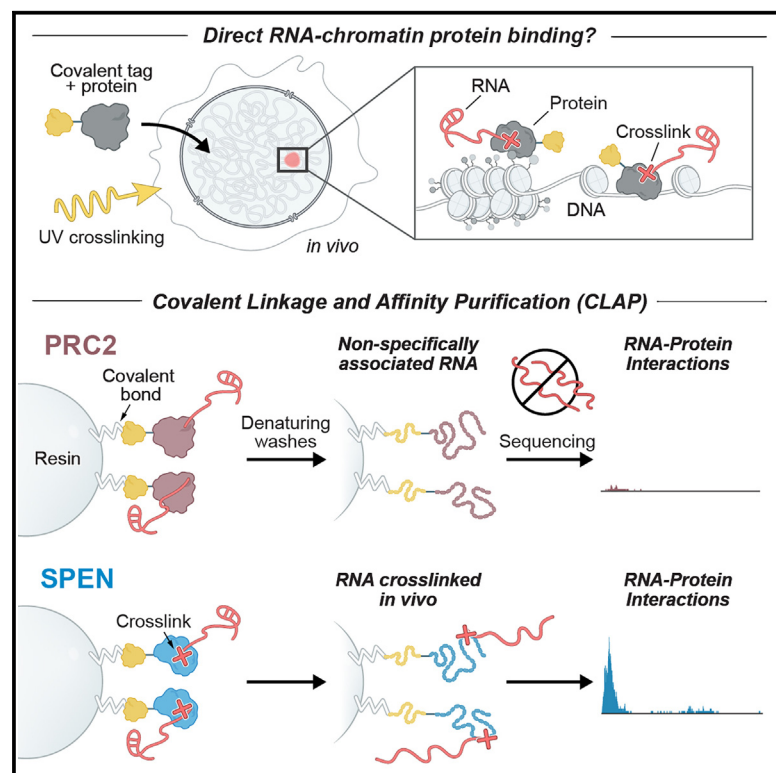


# Denaturing purifications demonstrate that PRC2 and other widely reported chromatin proteins do not appear to bind directly to RNA *in vivo*

## Graphical abstract



## Authors

Jimmy K. Guo, Mario R. Blanco, Ward G. Walkup IV, ..., Enrique Amaya, Vickie Trinh, Mitchell Guttman

## Correspondence

mblanco@incrna.caltech.edu (M.R.B.),  
mguttman@caltech.edu (M.G.)

## In brief

This work by Guo et al. challenges widespread reports that PRC2 and other chromatin proteins bind to RNA to control the regulation of gene expression. It highlights the discrepancy between biochemical and functional evidence and suggests that reports of a broad role for protein-RNA binding in controlling chromatin regulatory mechanisms should be re-evaluated.

## Highlights

- Many previously detected PRC2-RNA interactions do not occur *in vivo*
- Denaturing purifications eliminate PRC2-RNA interactions but retain known interactions
- Other reported chromatin regulators (CTCF and YY1) do not appear to bind RNA *in vivo*
- Broad role for RNA binding in chromatin regulation needs to be re-evaluated

Article

# Denaturing purifications demonstrate that PRC2 and other widely reported chromatin proteins do not appear to bind directly to RNA *in vivo*

Jimmy K. Guo,<sup>1,2,4</sup> Mario R. Blanco,<sup>1,4,\*</sup> Ward G. Walkup IV,<sup>1</sup> Grant Bonesteele,<sup>1</sup> Carl R. Urbinati,<sup>1,3</sup> Abhik K. Banerjee,<sup>1,2</sup> Amy Chow,<sup>1</sup> Olivia Ettlin,<sup>1</sup> Mackenzie Strehle,<sup>1</sup> Parham Peyda,<sup>1</sup> Enrique Amaya,<sup>1</sup> Vickie Trinh,<sup>1</sup> and Mitchell Guttman<sup>1,5,\*</sup>

<sup>1</sup>Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA 91125, USA

<sup>2</sup>Keck School of Medicine, University of Southern California, Los Angeles, CA 90089, USA

<sup>3</sup>Department of Biology, Loyola Marymount University, Los Angeles, CA 90045, USA

<sup>4</sup>These authors contributed equally

<sup>5</sup>Lead contact

\*Correspondence: [mblanco@incrna.caltech.edu](mailto:mblanco@incrna.caltech.edu) (M.R.B.), [mguttman@caltech.edu](mailto:mguttman@caltech.edu) (M.G.)

<https://doi.org/10.1016/j.molcel.2024.01.026>

## SUMMARY

Polycomb repressive complex 2 (PRC2) is reported to bind to many RNAs and has become a central player in reports of how long non-coding RNAs (lncRNAs) regulate gene expression. Yet, there is a growing discrepancy between the biochemical evidence supporting specific lncRNA-PRC2 interactions and functional evidence demonstrating that PRC2 is often dispensable for lncRNA function. Here, we revisit the evidence supporting RNA binding by PRC2 and show that many reported interactions may not occur *in vivo*. Using denaturing purification of *in vivo* crosslinked RNA-protein complexes in human and mouse cell lines, we observe a loss of detectable RNA binding to PRC2 and chromatin-associated proteins previously reported to bind RNA (CTCF, YY1, and others), despite accurately mapping bona fide RNA-binding sites across others (SPEN, TET2, and others). Taken together, these results argue for a critical re-evaluation of the broad role of RNA binding to orchestrate various chromatin regulatory mechanisms.

## INTRODUCTION

RNA-protein interactions are important for many aspects of RNA biogenesis, processing, and function. Recent efforts to catalog these interactions have led to the discovery of many novel RNA-binding proteins (RBPs) that do not contain canonical RNA-binding domains,<sup>1–4</sup> including chromatin and transcriptional regulators as well as metabolic proteins. This has led to intense interest in understanding the functional importance of non-canonical RNA-protein interactions. For example, RNA interactions have been proposed to be critical for the function of several chromatin proteins<sup>5–8</sup> and many chromatin regulators have been reported to act as central players in the mechanisms by which long non-coding RNAs (lncRNAs) regulate gene expression.<sup>9–17</sup>

One of the most widely studied chromatin complexes that is reported to interact with RNA is the polycomb repressive complex 2 (PRC2), which deposits the repressive trimethylation of lysine 27 on histone H3 (H3K27me3) modification.<sup>18,19</sup> PRC2 components have been reported to bind broadly to many RNAs, including lncRNAs and mRNAs.<sup>20–22</sup> These observations have led to the proposal that many canonical PRC2 functions are mediated by its interactions with RNA,<sup>5,20,23,24</sup> including PRC2 recruitment to genomic DNA sites,<sup>11,23,25–28</sup> tethering of PRC2 components,<sup>20</sup>

and activation of PRC2 enzymatic activity.<sup>29</sup> Additional proposals include that PRC2 binding to nascent pre-mRNAs precludes binding to DNA at active genes<sup>5,21,22</sup> and that PRC2 can act as a nuclease to degrade specific RNAs.<sup>30</sup> These observations have prompted studies of several other chromatin complexes and reports that they too bind broadly to many RNAs to achieve various functions; examples include DNA methylation enzymes (e.g., DNMT1),<sup>31–33</sup> PRC1 components (e.g., RING1 and CBX7),<sup>16,34</sup> trithorax components (e.g., WDR5),<sup>32,35</sup> transcription factors (e.g., SOX2),<sup>36–38</sup> and chromatin structure proteins (e.g., YY1 and CTCF).<sup>6–8,39–41</sup>

The paradigm example for the functional relevance of chromatin protein-RNA interactions is the Xist lncRNA.<sup>9,42</sup> Xist initiates X chromosome inactivation (XCI) by localizing across one of the two X chromosomes, recruiting numerous chromatin-modifying complexes (including PRC2 and its associated H3K27me3 mark<sup>43,44</sup>) and mediating chromosome-wide silencing.<sup>43,44</sup> PRC2 was reported to bind to the A-repeat of Xist,<sup>9,45</sup> a region that is required for Xist-mediated silencing.<sup>46</sup> This led to a model whereby Xist directly binds to PRC2 and recruits this repressive chromatin complex to the X to mediate silencing.<sup>9,42,47</sup> Other lncRNAs, such as HOTAIR, were similarly reported to directly bind PRC2 to silence transcription.<sup>10,48,49</sup>

However, genetic studies raised important questions about the functional relevance of many of these chromatin-RNA interactions. For example, deletion of PRC2 components that prevent its recruitment to, and deposition of H3K27me3 on, the X does not impact initiation of Xist-mediated silencing,<sup>50,51</sup> and deletion of the A-repeat from Xist does not prevent PRC2 recruitment to, or H3K7me3 accumulation on, the X.<sup>52–55</sup> Similarly, deletion of PRC2 does not impact HOTAIR-mediated silencing.<sup>52,56,57</sup> Importantly, this discrepancy is not limited to PRC2-RNA interactions: the YY1 transcriptional regulator was reported to bind to Xist to tether the RNA to chromatin,<sup>14</sup> yet neither deletion of the YY1 protein, nor deletion of the reported YY1-binding site from Xist (F-repeat) impacts the localization of Xist to chromatin or Xist-mediated silencing.<sup>46,55</sup> These examples highlight a critical discrepancy between the biochemical evidence supporting specific chromatin-RNA interactions and the genetic evidence demonstrating that these same interactions are often dispensable for lncRNA function.

Most biochemical evidence of RNA interactions with chromatin proteins comes from *in vitro* binding assays, which measure binding between purified proteins and RNA,<sup>5,58,59</sup> and RNA immunoprecipitation (RIP), which utilizes native purification conditions in crosslinked or non-crosslinked cells to immunoprecipitate a protein and measure its associated RNAs.<sup>60,61</sup> Because of the low stringency required to preserve native protein-RNA interactions, these methods can identify RNA-protein interactions that do not occur *in vivo*.<sup>62</sup>

More recently, several of the interactions detected by RIP<sup>5,15</sup> have been confirmed using crosslinking and immunoprecipitation (CLIP) methods.<sup>7,8,20,21</sup> CLIP utilizes UV-crosslinking to form covalent interactions in cells between directly interacting RNA and protein, followed by purification in stringent wash conditions (i.e., 1M salt), separation through a denaturing SDS-PAGE gel, transfer to a nitrocellulose membrane, and size extraction of the RNA-protein complex.<sup>63–67</sup> Because of its increased stringency, CLIP has emerged as the gold-standard for defining *in vivo* RNA-protein interactions and has been successfully used to define the precise RNA-binding sites of numerous RBPs.<sup>63</sup>

However, several observations suggest that many of the RNA-chromatin protein interactions reported, including by CLIP, may not represent interactions that occur *in vivo*: (1) we and others have purified Xist using denaturing conditions after *in vivo* crosslinking and failed to identify the previously reported interactions between Xist and any of the PRC2 components or YY1.<sup>55,68,69</sup> (2) PRC2 components have been shown to bind with measurable affinity to all RNAs *in vitro*, including bacterial RNAs that should not have endogenous affinity for mammalian proteins.<sup>5,58,59</sup> (3) visualization of labeled protein-RNA complexes separated on an SDS-PAGE gel after CLIP of PRC2 components fails to show the expected RNase-dependent size shift typically observed for RBPs.<sup>21,22</sup> (4) CLIP studies that have reported specific PRC2-RNA associations are often based on the analysis of low-complexity sequencing libraries, which may result in read redistribution at abundant RNAs and/or PCR duplicates of specific RNA fragments being mistaken for specific binding interactions.<sup>21,22,70</sup> (5) When CLIP is performed in the absence of *in vivo* crosslinking, there are still strong RNA associations observed for several chromatin proteins, such as CTCF.<sup>7</sup>

Based on these observations, we considered the possibility that many of the reported RNA-chromatin protein interactions may not occur *in vivo*. To explore this, we developed an experimental approach to unambiguously identify non-specific RNA associations that could not have occurred within the cell and applied a highly stringent, denaturing purification method to study chromatin proteins previously reported to bind RNA. Our results argue for a re-evaluation of the broad role of RNA binding in various chromatin regulatory mechanisms and provide a critical new framework for studying non-canonical RNA-protein interactions *in vivo*.

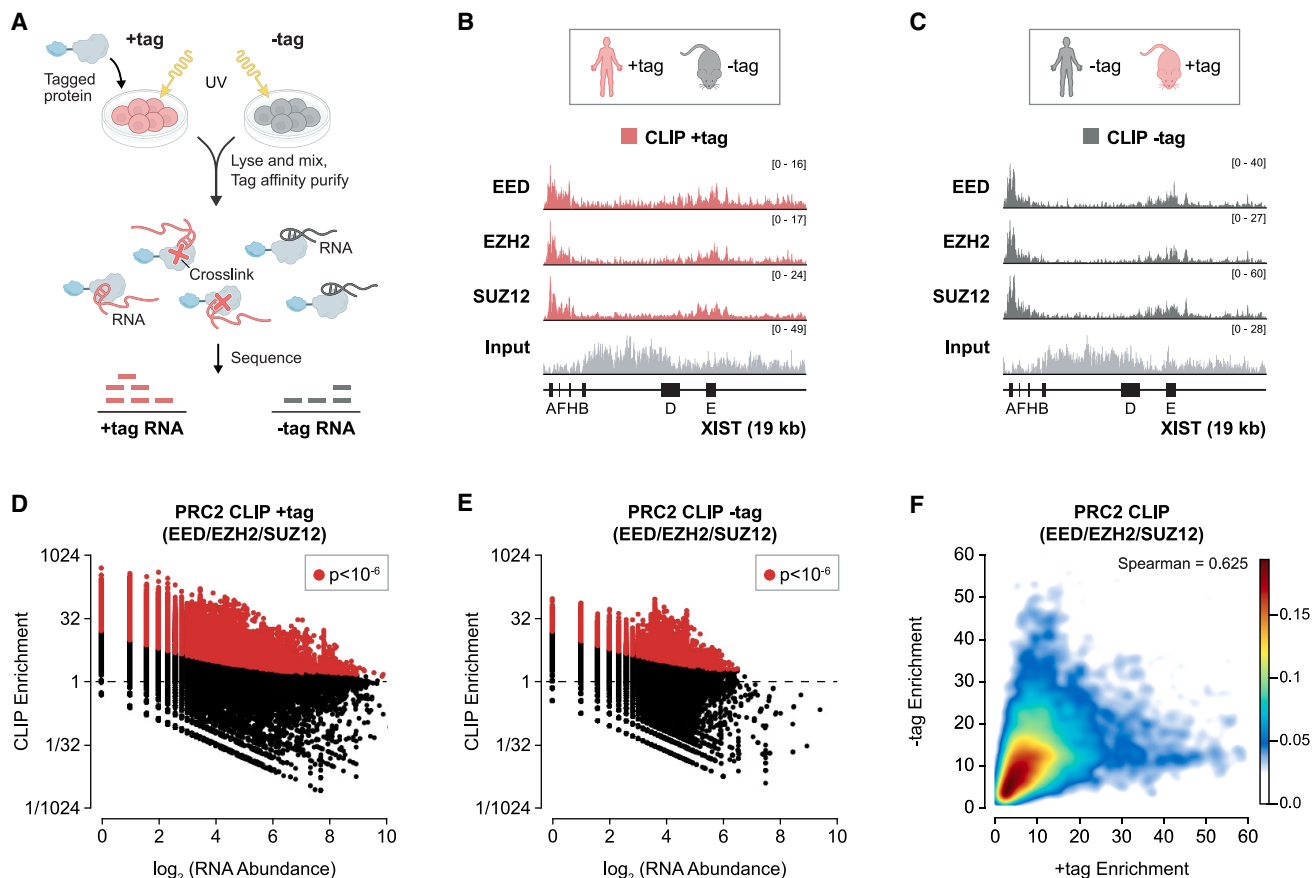
## RESULTS

### CLIP identifies many PRC2-RNA interactions that could not have occurred *in vivo*

Because PRC2 has been reported to bind directly and promiscuously to many RNAs, we considered the experimental and analytical challenges associated with distinguishing between promiscuous binding and lack of binding (see [STAR Methods, challenges associated with distinguishing between promiscuous binding and lack of binding](#)). To address this challenge, we designed an experiment—modeled after Mili and Steitz<sup>62</sup>—to unambiguously identify whether RNA detected from a RIP/CLIP experiment represents background that could not have occurred *in vivo*. Briefly, we generated V5-tagged versions of our proteins of interest, transfected them into cells and UV-crosslinked the cells to form covalent photo-crosslinks between directly interacting RNAs and proteins (+tag). We then mixed this lysate (+tag) with lysate from UV-crosslinked cells of a different species that do not express the V5-tagged proteins (–tag) and performed enhanced CLIP (eCLIP; a specific implementation of CLIP<sup>71</sup>) using a V5 antibody ([Figure 1A](#)). We focused on sequencing reads that can be mapped uniquely to one of the two species used. In this system, any sequencing reads that align to the genome of the species that did not contain the tagged protein (–tag RNA) must represent background interactions that could not occur *in vivo* because the affinity-purified V5-tagged protein was not expressed in the same cells as those RNAs ([Figure 1A](#)).

We expressed V5-tagged versions of each of the three PRC2 components—EED, EZH2, and SUZ12—that have been reported to bind directly to RNA by RIP,<sup>15–17,32,72</sup> CLIP,<sup>21,22</sup> and *in vitro*<sup>5,45,59</sup> experiments. We confirmed that the tagged proteins are (1) well expressed ([Figures S1A and S1B](#)), (2) specifically purified using a V5 antibody ([Figure S1C](#)), (3) properly incorporated into the endogenous PRC2 complex ([Figure S1D](#)), and (4) retain their RNA-binding activity *in vitro* ([Figure S2](#)). Consistent with previous observations,<sup>21,22</sup> we observed that radioactive isotope (<sup>32</sup>P)-labeled RNA that co-purified with each of the PRC2 components showed an enriched band near the expected protein size but did not display a clear RNase-dependent size shift ([Figure S3A](#)).

