HHS Public Access

Author manuscript

Nat Struct Mol Biol. Author manuscript; available in PMC 2016 May 18.

Published in final edited form as:

Nat Struct Mol Biol. 2015 December ; 22(12): 999–1007. doi:10.1038/nsmb.3122.

R-loops regulate promoter-proximal chromatin architecture and cellular differentiation

Poshen B. Chen[1,2], **Hsiuyi V. Chen**[3], **Diwash Acharya**[1,2], **Oliver J. Rando**[3], and **Thomas G. Fazzio**[1,2,4]

[1]Department of Molecular, Cell, and Cancer Biology, University of Massachusetts Medical School, Worcester, MA, USA

[2]Program in Molecular Medicine, University of Massachusetts Medical School, Worcester, MA, USA

[3]Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA, USA

Abstract

Numerous chromatin-remodeling factors are regulated by interactions with RNA, although the contexts and functions of RNA binding are poorly understood. Here we show that R-loops, RNA:DNA hybrids consisting of nascent transcripts hybridized to template DNA, modulate the binding of two key chromatin regulatory complexes, Tip60–p400 and polycomb repressive complex 2 (PRC2) in mouse embryonic stem cells (ESCs). Like PRC2, the Tip60–p400 histone acetyltransferase complex binds to nascent transcripts, but unlike PRC2, transcription promotes chromatin binding by Tip60–p400. Interestingly, we observed higher Tip60–p400 and lower PRC2 levels at genes marked by promoter-proximal R-loops. Furthermore, disruption of R-loops broadly reduced Tip60–p400 and increased PRC2 occupancy genome-wide. Consistent with these alterations, ESCs with reduced R-loops exhibited impaired differentiation. These results show that R-loops act both positively and negatively to modulate the recruitment of key pluripotency regulators.

INTRODUCTION

With the discovery of thousands of long non-coding RNAs (lncRNAs) that are expressed in mammalian cells, a considerable effort is underway to uncover the roles of specific lncRNAs in the nucleus, as well as to elucidate broadly generalizable mechanisms of action that

[4]Correspondence to: ; Email: thomas.fazzio@umassmed.edu

ACCESSION CODES

Sequencing data have been deposited in the Gene Expression Omnibus under accession number GSE67584.

The authors declare no competing interests.

AUTHOR CONTRIBUTIONS

P.B.C. performed all experiments and bioinformatics analyses, except as follows: H.V.C. prepared libraries for RIP-seq, D.A. generated several of the ESC lines, and T.G.F. performed Suz12 ChIP-seq and analysed the data. P.B.C., O.J.R., and T.G.F. designed the experiments. P.B.C. and T.G.F. wrote the paper with input from all authors.

govern their biological functions. LncRNAs function both *in cis* and *in trans* to regulate gene expression[1,2], raising the possibility that these transcripts act specifically to modulate the functions of individual transcription factors, the general transcription machinery, or other regulatory proteins. Indeed, numerous lncRNAs have been shown to interact with transcriptional regulatory proteins, consistent with this hypothesis[1–3].

Interestingly, in a survey of 74 lncRNAs expressed in ESCs, several chromatin regulatory complexes with key roles in ESC pluripotency were shown to bind lncRNAs[4]. Multiple complexes bound to more than 30% of lncRNAs tested, and numerous lncRNAs were bound by more than one complex, suggesting that either these factors are differentially regulated by dozens of individual lncRNAs, or these complexes bind lncRNAs relatively non-specifically. In the latter scenario, the distinct sequence of each lncRNA bound by a complex would not be predicted to impart a unique function (such as targeting the complex to specific genomic loci), but lncRNA binding in general may serve some structural or regulatory role within the complex.

Among the first chromatin regulatory complexes shown to bind lncRNAs was polycomb repressive complex 2 (PRC2)[5–7], a highly conserved histone H3 lysine-27 methyltransferase complex important for gene silencing during development[8]. PRC2 binding to the A-repeat of the *Xist* lncRNA is thought to play a role in recruitment of the complex to the inactive X-chromosome[6,9]. In addition to interacting with lncRNAs, PRC2 binds promiscuously to nascent RNA transcripts expressed from thousands of genes, and the level of RNA binding by the PRC2 catalytic subunit Ezh2 correlates with RNA abundance[10,11]. At first glance, PRC2 binding of nascent transcripts from active genes appears to conflict with models in which lncRNA-dependent PRC2 recruitment promotes gene silencing. However, RNA binding by PRC2 has been shown to inhibit its histone H3 lysine-27 methyltransferase activity[9,12]. Consistent with these findings, PRC2 components bind to both silent and active genes, and active genes bound by PRC2 are not marked by H3K27me3[10,11]. These findings support a revised model in which binding of nascent transcripts at active genes helps recruit PRC2 to these loci, but maintains the complex in an inactive state[9,12]. In this model, PRC2 is poised to generate repressive chromatin structure and enforce silencing at these genes at a later time, should their expression be silenced by an independent mechanism. On the other hand, chemical inhibition of transcription promotes binding of PRC2 to CpG islands (including numerous promoter-proximal regions) throughout the genome, arguing against a model in which nascent transcripts are necessary for recruitment of PRC2[13]. Therefore, the roles of nascent transcripts in regulation of PRC2 binding and chromatin structure appear to be complex and context-specific.

Tip60–p400 is another chromatin-remodeling complex with essential functions in ESC self-renewal and pluripotency reported to bind lncRNAs[4]. Tip60–p400 comprises a 17 subunit chromatin-remodeling complex with two catalytic subunits: the Tip60 (also known as Kat5) protein lysine acetyltransferase, which acetylates multiple lysines on histones H4 and H2A, among other proteins, and the p400 ATPase, which incorporates the H2A.Z histone variant into chromatin[14]. We previously found that Tip60–p400 was essential for normal ESC self-renewal and pluripotency, acting simultaneously to repress some differentiation genes and activate proliferation genes[15,16]. Although it is not clear how Tip60–p400 simultaneously

activates one group of genes and silences another, interaction with lncRNAs could potentially target the complex to specific regions of the genome and/or tune its catalytic activities at specific targets to favor activation or silencing.

Here, we address the role of RNA binding by Tip60–p400 in mouse ESCs. We find that, like PRC2, Tip60–p400 binds promiscuously to nascent RNAs from both coding and non-coding genes. However, unlike PRC2, whose binding to chromatin is inhibited by transcription[13], transcription promotes Tip60–p400 binding to many of its target promoters. Interestingly, we find that Tip60–p400 binding to many target genes is enhanced by promoter-proximal R-loops, RNA:DNA hybrid structures formed when G-rich sequences on RNA hybridize with their DNA template[17,18]. In contrast, binding of the PRC2 complex and histone H3 lysine-27 methylation were inhibited by R-loops. These results demonstrate that R-loops play a major role in regulation of chromatin structure near the 5′ regulatory regions of thousands of genes in ESCs, acting both positively and negatively to control binding of chromatin-remodeling factors. More broadly, these findings suggest that RNA binding can have different effects on chromatin regulators, depending on the molecular context in which the RNA is presented.

RESULTS

Tip60–p400 interacts with nascent transcripts

Previously, in a survey of chromatin-remodeling complexes with key roles in ESCs, Guttman et al. found that Tip60–p400 interacts with 9 of 74 long non-coding RNAs (lncRNAs) tested[4], raising the possibility that lncRNAs might be important for interaction of the complex with chromatin or remodeling of chromatin structure by the complex. Alternatively, Tip60–p400 might bind promiscuously to RNA, as shown for the well-studied chromatin regulatory complex, PRC2[10,11,19]. To distinguish between these possibilities, we first performed unbiased identification of Tip60–p400-interacting transcripts by deep sequencing of RNAs that co-immunoprecipitate with Tip60–p400 (RIP-seq). We performed biological replicate IPs of two different Tip60–p400 subunits, p400 and Ruvbl1, and observed significant correlations between replicates (Supplementary Fig. 1a, b). To elucidate the set of high-confidence Tip60–p400-binding RNAs, we focused on those enriched greater than two-fold in both replicates of p400 and Ruvbl1 RIPs compared to control RIPs, identifying approximately 2,500 transcripts in this category (Fig. 1a–d). Among these, we identified 608 enriched lncRNAs (Fig. 1c), confirming that Tip60–p400 binds to non-coding transcripts in ESCs. More interestingly, we observed that Tip60–p400 also interacts with 1,909 coding gene transcripts (Fig. 1d), suggesting Tip60–p400 does not bind specifically to lncRNAs, but rather interacts with a broad array of both coding and noncoding transcripts in ESCs.

Next, we considered whether this complex might interact with nascent transcripts and therefore examined the genomic locations of reads within our RIP-seq libraries. Aggregation of reads from p400 and Ruvbl1 RIP-seq experiments revealed significant peaks of interacting transcripts just downstream of transcription start sites (TSSs), compared to lower (but above background) levels near transcriptional termination sites (TTSs) (Fig. 1e–f). Consistent with this observation, we observed a significant overrepresentation of reads within the first exon and first intron of Tip60–p400-interacting RNAs (Supplementary Fig.

