RNAs from *LEF1* (MM.1S cell-specific), *VEGFC* (MDA-MB-231 cell-specific) and *FAM49B* (common) in comparison with mapped enhancers and promoters. Light blue overlay behind individual RNA binding tracks (dark blue) represents combined signals from all hit RNAs in the regions. RPM: RNA reads per million. Note that the commonly expressed hit RNAs from *FAM49B* showed distinct chromatin interactions between the two cell types. Also note different genomic and chromatin-interaction scales in different panels.

Fig. 4

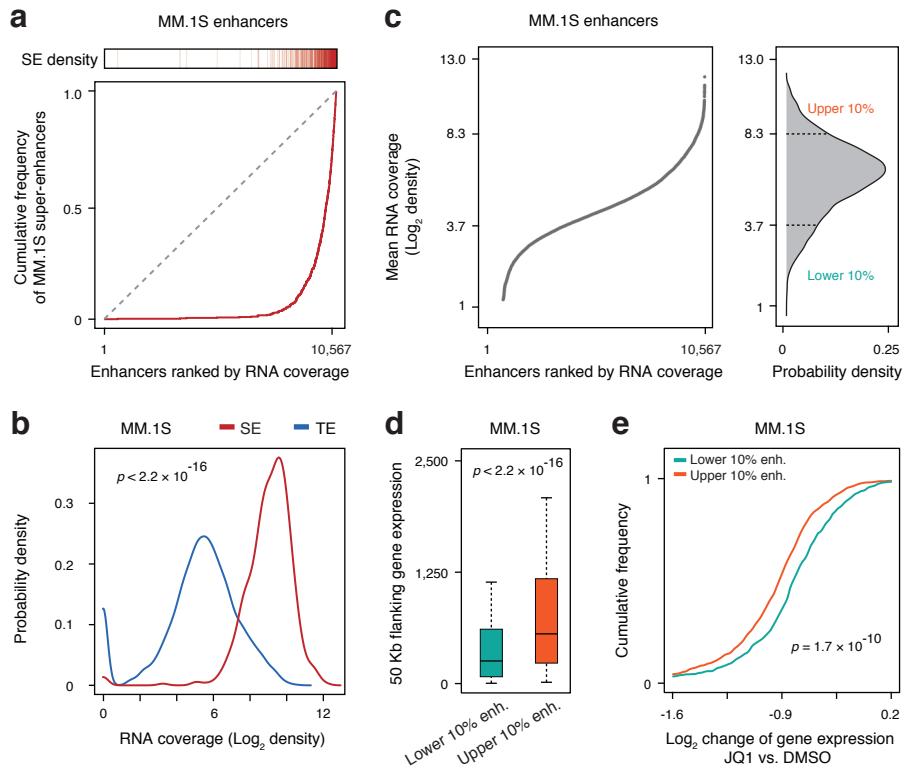


Fig. 4 | Preferential RNA decoration on super-enhancers in MM.1S cells

a, Super-enhancers in relationship to RNA-chromatin interactions signals detected by GRID-seq. 10,567 mapped enhancers in MM.1S cells (based on the mapped H3K27ac signals in the same cell from GSM894083) are ranked by hit RNA density. Each red bar on top represents a super-enhancer. Red line: The cumulative curve of rank-ordered RNA-chromatin interactions; Grey dashed line: Random distribution. **b**, Probability density map of hit RNA coverage on super-enhancers (SE, red) versus typical enhancers (TE, blue). **c**, Left: Rank-ordered RNA-chromatin interaction levels on all active enhancers. Right: Upper (SE enriched) and lower (TE enriched) 10 percentiles of enhancers selected for functional analysis. **d**, Expression of genes associated with top 10% RNA-interacting enhancers (orange box) relative to those associated with bottom 10% RNA-interacting enhancers (green box), both within the ± 50 Kb range (based on the data from GSM1094100 and GSM1094101). **e**, Fold changes in gene expression plotted in the accumulative fashion for the two groups of genes as defined in **c** in response to functional perturbation of enhancers on MM.1S cells by using the BRD4 inhibitor JQ1 (based on the data from GSM1094100, GSM1094101, GSM1094092, and GSM1094093). Statistical significance of comparison is estimated by t-test in panel **b**, **d** and **e**.