To account for potential experimental or analytical differences between species and to directly compare PRC2 binding with the same RNAs, we transfected the tagged protein into a human cell line (+tag) and mixed it with untransfected mouse cells ([Figure 1B](#)), and in parallel, transfected the tagged protein into a



**Figure 1. A method to identify RNA-protein associations that could not have formed *in vivo***

(A) Schematic of mixing experiment. An epitope-tagged protein is expressed in cells (+tag, red), UV-crosslinked, lysed, and mixed with UV-crosslinked cell lysate from cells of a different species not expressing the tagged protein (–tag, gray). The tagged protein is purified using an antibody against the epitope tag, and purified RNAs are sequenced and aligned to quantify the amount of RNA associated with +tag and –tag RNAs, respectively.

(B and C) CLIP enrichment profiles for each PRC2 protein (EED, EZH2, and SUZ12) are plotted across XIST in the +tag (red) samples (B) and in the –tag (gray) samples (C). Input reads for the EZH2 samples are plotted in light gray.

(D and E) Scatter plots of input RNA abundance (log scale, x axis) compared with CLIP enrichment (log scale, y axis) across 100-nt windows of all annotated human RNAs in +tag (left) and –tag (right). Windows with significant enrichment (binomial  $p < 10^{-6}$ ) are shown in red. Plots include all 3 PRC2 components; individual components are plotted in Figures S4A and S4B.

(F) Density scatter plot comparing the levels of significant ( $p < 10^{-6}$ ) +tag CLIP enrichments (x axis) with significant –tag CLIP enrichments (y axis) for all 3 PRC2 components across all human RNAs.

mouse cell line and mixed it with untransfected human cells (–tag) (Figure 1C). We performed CLIP on each PRC2 component using the same V5 antibody and RNase concentration, sequenced the purified RNAs, and compared the +tag and –tag samples within the same species (human).

We observed that ~40% of expressed RNAs are significantly enriched for binding of all 3 PRC2 components in the +tag samples relative to their expression levels in total RNA (~6,300 RNAs,  $p < 10^{-6}$ , Figures 1D and S4A). For example, we observe strong enrichment for all 3 PRC2 components across several lncRNAs that have previously been reported to bind to PRC2 including XIST, MALAT1,<sup>73</sup> and NEAT1<sup>74</sup> (Figures 1B and S4C). In addition, we observed strong binding to NORAD, a lncRNA that is predominantly localized in the cytoplasm (Figure S4C).<sup>75,76</sup>

Surprisingly, we also observed significant PRC2 binding in the –tag samples with >850 RNAs containing significantly enriched binding sites for all 3 PRC2 components (Figures 1E and S4B). Indeed, many of the same RNAs that were previously reported to bind to PRC2 and that were identified in the +tag samples also showed significant enrichment in the –tag samples (i.e., MALAT1, NEAT1, and NORAD; Figure S4C). Overall, we observed a strong global correlation between RNA regions that are highly enriched in the –tag samples and those that are enriched in the +tag samples (Spearman correlation = 0.625; Figure 1F). For example, when we focused on XIST, we observed that the 3 PRC2 components show highly comparable profiles in the +tag and –tag samples; both display broad enrichment across the RNA, with the strongest enrichment over the A-repeat as previously reported<sup>9,15</sup> (Figures 1B and 1C).



These results demonstrate that at least some of the RNA signal detected upon purification of PRC2 using CLIP represents non-specific signal that cannot reflect *in vivo* interactions. Furthermore, these non-specific associations are UV-crosslinking and PRC2 dependent as we did not detect RNA in the absence of UV-crosslinking or in cells lacking the immunoprecipitated protein (i.e., untransfected cells) (Figure S3B). Importantly, these non-specific (–tag) associations are not uniformly distributed across an RNA but often appear as “peaks” that could be mistaken for legitimate binding sites using standard analytical methods (Figures 1C and S4C). While these observations do not preclude the possibility that these PRC2 components may also bind to RNA *in vivo*, they highlight the challenge in accurately determining which of the detected PRC2-RNA interactions (if any) may represent bona fide interactions that occur *in vivo*.

### Denaturing purification removes non-specific associations

To determine whether PRC2-RNA interactions occur *in vivo*, we need to confidently exclude non-specific associations. Because CLIP utilizes UV-crosslinking to generate covalent RNA-protein interactions in cells, any detected non-specific associations must be due to the inability to fully separate crosslinked from non-crosslinked RNA-protein interactions. We considered several possible explanations for why UV- and protein-dependent non-specific associations might be detected. These include potential protein-dependent sources of background where (1) the captured protein may associate with other proteins that are crosslinked to RNA or (2) other proteins that are crosslinked to RNA may be retained after immunoprecipitation (Figure 2A). Additionally, we considered potential RNA-dependent sources of background where (3) the captured protein may interact with RNA that is crosslinked to another RBP or (4) the captured protein may interact with free, non-crosslinked RNA (Figure 2A). Any of these non-specific associations that remain after immunoprecipitation would be detected because the protein purification (immunoprecipitation) and denaturation steps (gel electrophoresis) are decoupled in the CLIP procedure. These potential sources of background binding would be especially problematic when the captured protein does not actually bind to RNA, or binds to rare RNA targets, *in vivo* because non-specific RNA targets would be present in vast excess relative to bona fide targets (see STAR Methods, possible explanations for UV- and protein-dependent non-specific associations).

To address these possibilities, we utilized a method that enables purification of RNA-protein interactions using fully denaturing conditions, called covalent linkage and affinity purification (CLAP)<sup>77,78</sup> (Figure 2B). CLAP, similar to other methods that utilize covalent linkage,<sup>79–81</sup> integrates an epitope tag that enables covalent coupling to a resin (e.g., HaloTag,<sup>82</sup> SpyTag<sup>83</sup>) into a protein of interest. Because proteins are covalently coupled (rather than captured through an antibody), we can purify using fully denaturing conditions—including high temperatures and high concentrations of denaturants and detergents—which disrupt protein and RNA folding. We found that CLAP-conditions generally increased the specificity of both (1) protein purification and (2) crosslinked RNA purification relative to CLIP-conditions (Figures 2C, 2D, S5A, and S5B). Specifically, we found that

CLAP washes significantly reduce non-specific protein binding (“protein-dependent background,” Figure 2C) and the amount of non-crosslinked RNA (“RNA-dependent background,” Figures 2D and S5B) that co-precipitates with the target protein relative to CLIP washes.

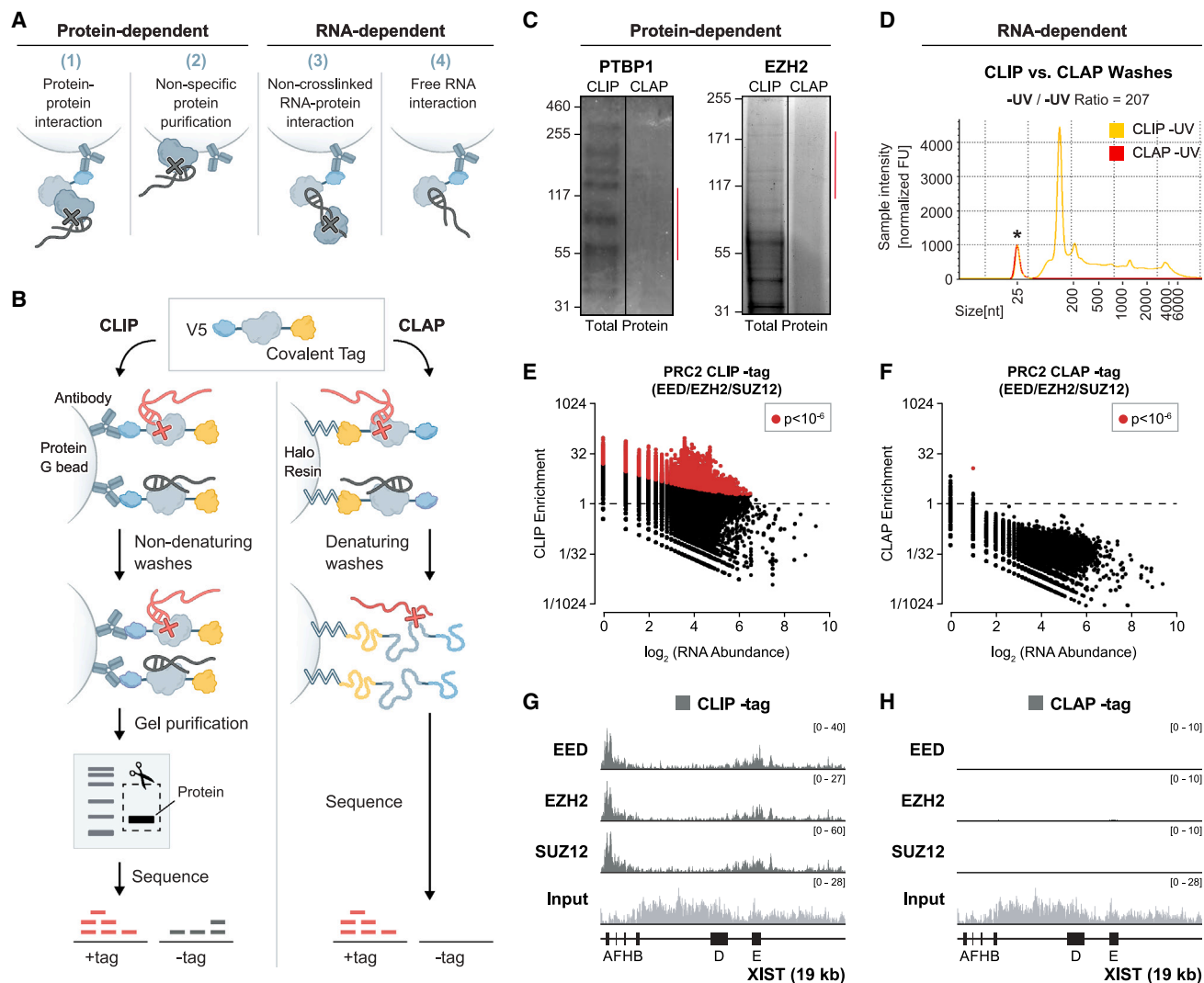
Having established that the increased stringency of CLAP can reduce both protein- and RNA-dependent sources of background, we explored whether CLAP could exclude non-specific RNA-protein associations that were previously detected in our –tag system. To do this, we expressed proteins tagged with both Halo and V5 tags, split the lysate, and performed CLIP and CLAP captures from the same mixture to directly compare the contribution of –tag RNAs in each experiment. We found that CLAP greatly reduces the proportion of –tag RNAs recovered for all 3 of the PRC2 components relative to CLIP (Figures 2E, 2F, and S5C). In fact, virtually all of the RNA regions that were significantly enriched in the –tag CLIP samples were depleted when measured by CLAP (Figures 2E–2H and S5C). These results demonstrate that CLAP accurately removes non-specific RNA-protein associations that do not occur *in vivo*.

### Denaturing purification accurately retains *in vivo* crosslinked RNA-protein interactions

To ensure that CLAP can still identify bona fide RNA-protein interactions that occur *in vivo*, we explored two well-defined RBPs that are known to interact with RNA through distinct binding modes: (1) PTBP1 is an RBP that contains multiple RNA recognition motif (RRM) domains, binds predominately within intronic regions, and has high selectivity toward a defined RNA sequence motif (HYUUUYU)<sup>84</sup> and (2) SAF-A (also known as hnRNP-U) is an RNA-binding protein that contains tandem RGG (arginine-glycine-glycine) motifs and binds promiscuously to many nascent pre-mRNAs with a broad localization profile.<sup>85</sup>

We performed CLIP and CLAP on each protein across replicates (Figures 3A, S6A, and S6B), sequenced the RNA, and observed the expected binding sites with PTBP1 binding primarily at intronic regions containing its known motif (Figures 3B and S3D) and SAF-A binding broadly across nascent pre-mRNAs (Figure 3C). Indeed, the vast majority of RNA regions significantly enriched by CLIP were also enriched by CLAP and the levels of enrichment were highly correlated between the two approaches (Figures 3D–3I) and highly reproducible between replicates (Figures S6A and S6B). Consistent with this, when we visualized <sup>32</sup>P-labeled RNA co-purifying with PTBP1, we observed a clear RNase-dependent size shift that resolves to the protein size (Figure S3C) with similar RNA sizes and amounts observed by CLIP and CLAP (Figure 3A).

These results demonstrate that CLAP accurately and sensitively identifies bona fide RNA-protein interactions that are crosslinked *in vivo*. Because both CLIP and CLAP utilize UV-crosslinking to form covalent RNA-protein interactions *in vivo* and are both premised on specifically detecting crosslinked interactions, there is no intrinsic difference between the two approaches in their ability to detect interactions of different affinities or stability (see STAR Methods, failure of CLAP to identify RNA-protein interactions identified by CLIP cannot be due to differences in assay sensitivity).



**Figure 2. CLAP removes non-specific RNA-protein associations**

(A) Protein-dependent background (left) and RNA-dependent background (right) that could lead to detection of RNA not crosslinked to the purified protein (gray RNAs) by CLIP.

(B) Comparison of CLIP (left) and CLAP (right). A protein tagged with both a covalent tag (HaloTag or SpyTag) and V5 epitope tag is expressed. The sample is split, and CLIP and CLAP are performed separately. CLIP is performed with an anti-V5 antibody followed by standard CLIP washes, gel electrophoresis, transfer to a nitrocellulose membrane, and size selection prior to RNA sequencing. CLAP is performed by covalently binding the protein to resin followed by washes in fully denaturing conditions prior to RNA sequencing.

(C) Halo-PTBP1-V5 (left) and Halo-EZH2-V5 (right) protein were captured on HaloLink resin, washed with either CLIP or CLAP wash buffers, and remaining associated proteins were eluted (via heat), separated by SDS-PAGE, and detected using SyproRuby total protein stain. Red lines indicate regions usually cut for CLIP (~70 kDa above protein molecular weight).

(D) Equivalent amounts of non-crosslinked HEK293T (–UV) whole-cell lysate were coupled to amine-reactive beads and washed with either CLIP or CLAP wash buffers. The remaining bound RNA-protein complexes was eluted using proteinase K and associated RNAs measured. \* denotes the lower marker used for sizing.

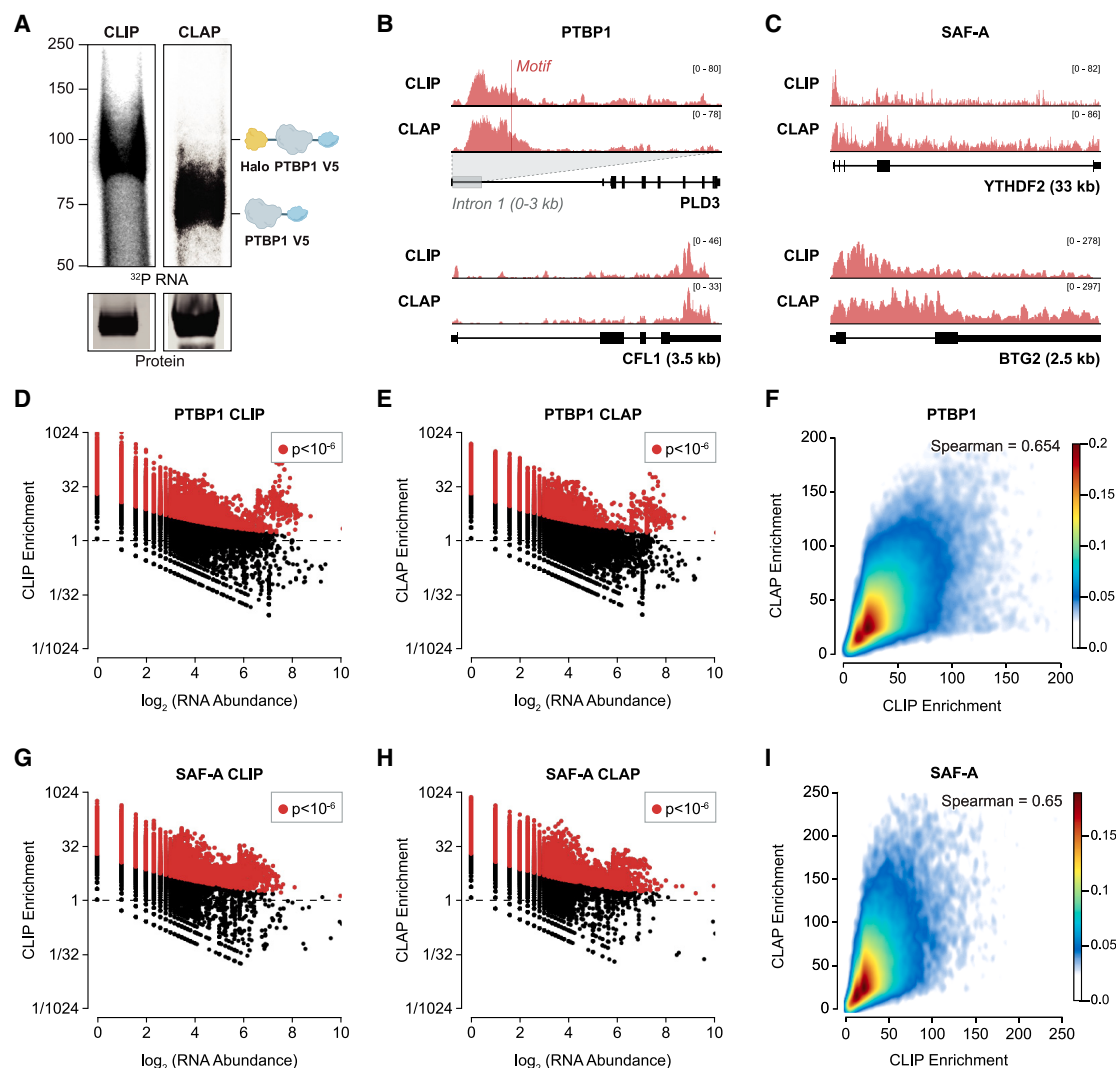
(E and F) Scatter plots (left, CLIP; right, CLAP) of input RNA abundance compared with enrichment across 100-nt windows of all human RNAs in the –tag experiments. Plots include all 3 PRC2 components; individual components are plotted in Figures S4B and S5C.

(G and H) Enrichment profiles for each PRC2 component (EED, EZH2, and SUZ12) in the –tag samples are plotted across the human XIST lncRNA for CLIP (left, same as Figure 1C) and CLAP (right). Input reads from EZH2 samples are plotted in light gray (CLIP and CLAP input are identical because they come from the same lysate).

