Published as: *Nat Struct Mol Biol*. 2013 November ; 20(11): 1250–1257.

Promiscuous RNA binding by Polycomb Repressive Complex 2

Chen Davidovich[1,2], **Leon Zheng**[3], **Karen J. Goodrich**[1,2], and **Thomas R. Cech**[1,2,4]

[1]Department of Chemistry & Biochemistry, Howard Hughes Medical Institute, University of Colorado, Boulder, Colorado, 80309, USA

[2]BioFrontiers Institute, University of Colorado, Boulder, Colorado, 80309, USA

[3]Medical Scientist Training Program, University of Colorado School of Medicine, Aurora, Colorado, 80045, USA

Abstract

Polycomb repressive complex-2 (PRC2) is a histone methyltransferase required for epigenetic silencing during development and cancer. Long non-coding RNAs (lncRNAs) recruit PRC2 to chromatin, but the general role of RNA in maintaining repressed chromatin is unknown. Here we measure the binding constant of human PRC2 to various RNAs and find comparable affinity for human lncRNAs targeted by PRC2 and irrelevant transcripts from ciliates and bacteria. PRC2 binding is size-dependent, with lower affinity for shorter RNAs. *In vivo*, PRC2 predominantly occupies repressed genes; PRC2 is also associated with active genes, but most of these are not regulated by PRC2. These findings support a model in which promiscuous binding of PRC2 to RNA transcripts allows it to scan for target genes that have escaped repression, leading to maintenance of the repressed state. Such RNAs may also provide a decoy for PRC2.

Polycomb repressive complex-2 (PRC2) is a histone methyltransferase that mono-, di- and tri-methylates lysine 27 of histone H3 (H3K27me3), inducing repressed chromatin[1]. EZH2 is the catalytic subunit of PRC2, and SUZ12 is an essential regulatory subunit[2]. EED is a histone-binding subunit that binds H3K27me3-modified histone tails, resulting in increased affinity to nucleosomes and stimulation of the catalytic activity of PRC2[3,4]. PRC2 is further activated by dense chromatin[5]. Marks for active chromatin, H3K4me3 and H3K36me3, are recognized by RBBP4 together with SUZ12 to inhibit PRC2 catalytically and reduce its affinity to nucleosomes[6,7]. Collectively, these activities allow PRC2 to functionally distinguish repressed from transcriptionally active chromatin.

---

[4]Correspondence to: thomas.cech@colorado.edu.

Note: Supplementary information is available on the Nature Structural & Molecular Biology website.

AUTHOR CONTRIBUTIONS

C.D. and T.R.C. designed the biochemical experiments, which were carried out by C.D., L.Z. and K.J.G. C.D. carried out the bioinformatics analysis. C.D. and T.R.C. wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Although PRC2 is typically shown to be associated with repressed chromatin and the H3K27me3 mark[8–10], an additional overlap between some PRC2 target genes and the H3K4me3 mark is observed in embryonic[11,12] and T-cells[13], forming bivalent chromatin suggested to be poised for rapid activation. Other recent evidence also tracks PRC2 to transcriptionally active genes[14,15] and suggests association with RNA polymerase II[16] and with Phf1 protein, which mediates recruitment of PRC2 to the H3K36me3 mark[17–19]. These mechanisms are neither fully understood nor unified.

In *Drosophila* PRC2 is recruited to chromatin through Polycomb Response Elements (PREs, reviewed in ref[20]). Despite the discovery of functionally similar elements in vertebrates[21–23], the understanding of PRC2-specific recruiters is far from complete. Evidence indicates that lncRNAs recruit PRC2 to loci designated for silencing. HOTAIR lncRNA recruits PRC2 *in trans* to the HOXD locus[24,25] and other loci[26]. RepA lncRNA recruits PRC2 during X-chromosome inactivation[27]. Moreover, PRC2 is associated with thousands of RNAs in various cell lines[25,27–31]. A two-hairpin motif has been suggested to be enriched in a subclass of ncRNA that associates with PRC2[28], inspired by a two-hairpin motif that was originally discovered in RepA RNA[27]. More complex RNA structures have also been proposed[32,33]. Yet, the lack of quantitative data on the affinity of PRC2 for its RNA binding partners has limited the understanding of binding specificity.

To measure the binding specificity of PRC2, we performed quantitative electrophoretic mobility shift assays (EMSAs) of reconstituted human PRC2 with various RNAs. We show that PRC2 binds RNA promiscuously *in vitro*, with sub-micromolar affinity. RNA immunoprecipitation coupled with next generation sequencing (RIP-seq), whole transcriptome shotgun sequencing (RNA-seq), Global Run-On Sequencing (GRO-seq) and chromatin immunoprecipitation combined with massively parallel DNA sequencing (ChIP-seq) datasets generated from various mouse cell lines published in multiple independent studies were subjected to comparative analysis, showing that PRC2 binds RNA promiscuously also *in vivo*. Furthermore, unlike the H3K27me3 mark, which exclusively associates with repressed chromatin, PRC2 associates with both repressed and active genes. Taken together, these results suggest a model for the maintenance of the repressed chromatin state by PRC2, directed by promiscuous binding to nascent RNA transcripts.

RESULTS

Promiscuous RNA binding by PRC2 *in vitro*

To quantitatively characterize the affinity of PRC2 to RNA *in vitro*, we expressed and purified nucleic-acid free human PRC2 with either four subunits (EZH2-SUZ12-EED-RBBP4, namely PRC2 4m complex) or five subunits (EZH2-SUZ12-EED-RBBP4-AEBP2, namely PRC2 5m) using the baculovirus system (Fig. 1a). Recombinant PRC2 was efficient in histone methyltransferase activity (Supplementary Fig. 1) and RNA binding; the apparent dissociation constants (K_d) for *in vitro* transcribed RNA comprising 400 bases from the 5' end of HOTAIR lncRNA (HOTAIR 400) were 136 ± 22 nM and 255 ± 3 nM for PRC2 5m and PRC2 4m, respectively (Fig 1b,c and Supplementary Fig. 2 for RNA construct design). This two-fold difference in apparent K_d in the presence or absence of AEBP2 may reflect small differences between protein preparations. Data were fit to binding curves with Hill

coefficients of 2.3 ± 0.2 for PRC2 5m and 1.7 ± 0.1 for PRC2 4m, suggesting positive binding cooperativity. The affinity of PRC2 4m was similar for HOTAIR 400 and HOTAIR 1–300, an RNA construct comprising 300 bases from the 5' domain of HOTAIR that was previously shown to bind PRC2[24,34] (Fig 2a). Increased incubation time did not reduce the apparent K_d (Supplementary Fig. 3), supporting binding equilibrium. In agreement with earlier studies[27,29], PRC2 4m demonstrated similar affinity to the sense and antisense strands of the complete A-region sequence found within the human RepA transcript (Supplementary Fig. 4a).

While previous studies provided valuable qualitative information regarding the association of PRC2 and its subunits with different RNAs[24,25,27–29,34], the complete binding curves obtained here provide the first measurements of sub-micromolar affinity.

Given the similar binding affinities of PRC2 to previously characterized RNA partners, we tested RNAs that would not be expected to bind. The *E. coli* maltose binding protein mRNA originates in an organism lacking polycomb group proteins. Unexpectedly, the first 300 bases of this mRNA (MBP 1–300) bound PRC2 4m with an apparent K_d of 110 ± 10 nM, comparable to that observed for HOTAIR RNA (Fig 2a). MBP 1–300 is not predicted to include the two-hairpin motif (Supplementary Fig. 5) previously observed in some PRC2-associated RNAs[27–29]. Similar affinities were also observed for the antisense (as) strand of HOTAIR 1–300 (Fig. 2a and Supplementary Fig. 4b), *Mus musculus* telomerase RNA and the P4–P6 domain of the group I intron from the ciliate *Tetrahymena* (Supplementary Fig. 4a). Excess MBP 1–300 competed HOTAIR 400 from PRC2 (Fig. 2b), suggesting that both RNAs interact with the same binding site on PRC2.

Collectively, these data indicate promiscuous RNA binding by PRC2. On the other hand, the affinity of PRC2 to RNA is quite high and, remarkably, is higher than the affinities of its subunit EED for repressive-mark histone-tail peptides H3K27me3, H3K9me3 and H1K26me3 (K_d > 20 μM[3,4]) or of the PRC2 recruiting factor Phf1 for H3K36me3 tails (K_d > 20 μM[17,18] or K_d > 180 μM[19] by different approaches), suggesting biological significance for promiscuous RNA binding by PRC2.

