Nuclear exosome targeting complexes modulate cohesin binding and enhancer-promoter interactions in 3D

Charbel Akkawi, Alexandre Heurteau, Xavier Contreras, Marion Helsmoortel, Stéphane Schaak, Julie Bossuyt, Olivier Fosseprez, David Dépierre, Olivier Cuvier, Rosemary Kiernan

Nuclear exosome targeting complexes modulate cohesin binding and enhancer-promoter interactions in 3D

Charbel Akkawi[1,4], Alexandre Heurteau[2,3,4], Xavier Contreras[1,5], Marion Helsmoortel[1], Stéphane Schaak[2], Julie Bossuyt[1], Olivier Fosseprez[2], David Dépierre[2], Olivier Cuvier[2*] and Rosemary Kiernan[1*]

[1]Institut de Génétique Humaine (IGH), Univ Montpellier, CNRS, Montpellier, 34396 France

[2] Center of Integrative Biology (CBI/MCD unit), CNRS; Université de Toulouse, F-31000; France.

[3] Gene2i, Toulouse

[4] These authors contributed equally to this work

*Corresponding authors: Rosemary.Kiernan@igh.cnrs.fr; Olivier.Cuvier@univ-tlse3.fr

Keywords : ncRNA, RNA processing, RNA degradation, nuclear exosome, genome architecture, chromatin contacts, cohesin

Running title: Nuclear RNA exosome modulates 3D chromatin interactions

Abstract

Three-dimensional long-range contacts between enhancers and promoters are thought to be largely determined by loop extrusion driven by the cohesin complex and insulator factors. However, recent evidence also suggests a role for noncoding RNAs (ncRNAs), such as enhancer-associated RNAs (eRNAs) and promoter upstream transcripts (PROMPTs), in shaping enhancer-promoter connectivity. While nuclear RNA exosome, together with targeting complexes, PAXT and NEXT, control the decay of ncRNAs, it has not yet been determined whether these complexes regulate 3D contacts. Chromatin recruitment maps of ZCCHC8 (NEXT), ZFC3H1 (PAXT) and MTR4 helicase revealed that these factors that associate with sites of enhancer-promoter interactions. Depletion of NEXT, PAXT or MTR4 induced the accumulation of ncRNAs, notably enhancer-associated RNAs (eRNAs) and promoter upstream transcripts (PROMPTs). Strikingly, this further increased cohesin levels at sites accumulating ncRNAs. Chromatin conformation capture analysis revealed that MTR4 modulates the 3D long-range contacts between enhancers with their distant TSS targets. Upon loss of MTR4, contacts at anchor points increase while intraloop contacts decrease, suggesting that MTR4 facilitates loop extrusion. These data highlight a key interplay between cohesin-mediated enhancer-promoter interactions and the regulation of ncRNAs by nuclear RNA exosome that is consistent with a role for RNA in genome folding.

Introduction

The human genome needs to be highly compacted in order to occupy a nucleus that measures approximately 10 μm in diameter. Compaction must occur in a manner that permits coordinated gene expression. To achieve this, the genome is organized at multiple scales. Chromosomes are sub-divided into distinct active and inactive (A/B, respectively) compartments, and more locally folded into topologically associating domains (TADS) that correspond to regions favoring high frequencies of contacts between distant sites, such as enhancers and their target promoters, that lie within the same TAD. The establishment of TADs and loop structures is thought to be determined by the coordinated action of CCCTC-binding factor (CTCF) and the cohesin complex, a ring-shaped multimeric structure. TADs are thought to be formed via 'loop extrusion', a mechanism in which the cohesin ring entraps chromatin and extrudes loops, until blocked at TAD borders by CTCF or other insulating factors [1-8].

While the organization of the genome in 3D has been implicated in the regulation of gene expression, it is less clear how gene expression affects genome folding. Recent evidence suggests that RNAPII and its RNA product may indeed influence 3D genome interactions[9]. Several subunits of the cohesin complex, as well as CTCF, bind RNA in addition to DNA. Deletion of the RNA-binding domain of CTCF abolished about 50% of chromatin loops in mESCs[10,11]. Notably, TAD boundaries often express ncRNAs, which stabilize cohesin-CTCF interactions or facilitate CTCF recruitment. RNA-guided recruitment of CTCF has been shown to increase the efficiency of target search by several fold[12]. RNAPII was recently shown to be required for certain enhancer-promoter contacts and to antagonize loop extrusion[13,14]. Moreover, the recent development of 3D interaction maps of RNA has revealed interactions between enhancer-associated RNAs (eRNAs) and promoter proximal transcripts (PROMPTs) that parallel the functional connectivity of enhancers and promoters[15]. Mechanisms that control the abundance of such RNAs is likely to affect RNA-RNA spatial interactions and, consequently, genome looping and gene expression.

Many ncRNAs, including enhancer-associated RNAs (eRNAs) and promoter proximal transcripts (PROMPTs) are intrinsically unstable and are degraded by the nuclear RNA

exosome, a multi-subunit barrel-shaped structure containing a helicase, MTR4 and 3' to 5' exoribonucleases, Dis3 and Exosc10[16]. RNA substrates must first be delivered to the nuclear exosome, which is carried out by two major targeting complexes: the Nuclear Exosome Targeting Complex (NEXT)[17] and the polyA tail exosome targeting connection (PAXT)[18]. NEXT consists of a zinc-finger protein, ZCCHC8, together with RBM7 and the shared RNA helicase, MTR4. NEXT targets mostly short RNAs that are not polyadenylated. The second targeting complex, PAXT, is comprised of a zinc-finger protein, ZFC3H1, together with RBM26/RBM27, ZC3H3 and PABPN1, and targets polyadenylated RNAs[18,19]. We recently identified an additional PAXT subunit, polyA polymerase gamma (PAPγ), further strengthening the link between PAXT and polyadenylated RNA substrates[20]. RNAs targeted by PAXT and/or NEXT include not only aberrant transcripts, but many RNAs that are associated with gene regulatory regions, such as enhancer RNAs (eRNAs) and promoter-proximal transcripts (PROMPTs), also known as upstream antisense transcripts (uaRNAs)[19]. While nuclear exosome and its targeting complexes have been well characterized, whether the degradation of certain targets, such as eRNAs but also PROMPTs and ncRNAs, might affect genome organization and gene expression is not clear.

We recently described the genome-wide mapping of PAXT subunits, ZFC3H1, RBM26 and PAPγ[20]. Here, we have identified the genome-wide localization of ZCCHC8 subunit of NEXT, as well as the shared helicase, MTR4. Chromatin immunoprecipitation and high throughput sequencing (ChIP-seq) mapping revealed that NEXT subunit, ZCCHC8 and PAXT subunit, ZFC3H1, were detected primarily at transcription start sites (TSSs) but also at active enhancers. Consistent with their localization, loss of these factors was associated with stabilization of PROMPTs and eRNAs, as expected. Loss of MTR4 also strongly stabilized ncRNAs including eRNAs and PROMPTs, as described previously[17-19]. However, ChIP-seq mapping revealed that MTR4 was present at active enhancers, but was notably absent at TSSs, suggesting that its interaction with TSSs may be transient. By integrating ChIP-seq, RNA-seq and chromatin conformation capture (Hi-C) data, our data show that MTR4 contacts TSSs via 3D interactions between enhancers and promoters. Hi-C and ChIP-seq data

furthermore showed that MTR4 modulates cohesin-dependent 3D contacts. Loss of MTR4 increases enhancer-promoter contacts at the anchor site but antagonizes loop extrusion. These data suggest that nuclear RNA surveillance factors contribute to chromatin loop formation and help shape genome architecture.

Results

PAXT, NEXT and MTR4 are recruited to specific sites on chromatin

We have previously shown that PAXT subunits, ZFC3H1, RBM27 and PAPγ, are associated with chromatin, frequently at TSSs that generate PROMPTs[20]. To gain further insight into nuclear exosome-dependent regulation of RNAs and ncRNAs, which may involve its recruitment to sites of RNA synthesis, we mapped the genomic localization of NEXT and the shared helicase, MTR4, by chromatin immunoprecipitation combined with high throughput sequencing (ChIP-seq). While MTR4, ZFC3H1 and ZCCHC8 recruitment sites frequently co-localized, they also showed some notable differences (Fig. 1a and Supplementary Fig. 1). Only a partial overlap between ZFC3H1 and ZCCHC8 recruitment sites was expected, which is consistent with their function in targeting specific RNAs[17-19] (Fig 1b-d, Supplementary Fig. 1). MTR4 recruitment sites overlapped with 1251 and 627 recruitment sites of ZCCHC8 and ZFC3H1, respectively (Fig. 1b; Fisher's exact test: p-values < 1e-126and <1e-235, respectively), consistent with the functions of PAXT or NEXT with MTR4 in targeting specific RNAs[17-19]. Surprisingly, however, the majority of MTR4 sites did not overlap with those of either ZFC3H1 or ZCCHC8 even though MTR4 is a subunit of both PAXT and NEXT macromolecular complexes. Further inspection using averaged profile analysis of ChIP-seq peaks showed that while recruitment of ZCCHC8, ZFC3H1 and MTR4 partially overlapped, ZFC3H1 and MTR4, in particular, showed some specificity in recruitment site preference (Fig. 1c-e, top panels). Ranking of ChIP-seq signal of ZCCHC8 or ZFC3H1 showed relatively moderate levels of MTR4 (Fig. 1c-d, bottom panels), though significantly above random signal as shown on average profile analysis (Fig. 1c-d, top panels). Similarly, averaged profile analysis and ranking of MTR4 ChIP-seq reads on heat maps further revealed only modest signal of either

ZFC3H1 or ZCCHC8 (Fig. 1e, Supplementary Fig. S1C). These results suggest that while binding sites of ZFC3H1, ZCCHC8 and MTR4 overlap to some extent, many sites appeared to be independent.