### PRC2 components do not detectably bind to RNA *in vivo*

To explore whether PRC2 components bind promiscuously to RNA *in vivo*, we performed CLAP on each of the three PRC2 components (EED, EZH2, and SUZ12), labeled co-purified

RNA with a radioactive isotope ( $^{32}\text{P}$ ), and visualized the absolute amount of labeled RNA bound to each protein on an SDS-PAGE gel. In all cases, we detected no labeled RNA for any of the PRC2 components (Figures 4A and S7A), despite successfully



**Figure 3. CLAP accurately maps *in vivo* crosslinked RNA-protein interactions**

(A) Visualization of radiolabeled RNA ( $^{32}\text{P}$ ) co-purified with Halo-PTBP1-V5 by either CLIP (left) or CLAP (right). Protein capture was verified by western blot (below). Lower molecular weight in CLAP due to TEV cleavage required to release from resin. Expected molecular weights are indicated.

(B) Examples of CLIP and CLAP enrichments for PTBP1 over PLD3 pre-mRNA (top, intronic region spanning 0–3,000 nt) and CFL1 mRNA (bottom). Location of PTBP1 motif is shown (red line).

(C) Examples of CLIP and CLAP enrichments for SAF-A over YTHDF2 mRNA and BTG2 mRNA. Exons are denoted by boxes and introns by connecting lines.

(D and E) Scatter plots of input RNA abundance compared with enrichment across 100-nt windows of all human RNAs identified for PTBP1 by CLIP (left) or CLAP (right).

(F) Density scatter plot comparing the levels of significant PTBP1 enrichment ( $p < 10^{-6}$ ) between CLIP (x axis) and CLAP (y axis) across all human RNAs.

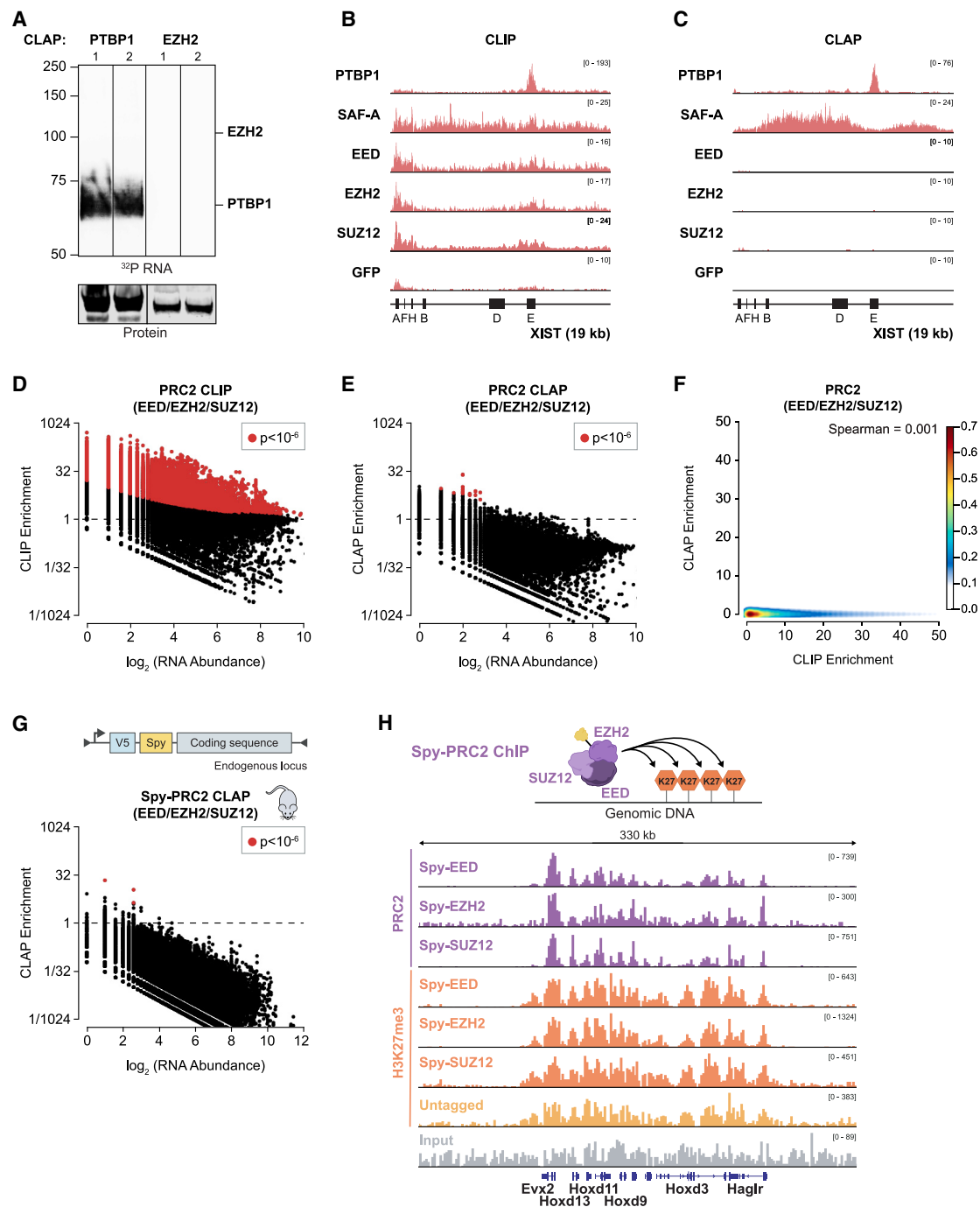
(G and H) Scatter plots of input RNA abundance compared with enrichment across 100-nt windows of all human RNAs identified for SAF-A by CLIP (left) or CLAP (right). Windows with significant enrichment (binomial  $p < 10^{-6}$ ) are shown in red.

(I) Density scatter plot comparing the levels of significant SAF-A enrichment ( $p < 10^{-6}$ ) between CLIP (x axis) and CLAP (y axis) across all human RNAs.

purifying each protein. In contrast, when we performed CLAP on PTBP1, we detected large amounts of radiolabeled RNA from a comparable amount of purified protein (Figures 4A, S7A, and S7B).

We next considered the possibility that PRC2 may bind to rare RNA targets that cannot be sensitively detected on a gel. To explore this, we performed CLAP followed by sequencing of the purified RNA. Importantly, we observed a strong global reduction

in binding of all 3 of the PRC2 components to RNA (Figures 4B–4E and S8A) with >99.97% of the RNA regions that were significantly enriched by CLIP being depleted in the CLAP samples (Spearman correlation = 0.001, Figure 4F). (The very few RNA regions that showed significant enrichment in any PRC2 sample tended to correspond to regions of low coverage and were not reproducible across replicates [Figure S8A].) For example, PRC2 binding that was observed over XIST by CLIP was depleted when measured



**Figure 4. PRC2 components purified using denaturing conditions do not appear to bind RNA**

(A) Visualization of radiolabeled RNA ( $^{32}$ P) purified by CLAP from Halo-V5-tagged versions of PTBP1 and EZH2 across independent biological replicates. Protein capture was verified by western blot (below).

(B and C) Enrichments for PTBP1, SAF-A, EED, EZH2, SUZ12, and GFP plotted across XIST in the +tag experiments for CLIP (left; EED, EZH2, and SUZ12 same as Figure 1B) or CLAP (right).

(D and E) Scatter plots of input RNA abundance compared with PRC2 enrichment across 100-nt windows of all human RNAs in the +tag experiments for CLIP (left; same as Figure 1D) or CLAP (right). The plot includes all 3 PRC2 components; individual components are plotted in Figure S8A.

(F) Density scatter plot comparing the levels of +tag CLIP enrichments (x axis) to +tag CLAP enrichments (y axis) for all 3 PRC2 components across 100-nt windows of all human RNAs.

(legend continued on next page)



by CLAP (Figures 4B, 4C, S8C). These results are strikingly different from the RNA-binding profiles observed by CLAP for PTBP1 or SAF-A (Figures 4B and 4C), regardless of the precise p-value cutoff utilized (Figure S7C). Instead, we find that the overall signal observed for each of the 3 PRC2 components is comparable to the level observed when we perform CLAP on GFP (Figures 4C and S8B), which does not bind RNA *in vitro* (Figures S2 and S11) and does not have any endogenous targets in mammalian cells.

This global reduction of PRC2 binding to RNA was observed even though CLIP and CLAP were performed on the same protein from the same lysate and CLAP purified more protein than CLIP in all cases (Figures S9A–S9D). In addition, we confirmed that the Halo-tagged PRC2 components (1) were successfully purified in all cases (Figure S10), (2) could still be properly incorporated into the endogenous PRC2 complex (Figure S1D), (3) were expressed at levels that exceed endogenous levels (Figures S1A and S1B), (4) retain their reported RNA-binding activity *in vitro* (Figure S2), and (5) can form UV-induced crosslinks with RNA when the interaction is assembled *in vitro* (Figure S11A; STAR Methods, Halo-tagged PRC2 components and purified PRC2 complexes associate and UV-crosslink with RNA *in vitro*).

Because PRC2 is a multi-component protein complex, we wanted to exclude the possibility that overexpression of exogenous PRC2 components might disrupt the complex and therefore impact RNA binding. While this possibility cannot explain why CLIP observes non-specific binding, we wanted to ensure that this was not the reason we failed to observe RNA binding by CLAP. To address this, we integrated an in-frame SpyTag (a distinct covalent tag that is ~10-fold smaller than the HaloTag; 3 versus 30 kDa) into both alleles of the endogenous *Eed*, *Ezh2*, or *Suz12* genes in mouse embryonic stem cells (mESCs) using CRISPR-Cas9 (Figures 4G and S12A). When we performed CLAP on these Spy-tagged PRC2 proteins, we observed a similar global depletion of RNA binding (Figure 4G). Despite the lack of PRC2-RNA binding, these PRC2 components properly localize to chromatin and deposit H3K27me3 at the expected locations across the genome (Figure 4H), including on the inactive X (Figure S12B). In contrast, when we performed CLAP on endogenously Spy-tagged PTBP1, we observed >10,000 enriched RNA regions corresponding to known PTBP1 binding sites (Figures S12C and S12D).

Together, these results indicate that PRC2 does not appear to bind directly to RNA *in vivo*—promiscuously or specifically—and that direct RNA interactions are not required for PRC2 to localize to chromatin or deposit H3K27me3.

### Several chromatin regulators reported to bind RNA do not appear to bind *in vivo*

Many additional chromatin proteins have been reported to bind to RNA,<sup>4,32,39,86–89</sup> including repressors and activators of transcription,<sup>16,36,37,90</sup> DNA-binding proteins,<sup>6,38</sup> and proteins

involved in 3D chromatin organization.<sup>8,40,41,91</sup> Given our observations about PRC2, we explored whether some of these other chromatin regulatory proteins may also fail to bind directly to RNA *in vivo*. We performed CLAP on three additional proteins—CTCF, YY1, and WDR5—that are known to play distinct roles in chromatin regulation and have been widely reported to bind RNA.

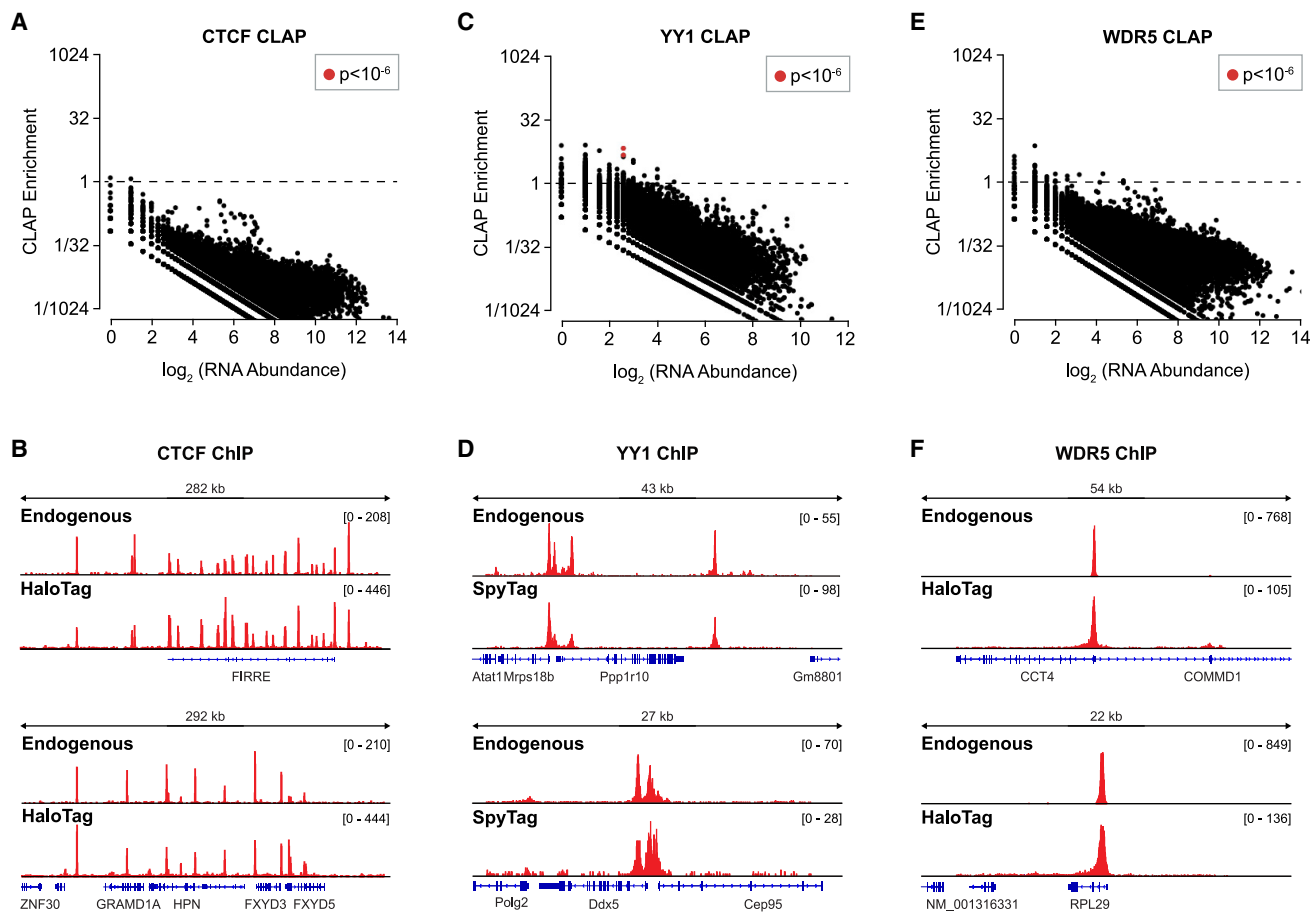
(1) CTCF is a zinc-finger-domain-containing protein that directly binds DNA and is critical for the formation of topologically associated domain structures.<sup>92,93</sup> Recently, CTCF has been reported to bind RNA to guide CTCF to specific locations on genomic DNA,<sup>7</sup> locally constrain CTCF mobility on chromatin,<sup>94</sup> and shape higher-order chromatin structure.<sup>7,8,40,41</sup> The evidence for CTCF as an RBP comes from CLIP data indicating that CTCF binds many different RNA targets including lncRNAs such as *Tsix*, *Firre*, and others.<sup>7</sup> Yet, in contrast to what is observed for other well-defined RBPs, CLIP analysis of CTCF shows strong RNA association even in the absence of *in vivo* UV-crosslinking.<sup>7</sup> Moreover, disruption of a specific region of CTCF that was reported to ablate interaction with RNA *in vitro* (termed the “RNA-binding region”<sup>8</sup>) still leads to detection of an appreciable amount of RNA by CLIP.<sup>41</sup>

To explore whether CTCF interacts with RNA *in vivo*, we generated a Halo-V5-tagged CTCF protein and expressed it in HEK293T cells. When we performed CLIP on this fusion protein, we observed comparable amounts of RNA in the presence and absence of UV-crosslinking, indicating that the HaloTag does not disrupt CTCF’s previously reported associations with RNA (Figure S13A). However, when we performed CLAP and sequenced the purified RNA, we did not identify a single mRNA or lncRNA region (including *FIRRE*) that was enriched in either of two independent CLAP experiments, even though we successfully purified the protein in both cases (Figures 5A, S13B, and S13E). Because RNA binding to CTCF has been proposed to impact CTCF localization, we performed genome-wide mapping of Halo-tagged CTCF by chromatin immunoprecipitation sequencing (ChIP-seq) and observed highly comparable localization patterns to those of endogenous CTCF proteins (Figure 5B). This indicates that the HaloTag does not disrupt CTCF localization to chromatin or its ability to bind DNA. Together, these results indicate that CTCF does not appear to bind directly to RNA *in vivo* and that direct RNA binding is not critical for CTCF localization on chromatin.

(2) YY1 is a DNA-binding protein that is thought to play an important role in mediating enhancer-promoter loop interactions.<sup>95</sup> YY1 was reported to bind broadly to RNA based on CLIP data, and these widespread RNA interactions were proposed to be important for tethering YY1 to DNA.<sup>6,39</sup> To explore if YY1 interacts directly with RNA *in vivo*, we generated a mESC line with an in-frame SpyTag integrated into the endogenous *Yy1* allele and performed CLAP (Figure S13C). When we sequenced the associated RNA, we failed to identify a single

(G) Integration strategy at endogenous locus for generating V5-Spy-tagged proteins (top). Scatter plot of input RNA abundance compared with CLAP enrichment for all endogenous Spy-tagged PRC2 proteins across 100-nt windows of all mouse RNAs.

(H) ChIP-seq against each V5-Spy-tagged PRC2 component (top, purple). ChIP-seq on H3K27me3 from each tagged (middle, orange) and untagged cell line (bottom, yellow). Read coverage is plotted across the *HOXD* cluster along with input (light gray).