Fig. 5

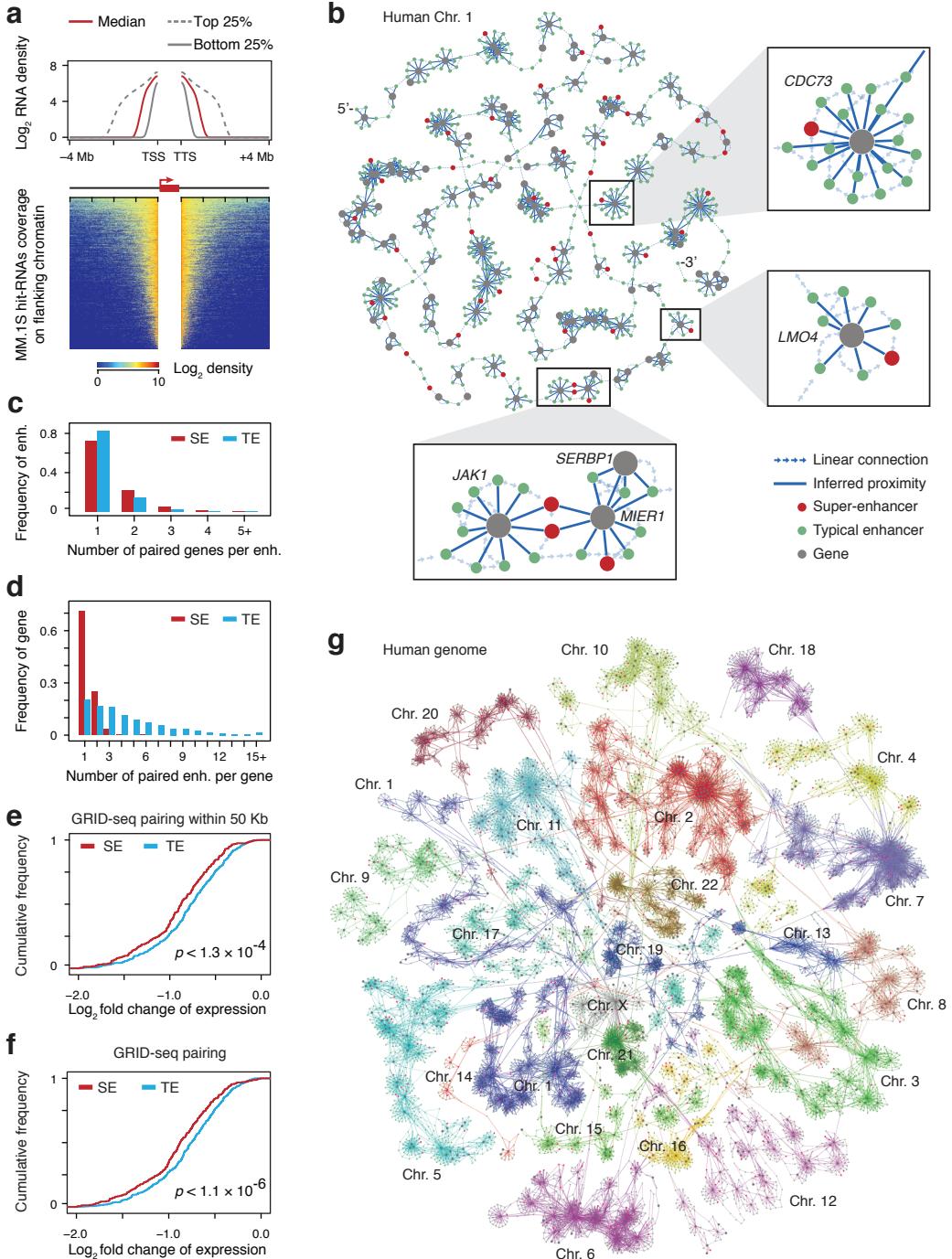
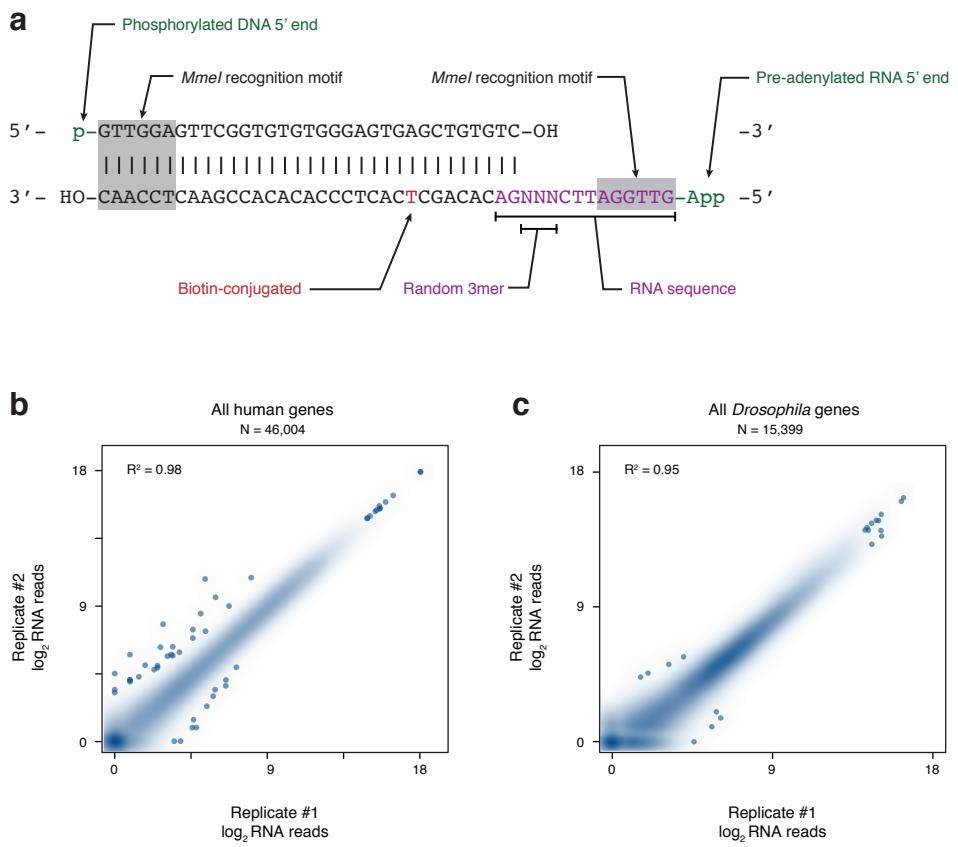


Fig. 5 | Inferring enhancer-promoter connectivity by enhancer-associated RNAs

a, The RNA-chromatin interaction range. Top: Meta-analysis of hit RNAs coverage relative to individual gene bodies. Dashed grey line and black line represent the distributions of top and bottom 25 percentile hit RNAs on chromatin relative to their sites of transcription, respectively. Red line shows the medium range. Bottom: A heatmap of hit RNA coverage on both sides of their respective genes loci. **b**, A representative intra-chromosomal RNA-chromatin interaction map on Chromosome 1 in MM.1S cells with three potential enhancer-promoter hubs highlighted. **c**, The number of genes controlled by typical (blue) or super- (red) enhancers calculated based on inferred enhancer-promoter relationships from network analysis in **b**. **d**, The number of typical (blue) and super-(red) enhancers involved in controlling a given gene calculated based on inferred enhancer-promoter relationships from network analysis in **b**. **e**, **f**, Fold changes in gene expression plotted in the accumulative fashion for the genes associated with typical (blue) and super-(red) enhancers inferred by GRID-seq signals, within the conventional 50 Kb range (**e**) or without setting any range (**f**), in response to functional perturbation of enhancers on MM.1S cells by using the BRD4 inhibitor JQ1 (based on the data from GSM1094100, GSM1094101, GSM1094092 and GSM1094093). P-values were determined by Kolmogorov–Smirnov test. **g**, Cytoscape visualization of global enhancer-promoter interaction networks inferred by all significant *cis*- and *trans*- RNA-chromatin interactions detected by GRID-seq in MM.1S cells.