1c), suggesting the complex interacts with unspliced (pre-mRNA) transcripts. This pattern was observed in both biological replicates of each RIP, although the relative heights and locations of RIP peaks were somewhat variable (Fig. 1g), suggesting Tip60–p400-bound pre-mRNAs may be heterogeneous. Finally, when we counted all reads within each gene rather than only those within spliced mRNAs, we observed greater enrichment of interacting RNAs in p400 and Ruvbl1 RIPs relative to controls (Fig. 1h–i). We therefore conclude that Tip60–p400, like PRC2, binds primarily to nascent transcripts near their initiation sites.

Transcription promotes chromatin binding by Tip60–p400

To dissect the role of RNA binding by Tip60–p400, we first tested whether the complex binds to the same regions of chromatin from which Tip60–p400-interacting RNAs are transcribed. To this end, we compared ChIP-seq maps of Tip60 and p400 localization near annotated TSSs to the set of RNAs bound by the complex. We observed significantly higher levels of Tip60 and p400 enrichment near the promoters of genes from which Tip60–p400-interacting RNAs are transcribed (Fig. 2a, b), and significant overrepresentation of Tip60–p400-target genes within the set of Tip60–p400-bound transcripts (Supplementary Fig. 1d), suggesting Tip60–p400 binds numerous transcripts *in cis*. These transcripts occupied a broad range of expression levels and functional categories (Supplementary Fig. 1e, f), consistent with the diverse set of genes bound and regulated by Tip60–p400[15,16].

These data suggested that interaction with RNA may promote chromatin binding by Tip60–p400. To address this possibility, we tested whether transcription was required for interaction of Tip60–p400 with its target genes by addition of transcription inhibitors DRB or Triptolide to culture media for 9 or 4 hours, respectively (optimization of treatment time described in Methods). Inhibition of transcription reduced the abundance of some short-lived transcripts, but did not affect protein levels of any of several Tip60–p400 subunits tested (Supplementary Fig. 2a, b). Interestingly, both inhibitors significantly reduced Tip60 and p400 binding to many of their genomic targets (Fig. 2c–f). We validated these data by ChIP-qPCR at several targets, obtaining results consistent with the genome-level data (Supplementary Fig. 2c). Together, these data demonstrate that binding of nascent transcripts by Tip60–p400, the act of transcription itself, or both, contribute to binding of the complex to many of its target genes in ESCs.

R-loops enhance promoter-proximal Tip60–p400 binding

Nascent transcripts can form R-loops near the 5′ and 3′ ends of transcribed genes in multiple cell types[17,18,20–22]. Although unresolved R-loops induce DNA damage and genomic instability[23–27], 3′ R-loops regulate transcription termination[18,20,21] and R-loop formation over CpG islands functions to keep these regions relatively free of DNA methylation[17,18]. Furthermore, R-loops have been implicated in regulation of chromosome condensation[28], regulation of sense-antisense transcript pairs[29], and other processes[30,31]. Since Tip60–p400 binds primarily near the 5′ ends of transcripts, we considered the possibility that Tip60–p400 binds to nascent transcripts in the form of R-loops, and that 5′ R-loops may play a role in recruitment or stabilization of Tip60–p400 binding at these loci.

To test this possibility, we first mapped the locations of R-loops across the genome of mouse ESCs. Immunoprecipitation of RNA:DNA hybrids using an antibody (S9.6) specific for these structures coupled to either quantitative PCR (DRIP-qPCR) or deep sequencing (DRIP-seq) has been used to map R-loops in multiple cell types[17,21,22]. To reduce the background and identify more precise boundaries of R-loops mapped using this technique (Supplementary Fig. 3a), we modified the DRIP-seq protocol to sequence only RNAs enriched within immunoprecipitates of RNA-DNA hybrids (Supplementary Fig. 3b, Methods). Using this DRIP-RNA-seq approach, we observed R-loops near the 5′ ends of 10,595 genes and the 3′ ends of 9,151 genes (Fig. 3a–d). Although R-loops were, in aggregate, elevated at highly expressed genes, we also observed R-loop formation at the 5′ ends of some low or moderately expressed genes (Fig. 3a, and compare lowly expressed, R-loop marked *Wipf2* to more highly expressed genes without R-loops in Fig. 3d). We confirmed the specificity of DRIP signals in two ways: First, signals were significantly reduced when samples were treated with RNaseH (which degrades RNA within RNA:DNA hybrids) prior to immunoprecipitation (Fig. 3a–d). Second, in our strand-specific DRIP-RNA-seq libraries, we observed mainly sense strand reads (Fig. 3d; Supplementary Fig. 3c).

Interestingly, we observed a high incidence of R-loops at Tip60–p400 target genes (Fig. 3e), and higher average enrichment of Tip60 and p400 at genes with associated R-loops than those without (Fig. 3f–g), consistent with the possibility that R-loops promote Tip60–p400 binding. To test this hypothesis, we utilized *Rnaseh1* overexpression in ESCs to disrupt R-loop formation. Overexpression of the RNaseH1 protein in multiple organisms is known to disrupt R-loops throughout the genome[21,25–27,29,32]. We found that overexpression of *Rnaseh1* in ESCs reduced bulk RNA:DNA hybrids approximately four-fold (Supplementary Fig. 3d). Interestingly, we observed a reduction in both Tip60 and p400 localization to most Tip60–p400 target genes in *Rnaseh1* overexpressing cells (Fig. 4a; Supplementary Fig. 4). At genes with high-confidence R-loops, we found that Tip60 binding was reduced an average of 63% (from peak to baseline) upon *Rnaseh1* overexpression (Fig. 4b). Tip60–p400 binding to genes lacking high-confidence R-loops was also reduced upon *Rnaseh1* overexpression, albeit to a lesser extent. Similar results were observed for p400 (Fig. 4c). These data indicate that high-confidence R-loop containing genes are bound at higher levels by Tip60–p400 in control cells, and exhibit a greater reduction in binding upon *Rnaseh1* overexpression. However, the smaller but significant reduction in binding at genes without high-confidence R-loops suggests that some of these genes have R-loops at levels below our detection threshold, some binding events might be indirectly affected by *Rnaseh1* expression, or both. We validated these data at a selection of Tip60–p400 targets by ChIP-qPCR (Supplementary Fig. 5a). Together, these data suggest that R-loops enhance chromatin association by Tip60–p400 complex.

Since RNA:DNA hybrids play roles in DNA replication, rRNA expression, and other processes[30,31], we tested the possibility that indirect effects of *Rnaseh1* overexpression may effect the interpretation of these data. We observed minimal effects of *Rnaseh1* overexpression on most cellular functions impacted by RNA:DNA hybrids: *Rnaseh1*-overexpressing ESCs self-renew normally (Supplementary Fig. 6a), and exhibit no apparent alterations in their cell cycle (Supplementary Fig. 6b–c), or rRNA levels (Supplementary Fig. 6d). *Rnaseh1* overexpression results in slower proliferation relative to control cells,

although this defect is less severe than in *Ep400* (the gene encoding the p400 protein) mutant ESCs generated by CRISPR/Cas9 cleavage and error-prone repair[33_35] (Supplementary Fig. 6e). To test the possibility that degradation of RNA:DNA hybrids may inhibit transcription, we examined the effects of *Rnaseh1* overexpression on promoter-proximal and gene body-associated RNAPII, observing no reduction in RNAPII association at either location (Supplementary Fig. 6f).

Although *Rnaseh1* overexpression directly disrupts R-loop formation by degrading RNAs within RNA:DNA hybrids, genome-wide disruption of R-loops can be indirectly achieved by global inhibition of transcription by RNAPII (Supplementary Fig. 3d). Since *Rnaseh1* overexpression does not inhibit transcription (Supplementary Fig. 6f), any potential indirect effects of transcription inhibitors and *Rnaseh1* overexpression are likely to be different. Therefore, if R-loops promote Tip60–p400 binding to chromatin, the sets of genes with reduced Tip60–p400 binding upon Rnaseh1 overexpression or treatment of cells with transcription inhibitors should significantly overlap. Consistent with this possibility, we observed significant overlap among genes with reduced Tip60 or p400 binding due to DRB treatment, Triptolide treatment, or *Rnaseh1* overexpression (Fig. 4d–g). We therefore conclude that promoter-proximal R-loops enhance Tip60–p400 binding to a large fraction of its target genes.

R-loops inhibit chromatin binding and methylation by PRC2

To test whether promoter-proximal R-loops function solely in Tip60–p400 recruitment, or are required for chromatin binding by additional regulatory complexes, we focused on PRC2, due to its established RNA-binding activity in multiple cell types[5_7,10,11,36]. Like Tip60–p400, PRC2 binds to nascent transcripts[10,11], the substrates for R-loop formation, consistent with the possibility that R-loops might promote PRC2 binding. However, since inhibition of transcription stimulates PRC2 association with chromatin[13], it was also possible that R-loops might inhibit PRC2 binding to a portion of its target genes, or have no effect at all. To distinguish among these possibilities, we first compared our maps of promoter-proximal R-loops to ChIP-seq maps of PRC2 subunit Suz12. Interestingly, DRIP-RNA-seq reads were poorly enriched near the promoter-proximal regions of genes highly bound by Suz12 (Fig. 5a), suggesting that moderate to high levels of promoter-proximal R-loops may inhibit PRC2 association. We tested this possibility directly by mapping Suz12 binding and H3K27me3 localization in the presence or absence of *Rnaseh1* overexpression, observing increased Suz12 and H3K27me3 occupancy in *Rnaseh1* overexpressing ESCs (Fig. 5b–e, Supplementary Fig. 7a). Some genes not significantly bound by Suz12 in control cells gained peaks of Suz12 binding (Fig. 5f–g), and Suz12 enrichment at promoter-proximal regions normally bound by the complex increased two-fold in aggregate upon *Rnaseh1* overexpression (Fig. 5b, Supplementary Fig. 7b–c). Consistent with these data, we confirmed a significant increase in Suz12 occupancy upon *Rnaseh1* overexpression by ChIP-qPCR (Supplementary Fig. 5b).