**Figure 5. Several chromatin regulators reported to bind RNA do not appear to bind *in vivo***

(A) Scatter plot of input RNA abundance compared with CLAP enrichment across 100-nt windows of all human RNAs for Halo-CTCF-V5.  
(B) ChIP-seq of endogenous CTCF versus Halo-V5-tagged CTCF.  
(C) Scatter plot of input RNA abundance compared with CLAP enrichment across 100-nt windows of all mouse RNAs for V5-Spy-YY1.  
(D) ChIP-seq of endogenous YY1 versus V5-Spy-tagged YY1.  
(E) Scatter plot of input RNA abundance compared with CLAP enrichment across 100-nt windows of all human RNAs for Halo-WDR5-V5.  
(F) ChIP-seq of endogenous WDR5 versus Halo-V5-tagged WDR5.

RNA that was enriched for YY1 binding (Figure 5C). Because YY1 RNA binding has been proposed to impact YY1 localization on chromatin, we performed genome-wide ChIP-seq experiments and observed localization patterns that were highly similar to those previously reported for the endogenous protein (Figure 5D). We observed a similar global depletion of RNA binding when we performed CLAP on an expressed Halo-tagged YY1 in HEK293T cells (Figures S13D and S13E). Together, these results indicate that YY1 does not appear to bind directly to RNA *in vivo* and that RNA binding is not critical for YY1 localization on chromatin.

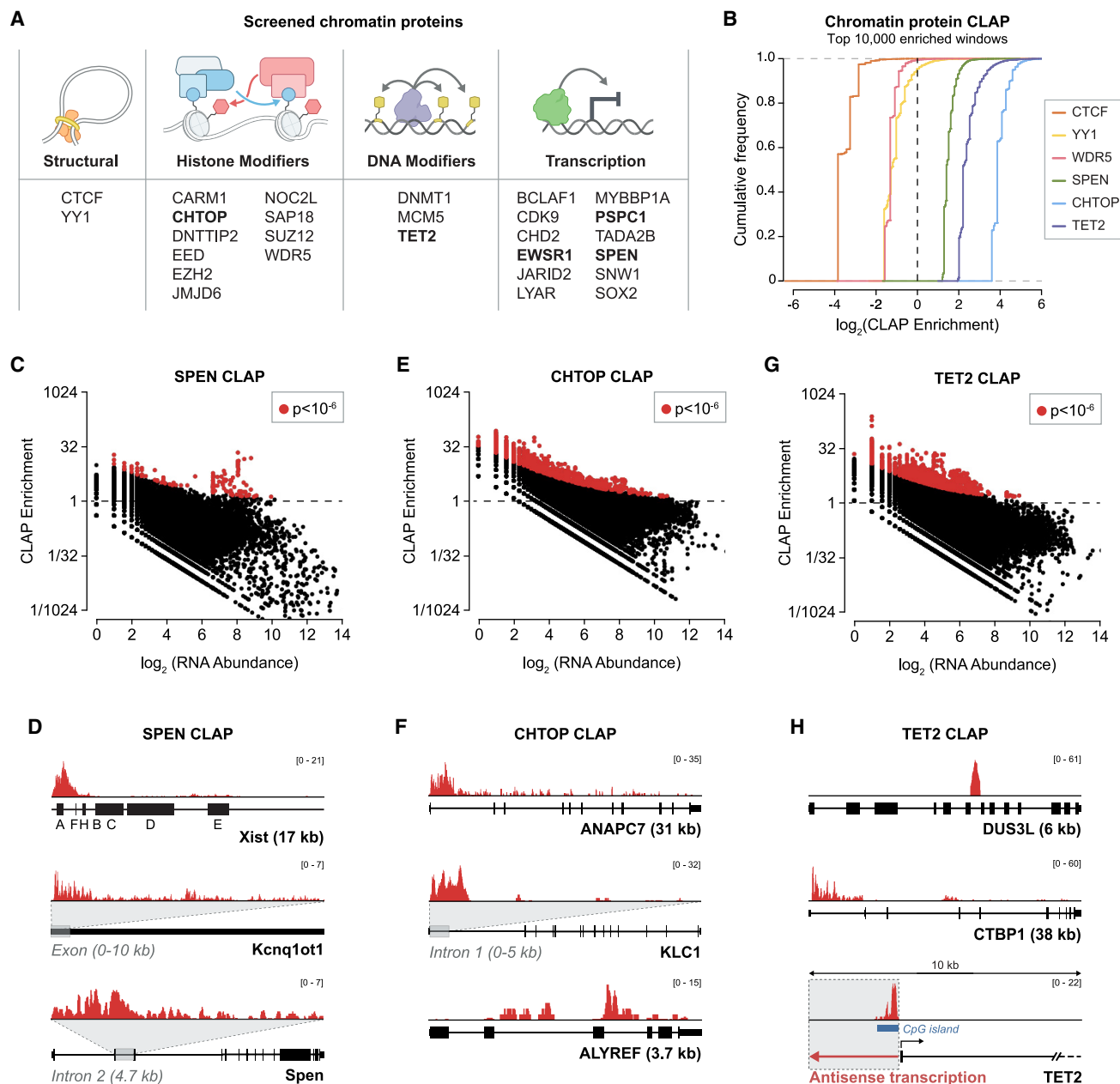
(3) WDR5 is a component of the MLL complex, which deposits the H3K4me3 modification, a mark of transcriptionally active chromatin.<sup>96</sup> RIP experiments have shown that WDR5 binds to many different lncRNAs, including HOTTIP, leading to the proposal that lncRNAs act to guide the MLL complex to maintain an active chromatin state.<sup>35</sup> To determine whether WDR5 binds directly to RNA, we expressed a Halo-V5-tagged WDR5 protein in human HEK293T cells and performed CLAP. Similar to our ob-

servations with PRC2, CTCF, and YY1, we did not identify any RNAs that were significantly enriched for interactions with WDR5, including HOTTIP, despite successfully purifying the protein (Figures 5E, S13B, and S13E). Given the proposed role for WDR5-RNA interactions in guiding WDR5 to chromatin, we performed ChIP-seq on Halo-tagged WDR5 and observed highly comparable localization patterns to those observed for the endogenous WDR5 complex (Figure 5F). These results demonstrate that RNA interactions are not essential for WDR5 localization and function on chromatin.

Taken together, these data suggest the need for a critical evaluation of the RNA-binding properties of these and other chromatin proteins.

### Specific chromatin proteins can bind directly to RNA *in vivo*

Although several chromatin proteins do not appear to bind directly to RNA *in vivo*, we explored whether other specific



**Figure 6. Denaturing purification identifies specific chromatin proteins that bind to RNA *in vivo***

(A) Functional categories of chromatin proteins tested by CLAP (see STAR Methods). Proteins identified as RNA-binding proteins by CLAP are bolded.

(B) Cumulative distribution plot for the top 10,000 enriched 100-nt windows across all RNAs for chromatin proteins measured by CLAP.

(C, E, and G) Scatter plots of input RNA abundance compared with CLAP enrichment across 100-nt windows of all RNAs for a set of Halo-V5-tagged chromatin proteins (SPEN, CHTOP, and TET2).

(D) Examples of CLAP enrichment profiles for SPEN across Xist (top), Kcnq1ot1 (middle, 0–10 kb), and Spn pre-mRNA (bottom, intron 2, ~4.7 kb).

(F) Examples of CLAP enrichment profiles for CHTOP over the ANAPC7 pre-mRNA (top), KLC1 pre-mRNA (middle, first intron, 0–5 kb), and ALYREF pre-mRNA.

(H) Examples of CLAP enrichment profiles for TET2 over the DUS3L pre-mRNA (top), CTBP1 pre-mRNA (middle), and an antisense RNA transcribed from the TET2 promoter (bottom, region spanning 10 kb). Blue box indicates a CpG island, and arrows indicate direction of transcription (black, sense; red, antisense).

chromatin proteins might. To do this, we analyzed previously published data generated from RNA purification in UV-cross-linked cells followed by global protein identification using mass spectrometry<sup>2,4,87,88,97–99</sup> to select a list of 20 additional proteins

with annotated roles in a range of chromatin functions, including 3D structure, histone modifications, DNA modifications, and transcription<sup>100</sup> (Figure 6A, see STAR Methods). We expressed Halo-V5-tagged versions of these proteins and performed

CLAP. While we did not detect significant RNA binding in most cases, because we did not characterize each protein in detail, we do not make any claims as to whether these represent bona fide RBPs (see [limitations of the study](#)). Instead, we focused on five chromatin-associated proteins—SPEN, EWSR1, CHTOP, PSPC1, and TET2—that are robustly expressed ([Figure S14](#)) and bind to specific RNA regions *in vivo* across multiple replicates using CLAP ([Figures 6B and S15A](#)).

(1) SPEN (also called SHARP) interacts with the SMRT/NCOR2 corepressor to activate histone deacetylase activity of HDAC3 and recruit it to specific genomic sites.<sup>55,101</sup> We and others identified SPEN as the critical protein that binds to the A-repeat of Xist and that is essential for initiation of silencing on the X.<sup>55,68,102,103</sup> While the interaction between Xist and SPEN was previously observed by CLIP experiments,<sup>55,104,105</sup> few other RNA interactions were identified. We reasoned that by reducing RBP association with non-specific RNA species, CLAP might increase the sensitivity for detecting bona fide RNA-protein interactions that occur with lower abundance RNAs. To explore this, we performed CLAP on SPEN in mESCs and identified ~500 significant binding sites within ~200 RNAs ([Figure 6C](#)). These included the known binding sites within the A-repeat region of Xist<sup>78,104–106</sup> ([Figure 6D](#)). We also identified binding sites within several other lncRNAs, including Kcnq1ot1 and Chaserr ([Figures 6D and S16A](#)). Because Kcnq1ot1 and Chaserr have been shown to repress expression of their neighboring genes,<sup>101,107–109</sup> their direct interaction with SPEN may in part explain how they achieve these roles. We also observed strong enrichments at specific sites within the introns of protein-coding genes, including within the second intron of the Spen pre-mRNA ([Figure 6D](#)); these interactions might act to modulate transcription of these genes.

(2) EWSR1 is a multifunctional protein belonging to the FET (FUS, EWSR1, and TAF15) family of proteins that bind DNA and regulate gene expression.<sup>110,111</sup> EWSR1 is known to associate with the basal transcriptional machinery<sup>112</sup> (transcription factor IID and RNA polymerase II<sup>113</sup>) and transcriptional coactivators (CREB-binding protein and histone acetyltransferase p300) to stimulate transcriptional activation.<sup>114</sup> EWSR1 functions have been attributed in part to its ability to undergo oligomerization through RNA-dependent phase separation in cells.<sup>113</sup> We performed CLAP on EWSR1 and observed >700,000 significantly enriched binding sites within thousands of RNAs. The majority of these EWSR1 binding sites occur within intronic regions of pre-mRNAs ([Figures S15B and S15C](#)), consistent with previous reports by CLIP.<sup>115–117</sup> Because RNA is thought to seed the binding of FET proteins to RNA polymerase II,<sup>118,119</sup> and the depletion of EWSR1 leads to a global reduction in nascent transcription,<sup>120</sup> these EWSR1-RNA interactions may explain how FET proteins mediate widespread transcriptional activation of most genes. In addition, we observe EWSR1 binding within the intron of its own mRNA as well as the intron of the TAF15 mRNA, another member of the FET family ([Figure S16B](#)), suggesting that EWSR1 might act to autoregulate its own expression as well as the closely related and functionally redundant FET proteins.<sup>121,122</sup>

(3) CHTOP is a component of an arginine methyltransferase complex<sup>123</sup> containing PRMT1 and PRMT5 that promotes methylation of arginine 3 of Histone H4 (H4R3), a histone modi-

fication associated with transcriptional activation.<sup>124</sup> CHTOP has also been reported to bind to DNA sequences containing 5-hydroxymethylcytosine (5hmC),<sup>123,125,126</sup> an intermediate of DNA methylation and an epigenetic mark that recruits DNA-binding proteins.<sup>127,128</sup> Beyond its reported DNA and chromatin roles, CHTOP was also recently identified as a novel component of the TREX mRNA export complex.<sup>125,126</sup> To explore CHTOP binding, we performed CLAP and identified >40,000 enriched sites within ~7,000 RNAs, including within the introns of ALYREF ([Figures 6E and 6F](#)), a core factor of the TREX complex.<sup>129,130</sup> Globally, CHTOP intronic binding sites were predominantly located at the 5' end of the first intron ([Figures 6F and S16C](#)), possibly due to co-transcriptional recruitment to pre-mRNAs via splicing, similar to previous observations for TREX binding to mRNA.<sup>131</sup> The association between CHTOP and nascent pre-mRNAs may act to facilitate its recruitment to the genomic DNA region of actively transcribed genes.<sup>126</sup>

(4) PSPC1 is a nuclear protein enriched within paraspeckles, a nuclear structure thought to fine-tune gene expression by sequestering proteins away from their target genes.<sup>132</sup> In addition, PSPC1 has been reported to localize at promoters<sup>120,133</sup> and guide chromatin regulators,<sup>134</sup> including TET2 and HDAC1/2, to transcriptionally active genomic loci through its interactions with nascent RNAs. We performed CLAP on PSPC1 and observed binding to NEAT1 ([Figures S15D and S16A](#)), a lncRNA that is essential for paraspeckle assembly.<sup>135</sup> Beyond NEAT1, we identified significant PSPC1-binding sites within the introns and 3' UTRs of >10,000 mRNAs ([Figures S15B and S16B](#)); these binding preferences are consistent with those observed by previous CLIP experiments.<sup>74,136</sup> These include specific enrichment within the introns of HDAC- and TET-related mRNAs, such as SIN3B and HDAC8 ([Figure S15A](#)). Because PSPC1 binds to a large number of pre-mRNAs and interacts with numerous chromatin regulators, these RNA interactions may be important to fine-tune gene expression at sites of active transcription.

(5) TET2 is a dioxygenase that catalyzes stepwise DNA demethylation by converting 5-methyl cytosine (5mC) to 5-hydroxymethyl cytosine (5hmC) on DNA.<sup>137</sup> To explore TET2 binding to RNA, we performed CLAP and uncovered >7,000 binding sites within ~1,500 RNAs ([Figures 6G and S17A](#)). For example, we observe highly specific, focal binding sites within introns of the DUS3L and CTBP1 pre-mRNAs ([Figure 6H](#)). Motif analysis of TET2 enriched RNA-binding sites revealed a strong preference for G/C-rich sequences, consistent with its DNA substrate<sup>138</sup> (i.e., CpG dinucleotides) ([Figure S17B](#)). Globally, we observed striking enrichment of TET2 binding to RNAs that are located near promoters ([Figure S17C](#)) with TET2 binding to hundreds of antisense RNAs at promoters, including at its own mRNA locus ([Figures 6H and S17D](#)). These RNA-binding sites often overlapped CpG islands (~80% overlap) at promoters and within coding sequences (~30% of all peaks, [Figure S17E](#)). This localization is similar to the known DNA localization of TET2 primarily at CpG islands and promoters.<sup>139–143</sup> Because TET2 functions as a DNA demethylase, and because TET2 lacks a known DNA-binding domain,<sup>144</sup> these RNA interactions may recruit TET2 to DNA to enact its enzymatic functions. Consistent with this, TET2 localization on chromatin has been shown to be



sensitive to transcriptional inhibition.<sup>134</sup> Importantly, our global RNA-binding maps for TET2 contrast with those previously reported by CLIP, which were predominantly enriched for transfer RNAs, ribosomal RNAs, and murine endogenous retrovirus-L (MERVL) transcripts.<sup>134,145,146</sup> The highly specific binding sites we identified on lower abundance RNAs will be critical for dissecting potential RNA-dependent functions of TET2.

Together, these results indicate that while not all previously reported chromatin regulators appear to directly bind RNA *in vivo*, some chromatin-associated proteins do, and CLAP can robustly identify their binding sites even when they lack canonical RNA-binding domains (Figure S15E).

## DISCUSSION

### Need for careful re-evaluation of the role of RNA in PRC2 functions

Our results argue for a critical re-evaluation of the idea that PRC2 and other chromatin regulators bind directly to many RNAs and that such binding is essential for their function.

While it remains unclear why PRC2 appears to bind promiscuously to RNA *in vitro* but does not appear to bind directly *in vivo*, there are several critical differences between *in vitro* reaction conditions and the *in vivo* context of PRC2 in the nucleus. For example, in contrast to *in vitro* experiments where protein and RNA are well mixed at defined concentrations, the nucleus is highly compartmentalized<sup>147</sup> such that these components might not occupy the same locations. As an example, PRC2 proteins are generally enriched in silenced domains of the nucleus with few actively transcribed genes (“polycomb bodies”).<sup>148–150</sup> Additionally, in contrast to *in vitro* reactions where a single purified protein and RNA of interest are present, there are hundreds of different proteins and RNAs in the nucleus that compete for binding. In the case of Xist, which is present on the inactive X (a location enriched for PRC2), the absence of PRC2 binding may reflect the fact that other RBPs (e.g., SPEN/SHARP) have stronger binding affinity and outcompete PRC2 binding. Finally, *in vitro* binding between PRC2 and RNA is highly dependent on the precise buffer composition, RNA templates, and other reaction conditions<sup>58</sup>; the specific conditions that promote these interactions do not fully mimic the cellular environment.

Currently, the only evidence indicating that PRC2 binds to RNA *in vivo* comes from methods using UV-based crosslinking.<sup>4,21,22,70,151</sup> Importantly, formaldehyde crosslinking and other proximity-based methods do not provide evidence for direct *in vivo* binding because they also detect associations that are in proximity. For example, Xist is in proximity to PRC2 because Xist localizes to the inactive X, which is enriched for PRC2. Because we and others<sup>59</sup> have shown that PRC2 can form UV crosslinks with RNA when assembled *in vitro*, the lack of observed binding by CLAP indicates that PRC2 does not bind to RNA *in vivo*. Nonetheless, even if this is caused by the inability of PRC2 to UV-crosslink to RNA, these results still indicate that previous reports of direct PRC2-RNA binding *in vivo* are problematic. Accordingly, there is currently no compelling biochemical evidence to support the notion that PRC2 binds to RNA *in vivo*.

