PRC2 binds active promoters and contacts nascent RNAs in embryonic stem cells

Syuzo Kaneko1,2, Jinsook Son1,2, Steven S Shen1–3, Danny Reinberg1,2 & Roberto Bonasio1,2,4

EZH2 is the catalytic subunit of PRC2, a central epigenetic repressor essential for development processes in vivo and for the differentiation of embryonic stem cells (ESCs) in vitro. The biochemical function of PRC2 in depositing repressive H3K27me3 marks is well understood, but how it is regulated and directed to specific genes before and during differentiation remains unknown. Here, we report that PRC2 binds at low levels to a majority of promoters in mouse ESCs, including many that are active and devoid of H3K27me3. Using in vivo RNA-protein cross-linking, we show that EZH2 directly binds the 5' region of nascent RNAs transcribed from a subset of these promoters and that these binding events correlate with decreased H3K27me3. Our findings suggest a molecular mechanism by which PRC2 senses the transcriptional state of the cell and translates it into epigenetic information.
supporting the use of in vivo cross-linking strategies to capture functionally relevant RNA-protein interactions on chromatin.

Because UV cross-linking and immunoprecipitation (CLIP) approaches are extremely sensitive to contamination from abundant RNA-interacting proteins that may remain undetected by western blotting, we generated ESC lines expressing, in an inducible manner, physiological levels of EZH2 fused to three different epitope tags (N3-EZH2, Supplementary Fig. 1a,b), thus allowing for tandem (and triple) affinity purifications, which, we reasoned, would minimize chances of contamination. As previously reported, N-terminal tagging of EZH2 did not perturb its incorporation into the PRC2 complex or its enzymatic activity (data not shown) and did not affect its distribution on chromatin (Supplementary Fig. 1c,d), consistent with our earlier studies13,22,23 and with a recent EM reconstruction of the PRC2 holoenzyme24.

UVC irradiation followed by hemagglutinin (HA) CLIP25 on N3-EZH2–expressing ESCs revealed the presence of a labeled protein migrating slightly above the predicted molecular weight for N3-EZH2. The radiolabeled band was visible only upon induction of N3-EZH2 expression and UVC irradiation (Fig. 1a), and therefore we concluded that it corresponded to RNA cross-linked to EZH2. Because the short-wavelength UVC irradiation appeared to cause extensive photodamage to EZH2 (Fig. 1a), in the following experiments we resorted to photoactivatable ribonucleoside–enhanced CLIP (PAR-CLIP)26, a modification of the CLIP technology that allows for cross-linking with UV light at longer wavelengths and lower energy. After anti-HA immunoprecipitation (IP) in stringent detergent conditions that disrupted the core PRC2 complex (data not shown), autoradiography revealed that EZH2 was cross-linked to 32P-labeled material that was sensitive to RNase treatment (Fig. 1b). We also noted that higher 4-thiouridine (4-SU) concentrations and irradiation with UVB (312 nm) rather than UVA resulted in more efficient cross-linking (Fig. 1b) and chose these conditions for the following experiments. Together, these data support the conclusion that EZH2 establishes direct physical contacts with RNA in ESCs, and these contacts can be identified by CLIP or PAR-CLIP approaches.

To recover RNA amounts sufficient to construct libraries for deep sequencing while minimizing contamination during the IP, we scaled up our PAR-CLIP strategy and performed tandem affinity purification with Streptactin-coupled resin and anti-HA IP (Fig. 1c), in four biological replicates. Deep sequencing of the CLIP tags eluted from the EZH2 bands (Fig. 1d) yielded ~60 million reads, of which ~30% could be aligned to the genome (Supplementary Table 1). Of these, 40% displayed the T-to-C transition diagnostic of 4-SU–mediated cross-linking (Fig. 1e). Initial validations of the alignments confirmed that the replicates were highly concordant (Supplementary Fig. 2a,b) and that known EZH2-interacting lncRNAs, such as Kcnq1ot1 and Meg3 (refs. 16,18), accumulated RCSs called by PARalyzer are shown as red bars. UCSC gene models are displayed. Rep1–rep4, biological replicates.

**Figure 1** EZH2 binds to RNA in mouse ESCs. (a) CLIP blots for HA-tagged EZH2 in control cells and cells induced with doxycycline and before or after irradiation with UVC. The autoradiography is shown at the top, and the approximate position of HA-EZH2 is indicated. The corresponding anti-HA immunoblot (IB) is shown at the bottom. (b) PAR-CLIP autoradiography (top) and western blot (bottom). Different 4-SU concentrations, UV wavelengths and RNase treatments are shown. (c) Scheme of the purification strategy used for PAR-CLIP-seq experiments. Dox, doxycycline. (d) Autoradiography of three biological replicates (rep) used for PAR-CLIP-seq library construction. The dashed red boxes indicate the position of the excised bands. (e) Histogram plot for the mutation frequencies in PAR-CLIP-seq reads. Bars represent the average percentage plus s.d. of unique mapped CLIP tags containing the indicated mutation from the four biological replicates. (f) Genome-browser view of EZH2 CLIP tags mapping to the Meg3 lncRNA (top) or Kcnq1ot1 antisense ncRNA (bottom). The four biological replicates are plotted separately. UCSC genes are displayed. Rep1–rep4, biological replicates.
reported several intronic CLIP tags associated with EZH2 despite shallower sequencing as compared to our PAR-CLIP.29