RNA length dependence of PRC2 binding

To further test the observation that PRC2 binds RNA promiscuously, rather than interacting with an unrecognized encrypted binding motif, we generated RNAs comprising 10, 20, 50, 100, 200, 300 and 800 bases from the 5' end of *E. coli* MBP mRNA. A complete binding curve was recorded for each RNA (Fig. 3a,b). The K_d decreased with increasing RNA length up to a size of around 300 bases, where the additional increment to 800 bases resulted in no further increase in affinity. Notably, plotting log(K_d) versus log(RNA length) revealed a linear dependence with slope = −1.04 (Fig. 3c). This linear dependency between the length of the RNA and 1/K_d, namely the association constant (K_a), is expected for a protein binding to an array of multiple binding sites[35–37]. On the contrary, if an encrypted binding motif were present in MBP 1–300, one would expect a step function or a sharp increase in K_a with the first RNA that included the motif. Finally, PRC2 binding exhibited increasing cooperativity for RNAs of 200–800 bases (Fig. 3d).

One plausible hypothesis to explain promiscuous PRC2-RNA binding is that PRC2 interacts electrostatically with the RNA phosphate backbone. To test the dependency of K_d on salt concentration, we performed EMSA experiments of HOTAIR 400 and PRC2 5m at various KCl concentrations (Fig. 4a,b). A linear dependence was observed between log(K_d) and log(KCl) with a slope of 0.7 (Fig. 4c), suggesting that at most one salt bridge is mediating binding[38]. This implies that although promiscuous, the interactions between PRC2 and RNA are not primarily electrostatic. Base stacking with aromatic amino acid side chains is an example of such an interaction that can occur in a sequence-independent manner, as observed in the exosome[39].

Widespread RNA binding by PRC2 *in vivo*

While our results clearly indicate promiscuous RNA binding by PRC2 *in vitro*, factors such as posttranslational modifications and presence of other RNA-binding proteins could modify this intrinsic property of PRC2 *in vivo*. If the complex bound RNA promiscuously also *in vivo*, we expected that PRC2 would track with RNA from transcriptionally active chromatin genome-wide. To test this hypothesis we analyzed publically available Ezh2 RIP-seq datasets generated using mouse embryonic stem cells (ESC), either wild type (wt) or an *Ezh2* knockout (*Ezh2*$^{(-/-)}$) control. Ezh2 fold enrichment (Ezh2-FE) was defined for each transcript separately as the degree of over-representation of a given RNA transcript in the wt dataset relative to the *Ezh2*$^{(-/-)}$ dataset. These data were originally used to identify the pool of transcripts associated with PRC2, namely the PRC2 transcriptome[29]. We used the same criteria: a given RNA was included in the PRC2 transcriptome only if observed with Ezh2-FE ≥ 3 and expression ≥ 0.4 $RPKM_e$ (Reads Per Kilobase of exon per Million fragments mapped[40]). In agreement with previous results[29], we found that these criteria identified key RNAs previously shown to associate with PRC2, including Xist, Kcnq1o1 and Gtl2 (also known as Meg3, Fig. 5a). As previously demonstrated[29], the Hotair transcript is lowly expressed in this cell line and therefore did not meet these criteria and was excluded from the PRC2 transcriptome, thereby serving as a negative control for our analysis. A positive internal control is Ezh2, which is under-represented in the *Ezh2*$^{(-/-)}$ dataset and therefore observed with Ezh2-FE of 20. Importantly, the pool of genes that meet the criteria defining the PRC2 transcriptome is rich with highly expressed genes. These include mRNAs for nuclear-coded mitochondrial ATP synthase F1 alpha subunit 1, alpha-actin, gamma-actin and three PRC2 subunits: Suz12, Rbbp7 and Aebp2 (Fig. 5a). Accordingly, Gene Ontology (GO) analysis of genes with Ezh2-FE > 3 and high wt Ezh2 RIP-seq coverage (> 10 $RPKM_e$, Supplementary Table 1) identified multiple GO terms, where the highly significant terms were related to metabolic processes (Supplementary Table 2).

If most of the transcripts associated with PRC2 were specifically recruiting PRC2 for repression, one would expect genome-wide positive correlation between ChIP-seq H3K27me3 mark and Ezh2-FE and negative correlation between Ezh2-FE and RNA expression level or the H3K36me3 mark for active chromatin. We tested these assumptions by comparative analysis using multiple independent published datasets (Fig. 5b and Supplementary Table 3). Short reads resulting from 35 ChIP-seq, RNA-seq and GRO-seq experiments performed in 8 independent studies and in multiple mouse cell lines were aligned to the mouse genome. To reduce bias, multiple independent approaches for data

acquisition and normalization were performed, resulting in 4 or 9 different data vectors for each ChIP-seq or RNA-seq experiment, respectively; these summed to a total of 180 data vectors, each describing 27,874 autosomal non-redundant mouse Refseq genes (See Methods section for detailed description).

Next, each of these 180 data vectors was correlated with the Ezh2-FE dataset, and each pair-wise comparison is represented by one bar in Fig. 5b. Strikingly, RNA-seq, GRO-seq and ChIP-seq for active chromatin marks H3K36me3, H3K79me2/3 and RNA polymerase II phosphorylated at Serine 2 (Pol II Ser2) were positively correlated with Ezh2-FE (blue bars are concentrated on the left part of the top graph of Fig. 5b), and repressed marks H3K27me3 and H3K9me3 were negatively correlated (red bars are concentrated on the right). This trend was consistent for almost all datasets that were subjected to this analysis, which spanned multiple cell lines and data-processing approaches. Interestingly, ChIP-seq datasets for SUZ12 and EZH2 did not follow this trend, and demonstrated either a positive or a negative correlation in different datasets. Classification of datasets based on cell lines showed no bias (Fig. 5b, bottom bar graph). These results suggest that the broad association of PRC2 with transcripts from active genes occurs in a cell-line independent manner. These data cannot exclude regulation *in trans* or specific recruitment of PRC2 by a subset of transcripts with greater affinity and specificity. Yet, the positive correlation observed between Ezh2-FE and active genes, in a genome-wide context and cell-line independent manner, implies that PRC2 predominantly binds RNA promiscuously *in vivo*.

PRC2 found at both repressed and active genes *in vivo*

ChIP-seq provides information complementary to that provided by RIP-seq. ChIP-seq signals could represent recruitment of PRC2 by RNA binding as well as other interactions, while RIP-seq signals represent RNA-protein binding that may or may not occur when the RNP is associated with chromatin.

We analyzed ChIP-seq datasets for EZH2, H3K27me3, H3K4me3 and H3K36me3 marks published for five different human cell lines (Fig. 6a,b). We defined genes as associated with a given mark only if a ChIP-seq peak was called in proximity to the corresponding TSS. In agreement with previous studies[8–10,41], our analysis revealed 1328 to 3443 genes that associate with EZH2 in different cell lines. Most of these EZH2-associated genes also included the H3K27me3 mark (Fig. 6a), either in the absence or presence of H3K4me3 marks (i.e. bivalent chromatin). Yet, in most cell lines we could also detect a substantial amount, up to 30%, of EZH2-associated genes that were devoid of the H3K27me3 mark and were associated with an H3K4me3 or H3K36me3 mark. Over 90% of EZH2-associated genes were also associated with other marks, either repressed or active. In contrast, only a small fraction of H3K27me3-associated genes were associated also with the H3K36me3 mark (Fig. 6b). These data suggest that although the dominant pattern *in vivo* is PRC2 residing at repressed chromatin domains, PRC2 is also present at transcriptionally engaged or active genes to a significant extent.

Analysis of mouse embryonic cell data provided additional support for the conclusion that EZH2 occupancy, unlike the H3K27me3 mark, is not restricted to inactive chromatin. When TSSs of the entire mouse genome were sorted by the occupancy of the repressed mark

H3K27me3, active genes (identified by a high density of reads for H3K4me3, H3K36me3 and RNA-seq) were ranked low in the heatmap, as expected (Fig. 6c, in blue). Accordingly, enrichment profiles generated for the top 20% ranked genes, resulting from this sorting, were observed to have a prominent camelback profile for H3K27me3 accumulated reads, with two broad peaks spanning the TSS position (Fig. 6c, bottom left enrichment profile, blue line). The same set of genes had low representation of active chromatin marks (Fig. 6c, blue line within enrichment plots labeled H3K4me3, H3K36me3 and RNA-seq). When mouse genes were sorted based on EZH2 occupancy, repressed genes were still ranked at the very top of the heatmap, but now active genes were ranked much higher (Fig. 6c, green heatmaps and green line within enrichment profiles) relative to sorting on the basis of H3K27me3. This analysis suggests distinct profiles for EZH2 and H3K27me3, where H3K27me3 is associated with repressed chromatin and EZH2 is predominantly associated with H3K27me3 but tracks also with active genes.