Recently, several other technical challenges to the PRC2-RNA model have emerged. For example, many of the published RIP and CLIP studies on EZH2 were shown to be confounded by cross-reactivity of the commonly used antibody with SAF-B,<sup>152</sup> a RBP that binds many RNAs.<sup>153</sup> Moreover, previous reports of a global role for RNA in PRC2 localization to chromatin<sup>23</sup> were shown to be primarily due to a technical issue and that, once corrected, PRC2 localization is independent of RNA.<sup>154,155</sup>

Beyond PRC2 and the chromatin proteins explored here, many additional proteins lacking classically defined RNA-binding domains—including transcription factors, chromatin regulators, and metabolic proteins—have been reported to bind RNA, suggesting the potential for a vast regulatory role for RNA. Yet, to date, few of these proteins have been demonstrated to function through their interaction with RNA.<sup>156</sup> This disconnect has led to an increasing body of literature that suggests putative functions for RNA-protein interactions that have proven difficult to build on and, conversely, has led to increasing skepticism about the functional relevance of ncRNAs more generally. Given these issues, it is essential for those reporting such interactions to robustly demonstrate that they occur in the cell and the functional roles of these interactions. A systematic understanding of which proteins are indeed bona fide RBPs and which specific RNAs they bind is required to enable rigorous functional and mechanistic studies.

### Considerations when evaluating RNA binding by CLIP and CLAP

CLIP has been used to accurately map the *in vivo* RNA-binding sites of numerous RBPs and has provided essential insights into their mechanisms of RNA recognition and functions in RNA processing and regulation.<sup>63–67,157,158</sup> Our results demonstrate that CLIP can quantitatively separate bona fide interactions that occur *in vivo* from non-specific associations for well-characterized RBPs (e.g., SAF-A and PTBP1) yet can fail to do so in cases where the proteins do not bind directly to RNA *in vivo* (e.g., PRC2 and GFP). We find that in the case of proteins that do not interact with RNA, CLIP can lead to detection of non-random associations that often show discrete UV-dependent and protein-specific “peaks” that could be mistaken for legitimate binding sites using standard analytical methods.

Despite these issues, the CLIP procedure itself is not the problem; rather, complications can arise from how the data are interpreted. Specifically, CLIP is generally performed on an individual protein and interpreted in isolation. When studying a known RBP, CLIP has proven to be a powerful and highly specific method for defining which specific regions of RNA are bound. Indeed, it was for this application that CLIP was initially developed<sup>67</sup> and for which it is most commonly used.<sup>159</sup> When it is applied to a protein that may not bind to any RNA, this approach can lead to non-random deviations that may appear as “significant” enrichments when analyzed in isolation. This is likely due to the fact that background signal is non-random and can be comparable even across different proteins.<sup>117,160,161</sup> For example, while PTBP1 and SAF-A exhibit distinct and specific binding patterns on XIST in +tag CLIP samples (Figure S18A), they display the same patterns as EED, EZH2, and SUZ12 in the –tag CLIP samples (Figure S18B).



Yet, even in these cases, we observe that there are clear qualitative and quantitative differences in the amount of RNA captured between bona fide RBPs and non-RBPs. For example, the absolute amount of RNA visually detected after CLIP for PRC2 or GFP is dramatically lower than for PTBP1 or SAF-A (Figure S19A). However, when PRC2 and GFP gels are contrasted in isolation, we observe some signal including at discrete bands corresponding to the expected protein size (Figure S19B). We observe a similar result when comparing different proteins within our +tag CLIP experiments; PTBP1 and SAF-A show strong enrichment relative to GFP (>5-fold) while PRC2 components were comparable to GFP (Figure S20A), similar to our observations by CLAP (Figure S20B).

These quantitative differences may explain some of the apparent discrepancies between conclusions from previous CLIP studies, including more stringent variants such as denaturing CLIP (dCLIP),<sup>34,70</sup> and those reported here. Accordingly, extreme care is needed especially when evaluating proteins that (1) lack canonical RNA-binding domains, (2) exhibit binding profiles that approximate those of input RNA, or (3) generate low-complexity sequencing libraries. In such cases, CLAP serves as a valuable orthogonal method to separate bona fide RNA-protein interactions that occur *in vivo* from potentially spurious background measurements.

### Limitations of the study

Our results demonstrate that CLAP significantly reduces non-specific signal while accurately mapping known RNA-protein interactions for well-characterized RBPs. CLAP has several additional technical advantages: (1) it takes significantly less time to perform because it does not require gel extraction, (2) it allows for fragmentation of RNA using heat and so eliminates structural biases associated with RNase digestion, and (3) it can be used with an exogenous protein expression system that eliminates the need for high-quality antibodies. However, one significant limitation is that it requires engineering a tagged protein. While this may be critical for studying specific proteins of interest, we appreciate that it is unlikely to be the ideal strategy for screening large numbers of putative RBPs.

Although our results show that PRC2 and other chromatin proteins do not appear to directly bind to RNA *in vivo*, they do not exclude the possibility that they may bind to specific RNAs in other contexts not explored here, or that they may bind to RNA indirectly through protein-protein interactions. Several orthogonal methods have been developed for exploring indirect RNA and protein association (adenosine-to-inosine editing,<sup>162</sup> proximity labeling,<sup>163–165</sup> and reverse transcribe and tagment (RT&Tag)<sup>166</sup>). In addition, because CLAP does not require size-based gel extraction, it can be used with orthogonal crosslinking methods (i.e., formaldehyde) to study RNA-protein assemblies that occur through indirect protein-protein contacts.

The absence of detectable RNA binding can be caused by multiple different factors and on its own does not indicate that a protein does not bind to RNA. For example, if a tagged protein is expressed at very low levels, it might not purify detectable amounts of RNA. Conversely, if a protein is expressed at very high levels such that it disrupts proper assembly into a multi-

component complex it might also fail to associate with RNA. Additionally, the integration of a tag into the protein could impact protein function and binding. Finally, not all protein-RNA contacts are capable of forming UV crosslinks<sup>167</sup> either because of the inherent UV bias for aromatic amino acids or due to differences in the type of RNA contacts that might occur (i.e., phosphodiester backbone), or because of the transient nature of specific interactions. Although these specific considerations do not impact our conclusions about PRC2 (because we specifically explored these aspects), these considerations are important when interpreting what the absence of RNA binding might mean when exploring other proteins.

### STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **RESOURCE AVAILABILITY**
  - Lead contact
  - Materials availability
  - Data and code availability
- **EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS**
  - Cell culture
- **METHOD DETAILS**
  - Challenges associated with distinguishing between promiscuous binding and lack of binding
  - Cloning of Halo-V5-tagged expression constructs
  - Expression, UV-crosslinking, and lysis of cells
  - Western blot of tagged proteins
  - Co-immunoprecipitation of PRC2 components and AlexaFluor labeling
  - Measurement of *in vitro* binding of PRC2 components to RNA
  - Enhanced crosslinking and immunoprecipitation
  - CLIP RNase titration
  - Possible explanations for UV- and protein-dependent non-specific associations
  - Comparison of CLIP-washes to CLAP-washes on NHS beads
  - Covalent linkage and affinity purification
  - Radiolabeling of captured RNA-protein complexes
  - Failure of CLAP to identify RNA-protein interactions identified by CLIP cannot be due to differences in assay sensitivity
  - Protein quantification by Halo-ligand
  - Elution and visualization of CLAP-purified proteins
  - Halo-tagged PRC2 components and purified PRC2 complexes associate and UV-crosslink with RNA *in vitro*
  - *In vitro* RNA crosslinking on HaloLink resin
  - *In vitro* RNA crosslinking and streptavidin-IR western blot
  - Cell line generation
  - CLAP on Spy-tagged proteins
  - Crosslinking for ChIP

- RNA-FISH + immunostaining
- Dox expression of Halo-tagged proteins in HEK293T cells for ChIP
- Expression confirmation
- CLAP on Halo-tagged SPEN
- IR-CLIP
- Library construction and sequencing
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
  - Read processing and alignment
  - Gene window enrichment calculations
  - Analysis of +tag and –tag samples
  - Plotting and visualization
  - Reproducibility of CLIP/CLAP replicates
  - Quantification of <sup>32</sup>P gels
  - Crosslink-induced truncation analysis
  - Definitions
  - Peak calling
  - Mean coverage of binding sites
  - Gene Ontology enrichment
  - Motif enrichment analysis

#### SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.molcel.2024.01.026>.

#### ACKNOWLEDGMENTS

We thank members of the Guttman lab for technical help, comments, and discussions; Tom Cech, John Rinn, and Ryan Flynn for feedback and discussions; and I.-M. Strazhnik for illustrations and S. Hiley for editing. This work was funded by grants from NIH (U01DK127420 [to M.G.] and R01DA053178 [to M.G.]) and the CZI Ben Barres Early Career Acceleration Award (to M.G.). J.K.G. is supported by NCI F30CA278005 and the University of Southern California MD/PhD program.

#### AUTHOR CONTRIBUTIONS

Conceptualization, J.K.G., M.R.B., and M.G.; methodology, M.R.B.; software, M.R.B. and M.G.; formal analysis, J.K.G., M.R.B., and M.G.; investigation, J.K.G., M.R.B., W.G.W., G.B., C.R.U., A.K.B., A.C., O.E., M.S., P.P., E.A., and V.T.; writing – original draft, J.K.G., M.R.B., and M.G.; writing – review & editing, J.K.G., M.R.B. and M.G.; visualization, J.K.G.; supervision, M.R.B. and M.G.; funding acquisition, J.K.G. and M.G.

#### DECLARATION OF INTERESTS

M.R.B., W.G.W., and M.G. are inventors on a patent covering the CLAP method.

Received: October 6, 2023

Revised: December 1, 2023

Accepted: January 30, 2024

Published: February 21, 2024

#### REFERENCES

1. Kwon, S.C., Yi, H., Eichelbaum, K., Föhr, S., Fischer, B., You, K.T., Castello, A., Krijgsvelde, J., Hentze, M.W., and Kim, V.N. (2013). The RNA-binding protein repertoire of embryonic stem cells. *Nat. Struct. Mol. Biol.* **20**, 1122–1130.
2. Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B.M., Strein, C., Davey, N.E., Humphreys, D.T., Preiss, T., Steinmetz, L.M., et al. (2012). Insights into RNA Biology from an Atlas of Mammalian mRNA-Binding Proteins. *Cell* **149**, 1393–1406.
3. Castello, A., Hentze, M.W., and Preiss, T. (2015). Metabolic Enzymes Enjoying New Partnerships as RNA-Binding Proteins. *Trends Endocrinol. Metab.* **26**, 746–757.
4. He, C., Sidoli, S., Warneford-Thomson, R., Tatomer, D.C., Wilusz, J.E., Garcia, B.A., and Bonasio, R. (2016). High-Resolution Mapping of RNA-Binding Regions in the Nuclear Proteome of Embryonic Stem Cells. *Mol. Cell* **64**, 416–430.
5. Davidovich, C., Zheng, L., Goodrich, K.J., and Cech, T.R. (2013). Promiscuous RNA binding by Polycomb repressive complex 2. *Nat. Struct. Mol. Biol.* **20**, 1250–1257.
6. Sigova, A.A., Abraham, B.J., Ji, X., Molin, B., Hannett, N.M., Guo, Y.E., Jangi, M., Giallourakis, C.C., Sharp, P.A., and Young, R.A. (2015). Transcription factor trapping by RNA in gene regulatory elements. *Science* **350**, 978–981.
7. Kung, J.T., Kesner, B., An, J.Y., Ahn, J.Y., Cifuentes-Rojas, C., Cognigni, D., Jeon, Y., Szanto, A., del Rosario, B.C., Pinter, S.F., et al. (2015). Locus-Specific Targeting to the X Chromosome Revealed by the RNA Interactome of CTCF. *Mol. Cell* **57**, 361–375.
8. Saldaña-Meyer, R., González-Buendía, E., Guerrero, G., Narendra, V., Bonasio, R., Recillas-Targa, F., and Reinberg, D. (2014). CTCF regulates the human p53 gene through direct interaction with its natural antisense transcript, *Wrap53*. *Genes Dev.* **28**, 723–734.
9. Zhao, J., Sun, B.K., Erwin, J.A., Song, J.J., and Lee, J.T. (2008). Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* **322**, 750–756.
10. Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Brugmann, S.A., Goodnough, L.H., Helms, J.A., Farnham, P.J., Segal, E., and Chnag, H.Y. (2007). Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* **129**, 1311–1323.
11. Rinn, J.L., and Chang, H.Y. (2012). Genome regulation by long noncoding RNAs. *Annu. Rev. Biochem.* **81**, 145–166.
12. Pandey, R.R., Mondal, T., Mohammad, F., Enroth, S., Redrup, L., Komorowski, J., Nagano, T., Mancini-Dinardo, D., and Kanduri, C. (2008). Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Mol. Cell* **32**, 232–246.
13. Kaneko, S., Li, G., Son, J., Xu, C.F., Margueron, R., Neubert, T.A., and Reinberg, D. (2010). Phosphorylation of the PRC2 component Ezh2 is cell cycle-regulated and up-regulates its binding to ncRNA. *Genes Dev.* **24**, 2615–2620.
14. Jeon, Y., and Lee, J.T. (2011). YY1 Tethers Xist RNA to the Inactive X Nucleation Center. *Cell* **146**, 119–133.
15. Zhao, J., Ohsumi, T.K., Kung, J.T., Ogawa, Y., Grau, D.J., Sarma, K., Song, J.J., Kingston, R.E., Borowsky, M., and Lee, J.T. (2010). Genome-wide Identification of Polycomb-Associated RNAs by RIP-seq. *Mol. Cell* **40**, 939–953.
16. Guttman, M., Donaghey, J., Carey, B.W., Garber, M., Grenier, J.K., Munson, G., Young, G., Lucas, A.B., Ach, R., Bruhn, L., et al. (2011). lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature* **477**, 295–300.
17. Khalil, A.M., Guttman, M., Huarte, M., Garber, M., Raj, A., Rivea Morales, D., Thomas, K., Presser, A., Bernstein, B.E., van Oudenaarden, A., et al. (2009). Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc. Natl. Acad. Sci. USA* **106**, 11667–11672.
18. Margueron, R., and Reinberg, D. (2011). The Polycomb complex PRC2 and its mark in life. *Nature* **469**, 343–349.
19. Simon, J.A., and Kingston, R.E. (2009). Mechanisms of polycomb gene silencing: knowns and unknowns. *Nat. Rev. Mol. Cell Biol.* **10**, 697–708.

