R loops regulate promoter-proximal chromatin architecture and cellular differentiation

Poshen B Chen1,2, Hsiuyi V Chen3, Diwash Acharya1,2, Oliver J Rando3 & Thomas G Fazzio1,2

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; however, transcription promotes chromatin binding of Tip60–p400 but not PRC2. Interestingly, we observed higher Tip60–p400 and lower PRC2 levels at genes marked by promoter-proximal R loops. Furthermore, disruption of R loops broadly decreased Tip60–p400 occupancy and increased PRC2 occupancy genome wide. In agreement with these alterations, ESCs partially depleted of R loops exhibited impaired differentiation. These results show that R loops act both positively and negatively in modulating the recruitment of key pluripotency regulators.

Owing to the discovery of thousands of long noncoding RNAs (lncRNAs) that are expressed in mammalian cells, considerable efforts are underway to uncover the roles of specific lncRNAs in the nucleus, as well as to broadly elucidate generalizable mechanisms of action that govern their biological functions. lncRNAs function both in cis and in trans in regulating gene expression, thus raising the possibility that these transcripts specifically modulate the functions of individual transcription factors, the general transcription machinery or other regulatory proteins. In agreement with this hypothesis, numerous lncRNAs have been shown to interact with transcriptional-regulatory proteins.

Interestingly, in a survey of 74 lncRNAs expressed in ESCs, several chromatin-remodeling complexes with key roles in ESC pluripotency have been shown to bind lncRNAs. Multiple complexes bound to more than 30% of the lncRNAs tested, and numerous lncRNAs were bound by more than one complex, thus suggesting either that these factors are differentially regulated by dozens of individual lncRNAs or that these complexes bind lncRNAs relatively nonspecifically. 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-remodeling complexes shown to bind lncRNAs was PRC2 (refs. 5–7), a highly conserved histone H3 Lys27 methyltransferase complex important for gene silencing during development8. PRC2 binding to the A repeat of the Xist lncRNA is thought to have a role in recruitment of the complex to the inactive X chromosome9,10. 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 abundance10,11. At first glance, PRC2’s 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 H3 Lys27 methyltransferase activity9,12. In agreement with these findings, PRC2 components bind to both silent and active genes, and active genes bound by PRC2 are not marked by trimethylated H3 Lys27 (H3K37me3)10,11. These findings support a revised model in which binding of nascent transcripts at active genes helps to recruit PRC2 to these loci but maintains the complex in an inactive state9,12. In this model, PRC2 is poised to generate a repressive chromatin structure and to enforce silencing at these genes at a later time, should their expression be silenced by an independent mechanism. However, chemical inhibition of transcription promotes binding of PRC2 to CpG islands (including numerous promoter-proximal regions) throughout the genome, thus conflicting with a model in which nascent transcripts are necessary for recruitment of PRC2 (ref. 13). Therefore, the roles of nascent transcripts in the 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 that has been reported to bind lncRNAs4. 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.

1Department of Molecular, Cell, and Cancer Biology, University of Massachusetts Medical School, Worcester, Massachusetts, USA. 2Program in Molecular Medicine, University of Massachusetts Medical School, Worcester, Massachusetts, USA. 3Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, Massachusetts, USA. Correspondence should be addressed to T.G.F. (thomas.fazzio@umassmed.edu).

Received 27 April; accepted 9 October; published online 9 November 2015; doi:10.1038/nsmb.3122
variant into chromatin\textsuperscript{14}. We have previously found that Tip60–p400 is essential for normal ESC self-renewal and pluripotency and that it acts simultaneously to repress some differentiation genes and activate proliferation genes\textsuperscript{15,16}. Although it is not clear how Tip60–p400 simultaneously activates one group of genes and silences another, interactions with lncRNAs could potentially target the complex to specific regions of the genome and/or tune its catalytic activities at specific targets, thereby favoring activation or silencing.

Here, we address the role of RNA binding by Tip60–p400 in mouse ESCs. We found that, like PRC2, Tip60–p400 binds promiscuously to nascent RNAs from both coding and noncoding genes. However, unlike binding of PRC2 to chromatin, which is inhibited by transcription\textsuperscript{13}, binding of Tip60–p400 to many of its target promoters is enhanced by promoter-proximal R loops, RNA-DNA hybrid structures formed when G-rich sequences bind to many target genes was enhanced by promoter-proximal R loops. These results demonstrate that R loops play a major part 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 interacted with 9 of 74 of the lncRNAs tested\textsuperscript{4}, thus raising the possibility that lncRNAs might be important for interaction of the complex with chromatin or for remodeling of chromatin structure by the complex. Alternatively, Tip60–p400 might bind promiscuously to RNA, as has been shown for the well-studied chromatin-regulatory complex, PRC2 (refs. 10,11,19).