Recruitment of PAXT, NEXT and MTR4 at ncRNA target sites

To further investigate binding profiles of NEXT, PAXT and MTR4, we compared sites of recruitment to those of synthesis of their ncRNA targets. To do so, we generated atlases for coding and non-coding RNA targets, using strand-specific RNA sequencing data obtained from cells depleted of MTR4 (Supplementary Fig. 2A), ZFC3H1 subunit of PAXT and ZCCHC8 subunit of NEXT[18,21,22]. As expected, depletion of MTR4, ZCCHC8 or ZFC3H1 led to the accumulation of thousands of ncRNAs (Supplementary Fig. 2B), using a high confidence threshold for detection (FDR p-value of 1e-6). The ncRNAs detected in each depletion condition showed significant overlap, with more than 500 ncRNAs detected upon depletion of either ZFC3H1 or ZCCHC8 that were also detected upon MTR4 depletion (Supplementary Fig. 2B; p-value of 1e-102 and 1e-149, respectively) and levels of >200 ncRNAs were increased upon depletion of ZFC3H1, ZCCH8 or MTR4 (Supplementary Fig. 2B; p-value of 1e-121).

The ncRNAs detected were classified into categories based on their localization with respect to genes (Supplementary Fig. 2C). A significant fraction of ncRNAs detected occur immediately upstream (<2 kb) from transcription start sites (TSSs) of coding genes (Supplementary Fig. 2C-D), corresponding to PROMPTs, in agreement with previous reports[18,19,21]. Approximately 30% of the ncRNAs increased upon depletion ZFC3H1, MTR4 or ZCCHC8 were identified as PROMPTs (Supplementary Fig. 2C). Other categories of ncRNAs increased upon depletion of ZFC3H1, ZCCHC8 or MTR4 include eRNAs, antisense RNAs within gene bodies, and RNAs in intergenic regions (Supplementary Fig. 2C). Among PROMPTs, 62.5% of ZFC3H1-dependent PROMPTs and 51.6% of ZCCHC8-dependent PROMPTs were also increased following MTR4 depletion (Supplementary Fig. 2E). Similarly, 54.6% of ZFC3H1-dependent eRNAs and 42.5% of ZCCHC8-dependent eRNAs were also increased following MTR4 depletion **(**Supplementary Fig. 2F**)**. Our data are thus similar to

previous analyses showing the contribution of MTR4-containing complexes in regulating ncRNAs including eRNAs and PROMPTs.

We then analyzed the genomic distribution of MTR4, ZFC3H1 and ZCCHC8 ChIP-seq reads with respect to their ncRNA targets. Surprisingly, less than 2% of MTR4 sites localized near a TSS, in contrast to ZFC3H1 sites that were more frequently associated with a TSS, and to a lesser extent ZCCHC8 sites (Fig. 2a). Accordingly, MTR4 recruitment sites accounted for < 5 % of the PROMPTs detected upon the depletion of ZFC3H1, ZCCHC8, or MTR4 itself (Fig. 2b; p-values of 1). In contrast, ZFC3H1 was detected at 27% of the PROMPTs detected upon its depletion (Fig. 2b; bottom left Venn diagram, p-value < 1e-11). Similarly, 28% of ZCCHC8-dependent PROMPTS co-localized with a ZCCHC8 recruitment site (Fig. 2B; bottom right Venn diagram). Consistent with the poor overlap between MTR4 binding sites and PROMPTs, recruitment of MTR4 was essentially not detected at TSSs bound by either ZFC3H1 or ZCCHC8 (Fig. 2c). Accordingly, the histone mark H3K4me3 that is typical of active TSSs was enriched at ZFC3H1 or ZCCHC8 sites, but not at MTR4 sites (Supplementary Fig. 2G). In contrast, MTR4 recruitment sites were enriched in the histone H3K4me1 mark that is associated with enhancers (Supplementary Fig. 2G). In support of this finding, ZFC3H1- and ZCCHC8-bound enhancers were associated with MTR4 (Fig. 2d). In summary, MTR4 co-localizes with ZFC3H1 or ZCCHC8 primarily at sites that bear marks of active enhancers, in stark contrast with TSSs where MTR4 recruitment is poorly detected.

Distal MTR4 sites establish long-range interactions with TSSs that generate PROMPTs

To understand how MTR4 may influence the accumulation of TSS-associated PROMPTs despite being poorly detectable at sites of PROMPT synthesis, we first determined the expression level of flanking genes in the presence and absence of MTR4, reasoning that increased expression of the flanking gene might explain the increase in PROMPTs. While the expression of genes flanking PROMPTs was sometimes deregulated in MTR4-depleted cells (126 genes) (Supplementary Fig. 3A), MTR4-dependent PROMPTs were associated with both up-regulated and down-regulated genes, similarly to ZFC3H1- or ZCCHC8-dependent

PROMPTs (Supplementary Fig. 3B-E). This result argues against the possibility that the accumulation of MTR4-dependent PROMPTs was linked to increased expression of the flanking gene. Moreover, these results suggest that the differential expression of genes does not explain the effect of MTR4 on TSS-associated PROMPTs.

As MTR4 was recruited to distant sites marked by H3K4me1 (Supplementary Fig. 2G), our data raised an alternative possibility that MTR4, together with ZFC3H1 or ZCCHC8, may be involved in regulating PROMPTs depending on enhancer-promoter long-range interactions. Cohesin preferentially binds to H3K4me1-associated enhancers[23] to mediate long-range interactions (LRIs) with distant TSSs that may be more or less transient[7,24]. Remarkably, more than half of all MTR4 binding sites co-localized with the Rad21 subunit of cohesin, and a significant proportion co-localized with both Rad21 and H3K4me1 (Fig. 3a and Supplementary Fig. 4A-B; Fisher exact test p-value of 1e-225 and 1e-152). Assessing long-range interactions in 3D by Chromatin Conformation Capture (Hi-C) using 2D interaction maps indicated that many recruitment sites of PAXT/NEXT/MTR4 established LRIs with distant TSSs that accumulate PROMPTs upon nuclear exosome depletion (Fig. 3b). To further test a possible role of LRIs in juxtaposing MTR4 to its TSS-associated target, we first performed a genome-wide ranking of all pairs of enhancers/TSSs from the highest to lowest interaction, based on normalized levels of Hi-C counts[25-27] (Fig. 3c). TSSs were further scored for the presence or absence of PROMPTs. Of interest, PROMPTs were largely enriched at TSSs scoring high interaction levels with distal enhancers bound by MTR4 (Fig. 3c, bottom row of matrix). In contrast, TSSs having the least contact with distal MTR4 sites were significantly depleted of PROMPTs (Fig. 3c, bottom row). Moreover, TSSs with high levels of contacts with enhancers not bound by MTR4, were not significantly associated with PROMPTs (Fig. 3c, top row), indicating that stronger enhancer-promoter interactions alone do not explain the enrichment of PROMPTs. The differential levels of LRIs was validated by assessing the averaged physical contacts by aggregation of Hi-C data (aggregated plot analysis, APA), which readily estimates genome-wide LRIs[26,27]. APA showed that LRIs between enhancer/TSSs were clearly higher for the upper quartile that are enriched in MTR4-bound enhancers and PROMPTs compared

to the lower quartile that are depleted of PROMPTs (Fig. 3d). These data suggest that PROMPTs are enriched at promoters engaged in high levels of LRIs with distal enhancer-bound MTR4. Moreover, these data fully agree with the formation of physical contacts between distant enhancers associated with nuclear exosome-associated sites and their target TSSs harboring PROMPTs (Fig. 3c, schematic, top panel). As such, these data raise the possibility that regulation of ncRNAs may involve the juxtaposition of enhancer-bound MTR4 complexes to distal TSSs, in 3D.

MTR4 is associated with sites of cohesin-dependent long-range interactions and modulates cohesin binding.

Since MTR4 co-localized significantly with the Rad21 subunit of cohesin and the H3K4Me1 mark associated with active enhancers (Fig. 3a and Supplementary Fig. 4A-B), we tested if MTR4 might be recruited to cohesin-associated LRIs genome-wide. Physical contacts were probed through aggregation of Hi-C data (APA). LRIs were further oriented in the direction of sense transcription, which allows an estimate of loop extrusion associated with the enhancer-promoter contacts. Hi-C data and MTR4 ChIP-seq data from HeLa cells were analyzed to determine the frequency of cohesin-associated LRIs and loop extrusion, associated with MTR4 co-localization or not (Fig. 4a). Firstly, the analysis revealed that MTR4 is present at distal enhancers that are engaged in 3D interactions with TSSs and that these LRIs are associated with loop extrusion (Fig. 4a, left panel). Although this phenomenon could be general, we noted that LRIs involving enhancers that were not associated with MTR4 (No MTR4) showed significantly less loop extrusion (Fig. 4a; right panel, Fig 4b). This is likely to be due at least in part to low Rad21 association at these sites, since MTR4 recruitment sites frequently co-localize with Rad21 (Fig. 3a and Supplementary Fig. 4A-B). Indeed, LRIs between Rad21-associated enhancers and TSSs showed strong loop extrusion while enhancers that were not associated with Rad21 showed significantly lower levels of loop extrusion, as expected (Supplementary Fig. 4C-D). Thus, MTR4 is associated with enhancers that engage in LRIs with TSSs and form extruded genome loops.

Since MTR4 frequently co-localizes with Rad21, we sought to test whether MTR4 might influence cohesin binding, as measured by ChIP-seq of cohesin in cells depleted of MTR4 compared to controls. To this end, we measured the association of cohesin with chromatin in cells depleted of MTR4 compared to controls by ChIP-seq of the Rad21 subunit. Strikingly, MTR4 depletion increased the accumulation of Rad21 that was bound at several hundred MTR4 sites (Fig. 5c and Supplementary Fig. 5A-B). Of note, Rad21 protein levels were unchanged in extracts of MTR4-depleted cells compared to controls (Supplementary Fig. 5C), indicating that the observed accumulation was not due to increased Rad21 expression. To determine whether such increases might be related to the abundance of ncRNAs, TSSs were then ranked by the accumulation of PROMPTs following MTR4 depletion (Fig. 4d, left heatmaps, highest to lowest PROMPT accumulation). Accumulation of PROMPTs at TSSs correlated strikingly with gain in Rad21 ChIP-seq signal in MTR4-depletion condition (Fig. 4d, right heatmaps). Furthermore, TSSs at which Rad21 most accumulated upon MTR4 depletion were also those at which PROMPTs increased upon depletion of ZFC3H1 or ZCCHC8, in addition to MTR4 (Supplementary Fig. 5D). We also observed that the most significant effect was associated with the differential expression of the flanking genes (Supplementary Fig. S5E), suggesting an association with gene expression. In addition, repeating the same analysis upon depletion of ZFC3H1 (Supplementary Fig. 5F) confirmed the impact on Rad21 signal (Fig. 4e-f), indicating that cohesin association is modulated by MTR4, likely in the context of PAXT. Accumulation of cohesin was similarly found over distant MTR4 sites and TSSs, depending on ZFC3H1 depletion. As observed for MTR4, depletion of ZFC3H1 did not alter the expression of Rad21 in cell extracts (Supplementary Fig. 5G). Therefore, our results show that nuclear RNA exosome, which regulates the abundance of non-coding RNAs, also modulates the accumulation of cohesin at the same sites.