20. Kaneko, S., Bonasio, R., Saldaña-Meyer, R., Yoshida, T., Son, J., Nishino, K., Umezawa, A., and Reinberg, D. (2014). Interactions between JARID2 and noncoding RNAs regulate PRC2 recruitment to chromatin. *Mol. Cell* 53, 290–300.
21. Kaneko, S., Son, J., Shen, S.S., Reinberg, D., and Bonasio, R. (2013). PRC2 binds active promoters and contacts nascent RNAs in embryonic stem cells. *Nat. Struct. Mol. Biol.* 20, 1258–1264.
22. Beltran, M., Yates, C.M., Skalska, L., Dawson, M., Reis, F.P., Viiri, K., Fisher, C.L., Sibley, C.R., Foster, B.M., Bartke, T., et al. (2016). The interaction of PRC2 with RNA or chromatin is mutually antagonistic. *Genome Res.* 26, 896–907.
23. Long, Y., Hwang, T., Gooding, A.R., Goodrich, K.J., Rinn, J.L., and Cech, T.R. (2020). RNA is essential for PRC2 chromatin occupancy and function in human pluripotent stem cells. *Nat. Genet.* 52, 931–938.
24. Song, J., Gooding, A.R., Hemphill, W.O., Love, B.D., Robertson, A., Yao, L., Zon, L.I., North, T.E., Kasinath, V., and Cech, T.R. (2023). Structural basis for inactivation of PRC2 by G-quadruplex RNA. *Science* 381, 1331–1337.
25. Spitale, R.C., Tsai, M.C., and Chang, H.Y. (2011). RNA templating the epigenome: long noncoding RNAs as molecular scaffolds. *Epigenetics* 6, 539–543.
26. Lee, J.T. (2009). Lessons from X-chromosome inactivation: long ncRNA as guides and tethers to the epigenome. *Genes Dev.* 23, 1831–1842.
27. Guttman, M., and Rinn, J.L. (2012). Modular regulatory principles of large non-coding RNAs. *Nature* 482, 339–346.
28. Koziol, M.J., and Rinn, J.L. (2010). RNA traffic control of chromatin complexes. *Curr. Opin. Genet. Dev.* 20, 142–148.
29. Wang, D., Ding, L., Wang, L., Zhao, Y., Sun, Z., Karnes, R.J., Zhang, J., and Huang, H. (2015). LncRNA MALAT1 enhances oncogenic activities of EZH2 in castration-resistant prostate cancer. *Oncotarget* 6, 41045–41055.
30. Zovoilis, A., Cifuentes-Rojas, C., Chu, H.P., Hernandez, A.J., and Lee, J.T. (2016). Destabilization of B2 RNA by EZH2 Activates the Stress Response. *Cell* 167, 1788–1802.e13.
31. Di Ruscio, A., Ebraldize, A.K., Benoukraf, T., Amabile, G., Goff, L.A., Terragni, J., Figueroa, M.E., De Figueiredo Pontes, L.L., Alberich-Jorda, M., Zhang, P., et al. (2013). DNMT1-interacting RNAs block gene-specific DNA methylation. *Nature* 503, 371–376.
32. G Hendrickson, D., Kelley, D.R., Tenen, D., Bernstein, B., and Rinn, J.L. (2016). Widespread RNA binding by chromatin-associated proteins. *Genome Biol.* 17, 28.
33. Jansson-Fritzberg, L.I., Sousa, C.I., Smallegan, M.J., Song, J.J., Gooding, A.R., Kasinath, V., Rinn, J.L., and Cech, T.R. (2022). DNMT1 inhibition by pUG-fold quadruplex RNA. *RNA* 29, 346–360.
34. Rosenberg, M., Blum, R., Kesner, B., Maier, V.K., Szanto, A., and Lee, J.T. (2017). Denaturing CLIP, dCLIP, Pipeline Identifies Discrete RNA Footprints on Chromatin-Associated Proteins and Reveals that CBX7 Targets 3' UTRs to Regulate mRNA Expression. *Cell Syst.* 5, 368–385.e15.
35. Yang, Y.W., Flynn, R.A., Chen, Y., Qu, K., Wan, B., Wang, K.C., Lei, M., and Chang, H.Y. (2014). Essential role of lncRNA binding for WDR5 maintenance of active chromatin and embryonic stem cell pluripotency. *Elife* 3, e02046.
36. Holmes, Z.E., Hamilton, D.J., Hwang, T., Parsonnet, N.V., Rinn, J.L., Wuttke, D.S., and Batey, R.T. (2020). The Sox2 transcription factor binds RNA. *Nat. Commun.* 11, 1805.
37. Hou, L., Wei, Y., Lin, Y., Wang, X., Lai, Y., Yin, M., Chen, Y., Guo, X., Wu, S., Zhu, Y., et al. (2020). Concurrent binding to DNA and RNA facilitates the pluripotency reprogramming activity of Sox2. *Nucleic Acids Res.* 48, 3869–3887.
38. Oksuz, O., Henninger, J.E., Warneford-Thomson, R., Zheng, M.M., Erb, H., Vancura, A., Overholt, K.J., Hawken, S.W., Banani, S.F., Lauman, R., et al. (2023). Transcription factors interact with RNA to regulate genes. *Mol. Cell* 83, 2449–2463.e13.
39. Xiao, R., Chen, J.-Y., Liang, Z., Luo, D., Chen, G., Lu, Z.J., Chen, Y., Zhou, B., Li, H., Du, X., et al. (2019). Pervasive Chromatin-RNA Binding Protein Interactions Enable RNA-Based Regulation of Transcription. *Cell* 178, 107–121.e18.
40. Hansen, A.S., Hsieh, T.-H.S., Cattoglio, C., Pustova, I., Saldaña-Meyer, R., Reinberg, D., Darzacq, X., and Tjian, R. (2019). Distinct Classes of Chromatin Loops Revealed by Deletion of an RNA-Binding Region in CTCF. *Mol. Cell* 76, 395–411.e13.
41. Saldaña-Meyer, R., Rodriguez-Hernandez, J., Escobar, T., Nishana, M., Jácome-López, K., Nora, E.P., Bruneau, B.G., Tsigos, A., Furlan-Magaril, M., Skok, J., and Reinberg, D. (2019). RNA Interactions Are Essential for CTCF-Mediated Genome Organization. *Mol. Cell* 76, 412–422.e5.
42. Froberg, J.E., Yang, L., and Lee, J.T. (2013). Guided by RNAs: X-Inactivation as a Model for lncRNA Function. *J. Mol. Biol.* 425, 3698–3706.
43. Wutz, A. (2011). Gene silencing in X-chromosome inactivation: advances in understanding facultative heterochromatin formation. *Nat. Rev. Genet.* 12, 542–553.
44. Galupa, R., and Heard, E. (2015). X-chromosome inactivation: new insights into cis and trans regulation. *Curr. Opin. Genet. Dev.* 31, 57–66.
45. Cifuentes-Rojas, C., Hernandez, A.J., Sarma, K., and Lee, J.T. (2014). Regulatory Interactions between RNA and Polycomb Repressive Complex 2. *Mol. Cell* 55, 171–185.
46. Wutz, A., Rasmussen, T.P., and Jaenisch, R. (2002). Chromosomal silencing and localization are mediated by different domains of Xist RNA. *Nat. Genet.* 30, 167–174.
47. Lee, J.T. (2012). Epigenetic regulation by long noncoding RNAs. *Science* 338, 1435–1439.
48. Gupta, R.A., Shah, N., Wang, K.C., Kim, J., Horlings, H.M., Wong, D.J., Tsai, M.C., Hung, T., Argani, P., Rinn, J.L., et al. (2010). Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* 464, 1071–1076.
49. Terranova, R., Yokobayashi, S., Stadler, M.B., Otte, A.P., van Lohuizen, M., Orkin, S.H., and Peters, A.H.F.M. (2008). Polycomb group proteins Ezh2 and Rnf2 direct genomic contraction and imprinted repression in early mouse embryos. *Dev. Cell* 15, 668–679.
50. Kalantry, S., and Magnuson, T. (2006). The Polycomb Group Protein EED Is Dispensable for the Initiation of Random X-Chromosome Inactivation. *PLoS Genet.* 2, e66.
51. Schoeftner, S., SenGupta, A.K., Kubicek, S., Mechtler, K., Spahn, L., Koseki, H., Jenuwein, T., and Wutz, A. (2006). Recruitment of PRC1 function at the initiation of X inactivation independent of PRC2 and silencing. *EMBO J.* 25, 3110–3122.
52. Plath, K., Fang, J., Mlynarczyk-Evans, S.K., Cao, R., Worringer, K.A., Wang, H., de la Cruz, C.C., Otte, A.P., Panning, B., and Zhang, Y. (2003). Role of histone H3 lysine 27 methylation in X inactivation. *Science* 300, 131–135.
53. da Rocha, S.T., Boeva, V., Escamilla-Del-Arenal, M., Ancelin, K., Granier, C., Matias, N.R., Sanulli, S., Chow, J., Schulz, E., Picard, C., et al. (2014). Jarid2 Is Implicated in the Initial Xist-Induced Targeting of PRC2 to the Inactive X Chromosome. *Mol. Cell* 53, 301–316.
54. Kohlmaier, A., Savarese, F., Lachner, M., Martens, J., Jenuwein, T., and Wutz, A. (2004). A chromosomal memory triggered by Xist regulates histone methylation in X inactivation. *PLoS Biol.* 2, E171.
55. McHugh, C.A., Chen, C.K., Chow, A., Surka, C.F., Tran, C., McDonel, P., Pandya-Jones, A., Blanco, M., Burghard, C., Moradian, A., et al. (2015). The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature* 521, 232–236.
56. Brockdorff, N. (2013). Noncoding RNA and Polycomb recruitment. *RNA* 19, 429–442.

57. Portoso, M., Ragazzini, R., Brenčić, Ž., Moiani, A., Michaud, A., Vassilev, I., Wassef, M., Servant, N., Sargueil, B., and Margueron, R. (2017). PRC2 is dispensable for HOTAIR-mediated transcriptional repression. *EMBO J.* **36**, 981–994.
58. Davidovich, C., Wang, X., Cifuentes-Rojas, C., Goodrich, K.J., Gooding, A.R., Lee, J.T., and Cech, T.R. (2015). Toward a Consensus on the Binding Specificity and Promiscuity of PRC2 for RNA. *Mol. Cell* **57**, 552–558.
59. Wang, X., Goodrich, K.J., Gooding, A.R., Naeem, H., Archer, S., Paucek, R.D., Youmans, D.T., Cech, T.R., and Davidovich, C. (2017). Targeting of Polycomb Repressive Complex 2 to RNA by Short Repeats of Consecutive Guanines. *Mol. Cell* **65**, 1056–1067.e5.
60. Peritz, T., Zeng, F., Kannanayakal, T.J., Kilk, K., Eiríksdóttir, E., Langel, U., and Eberwine, J. (2006). Immunoprecipitation of mRNA-protein complexes. *Nat. Protoc.* **1**, 577–580.
61. Keene, J.D., Komisarow, J.M., and Friedersdorf, M.B. (2006). RIP-Chip: the isolation and identification of mRNAs, microRNAs and protein components of ribonucleoprotein complexes from cell extracts. *Nat. Protoc.* **1**, 302–307.
62. Mili, S., and Steitz, J.A. (2004). Evidence for reassociation of RNA-binding proteins after cell lysis: implications for the interpretation of immunoprecipitation analyses. *RNA* **10**, 1692–1694.
63. Darnell, R.B. (2010). HITS-CLIP: panoramic views of protein-RNA regulation in living cells. *Wiley Interdiscip. Rev. RNA* **1**, 266–286.
64. König, J., Zarnack, K., Rot, G., Curk, T., Kayikci, M., Zupan, B., Turner, D.J., Luscombe, N.M., and Ule, J. (2010). iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat. Struct. Mol. Biol.* **17**, 909–915.
65. Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jr., Jungkamp, A.C., Munschauer, M., et al. (2010). Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141**, 129–141.
66. Licatalosi, D.D., Mele, A., Fak, J.J., Ule, J., Kayikci, M., Chi, S.W., Clark, T.A., Schweitzer, A.C., Blume, J.E., Wang, X., et al. (2008). HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* **456**, 464–469.
67. Ule, J., Jensen, K.B., Ruggiu, M., Mele, A., Ule, A., and Darnell, R.B. (2003). CLIP Identifies Nova-Regulated RNA Networks in the Brain. *Science* **302**, 1212–1215.
68. Chu, C., Zhang, Q.C., da Rocha, S.T., Flynn, R.A., Bharadwaj, M., Calabrese, J.M., Magnuson, T., Heard, E., and Chang, H.Y. (2015). Systematic Discovery of Xist RNA Binding Proteins. *Cell* **161**, 404–416.
69. Minajigi, A., Froberg, J.E., Wei, C., Sunwoo, H., Kesner, B., Colognori, D., Lessing, D., Payer, B., Boukhali, M., Haas, W., and Lee, J.T. (2015). Chromosomes. A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation. *Science* **349**, aab2276.
70. Rosenberg, M., Blum, R., Kesner, B., Aeby, E., Garant, J.M., Szanto, A., and Lee, J.T. (2021). Motif-driven interactions between RNA and PRC2 are rheostats that regulate transcription elongation. *Nat. Struct. Mol. Biol.* **28**, 103–117.
71. Van Nostrand, E.L., Pratt, G.A., Shishkin, A.A., Gelboin-Burkhart, C., Fang, M.Y., Sundaraman, B., Blue, S.M., Nguyen, T.B., Surka, C., Elkins, K., et al. (2016). Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat. Methods* **13**, 508–514.
72. Kanhere, A., Viiri, K., Araújo, C.C., Rasaiyaah, J., Bouwman, R.D., Whyte, W.A., Pereira, C.F., Brookes, E., Walker, K., Bell, G.W., et al. (2010). Short RNAs are transcribed from repressed polycomb target genes and interact with polycomb repressive complex-2. *Mol. Cell* **38**, 675–688.
73. Hirata, H., Hinoda, Y., Shahryari, V., Deng, G., Nakajima, K., Tabatabai, Z.L., Ishii, N., and Dahiya, R. (2015). Long Noncoding RNA MALAT1 Promotes Aggressive Renal Cell Carcinoma through Ezh2 and Interacts with miR-205. *Cancer Res.* **75**, 1322–1331.
74. Huang, X., Bashkenova, N., Hong, Y., Lyu, C., Guallar, D., Hu, Z., Malik, V., Li, D., Wang, H., Shen, X., et al. (2022). A TET1-PSPC1-Neat1 molecular axis modulates PRC2 functions in controlling stem cell bivalency. *Cell Rep.* **39**, 110928.
75. Tichon, A., Gil, N., Lubelsky, Y., Havkin Solomon, T., Lemze, D., Itzkovitz, S., Stern-Ginossar, N., and Ulitsky, I. (2016). A conserved abundant cytoplasmic long noncoding RNA modulates repression by Pumilio proteins in human cells. *Nat. Commun.* **7**, 12209.
76. Lee, S., Kopp, F., Chang, T.C., Sataluri, A., Chen, B., Sivakumar, S., Yu, H., Xie, Y., and Mendell, J.T. (2016). Noncoding RNA NORAD Regulates Genomic Stability by Sequestering PUMILIO Proteins. *Cell* **164**, 69–80.
77. Banerjee, A.K., Blanco, M.R., Bruce, E.A., Honson, D.D., Chen, L.M., Chow, A., Bhat, P., Ollikainen, N., Quinodoz, S.A., Loney, C., et al. (2020). SARS-CoV-2 Disrupts Splicing, Translation, and Protein Trafficking to Suppress Host Defenses. *Cell* **183**, 1325–1339.e21.
78. Jachowicz, J.W., Strehle, M., Banerjee, A.K., Blanco, M.R., Thai, J., and Guttman, M. (2022). Xist spatially amplifies SHARP/SPEN recruitment to balance chromosome-wide silencing and specificity to the X chromosome. *Nat. Struct. Mol. Biol.* **29**, 239–249.
79. Gu, J., Wang, M., Yang, Y., Qiu, D., Zhang, Y., Ma, J., Zhou, Y., Hannon, G.J., and Yu, Y. (2018). GoldCLIP: Gel-omitted Ligation-dependent CLIP. *Genomics Proteomics Bioinformatics* **16**, 136–143.
80. Li, X., Pritykin, Y., Concepcion, C.P., Lu, Y., La Rocca, G., Zhang, M., King, B., Cook, P.J., Au, Y.W., Popow, O., et al. (2020). High-Resolution In Vivo Identification of miRNA Targets by Halo-Enhanced Ago2 Pull-Down. *Mol. Cell* **79**, 167–179.e11.
81. Maticzka, D., Ilik, I.A., Aktas, T., Backofen, R., and Akhtar, A. (2018). uvCLAP is a fast and non-radioactive method to identify in vivo targets of RNA-binding proteins. *Nat. Commun.* **9**, 1142.
82. Los, G.V., Encell, L.P., McDougall, M.G., Hartzell, D.D., Karassina, N., Zimprich, C., Wood, M.G., Learish, R., Ohana, R.F., Urh, M., et al. (2008). HaloTag: A Novel Protein Labeling Technology for Cell Imaging and Protein Analysis. *ACS Chem. Biol.* **3**, 373–382.
83. Zakeri, B., Fierer, J.O., Celik, E., Chittock, E.C., Schwarz-Linek, U., Moy, V.T., and Howarth, M. (2012). Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin. *Proc. Natl. Acad. Sci. USA* **109**, E690–E697.
84. Ray, D., Kazan, H., Chan, E.T., Peña Castillo, L.P., Chaudhry, S., Talukder, S., Blencowe, B.J., Morris, Q., and Hughes, T.R. (2009). Rapid and systematic analysis of the RNA recognition specificities of RNA-binding proteins. *Nat. Biotechnol.* **27**, 667–670.
85. Ozdilek, B.A., Thompson, V.F., Ahmed, N.S., White, C.I., Batey, R.T., and Schwartz, J.C. (2017). Intrinsically disordered RGG/RG domains mediate degenerate specificity in RNA binding. *Nucleic Acids Res.* **45**, 7984–7996.
86. Conrad, T., Albrecht, A.S., de Melo Costa, V.R., Sauer, S., Meierhofer, D., and Ørom, U.A. (2016). Serial interactome capture of the human cell nucleus. *Nat. Commun.* **7**, 11212.
87. Castello, A., Fischer, B., Frese, C.K., Horos, R., Alleaume, A.M., Foehr, S., Curk, T., Krijgsvel, J., and Hentze, M.W. (2016). Comprehensive Identification of RNA-Binding Domains in Human Cells. *Mol. Cell* **63**, 696–710.
88. Brannan, K.W., Jin, W., Huelga, S.C., Banks, C.A.S., Gilmore, J.M., Florens, L., Washburn, M.P., Van Nostrand, E.L., Pratt, G.A., Schwin, M.K., et al. (2016). SONAR Discovers RNA-Binding Proteins from Analysis of Large-Scale Protein-Protein Interactomes. *Mol. Cell* **64**, 282–293.
89. Trendel, J., Schwarzl, T., Horos, R., Prakash, A., Bateman, A., Hentze, M.W., and Krijgsvel, J. (2019). The Human RNA-Binding Proteome and Its Dynamics during Translational Arrest. *Cell* **176**, 391–403.e19.