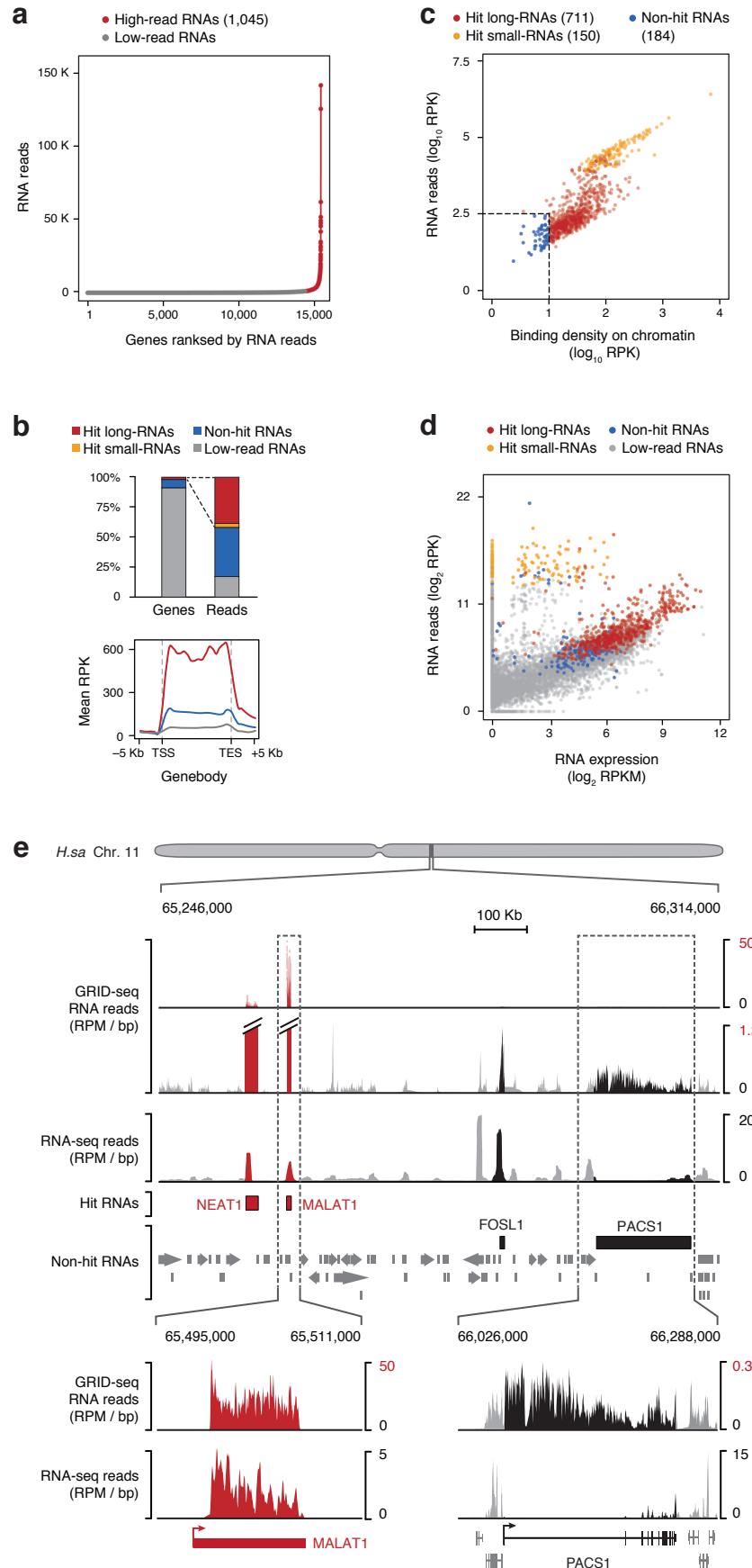
Extended Data Fig. 1



Extended Data Fig. 1 | GRID-seq linker design and reproducibility of RNA capture

a, The design of a bivalent linker for GRID-seq. The top strand is a 5' phosphorylated DNA sequence (black) and the bottom strand consists of both DNA and RNA bases (purple) with a biotinylated T residue (red) in the middle. Randomized bases (N) could serve as barcodes for filtering PCR duplicates generated during library amplification and both ends of the linker also carry the MmeI restriction site (grey-shaded). The linker is pre-adenylated for ligation to RNA in the absence of ATP. **b,c**, Reproducibility of RNAs captured by GRID-seq. RNA reads corresponding to individual annotated genes were plotted to compare between the biological replicates of human MDA-MB-231 cells (**b**) or *Drosophila* S2 cells (**c**).

Extended Data Fig. 2

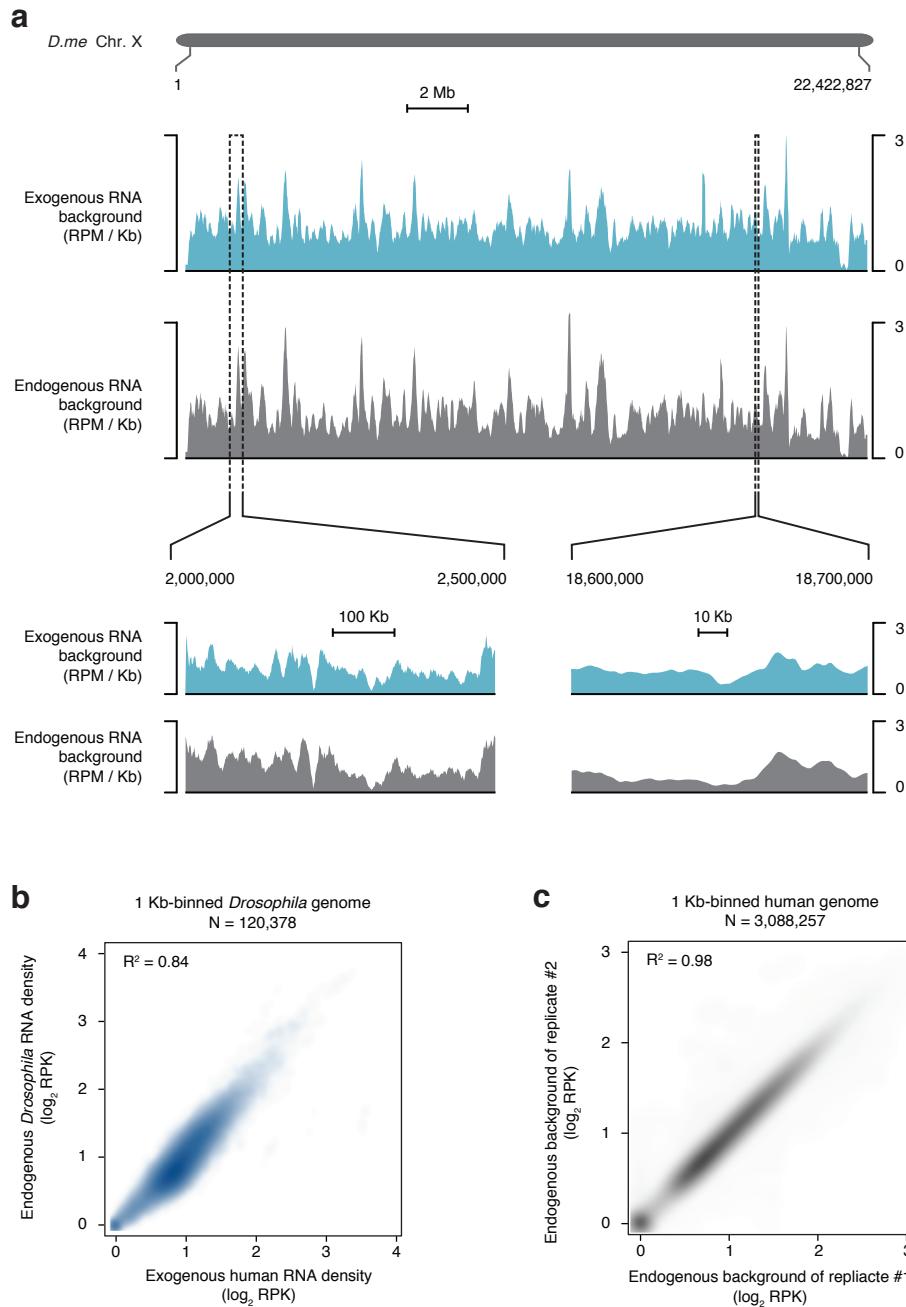


Extended Data Fig. 2 | RNAs captured by GRID-seq in *Drosophila* S2 cells

a, *Drosophila* genes rank-ordered by GRID-seq RNA reads in S2 cells. Red-labeled are genes that passed the inflection. **b**, Top: Percentages of genes that gave to GRID-seq RNA reads in different classes. Bottom: The distribution of GRID-seq detected RNAs along the gene body in MDA-MB-231 cells. Small RNAs were excluded from this analysis due to their short gene body.

c, Scatterplot of length-normalized RNA reads from each annotated gene (y-axis) and the read density of the largest RNA peak on DNA (x-axis) in *Drosophila* S2 cells. RPK: read per Kb. Dashed lines are the set thresholds for specific hits with red dots for long RNAs and orange dots for small RNAs. **d**, Comparison between gene expression detected by RNA-seq (GSM480160) and chromatin-interacting RNAs detected by GRID-seq for all expressed genes in *Drosophila* S2 cells. Grey dots are genes not showing frequently interactions with DNA as in **a**, and colored genes correspond to those in **c**. **e**, Visualization of GRID-seq RNA reads in two scales (first two tracks) in comparison with RNA-seq reads (third track) on human Chromosome 11 in MDA-MB-231 cells. Red-highlighted are hit RNAs from the two long non-coding RNA genes *NEAT1* and *MALAT1*. Black-highlighted are non-hit RNAs from protein-coding *FOSL1* and *PACSI* genes. Grey indicates collective reads from all other non-hit RNAs in this region (fifth track). Note that *FOSL1* was expressed at a higher level than *NEAT1* and *MALAT1* based on the RNA-seq data (third track), indicating that hit RNAs are not necessarily from highly expressed genes. Boxed regions are enlarged in bottom panels to further show the GRID-seq detected RNA signals of hit RNA of *MALAT1* and non-hit RNA of *PACSI* on DNA in comparison with gene expression.