To distinguish between these possibilities, we first performed unbiased identification of Tip60–p400–interacting transcripts by deep sequencing of RNAs that coimmunoprecipitate with Tip60–p400 (RNA immunoprecipitation–sequencing (RIP-seq)). We performed immunoprecipitations of two different Tip60–p400 subunits, p400 and Ruvbl1, in replicate cell cultures 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 on RNA hybridize with their DNA template\textsuperscript{17,18}. In contrast, binding of the PRC2 complex and H3 Lys27 methylation were inhibited by R loops. These results demonstrate that R loops play a major part 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.

![Figure 1](image1.png)

**Figure 1** Tip60–p400 binds nascent transcripts. (a,b) Enrichment of transcripts in p400 or Ruvbl1 RIP-seq libraries relative to control (IgG) RIP-seq. Normalized reads (per million (RPMM) from n = 2 immunoprecipitations from separate cultures. RIP-seq experiments are shown averaged and plotted for lncRNAs (a) or coding RNAs (b). (c,d) Overlap of lncRNAs (c) or coding RNAs (d) enriched in each RIP-seq data set, shown as Venn diagrams with significance of overlap (determined by hypergeometric tests) and numbers of overlapping and nonoverlapping genes indicated. (e,f) Aggregation plot of RIP-seq data over annotated TSSs (e) and TTSs (f). ***P < 2.2 × 10\textsuperscript{−16} by two-sample Kolmogorov-Smirnov (K-S) test after summing promoter-proximal reads (n = 31,576 TSSs in e) or TTS-proximal reads (n = 31,593 TTSs in f) for each gene. (g) Example browser track showing locations of RIP-seq reads for the Taf1d gene relative to introns (thin lines) and exons (black boxes). (h,i) Cumulative distribution plots showing enrichment of reads over the entire genome (red) or only within exons (blue) in p400 (h) and Ruvbl1 (i) RIP-seq compared to IgG, expressed as log\textsubscript{2} ratio. ***P < 2.2 × 10\textsuperscript{−16} by two-sample K-S test, after summing reads over each of n = 20,314 annotated transcripts.
more 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 noncoding transcripts in ESCs. More interestingly, we observed that Tip60–p400 also interacts with 1,909 coding gene transcripts (Fig. 1d), results suggesting that Tip60–p400 does not bind specifically to lncRNAs but instead interacts with a broad array of both coding and noncoding transcripts in ESCs.

Next, we considered whether this complex might interact with nascent transcripts; therefore, we 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), as compared to lower (but above background) levels near transcriptional termination sites (TTSs) (Fig. 1e,f). In agreement with this result, we observed a significant overrepresentation of reads within the first exon and first intron of Tip60–p400–interacting RNAs (Supplementary Fig. 1c), thus suggesting that the complex interacts with unspliced precursor mRNA (pre-mRNA) transcripts. This pattern was present in both cell-culture replicates of each RIP, although the relative heights and locations of RIP peaks were somewhat variable (Fig. 1g), thus suggesting that 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 chromatin immunoprecipitation–sequencing (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). These results suggest that Tip60–p400 binds numerous transcripts in cis. The bound transcripts occupied a broad range of expression levels and functional categories (Supplementary Fig. 1e,f), in agreement with the diverse set of genes bound and regulated by Tip60–p400 (refs. 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...
R loops induce DNA damage and genomic instability23–27, and R-loop formation 
regulate transcription termination 18,20,21, and R-loop formation 
1002, described genes in multiple cell types 17,18,20–22. Although unresolved 
Nascent transcripts can form R loops 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 the recruit-
ment 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 (DNA–RNA immunoprecipitation (DRIP)) which uses 
an antibody (S9.6) specific for these structures, coupled to either qPCR 
(DRIP-qPCR) or deep sequencing (DRIP-seq) has been used to map 
R loops in multiple cell types17,21,22. To reduce the background and 
identify more precise boundaries of R loops mapped with this tech-
nique (Supplementary Fig. 3a), we modified the DRIP-seq protocol 
to sequence only RNAs enriched within immunoprecipitates of RNA-
DNA hybrids (Supplementary Fig. 3b and Online 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 weakly or