90. Wang, K.C., Yang, Y.W., Liu, B., Sanyal, A., Corces-Zimmerman, R., Chen, Y., Lajoie, B.R., Protacio, A., Flynn, R.A., Gupta, R.A., et al. (2011). A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* 472, 120–124.
91. Mumbach, M.R., Granja, J.M., Flynn, R.A., Roake, C.M., Satpathy, A.T., Rubin, A.J., Qi, Y., Jiang, Z., Shams, S., Louie, B.H., et al. (2019). HiChIRP reveals RNA-associated chromosome conformation. *Nat. Methods* 16, 489–492.
92. Sanborn, A.L., Rao, S.S.P., Huang, S.C., Durand, N.C., Huntley, M.H., Jewett, A.I., Bochkov, I.D., Chinnappan, D., Cutkosky, A., Li, J., et al. (2015). Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc. Natl. Acad. Sci. USA* 112, E6456–E6465.
93. Fudenberg, G., Imakaev, M., Lu, C., Goloborodko, A., Abdennur, N., and Mirny, L.A. (2016). Formation of Chromosomal Domains by Loop Extrusion. *Cell Rep.* 15, 2038–2049.
94. Oh, H.J., Aguilar, R., Kesner, B., Lee, H.G., Kriz, A.J., Chu, H.P., and Lee, J.T. (2021). Jpx RNA regulates CTCF anchor site selection and formation of chromosome loops. *Cell* 184, 6157–6173.e24.
95. Weintraub, A.S., Li, C.H., Zamudio, A.V., Sigova, A.A., Hannett, N.M., Day, D.S., Abraham, B.J., Cohen, M.A., Nabet, B., Buckley, D.L., et al. (2017). YY1 Is a Structural Regulator of Enhancer-Promoter Loops. *Cell* 171, 1573–1588.e28.
96. Li, Y., Han, J., Zhang, Y., Cao, F., Liu, Z., Li, S., Wu, J., Hu, C., Wang, Y., Shuai, J., et al. (2016). Structural basis for activity regulation of MLL family methyltransferases. *Nature* 530, 447–452.
97. Caudron-Herger, M., Jansen, R.E., Wassmer, E., and Diederichs, S. (2021). RBP2GO: a comprehensive pan-species database on RNA-binding proteins, their interactions and functions. *Nucleic Acids Res.* 49, D425–D436.
98. Caudron-Herger, M., Rusin, S.F., Adamo, M.E., Seiler, J., Schmid, V.K., Barreau, E., Kettenbach, A.N., and Diederichs, S. (2019). R-DeeP: Proteome-wide and Quantitative Identification of RNA-Dependent Proteins by Density Gradient Ultracentrifugation. *Mol. Cell* 75, 184–199.e10.
99. Mallam, A.L., Sae-Lee, W., Schaub, J.M., Tu, F., Battenhouse, A., Jang, Y.J., Kim, J., Wallingford, J.B., Finkelstein, I.J., Marcotte, E.M., and Drew, K. (2019). Systematic Discovery of Endogenous Human Ribonucleoprotein Complexes. *Cell Rep.* 29, 1351–1368.e5.
100. Harris, M.A., Clark, J., Ireland, A., Lomax, J., Ashburner, M., Foulger, R., Eilbeck, K., Lewis, S., Marshall, B., Mungall, C., et al. (2004). The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res.* 32, D258–D261.
101. Quinodoz, S.A., Jachowicz, J.W., Bhat, P., Ollikainen, N., Banerjee, A.K., Goronzy, I.N., Blanco, M.R., Chovanec, P., Chow, A., Markaki, Y., et al. (2021). RNA promotes the formation of spatial compartments in the nucleus. *Cell* 184, 5775–5790.e30.
102. Moindrot, B., Cerase, A., Coker, H., Masui, O., Grijzenhout, A., Pintacuda, G., Schermelleh, L., Nesterova, T.B., and Brockdorff, N. (2015). A Pooled shRNA Screen Identifies Rbm15, Spen, and Wtap as Factors Required for Xist RNA-Mediated Silencing. *Cell Rep.* 12, 562–572.
103. Monfort, A., Di Minin, G., Postlmayr, A., Freimann, R., Arieti, F., Thore, S., and Wutz, A. (2015). Identification of Spen as a Crucial Factor for Xist Function through Forward Genetic Screening in Haploid Embryonic Stem Cells. *Cell Rep.* 12, 554–561.
104. Lu, Z., Guo, J.K., Wei, Y., Dou, D.R., Zarnegar, B., Ma, Q., Li, R., Zhao, Y., Liu, F., Choudhry, H., et al. (2020). Structural modularity of the XIST ribonucleoprotein complex. *Nat. Commun.* 11, 6163.
105. Carter, A.C., Xu, J., Nakamoto, M.Y., Wei, Y., Zarnegar, B.J., Shi, Q., Broughton, J.P., Ransom, R.C., Salhotra, A., Nagaraja, S.D., et al. (2020). Spen links RNA-mediated endogenous retrovirus silencing and X chromosome inactivation. *Elife* 9, e54508.
106. Chen, C.K., Blanco, M., Jackson, C., Aznauryan, E., Ollikainen, N., Surka, C., Chow, A., Cerase, A., McDonel, P., and Guttman, M. (2016). Xist recruits the X chromosome to the nuclear lamina to enable chromosome-wide silencing. *Science* 54, 468–472.
107. Kanduri, C. (2011). Kcnq1ot1: A chromatin regulatory RNA. *Semin. Cell Dev. Biol.* 22, 343–350.
108. Nagano, T., and Fraser, P. (2009). Emerging similarities in epigenetic gene silencing by long noncoding RNAs. *Mamm. Genome* 20, 557–562.
109. Rom, A., Melamed, L., Gil, N., Goldrich, M.J., Kadir, R., Golan, M., Biton, I., Perry, R.B.-T., and Ulitsky, I. (2019). Regulation of CHD2 expression by the Chaserr long noncoding RNA gene is essential for viability. *Nat. Commun.* 10, 5092.
110. Schwartz, J.C., Cech, T.R., and Parker, R.R. (2015). Biochemical Properties and Biological Functions of FET Proteins. *Annu. Rev. Biochem.* 84, 355–379.
111. Bertolotti, A., Lutz, Y., Heard, D.J., Chambon, P., and Tora, L. (1996). hTAF(II)68, a novel RNA/ssDNA-binding protein with homology to the pro-oncoproteins TLS/FUS and EWS is associated with both TFIID and RNA polymerase II. *EMBO J.* 15, 5022–5031.
112. Bertolotti, A., Melot, T., Acker, J., Vigneron, M., Delattre, O., and Tora, L. (1998). EWS, but Not EWS-FLI-1, Is Associated with Both TFIID and RNA Polymerase II: Interactions between Two Members of the TET Family, EWS and hTAF II 68, and Subunits of TFIID and RNA Polymerase II Complexes. *Mol. Cell Biol.* 18, 1489–1497.
113. Ahmed, N.S., Harrell, L.M., Wieland, D.R., Lay, M.A., Thompson, V.F., and Schwartz, J.C. (2021). Fusion protein EWS-FLI1 is incorporated into a protein granule in cells. *RNA* 27, 920–932.
114. Rossow, K.L., and Janknecht, R. (2001). The Ewing's sarcoma gene product functions as a transcriptional activator. *Cancer Res.* 61, 2690–2695.
115. Ouyang, H., Zhang, K., Fox-Walsh, K., Yang, Y., Zhang, C., Huang, J., Li, H., Zhou, Y., and Fu, X.D. (2017). The RNA binding protein EWS is broadly involved in the regulation of pri-miRNA processing in mammalian cells. *Nucleic Acids Res.* 45, 12481–12495.
116. Van Nostrand, E.L., Freese, P., Pratt, G.A., Wang, X., Wei, X., Xiao, R., Blue, S.M., Chen, J.Y., Cody, N.A.L., Dominguez, D., et al. (2020). A large-scale binding and functional map of human RNA-binding proteins. *Nature* 583, 711–719.
117. Van Nostrand, E.L., Pratt, G.A., Yee, B.A., Wheeler, E.C., Blue, S.M., Mueller, J., Park, S.S., Garcia, K.E., Gelboin-Burkhart, C., Nguyen, T.B., et al. (2020). Principles of RNA processing from analysis of enhanced CLIP maps for 150 RNA binding proteins. *Genome Biol.* 21, 90.
118. Schwartz, J.C., Wang, X., Podell, E.R., and Cech, T.R. (2013). RNA Seeds Higher-Order Assembly of FUS Protein. *Cell Rep.* 5, 918–925.
119. Kwon, I., Kato, M., Xiang, S., Wu, L., Theodoropoulos, P., Mirzaei, H., Han, T., Xie, S., Corden, J.L., and McKnight, S.L. (2013). Phosphorylation-Regulated Binding of RNA Polymerase II to Fibrillar Polymers of Low-Complexity Domains. *Cell* 155, 1049–1060.
120. Shao, W., Bi, X., Pan, Y., Gao, B., Wu, J., Yin, Y., Liu, Z., Peng, M., Zhang, W., Jiang, X., et al. (2022). Phase separation of RNA-binding protein promotes polymerase binding and transcription. *Nat. Chem. Biol.* 18, 70–80.
121. Spitzer, J.I., Ugras, S., Runge, S., Decarolis, P., Antonescu, C., Tuschl, T., and Singer, S. (2011). mRNA and protein levels of FUS, EWSR1, and TAF15 are upregulated in liposarcoma. *Genes Chromosomes Cancer* 50, 338–347.
122. Luo, Y., Blechinger, J., Fernandes, A.M., Li, S., Fryland, T., Børglum, A.D., Bolund, L., and Nielsen, A.L. (2015). EWS and FUS bind a subset of transcribed genes encoding proteins enriched in RNA regulatory functions. *BMC Genomics* 16, 929.
123. Takai, H., Masuda, K., Sato, T., Sakaguchi, Y., Suzuki, T., Suzuki, T., Koyama-Nasu, R., Nasu-Nishimura, Y., Katou, Y., Ogawa, H., et al. (2014). 5-Hydroxymethylcytosine Plays a Critical



- Role in Glioblastomagenesis by Recruiting the CHTOP-Methylosome Complex. *Cell Rep.* 9, 48–60.
124. Zhao, Q., Rank, G., Tan, Y.T., Li, H., Moritz, R.L., Simpson, R.J., Cerruti, L., Curtis, D.J., Patel, D.J., Allis, C.D., et al. (2009). PRMT5-mediated methylation of histone H4R3 recruits DNMT3A, coupling histone and DNA methylation in gene silencing. *Nat. Struct. Mol. Biol.* 16, 304–311.
  125. Viphakone, N., Sudbery, I., Griffith, L., Heath, C.G., Sims, D., and Wilson, S.A. (2019). Co-transcriptional Loading of RNA Export Factors Shapes the Human Transcriptome. *Mol. Cell* 75, 310–323.e8.
  126. Chang, C.T., Hautbergue, G.M., Walsh, M.J., Viphakone, N., van Dijk, T.B., Philipsen, S., and Wilson, S.A. (2013). Chtop is a component of the dynamic TREX mRNA export complex. *EMBO J.* 32, 473–486.
  127. Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L., and Rao, A. (2009). Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. *Science* 324, 930–935.
  128. Yildirim, O., Li, R., Hung, J.H., Chen, P.B., Dong, X., Ee, L.S., Weng, Z., Rando, O.J., and Fazio, T.G. (2011). Mbd3/NURD Complex Regulates Expression of 5-Hydroxymethylcytosine Marked Genes in Embryonic Stem Cells. *Cell* 147, 1498–1510.
  129. Pühringer, T., Hohmann, U., Fin, L., Pacheco-Fiallos, B., Schellhaas, U., Brennecke, J., and Plaschka, C. (2020). Structure of the human core transcription-export complex reveals a hub for multivalent interactions. *Elife* 9, e61503.
  130. Shi, M., Zhang, H., Wu, X., He, Z., Wang, L., Yin, S., Tian, B., Li, G., and Cheng, H. (2017). ALYREF mainly binds to the 5' and the 3' regions of the mRNA *in vivo*. *Nucleic Acids Res.* 45, 9640–9653.
  131. Cheng, H., Dufu, K., Lee, C.S., Hsu, J.L., Dias, A., and Reed, R. (2006). Human mRNA Export Machinery Recruited to the 5' End of mRNA. *Cell* 127, 1389–1400.
  132. Hirose, T., Vimicchi, G., Tanigawa, A., Naganuma, T., Li, R., Kimura, H., Yokoi, T., Nakagawa, S., Bénard, M., Fox, A.H., and Pierron, G. (2014). NEAT1 long noncoding RNA regulates transcription via protein sequestration within subnuclear bodies. *Mol. Biol. Cell* 25, 169–183.
  133. Knott, G.J., Bond, C.S., and Fox, A.H. (2016). The DBHS proteins SFPQ, NONO and PSPC1: a multipurpose molecular scaffold. *Nucleic Acids Res.* 44, 3989–4004.
  134. Guallar, D., Bi, X., Pardavila, J.A., Huang, X., Saenz, C., Shi, X., Zhou, H., Faiola, F., Ding, J., Haruehanroengra, P., et al. (2018). RNA-dependent chromatin targeting of TET2 for endogenous retrovirus control in pluripotent stem cells. *Nat. Genet.* 50, 443–451.
  135. Clemson, C.M., Hutchinson, J.N., Sara, S.A., Ensminger, A.W., Fox, A.H., Chess, A., and Lawrence, J.B. (2009). An Architectural Role for a Nuclear Noncoding RNA: NEAT1 RNA Is Essential for the Structure of Paraspeckles. *Mol. Cell* 33, 717–726.
  136. Wang, J., Rajbhandari, P., Damianov, A., Han, A., Sallam, T., Waki, H., Villanueva, C.J., Lee, S.D., Nielsen, R., Mandrup, S., et al. (2017). RNA-binding protein PSPC1 promotes the differentiation-dependent nuclear export of adipocyte RNAs. *J. Clin. Invest.* 127, 987–1004.
  137. Tan, L., and Shi, Y.G. (2012). Tet family proteins and 5-hydroxymethylcytosine in development and disease. *Development* 139, 1895–1902.
  138. Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C., and Zhang, Y. (2011). Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine. *Science* 333, 1300–1303.
  139. Chen, Q., Chen, Y., Bian, C., Fujiki, R., and Yu, X. (2013). TET2 promotes histone O-GlcNAcylation during gene transcription. *Nature* 493, 561–564.
  140. Deplus, R., Delatte, B., Schwinn, M.K., Defrance, M., Méndez, J., Murphy, N., Dawson, M.A., Volkmar, M., Putmans, P., Calonne, E., et al. (2013). TET2 and TET3 regulate GlcNAcylation and H3K4 methylation through OGT and SET1/COMPASS. *EMBO J.* 32, 645–655.
  141. Peng, L., Li, Y., Xi, Y., Li, W., Li, J., Lv, R., Zhang, L., Zou, Q., Dong, S., Luo, H., et al. (2016). MBD3L2 promotes Tet2 enzymatic activity for mediating 5-methylcytosine oxidation. *J. Cell Sci.* 129, 1059–1071.
  142. Wang, L., Ozark, P.A., Smith, E.R., Zhao, Z., Marshall, S.A., Rendleman, E.J., Piunti, A., Ryan, C., Whelan, A.L., Helmin, K.A., et al. (2018). TET2 coactivates gene expression through demethylation of enhancers. *Sci. Adv.* 4, eaa06986.
  143. Wang, X., Rosikiewicz, W., Sedkov, Y., Martinez, T., Hansen, B.S., Schreiner, P., Christensen, J., Xu, B., Pruett-Miller, S.M., Helin, K., and Herz, H.M. (2022). PROSER1 mediates TET2 O-GlcNAcylation to regulate DNA demethylation on UTX-dependent enhancers and CpG islands. *Life Sci. Alliance* 5, e202101228.
  144. Ko, M., An, J., Bandukwala, H.S., Chavez, L., Åijö, T., Pastor, W.A., Segal, M.F., Li, H., Koh, K.P., Lähdesmäki, H., et al. (2013). Modulation of TET2 expression and 5-methylcytosine oxidation by the CXXC domain protein IDAX. *Nature* 497, 122–126.
  145. He, C., Bozler, J., Janssen, K.A., Wilusz, J.E., Garcia, B.A., Schorn, A.J., and Bonasio, R. (2021). TET2 chemically modifies tRNAs and regulates tRNA fragment levels. *Nat. Struct. Mol. Biol.* 28, 62–70.
  146. Lan, J., Rajan, N., Bizet, M., Penning, A., Singh, N.K., Guallar, D., Calonne, E., Li Greci, A., Bonvin, E., Deplus, R., et al. (2020). Functional role of Tet-mediated RNA hydroxymethylcytosine in mouse ES cells and during differentiation. *Nat. Commun.* 11, 4956.
  147. Quinodoz, S.A., Ollikainen, N., Tabak, B., Palla, A., Schmidt, J.M., Detmar, E., Lai, M.M., Shishkin, A.A., Bhat, P., Takei, Y., et al. (2018). Higher-Order Inter-chromosomal Hubs Shape 3D Genome Organization in the Nucleus. *Cell* 174, 744–757.e24.
  148. Kundu, S., Ji, F., Sunwoo, H., Jain, G., Lee, J.T., Sadreyev, R.I., Dekker, J., and Kingston, R.E. (2017). Polycomb Repressive Complex 1 Generates Discrete Compacted Domains that Change during Differentiation. *Mol. Cell* 65, 432–446.e5.
  149. Cheutin, T., and Cavalli, G. (2014). Polycomb silencing: from linear chromatin domains to 3D chromosome folding. *Curr. Opin. Genet. Dev.* 25, 30–37.
  150. Bantignies, F., Roure, V., Comet, I., Leblanc, B., Schuettengruber, B., Bonnet, J., Tixier, V., Mas, A., and Cavalli, G. (2011). Polycomb-dependent regulatory contacts between distant Hox loci in *Drosophila*. *Cell* 144, 214–226.
  151. Zhang, Q., McKenzie, N.J., Warneford-Thomson, R., Gail, E.H., Flanigan, S.F., Owen, B.M., Lauman, R., Levina, V., Garcia, B.A., Schittenhelm, R.B., et al. (2019). RNA exploits an exposed regulatory site to inhibit the enzymatic activity of PRC2. *Nat. Struct. Mol. Biol.* 26, 237–247.
  152. Cherney, R.E., Mills, C.A., Herring, L.E., Bracer, A.K., and Calabrese, J.M. (2023). A monoclonal antibody raised against human EZH2 cross-reacts with the RNA-binding protein SAFB. *Biol. Open* 12, bio059955.
  153. Huo, X., Ji, L., Zhang, Y., Lv, P., Cao, X., Wang, Q., Yan, Z., Dong, S., Du, D., Zhang, F., et al. (2020). The Nuclear Matrix Protein SAFB Cooperates with Major Satellite RNAs to Stabilize Heterochromatin Architecture Partially through Phase Separation. *Mol. Cell* 77, 368–383.e7.
  154. Healy, E., Zhang, Q., Gail, E.H., Agius, S.C., Sun, G., Bullen, M., Pandey, V., Das, P.P., Polo, J.M., and Davidovich, C. (2023). The apparent loss of PRC2 chromatin occupancy as an artefact of RNA depletion. Preprint at bioRxiv. <https://doi.org/10.1101/2023.08.16.553488>.
  155. Hickman, A.H., and Jenner, R.G. (2023). Apparent RNA bridging between PRC2 and chromatin is an artefact of non-specific chromatin precipitation upon RNA degradation. Preprint at bioRxiv. <https://doi.org/10.1101/2023.08.16.553503>.
  156. Guo, J.K., and Guttman, M. (2022). Regulatory non-coding RNAs: everything is possible, but what is important? *Nat. Methods* 19, 1156–1159.
  157. Yeo, G.W., Coufal, N.G., Liang, T.Y., Peng, G.E., Fu, X.D., and Gage, F.H. (2009). An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells. *Nat. Struct. Mol. Biol.* 16, 130–137.