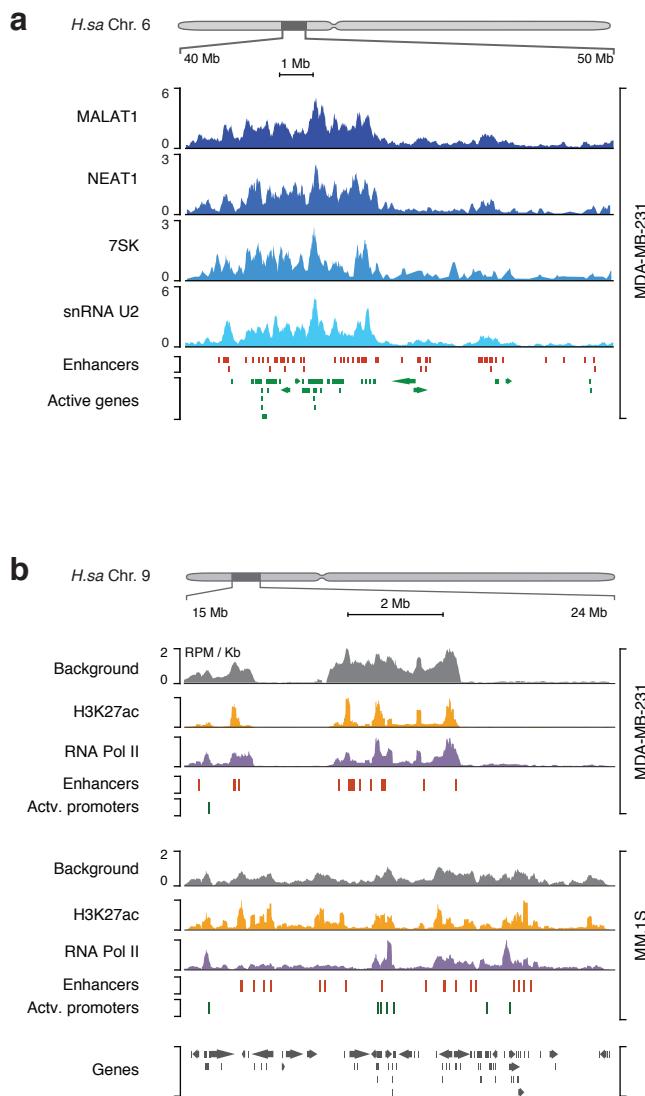
Extended Data Fig. 3



Extended Data Fig. 3 | Background established with exogenous or endogenous RNAs

a, Background deduced with exogenous or endogenous RNAs on *Drosophila* Chromosome X. Top two tracks: The distribution of human RNAs linked to *Drosophila* DNA, which represents the true background (green), and the distribution of collective endogenous *Drosophila* trans-chromosomal interacting RNAs from all protein-coding genes, which corresponds to the deduced background (grey). Two representative genomic regions were enlarged in the bottom tracks. y-axis indicates RNA reads per million. **b**, Comparison between exogenous and endogenous background RNA densities on the 1 Kb-binned *Drosophila* genome. **c**, Comparison between the deduced backgrounds based on endogenous RNAs from two independent GRIP-seq experiments on human MDA-MB-231 cells. RPK: reads per Kb.