To emphasize this observation, genes were further sorted based on differential ChIP-seq coverage, obtained by subtracting the number of normalized H3K27me3 reads from the number of normalized EZH2 reads within the same chromosomal location (EZH2 – H3K27me3, Fig. 6c, red heatmap). This operation aimed to identify genes where EZH2 was present in the absence of a significant H3K27me3 mark. Indeed, the top 20% of genes (based on EZH2 – H3K27me3 differential coverage) were observed to have a prominent accumulation of the active histone mark over their TSSs as well as RNA production (Fig. 6c, red line within enrichment plots).

EZH2-associated active genes are not PRC2-regulated

We hypothesized that the association of PRC2 with transcriptionally active genes would not necessarily force repression, as transcriptionally active chromatin marks are known to inhibit PRC2 catalytic activity[6,7]. As a test, we knocked down the PRC2 essential subunit SUZ12[2] in HEK293T/17 cells (Fig. 7a). Of the 736 genes whose expression changed significantly (false discovery rate < 0.05) after SUZ12 knockdown, 338 were up-regulated (Fig. 7b and Supplementary Table 4). Gene Ontology (GO) analysis showed enrichment for regulation and development terms for up-regulated genes (Supplementary Table 5), with a notable specialization in neuron development (Supplementary Fig. 6). This finding correlates with the proposed neuronal origin of HEK293 cells[42].

To identify genes that are physically associated with PRC2, we performed ChIP-seq for EZH2. Genes that were engaged in transcription were identified by ChIP-seq for H3K4me2/3 and ChIP-seq for RNA polymerase II phosphorylated at serine 5 (Pol II Ser5, performed using the same cell line within the scope of another study[43]). 2355 genes were associated with EZH2 (Fig. 7b). A large overlap was observed between EZH2-associated genes and H3K4me2/3 and Pol II Ser5 (hypergeometric p-value 5.43e–164 and 9.87e–183, respectively), in accord with previous reports[11–13,28]. Yet, only 51 of the 2355 genes that were physically associated with EZH2 responded to SUZ12 knockdown (Fig. 7b), a statistically insignificant overlap (hypergeometric p-value 0.813). Even if our ChIP-seq analysis is not sensitive enough to detect all EZH2-associated genes within repressed

chromatin, we can conclude that the majority of EZH2-associated genes that were engaged in transcription were not regulated by PRC2 (Fig. 7c).

DISCUSSION

Our *in vitro* PRC2-RNA binding assays provide the first quantitative measurements of sub-micromolar affinity of PRC2 to RNA; such affinity is likely to be biologically meaningful. Yet, our data also show that a fundamental property of PRC2 is to bind RNA promiscuously. Although post-translational modification or partner proteins may modulate this fundamental activity of PRC2 *in vivo*, we find that it correlates with results obtained by Ezh2 RIP-seq experiments *in vivo*. The PRC2 transcriptome includes thousands of transcripts[29], among which one can pinpoint selected RNAs that were previously reported to recruit PRC2. Yet our comparative analysis suggests that Ezh2-FE correlates positively with active genes and negatively with repressed genes, in a cell line independent manner. Importantly, this does not rule out other PRC2 recruitment models that were previously suggested but implies that, in parallel to them, a strong driving force directs PRC2 to transcripts of a wide variety of genes, both active and insufficiently repressed.

Our ChIP-seq analysis is in good agreement with previous studies[8–10,44] and shows that most of gene promoters that associate with EZH2 are also associated with H3K27me3 (Fig. 6). The differences between association patterns of PRC2 with RNA (RIP-seq) and chromatin (ChIP-seq) suggest that the majority of PRC2 that associates with RNA transcripts is never deposited to promoter regions of active genes. Alternatively, PRC2 may not stay there long enough to be cross-linked in a ChIP-seq experiment. There was a small, but yet notable, fraction of genes that tracked with EZH2 ChIP-seq peaks at promoter regions together with H3K36me3 or H3K4me3 marks and in the absence of H3K27me3; this suggests that such deposition of PRC2 to TSS of active genes does occur to some degree. The lack of H3K27me3 marks in the vicinity of these TSSs indicates that PRC2 is not actively repressing these genes.

PRC2 has the remarkable ability to recognize and be regulated by chromatin compaction[5] and the epigenetic marks H3K27me3[3,4], H3K4me3 and H3K36me3[6,7]. This allows the complex to functionally discriminate silent chromatin domains, or those that are in the process of being repressed, from transcriptionally active chromatin. Here we suggest an additional layer to this model, in which promiscuous RNA binding by PRC2 allows it to recognize transcriptionally active chromatin domains and scan them for genuine target genes that are already decorated with some H3K27me3 mark, but are not fully silenced. As both the affinity and cooperativity of promiscuous RNA binding by PRC2 increases with the RNA length (Fig. 3c,d), we speculate that PRC2 has evolved to track nascent RNA transcripts in genes that escaped from repression. Once PRC2 recognizes the H3K27me3 mark within a transcriptionally active gene, it will bind nucleosomes and be catalytically stimulated[3,4], restoring repression. Contrarily, if PRC2 recognizes active marks H3K4me3 and H3K36me3, its histone methyltransferase activity will be inhibited and it will not be deposited from the RNA to nucleosomes[6,7]. Further studies are required to identify and distinguish among different recruitment modes of PRC2 to chromatin for transcriptional reprogramming and repression maintenance.

While promiscuous RNA binding by PRC2 can facilitate scanning of chromatin, in highly expressed genes the RNA could additionally serve as a decoy, or natural sink, that can strip PRC2 away from chromatin as transcripts roll off of the DNA template. Similarly, exclusion of PRC2 from chromatin by Tsix lncRNA[27] and removal of CTCF from chromatin by Jpx lncRNA[45] have been suggested. The observations that Ezh2-FE is positively correlated with active genes and negatively correlated with repressed genes (Fig. 5b) together with the fact that EZH2 ChIP-seq tags are devoid of most highly expressed genes (Fig. 6a,c) fit with this idea. In the absence of previously deposited H3K27me3 mark, PRC2 is unlikely to be transferred from nascent transcripts to chromatin, leaving these genes active. Therefore, most highly expressed genes whose transcripts interact with PRC2, based on high Ezh2-FE (Fig. 5a), are unlikely to be repressed by the complex, as deposition to chromatin will never occur.

Thus, promiscuous RNA binding is likely to boost the process of polycomb-mediated repression and serve as a checkpoint to prevent escape from silencing. In this model, this simple yet robust process allows PRC2 to use the most direct outcome of transcription, namely RNA, to maintain the repressed chromatin state. Although RNA-bound PRC2 is poised for repression, the actual choice is directed by the local chromatin context. An analogy occurs with commercial organizations that use rough and general criteria, such as annual income[46], to target ‘junk mail’ advertising to mailboxes of people who could potentially, but not necessarily, be their customers. Whether to become functionally involved is the choice of the recipient. In this ‘Junk Mail Model’ (Fig. 8), PRC2 is the junk mail, chromosomal loci are mailboxes, transcriptional activity is the criterion for delivery, and the local chromatin context (active or repressive histone marks and chromatin compaction) dictates the response.

While earlier studies provided insights for target-specific recruitment of PRC2 for *de novo* repression by ncRNAs[24–29], our model proposes a new role for promiscuous RNA binding in maintaining the repressed chromatin state. Such functional duality has been well established for promiscuous enzymes that can process various substrates, some with increased affinity and catalytic efficiency (reviewed in ref[47]). Accordingly, previously described RNA motifs may still be preferred sites of PRC2 binding, and specific binding may be enhanced by post-transcriptional modifications of the RNA, by post-translational modifications of PRC2 subunits, as previously suggested[34], or by bridging proteins.