PRC2 binds strongly to relatively unmethylated CpG islands[37_39], which make up a large fraction of mammalian promoters and regulatory elements. CpG islands are kept unmethylated, in part, by the presence of R-loops[17,18], suggesting R-loops may help recruit

PRC2 complex to these regions. However, we observed a significant increase in Suz12 association with CpG islands in *Rnaseh1* overexpressing cells (Fig. 5e), suggesting that R-loops produced from nascent transcripts inhibit PRC2 binding to these sites. Finally, we observed examples of genes bound by Tip60–p400 complex in control ESCs that, upon disruption of R-loops by *Rnaseh1* overexpression, exhibited reduction of Tip60–p400 binding to background levels and ectopic PRC2 binding, representing a substantial restructuring of their chromatin architecture (Supplementary Fig. 7d). Together, these data reveal that R-loop formation contributes to differential recruitment of chromatin regulatory complexes at thousands of genes in ESCs, promoting Tip60–p400 association and inhibiting PRC2 association with numerous R-loop-associated genes.

R-loops are necessary for robust ESC differentiation

Knockdown of *Kat5* or *Ep400* (encoding the Tip60 and p400 proteins) in ESCs results in partial defects in both ESC self-renewal and differentiation[15,16]. In addition, knockdown of the *Hdac6* gene, which encodes a cell type-specific Tip60–p400 binding protein, results in a partial loss of Tip60–p400 binding to many target genes, and a defect in ESC differentiation, but has no effect on self-renewal[16]. These findings raise the possibility that R-loop-deficient ESCs might also be defective in differentiation. To test this possibility, *Rnaseh1* overexpressing ESCs were grown in differentiation medium alongside control ESCs and homozygous *Ep400* mutant ESCs (see Methods for details). Consistent with the differentiation defect previously observed upon KD of *Ep400* or other Tip60–p400 subunits[15,16], we observed a higher abundance of *Ep400* mutant cells with clustered (ESC-like) morphology after 14 days that stained positive for alkaline phosphatase (Fig. 6a) and the ESC-specific transcription factor Nanog (Fig. 6b), relative to control cells. Interestingly, we also observed an increase in both alkaline phosphatase and Nanog staining upon *Rnaseh1* overexpression (Fig. 6a–b). As a more stringent test of ESC differentiation, we examined the ability of *Rnaseh1* overexpressing ESCs to form teratomas with differentiated cell types from all three germ layers when injected into nude mice. As previously observed upon knockdown of the gene encoding Tip60–p400 subunit *Dmap1*[15], *Rnaseh1* overexpression resulted in smaller teratomas (Fig. 6c), which were poorly differentiated relative to control cells (Fig. 6d). Together, these data suggest one major role of R-loops in ESCs is to enable their efficient response to differentiation cues, in part by promoting high levels of Tip60–p400 association and limiting levels of PRC2 association with specific sets of target genes. However, it is also possible that disruption of R-loops by overexpression of *Rnaseh1* causes additional, Tip60–p400- and PRC2-independent perturbations that impair ESC differentiation.

DISCUSSION

In mammalian cells, R-loops are most abundant at the 5′ ends of genes with G-rich transcripts, as well as near Pol II pause sites at transcriptional termini[17,18,40]. In addition, formation of R-loops *in trans* has been observed in some systems[41], which may contribute to the functions of some lncRNAs[20]. Several proteins that resolve or stabilize R-loops have been described, suggesting formation and persistence of R-loops is highly regulated[42]. Thus, R-loop accumulation appears to be a function of both the transcription and sequence of RNA

species, along with trans-acting factors. It remains to be determined how the positions and abundance of R-loops change in different cell types or during cellular differentiation.

Here, we have uncovered a role for R-loops in shaping the chromatin landscape and controlling the differentiation program in ESCs. We show that R-loops promote elevated levels of promoter-proximal chromatin binding by Tip60–p400, but inhibit binding of PRC2 to its targets. Therefore, with regards to these key regulators of ESC pluripotency, R-loops help segregate genes into classes that are highly bound by Tip60–p400 but not PRC2, those highly bound by PRC2 but not Tip60–p400, and some genes that are lowly bound by both complexes (Fig. 7). Interestingly, at some genes with low DRIP-RNA-seq signals, we see a significant increase in Suz12 binding upon *Rnaseh1* overexpression, suggesting PRC2 complex may be very sensitive to the presence of R-loops, even when present at low levels (Supplementary Fig. 5b). Conversely, at some genes with high DRIP-RNA-seq signals, we do not see increased PRC2 binding upon *Rnaseh1* overexpression, suggesting either that the residual R-loops at these loci are sufficient to inhibit PRC2 association, or that additional features of chromatin structure at these sites impair PRC2 binding. Whether additional chromatin regulators are affected positively or negatively by the presence of R-loops to further compartmentalize the chromatin structure of genes in ESCs remains to be tested. However, given the large number of chromatin regulatory complexes found to bind lncRNAs[1–4], it seems likely additional factors will bind nascent transcripts in the form of R-loops.

Context-dependent effects of RNA binding on PRC2 function

While the effects of RNA on Tip60–p400 function have not been studied in detail, transcription appears to exert both positive and negative effects on the functions of polycomb complexes in multiple systems[6,9–13,19,36,43–45]. PRC2 binds to the A-repeat of the *Xist* lncRNA, and this is thought to help recruit the complex to the inactive X-chromosome[6,9]. Ezh2 binds nascent transcripts from numerous active genes, and has been shown to bind near the promoters of most active genes at low levels[10,11]. In addition, RNA binding inhibits the histone methyltransferase activity of PRC2, suggesting that binding of nascent transcripts holds PRC2 activity in check at active promoters, and PRC2 remains poised for histone methylation at these genes once transcription is silenced by another mechanism[9,12]. However, as with inhibition of transcription[13], we find that disruption of R-loops broadly stimulates PRC2 binding, suggesting that the effects of nascent transcription on PRC2 recruitment may be context-dependent. For example, nascent transcripts with G-rich sequences prone to R-loop formation may prevent PRC2 binding while different nascent transcripts that do not form R-loops may allow some PRC2 binding, while inhibiting its methyltransferase activity.

Multifaceted recruitment of Tip60–p400 to target genes

Although inhibition of transcription enhances PRC2 binding at CpG islands, including many promoter regions[13], transcriptional inhibitors significantly reduced Tip60–p400 association with target gene promoters. Importantly, *Rnaseh1* overexpression mimicked the effect of transcription inhibition on both complexes – enhancing Suz12 association[13] and inhibiting Tip60–p400 association. Since RNaseH1 degrades RNA species only within RNA:DNA

hybrids, this finding demonstrates that nascent transcripts, rather than the act of transcription itself, promote chromatin association by Tip60–p400 and inhibit chromatin association by PRC2. In addition, these data suggest that chromatin regulatory complexes encounter nascent transcripts at many genes in the form of R-loops, rather than free RNA.

While we observed a significant correlation between promoter-proximal R-loops and Tip60–p400 binding, several lines of data indicate R-loops are not sufficient for Tip60–p400 recruitment. First, R-loops are also prevalent at transcriptional termini[18,21,22] (also see Fig. 3a, c), which are not highly bound by Tip60–p400 (data not shown). Secondly, the PHD domain of the Ing3 subunit of Tip60–p400 was previously shown to bind histone H3 methylated on lysine-4 (H3K4me3)[46], and knockdown of genes required for H3K4me3 deposition leads to a moderate reduction of Tip60–p400 binding[15]. These data suggest that recruitment of Tip60–p400 to target sites on chromatin is a function of multiple different mechanisms. In addition, whether recruitment of Tip60–p400 to R-loop containing genes functions via direct binding of the complex to RNA:DNA hybrids or single-stranded DNA, or whether this interaction is bridged by another protein that is yet to be discovered, is not known.

Disruption of R-loops impairs ESC differentiation

Like *Ep400* mutant ESCs, *Rnaseh1* overexpression resulted in impaired differentiation, consistent with the reduction in Tip60–p400 binding observed in these cells. However, given the differences in proliferation observed between *Ep400* mutant and *Rnaseh1* overexpressing ESCs, the phenotypes observed upon disruption of R-loops likely reflect more than just the effects of reduced Tip60–p400 activity. Accordingly, while the precise effects of enhanced PRC2 binding on proliferation and differentiation of *Rnaseh1* overexpressing cells are difficult to predict, they are likely to contribute to the observed phenotypes. Furthermore, it is also possible R-loops modulate the binding of additional factors that regulate ESC differentiation. Nonetheless, the opposing effects of R-loops on Tip60–p400 and PRC2, and their importance for normal ESC differentiation, suggest an additional layer of complexity controlling gene regulation and cell identity in ESCs. These findings also suggest that factors regulating R-loop formation or clearance may have additional roles in gene regulation in multiple cell types.