We confirmed the 5′ bias in EZH2-RNA contacts at the genome-wide level by mapping the RCSs identified above to all mouse transcripts and plotting their distribution over the bodies of the 250 genes that exhibited the strongest signal and thus were more likely to reflect true in vivo interactions (Fig. 2c). This bias was not an artifact of RCS identification, as it was readily visible also when we mapped the raw CLIP tags to the gene models in a similar fashion (Supplementary Fig. 2d). Both CLIP tags containing the T-to-C mutation (diagnostic of RNA-protein cross-links) and CLIP tags not containing the mutation (dashed gray line). (e) Distribution of CLIP tags and RNA-seq reads on the first 50 bp of first exons (E1), first introns (i1), last introns (i2) and last exons (E2) from RefSeq transcripts with at least two exons. Data are shown as mean percentage of total reads mapping to these features from two (RNA-seq) or four (CLIP-seq) biological replicates ± numerical range (RNA-seq) or s.e.m. (CLIP-seq). (f) PAR-CLIP as in Fig. 2d after the 4-SU was chased with U nucleoside for the indicated time. Autoradiography (top) and anti-HA blot (bottom) demonstrating equal protein loading.

EZH2-bound RNAs originate from genes with low PRC2 occupancy

Because CLIP tags mapped equally well to introns and exons (Fig. 2e), we defined a moderately conservative set of EZH2-bound nascent RNAs (henceforth referred to as ezRNAs), on the basis of the presence of at least three separate RCSs in their introns. Using this parameter, we identified 1,108 transcripts originating from 784 genes (Supplementary Table 2). Genes encoding ezRNAs were enriched in gene ontology (GO) terms suggestive of transcriptional control and nuclear processes, including ‘chromatin modification’ and ‘regulation of transcription’ (Fig. 3a and Supplementary Table 3).

Having observed EZH2 peaks at the promoters of visually inspected ezRNA+ genes (Fig. 2b and Supplementary Fig. 2c), we wished to determine whether this was a general feature of RNAs cross-linked to EZH2, despite the fact that, at least on the basis of GO enrichment, ezRNA+ genes constituted a set distinct from traditional Polycomb-target genes.30 Indeed, EZH2 CLIP tags tended to accumulate near chromatin regions enriched for EZH2 (Fig. 3b). However, plotting of the distribution of ezRNAs on a heat map of all RefSeq TSSs sorted by PRC2 occupancy revealed that they were missing not only from regions with no detectable PRC2 (Fig. 3c), as expected, but also from regions with very high PRC2 levels (Fig. 3c). In addition, they originated mostly from TSSs with intermediate levels of PRC2, which, to our surprise, comprised a majority of all annotated TSSs (Fig. 3c). However, when the TSSs were sorted by H3K27me3 density, the ezRNA-producing TSSs clustered in the rightmost section of the heat map, in regions depleted for this histone mark (Fig. 3d). We validated this observation by analyzing a different ChIP-seq dataset generated in the same cell line and under similar culture conditions. Consistent with our findings above, ezRNAs mapped to genes with intermediate levels of EZH2 and low levels of H3K27me3 also in this data set (Supplementary Fig. 4a,b). Despite the presence of PRC2, genes giving rise to ezRNAs were transcriptionally active, as demonstrated by the enrichment of H3K4me3 and H3K36me3 relative to that in
Figure 3  EZH2-bound nascent RNAs originate from PRC2+ and H3K27me3− promoters. (a) Bar plot for the −log10 of the P value of the top 15 most-enriched GO terms in genes producing ezRNAs, as determined by the hypergeometric distribution. Nonmembrane-b. org., nonmembrane-bound organelle; intracell., intracellular; neg. reg., negative regulation. (b) Density profile of raw CLIP tags (after duplicate removal and repeat masking) relative to the top 20,000 EZH2 ChIP enriched regions (ERs). As a control, the density of the same CLIP tags calculated on 20,000 random genomic regions of comparable size (dashed gray line) is shown. (c) Heat map (bottom) showing EZH2 densities at all unique RefSeq TSSs ± 2.5 kb, sorted by EZH2 occupancy. The distribution of ezRNA-producing TSSs within this heat map is indicated by individual bars (middle) and a density plot (top, black line) and compared to their density over a randomly permuted heat map (top, dashed gray line). Data are from two biological replicates. (d) Same as c but for H3K27me3. The heat map shows the average of three biological replicates. (e) Heat maps for H3K4me3 and H3K36me3 occupancy, comparing the 1,108 genes producing ezRNAs and a random gene set. Data were obtained from GSM590111 and GSM590119 (ref. 31). (f) PRC2 occupancy in 5-kb windows at TSSs from different positions along the EZH2 gradient in the heat map (top). The y axis represents reads per 10 million mapped (RP10M), and the scale is indicated by the number to the left of each profile. In all cases, the input was scaled at the highest magnification for meaningful comparison.