As was previously shown for PRC2 recruitment by short ncRNAs[28], association of PRC2 with chromatin is predominantly around TSSs, rather than gene bodies (Figure 6c). While earlier results explained the association of PRC2 with promoters of repressed genes, our data show that a substantial amount of promoters that associate with EZH2 are active (Figure 6). Based on the Junk Mail Model, the association of PRC2 with TSSs of active genes, rather than genes bodies, can be attributed to the relatively open chromatin in these regions or the exclusion of PRC2 from gene bodies by other factors (e.g., elongating polymerases, histone acetylation marks or splicing factors). Above all, the massive amount of elongating RNA can serve as decoy to strip PRC2 away from chromatin while transcription machinery progresses into the gene body. The relative contributions of these various processes should be addressed in future studies.

Although our data and analysis are restricted to PRC2, we suggest that the Junk Mail Model may potentially be generalized to other systems. Thus, promiscuous RNA binding may facilitate the recruitment of chromatin remodeling and other factors to transcriptionally active chromatin by direct binding to the most immediate product of the transcriptional machinery, RNA.

METHODS

Methods and associated references are available in the online version of the paper at http://www.nature.com/nsmb/.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

We thank R. Dowell (CU Boulder) for computational resources, J. Huntley (BioFrontiers Next-Gen Sequencing Facility, CU Boulder) for discussion and assistance with sequencing, R. Kingston (Harvard Medical School, Boston, Massachusetts, USA) for kindly providing plasmids with PRC2 subunit genes, and J. T. Lee (Harvard Medical School) and D. Reinberg (New York University) for discussions. T.R.C. is an investigator of the Howard Hughes Medical Institute. C.D. is supported by the Fulbright Postdoctoral Fellowship and the Machiah Foundation Program.

REFERENCES

1. Margueron R, Reinberg D. The Polycomb complex PRC2 and its mark in life. Nature. 2011; 469:343–349. [PubMed: 21248841]
2. Cao R, Zhang Y. SUZ12 is required for both the histone methyltransferase activity and the silencing function of the EED-EZH2 complex. Mol Cell. 2004; 15:57–67. [PubMed: 15225548]
3. Margueron R, et al. Role of the polycomb protein EED in the propagation of repressive histone marks. Nature. 2009; 461:762–767. [PubMed: 19767730]
4. Xu C, et al. Binding of different histone marks differentially regulates the activity and specificity of polycomb repressive complex 2 (PRC2). Proc Natl Acad Sci U S A. 2010; 107:19266–19271. [PubMed: 20974918]
5. Yuan W, et al. Dense chromatin activates Polycomb repressive complex 2 to regulate H3 lysine 27 methylation. Science. 2012; 337:971–975. [PubMed: 22923582]
6. Schmitges FW, et al. Histone methylation by PRC2 is inhibited by active chromatin marks. Mol Cell. 2011; 42:330–341. [PubMed: 21549310]
7. Yuan W, et al. H3K36 methylation antagonizes PRC2-mediated H3K27 methylation. J Biol Chem. 2011; 286:7983–7989. [PubMed: 21239496]
8. Boyer LA, et al. Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature. 2006; 441:349–353. [PubMed: 16625203]
9. Lee TI, et al. Control of developmental regulators by Polycomb in human embryonic stem cells. Cell. 2006; 125:301–313. [PubMed: 16630818]
10. Ram O, et al. Combinatorial patterning of chromatin regulators uncovered by genome-wide location analysis in human cells. Cell. 2011; 147:1628–1639. [PubMed: 22196736]
11. Bernstein BE, et al. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell. 2006; 125:315–326. [PubMed: 16630819]
12. Azuara V, et al. Chromatin signatures of pluripotent cell lines. Nat Cell Biol. 2006; 8:532–538. [PubMed: 16570078]

13. Roh TY, Cuddapah S, Cui K, Zhao K. The genomic landscape of histone modifications in human T cells. Proc Natl Acad Sci U S A. 2006; 103:15782–15787. [PubMed: 17043231]
14. Mousavi K, Zare H, Wang AH, Sartorelli V. Polycomb protein Ezh1 promotes RNA polymerase II elongation. Mol Cell. 2012; 45:255–262. [PubMed: 22196887]
15. Herz HM, et al. Polycomb repressive complex 2-dependent and -independent functions of Jarid2 in transcriptional regulation in Drosophila. Mol Cell Biol. 2012; 32:1683–1693. [PubMed: 22354997]
16. Brookes E, et al. Polycomb associates genome-wide with a specific RNA polymerase II variant, and regulates metabolic genes in ESCs. Cell Stem Cell. 2012; 10:157–170. [PubMed: 22305566]
17. Ballare C, et al. Phf19 links methylated Lys36 of histone H3 to regulation of Polycomb activity. Nat Struct Mol Biol. 2012
18. Musselman CA, et al. Molecular basis for H3K36me3 recognition by the Tudor domain of PHF1. Nat Struct Mol Biol. 2012
19. Brien GL, et al. Polycomb PHF19 binds H3K36me3 and recruits PRC2 and demethylase NO66 to embryonic stem cell genes during differentiation. Nat Struct Mol Biol. 2012
20. Schwartz YB, Pirrotta V. Polycomb silencing mechanisms and the management of genomic programmes. Nat Rev Genet. 2007; 8:9–22. [PubMed: 17173055]
21. Cuddapah S, et al. A novel human polycomb binding site acts as a functional polycomb response element in Drosophila. PLoS One. 2012; 7:e36365. [PubMed: 22570707]
22. Sing A, et al. A vertebrate Polycomb response element governs segmentation of the posterior hindbrain. Cell. 2009; 138:885–897. [PubMed: 19737517]
23. Woo CJ, Kharchenko PV, Daheron L, Park PJ, Kingston RE. A region of the human HOXD cluster that confers polycomb-group responsiveness. Cell. 2010; 140:99–110. [PubMed: 20085705]
24. Tsai MC, et al. Long noncoding RNA as modular scaffold of histone modification complexes. Science. 2010; 329:689–693. [PubMed: 20616235]
25. 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]
26. Chu C, Qu K, Zhong FL, Artandi SE, Chang HY. Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. Mol Cell. 2011; 44:667–678. [PubMed: 21963238]
27. 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]
28. 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]
29. Zhao J, et al. Genome-wide identification of polycomb-associated RNAs by RIP-seq. Mol Cell. 2010; 40:939–953. [PubMed: 21172659]
30. Khalil AM, et al. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. Proc Natl Acad Sci U S A. 2009; 106:11667–11672. [PubMed: 19571010]
31. Guttman M, et al. lincRNAs act in the circuitry controlling pluripotency and differentiation. Nature. 2011; 477:295–300. [PubMed: 21874018]
32. Maenner S, et al. 2-D structure of the A region of Xist RNA and its implication for PRC2 association. PLoS Biol. 2010; 8:e1000276. [PubMed: 20052282]
33. Duszczyk MM, Wutz A, Rybin V, Sattler M. The Xist RNA A-repeat comprises a novel AUCG tetraloop fold and a platform for multimerization. RNA. 2011; 17:1973–1982. [PubMed: 21947263]
34. Kaneko S, et al. Phosphorylation of the PRC2 component Ezh2 is cell cycle-regulated and up-regulates its binding to ncRNA. Genes Dev. 2010; 24:2615–2620. [PubMed: 21123648]
35. Kowalczykowski SC, et al. Cooperative and noncooperative binding of protein ligands to nucleic acid lattices: experimental approaches to the determination of thermodynamic parameters. Biochemistry. 1986; 25:1226–1240. [PubMed: 3486003]
36. Epstein IR. Kinetics of nucleic acid-large ligand interactions: exact Monte Carlo treatment and limiting cases of reversible binding. Biopolymers. 1979; 18:2037–2050. [PubMed: 497353]