ONLINE METHODS

Antibodies

Anti-p400 (A300-541A; Bethyl), anti-RUVBL1 (10210-2-AP; Proteintech), anti-FLAG-M2 (F1804; Sigma), anti-DNA-RNA hybrid S9.6 (ENH001; KeraFAST), anti-SUZ12 (A302-407A; Bethyl), anti-NANOG (A300-398A; Bethyl), rabbit-IgG (ab37415; Abcam), anti-HDAC6 (07-732; Millipore), anti-DMAP1 (10411-1-AP; Proteintech), anti-RNA Polymerase II (sc-899; Santa Cruz Biotechnology), anti-ACTIN (A5316; Sigma) antibodies were used. All primary antibodies against mammalian proteins used in this study are reported by the manufacturers to recognize protein from mouse. Information regarding antibody validation can be found on the manufacturers' websites.

Cell culture and treatment

Mouse ESCs were derived from E14[47] and previously obtained from Dr. Barbara Panning (University of California, San Francisco). Cells have been subjected to extensive sequencing in the course of this and previous studies, verifying they are male cells from mouse, and the pluripotency experiments reported in this and previous studies verify their ESC identity. ESCs were previously tested to ensure they are free of mycoplasma, and grown under feeder-free conditions as described[16]. The homozygous *Tip60-FLAG* line was generated by CRISPR/Cas-mediated genome editing into E14 using homology arms of approximately 900 bp surrounding a 6-Histidine-3XFLAG tag described previously[16], immediately 5′ of the endogenous stop codon (guide RNA sequence: AAGCCAGTTATCCTCGGAGT). The *Ep400* homozygous mutant line was made similarly, using a guide RNA (TGGCTGATGAAGCAGGGCTT) specific for the Walker A box of the ATPase domain of the *Ep400* gene and no homology template. Sequencing revealed a 135 bp deletion in exon 15 of both alleles. Full-length mouse *Rnaseh1* (NM_001286865.1) including an N-terminal 3XHA tag was synthesized (gBlocks, Integrated DNA Technologies) and cloned into the EcoRI-XhoI fragment of the pCAGGS-ires-Hygro vector. *Rnaseh1* overexpressing cells were generated by transfection of pCAGGS-*Rnaseh1*-ires-Hygro plasmid into the *Tip60-H3F* line and selection with Hygromycin B (Roche). For inhibition of transcription, cells were treated with 100 μM or 10 μM of DRB or Triptolide (Sigma), respectively, as described[13]. We tested several time points of treatment for inhibition of transcription by RT-qPCR and the protein levels of several subunits in Tip60–p400 by western blotting. We determined the optimal time of inhibitor treatment based on the shortest time we observed efficient inhibition of transcription while having no effect on protein levels of Tip60 subunits. For ESC differentiation, 10^6 ESC cells were suspended in medium lacking LIF and cultured in non-cell culture treated petri dishes for 2 days. Subsequently, cells were transferred to gelatin coated cell culture dishes in medium lacking LIF for the number of days indicated. Cells were fixed or RNA was isolated at the indicated time points.

Alkaline phosphatase staining

After 14-days of differentiation, cells were stained for AP activity using an Alkaline Phosphatase detection kit (EMD Millipore) according to the manufacturer's instructions.

Immunofluorescence staining

Cells were fixed with 4% paraformaldehyde, blocked with blocking buffer (10% normal goat serum, 0.3% Triton X-100 in PBS) for 1 hour, and stained with anti-Nanog antibody (1:100 dilution) overnight at 4°C. The next day, cells were washed and stained with Alexa Fluor 488-conjugated secondary antibodies (1:1,000) (Life Technologies). The nuclei were stained with DAPI, and the slides were imaged on an EVOS FL microscope (Life Technologies).

Cell cycle analysis

Propidium iodide staining and FACS analysis of DNA content were performed as described[15].

Dot Blotting

Indicated amounts of DNA were spotted onto a nitrocellulose membrane. After drying, the membrane was blocked in 5% milk for 30 minutes at room temperature and incubated with anti-S9.6 antibody (1:2,000 dilution) overnight at 4°C. The next day, the membrane was washed and stained with HRP conjugated anti-mouse secondary antibody (1:10,000).

Chromatin immunoprecipitation (ChIP)

Chromatin immunoprecipitation and library construction (ChIP-seq) were performed as described previously[16]. Libraries with different barcodes were pooled, and single-end sequencing (50bp) was performed on an Illumina HiSeq2000 at the UMass Medical School deep sequencing core facility. ChIP-qPCR was performed as described[16].

RIP-seq

Cells were lysed using an NE-PER Extraction kit (Thermo Fisher) to isolate nuclear fractions. For immunoprecipitation, 1.5 mg of nuclear extracts were treated with DNase I (New England Biolabs) and pre-cleared with Protein A magnetic beads (New England Biolabs) for 3 hours. Cleared nuclear extract was incubated with specific antibodies in IP buffer (50 mM Tris-HCl pH7.4, 250 mM NaCl, 0.1% Triton X-100) plus 1X HALT protease inhibitors (Thermo Fisher) and SUPERaseIn (Life Technologies) overnight at 4°C. The next day, pre-washed Protein A magnetic beads were added to IP samples and incubated for another 4 hours at 4°C. The magnetic beads were sequentially washed with IP buffer twice, high-salt IP buffer (50 mM Tris-HCl pH7.4, 500 mM NaCl, 0.1% Triton X-100, 0.5 % sodium deocycholate) four times, and IP buffer two more times. RNA was eluted from beads, purified by TRIzol (Life Technologies) extraction and precipitated at −80°C for at least 2 hours. For RIP-seq, 10–50 ng of RIP enriched RNA and Adaptor 1 (5′-CTGAACCGCTCTTCCGATCTNNNNNN-3′) were used for first-strand cDNA synthesis with Superscript III Reverse Transcription Kit (Life Technologies). After first-strand cDNA synthesis, RNA was degraded by sodium hydroxide, and cDNA was purified by SILAN beads (Life Technologies). To preserve strand information, Adaptor 2 with the modification of 5′ phosphorylated and 3′ dideoxy-C (5′-p-NNACGTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT-3′ddC) was ligated to the 3′ end of first-strand cDNA using T4 RNA ligase 1 (New England Biolabs). The ligated material was purified by SILAN beads and PCR amplified with Illumina primers using 18 cycles of PCR. To remove PCR primers, libraries were purified by AMPure XP beads (Beckman Coulter). Libraries with different barcodes were pooled together and sequenced as described for the ChIP-seq libraries.

DRIP-RNA-seq

Nucleic acid extraction, immunoprecipitation, and library preparation were performed as described previously[17] with the following modifications (cartooned in Supplementary Fig. 3b). The immunoprecipitated material (with and without RNaseH treatment) was denatured at 94°C for 1 min and cooled on ice. To reduce DNA background, the samples were treated with DNaseI at 37°C for 30 minutes and RNA was purified using phenol/chloroform/isoamylalcohol extraction. 38 pmol of Adaptor 1

(CTGAACCGCTCTTCCGATCTNNNNNN) was combined with 50 ng of S9.6 enriched RNA for first-strand cDNA synthesis with a Superscript III Reverse Transcription Kit (Life Technologies). After first-strand cDNA synthesis, RNA was degraded by sodium hydroxide, and strand-specific RNA-seq libraries were prepared as described above for RIP-seq libraries. Libraries with different barcodes were pooled together and sequenced as described above.

RNA-seq

Strand-specific RNA-seq libraries for ESCs and differentiated ESCs were performed as described previously[48].

Sequencing data analysis

Barcodes were removed, and reads were mapped to the mouse genome (mm9) using Bowtie-1.0.0[49] for ChIP-seq and TopHat-2[50] for RNA-seq, RIP-seq, and DRIP-RNA-seq. For ChIP-seq and DRIP-RNA-seq, aligned sequences were processed in HOMER[51] by using the “annotatePeaks” command to bin the regions of interest in 20 bp windows and sum the reads within each window. Average enrichment was calculated by normalizing the reads in each window to total reads, dividing by the number of regions of interest, and presented in reads per million (rpm). For RIP-seq data, aligned sequences were processed in HOMER by using the “analyzeRNA” command to calculate, normalize, and present in reads per kilobase per million mapped reads (rpkm) for each reference gene. Fold change was calculated by dividing rpkm from experimental IPs by rpkm IgG control IPs. Noncoding RNA data was obtained from the GENCODE release M1 dataset[52] and previously published lncRNAs[4]. For RNA-seq, rRNA sequences were removed before transcript quantification using RSEM[53]. Differentially expressed genes were identified by DESeq2[54] and significantly changed genes were selected using a cutoff of adjusted p-value < 0.05, comparing *Rnaseh1* overexpressing cells to control cells at each time point during differentiation.

Teratoma Formation Assays

We injected one million cells into both hind flanks of 5 *nu/nu* (nude) mice (male, aged 6–8 weeks) each for control and *Rnaseh1* overexpressing ESCs, and allowed tumors to grow for 21 days. Mice were sacrificed, tumors were weighed, followed by fixation and staining as described[15]. All animal experiments were performed according to an approved UMMS animal care and use protocol (2165-13). No statistical method was used to predetermine sample size. The experiments were not randomized and were not performed with blinding to the conditions of the experiments.