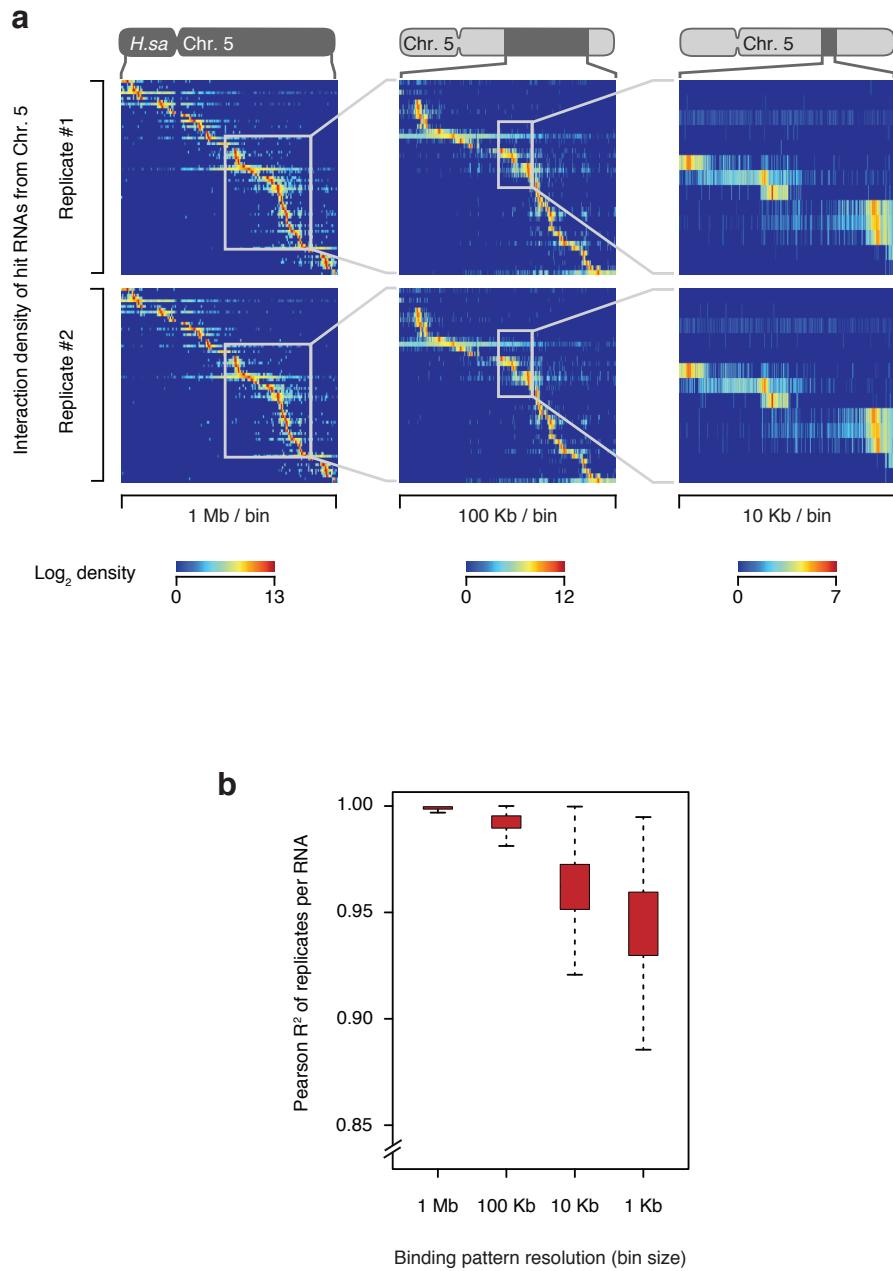
Extended Data Fig. 4



Extended Data Fig. 4 | Examples of hit RNA signals relative to the background

a, Four major *trans*-chromosomal hit RNAs mapped to a representative region on human Chromosome 6 in MDA-MB-231 cells. y-axis indicates reads per million for each RNA. Active enhancers and genes in the region are also shown for comparison. **b**, The background deduced from endogenous *trans*-chromosomal interacting RNAs from all protein-coding genes in comparison with the enhancer mark H3K27ac, RNA Pol II binding, and active enhancers and promoters in MDA-MB-231 (top panel) or MM.1S (bottom panel) cells on a representative region of human Chromosome 9. These data indicate that the general background for RNA-chromatin interactions tends to occur on open chromatin regions.

Extended Data Fig. 5

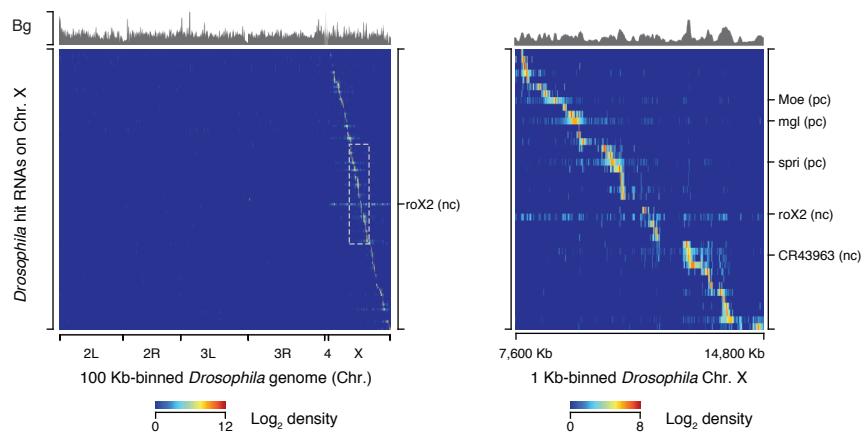


Extended Data Fig. 5 | Gradually expanded views of RNA-chromatin interactions

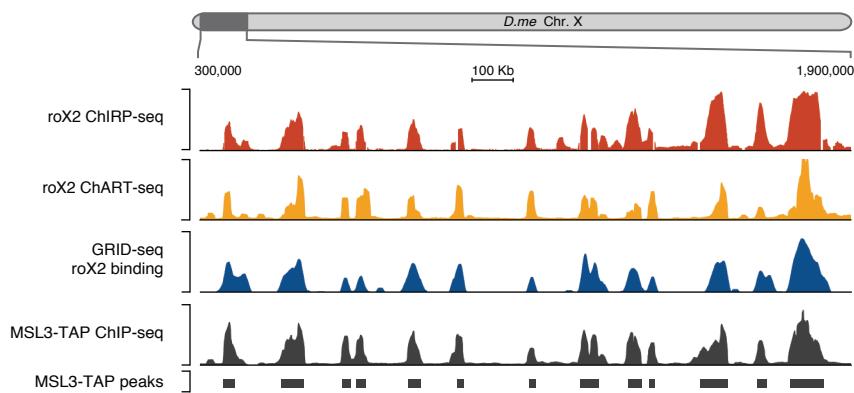
a, RNA-chromatin interaction heatmaps on Chromosome 5 constructed from two independent GRID-seq experiments on MDA-MB-231 cells. Boxed regions in each panel were enlarged with increasing resolution in the next panel on the right. **b**, Correlation of RNA read densities at increasing resolution (decreasing bin size) based on total hit RNAs across the human genome between the replicates performed on MDA-MB-231 cells.