Since the accumulation of cohesin, that is involved in the formation of long-range contacts, was modulated by MTR4 at its recruitment sites, we wondered whether its depletion might also affect 3D genome interactions. Thus, aggregation of Hi-C data obtained from control cells and cells depleted of MTR4 was performed (Supplementary Fig. 6A). TSSs showing

accumulation of Rad21 following loss of MTR4 ($p < 10^{e}$-2) were selected for analysis. Consistent with an increased association of Rad21, LRIs between TSSs and distant sites were significantly enhanced at the central region (Fig 5a and Fig. 5b, left panel). Of interest, contacts in the intra-loop region were significantly decreased following loss of MTR4 compared to the control condition (Fig. 5a and Fig. 5b, right panel). Interestingly, this pattern resembles that of LRI sites that are not associated with MTR4 in control conditions (Fig. 4a-b). Changes in LRIs were visible at specific MTR4-associated genomic regions (Fig. 5c). Taken together, these data show that LRIs are influenced by nuclear RNA exosome complexes containing MTR4.

Discussion

Here, we identify an intricate regulation of 3D long-range interactions between enhancers and TSSs by nuclear RNA exosome complexes. We began by mapping the genomic localization of targeting complexes, PAXT and NEXT, via ZFC3H1 and ZCCHC8 subunits, respectively, and their shared RNA helicase subunit, MTR4. Comparison between the localizations of the complexes and ncRNA accumulated upon their depletions, revealed an interplay between the stability of ncRNAs, including PROMPTs and eRNAs, and 3D long-range interactions. Nuclear exosome subunits were found to be localized at sites that engage in long-range genomic interactions. Moreover, loss of nuclear exosome subunits, MTR4 or ZFC3H1, led to significant changes in the accumulation of cohesin, a key player in long-range interactions, at both TSSs and enhancers and modified the dynamics of chromatin looping (Fig. 5D).

Increasing evidence suggests that RNAPII and its RNA product may play a role in genome organization. Treatment with RNase abolished almost half of all genome contacts[10]. CTCF and cohesin subunits are RNA-binding proteins and mutation of the RNA binding region (RBR) of CTCF significantly disrupted CTCF-dependent insulation of chromatin domains and loop extrusion[10,11]. More recently, RNAPII has been shown to shape genome architecture by blocking loop extrusion while facilitating enhancer-promoter contacts[13,14]. Our data show that depletion of nuclear RNA exosome targeting complexes impacts 3D genome architecture and

binding of cohesin. It seems unlikely that MTR4 or associated complexes, such as PAXT, are directly responsible for the effects observed on LRIs. Instead, it seems more probable that ncRNA targets of nuclear exosome, particularly eRNAs and PROMPTs, that were concomitantly stabilized at the enhancers and promoters in contact, may be responsible for the effect on LRIs and loop extrusion. Indeed, these data are consistent with recent studies showing that enhancer-promoter interactions are facilitated by RNAPII[13,14] as well as pair-wise interactions between eRNAs and PROMPTs[15]. Thus, stabilization of ncRNAs, such as eRNAs and PROMPTs, following depletion of nuclear exosome targeting complexes would be expected to modify RNA-RNA spatial interactions and consequently, affect genome looping and gene expression.

Interestingly, our analysis revealed that although interactions between enhancers and TSSs and association with the Rad21 subunit of cohesin were increased, loop extrusion appeared to be less efficient. We speculate that insulator proteins, such as CTCF, Mediator[28] or RNA-binding proteins, such as hnRNPK[15], may be retained at the LRI sites, which may impede or contribute to a reduction in loop extrusion. The genome-wide mapping of nuclear exosome targeting complexes, described previously[20] and in the present study reveal a strong association with TSS regions of actively transcribing genes and RNAPII. While still speculative, these findings raise the possibility that an important function of these complexes may be to fine-tune the abundance of TSS-associated and enhancer-associated RNAs and, consequently, modulate genome architecture.

An unexpected finding of genome-wide mapping was the absence of detectable MTR4 at TSSs, even though subunits of PAXT and NEXT are readily detectable at TSSs. MTR4 was more stably associated with enhancer regions and was present at sites engaged in long-range interactions with TSSs. This might suggest that cohesin-mediated LRIs may influence ncRNAs post-transcriptionally, depending on spatial targeting of the nuclear RNA exosome. Interestingly, Rahman et al. demonstrated that eRNA synthesis precedes looping and that their production is strongly reduced in the closed loop conformation[29]. Thus, chromatin loops might help degrade eRNAs when they are no longer required, by increasing nuclear exosome

recruitment, or facilitating the activation of nuclear exosome locally, by juxtaposing the targeting complexes with MTR4 helicase whose binding is required to activate nuclear exosome. LRIs may thus define the context in which the nuclear exosome degrades ncRNAs, notably eRNAs and antisense PROMPTs. Improper accumulation of these ncRNAs may deregulate, either directly or indirectly, expression of the neighboring genes.

Materials and Methods

Cell culture and reagents

HeLa cells were grown in Dulbecco's modified Eagle's minimal essential medium (DMEM) (Sigma-Aldrich, D6429), supplemented with 10% fetal calf serum (FCS; Eurobio Scientific, CVFSVF00-01) and containing 1% penicillin-streptomycin (Sigma-Aldrich, P4333). HEK-293T were grown in Hepes-modified DMEM (Sigma-Aldrich, D6171), supplemented with 10% FCS (Eurobio Scientific, CVFSVF00-01) and containing 1% penicillin-streptomycin (Sigma-Aldrich, P4333). All cells were grown in a humidified incubator at 37°C with 5% CO2.

Antibodies

Antibodies used in this study are shown in Table S1.

RNAi

Production of short hairpin RNA (shRNA)-expressing lentiviral particles was performed as described previously (20) using plasmids expressing shRNAs targeting MTR4 (Sigma-Aldrich MISSION shRNA, TRCN0000296301), ZFC3H1 (Sigma-Aldrich MISSION shRNA, TRCN0000130498) or a non-targeting control (Addgene, plasmid 1864). For knockdown experiments, HeLa cells were either transduced with lentiviral particles and harvested 5 days later, or transfected with siRNAs (10 nM) shown in Table S2 using Interferin (PolyPlus) and harvested 72 h later, as described previously[30]. RNAi depletions were validated by immunoblot of total cell extracts, as described previously[30].

358

359 **Sample preparation for RNA-seq, ChIP-seq and Hi-C**

360 For RNA-seq, total RNA was extracted from HeLa cells using TRIzol (Thermo Fisher Scientific)
361 according to the manufacturer's instructions. RNA-seq (paired end, 125 bp) was carried out by
362 BGI Genomic Services in triplicates.

363 Chromatin immunoprecipitation followed by high throughput sequencing (ChIP-seq)
364 was performed from HeLa cells as described previously[31] using the ChIP-IT High Sensitivity®
365 kit from Active motif (ref #53040) according to the manufacturer's instructions. Each ChIP used
366 30 μg of chromatin along with 4 μg of antibody detecting MTR4, ZFC3H1, ZCCHC8 or Rad21
367 (Supplementary Table S1). ChIP-seq libraries were constructed using the Next Gen DNA
368 Library Kit (Active Motif 53216 and 53264). Library quality was assessed using Agilent 2100
369 Bioanalyzer and Agilent High Sensitivity DNA assay. High throughput sequencing (PE75) was
370 performed by Sequencing-By-Synthesis technique using a NextSeq 500 (Illumina) at Genom'ic
371 facility, Institut Cochin, Paris.

372 Hi-C was performed from HeLa cells using the Arima Hi-C ® kit from Arima Genomics
373 (ref #A510008) according to the manufacturer's instructions. Libraries were prepared using the
374 KAPA® Hyper Prep kit (KK8500 and KK4824) and Illumina adapters (ref 20016329). High
375 throughput sequencing (PE150) was performed by Novogene (NovaSeq) or BGI (DNBseq).

376

377 **RNA-seq alignment and ChIP-seq data integration**

378 Raw RNA-seq data were aligned on hg19 reference genome using STAR aligner in paired-
379 end and strand-aware mode. Resulting BAM files were used as such for differential analysis.
380 For the detection of non-coding RNAs, all aligned BAM files were further split in two files
381 corresponding to the strands of alignment.

382 After quality control using fastQC tool, ChIP-seq reads were aligned on hg19 reference
383 genome using Burrows-Wheeler Aligner (BWA, http://bio-bwa.sourceforge.net/)[32] and then
384 normalized on the corresponding input as described previously [33]. ChIP-Seq peak calling were
385 performed in all conditions (depleted or not) with MACS2 with R 3.4.2 version. Gene TSSs

associated with such peaks were identified by overlapping within the +/-500bp region around TSSs. Enhancers associated with peaks were identified by overlapping peaks summit with each enhancer range (+/-1000bp). Overlapping analyses were performed using 'GenomicRanges' package functions of R (https://bioconductor.org/packages/release/bioc/html/ GenomicRanges.html). Venn Diagrams and statistical enrichment tests were performed using Fisher's exact test or proportional tests. Boxplots showing quantification of ChIP-seq reads at the indicated genomic coordinates as performed using a Wilcoxon test comparing WT, MTR4-KD, ZFC3H1-KD or ZCCHC8-KD conditions, over the lists of gene loci. Boxplots were generated using ggplot2[34] R package functions (https://cran.r-project.org/web/packages/ggplot2/index.html). Heatmaps were generated by aligning genes (oriented towards the right) on TSSs (position 0). Genes were ranked according to normalized ChIP-seq reads (TSS +/- 500 bp) of ZFC3H1 followed by visualization of the indicated ChIP-seq reads (of ZFC3H1, MTR4 or ZCCHC8), or of ncRNAs. Visualisation of heatmaps was performed with SeqPlot package[35]. H3K4me1, and Rad21 peaks on Hela cells were obtained from Encode Project (GEO accession GSM798322 and GSM935571 respectively).