![Image](224x557 to 416x686)

**Figure 3** Promoter-proximal R loops colocalize with Tip60–p400. (a–e) R-loop localization 
in mouse ESCs by DRIP-RNA-seq. (a) DRIP-
RNA-seq data, represented as heat maps for 
both TSSs and TTSs, sorted by gene expression 
in ESCs from high to low (expression levels 
indicated at left). RNA-seq read densities from 
DRIP-RNA-seq libraries are indicated in white. 
(b,c) R-loop enrichment in reads per million 
(RPM) aggregated over annotated TSSs or TTSs 
in control samples or samples pretreated with 
RNase H in vitro before DRIP (Online Methods). 
***P < 2.2 × 10−16 by two-sample K-S test, after 
summing reads over n = 31,576 TSSs or 
n = 31,593 TSSs, as in Figure 1. (d) R-loop 
localization at an example genomic location. 
DRIP-RNA-seq reads were split into plus and 
minus strands (with direction of transcription for 
each gene indicated at bottom). (e) Heat maps 
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. ChIP-seq libraries were described in 
Figure 2. **P < 2.2 × 10−16 by two-sample 
K-S test, after summing TSS-proximal reads for 
each gene (n = 31,576 TSSs).
moderately expressed genes (Fig. 3a, and comparison of the weakly expressed R loop marked Wip2 to more highly expressed genes without R loops in Fig. 3d). We confirmed the specificity of DRIP signals in two ways. First, we found that signals were significantly reduced when we treated samples with RNase H (which degrades RNA within RNA-DNA hybrids) before immunoprecipitation (Fig. 3a–d). Second, in our strand-specific DRIP-RNA-seq libraries, we observed mainly sense-strand reads (Fig. 3d and Supplementary Fig. 3c).

Interestingly, we observed a high incidence of R loops at Tip60–p400–target genes (Fig. 3e) as well as higher overall enrichment of Tip60 and p400 at genes with associated R loops than those without (Fig. 3f,g), in agreement with the possibility that R loops promote Tip60–p400 binding. To test this hypothesis, we used Rnaseh1 overexpression in ESCs to disrupt R-loop formation. Overexpression of the RNase H1 protein in multiple organisms is known to disrupt R loops throughout the genome. We found that overexpression of Rnaseh1 in ESCs reduced bulk RNA-DNA hybrids approximately four-fold (Supplementary Fig. 3d). Interestingly, we observed a decrease in localization of both Tip60 and p400 to most Tip60–p400–target genes in Rnaseh1-overexpressing cells (Fig. 4a and Supplementary Fig. 4). At genes with high-confidence R loops, we found that Tip60 binding was reduced by an average of 63% (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. We observed similar results 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, that some binding events might be indirectly affected by Rnaseh1 expression or that both possibilities occur. 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 the chromatin association of the Tip60–p400 complex.

Because RNA-DNA hybrid have roles in DNA replication, rRNA expression and other processes, Rnaseh1 overexpression might affect the interpretation of these data. We observed minimal effects of Rnaseh1 overexpression on most cellular functions affected by RNA-DNA hybrids: Rnaseh1-overexpressing ESCs self-renewed normally (Supplementary Fig. 6a) and exhibited no apparent alterations in their cell cycle (Supplementary Fig. 6b,c) or rRNA levels (Supplementary Fig. 6d). Rnaseh1 overexpression resulted in slower proliferation relative to control cells, although this defect was less severe than in Ep400 (the gene encoding the p400 protein)-mutant ESCs generated by clustered regularly interspaced short palindromic repeats (CRISPR)-Cas9 cleavage