A random gene set of equal size (Fig. 3c), and they were largely distinct from bivalent genes31,32 (Supplementary Fig. 4c,d). The presence of EZH2 at such a large proportion of genes, including actively transcribed ones, was unexpected; therefore, we sought confirmatory evidence for this finding. Consistent with the results above, SUZ12, another core subunit of PRC2, was also present at a majority of TSSs (Supplementary Fig. 4e), although the overall enrichment with antibody to SUZ12 was considerably lower (Fig. 3f). Moreover, the ChIP-seq profiles obtained with antibodies to HA in N3-EZH2–expressing mouse ESCs were indistinguishable from those obtained with antibodies to EZH2 in the parental cell line (Supplementary Fig. 1c,d), and in both cases peaks of considerable intensity were easily discernible in the promoter regions of genes that would be traditionally considered PRC2 negative (Fig. 3f).

In summary, the results above show that ezRNAs originate from a subset of genes that, although occupied by intermediate levels of PRC2, are actively transcribed in mouse ESCs and are devoid of H3K27me3.

Genes producing ezRNAs have lower H3K27me3 levels in ESCs Because genes producing ezRNAs were decorated with chromatin marks consistent with productive transcription (Fig. 3e), and their distribution correlated negatively with H3K27me3 density at a genomewide level (Fig. 3d), we hypothesized that, despite the presence of PRC2, their promoters might be devoid of H3K27me3. Indeed, when we compared the set of genes producing ezRNAs to a control gene set, matched gene by gene for the amount of PRC2 at their promoter, we found that the latter had higher amounts of H3K27me3 (Fig. 4a) despite having equal amounts of PRC2 (Fig. 4a). In accordance with the lower levels of H3K27me3, these genes were expressed at higher levels than in the control set, as shown by an average increase in RNA-seq reads per kilobase per million fragments mapped (RPKM) (Fig. 4a). When we matched the control by equalizing RPKMs rather than PRC2 levels, the difference in H3K27me3 was reduced but still noticeable (Fig. 4b), thus suggesting that differences in RNA transcription alone are not sufficient to explain the changes in H3K27me3 levels and that contacts between EZH2 and nascent RNAs may influence the function of PRC2 on chromatin. Differences in the average H3K27me3 profiles were not due to a few outliers, because regression curves fitted through the individual points had different slopes, thus confirming that in general ezRNA+ genes contained less H3K27me3 than expected on the basis of PRC2 levels alone (Supplementary Fig. 5a). The differences in H3K27me3 at these genes have also been observed in ChIP-seq profiles generated in different laboratories and even with different ESC lines and different culture conditions31 (Supplementary Fig. 5b).

Finally, we wished to determine the fate of genes producing ezRNAs in committed cells that had lost the pluripotent characteristics of ESCs. To this end, we compared the H3K27me3 distribution on genes producing ezRNAs in a different pluripotent mouse ESC line (V6.5) and in differentiated mouse embryonic fibroblasts (MEFs), using data sets from Mikkelsen et al.32. The difference in H3K27me3 densities on ezRNAs genes versus on EZH2-matched controls was pronounced in the pluripotent V6.5 ESCs (Fig. 4c), but it was much decreased in MEFs (Fig. 4c), thus suggesting that upon differentiation...
several of the genes that produced ezRNAs in ESCs can become fully repressed by acquiring H3K27me3. Presumably, repression caused decreased synthesis of ezRNAs and consequently increased deposition of H3K27me3 by PRC2, but only EZH2 PAR-CLIP in MEFs could confirm this hypothesis. Nonetheless, the accumulation of H3K27me3 at ezRNA+ genes proves that the lack of PRC2 activity at those genes in ESCs was not due to any intrinsic property of their promoters but rather was a consequence of dynamic processes, specific to ESCs, possibly involving the production and recognition of ezRNAs.

**DISCUSSION**

By performing PAR-CLIP-seq for EZH2 in undifferentiated mouse ESCs (Fig. 1), we have identified a family of nascent transcripts that bind to EZH2 while they are being transcribed and before they are spliced and processed into mature mRNAs (Fig. 2). The majority of these nascent transcripts originate from transcriptionally active regions containing intermediate or low levels of PRC2 (Fig. 2), as detected in multiple biological replicates and with different antibodies in our data (Fig. 3f), as well as in data sets generated by other groups (Supplementary Fig. 4). A peculiar characteristic of these ezRNA-producing genes is the lack of H3K27me3 despite measurable levels of PRC2 (Fig. 4); this hints at the possibility that PRC2 might function as an RNA ‘sensor’ (described below).

One conclusion of our study is that PRC2, at least in mouse ESCs, binds to a majority of promoters. In addition to being present at regions where PRC2 is entrenched and large amounts of H3K27me3 accumulate (Fig. 3c), low but detectable levels of EZH2 are present at a majority of TSSs (Fig. 3c,f), including many that have no detectable H3K27me3. The regions highly enriched in PRC2 correlate with strong repression and contain the gene targets that have traditionally been associated with PcG function in ESCs, including developmental regulators such as the Hox genes. At these genes, heterochromatin

![Figure 4](image)

**Figure 4** Decreased levels of H3K27me3 at ezRNA+ genes. (a) Density profiles for H3K27me3 in E14 mouse ESCs (left), spanning the promoters of 1,108 ezRNA+ transcripts (black line) or an equally sized control set of promoters (gray dashed line), matched individually for EZH2 occupancy (middle). Log-converted RPKM values for the two sets of transcripts are shown in the violin plot (right). The plots show the average of three (H3K27me3) or two (EZH2, RNA-seq) biological replicates. (b) Same as a, but control genes were selected by equalizing RPKM levels on a transcript-by-transcript basis. (c) H3K27me3 density plot calculated as in a for V6.5 ESCs (left) and MEFs (right). Data were obtained from GSE12241 (ref. 32).