Detection of non-coding RNAs

Differential non-coding RNA levels were detected in MTR4 (this study), ZFC3H1 or ZCCHC8 depleted cells[18,21,22]. NcRNAs levels were compared between each condition and with control mock-depleted ('WT control') wild-type cells using enrichR function from NormR (binsize of 500 bp, FDR of 1e-5)[36]. Preliminary analyses were performed in comparison with analyses by mNET-seq and chromatin RNA-seq data in siLUC control and EXOSC3 siRNA conditions escribed previously[37,38] .

Heatmaps of RNA-seq and ChIP-seq data

Heatmaps were generated by aligning genes (oriented to the right) on TSSs (position 0) and by plotting the normalized bigwig signal, on reverse and forward strand for RNA-seq, simultaneously for each condition. Genes were ranked according to normalized ChIP-seq

reads (TSS +/- 500 bp) of a given feature as indicated in the figure. Visualisation of heatmaps was performed with SeqPlot package[35].

Plots

Boxplot, Scatterplots, histograms and profiles were generated with ggpubr package based on ggplot2 package[34]. Venn diagrams were performed using Vennerable v1.0 package, developed by Jonathan Swinton.

Computational integration of Hi-C data

Hi-C data at 5 Kb resolution were obtained from WT Hela cells or from cells depleted of MTR4. Hi-C data were validated by comparison with Hi-C in WT cells from Rao and colleagues (GSE63525)[25] or Hi-C in wild-type and PDS5 depletion condition from J.M. Peter's lab (GSE102884)[39]. Extraction of long-range contacts was performed using dump command from juicer tools in order to obtain normalized contacts (observed/expected and Knight-Ruiz for matrix balancing normalization) and filtered to fit inside Hela TADs. APA plots were generated as described [24] with heatmap function from R stats library after importing dumped matrices in 2D/3D, as previously described by 'Aggregate Peak Analysis' (APA)[27,40], run on the normalized Hi-C matrices at a resolution of 10 kb as described[26,27,40] with some adaptations. A minimal/maximal distance threshold of 200 kb to 1 Mbp was applied and each pair of sites was oriented to align the interactions between distant enhancers +/- bound by Cohesin/MTR4 sites and transcription start sites (TSSs) with their gene body towards the right of the APA. For statistical validation, a sampling of the interactions was performed to assess 50 independent times the interactions of 5,000 pairs in every condition. The variation was then tested using a Wilcoxon pairwise test for the central square values (3x3 central square (CS) of each of the 50 aggregated matrices) between the top and control group (e.g. top left corner of the APA matrix) applying a Wilcoxon test (stat_compare_mean from ggpubr, R package version 0.6.0). We estimated loop extrusion (all squares on the right/3' that align with the central pixel) in parallel to the loop values (by taking the same number of pixels from the APA matrix (either

3x3 central square or 9 pixels on the right; each of the 50 aggregated matrices) and compared values using a Wilcoxon test (stat_compare_mean from ggpubr, R package version 0.6.0). Enrichment of a given site was performed by ranking all enhancer-TSSs interactions within a range of 200 kb to 1 Mbp according to their interactions strength (Normalized Hi-C counts) to define quartiles of enhancer-TSSs pairs from low-to-high (q1 to q4) interaction levels. Then, the proportion of a given feature (e.g. MTR4 sites) was scored for each quartile and enrichment was assessed using a Fisher's exact test with respect to all combinations. Visualization of Hi-C signal was performed using Juicebox using default settings[41].

Statistical analyses of ChIP-Seq data

After quality control using fastQC tool, ChIP-seq reads of MTR4, ZFC3H1, ZCCHC8 and Rad21 and were aligned on hg19 reference genome using Burrows-Wheeler Aligner (BWA, http://bio-bwa.sourceforge.net/)[32] and then normalized using the corresponding input. ChIP-Seq peaks were identified using MACS2 with normalization to the corresponding input sequenced in parallel generated in R. All subsequent analyses were done with R 3.4.2 version. Genes associated with peaks were identified by overlapping within the +/-500bp region around TSSs as defined by human TxdB database based on hg19 annotations (UCSC). Enhancers were identified by CAGE-seq (Encode). Overlapping analyses were performed using « GenomicRanges » R functions (https://bioconductor.org/packages/release/bioc/html/GenomicRanges.html), as described previously[33]. Enrichments in Venn diagrams and enrichment matrices were performed using Fisher's exact test, p-values are either one-sided or two-sided to reflect enrichment and under-representation (i.e. positive Odds ratio, alternative = « greater » of R Fisher Test). Proportional Venn diagrams were plotted with « Vennerable » R package (https://github.com/js229/Vennerable). Strand-specific normalized read quantifications were performed by aligning stranded reads over oriented TSSs (+/-5000bp). For analyses of enhancer-promoter associations, enhancers were those defined by Cap analysis of gene expression (CAGE-seq) by the Functional annotation of the mammalian genome (Fantom5) and Encode projects[42,43].

Statistical analyses of gene expression data

Gene expression analyses were first performed by polyA+ RNAseq analysis from HeLa cells depleted of MTR4 (MTR4-KD), ZFC3H1 (ZFC3H1-KD) as compared to siRNA treated control cells ('WT') in HeLa cells. Normalized strand-specific RNAseq reads were visualized using the IGV Gbrowser. Normalized strand-specific reads were further quantified in duplicates and directly plotted onto the heatmaps depending on gene ranking according to the indicated levels of ChIP-seq reads. After filtering low reads genes using R package HTSfilter, a differential analysis was performed with DESeq2 R package for comparing expression level of each gene in MTR4-KD or ZFC3H1-KD compared to siRNA treated WT conditions. Up-regulated and down-regulated genes subsets were defined according to a p-value threshold <0,05 and a LogFoldChange threshold > 1.5 or < 1.5 respectively. Enrichment tests between differentially expressed genes and other subsets of genes, such as those associated with peaks, were done using Fisher's Exact test. Intersection matrices were performed using GRange with Fisher's exact test.

Data availability

RNA-seq, ChIP-seq and Hi-C data have been deposited at GEO (GSM4263550, GSE154100 and GSE283061). All data are available in the main text or the supplementary materials.

References

1. Rowley, M.J. & Corces, V.G. Organizational principles of 3D genome architecture. *Nat Rev Genet* **19**, 789-800 (2018).
2. Bonev, B. & Cavalli, G. Organization and function of the 3D genome. *Nat. Rev. Genet.* **17**, 661-678 (2016).
3. Ea, V., Baudement, M.O., Lesne, A. & Forne, T. Contribution of Topological Domains and Loop Formation to 3D Chromatin Organization. *Genes (Basel)* **6**, 734-750 (2015).

4. Merkenschlager, M. & Nora, E.P. CTCF and Cohesin in Genome Folding and Transcriptional Gene Regulation. *Annu. Rev. Genomics Hum. Genet.* **17**, 17-43 (2016).

5. Rowley, M.J. & Corces, V.G. The three-dimensional genome: principles and roles of long-distance interactions. *Curr. Opin. Cell Biol.* **40**, 8-14 (2016).

6. Szabo, Q., Bantignies, F. & Cavalli, G. Principles of genome folding into topologically associating domains. *Sci. Adv.* **5**, eaaw1668 (2019).

7. Kagey, M.H. et al. Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **467**, 430-435 (2010).

8. Dekker, J. & Mirny, L. The 3D Genome as Moderator of Chromosomal Communication. *Cell* **164**, 1110-1121 (2016).

9. Fosseprez, O. & Cuvier, O. Uncovering the functions and mechanisms of regulatory elements-associated non-coding RNAs. *Biochim Biophys Acta Gene Regul Mech* **1867**, 195059 (2024).

10. Hansen, A.S. et al. Distinct Classes of Chromatin Loops Revealed by Deletion of an RNA-Binding Region in CTCF. *Mol. Cell* **76**, 395-411 e13 (2019).

11. Saldana-Meyer, R. et al. RNA Interactions Are Essential for CTCF-Mediated Genome Organization. *Mol Cell* **76**, 412-422 e5 (2019).

12. Hansen, A.S., Amitai, A., Cattoglio, C., Tjian, R. & Darzacq, X. Guided nuclear exploration increases CTCF target search efficiency. *Nat. Chem. Biol.* **16**, 257-266 (2020).

13. Barshad, G. et al. RNA polymerase II dynamics shape enhancer-promoter interactions. *Nat Genet* **55**, 1370-1380 (2023).

14. Zhang, S., Ubelmesser, N., Barbieri, M. & Papantonis, A. Enhancer-promoter contact formation requires RNAPII and antagonizes loop extrusion. *Nat Genet.* **55**, 832-840 (2023).

15. Cai, Z. et al. RIC-seq for global in situ profiling of RNA-RNA spatial interactions. *Nature* **582**, 432-437 (2020).

16. Schmid, M. & Jensen, T.H. Controlling nuclear RNA levels. *Nat. Rev. Genet.* **19**, 518-529 (2018).

17. Lubas, M. et al. Interaction profiling identifies the human nuclear exosome targeting complex. *Mol. Cell* **43**, 624-637 (2011).

18. Meola, N. et al. Identification of a Nuclear Exosome Decay Pathway for Processed Transcripts. *Mol. Cell* **64**, 520-533 (2016).

19. Ogami, K. et al. An Mtr4/ZFC3H1 complex facilitates turnover of unstable nuclear RNAs to prevent their cytoplasmic transport and global translational repression. *Genes Dev.* **31**, 1257-1271 (2017).

20. Contreras, X. et al. PAPgamma associates with PAXT nuclear exosome to control the abundance of PROMPT ncRNAs. *Nat Commun* **14**, 6745 (2023).

21. Meola, N. & Jensen, T.H. Targeting the nuclear RNA exosome: Poly(A) binding proteins enter the stage. *RNA Biol.* **14**, 820-826 (2017).