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 (oex), expressed as heat maps sorted by p400 enrichment (a) or aggregated over all TSSs for Tip60 binding (b) or p400 binding (c). * * * P < 2.2 × 10⁻¹⁰ by two-sample K-S tests, after summing TSS-proximal reads for each gene (n = 31,576 TSSs), comparing control R loop–marked genes to all other groups. For each ChIP-seq, one of two representative independent experiments 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 the transcription inhibitors DRB or triptolide. P values indicating significance of all pairwise overlaps were calculated with hypergeometric tests; numbers of overlapping and nonoverlapping genes are indicated. (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.
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 because of its established RNA-binding activity in multiple cell types. Like Tip60–p400, PRC2 binds to nascent transcripts, the substrates for R-loop formation, in agreement with the possibility that R loops might promote PRC2 binding. However, because inhibition of transcription stimulates the association of PRC2 with chromatin, 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 the PRC2 subunit Suz12. Interestingly, DRIP-RNA-seq reads were poorly enriched near the promoter-proximal regions of genes highly bound by Suz12 (Fig. 5a), thus suggesting that moderate to high levels of promoter-proximal R loops may inhibit PRC2 association. We tested this possibility directly.

Figure 5 R loops inhibit chromatin binding by PRC2. (a) Heat map showing DRIP-RNA-seq data sorted by Suz12 enrichment, as measured by ChIP-seq. (b) Effect of R-loop disruption on Suz12 binding upon Rnaseh1 overexpression, expressed as a density plot. The red line marks equal enrichment in both cell types. (c–e) Heat maps 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 (oex) cells (d) or annotated CpG islands (e). All heat maps are sorted by Suz12 binding in control cells, and one of two independent experiments from separate cultures with similar results is shown. (f,g) Browser tracks of Gpr83 (f) and Inhbb (g), genes that gain ectopic Suz12 binding upon Rnaseh1 overexpression.
by mapping Suz12 binding and H3K27me3 localization in the presence or absence of Rnaseh1 overexpression and observed increased Suz12 and H3K27me3 occupancy in Rnaseh1-overexpressing ESCs (Fig. 5b–e and Supplementary Fig. 7a). Some genes not significantly bound by Suz12 in control cells gained peaks of Suz12 binding (Fig. 5f,g). Moreover, Suz12 enrichment at promoter-proximal regions normally bound by the complex increased two-fold in aggregate upon Rnaseh1 overexpression (Fig. 5b and Supplementary Fig. 7b,c). In support of 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 islands37–39, which make up a large fraction of mammalian promoters and regulatory elements. CpG islands are kept unmethylated, in part, by the presence of R loops17,18, thus suggesting that 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), a result suggesting that R loops produced from nascent transcripts inhibit PRC2 binding to these sites. Finally, we observed examples of genes bound by the Tip60–p400 complex in control ESCs that, upon disruption of R loops by Rnaseh1 overexpression, exhibited both reduction of Tip60–p400 binding to background levels and ectopic PRC2 binding, thus 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, thereby promoting Tip60–p400 association and inhibiting PRC2 association with numerous R loop–associated genes.

R loops are necessary for robust ESC differentiation

Knockdown of Tip60 (official symbol Kat5) or Ep400 in ESCs results in partial defects in both ESC self-renewal and differentiation15,16. 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, as well as a defect in ESC differentiation, but has no effect on self-renewal16. These findings raise the possibility that R loop–deficient ESCs might also be defective in differentiation. To test this possibility, we grew Rnaseh1-overexpressing ESCs in differentiation medium alongside control ESCs and homozygous Ep400-mutant ESCs (Online Methods). In agreement with the differentiation defect previously observed upon knockdown of Ep400 or other Tip60–p400 subunits15,16, after 14 d, a higher abundance of Ep400-mutant cells exhibited clustered (ESC-like) morphology 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). In a more stringent

Figure 6 Disruption of R loops impairs ESC differentiation. (a,b) Alkaline phosphatase staining (a) or Nanog immunofluorescence staining (b) of control, Rnaseh1-overexpressing (oex) or Ep400-mutant ESCs after culture in differentiation-promoting medium for 14 d. White arrowheads in a indicate clusters of robustly Alkaline phosphatase–stained cells in Rnaseh1-overexpressing cells. Scale bars, 200 μm (bottom) in a and 100 μm in b. The indicated percentages of Nanog-expressing cells were averaged from two replicate differentiation experiments (independent cultures on different days). DAPI, 4′,6-diamidino-2-phenylindole. (c) Weights of teratomas derived from subcutaneous injection of control or Rnaseh1-overexpressing ESCs into nude mice (n = 8 tumors per condition). The weight of each tumor is shown as a black dot, and the mean is shown as a red bar. *P < 0.05 by Student’s two-tailed t test. (d) Representative examples of sections (n = 32 per genotype) from teratomas derived from control or Rnaseh1-overexpressing ESCs. Scale bars, 200 μm.