Extended Data Fig. 6

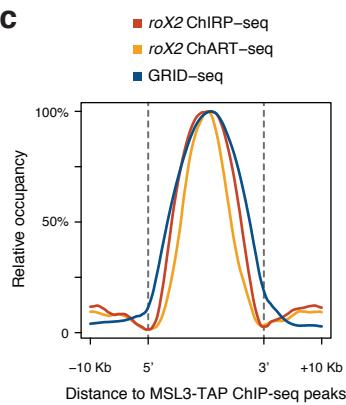
a



b



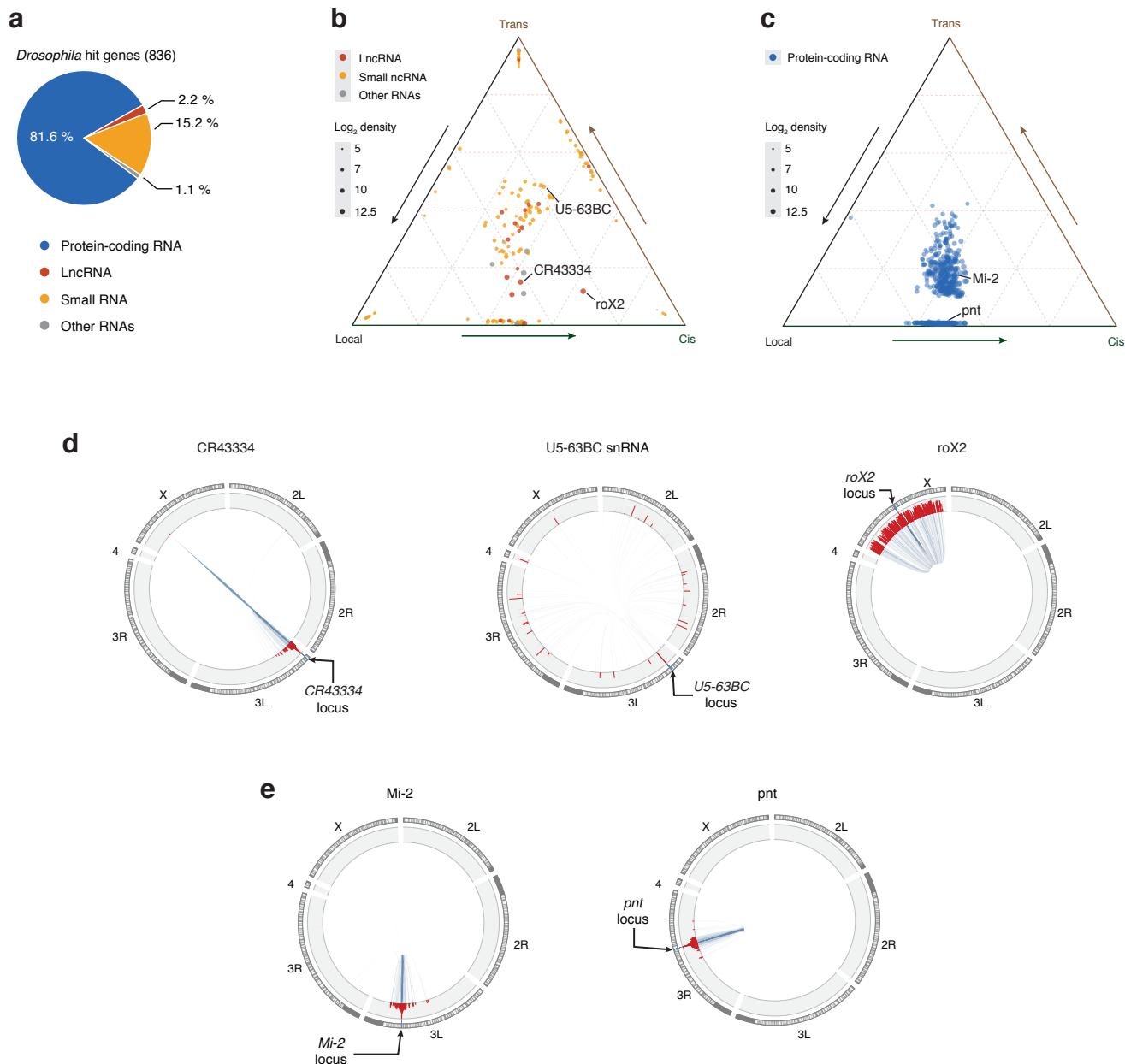
c



Extended Data Fig. 6 | *roX2* RNA-chromatin interaction map in *Drosophila* S2 cells

a, A heatmap showing the interaction of hit RNAs from Chromosome X across the whole *Drosophila* genome. Row: hit RNAs from their origins of transcription. Column: hit RNAs linked to DNA in the 100 Kb-binned *Drosophila* genome. Top: The background (Bg) deduced from endogenous *trans*-chromosomal interacting RNAs. Right is an enlarged view of the boxed region in panel **a**, showing decoration of *roX2* RNA on Chromosome X in S2 cells. Labeled on the right are representative non-coding hit RNAs (nc) and protein-coding hit RNAs (pc). **b**, A representative region of *Drosophila* Chromosome X, illustrating the interactions of *roX2* RNA with chromatin detected by ChIRP (first track) or ChART (second track) in comparison with GRID-seq signals or with the ChIP-seq signals for the TAP-tagged *roX2* binding protein MSL3. **c**, Meta-analysis of *roX2* chromatin-interacting signals detected by ChIRP (from GSM820427 and GSM820428), ChART (from GSM833475 and GSM833476) and GRID-seq relative to MSL3-TAP ChIP-seq peaks (from GSM296247).

Extended Data Fig. 7

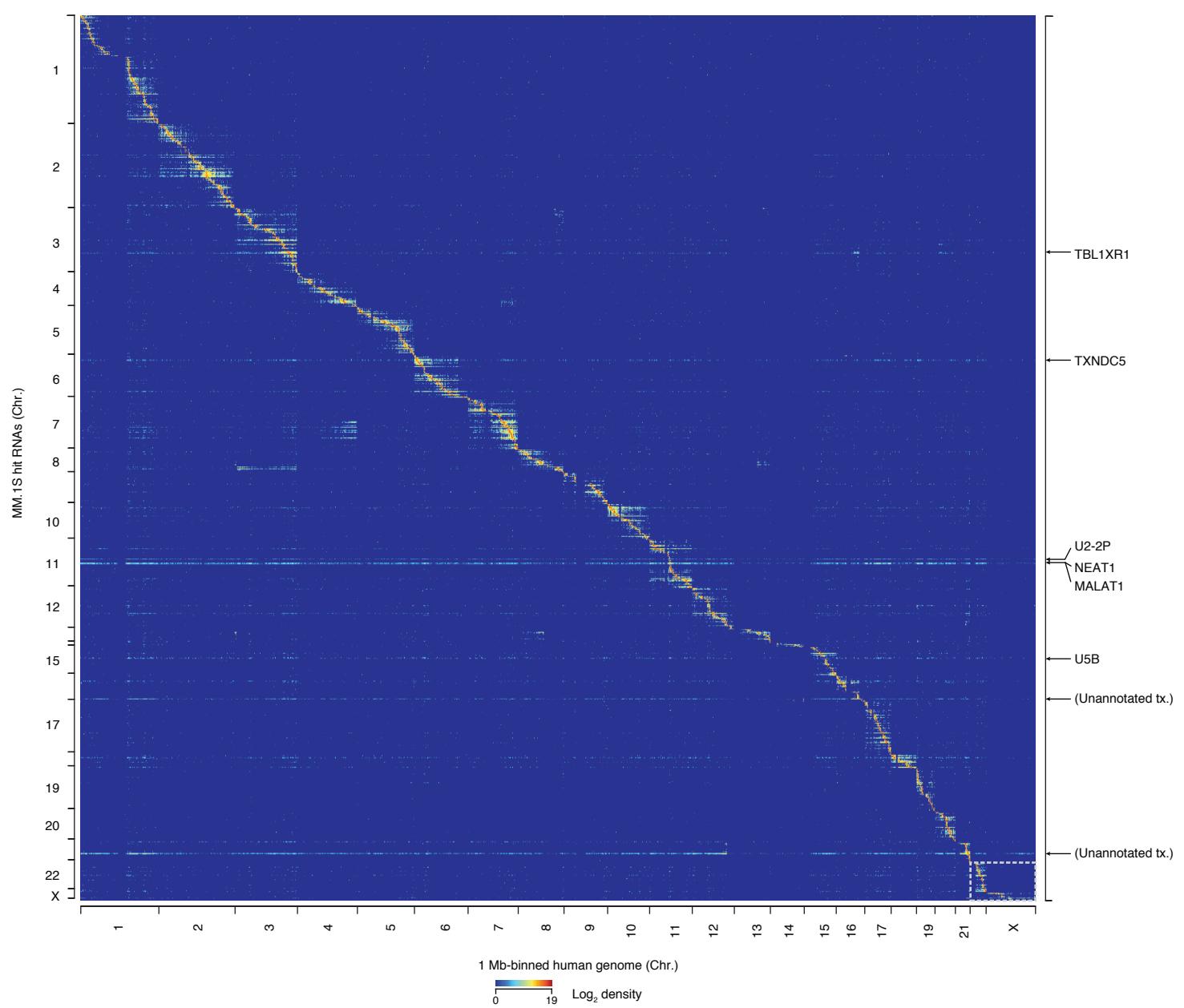


Extended Data Fig. 7 | Patterns of RNA-chromatin interactions in *Drosophila* S2 cells