![Figure 5](image)

**Figure 5** PRC2 senses transcriptional activity at ESC promoters. (a) In ESCs, PRC2 binds to a majority of TSSs independently of their transcriptional state, sampling their activity. (b) Regulation of PRC2 activity occurs during the sensing phase, when PRC2 binds to nascent ezRNAs that somehow impede the deposition of H3K27me3 on chromatin. This could happen by either (1) inhibition of PRC2 activity, possibly dependent on additional in vivo factors or (2) eviction of PRC2 from chromatin or of other factors required for H3K27me3 deposition. This allows the cell to continue expressing genes for which activating transcription factors are present, despite the constant presence of PRC2. (c) During differentiation, lineage-specific transcriptional repressors are upregulated and silence transcriptional activity at selected target genes. Lacking the inhibitory ezRNAs, PRC2 initiates deposition of H3K27me3, thus establishing the nucleating conditions for the formation of facultative heterochromatin. (d) By the time the differentiation program is completed, a positive feedback loop of H3K27me3 deposition, EED binding, PRC2 recruitment and more H3K27me3 deposition gives rise to stable and self-perpetuating facultative heterochromatin, in which transcriptional repression is maintained even after the original transcription factor ceases to be expressed, PRC2hi, high occupancy of PRC2; H3K27me3hi, high density of H3K27me3; PRC0, low occupancy of PRC2.
has already been nucleated and is probably propagated by the known positive feedback cycles affecting PRC2 (ref. 8). However, the promoters with low PRC2 occupancy appear to be transcriptionally competent if not completely activated, and this is in contrast with the traditional view that PRC2 enforces epigenetic repression. What might be the reason for this constitutive and broad TSS occupancy, and how can the transcription of these genes despite the presence of PRC2 be explained?

Given that the function of the PcG axis is that of maintaining—not establishing—transcriptional repression33, the results presented above suggest a model wherein PRC2 exerts its epigenetic function by sensing the activation states of promoters through contacts with nascent RNAs (Fig. 5). The first step of this model is ‘promoter sampling’, i.e., the interaction of PRC2 with a majority of TSSs, at least in pluripotent mouse ESCs, through low-affinity or low-frequency binding events (Fig. 5a), possibly directed by GC-rich sequences34, other protein factors22,35,36 and even ncRNAs34,14,16,18. This broad promoter surveillance gives PRC2 the opportunity to survey the promoter state of ezRNA + promoters by not depositing H3K27me3, and this is open questions of H3K27me3, thus allowing chromatin to stay open and the promoter to remain active. Whether and how exactly ezRNA-EZH2 contacts directly cause decreased H3K27me3 are open questions that will require further investigation. It is tempting to speculate that RNA contacts directly inhibit the enzymatic activity of EZH2 (Fig. 5b, step 1), but we can also foresee more complex scenarios, in which one or more components of the PRC2 complex might be evicted from chromatin through interactions with or competition from ezRNAs (Fig. 5b, step 2).

Whatever the mechanism may be, PRC2 responds to the activated state of ezRNA promoter by not depositing H3K27me3, and this allows the gene to remain active; however, the Polycomb axis is better known for maintenance of repression, not activation. How does this fit our model? We propose that when the cell detects a new stimulus, for example during differentiation, DNA-binding transcriptional repressors are expressed de novo or are otherwise activated, enter the nucleus and bind to chromatin and silence their target genes, thus abolishing the production of nascent RNAs at these loci. Without inhibitory ezRNAs, PRC2 complexes that are still surveying these loci deposit increased amounts of H3K27me3 to establish repressed chromatin structures (Fig. 5c). As H3K27me3 marks accumulate, a tipping point is reached when heterochromatic structures are maintained in an epigenetic manner3, even in the absence of the original transcriptional repressor (Fig. 5d).

Although the physical contacts between RNA and EZH2 reported herein are highly suggestive of a direct regulatory mechanism, we cannot exclude that other molecular features associated with ongoing transcription may also contribute to the attenuation of PRC2 activity. For example, we and others recently showed that, in certain contexts, the presence of H3K4me3 inhibits the enzymatic activity of PRC2 (refs. 37,38). This regulatory mechanism may function in addition to or in cooperation with the recognition of ezRNAs.

Our results and the model presented above are consistent with the genetic evidence that PRC2 does not repress transcription de novo but rather senses loci that are already silenced and maintains their repressed state over time and throughout cell division33,39. In Drosophila, PRC2 binds to the Polycomb-responsive element of the Ultrabithorax gene (Ubx) independently of its activation state, but H3K27me3 spreads to the promoter and body of the gene only in imaginal discs that repress Ubx40. In fact, transient activation of a reporter locus controlled by Polycomb is sufficient to switch its epigenetic state from repressed to active41, and this fits well with our proposed model in which PRC2 is inhibited by active transcription.