Figure 7 Model of R-loop function. Genes that do not form R loops, owing to either lack of expression or G-poor sequence within the 5′ region of the transcript, are good substrates for PRC2 binding but are 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 moderate levels, owing to low expression and/or weak or moderate G enrichment within the 5′ region of the transcript, are predicted to be relatively poor substrates for both complexes. Red curved lines indicate RNA.
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 the Tip60–p400 subunit Dmap1 (ref. 15), Rnaseh1 overexpression resulted in smaller teratomas (Fig. 6c), which were poorly differentiated in comparison with controls (Fig. 6d). Together, these data suggest that 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 RNAI pause sites at transcriptional termini. In addition, formation of R loops in trans has been observed in some systems and may contribute to the functions of some lncRNAs. Several proteins that resolve or stabilize R loops have been described, thus suggesting that the formation and persistence of R loops is highly regulated. Thus, R-loop accumulation appears to be a function of transcription, RNA sequence and trans-acting DNA-binding 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 regard to these key regulators of ESC pluripotency, R loops help to segregate genes into classes that are highly bound by Tip60–p400 but not PRC2, are highly bound by PRC2 but not Tip60–p400 or are weakly bound by both complexes. Interestingly, at some genes with low DRIP-RNA-seq signals, we observed a significant increase in Suz12 binding upon Rnaseh1 overexpression, thus suggesting that the PRC2 complex may be very sensitive to the presence of R loops, even when they are present at low levels. Conversely, at some genes with high DRIP-RNA-seq signals, we did not observe increased PRC2 binding upon Rnaseh1 overexpression, thus suggesting 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, and whether they further compartmentalize the chromatin structure of genes in ESCs, remain to be tested. However, given the large number of chromatin-regulatory complexes found to bind lncRNAs, it seems likely additional factors will bind nascent transcripts in the form of R loops.

**Context-dependent effects of RNA binding on PRC2 function**

Although 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. Rnaseh1 overexpression mimicked the effect of transcription inhibition on both complexes, enhancing Suz12 association, and inhibiting Tip60–p400 association. Because RNase H1 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.

Although we observed a significant correlation between promoter-proximal R loops and Tip60–p400 binding, several lines of data indicate that R loops are not sufficient for Tip60–p400 recruitment. First, R loops are also prevalent at transcriptional termini, which are not highly bound by Tip60–p400. Together, these data suggest that recruitment of Tip60–p400 to target sites on chromatin is a function of multiple 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 to 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-overexpressing ESCs exhibited impaired differentiation, in agreement 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 probably reflect more than just the effects of reduced Tip60–p400 activity. Accordingly, although 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 that 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 in gene regulation and control of 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.
METHODS

Methods and any associated references are available in the online version of the paper.

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

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

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 (both to T.G.F.) and NIH grant R01HD080224 (to O.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.

AUTHOR CONTRIBUTIONS

P.B.C. performed all experiments and bioinformatics analyses, except as otherwise indicated. 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 analyzed 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.