22. Silla, T., Karadoulama, E., Makosa, D., Lubas, M. & Jensen, T.H. The RNA Exosome Adaptor ZFC3H1 Functionally Competes with Nuclear Export Activity to Retain Target Transcripts. *Cell Rep.* **23**, 2199-2210 (2018).

23. Yan, J. et al. Histone H3 lysine 4 monomethylation modulates long-range chromatin interactions at enhancers. *Cell Res* **28**, 204-220 (2018).

24. Schaaf, C.A. et al. Genome-wide control of RNA polymerase II activity by cohesin. *PLoS Genet* **9**, e1003382 (2013).

25. Cubenas-Potts, C. et al. Different enhancer classes in Drosophila bind distinct architectural proteins and mediate unique chromatin interactions and 3D architecture. *Nucleic Acids Res.* **45**, 1714-1730 (2017).

26. Liang, J. et al. Chromatin immunoprecipitation indirect peaks highlight long-range interactions of insulator proteins and Pol II pausing. *Mol Cell* **53**, 672-681 (2014).

27. Rao, S.S. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665-1680 (2014).

28. Lai, F. et al. Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. *Nature* **494**, 497-501 (2013).

29. Rahman, S. et al. Single-cell profiling reveals that eRNA accumulation at enhancer-promoter loops is not required to sustain transcription. *Nucleic Acids Res.* **45**, 3017-3030 (2017).

30. Contreras, X. et al. Nuclear RNA surveillance complexes silence HIV-1 transcription. *PLoS Pathog* **14**, e1006950 (2018).

31. Salifou, K. et al. Chromatin-associated MRN complex protects highly transcribing genes from genomic instability. *Sci Adv.* **7**(2021).

32. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009).

33. Stadelmayer, B. et al. Integrator complex regulates NELF-mediated RNA polymerase II pause/release and processivity at coding genes. *Nat. Commun.* **5**, 5531 (2014).

34. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis.*, (Springer-Verlag, New York, 2016).

35. Stempor, P. & Ahringer, J. SeqPlots - Interactive software for exploratory data analyses, pattern discovery and visualization in genomics. *Wellcome Open Res* **1**, 14 (2016).

36. Kinkley, S. et al. reChIP-seq reveals widespread bivalency of H3K4me3 and H3K27me3 in CD4(+) memory T cells. *Nat Commun.* **7**, 12514 (2016).

37. Bhardwaj, S., Schlackow, M., Rabajdova, M. & Gullerova, M. Transcription facilitates sister chromatid cohesion on chromosomal arms. *Nucleic Acids Res.* **44**, 6676-6692 (2016).

38. Schlackow, M. et al. Distinctive Patterns of Transcription and RNA Processing for Human lincRNAs. *Mol. Cell* **65**, 25-38 (2017).

39. Wutz, G. et al. Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. *EMBO J.* **36**, 3573-3599 (2017).

40. Heurteau, A. et al. Insulator-based loops mediate the spreading of H3K27me3 over distant micro-domains repressing euchromatin genes. *Genome Biol.* **21**, 193 (2020).

41. Durand, N.C. et al. Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. *Cell Syst* **3**, 99-101 (2016).

42. Andersson, R. et al. An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455-461 (2014).

43. Arner, E. et al. Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells. *Science* **347**, 1010-1014 (2015).

Acknowledgements

The work carried out in this study was supported by ERC CoG ‘RNAmedTGS’, MSD Avenir ‘EpiMum3D’, Sidaction, LabUM EpiGenMed, to RK, MSD Avenir ‘Hit-Hidden-HBV’ to OC, and ANR ‘Helico’ and ANRS (#229035) to OC and RK. CA was supported by a fellowship from ANRS and DD by scholarships from La Ligue Nationale Contre le Cancer (LNCC) and ANRS.

We wish to thank members of the Kiernan and Cuvier labs as well as Matthias Merkenschlager for helpful discussions.

Author contributions

XC and MH prepared ChIP-seq samples; CA performed Hi-C, CA and JB performed cell culture, RNAi and immunoblotting; AH, DD, SS, OF and OC performed data analysis; OC and RK conceived and supervised the project; OC and RK wrote the paper with input from all authors.

Competing Interests

The authors declare no competing interests

Additional Information

Correspondence and requests for materials should be addressed to Olivier Cuvier or Rosemary Kiernan.

Figures Legends

Fig. 1. Genome-wide identification of sites bound by MTR4, ZFC3H1 or ZCCHC8. a Browsershot showing ChIP-seq reads following precipitation with anti-RNAPII, anti-ZCCHC8, anti-ZFC3H1 or anti-MTR4 specific antibodies, over the indicated region of the human genome. **b** Venn diagram showing the intersection between sites bound by MTR4, ZFC3H1 or ZCCHC8, as indicated. P-values were calculated by Fisher's exact test. **c-e Top** Average plots of ChIP-seq reads of ZCCHC8, ZFC3H1 and MTR4, as indicated, at ZFC3H1, ZCCHC8 or MTR4 recruitment sites compared to control sites (see Methods). P-values were calculated using a Wilcoxon test for specific sites compared to control (random) sites. **bottom** Heatmaps centered and rank-ordered on normalized ChIP-seq reads of ZCCHC8, ZFC3H1 and MTR4, as indicated. ChIP-seq reads of ZCCHC8, ZFC3H1 or MTR4 were plotted respecting the same ranking, as indicated.

Fig. 2. Recruitment sites of MTR4 overlap with those of ZFC3H1 and ZCCHC8 but not with PROMPTs. **a** Histograms showing the genomic distribution of ChIP-seq peaks of ZFC3H1, ZCCHC8 and MTR4, as indicated, overlapping with transcription start sites (TSSs), transcription end sites (TESs), gene bodies or intergenic (not surrounding gene units +/- 2kb) regions. **b** Venn diagrams showing the intersection between PROMPTs detected in cells depleted of MTR4, ZFC3H1 or ZCCHC8, as indicated, and binding sites of MTR4, ZFC3H1 or ZCCHC8, as indicated (P-values were obtained using Fisher's exact test). **c** ChIP-seq heatmaps centered and rank-ordered by ZFC3H1 (left) or ZCCHC8 (right) signal at TSSs (+/- 10 kb). ChIP-seq reads of MTR4 were plotted respecting the same ranking, as indicated. **d** ChIP-seq heatmaps centered and rank-ordered by ZFC3H1 (left) or ZCCHC8 (right) signal at distant enhancer sites (+/- 10 kb). ChIP-seq reads of MTR4 were plotted respecting the same ranking, as indicated.

628 **Fig. 3. MTR4 influences PROMPTs through long-range interactions. a** Venn diagram
629 showing the intersection between binding sites of MTR4, Rad21 and the enhancer-associated
630 H3K4me1 histone mark. P-values were obtained using Fisher's exact test. **b** Top: Genomic
631 view of the binding profiles of RNAPII, MTR4, ZFC3H1 and Rad21. Bottom: 2D matrix
632 representing the levels of normalized long-range interactions within the region, determined
633 from Hi-C data. **c** Top: Schematic representation of the enrichment test performed for
634 enhancer-promoter LRIs, determined by systematic genome-wide interrogation of TSSs from
635 Hi-C data, with the indicated features. Bottom: Matrices showing enrichment for the presence
636 of PROMPTs, depending on levels of LRIs (ordered in quartiles) and co-localization of MTR4
637 with distant enhancers (**, p-value < 1e-5, ***, p-value < 1e-6, NS= not significant, Wilcoxon
638 test). **d** 2D APA plots representing normalized Hi-C counts surrounding the MTR4 site (middle
639 of the vertical y-axis) with sites surrounding TSSs (middle of the horizontal x-axis) from the
640 lowest and highest quartiles of LRIs represented in c.
641
642 **Fig. 4. MTR4 is associated with sites of cohesin-dependent long-range interactions and**
643 **modulates cohesin binding. a** 2D APA plots representing normalized Hi-C counts at
644 enhancers bound by MTR4 or not (left and right plot, respectively). Aggregation analysis was
645 performed between enhancers and all active TSSs within 1 Mbp. **b** Box plot showing the
646 distribution of long-range interactions along the bodies of genes contacted by distant
647 enhancers depending on the presence or absence of MTR4. The statistical variations were
648 tested by Wilcoxon test (*** p-value < 1e-6). **c** Heatmaps centred on MTR4 binding sites -/+ 5
649 kB, and rank ordered on differential Rad21 ChIP-seq reads in cells depleted of MTR4
650 (shMTR4) compared to control cells (shC) (left). MTR4 ChIP-seq reads were plotted respecting
651 the same ranking (right) (*** p-value < 1e-6, Wilcoxon test). **d** Heatmaps rank ordered by
652 increased PROMPTs detected by RNA-seq upon depletion of MTR4 compared to controls,
653 and centred on TSSs -/+ 5 kB (left). Rad21 ChIP-seq reads were plotted respecting the same
654 ranking. **e** Heatmaps centred on distant MTR4 binding sites -/+ 5 kB, and rank ordered on

differential Rad21 ChIP-seq reads in cells depleted of ZFC3H1 (shZFC3H1) compared to control cells (shC) (left). MTR4 ChIP-seq reads were plotted respecting the same ranking (right) (*** p-value < 1e-6, Wilcoxon test). **f** Heatmaps of Rad21 upon depletion of ZFC3H1 compared to control, and centred on TSSs -/+ 5 kB (left). ZFC3H1 ChIP-seq reads were plotted respecting the same ranking (variation in Rad21 binding).

Fig. 5. Nuclear exosome targeting complexes influence cohesin-mediated long-range interactions. a 2D APA plots representing normalized Hi-C counts at distant MTR4 sites in control or MTR4-depleted conditions (left and right plot, respectively). Aggregation analysis was performed between enhancers and all active TSSs within 1 Mbp. **b** Box plot showing the distribution of long-range interactions at interaction sites (central area, left box plot) or along the bodies of genes contacted by distant enhancers (loop extrusion, right box plot) in control or MTR4-depleted conditions, as indicated. The statistical variations were tested by Wilcoxon test (*** p-value < 1e-6). **c** Browser shots of Hi-C signal over representative regions in control and MTR4 knock-down cells, as indicated. Genomic coordinates in GRCh37 are shown on left. **d** Schematic representing a model where cohesin-mediated looping is influenced by PAXT, NEXT and MTR4. In wild-type cells (WT), ncRNAs including PROMPTs and eRNAs are degraded by nuclear exosome complexes found at loop contact sites. MTR4 depleted cells (MTR4-KD) show accumulation of eRNAs and PROMPTs and stabilization of cohesin-mediated loops (thick black line). Note that eRNAs may be still degraded at enhancers since MTR4 binds to enhancers with PAXT/NEXT, independently of cohesin-mediated looping, in contrast to MTR4-mediated degradation of PROMPTs at promoters that is associated with chromatin looping.