Although we focused our attention on the 784 top ezRNA+ candidates, to our knowledge this is the first report of genome-wide PAR-CLIP for a noncanonical RNA-binding protein (the second counting Gull et al.29, who performed CLIP without 4-SU), and we must consider the possibility that other nascent RNAs that we cannot detect with the current protocol might bind to EZH2 in vivo. In fact, according to our model, binding to RNA is likely to be very promiscuous in vivo because the presence of RNA, not necessarily specific sequences or structures, would suffice to deliver a regulatory signal to PRC2. This would help explain the large number of EZH2-interacting RNA species recovered by previous studies17,18, and the promiscuity of EZH2-RNA interactions in vitro42, but it does not exclude that other RNAs, particularly IncRNAs, may have evolved to further exploit this property of EZH2 in the context of complex regulatory scenarios such as in X-chromosome inactivation43 or imprinting16,18.

In conclusion, we propose that, at least in mammals, PRC2 might perform its epigenetic function by sensing active transcription through interactions with nascent RNAs and curtailing its activity on chromatin in their presence. The increased facility by which the genomes of ESCs can be edited will soon offer the opportunity to test this hypothesis experimentally.

METHODS

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

Accession codes. Sequencing data have been deposited in the National Center for Biotechnology Information Gene Expression Omnibus as SuperSeries GSE49435.

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

ACKNOWLEDGMENTS

We thank W.W. Tee, L. Vales and P. Voigt for their critical assessment of the manuscript, T. Cech (University of Colorado Boulder) for discussions and the Genome Technology Center at New York University for help with sequencing. This work was supported by grants from the US National Institutes of Health (GM-6844 and R37-37120 to D.R.) and the Howard Hughes Medical Institute (to D.R.). R.B. was supported by a Helen Hay Whitney Foundation postdoctoral fellowship and by the Helen L. and Martin S. Kimmel Center for Stem Cell Biology postdoctoral fellow award.

AUTHOR CONTRIBUTIONS

S.K., J.S., S.S.S., D.R. and R.B. designed and performed experiments, analyzed data and wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

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

1. Bonasio, R., Tu, S. & Reinberg, D. Molecular signals of epigenetic states. Science 330, 612–616 (2010).
2. Margueron, R. & Reinberg, D. The Polycomb complex PRC2 and its mark in life. Nature 469, 343–349 (2011).
3. Ringrose, L. & Paro, R. Epigenetic regulation of cellular memory by the Polycomb and Trithorax group proteins. Annu. Rev. Genet. 38, 413–443 (2004).
4. Cao, R. et al. Role of histone H3 lysine 27 methylation in Polycomb-group silencing. Science 298, 1039–1043 (2002).
5. Czermin, B. et al. Drosophila enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. Cell 111, 185–196 (2002).
ARTICLES

6. Kuzmichev, A., Nishioka, K., Erdjument-Bromage, H., Tempst, P. & Reinberg, D. Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. Genes Dev. 16, 2893–2905 (2002).

7. Müller, J. et al. Histone methyltransferase activity of a Drosophila Polycomb group repressor complex. Cell 111, 197–208 (2002).

8. Margueron, R. et al. Role of the polycomb protein EED in the propagation of repressive histone marks. Nature 461, 762–767 (2009).

9. Lewis, P.W. et al. Inhibition of PRC2 activity by a gain-of-function H3 mutation found in pediatric glioblastoma. Science 340, 857–861 (2013).

10. Ernst, T. et al. Inactivating mutations of the histone methyltransferase gene EZH2 in myeloid disorders. Nat. Genet. 42, 722–726 (2010).

11. Nikoloski, G. et al. Somatic mutations of the histone methyltransferase gene EZH2 in myelodyplastic syndromes. Nat. Genet. 42, 665–667 (2010).

12. Shih, A.H., Abdel-Wahab, O., Patel, J.P. & Levine, R.L. The role of mutations in epigenetic mechanisms. Cell 149, 590–604 (2012).

13. Margueron, R. et al. Enhancer of Zeste protein and Trithorax group proteins. Annu. Rev. Genet. 48, 413–443 (2004).

14. Ringrose, L. & Paro, R. Epigenetic regulation of cellular memory by the Polycomb and Trithorax group proteins. Annu. Rev. Genet. 38, 1166–670 (2012).

15. Shen, J. & Levine, R.L. Histone demethylase NO66 to embryonic stem cell genes during differentiation. Nat. Struct. Mol. Biol. 19, 1273–1281 (2012).

16. Kim, H., Kang, K. & Kim, J. AEBP2 as a potential targeting protein for Polycomb complex repressor complex 2 and demethylase NO66 to embryonic stem cell genes during differentiation. Nat. Struct. Mol. Biol. 19, 1273–1281 (2012).

17. Davidovich, C., Zheng, L., Goodrich, K.J. & Cech, T.R. Short RNAs are transcribed from repressed polycomb target genes and up-regulates its binding to ncRNA. Genes Dev. 24, 2615–2620 (2010).

18. Ciferri, C. et al. Molecular architecture of human polycomb repressive complex 2. eLife 1, e00005 (2012).

19. Klattenhoff, C.A. & Rinn, J.L. Short-read sequence data. Genome Biol. 12, R79 (2011).

20. Guttman, M. et al. lincRNAs act in the circuitry controlling pluripotency and differentiation. Nature 477, 295–300 (2011).