COMPETING INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Bonasio, R. & Shiekhattar, R. Regulation of transcription by long noncoding RNAs. Annu. Rev. Genet. 48, 433–455 (2014).
2. Rinn, J.L. & Chang, H.Y. Genome regulation by long noncoding RNAs. Annu. Rev. Biochem. 81, 145–166 (2012).
3. Flynn, R.A. & Chang, H.Y. Long noncoding RNAs in cell fate programming and reprogramming. Cell Stem Cell 14, 752–761 (2014).
4. Guttman, M. et al. IncRNAs act in the circuitry controlling pluripotency and differentiation. Nature 477, 295–300 (2011).
5. Rinn, J.L. et al. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Science 322, 750–756 (2008).
6. Zhao, J., Sun, B.K., Erwin, J.A., Song, J.-J. & Lee, J.T. Polycistrons proteins targeted by a short reverse RNA to the mouse X chromosome. Science 322, 750–756 (2008).
7. Pandey, R.R. et al. Ncnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. Mol. Cell 32, 232–246 (2008).
8. Aloia, L., Di Stefano, B. & Di Croce, L. Polycistronic complexes in stem cells and embryonic development. Development 140, 2525–2534 (2013).
9. Cifuentes-Rojas, C., Hernandez, A.J., Sarma, K. & Lee, J.T. Regulatory interactions between RNA and polycistronic repressive complex 2. Mol. Cell 55, 171–185 (2014).
10. Kaneko, S., Son, J., Shen, S.S., Reineberg, D. & Bonasio, R. PRC2 binds active promoters and contacts nascent RNAs in embryonic stem cells. Nat. Struct. Mol. Biol. 20, 1258–1264 (2013).
11. Davidovich, C., Zheng, L., Goodrich, K.J. & Cech, T.R. Promiscuous RNA binding by Polycistronic repressive complex 2. Nat. Struct. Mol. Biol. 20, 1250–1257 (2013).
12. Kaneko, S., Son, J., Bonasio, R., Shen, S.S. & Reineberg, D. Nascent RNA interaction keeps PRC2 activity poised and in check. Genes Dev. 28, 1983–1988 (2014).
13. Riising, E.M. et al. Gene silencing triggers polycistronic repressive complex 2 recruitment to CpG islands genome wide. Mol. Cell 55, 347–360 (2014).
14. Squarrito, M., Gorrini, C. & Anati, B. Tip60 in DNA damage response and growth control: many tricks in one hat. Trends Cell Biol. 16, 433–442 (2006).
15. Fazzio, T.G., Huff, J.T. & Panning, B. An RNAi screen of chromatin proteins identifies Tip60-p400 as a regulator of embryonic stem cell identity. Cell 134, 162–174 (2008).
16. Chen, P.B. et al. Hdac6 regulates Tip60-p400 function in stem cells. eLife 2, e01557 (2013).
of approximately 900 bp surrounding a His6–triple FLAG tag, as described by CRISPR/Cas–mediated genome editing into E14 by using homology arms for the Walker A box of the ATPase domain of the

made similarly, with a guide RNA (TGGCTGATGAAGCAGGGCTT) specific length mouse

and library construction (ChIP-seq) were performed as described previously16. ChIP.

conjugated anti-mouse secondary antibody (1:10,000) (1706515; Bio-Rad).

overnight at 4 °C. The next day, the membrane was washed and stained with HRP

at room temperature and incubated with anti-S9.6 antibody (1:2,000 dilution)

which we observed efficient inhibition of transcription with no effect on protein

the optimal time of inhibitor treatment on the basis of the shortest time within

levels of several subunits in Tip60–p400 by western blotting. We determined

Cell culture and treatment. Mouse ESCs were derived from E14 (ref. 47) and were previously obtained from B. Panning (University of California, San Francisco). Cells have been subjected to extensive sequencing in the course of this and previous studies, thus verifying that they are of mouse origin, and the pluripotency experiments reported in this and previous studies have verified their ESC identity. ESCs were previously tested to ensure that they were free of mycoplasma, and they were grown under feeder-free conditions as previously described14. The homozygous Tip60–FLAG tag was generated by CRISPR/Cas–mediated genome editing into E14 by using homology arms previously19; immediately 5′ of the endogenous stop codon (guide RNA sequence: AAAGTATACGTCTCCATGAGT). The Ep400 homozygous-mutant line was made similarly, with a guide RNA (TGCTTGATGAAGCAGGGCTT) 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 Rasna1 (NM_001286651.1) containing an N-terminal triple-HA tag was synthesized (gBlocks, Integrated DNA Technologies) and cloned into the EcoRI-XhoI fragment of the pCAGGS-ires-Hygro vector. Rasna1-overexpressing cells were generated by transfection of the pCAGGS-Rasna1-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 previously described13. 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 on the basis of the shortest time within which we observed efficient inhibition of transcription with no effect on protein levels of Tip60 subunits. For ESC differentiation, 10% ESC cells were suspended in medium lacking LIF and cultured in non–cell culture treated Petri dishes for 2 d. Subsequently, cells were transfected 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 d of differentiation, cells were stained for AP activity with 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 h, 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) (A11034; 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 previously described15.

Dot blotting. Indicated amounts of DNA were spotted onto a nitrocellulose membrane. After drying, the membrane was blocked in 5% milk for 30 min 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) (1706515; Bio-Rad).

Chromatin immunoprecipitation (ChIP). Chromatin immunoprecipitation and library construction (ChIP-seq) were performed as described previously16. Libraries with different barcodes were pooled, and single-end sequencing (50 bp) was performed on an Illumina HiSeq2000 at the University of Massachusetts Medical School deep-sequencing core facility. ChIP-qPCR was performed as previously described18.