37. Broderick JA, Salomon WE, Ryder SP, Aronin N, Zamore PD. Argonaute protein identity and pairing geometry determine cooperativity in mammalian RNA silencing. RNA. 2011; 17:1858–1869. [PubMed: 21878547]
38. Record MT Jr, Lohman ML, De Haseth P. Ion effects on ligand-nucleic acid interactions. J Mol Biol. 1976; 107:145–158. [PubMed: 1003464]
39. Makino DL, Baumgartner M, Conti E. Crystal structure of an RNA-bound 11-subunit eukaryotic exosome complex. Nature. 2013; 495:70–75. [PubMed: 23376952]
40. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods. 2008; 5:621–628. [PubMed: 18516045]
41. Marks H, et al. The transcriptional and epigenomic foundations of ground state pluripotency. Cell. 2012; 149:590–604. [PubMed: 22541430]
42. Shaw G, Morse S, Ararat M, Graham FL. Preferential transformation of human neuronal cells by human adenoviruses and the origin of HEK 293 cells. FASEB J. 2002; 16:869–871. [PubMed: 11967234]
43. Schwartz JC, et al. FUS binds the CTD of RNA polymerase II and regulates its phosphorylation at Ser2. Genes Dev. 2012; 26:2690–2695. [PubMed: 23249733]
44. Ku M, et al. Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. PLoS Genet. 2008; 4:e1000242. [PubMed: 18974828]
45. Sun S, et al. Jpx RNA Activates Xist by Evicting CTCF. Cell. 2013; 153:1537–1551. [PubMed: 23791181]
46. Mazzone J, Pickett J. The Household Diary Study, Mail Use & Attitudes in FY 2010. 2011
47. Khersonsky O, Tawfik DS. Enzyme promiscuity: a mechanistic and evolutionary perspective. Annu Rev Biochem. 2010; 79:471–505. [PubMed: 20235827]
48. Murphy FL, Wang YH, Griffith JD, Cech TR. Coaxially stacked RNA helices in the catalytic center of the Tetrahymena ribozyme. Science. 1994; 265:1709–1712. [PubMed: 8085157]
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. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010; 26:841–842. [PubMed: 20110278]
51. Li H, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009; 25:2078–2079. [PubMed: 19505943]
52. Robinson JT, et al. Integrative genomics viewer. Nat Biotechnol. 2011; 29:24–26. [PubMed: 21221095]
53. Zhang Y, et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 2008; 9:R137. [PubMed: 18798982]
54. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B-Methodological. 1995; 57:289–300.
55. Schwartz JC, et al. FUS binds the CTD of RNA polymerase II and regulates its phosphorylation at Ser2. Genes Dev. 2012:26.
56. Reimand J, Kull M, Peterson H, Hansen J, Vilo J. g:Profiler--a web-based toolset for functional profiling of gene lists from large-scale experiments. Nucleic Acids Res. 2007; 35:W193–W200. [PubMed: 17478515]
57. Gupta RA, et al. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature. 2010; 464:1071–1076. [PubMed: 20393566]
58. Consortium, E.P. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012; 489:57–74. [PubMed: 22955616]
59. Consortium, E.P. A user's guide to the encyclopedia of DNA elements (ENCODE). PLoS Biol. 2011; 9:e1001046. [PubMed: 21526222]

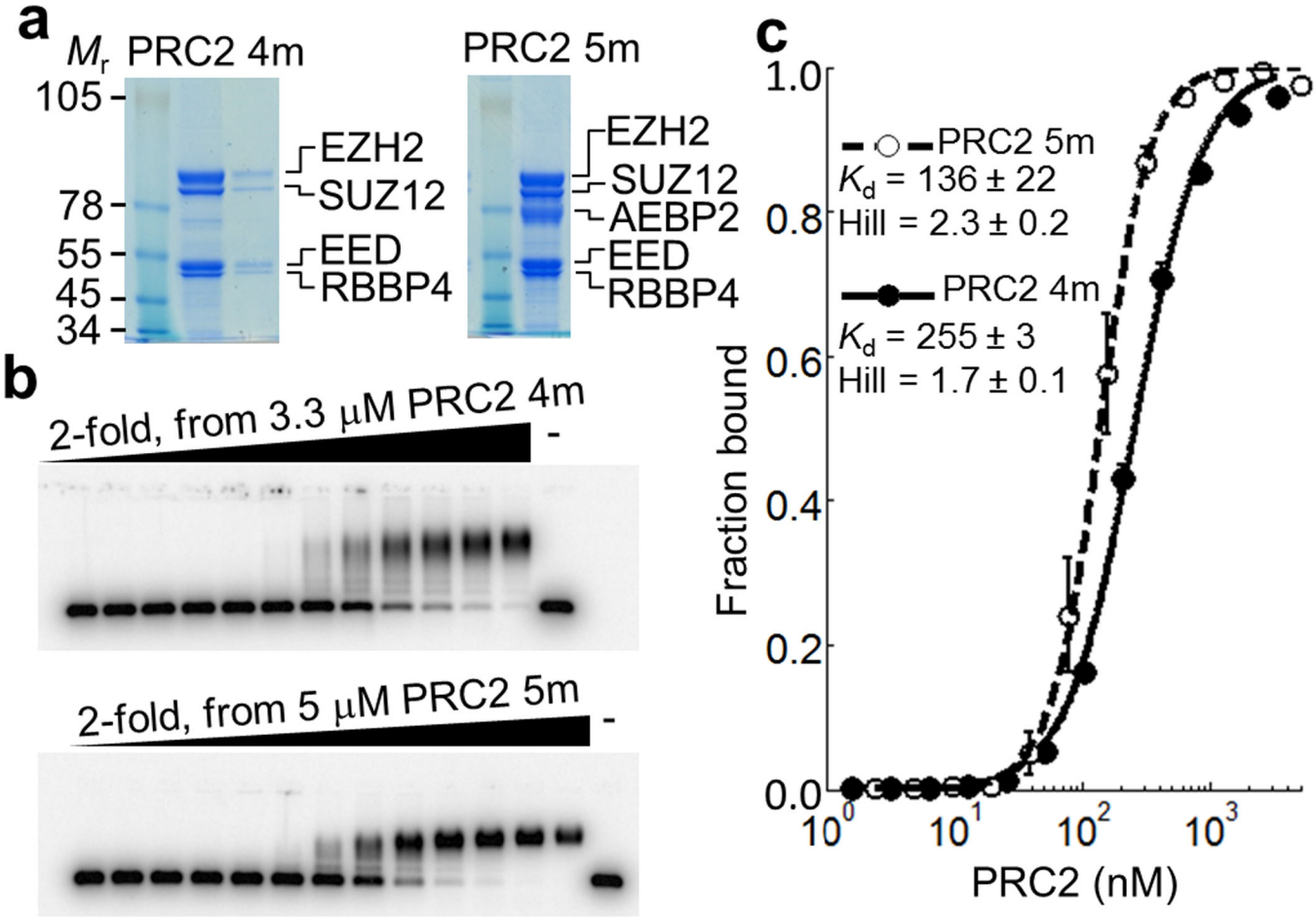

Figure 1.

PRC2 binds the 5' domain of HOTAIR RNA with sub-micromolar affinity, in the presence or absence of AEBP2. (**a**) Recombinant human PRC2 complexes reconstituted using the baculovirus system to include SUZ12, EZH2, EED and RBBP4 in the presence or absence of the zinc-finger subunit AEBP2 (PRC2 5m and 4m, respectively), analyzed by SDS-PAGE and stained with Coomassie blue. (**b**) Electrophoretic Mobility Shift Assays (EMSAs) performed using *in vitro* transcribed RNA including 400 nucleotides from the 5' of HOTAIR RNA (HOTAIR 400) in the presence of various concentrations of PRC2 4m and 5m. The two gels were run for different times, so the extent of the mobility shift upon protein binding is not meaningful. (**c**) Complete binding curves for HOTAIR 400 with PRC2 4m and PRC2 5m. Error bars for K_d, Hill and each data point represent range of two independent experiments performed on different days.