21. Zhang, J., Lu, J., Wang, Y., et al. Inhibition of PRC2 activity by a gain-of-function H3 mutation found in pediatric glioblastoma. Science 340, 857–861 (2013).

22. Li, G. et al. Short RNA targets of the polycomb repressive complex 2. Mol. Cell 1311–1323 (2007).

23. Li, G. et al. Short repeat RNA to the mouse X chromosome. Science 322, 750–756 (2008).

24. Panide, R.R. et al. Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. Mol. Cell 32, 232–246 (2008).

25. Jensen, K.B. & Darnell, R.B. CLIP: crosslinking and immunoprecipitation of in vivo RNA targets of RNA-binding proteins. Methods Mol. Biol. 488, 85–98 (2008).

26. Hafner, M. et al. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell 141, 129–141 (2010).

27. Corcoran, D.L. et al. PARalyzer: definition of RNA binding sites from PAR-CLIP short-read sequence data. Genome Biol. 12, R79 (2011).

28. Gutman, M. et al. lincRNAs act in the circuitry controlling pluripotency and differentiation. Nature 477, 295–300 (2011).

29. Guil, S. et al. Intronic RNAs mediate EZH2 regulation of epigenetic targets. Nat. Struct. Mol. Biol. 19, 664–670 (2012).

30. Boyer, L.A. et al. Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature 441, 349–353 (2006).

31. Marks, H. et al. The transcriptional and epigenomic foundations of ground state pluripotency. Cell 149, 590–604 (2012).

32. Mikkelsen, T.S. et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature 448, 553–560 (2007).

33. Ernst, T. et al. Genome-wide identification of polycomb-associated RNAs by RIP-seq. Mol. Cell 40, 939–953 (2010).

34. Zha, J. et al. Genome-wide identification of polycomb-associated RNAs by RIP-seq. Mol. Cell 40, 939–953 (2010).

35. Zheng, L., Goodrich, K.J. & Cech, T.R. Short RNAs are transcribed from repressed polycomb target genes and up-regulates its binding to ncRNA. Genes Dev. 24, 2615–2620 (2010).

36. Papp, B. & Müller, J. Histone trimethylation and the maintenance of transcriptional ON and OFF states by trxG and PcG proteins. Genes Dev. 20, 2041–2054 (2006).

37. Schmitz, F., Stoddard, B., et al. Histone methylation by PRC2 is inhibited by active chromatin marks. Mol. Cell 42, 330–341 (2011).

38. Schmitz, F., Stoddard, B., et al. Histone methylation by PRC2 is inhibited by active chromatin marks. Mol. Cell 42, 330–341 (2011).

39. Struhl, G. & Akam, M. Altered distributions of Ultrabithorax transcripts in extra sex comb mutants embryos of Drosophila. EMBO J. 4, 3259–3264 (1985).

40. Papp, B. & Müller, J. Histone trimethylation and the maintenance of transcriptional ON and OFF states by trxG and PcG proteins. Genes Dev. 20, 2041–2054 (2006).

41. Kim, H., Kang, K. & Kim, J. AEBP2 as a potential targeting protein for Polycomb Repression Complex 2. Nucleic Acids Res. 37, 2940–2950 (2009).

42. Voigt, P. et al. Asymmetrically modified nucleosomes. Cell 151, 181–193 (2012).

43. Schmitz, F., Stoddard, B., et al. Histone methylation by PRC2 is inhibited by active chromatin marks. Mol. Cell 42, 330–341 (2011).

44. Struhl, G. & Akam, M. Altered distributions of Ultrabithorax transcripts in extra sex comb mutants embryos of Drosophila. EMBO J. 4, 3259–3264 (1985).

45. Papp, B. & Müller, J. Histone trimethylation and the maintenance of transcriptional ON and OFF states by trxG and PcG proteins. Genes Dev. 20, 2041–2054 (2006).

46. Schmitz, F., Stoddard, B., et al. Histone methylation by PRC2 is inhibited by active chromatin marks. Mol. Cell 42, 330–341 (2011).

47. Struhl, G. & Akam, M. Altered distributions of Ultrabithorax transcripts in extra sex comb mutants embryos of Drosophila. EMBO J. 4, 3259–3264 (1985).

48. Papp, B. & Müller, J. Histone trimethylation and the maintenance of transcriptional ON and OFF states by trxG and PcG proteins. Genes Dev. 20, 2041–2054 (2006).

49. Schmitz, F., Stoddard, B., et al. Histone methylation by PRC2 is inhibited by active chromatin marks. Mol. Cell 42, 330–341 (2011).

50. Struhl, G. & Akam, M. Altered distributions of Ultrabithorax transcripts in extra sex comb mutants embryos of Drosophila. EMBO J. 4, 3259–3264 (1985).
ONLINE METHODS

Original images of gels, autoradiographs and blots used in this study can be found in Supplementary Figure 6. Publicly available data sets used in this paper were obtained from series GSE23943 and GSE12241 at the NCBI GEO and from the mouse ENCODE project.