Statistical analysis and design

For most genomic datasets, we did not assume equal variances or similar distributions and therefore performed nonparametric tests, such as the Kolmogorov–Smirnov test to assess statistical significance of observed differences in distributions. Specific applications of statistical tests are discussed in the figure legends. For other experiments comparing individual genes or loci where we could assume similar variance and normally distributed values, we performed two tailed students t-tests. Due to the nature of genome-wide

experiments, we did not perform power analyses to determine sample sizes. For teratoma assays, we examined eight tumors for each condition out of ten injections, excluding the largest and smallest tumor in each group (by prior design) to reduce biases due to poor engraftment/injection. This sample size has been sufficient to clearly elucidate differentiation defects in our prior experience. Histograms indicate averages and error bars indicate standard deviations in all cases. Injections were performed into genetically identical nude mice, selected at random. Investigators were not blinded during injection of mice or downstream analyses of tumors.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank W. Hardy and M. Green (University of Massachusetts Medical School) for the pCAGGS-ires-Hygro expression vector and S. Hainer, L. Ee, and K. McCannell for input on the manuscript. This work was supported by US National Institutes of Health (NIH) grant R01HD072122 and American Cancer Society grant RSG-14-220-01 (T.G.F.), and NIH grant R01HD080224 (O.J.R.). T.G.F. was supported as a Pew Scholar in the Biological Sciences and is supported as a Leukemia and Lymphoma Society Scholar. Deep sequencing was performed at the University of Massachusetts Medical School Core facility on a HiSeq2000 supported by NIH grant 1S10RR027052-01.

References

1. Bonasio R, Shiekhattar R. Regulation of transcription by long noncoding RNAs. Annu Rev Genet. 2014; 48:433–455. [PubMed: 25251851]
2. Rinn JL, Chang HY. Genome regulation by long noncoding RNAs. Annu Rev Biochem. 2012; 81:145–166. [PubMed: 22663078]
3. Flynn RA, Chang HY. Long noncoding RNAs in cell-fate programming and reprogramming. Cell Stem Cell. 2014; 14:752–761. [PubMed: 24905165]
4. Guttman M, et al. lincRNAs act in the circuitry controlling pluripotency and differentiation. Nature. 2011; 477:295–300. [PubMed: 21874018]
5. Rinn JL, et al. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell. 2007; 129:1311–1323. [PubMed: 17604720]
6. Zhao J, Sun BK, Erwin JA, Song JJ, Lee JT. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. Science. 2008; 322:750–756. [PubMed: 18974356]
7. Pandey RR, et al. Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. Mol Cell. 2008; 32:232–246. [PubMed: 18951091]
8. Aloia L, Di Stefano B, Di Croce L. Polycomb complexes in stem cells and embryonic development. Development. 2013; 140:2525–2534. [PubMed: 23715546]
9. Cifuentes-Rojas C, Hernandez AJ, Sarma K, Lee JT. Regulatory interactions between RNA and polycomb repressive complex 2. Mol Cell. 2014; 55:171–185. [PubMed: 24882207]
10. Kaneko S, Son J, Shen SS, Reinberg D, Bonasio R. PRC2 binds active promoters and contacts nascent RNAs in embryonic stem cells. Nature Structural & Molecular Biology. 2013; 20:1258–1264.
11. Davidovich C, Zheng L, Goodrich KJ, Cech TR. Promiscuous RNA binding by Polycomb repressive complex 2. Nature Structural & Molecular Biology. 2013; 20:1250–1257.
12. Kaneko S, Son J, Bonasio R, Shen SS, Reinberg D. Nascent RNA interaction keeps PRC2 activity poised and in check. Genes & Development. 2014; 28:1983–1988. [PubMed: 25170018]
13. Riising EM, et al. Gene silencing triggers polycomb repressive complex 2 recruitment to CpG islands genome wide. Mol Cell. 2014; 55:347–360. [PubMed: 24999238]

14. Squatrito M, Gorrini C, Amati B. Tip60 in DNA damage response and growth control: many tricks in one HAT. Trends in Cell Biology. 2006; 16:433–442. [PubMed: 16904321]
15. Fazzio TG, Huff JT, Panning B. An RNAi screen of chromatin proteins identifies Tip60-p400 as a regulator of embryonic stem cell identity. Cell. 2008; 134:162–174. [PubMed: 18614019]
16. Chen PB, et al. Hdac6 regulates Tip60-p400 function in stem cells. Elife. 2013; 2:e01557. [PubMed: 24302573]
17. Ginno PA, Lott PL, Christensen HC, Korf I, Chédin F. R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. Mol Cell. 2012; 45:814–825. [PubMed: 22387027]
18. Ginno PA, Lim YW, Lott PL, Korf I, Chédin F. GC skew at the 5′ and 3′ ends of human genes links R-loop formation to epigenetic regulation and transcription termination. Genome Res. 2013; 23:1590–1600. [PubMed: 23868195]
19. Davidovich C, et al. Toward a Consensus on the Binding Specificity and Promiscuity of PRC2 for RNA. Mol Cell. 2015; doi: 10.1016/j.molcel.2014.12.017
20. Sun Q, Csorba T, Skourti-Stathaki K, Proudfoot NJ, Dean C. R-loop stabilization represses antisense transcription at the Arabidopsis FLC locus. Science. 2013; 340:619–621. [PubMed: 23641115]
21. Skourti-Stathaki K, Kamieniarz-Gdula K, Proudfoot NJ. R-loops induce repressive chromatin marks over mammalian gene terminators. Nature. 2015; 516:436–439. [PubMed: 25296254]
22. Castel SE, et al. Dicer Promotes Transcription Termination at Sites of Replication Stress to Maintain Genome Stability. Cell. 2014; 159:572–583. [PubMed: 25417108]
23. Britton S, et al. DNA damage triggers SAF-A and RNA biogenesis factors exclusion from chromatin coupled to R-loops removal. Nucleic Acids Res. 2014; 42:9047–9062. [PubMed: 25030905]
24. Groh M, Lufino MMP, Wade-Martins R, Gromak N. R-loops associated with triplet repeat expansions promote gene silencing in Friedreich ataxia and fragile X syndrome. PLoS Genet. 2014; 10:e1004318. [PubMed: 24787137]
25. Helmrich A, Ballarino M, Tora L. Collisions between replication and transcription complexes cause common fragile site instability at the longest human genes. Mol Cell. 2011; 44:966–977. [PubMed: 22195969]
26. Huertas P, Aguilera A. Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. Mol Cell. 2003; 12:711–721. [PubMed: 14527416]
27. Li X, Manley JL. Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. Cell. 2005; 122:365–378. [PubMed: 16096057]
28. Castellano-Pozo M, et al. R loops are linked to histone H3 S10 phosphorylation and chromatin condensation. Mol Cell. 2013; 52:583–590. [PubMed: 24211264]
29. Boque-Sastre R, et al. Head-to-head antisense transcription and R-loop formation promotes transcriptional activation. Proc Natl Acad Sci USA. 2015; 112:5785–5790. [PubMed: 25902512]
30. Costantino L, Koshland D. The Yin and Yang of R-loop biology. Current Opinion in Cell Biology. 2015; 34:39–45. [PubMed: 25938907]
31. Aguilera A, García-Muse T. R loops: from transcription byproducts to threats to genome stability. Mol Cell. 2012; 46:115–124. [PubMed: 22541554]
32. Shen YJ, et al. Genome-derived cytosolic DNA mediates type I interferon-dependent rejection of B cell lymphoma cells. CellReports. 2015; 11:460–473.
33. Wang H, et al. One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. Cell. 2013; 153:910–918. [PubMed: 23643243]
34. Mali P, et al. RNA-guided human genome engineering via Cas9. Science. 2013; 339:823–826. [PubMed: 23287722]
35. Cong L, et al. Multiplex genome engineering using CRISPR/Cas systems. Science. 2013; 339:819–823. [PubMed: 23287718]
36. Kanhere A, et al. Short RNAs are transcribed from repressed polycomb target genes and interact with polycomb repressive complex-2. Mol Cell. 2010; 38:675–688. [PubMed: 20542000]