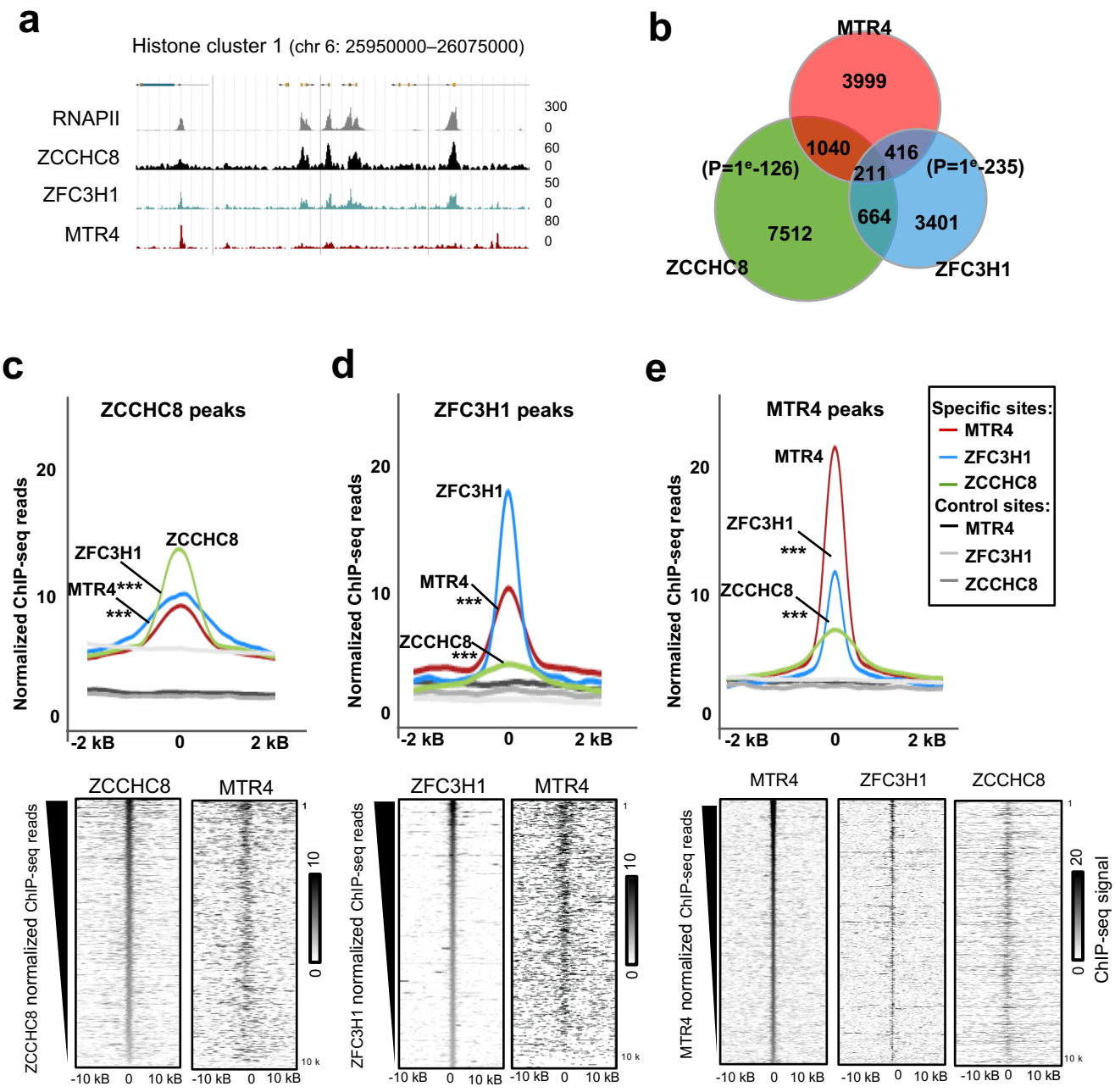
a
Histone cluster 1 (chr 6: 25950000–26075000)
RNAPII
ZCCHC8
ZFC3H1
MTR4
300
0
60
0
50
0
80
0
b
MTR4
3999
1040
416
211
(P=1e-126)
(P=1e-235)
664
3401
7512
ZCCHC8
ZFC3H1
c
ZCCHC8 peaks
Normalized ChIP-seq reads
ZFC3H1
ZCCHC8
MTR4***
***
-2 kB
0
2 kB
d
ZFC3H1 peaks
ZFC3H1
MTR4
***
ZCCHC8
***
e
MTR4 peaks
MTR4
ZFC3H1
***
ZCCHC8
***
Specific sites:
MTR4
ZFC3H1
ZCCHC8
Control sites:
MTR4
ZFC3H1
ZCCHC8
ZCCHC8
MTR4
ZCCHC8 normalized ChIP-seq reads
-10 kB
10 kB
ZFC3H1
MTR4
ZFC3H1 normalized ChIP-seq reads
MTR4
ZFC3H1
ZCCHC8
MTR4 normalized ChIP-seq reads
ChIP-seq signal

Fig. 1

Fig. 1. Genome-wide identification of sites bound by MTR4, ZFC3H1 or ZCCHC8. a Browsershot showing ChIP-seq reads following precipitation with anti-RNAPII, anti-ZCCHC8, anti-ZFC3H1 or anti-MTR4 specific antibodies, over the indicated region of the human genome. **b** Venn diagram showing the intersection between sites bound by MTR4, ZFC3H1 or ZCCHC8, as indicated. P-values were calculated by Fisher's exact test. **c-e Top** Average plots of ChIP-seq reads of ZCCHC8, ZFC3H1 and MTR4, as indicated, at ZFC3H1, ZCCHC8 or MTR4 recruitment sites compared to control sites (see Methods). P-values were calculated using a Wilcoxon test for specific sites compared to control (random) sites. **bottom** Heatmaps centered and rank-ordered on normalized ChIP-seq reads of ZCCHC8, ZFC3H1 and MTR4, as indicated. ChIP-seq reads of ZCCHC8, ZFC3H1 or MTR4 were plotted respecting the same ranking, as indicated.

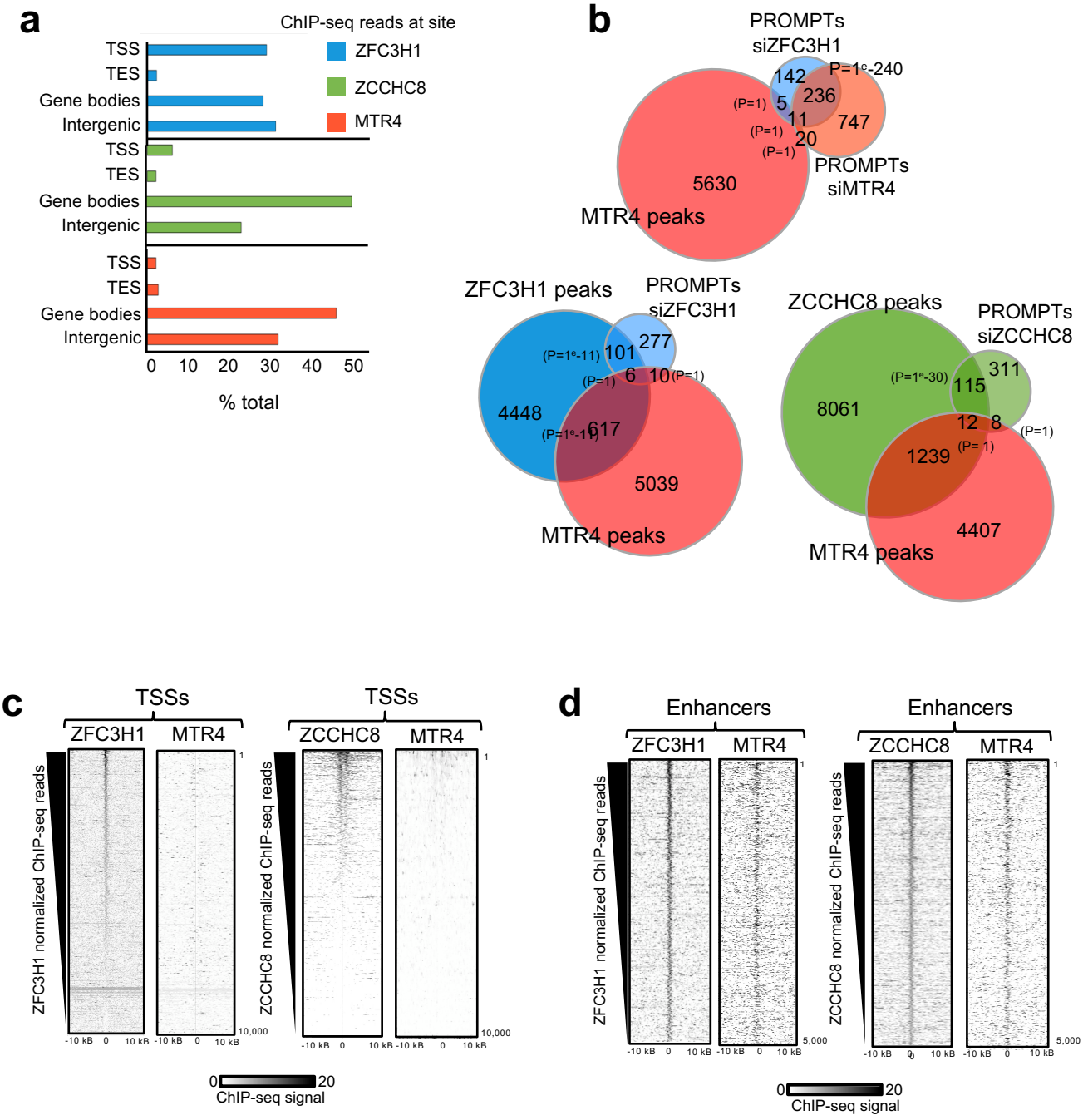
a
ChIP-seq reads at site
ZFC3H1
ZCCHC8
MTR4
TSS
TES
Gene bodies
Intergenic
0 10 20 30 40 50
% total
b
PROMPTs siZFC3H1
PROMPTs siMTR4
MTR4 peaks
142
236
747
5630
ZFC3H1 peaks
PROMPTs siZFC3H1
277
101
6
10
4448
617
5039
MTR4 peaks
ZCCHC8 peaks
PROMPTs siZCCHC8
311
115
8061
12
8
1239
4407
MTR4 peaks
c
TSSs
ZFC3H1
MTR4
ZCCHC8
ZFC3H1 normalized ChIP-seq reads
ZCCHC8 normalized ChIP-seq reads
-10 kB 0 10 kB
10,000
0 20
ChIP-seq signal
d
Enhancers
ZFC3H1
MTR4
ZCCHC8
ZFC3H1 normalized ChIP-seq reads
ZCCHC8 normalized ChIP-seq reads
5,000
0 20
ChIP-seq signal