Antibodies. For the purpose of CLIP and PAR-CLIP, we used 2–10 µg of anti-HA antibody (Abcam ab9110, lot GR98618-2). For ChIP-seq, we used 2 µg of anti-HA (Abcam ab9110, lot GR98618-2), EZH2 (ref. 13), SUZ12 (Cell signaling 3737, lot 2) and H3K27me3 (Millipore 07-449, lot 1999681). For western blotting, we used an antibody against EZH2 (Abcam ab9110, lot GR98618-2), EZH2 (ref. 13), SUZ12 (Cell signaling 3737), β-tubulin (Abcam ab6046), and HA (Covance MMS-101P), diluted 1:1000. More detailed information about antibody validation and references are available on the manufacturers’ websites.

Cell culture. E14Tg2A.4 mouse ESCs were cultured as described.

Lentiviral transduction. For production of lentiviruses, viral vector and packaging plasmids were cotransfected in HEK293T cells with Lipofectamine 2000 (Invitrogen) according to the manufacturer’s instructions. After 48–72 h, cell culture medium containing lentiviruses was collected and purified with Fast-Track Virus Purification and Concentration Kit (Millipore). Lentiviral transduction of mESCs was carried out in the presence of polybrene (5 µg/ml).

Knockdown of Ezh2 and expression of epoxide-tagged EZH2. To establish conditional knockdown of Ezh2 in E14 mESCs, pTRIPZ lentiviral-inducible shRNAAmir targeting human and mouse Ezh2 were purchased from Open Biosystems (RH56966-96935303). Lentiviruses were prepared as described above. Selection was performed with puromycin (1 µg/ml), and clones were screened by RFP expression 3–4 d after doxycycline (1 µg/ml) induction. To replace the endogenous copy with an epoxide tag–encoding Ezh2, we cloned mouse Ezh2 into a pINTA-N3 vector, based on the Tet-On 3G system (Clontech) and encoding a pINTA-N3-EZH2; two biological replicates for SUZ12 (two separate ChIPs from the E14 parental line); three biological replicates for H3K27me3 (two from separate ChIPs from the E14 parental line and one from dox-induced E14::N3-EZH2).

CLIP, PAR-CLIP and PAR-CLIP-seq. For CLIP experiments, ESCs were cross-linked with 400 µl/cm² UVC (254 nm) and then processed as described below. For PAR-CLIP, ESCs were grown in standard conditions and pulsed with 500 µM 4-SU (Sigma) for 2 h. In some experiments the 4-SU was chased with 500 µM U nucleoside. After plates were washed with PBS, RNA-protein cross-links were generated by irradiation of 4-SU–treated cells with 400 µl/cm² UVB (312 nm) with a Stratalinker UV cross-linker (Stratagene, CA). These conditions are different from those previously reported and were optimized to maximize cross-linking of RNA to EZH2 (Fig. 1b), a noncanonical RNA-binding protein. Although Rabani et al. observed 5’-biased 4-SU incorporation at the ifih1 gene after a 45-min pulse, it was only at the first time point after gene induction, thus suggesting that the 5’ bias reflected the increased rate of initiation events rather than a predisposition of RNAPII to incorporate 4-SU at the beginning of transcriptional units. Indeed, in the same experiment, later time points showed homogeneous 4-SU incorporation across the body of the same transcript. Incubation with 4-SU for 2 h was previously shown to label RNA efficiently in a broad range of cells, and it is considerably longer than the half-life of a majority of mature mRNAs and virtually all pre-mRNAs. These optimized PAR-CLIP conditions may be useful to determine contacts with noncoding RNAs for other noncanonical RNA-binding proteins found in chromatin.

Whole cell extracts were obtained by incubation of cells for 10 min at 37 °C in an appropriate volume of CLIP buffer (20 mM HEPES, pH 7.4, 5 mM EDTA, 150 mM NaCl and 2% lauryl-dimethylbetaine) supplemented with protease inhibitors, 20 U/ml Turbo DNase (Life Technologies), and 200 U/ml murine RNase inhibitor (New England Biolabs), and lysates were cleared by centrifugation. Prior to immunoprecipitation, epoxide-tagged EZH2 was affinity purified with StrepTactin beads (IBA) and eluted with CLIP buffer containing 2 mM biotin (Sigma). Immunoprecipitations were carried out with ChIP-grade anti-HA antibody (Abcam) in the same CLIP buffer for 16 h at 4 °C, after which, when required, the extracts were treated with various concentrations of RNaseA A + T1 cocktail (Ambion) for 5 min at 37 °C. Immunocomplexes were recovered by adding protein G–coupled Dynabeads (Life Technologies) for 45 min at 4 °C. Contaminating DNA was removed by treatment of the beads with Turbo DNase (2 U in 20 µl). Cross-linked RNA was labeled by successive incubation with 5 U Antarctic phosphatase (New England Biolabs) and 5 U T4 PNK (New England Biolabs) for 30 min at 37 °C. After centrifugation, supernatants were incubated with Dynabeads protein G (Invitrogen) for 30 min at 4 °C. Beads were washed twice with ChIP buffer, treated with RNase A at 37 °C for 15 min, and washed three more times with ChIP buffer. Immunocomplexes were heated at 65 °C for ~16 h and incubated with proteinase K for 15 min at 55 °C. Immunoprecipitated DNA was extracted by phenol-chloroform and precipitated with ethanol.