RIP-seq. Cells were lysed with 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 precleared with Protein A magnetic beads (New England Biolabs) for 3 h. Cleared nuclear extract was incubated with specific antibodies in IP buffer (50 mM Tris-HCl, pH 7.4, 250 mM NaCl, and 0.1% Triton X-100) plus 1× HALT protease inhibitors (Thermo Fisher) and SUPERaseIn (Life Technologies) overnight at 4 °C. The next day, prewashed Protein A magnetic beads were added to IP samples and incubated for another 4 h at 4 °C. The magnetic beads were sequentially washed with IP buffer twice, high-salt IP buffer (50 mM Tris-HCl, pH 7.4, 500 mM NaCl, 0.1% Triton X-100, and 0.5% sodium deoxycholate) 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 h. For RIP-seq, 10–50 ng of RIP-enriched RNA and adaptor 1 (5′-CTGAAACGCTTTCCTCCAG CTNNNNNNN-3′) were used 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 cDNA was purified by SILAN beads (Life Technologies). To preserve strand information, adaptor 2 with 5′ phosphorylation and 3′ dideoxy-C modifications (5′-p-NNAAGTAGA TCAGGAAGAGCGTCGTGT -3′) was ligated to the 3′ end of the first-strand cDNA with T4 RNA ligase 1 (New England Biolabs). The ligated material was purified by SILAN beads and PCR-amplified with Illumina primers with 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 previously17 with the following modifications (shown schematically in Supplementary Fig. 3b). The immunoprecipitated material (with and without RNase H treatment) was denatured at 94 °C for 1 min and cooled on ice. To reduce DNA background, the samples were treated with DNase I at 37 °C for 30 min, and RNA was purified with phenol/chloroform/isoamyl alcohol extraction. 38 pmol of adaptor 1 (CTGAACGCTTCCTCCAGCTNN NNNN) was combined with 50 ng of 5′-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 as described previously48.

Sequencing data analysis. Barcodes were removed, and reads were mapped to the mouse genome (mm9 with Bowtie-1.0.0 (ref. 49) for ChIP-seq and TopHat2 (ref. 50) for RNA-seq, RIP-seq, and DRIP-RNA-seq. For ChIP-seq and DRIP-RNA-seq, aligned sequences were processed in HOMER2 with 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 presenting the results in reads per million (RPM). For RIP-seq data, aligned sequences were processed in HOMER with the ‘analyzeRNA’ command to calculate, normalize, and present the results in reads per kilobase per million mapped reads (RPKM) for each reference gene. Fold changes were calculated by dividing the RPKM from experimental IPs by the RPKM from IgG control IPs. Noncoding RNA data were obtained from the GENCODE release M data set52 and previously published IncRNAs4. For RNA-seq, rRNA sequences were removed before transcript quantification with RSEM4. Differentially expressed genes were identified by DESeq2 (ref. 54), and significantly changed genes were selected with a cutoff of adjusted P value <0.05, comparing Rnaseh1-overexpressing cells to control cells at each time point during differentiation.

© 2015 Nature America, Inc. All rights reserved.
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 d. Mice were sacrificed, and tumors were weighed; this was followed by fixation and staining as previously described. All animal experiments were performed according to an approved UMMS animal care and use protocol (2165–13). No statistical method was used to determine 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 data sets, we did not assume equal variances or similar distributions and therefore performed nonparametric tests, such as the Kolmogorov-Smirnov test to assess the 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 Student’s t tests. Because of 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 s.d. in all cases. Injections were performed on genetically identical nude mice, selected at random. Investigators were not blinded during injection of mice or downstream analyses of tumors.

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 326, 292–295 (1987).
48. Hainer, S.J. et al. Suppression of pervasive noncoding transcription in embryonic stem cells by esBAF. Genes Dev. 29, 362–378 (2015).
49. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10, R25 (2009).
50. Kim, D. et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 14, R36 (2013).
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 38, 576–589 (2010).
52. Pervouchine, D.D. et al. Enhanced transcriptome maps from multiple mouse tissues reveal evolutionary constraint in gene expression. Nat. Commun. 6, 5903 (2015).
53. Li, B. & Dewey, C.N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12, 323 (2011).
54. Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550 (2014).