Fig. 2

Fig. 2. Recruitment sites of MTR4 overlap with those of ZFC3H1 and ZCCHC8 but not with PROMPTs. **a** Histograms showing the genomic distribution of ChIP-seq peaks of ZFC3H1, ZCCHC8 and MTR4, as indicated, overlapping with transcription start sites (TSSs), transcription end sites (TESs), gene bodies or intergenic (not surrounding gene units +/- 2kb) regions. **b** Venn diagrams showing the intersection between PROMPTs detected in cells depleted of MTR4, ZFC3H1 or ZCCHC8, as indicated, and binding sites of MTR4, ZFC3H1 or ZCCHC8, as indicated (P-values were obtained using Fisher's exact test). **c** ChIP-seq heatmaps centered and rank-ordered by ZFC3H1 (left) or ZCCHC8 (right) signal at TSSs (+/- 10 kb). ChIP-seq reads of MTR4 were plotted respecting the same ranking, as indicated. **d** ChIP-seq heatmaps centered and rank-ordered by ZFC3H1 (left) or ZCCHC8 (right) signal at distant enhancer sites (+/- 10 kb). ChIP-seq reads of MTR4 were plotted respecting the same ranking, as indicated.

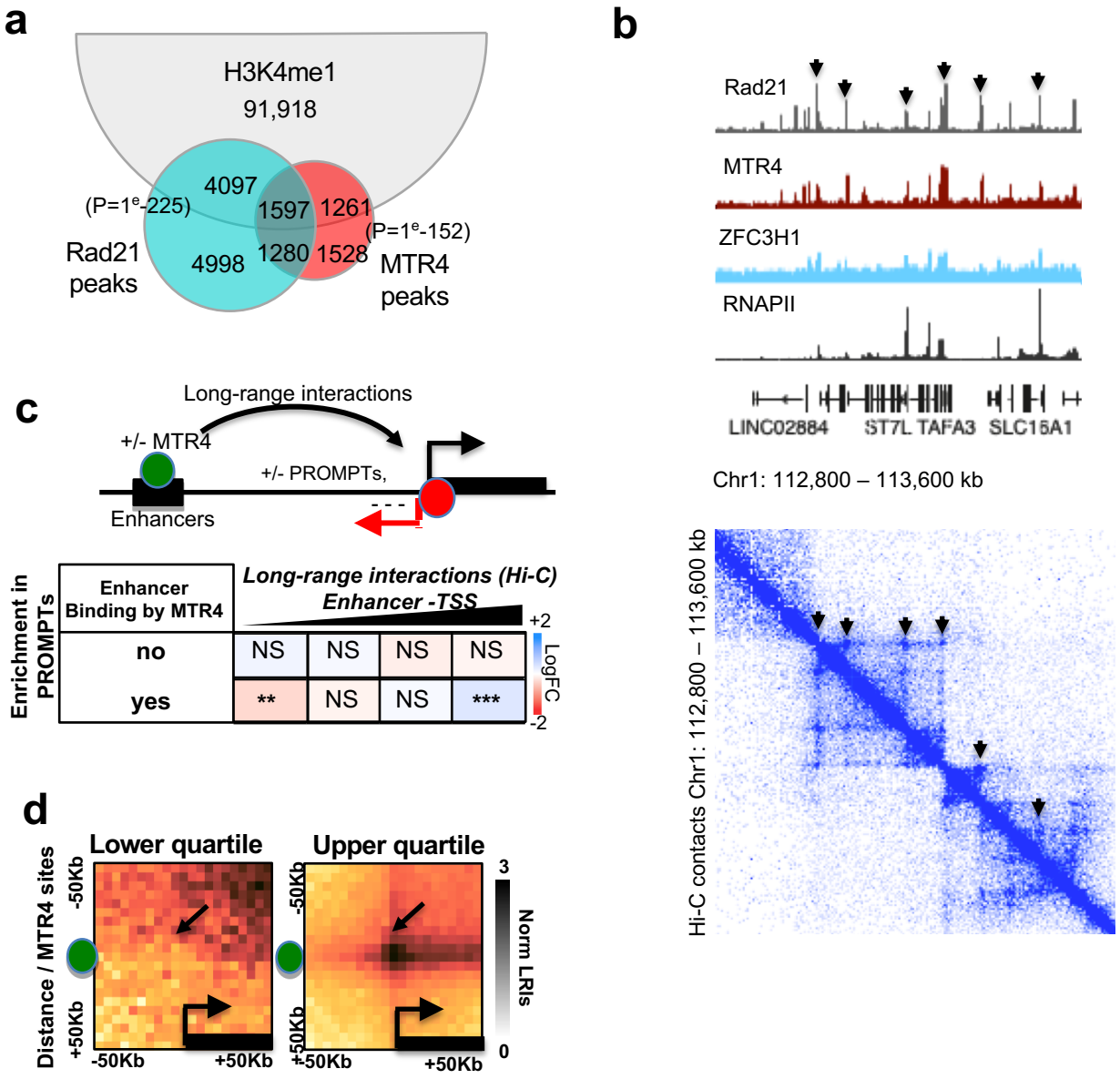
a
H3K4me1
91,918
4097
(P=1e-225)
1597
1261
(P=1e-152)
Rad21
peaks
4998
1280
1528
MTR4
peaks
b
Rad21
MTR4
ZFC3H1
RNAPII
LINC02884
ST7L
TAFA3
SLC15A1
Chr1: 112,800 – 113,600 kb
Hi-C contacts Chr1: 112,800 – 113,600 kb
c
Long-range interactions
+/- MTR4
+/- PROMPTs,
Enhancers
Enrichment in PROMPTs
Enhancer Binding by MTR4
Long-range interactions (Hi-C)
Enhancer -TSS
+2
-2
LogFC
no
yes
NS
**
***
d
Lower quartile
Upper quartile
Distance / MTR4 sites
-50Kb
+50Kb
3
0
Norm LRIs

Fig. 3

Fig. 3. MTR4 influences PROMPTs through long-range interactions. a Venn diagram showing the intersection between binding sites of MTR4, Rad21 and the enhancer-associated H3K4me1 histone mark. P-values were obtained using Fisher's exact test. **b** Top: Genomic view of the binding profiles of RNAPII, MTR4, ZFC3H1 and Rad21. Bottom: 2D matrix representing the levels of normalized long-range interactions within the region, determined from Hi-C data. **c** Top: Schematic representation of the enrichment test performed for enhancer-promoter LRIs, determined by systematic genome-wide interrogation of TSSs from Hi-C data, with the indicated features. Bottom: Matrices showing enrichment for the presence of PROMPTs, depending on levels of LRIs (ordered in quartiles) and co-localization of MTR4 with distant enhancers (**, p-value < 1e-5, ***, p-value < 1e-6, NS= not significant, Wilcoxon test). **d** 2D APA plots representing normalized Hi-C counts surrounding the MTR4 site (middle of the vertical y-axis) with sites surrounding TSSs (middle of the horizontal x-axis) from the lowest and highest quartiles of LRIs represented in c.