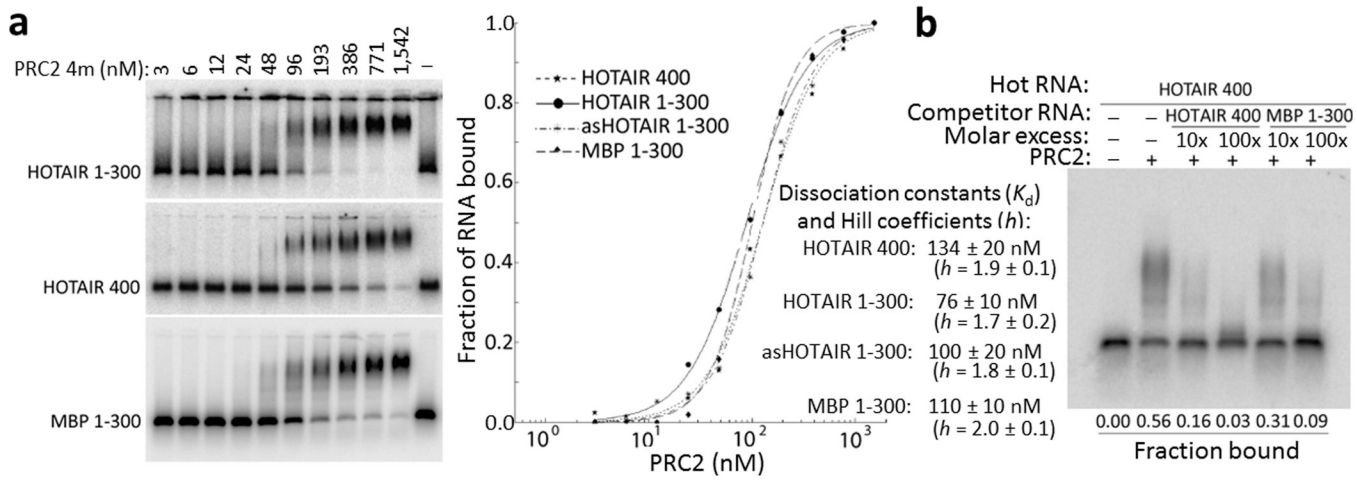

Figure 2.

Promiscuous binding of RNA by PRC2 *in vitro*. (**a**) Representative EMSAs and the corresponding binding curves show binding of PRC2 to HOTAIR 1–300, HOTAIR 400, antisense (as) HOTAIR 1–300 and *E. coli* MBP 1–300. K_d and Hill coefficients represented with data-range from two independent experiments performed on different days. (**b**) Binding competition experiment with unlabeled competitors MBP 1–300 and HOTAIR 400, radiolabeled HOTAIR 400 and PRC2 4m.

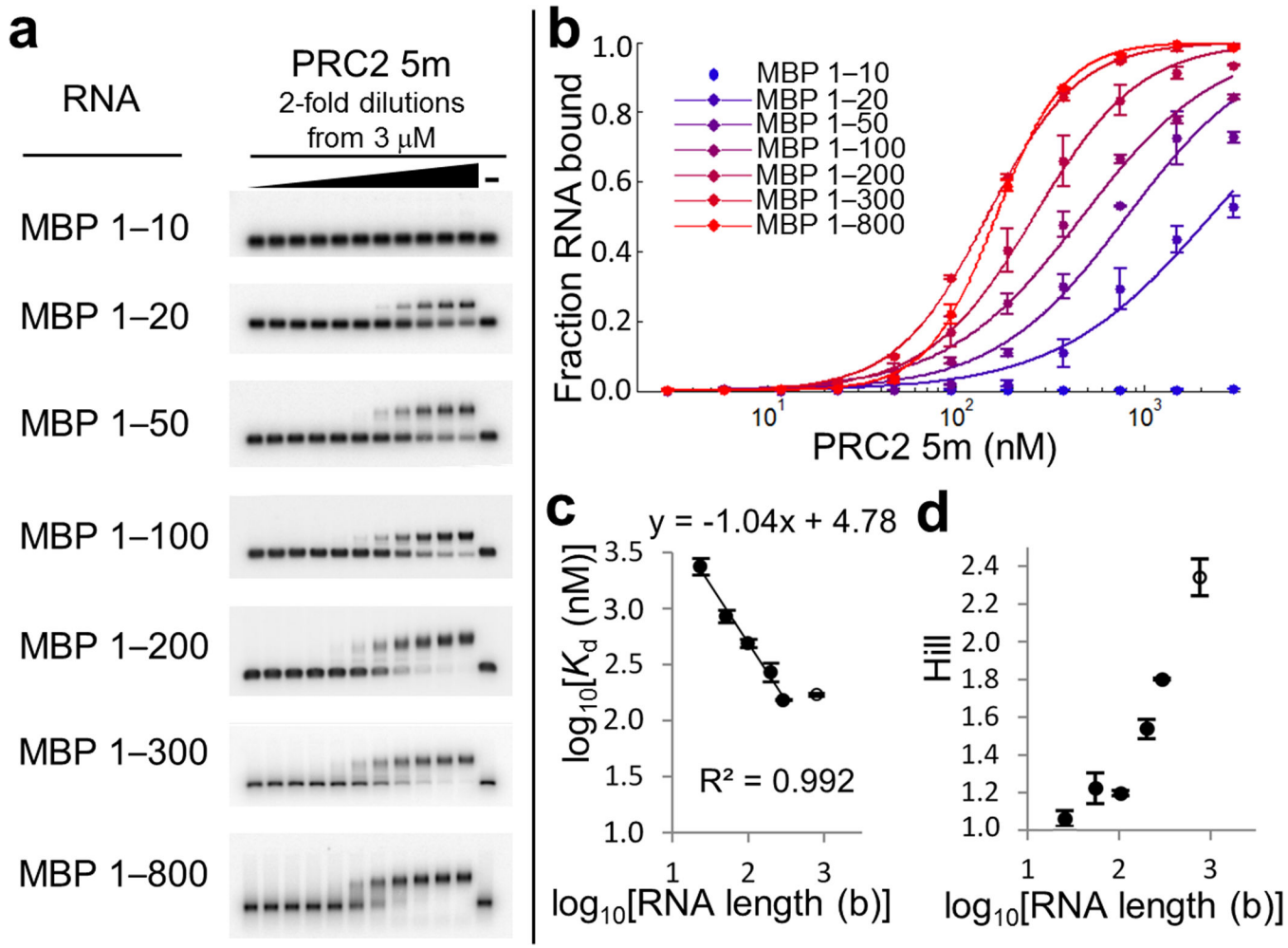

Figure 3.

Binding of PRC2 to RNA is size dependent. (**a**) Representative EMSAs of different *in vitro* transcribed RNAs that comprise 10, 20, 50, 100, 300 and 800 bases from the 5' end of *E. coli* MBP mRNA. The total size of each RNA includes additional five bases that were added through *in vitro* transcription. (**b**) Corresponding binding curves. (**c**) Linear correlation between $\log_{10}(K_d)$ and $\log_{10}$(RNA length (bases)). (**d**) Dependency between RNA length and binding cooperativity (Hill coefficient). Error bars in panels **b**, **c** and **d** represent range of data from two independent experiments performed on different days.