37. Lynch MD, et al. An interspecies analysis reveals a key role for unmethylated CpG dinucleotides in vertebrate Polycomb complex recruitment. EMBO J. 2012; 31:317–329. [PubMed: 22056776]
38. Brinkman AB, et al. Sequential ChIP-bisulfite sequencing enables direct genome-scale investigation of chromatin and DNA methylation cross-talk. Genome Res. 2012; 22:1128–1138. [PubMed: 22466170]
39. Mendenhall EM, et al. GC-rich sequence elements recruit PRC2 in mammalian ES cells. PLoS Genet. 2010; 6:e1001244. [PubMed: 21170310]
40. Skourti-Stathaki K, Proudfoot NJ, Gromak N. Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. Mol Cell. 2011; 42:794–805. [PubMed: 21700224]
41. Wahba L, Gore SK, Koshland D. The homologous recombination machinery modulates the formation of RNA-DNA hybrids and associated chromosome instability. Elife. 2013; 2:e00505. [PubMed: 23795288]
42. Skourti-Stathaki K, Proudfoot NJ. A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression. Genes & Development. 2014; 28:1384–1396. [PubMed: 24990962]
43. Cavalli G, Paro R. The Drosophila Fab-7 chromosomal element conveys epigenetic inheritance during mitosis and meiosis. Cell. 1998; 93:505–518. [PubMed: 9604927]
44. Klymenko T, Müller J. The histone methyltransferases Trithorax and Ash1 prevent transcriptional silencing by Polycomb group proteins. EMBO Reports. 2004; 5:373–377. [PubMed: 15031712]
45. Schmitt S, Prestel M, Paro R. Intergenic transcription through a polycomb group response element counteracts silencing. Genes & Development. 2005; 19:697–708. [PubMed: 15741315]
46. Shi X, et al. ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. Nature. 2006; 442:96–99. [PubMed: 16728974]
47. Hooper M, Hardy K, Handyside A, Hunter S, Monk M. HPRT-deficient (Lesch-Nyhan) mouse embryos derived from germline colonization by cultured cells. Nature. 1987; 326:292–295. [PubMed: 3821905]
48. Hainer SJ, et al. Suppression of pervasive noncoding transcription in embryonic stem cells by esBAF. Genes & Development. 2015; 29:362–378. [PubMed: 25691467]
49. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009; 10:R25. [PubMed: 19261174]
50. Kim D, et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 2013; 14:R36. [PubMed: 23618408]
51. Heinz S, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell. 2010; 38:576–589. [PubMed: 20513432]
52. Pervouchine DD, et al. Enhanced transcriptome maps from multiple mouse tissues reveal evolutionary constraint in gene expression. Nat Commun. 2015; 6:5903. [PubMed: 25582907]
53. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011; 12:323. [PubMed: 21816040]
54. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014; 15:550. [PubMed: 25516281]

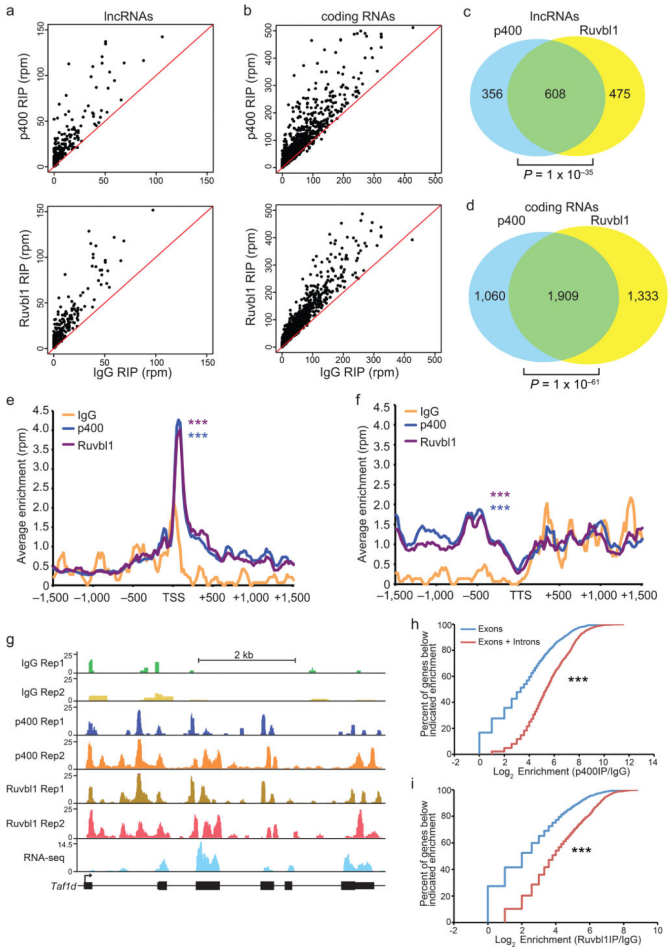

Figure 1. Tip60–p400 binds nascent transcripts

a–b, Enrichment of transcripts in p400 (upper) or Ruvbl1 (lower) RIP-seq libraries relative to control (IgG) RIP-seq. Normalized reads (reads per million; rpm) from two biological replicate (IPs from separate cultures) RIP-seq experiments were averaged and plotted for lncRNAs RNAs (a) or coding RNAs (b). **c–d,** Overlaps of lncRNAs (c) or coding RNAs (d) enriched in each RIP-seq dataset are shown as Venn diagrams with significance of overlaps (hypergeometric tests) indicated. **e–f,** Aggregation plot of RIP-seq data over annotated TSSs (e) and TTSs (f). ***$P < 2.2 \times 10^{-16}$, calculated using a two-sample Kolmogorov-Smirnov (K-S) test after summing promoter-proximal (e) or TTS-proximal (f) reads for each gene. **g,** Example browser track showing locations of RIP-seq reads for the *Taf1d* gene relative to introns (thin line) and exons (black boxes). **h–i,** Cumulative distribution plots showing enrichment of reads over the entire gene (red) or only within exons (blue) in p400 and Ruvbl1 RIP-seq compared to IgG, expressed as a log2 ratio. ***$P < 2.2 \times 10^{-16}$ using a two-sample K-S test, as above.

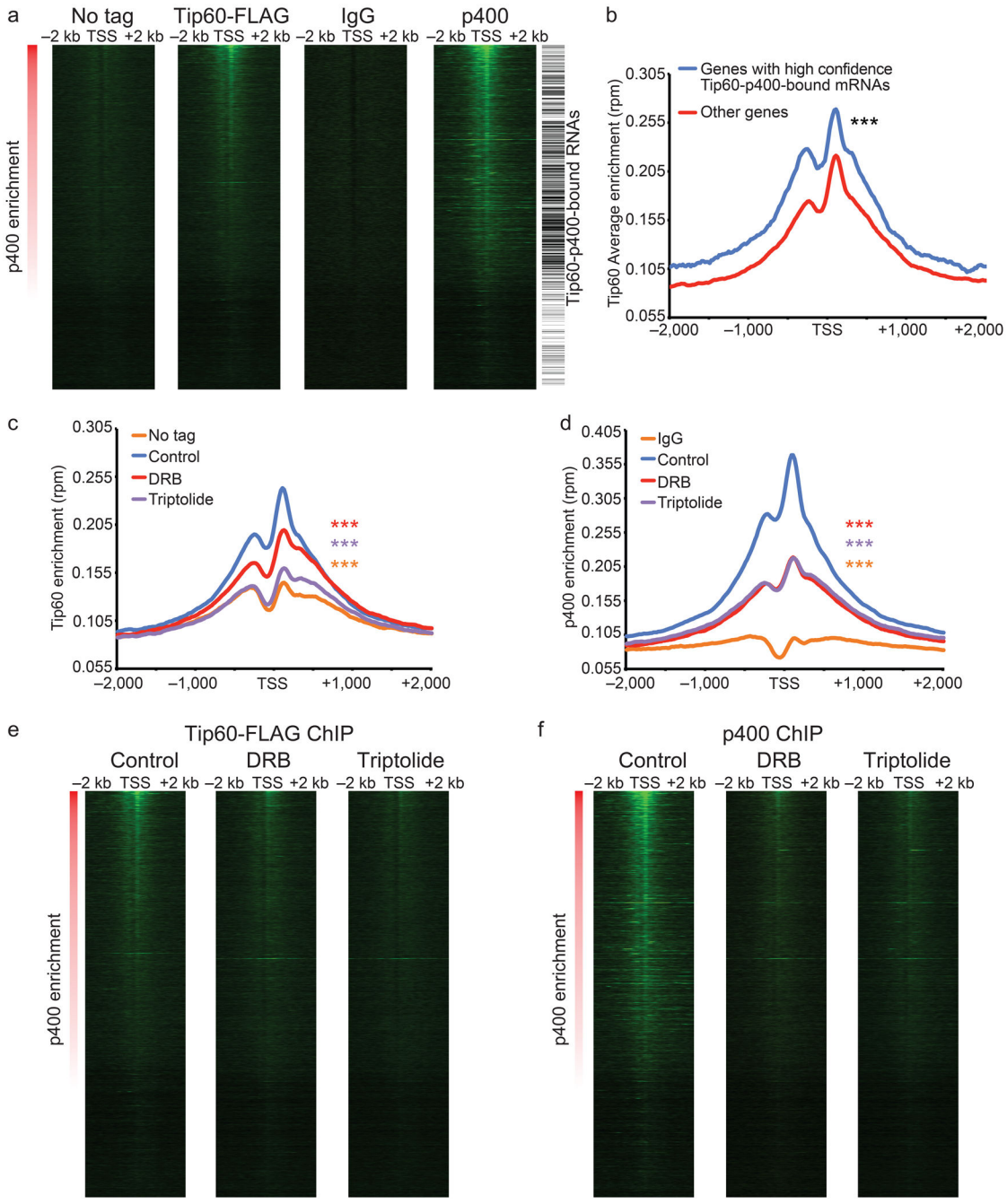

Figure 2. Transcription promotes promoter-proximal association of Tip60–p400 with chromatin

a, Comparison of ChIP-seq maps of Tip60 (C-terminally FLAG-tagged at the endogenous *Tip60* locus), p400, or control IPs (anti-FLAG ChIP of ESCs lacking FLAG-tagged Tip60 and IgG ChIP) with Tip60–p400-interacting RNAs (overlapping in p400 and Ruvbl1 RIP-seq libraries). ChIP data is shown as heatmaps extending from −2 kb to + 2 kb from each TSS, with each row representing a gene and enrichment denoted in green. Heatmaps are sorted by previously published p400 ChIP-chip data[15]. **b,** Tip60 enrichment in reads per million (rpm) aggregated over TSSs of genes whose transcripts are bound by both p400 and

Ruvbl1 and those that are not. ***$P < 2.2 \times 10^{-16}$, calculated using a two-sample K-S test, as in Fig. 1. **c–d,** Aggregate Tip60 (c) or p400 (d) ChIP-seq enrichment at TSSs in control treated ESCs or ESCs treated with indicated transcription inhibitors. ***$P < 2.2 \times 10^{-16}$ for all treatments vs. controls, using two-sample K-S tests. **e–f**, Heatmaps of Tip60 (e) or p400 (f) enrichment by ChIP-seq in control treated ESCs or ESCs treated with indicated transcription inhibitors. All heatmaps are sorted by previously published p400 ChIP-chip data to show concordance of Tip60 and p400 binding sites with each other and with previous studies. One ChIP-seq was performed with or without each of two independent transcription inhibitors (treatments performed in independent cultures) for each Tip60-p400 subunit indicated (and controls).