ChIP and ChIP-seq. ChIP from mouse ESCs was performed as described with minor modifications. Briefly, cells were cross-linked with 1% formaldehyde for 10 min at room temperature. Cross-linking was quenched by addition of glycine, pH 7.0, to 200 mM final concentration. Cells were washed three times with ice-cold PBS and harvested in ChIP buffer (50 mM Tris-HCl, pH 7.9, 150 mM NaCl, 1% Triton X-100, 0.5% NP-40, 5 mM EDTA, pH 8.0, 1 µg/ml PMSF and protease inhibitor (Sigma, P8849)). Samples were sonicated with a Diagenode Bioruptor to generate DNA fragments of 400 base pairs for quantitative PCR (qPCR) and ~200 base pairs for library construction. For the purpose of ChIP and ChIP-seq, we used 2–10 µg of anti-HA (Abcam ab9110, lot GR98618-2), EZH2 (ref. 13), SUZ12 (Cell signaling 3737), β-tubulin (Abcam ab6046), and HA (Covance MMS-101P), diluted 1:1000. Knockdown of endogenous Ezh2 and simultaneous rescue with epoxide tag–encoding Ezh2 were induced by treatment with doxycycline (1 µg/ml) for 72 h.

ChIP-seq. RNA-seq was performed in duplicates in two different E14 subclones. RNA was isolated with TRizol (Invitrogen). Poly(A)+ RNA was purified with Oligo(dT)25 Dynabeads (Invitrogen) and converted to cDNA with preservation of strand information by the dUTP method. Libraries were constructed from cDNA with the same protocol as described above for ChIP-seq libraries.

Bioinformatic analyses. In all cases, reads were mapped to the mm9 version of the mouse genome, and gene annotations were obtained from ENSEMBL (v67) or RefSeq (downloaded on 10 July 2013). Except for Figure 2a, in which all features from ENSEMBL were considered, all other analyses concentrate on the set of 26,925 protein-coding transcripts present in both ENSEMBL and RefSeq databases and having unique TSSs and TTSs.

ChIP-seq analysis was performed as described before with modifications. Sequenced reads from ChIP-seq experiments were mapped with Bowtie with
parameters -v2 -m4–best. Normalized genome-wide read-count densities (typically reads per 10 million mapped, RP10M) were computed with samtools and visualized on the UCSC genome browser (http://genome.ucsc.edu/) after reads were extended to the estimated size of the ChIP fragment (typically 200 nt). Enriched regions (ERs) were identified with MACS 2.0 with the relevant input as control and default parameters. For the heat maps shown in Figure 3 and Supplementary Figure 4, the extended reads were mapped to nonoverlapping 25-bp bins spanning 5 kb centered around the TSS and then normalized binwise to percentage of maximum. Density plots (for example, in Fig. 4) were generated similarly, except that the window spanned 10 kb (divided in 50-bp bins) around the TSS, and the mean for each given bin in all genes considered was plotted on the y axis. The correlation of binding affinities in Supplementary Figure 1d was calculated with DiffBind.

For the initial PAR-CLIP-seq mapping and RCS identification, we followed Corcoran et al. Briefly, we clipped adapter sequences from PAR-CLIP reads and kept those longer than 17 nt. The resulting reads were collapsed to remove duplicate sequences with the fastx toolkit, then mapped with bowtie -v2 -m40–best–strata to the mm9 assembly. We pooled the four separate replicates and processed them with PARalyzer, requiring at least two independent T-to-C conversions; this resulted in the identification of 13,764 RCSs. Of these, 2,261 overlapped repetitive elements listed by repeat masker (as downloaded from the UCSC genome browser website on 29 May 2013) and were discarded, leaving 11,503 nonrepetitive RCSs. For CLIP-tag analyses (Figs. 2d,e and 3b; Supplementary Figs. 2a,b,d and 3b), reads were preprocessed in a similar way, with the exception that duplicate removal was performed after bowtie mapping.

RNA-seq reads were trimmed to 30 nt, mapped with bowtie -v2 -m40–best and assigned to gene models with DEGseq.

GO enrichment analyses were performed with DAVID (http://david.abcc.ncifcrf.gov/) with default parameters.

44. Kaneko, S., Rozenblatt-Rosen, O., Meyerson, M. & Manley, J.L. The multifunctional protein p54nrb/PSF recruits the exonuclease XRN2 to facilitate pre-mRNA 3′ processing and transcription termination. *Genes Dev.* 21, 1779–1789 (2007).
45. Gao, Z. et al. PCGF homologs, CBX proteins, and RBBP define functionally distinct prc1 family complexes. *Mol. Cell* 45, 344–356 (2012).
46. Hafner, M. et al. PAR-CLIP: a method to identify transcriptome-wide the binding sites of RNA binding proteins. *J. Vis. Exp.* 41, 2034 (2010).
47. Rabani, M. et al. Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. *Nat. Biotechnol.* 29, 436–442 (2011).
48. Dölken, L. et al. High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay. *RNA* 14, 1959–1972 (2008).
49. Parkhomchuk, D. et al. Transcriptome analysis by strand-specific sequencing of complementary DNA. *Nucleic Acids Res.* 37, e123 (2009).
50. 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).
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. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 9, R137 (2008).
53. Ross-Innes, C.S. et al. Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* 481, 389–393 (2012).
54. Wang, L., Feng, Z., Wang, X. & Zhang, X. DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics* 26, 136–138 (2010).
55. Huang, D., Sherman, B.T. & Lempicki, R.A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4, 44–57 (2009).