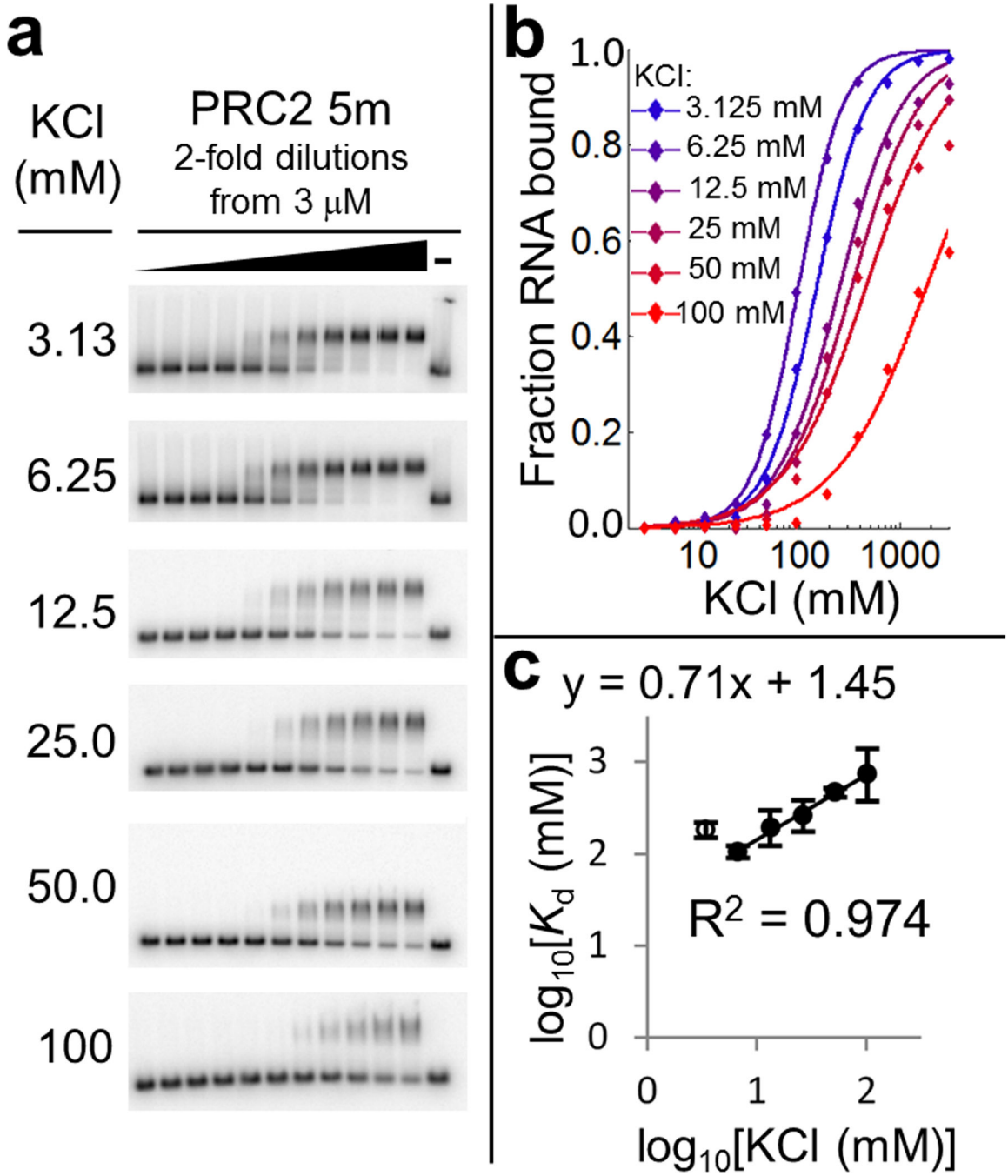

Figure 4.

Interaction between PRC2 and RNA shows little salt dependence. (**a**) Representative EMSAs of HOTAIR 400 in the presence PRC2 5m performed under different KCl concentrations. (**b**) Corresponding binding curves. (**c**) Linear correlation between $\log_{10}(K_d)$ and $\log_{10}$(KCl concentration). The apparent number of salt bridges between the RNA and the protein is calculated from the slope of the fitted line[38] and is less than one. Error bars, same as Fig. 3.

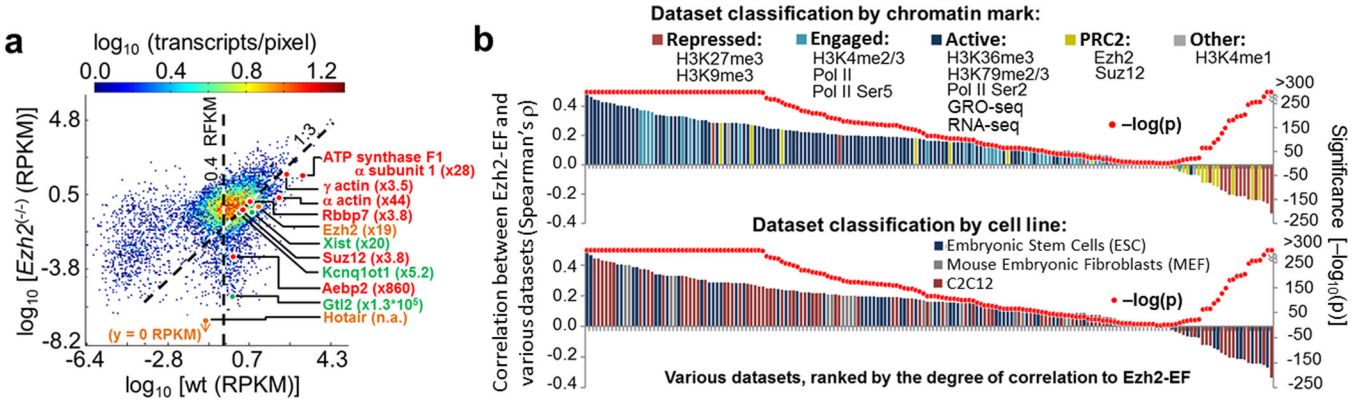

Figure 5.

Widespread binding of RNA by PRC2 *in vivo*. (**a**) Analysis of Ezh2 RIP-seq results from mouse ESC determines the PRC2 transcriptome, using the previous criteria[29]. Diagonal line represents Ezh2 RIP-seq Fold Enrichment (Ezh2-FE) of 3-fold for transcripts associated with Ezh2 in wt cells over *Ezh2* knockout cells (*Ezh2*$^{(-/-)}$). Vertical line represents a cutoff of 0.4 $RPKM_e$. Transcripts belonging to the PRC2 transcriptome lie under the diagonal line and to the right of the vertical line. Green, selected transcripts previously suggested as associated with PRC2. Orange, negative and positive internal controls, Hotair RNA and Ezh2 mRNA, respectively. Red, selected transcripts present in the PRC2 transcriptome that belong to either highly expressed genes or PRC2 subunit mRNAs. Parentheses, Ezh2-FE values. (**b**) High throughput sequencing data from 35 published ChIP-seq, RNA-seq and GRO-seq experiments were subjected to different types of data acquisition and analysis using identical pipelines. Analysis resulted in 180 data vectors, each describing any of 27,874 autosomal non-redundant mouse Refseq genes. Next, data in each of these 180 data vectors were correlated against Ezh2-FE values of the corresponding genes. Finally, all 180 data vectors were ranked based on the degree of correlation and plotted as bar graph. Red dots, correlation significance as -$\log_{10}$(p-value). Bar size, Spearman's rho correlation coefficient. Bar color, data classification. Top: bars were color-coded based on chromatin state recorded by the data. Bottom: the same bar graph as described above was color-coded based on different cell types. See Supplementary Table 6 for full description of each dataset.

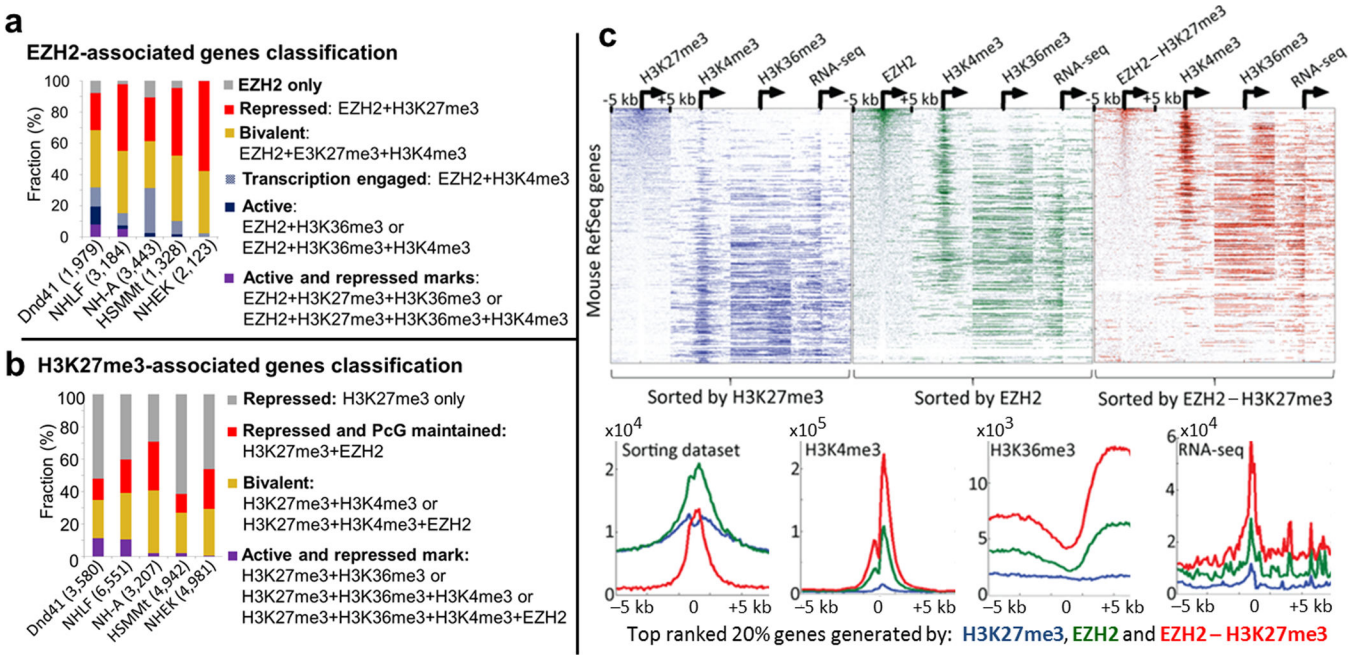

Figure 6.