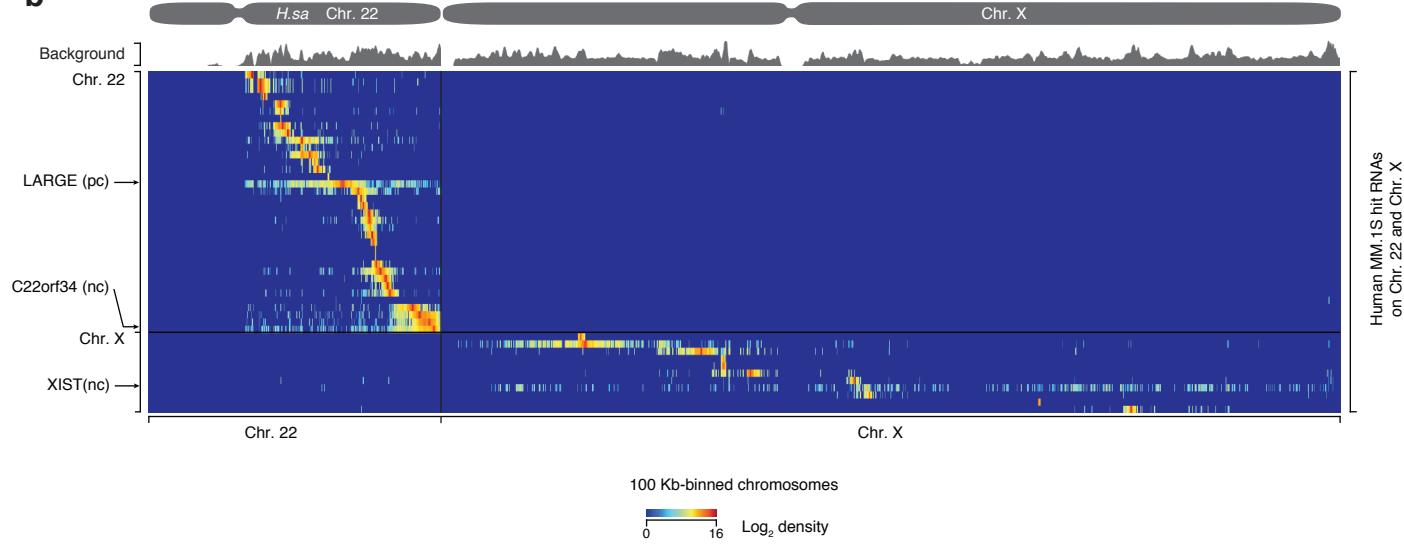
a, Relative representations of different hit RNA types in *Drosophila* S2 cells. **b**, Ternary plot of non-coding hit RNAs based their relative interactions in local (± 1 Kb from their genes), *cis* (the same chromosome the gene resides except local), and *trans* (all other chromosomes except its own chromosome) modes. Colors of dots represent different types of RNAs and sizes represent the levels of chromatin-interacting RNAs. **c**, Ternary plot of protein-coding hit RNAs as similarly analyzed in **b**. **d**, Circos plots of chromatin interactions of non-coding hit RNAs *CR43334* (left), *U5-63BC snRNA* (middle) and *roX2* (right) in the *Drosophila* genome. **e**, Circos plots of chromatin interactions of protein-coding hit RNAs *Mi-2* (left) and *pnt* (right) in the *Drosophila* genome.

Extended Data Fig. 8

a



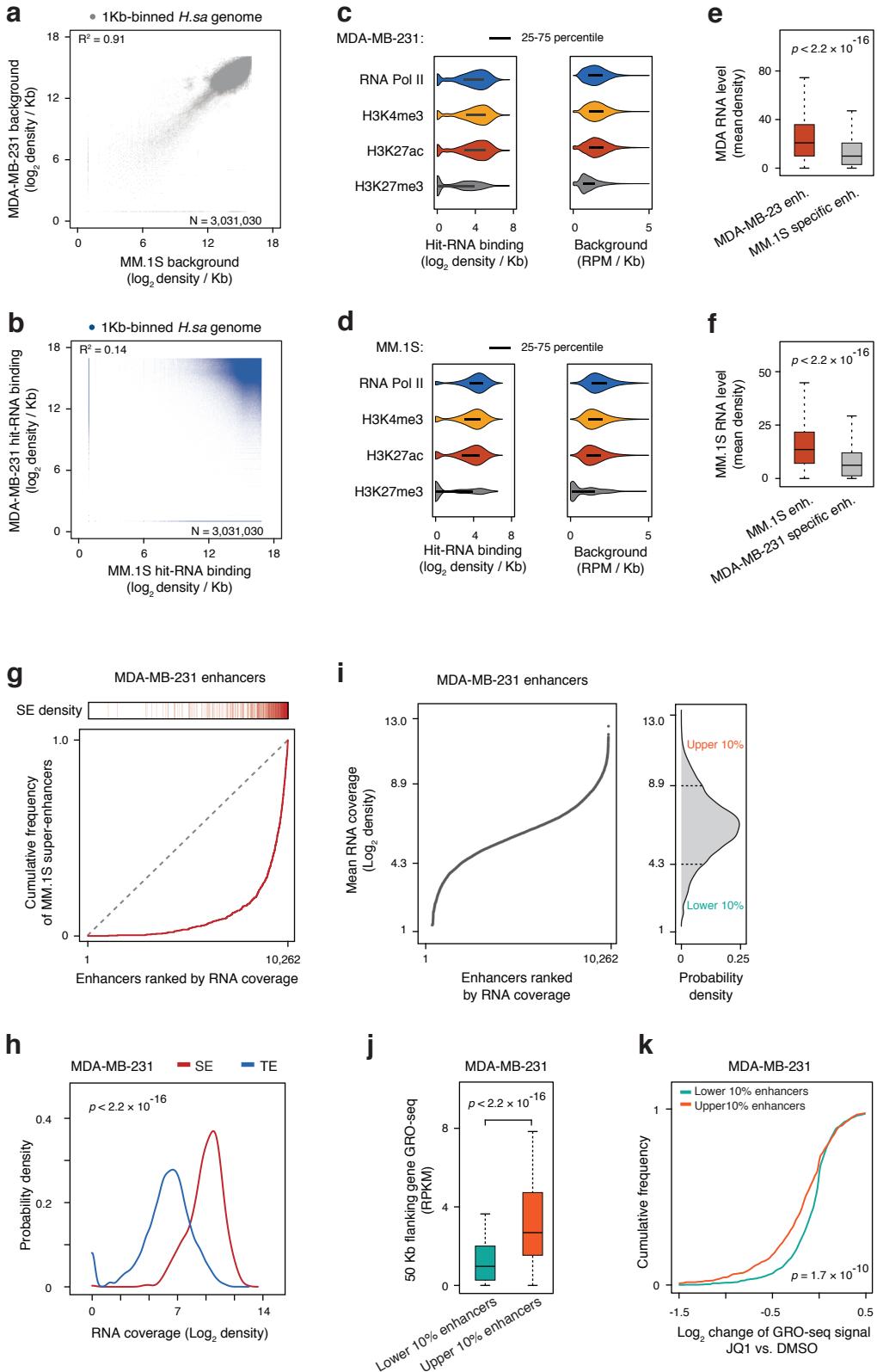
b



Extended Data Fig. 8 | Global view of RNA-chromatin interactions in human MM.1S cells

a, A heatmap showing the chromatin interactions of all hit RNAs across the whole human genome in MM.1S cells. Row: hit RNAs from their origins of transcription. Column: hit RNAs linked to DNA in the 1Mb-binned human genome. Labeled on the right are representative *trans*-chromosomal interacting RNAs.. **b**, Enlarged heatmaps of boxed Chromosome 22 and X in **a**, showing detailed chromatin interactions of hit RNAs from Chromosome 22 (left) and Chromosome X (right). Representative hit RNAs are labeled on the left (pc: protein-coding RNAs, nc: non-coding RNAs), showing that the hit non-coding RNA *XIST* interacts predominantly with Chromosome X (note that *XIST* is expressed in MM.1S cells, but not in MDA-MD-231 cells). Top: The background deduced from endogenous *trans*-chromosomal chromatin-interacting RNAs from all protein-coding genes.