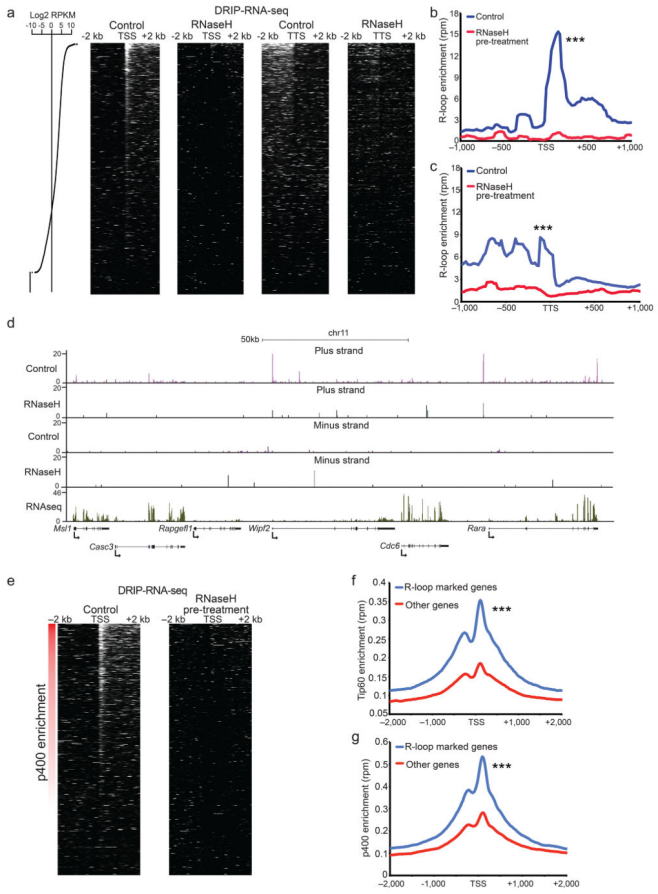

Figure 3. Promoter-proximal R-loops co-localize with Tip60–p400

a–e, R-loop localization in mouse ESCs using DRIP-RNA-seq. **a,** DRIP-RNA-seq data represented as heatmaps for both transcription start sites (TSSs) and transcription termination sites (TTSs), and sorted by gene expression in ESCs from high to low (expression level indicated to left). RNA-seq read density from DRIP-RNA-seq libraries is indicated in white. **b–c,** R-loop enrichment in reads per million (rpm) aggregated over annotated TSSs or TTSs in control samples or samples pre-treated with RNaseH in vitro prior to DRIP (see Methods). ***$P < 2.2 \times 10^{-16}$, calculated using a two-sample K-S test, as in Fig. 1. **d,** R-loop localization at an example genomic location. DRIP-RNA-seq reads were split into plus and minus strands (direction of transcription for each gene noted at bottom). **e,** Heatmaps as in (a) of DRIP-RNA-seq data sorted by p400 enrichment. **f–g,** Average Tip60 (f) or p400 (g) binding measured by ChIP-seq over promoters with highly enriched DRIP-RNA-seq levels (blue) and all other promoters (red). One DRIP-RNA-seq library per condition was analyzed after several pilot DRIP experiments were performed in the lab. ChIP-seq libraries were described in Fig. 2. ***$P < 2.2 \times 10^{-16}$, calculated using a two-sample KS test.

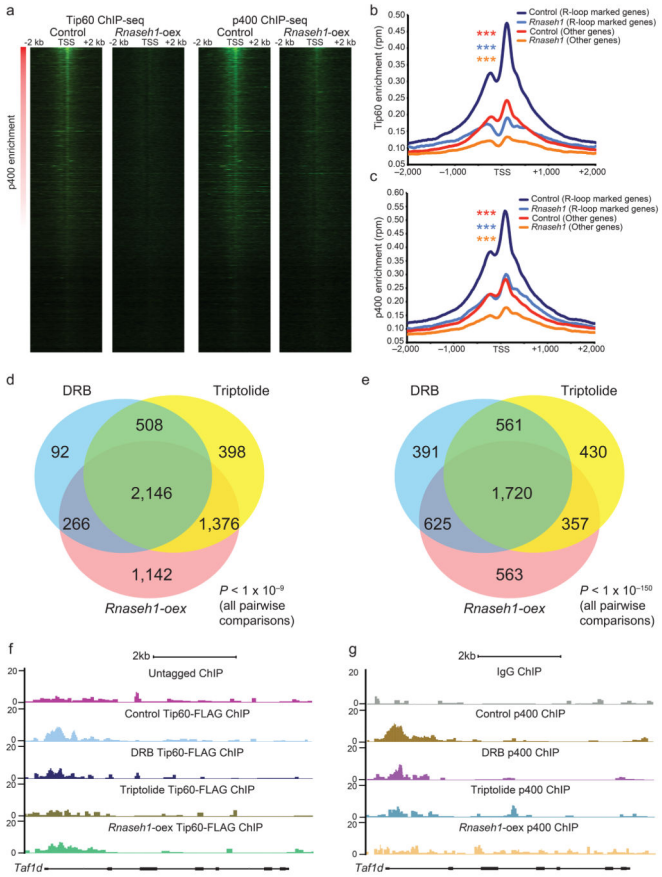

Figure 4. R-loops promote elevated Tip60 and p400 binding to many target genes

a–c, Changes in Tip60 and p400 chromatin binding in ESCs overexpressing *Rnaseh1*, expressed as heatmaps (a) sorted by p400 enrichment, or aggregated over all TSSs for Tip60 binding (b) or p400 binding (c). ***$P < 2.2 \times 10^{-6}$, calculated using two-sample K-S tests, as in Fig. 1, comparing Control R-loop marked genes to all other groups. For each ChIP-seq, one of two representative biological replicate ChIPs (independent IPs from separate cultures) with similar results is shown. **d–e,** Overlap of genes with at least two-fold reduced binding of Tip60 (d) or p400 (e) upon *Rnaseh1* overexpression or addition of transcription inhibitors DRB or Triptolide. *P* values indicating significance of all pairwise overlaps were calculated using hypergeometric tests. **f–g,** Browser tracks for one example locus with reduced Tip60 (f) and p400 (g) binding upon addition of DRB or Triptolide, or overexpression of *Rnaseh1*.

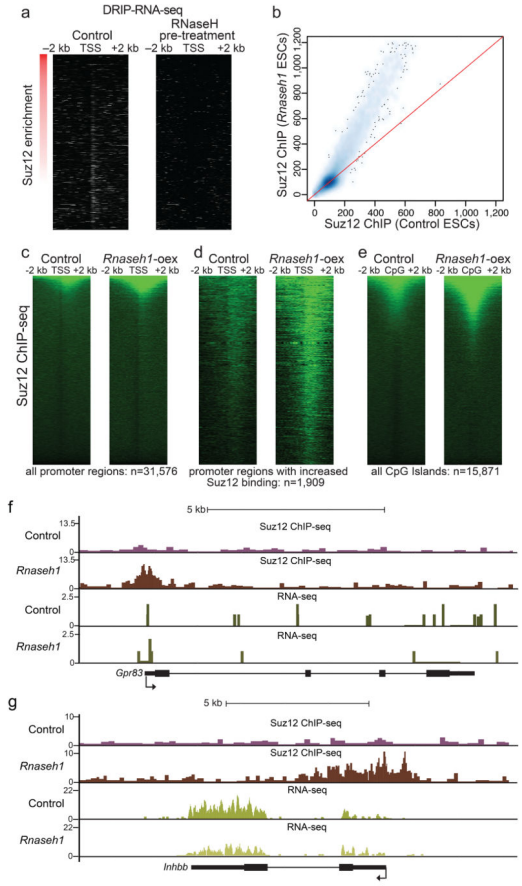

Figure 5. R-loops inhibit chromatin binding by PRC2

a, Heatmap showing DRIP-RNA-seq data sorted by Suz12 enrichment, as measured by ChIP-seq. **b,** Effect of R-loop disruption on Suz12 enrichment, expressed as a density plot. The red line marks equal enrichment in both cell types. **c–e,** Heatmaps illustrating changes in Suz12 chromatin binding upon *Rnaseh1* overexpression over the promoter proximal regions of all promoters (c), genes with increased Suz12 association in *Rnaseh1* overexpressing cells (d), or annotated CpG islands (e). All heatmaps are sorted by Suz12 binding in control cells, and one of two biological replicate ChIP-seq experiments (independent IPs from separate cultures) with similar results is shown. **f–g**, browser tracks of genes that gain ectopic Suz12 binding upon *Rnaseh1* overexpression.