PRC2 associates with active genes, in addition to its predominant association with repressed chromatin. (**a**) EZH2-associated genes were classified based on their association with other chromatin marks. Numbers in parentheses represent the number of EZH2-associated genes identified in each cell line. (**b**) H3K27me3-associated genes were classified based on their association with other chromatin marks. Numbers in parentheses represent the number of H3K27me3-associated genes identified in each cell line. (**c**) Heatmaps for chromatin marks H3K4me3 and H3K36me3 (ChIP-seq data) and for RNA-seq in mouse E14 cell line were generated using the same datasets, but presented using different types of sorting. Refseq mouse genes were sorted by three different criteria: reads (from 0.5 kb upstream of TSSs to 0.5 kb downstream) of either H3K27me3 (blue), EZH2 (red) or differential coverage obtained by subtracting the number of normalized H3K27me3 reads from the number of normalized EZH2 reads at the same position (EZH2 - H3K27me3, red). Below, top-ranked 20% of genes, resulting from these three different types of sorting, were used to generate enrichment profiles (same color key used). Titles of enrichment profiles represent the corresponding epigenetic mark. ‘Sorting dataset’ indicates the datasets that determined the three types of sorting, namely top 20% of genes ranked according to read occupancy around TSS of either H3K27me3 (blue), EZH2 (red) or EZH2 – H3K27me3 (red).

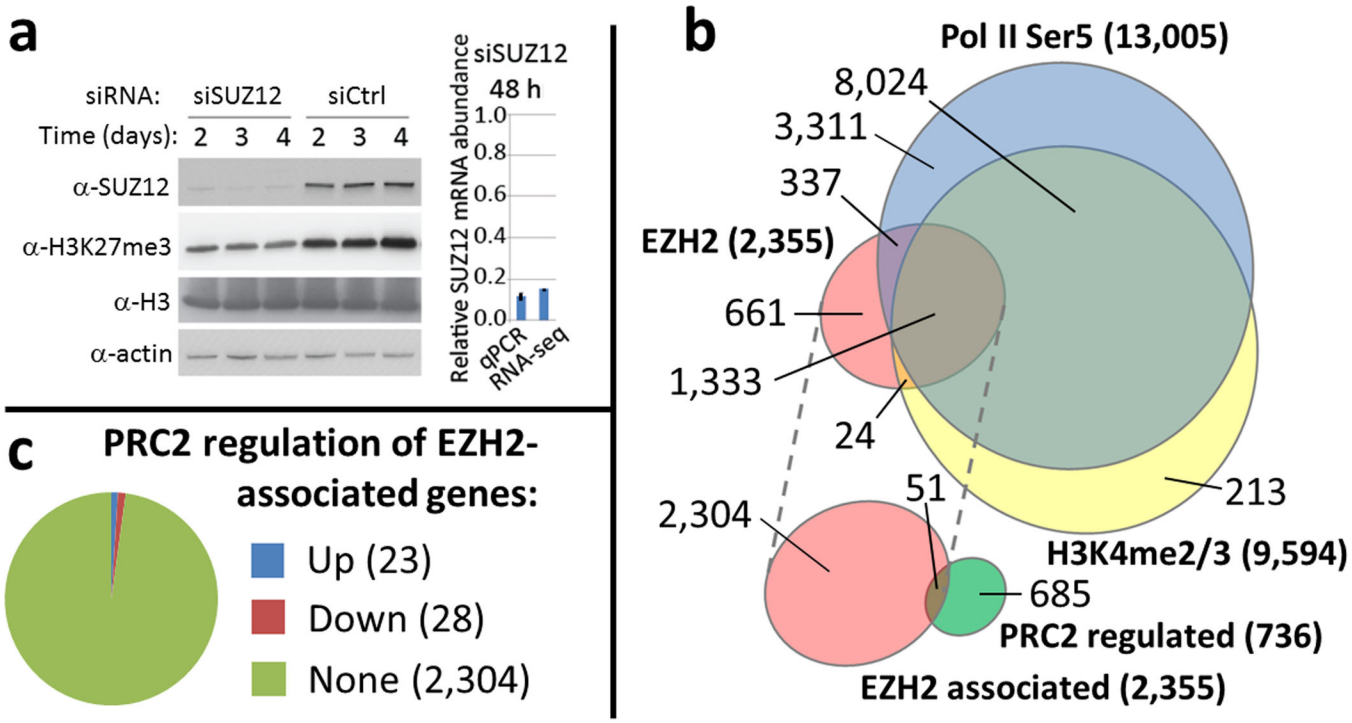

Figure 7.

Knockdown of SUZ12 in HEK293T/17 cells, combined with ChIP-seq and RNA-seq, confirms widespread association of EZH2 with H3K4me2/3 and Pol II Ser5, but expression of the vast majority of these genes does not respond to SUZ12 knockdown. (**a**) Knockdown of SUZ12 achieved after 48 hours, confirmed by qPCR and RNA-seq (right) and immunoblotting (left) and accompanied by global depletion of the H3K27me3 mark (left). (**b**) Top Euler diagram demonstrates large overlap between genes associated with EZH2, H3K4me2/3 and Pol II Ser5 in HEK293T/17 cells. Bottom Euler diagram demonstrates the small intersect between genes that responded to SUZ12 knockdown, identified based on two biological replicates, and genes associated with EZH2. (**c**) Only a small fraction of EZH2-associated genes is transcriptionally regulated by PRC2.

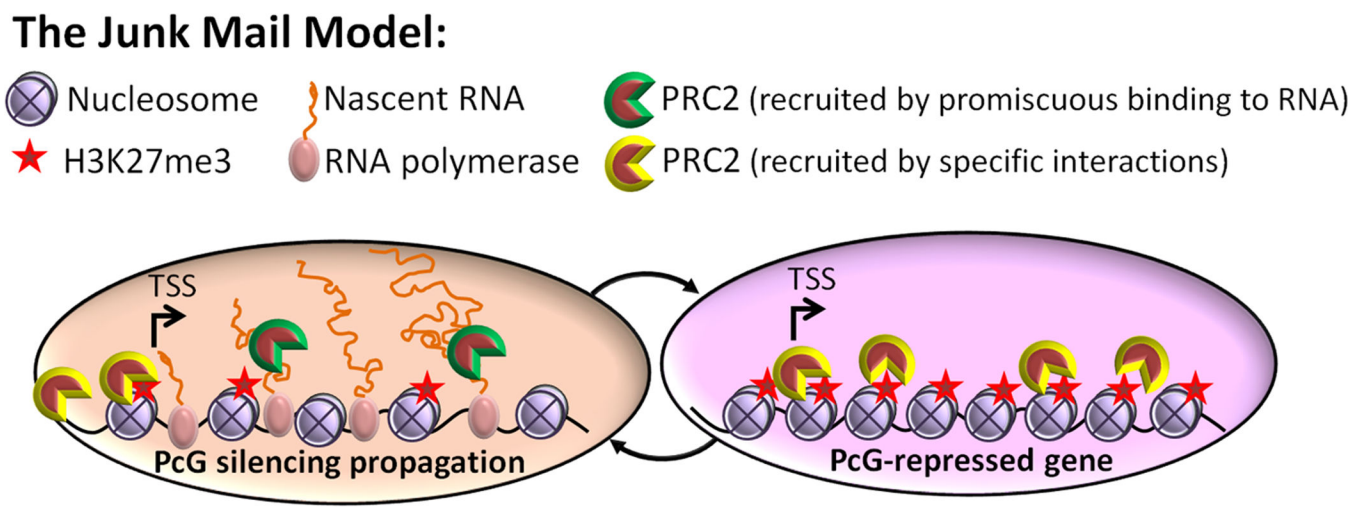

Figure 8.

The Junk Mail Model for repressed chromatin maintenance by PRC2 utilizing promiscuous RNA binding. Upon inappropriate transcription of a genuine polycomb target gene, recruitment of PRC2 through promiscuous RNA binding results in recognition of previously deposited H3K27me3 marks, stimulation of PRC2 histone methyltransferase activity and restoration of repression. Contrarily, if PRC2 binds to a nascent transcript of an active gene, in the absence of the H3K27me3 mark and in the presence of H3K36me3 and H3K4me3 marks, the histone methyltransferase activity of PRC2 is inhibited and its affinity to nucleosomes reduced, resulting in inefficient deposition to chromatin.