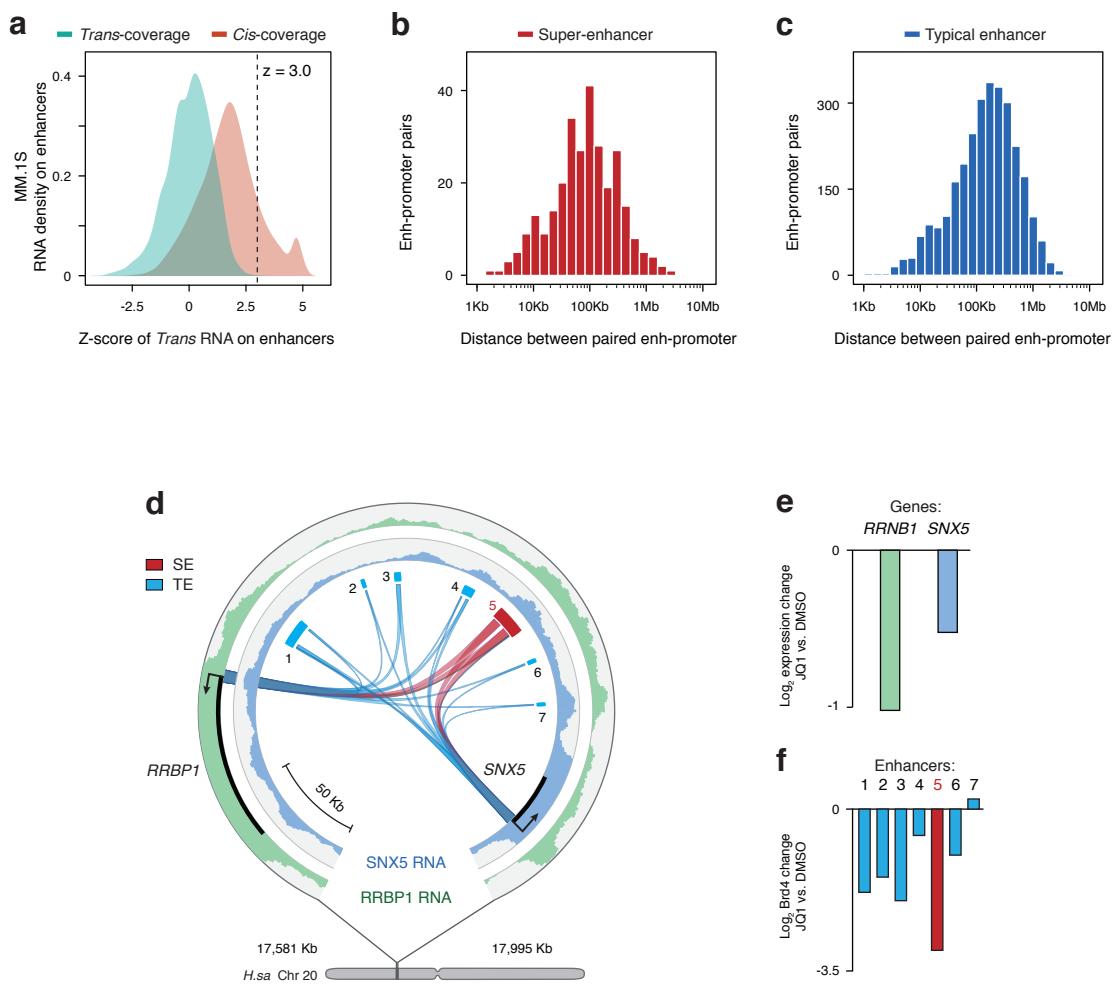
Extended Data Fig. 9



Extended Data Fig. 9 | RNA-chromatin interactions on cell type-specific enhancers

a, b, Scatter plots of background (**a**) and foreground (**b**) GRIP-seq signals between MDA-MD-231 and MM.1S cells in the 1 Kb-binned human genome. **c, d**, Violin plots showing co-enrichment of specific RNA-chromatin interactions and key chromatin marks in MDA-MB-231 cells (**c**) and MM.1S (**d**). Left: Enrichment of mean chromatin interaction signals of hit RNAs relative to ChIP-seq peaks of RNA Pol II, H3K4me3, H3K27ac and H3K27me3 (all based on the published ChIP-seq data, see Extended Methods). Right: Background signals. Bars represent the range from 25 to 75 percentile. **e, f**, Quantification of mean hit RNA densities on enhancers in the same cell type (orange) relative to specific enhancers in a different cell type (grey). **g**, Super-enhancers relative to RNA-chromatin interactions signals detected by GRID-seq. Enhancers in MDA-MD-231 cells were defined based on the mapped H3K27ac signals (from the data in GSM1204474 and GSM1204475). Each red bar on top represents a super-enhancer. Red line: The cumulative curve of rank-ordered RNA-chromatin interactions; Grey dashed line: Random distribution. **h**, Probability density map of hit RNA coverage on super-enhancers (SE, red) versus typical enhancers (TE, blue). **i**, Left: Rank-ordered RNA-chromatin interaction levels on all active enhancers. Right: Upper (SE enriched) and lower (TE enriched) 10 percentiles of enhancers selected for functional analysis. **j**, Expression of genes associated with top 10% RNA-interacting enhancers (red box) relative to those associated with bottom 10% RNA-interacting enhancers (green box), both within the ± 50 Kb range based on the GRO-seq assay performed on the same cell type. **k**, Fold changes in GRO-seq plotted in the accumulative fashion for the two groups of genes as defined in **c** in response to functional perturbation of enhancers on MDA-MD-231 cells by using the BRD4 inhibitor JQ1. Statistical significance of comparison is estimated by t-test in panel **h, j** and **k**.

Extended Data Fig. 10



Extended Data Fig. 10 | Inferred enhancer-promoter distance and example

a, Densities of hit RNA signals on typical and super-enhancers in MM.1S cells. x-axis: Z-scores of all *trans*-chromosomal RNA signals (teal). Z ≥ 3 was set to define significant RNA coverage on enhancers. **b**, Distribution of linear DNA distance between super-enhancers and genes. **c**, Distribution of linear DNA distance between typical enhancers and genes. The data indicate similar ranges reached out by typical and super-enhancers. **d**, A Circos plot, showing a representative case of two hit RNAs from two genes *RRBPI* and *SNX5* on nearby seven enhancers, one of which corresponds to a super-enhancer in MM.1S cells. The *RRBPI* RNA binding profile is shown on the outer track (green) and the *SNX5* RNA binding profile on the inter track (blue). Ribbons connecting with enhancers illustrate inferred enhancer-promoter interactions. **e**, **f**, Upon JQ1 treatment, fold changes in gene expression are shown in **e** (based on the data from GSM1094100, GSM1094101, GSM1094092 and GSM1094093) and fold changes in BRD4 binding on individual enhancers in **f** (based on the data from GSM1038271 and GSM1038275).