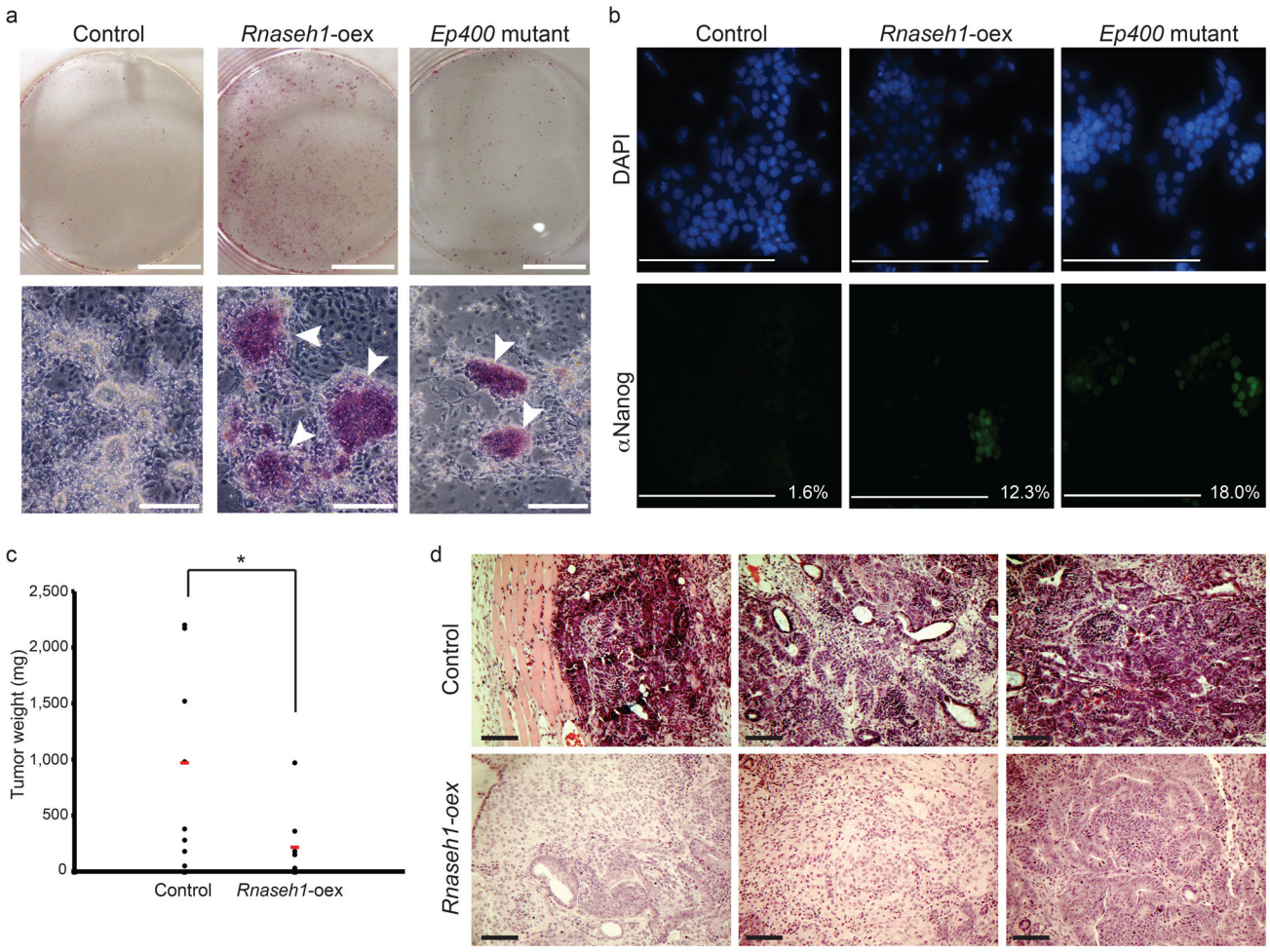

Figure 6. Disruption of R-loops impairs ESC differentiation

a–b, AP staining (a) or Nanog immunofluorescence staining (b) of control, *Rnaseh1* overexpressing, or *Ep400* mutant ESCs after culture in differentiation-promoting medium for 14 days. White arrows in (a) indicate clusters of robustly AP stained cells in *Rnaseh1* overexpressing cells. Scale bars in (a) measure 1 cm (upper panels) and 200 μm (lower panels). Scale bars in (b) measure 100 μm. The percentages of Nanog-expressing cells (indicated by white text in lower right of each panel) were averaged from two biological replicate differentiation experiments (independent cultures on different days). **c,** Weight of teratomas derived from subcutaneous injection of control or *Rnaseh1* overexpressing ESCs (n = 8 tumors each) into nude mice. Each tumor weight is shown as a black dot, and the mean is shown in red. $*P < 0.05$ by student's two-tailed t-test. **d,** Representative examples of sections from teratomas derived from control or *Rnaseh1* overexpressing ESCs. Scale bars measure 200 μm.

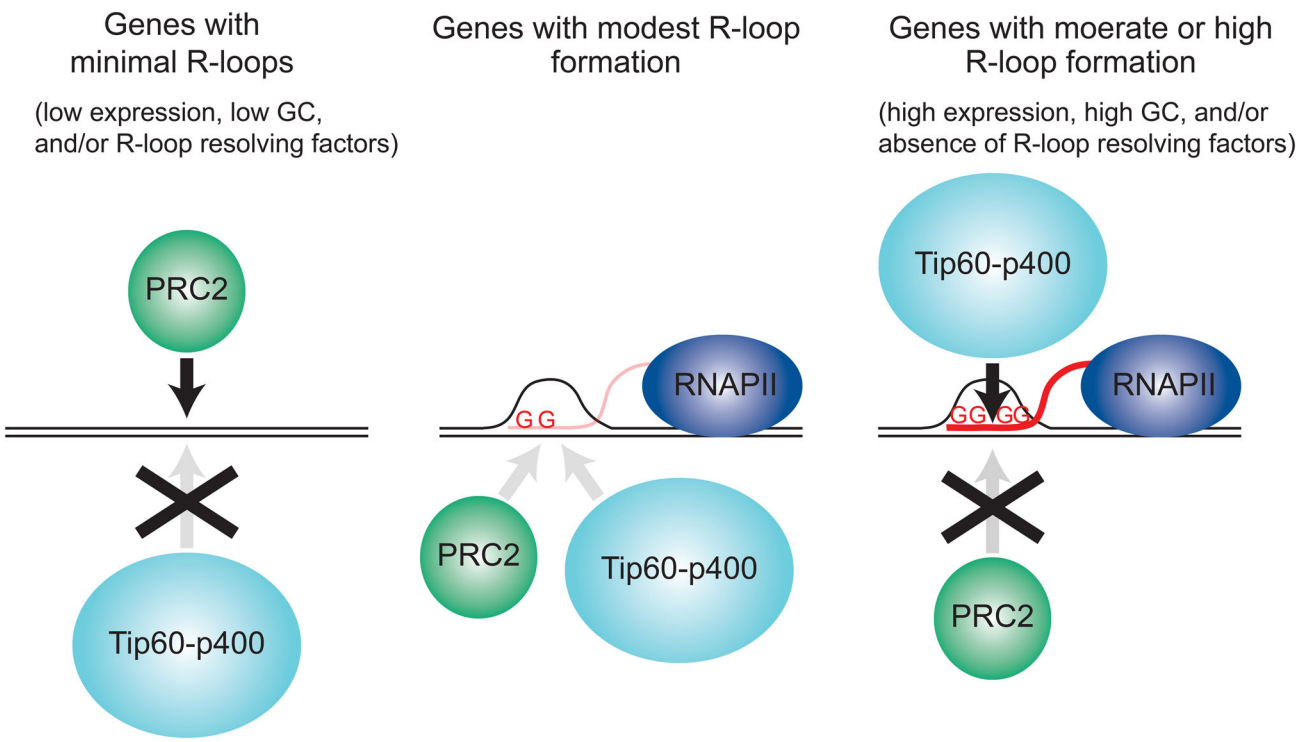

Figure 7. Model of R-loop function

Genes that do not form R-loops, due to either lack of expression or G-poor sequence within the 5′ region of the transcript, make good substrates for PRC2 binding, but poor Tip60–p400 substrates. Conversely, genes that form moderate to high levels of R-loops are good Tip60–p400 substrates, but poor PRC2 substrates. Genes that form R-loops at modest levels, due to low expression and/or weak to moderate G-richness within the 5′ region of the transcript, are predicted to be relatively poor substrates for both complexes. RNA: Red curved line.