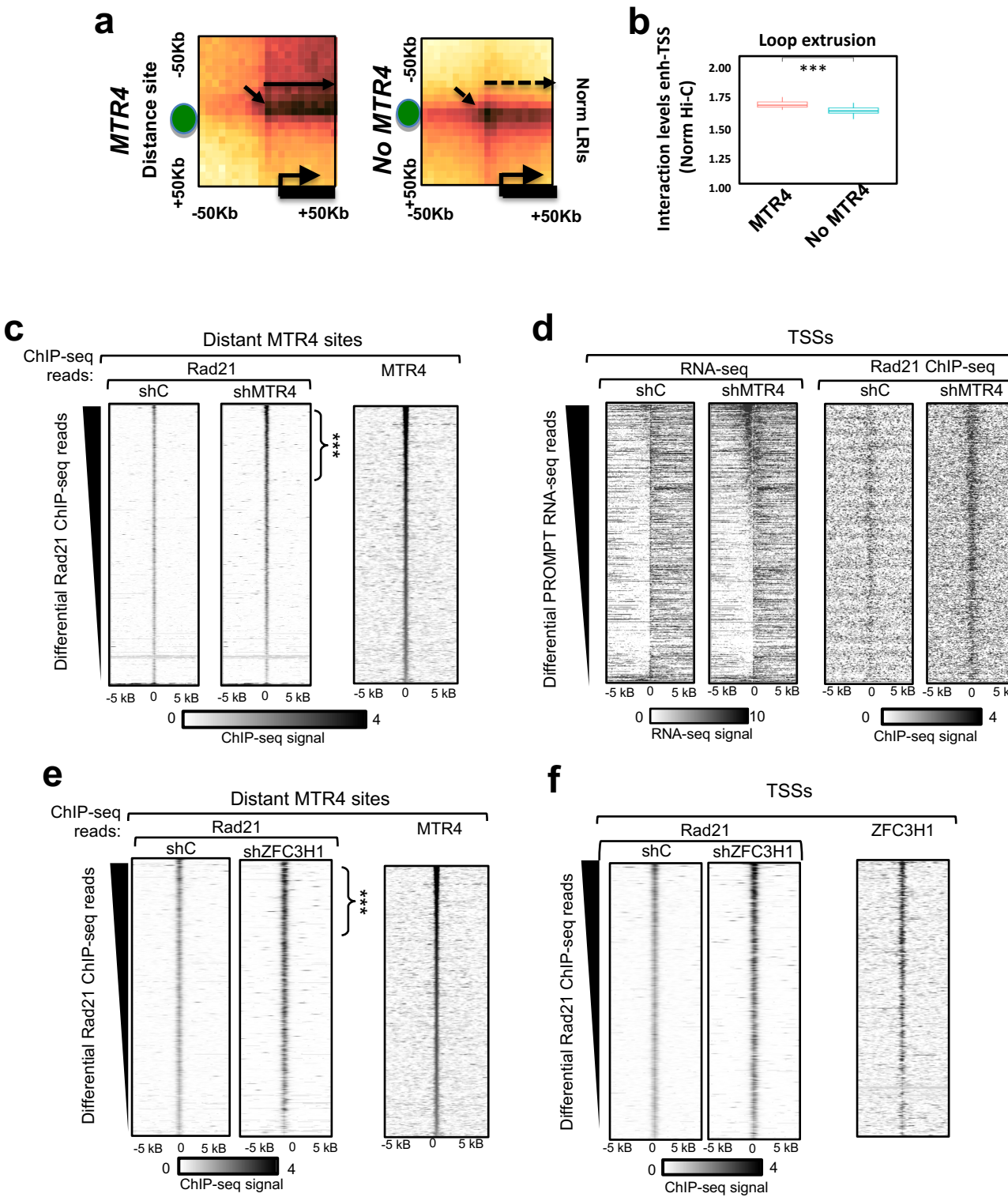
a
MTR4
Distance site
-50Kb
+50Kb
-50Kb
+50Kb
No MTR4
Norm LRIs
b
Loop extrusion
Interaction levels enh-TSS (Norm Hi-C)
***
MTR4
No MTR4
c
Distant MTR4 sites
ChIP-seq reads:
Rad21
MTR4
shC
shMTR4
Differential Rad21 ChIP-seq reads
***
-5 kB 0 5 kB
ChIP-seq signal
d
TSSs
RNA-seq
Rad21 ChIP-seq
Differential PROMPT RNA-seq reads
RNA-seq signal
e
Distant MTR4 sites
shZFC3H1
f
TSSs
ZFC3H1

Fig. 4

Fig. 4. MTR4 is associated with sites of cohesin-dependent long-range interactions and modulates cohesin binding. a 2D APA plots representing normalized Hi-C counts at enhancers bound by MTR4 or not (left and right plot, respectively). Aggregation analysis was performed between enhancers and all active TSSs within 1 Mbp. **b** Box plot showing the distribution of long-range interactions along the bodies of genes contacted by distant enhancers depending on the presence or absence of MTR4. The statistical variations were tested by Wilcoxon test (*** p-value < 1e-6). **c** Heatmaps centred on MTR4 binding sites -/+ 5 kB, and rank ordered on differential Rad21 ChIP-seq reads in cells depleted of MTR4 (shMTR4) compared to control cells (shC) (left). MTR4 ChIP-seq reads were plotted respecting the same ranking (right) (*** p-value < 1e-6, Wilcoxon test). **d** Heatmaps rank ordered by increased PROMPTs detected by RNA-seq upon depletion of MTR4 compared to controls, and centred on TSSs -/+ 5 kB (left). Rad21 ChIP-seq reads were plotted respecting the same ranking. **e** Heatmaps centred on distant MTR4 binding sites -/+ 5 kB, and rank ordered on differential Rad21 ChIP-seq reads in cells depleted of ZFC3H1 (shZFC3H1) compared to control cells (shC) (left). MTR4 ChIP-seq reads were plotted respecting the same ranking (right) (*** p-value < 1e-6, Wilcoxon test). **f** Heatmaps of Rad21 upon depletion of ZFC3H1 compared to control, and centred on TSSs -/+ 5 kB (left). ZFC3H1 ChIP-seq reads were plotted respecting the same ranking (variation in Rad21 binding).

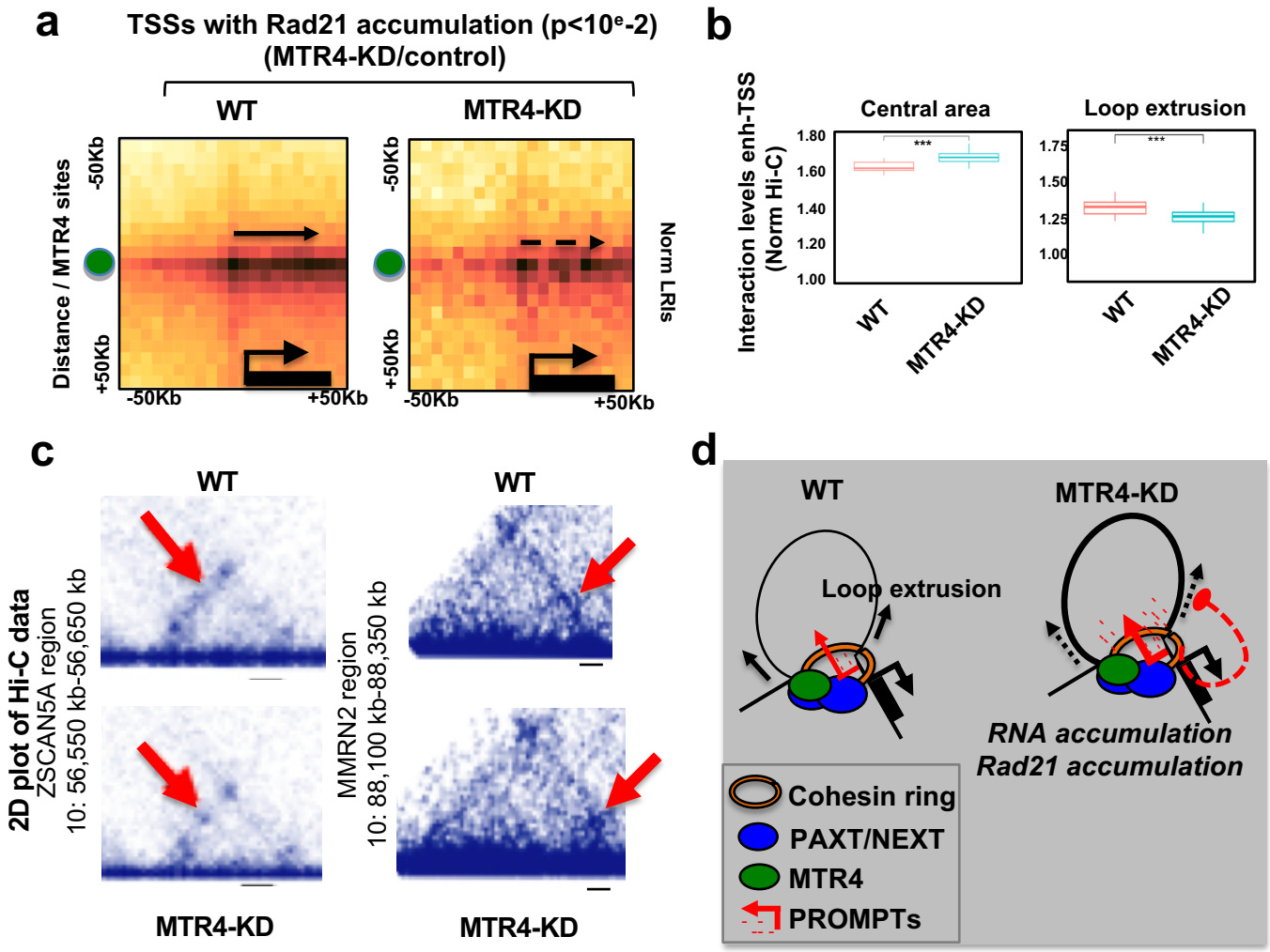
a
TSSs with Rad21 accumulation (p<10e-2)
(MTR4-KD/control)
WT
MTR4-KD
Distance / MTR4 sites
-50Kb
+50Kb
-50Kb
+50Kb
-50Kb
+50Kb
-50Kb
+50Kb
Norm LRIs
b
Central area
Loop extrusion
Interaction levels enh-TSS
(Norm Hi-C)
***
***
1.80
1.60
1.40
1.20
1.00
1.75
1.50
1.25
1.00
WT
MTR4-KD
WT
MTR4-KD
c
WT
WT
2D plot of Hi-C data
ZSCAN5A region
10: 56,550 kb-56,650 kb
MMRN2 region
10: 88,100 kb-88,350 kb
MTR4-KD
MTR4-KD
d
WT
MTR4-KD
Loop extrusion
RNA accumulation
Rad21 accumulation
Cohesin ring
PAXT/NEXT
MTR4
PROMPTs

Fig. 5

Fig. 5. Nuclear exosome targeting complexes influence cohesin-mediated long-range interactions. a 2D APA plots representing normalized Hi-C counts at distant MTR4 sites in control or MTR4-depleted conditions (left and right plot, respectively). Aggregation analysis was performed between enhancers and all active TSSs within 1 Mbp. **b** Box plot showing the distribution of long-range interactions at interaction sites (central area, left box plot) or along the bodies of genes contacted by distant enhancers (loop extrusion, right box plot) in control or MTR4-depleted conditions, as indicated. The statistical variations were tested by Wilcoxon test (*** p-value < 1e-6). **c** Browser shots of Hi-C signal over representative regions in control and MTR4 knock-down cells, as indicated. Genomic coordinates in GRCh37 are shown on left. **d** Schematic representing a model where cohesin-mediated looping is influenced by PAXT, NEXT and MTR4. In wild-type cells (WT), ncRNAs including PROMPTs and eRNAs are degraded by nuclear exosome complexes found at loop contact sites. MTR4 depleted cells (MTR4-KD) show accumulation of eRNAs and PROMPTs and stabilization of cohesin-mediated loops (thick black line). Note that eRNAs may be still degraded at enhancers since MTR4 binds to enhancers with PAXT/NEXT, independently of cohesin-mediated looping, in contrast to MTR4-mediated degradation of PROMPTs at promoters that is associated with chromatin looping.