158. Sundararaman, B., Zhan, L., Blue, S.M., Stanton, R., Elkins, K., Olson, S., Wei, X., Van Nostrand, E.L., Pratt, G.A., Huelga, S.C., et al. (2016). Resources for the Comprehensive Discovery of Functional RNA Elements. *Mol. Cell* 61, 903–913.
159. Hafner, M., Katsantoni, M., Köster, T., Marks, J., Mukherjee, J., Staiger, D., Ule, J., and Zavolan, M. (2021). CLIP and complementary methods. *Nat. Rev. Methods Primers* 1, 20.
160. Friedersdorf, M.B., and Keene, J.D. (2014). Advancing the functional utility of PAR-CLIP by quantifying background binding to mRNAs and lncRNAs. *Genome Biol.* 15, R2.
161. Wolin, E., Guo, J.K., Blanco, M.R., Perez, A.A., Goronzy, I.N., Abdou, A.A., Gorhe, D., Guttman, M., and Jovanovic, M. (2023). SPIDR: a highly multiplexed method for mapping RNA-protein interactions uncovers a potential mechanism for selective translational suppression upon cellular stress. Preprint at bioRxiv. <https://doi.org/10.1101/2023.06.05.543769>.
162. McMahon, A.C., Rahman, R., Jin, H., Shen, J.L., Fieldsend, A., Luo, W., and Rosbash, M. (2016). TRIBE: Hijacking an RNA-Editing Enzyme to Identify Cell-Specific Targets of RNA-Binding Proteins. *Cell* 165, 742–753.
163. Fazal, F.M., Han, S., Parker, K.R., Kaewsapsak, P., Xu, J., Boettiger, A.N., Chang, H.Y., and Ting, A.Y. (2019). Atlas of Subcellular RNA Localization Revealed by APEX-Seq. *Cell* 178, 473–490.e26.
164. Padrón, A., Iwasaki, S., and Ingolia, N.T. (2019). Proximity RNA Labeling by APEX-Seq Reveals the Organization of Translation Initiation Complexes and Repressive RNA Granules. *Mol. Cell* 75, 875–887.e5.
165. Kaewsapsak, P., Shechner, D.M., Mallard, W., Rinn, J.L., and Ting, A.Y. (2017). Live-cell mapping of organelle-associated RNAs via proximity biotinylation combined with protein-RNA crosslinking. *Elife* 6, e29224.
166. Khyzha, N., Henikoff, S., and Ahmad, K. (2022). Profiling RNA at chromatin targets *in situ* by antibody-targeted tagmentation. *Nat. Methods* 19, 1383–1392.
167. Wheeler, E.C., Van Nostrand, E.L., and Yeo, G.W. (2018). Advances and challenges in the detection of transcriptome-wide protein–RNA interactions. *Wiley Interdiscip. Rev. RNA* 9, e1436.
168. Engreitz, J.M., Pandya-Jones, A., McDonel, P., Shishkin, A., Sirokman, K., Surka, C., Kadri, S., Xing, J., Goren, A., Lander, E.S., et al. (2013). The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science* 341, 1237973.
169. Yang, X., Boehm, J.S., Yang, X., Salehi-Ashtiani, K., Hao, T., Shen, Y., Lubonja, R., Thomas, S.R., Alkan, O., Bhimdi, T., et al. (2011). A public genome-scale lentiviral expression library of human ORFs. *Nat. Methods* 8, 659–661.
170. Wang, Y., and Zhang, Y. (2014). Regulation of TET Protein Stability by Calpains. *Cell Rep.* 6, 278–284.
171. Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. *Nat. Biotechnol.* 29, 24–26.
172. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21.
173. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359.
174. Uyar, B., Yusuf, D., Wurmus, R., Rajewsky, N., Ohler, U., and Akalin, A. (2017). RCAS: an RNA centric annotation system for transcriptome-wide regions of interest. *Nucleic Acids Res.* 45, e91.
175. Grant, C.E., Bailey, T.L., and Noble, W.S. (2011). FIMO: scanning for occurrences of a given motif. *Bioinformatics* 27, 1017–1018.
176. Rueden, C.T., Schindelin, J., Hiner, M.C., DeZonia, B.E., Walter, A.E., Arena, E.T., and Eliceiri, K.W. (2017). ImageJ2: ImageJ for the next generation of scientific image data. *BMC Bioinformatics* 18, 529.
177. Porter, D.F., Garg, R.M., Meyers, R.M., Miao, W., Ducoli, L., Zarnegar, B.J., and Khavari, P.A. (2023). Analyzing RNA-Protein Interactions by Cross-Link Rates and CLIP-seq Libraries. *Curr. Protoc.* 3, e659.
178. Porter, D.F., Miao, W., Yang, X., Goda, G.A., Ji, A.L., Donohue, L.K.H., Aleman, M.M., Dominguez, D., and Khavari, P.A. (2021). easyCLIP analysis of RNA-protein interactions incorporating absolute quantification. *Nat. Commun.* 12, 1569.
179. Gagnon, K.T., Li, L., Janowski, B.A., and Corey, D.R. (2014). Analysis of nuclear RNA interference in human cells by subcellular fractionation and Argonaute loading. *Nat. Protoc.* 9, 2045–2060.
180. Cirillo, D., Blanco, M., Armaos, A., Buness, A., Avner, P., Guttman, M., Cerase, A., and Tartaglia, G.G. (2016). Quantitative predictions of protein interactions with long noncoding RNAs. *Nat. Methods* 14, 5–6.

## STAR★METHODS

### KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Antibodies</b>		
Goat anti-V5 antibody	Bethyl	Cat# A190-119A; RRID:AB_67317
Rabbit anti-V5 antibody	Bethyl	Cat# A190-120A; RRID:AB_67586
Anti-HaloTag antibody	Promega	Cat# G9211; RRID:AB_2688011
Anti-EED antibody	Cell Signaling Technology	Cat# 85322; RRID:AB_2923355
Anti-EZH2 antibody	Cell Signaling Technology	Cat# 5246; RRID:AB_10694683
Anti-SUZ12 antibody	Cell Signaling Technology	Cat# 3737; RRID:AB_2196850
Anti-PTBP1 antibody	Cell Signaling Technology	Cat# 72669; RRID:AB_2799826
Goat anti-Rabbit IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor 568	Invitrogen	Cat# A-11011; RRID: AB_143157
Anti-FLAG mouse monoclonal antibody	Sigma	Cat# F3165; RRID:AB_259529
V5 rabbit polyclonal antibody	Santa Cruz Biotechnology	Cat# sc-83849-R; RRID:AB_2019669
IRDye 800CW Goat anti-Rabbit IgG	LI-COR Biosciences	Cat# 925-32211; RRID:AB_2651127
IRDye 680DR Goat anti-Mouse IgG	LI-COR Biosciences	Cat# 925-68070; RRID:AB_2651128
IRDye 800CW Streptavidin	LI-COR Biosciences	Cat# 925-32230
Normal Rabbit IgG	Cell Signaling Technology	Cat# 2729; RRID:AB_1031062
Anti-EZH2 antibody (pAb)	Active Motif	Cat# 39933; RRID:AB_2793397
Anti-Suz12 antibody (pAb)	Active Motif	Cat# 39357; RRID:AB_2614929
Anti-H3K27me3 antibody	Active Motif	Cat# 39155; RRID:AB_2561020
<b>Bacterial and virus strains</b>		
NEB 10-beta Competent <i>E. coli</i> (High Efficiency)	NEB	Cat# C3019H
BL21(DE3)-Gold competent cells	Agilent	Cat# 230132
<b>Chemicals, peptides, and recombinant proteins</b>		
HaloTag Alexa Fluor 660 Ligand	Promega	Cat# G8471
HaloLink Resin	Promega	Cat# G1915
Doxycycline	Sigma	Cat# D9891
16% Formaldehyde (w/v), Methanol-free	Pierce	Cat# 28908
Disuccinimidyl glutarate (DSG)	Thermo Fisher	Cat# 20593
Recombinant EZH2 protein complex	Active Motif	Cat# 31337
M-PER Mammalian Protein Extraction Reagent	Thermo Fisher	Cat# 78501
pCp-IR680LT	Jena Bioscience	Cat# NU-1706-IR680LT
<b>Critical commercial assays</b>		
BioT transfection reagent	Bioland	Cat# B01-03
NEBNext Ultra II End Repair/dA-Tailing Module	NEB	Cat# E7546L
T7 RiboMAX Express Large Scale RNA Production System	Promega	Cat# P1320
<b>Deposited data</b>		
CLIP- and CLAP-seq data	This study	GEO: GSE253477
ChIP-seq data	This study	GEO: GSE253477
Processed CLIP and CLAP enrichments and full-length <sup>32</sup> P gel images	This study	Mendeley Data: <a href="https://doi.org/10.17632/wmsbzbv6kg5.2">https://doi.org/10.17632/wmsbzbv6kg5.2</a>
<b>Experimental models: Cell lines</b>		
Human: HEK293T cell line	ATCC	Cat. # CRL-3216
Mouse: pSM33 ES cell line	Engreitz et al. <sup>168</sup> (K. Plath)	pSM33 (dox-inducible Xist)
Mouse: Spy-tagged Eed ES cell line	This study	pSM33 (dox-inducible Xist)

(Continued on next page)

### Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Mouse: Spy-tagged Ezh2 ES cell line	This study	pSM33 (dox-inducible Xist)
Mouse: Spy-tagged Suz12 ES cell line	This study	pSM33 (dox-inducible Xist)
Mouse: Spy-tagged Ptbp1 ES cell line	This study	pSM33 (dox-inducible Xist)
Mouse: Spy-tagged Yy1 ES cell line	This study	pSM33 (dox-inducible Xist)
<b>Oligonucleotides</b>		
gRNAs, ultramer sequences, and genotyping primers used for CRISPR-targeting	This study	See Table S3
<b>Recombinant DNA</b>		
Entry clone expression vectors	Yang et al. <sup>169</sup>	See Table S1
pcDNA3-Tet2	Wang et al. <sup>170</sup>	RRID: Addgene_60939
pCAG-Halo-TEV-DEST-V5-IRES-puroR destination vector	This study	N/A
PyDox-Halo-TEV-DEST-V5-EFS-hygR destination vector	This study	N/A
T7-Halo-TEV-6xHis-SpyCatcher001 expression vector	This study	N/A
<b>Software and algorithms</b>		
CLAP pipeline (v0.1)	This study	<a href="https://github.com/GuttmanLab/CLAPAnalysis">https://github.com/GuttmanLab/CLAPAnalysis</a> <a href="https://doi.org/10.5281/zenodo.10535123">https://doi.org/10.5281/zenodo.10535123</a>
Integrative Genomics Viewer	Robinson et al. <sup>171</sup>	<a href="http://software.broadinstitute.org/software/igv/">http://software.broadinstitute.org/software/igv/</a>
STAR aligner	Dobin et al. <sup>172</sup>	<a href="https://github.com/alexdobin/STAR">https://github.com/alexdobin/STAR</a>
Bowtie2	Langmead and Salzberg <sup>173</sup>	<a href="http://bowtie-bio.sourceforge.net/bowtie2/index.shtml">http://bowtie-bio.sourceforge.net/bowtie2/index.shtml</a>
Picard	"Picard Toolkit." 2019. Broad Institute, GitHub Repository.	<a href="https://broadinstitute.github.io/picard/">https://broadinstitute.github.io/picard/</a>
Trim Galore! (v0.6.2)	Felix Krueger (The Babraham Institute)	<a href="https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/">https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/</a>
FastQC (v0.11.8)	Simon Andrews (The Babraham Institute)	<a href="https://www.bioinformatics.babraham.ac.uk/projects/fastqc/">https://www.bioinformatics.babraham.ac.uk/projects/fastqc/</a>
R Package (v4.3.1)	R Project	<a href="https://www.r-project.org/">https://www.r-project.org/</a>
RNA Centric Analysis System Report (v.3.18)	Uyar et al. <sup>174</sup>	<a href="https://github.com/BIMSBbioinfo/RCAS">https://github.com/BIMSBbioinfo/RCAS</a>
FIMO, MEME suite (v5.5.5)	Grant et al. <sup>175</sup>	<a href="https://meme-suite.org/meme/doc/fimo.html">https://meme-suite.org/meme/doc/fimo.html</a>
HOMER	Benner Lab	<a href="http://homer.ucsd.edu/homer/ngs/">http://homer.ucsd.edu/homer/ngs/</a>
ImageJ	Rueden et al. <sup>176</sup>	<a href="https://imagej.nih.gov/ij/">https://imagej.nih.gov/ij/</a>

## RESOURCE AVAILABILITY

### Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead contact, Mitchell Guttman ([mguttman@caltech.edu](mailto:mguttman@caltech.edu)).

### Materials availability

Plasmids, cell lines, and reagents generated in this study are available upon request to the [lead contact](#).

### Data and code availability

- CLIP and CLAP sequencing data have been deposited to GEO with identifier GEO: GSE253477. Original images for <sup>32</sup>P gels and processed CLIP and CLAP data are available through Mendeley Data and are publicly available as of the date of publication. The DOI is listed in the [key resources table](#).
- Software generated for this study is available through GitHub at <https://github.com/GuttmanLab/CLAPAnalysis>. The DOI is listed in the [key resources table](#).
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

## EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

### Cell culture

CLAP experiments were performed on either Human Embryonic Kidney Cells expressing T-antigen (HEK293T cell line) or male mouse embryonic stem cells containing a doxycycline-inducible Xist (bs/ps pSM33 cell line). HEK293T cells were cultured in HEK293T media consisting of 1X DMEM media (Gibco), 1 mM MEM non-essential amino acids (Gibco), 1 mM Sodium Pyruvate (Gibco), 2 mM L-Glutamine (Gibco), 1X FBS (Seradigm). bs/ps pSM33 were cultured in serum-free 2i/LIF medium as previously described.<sup>168</sup>

## METHOD DETAILS

### Challenges associated with distinguishing between promiscuous binding and lack of binding

It is well documented in RIP- and CLIP-Seq experiments that the overall coverage of an RNA detected in the capture sample is proportional to the abundance of the RNA in the input sample (background) and the enrichment of protein binding to that RNA (signal).<sup>159</sup> When a protein binds to specific targets, the proportion of reads for those RNAs will significantly exceed the level of RNA in the initial sample (signal > background). When a protein does not bind RNA *in vivo*, the proportion of reads will be similar to its abundance in the input sample (signal ~ background). Similarly, when a protein binds promiscuously to a majority of, or all, RNAs *in vivo* (i.e., SAF-A), the proportion of reads is expected to approximate that of the input sample, despite representing true binding events (signal ~ background); this is because the enrichment is constant across all RNAs, and therefore abundance will be the primary determinant of its proportions. Accordingly, it is difficult to distinguish between proteins that do not bind to RNA *in vivo* and those that bind promiscuously to many RNA targets without employing additional experimental controls.

### Cloning of Halo-V5-tagged expression constructs

Protein constructs were obtained from DNASU (<https://dnasu.org/DNASU/Home.do>) and LR-cloned (Invitrogen Gateway Cloning, Thermo Fisher Scientific) into either the mammalian expression destination vector pCAG-Halo-TEV-DEST-V5-IRES-puroR or PyDox-Halo-TEV-DEST-V5-EFS-hygR as previously described.<sup>77</sup>

Additional chromatin proteins were selected based on the criteria that they were either (i) enriched in at least one of 40+ global proteomics studies identifying RBPs, as catalogued by the comprehensive RBP2GO database<sup>97</sup> or (ii) previously reported to bind to RNA to enact chromatin regulatory functions. We further filtered this list to focus only on proteins that were present in the ORFeome entry clone library<sup>169</sup> (with the exception of TET2). Meta-analysis including the number of times a candidate protein was identified as an RBP, which studies (including method, authors, and year when study was published) identified them, and complete Gene Ontology annotations, was downloaded directly from the RBP2GO website (<https://rbp2go.dkfz.de/>) and included in Table S1, along with DNASU Clone ID numbers.

For TET2, an entry clone was generated by BP cloning (Invitrogen Gateway Cloning, Thermo Fisher Scientific) using a PCR amplicon (primers: GGGGACAAGTTTGTACAAAAAAGCAGGCTTTatggaacaggacagaaccacc and GGGGACCACTTTGTACAAGAAAGCTGGGTTtacaatgtgtgtgaaggccc; template: pcDNA3-Tet2, a gift from Yi Zhang<sup>170</sup> (see Table S1 for Addgene information) into pDONR223.

### Expression, UV-crosslinking, and lysis of cells

Expression, UV-crosslinking, and lysis of HEK293T cells were performed as previously described.<sup>77</sup> pSM33 cells were trypsinized using 0.025% trypsin (Gibco), pelleted, and transferred to tubes at a ratio of 2 million cells/transfection and pelleted by centrifugation. Cells were then resuspended in resuspension buffer R (Invitrogen) and mixed with 12  $\mu$ g of DNA. The mixture was transfected with the following settings using 100  $\mu$ L tips on the Neon Transfection Device (Invitrogen): 1400 V, 3 pulses, and a 10 ms pulse width. Transfected cells were pipetted directly onto a 10 cm culture plate. After 24 hours, the media was changed on the samples and 1  $\mu$ g/mL puromycin was added in order to select for cells that contained the transfected expression cassette. At time of transfection, cells were washed once with PBS and then crosslinked on ice using 0.25 J cm<sup>-2</sup> (UV 2.5k) of UV at 254 nm in a Spectrolinker UV Crosslinker. Cells were then scraped from culture dishes, washed once with PBS, pelleted by centrifugation at 1500g for 4 min, and flash-frozen in liquid nitrogen for storage at -80 °C.

### Western blot of tagged proteins

HEK293T cells expressing the Halo-tagged proteins were harvested and lysed as described before in 1 mL of cold lysis buffer (50 mM HEPES pH 7.4, 100 mM NaCl, 1% NP-40, 0.1% SDS, 0.5% sodium deoxycholate). 20  $\mu$ L of lysate was diluted to 1X final concentration of LDS loading buffer (4  $\mu$ L 10X Bolt reducing agent, 10  $\mu$ L 4X NuPAGE LDS Sample Buffer, 4  $\mu$ L H<sub>2</sub>O), denatured at 80 °C for 6 minutes, run on a Bolt 4-12% Bis-Tris Plus Gel (Thermo Fisher Scientific), and then transferred to a nitrocellulose membrane using the iBlot transfer system. Proteins were visualized by western blotting using the following primary antibodies and dilutions: anti-V5 (Bethyl, A190-119A; 1:2000 dilution), anti-HaloTag (Promega, G9211, 1:1000 dilution), anti-EED (CST, E4L6E, 1:1000 dilution), anti-EZH2 (CST, D2C9, 1:1000 dilution), anti-SUZ12 (CST, D39F6, 1:1000 dilution), and anti-PTBP1 (CST, E5O2S, 1:1000 dilution).



### Co-immunoprecipitation of PRC2 components and AlexaFluor labeling

5  $\mu$ g of Rabbit IgG (CST, 2729S), anti-EZH2 (Active Motif, 39933), or anti-SUZ12 (Active Motif, 39357) antibodies were coupled to 50  $\mu$ L of Protein G beads (Dynabeads) at room temperature for 30 minutes. Beads were washed three times with 300  $\mu$ L mammalian lysis buffer (50 mM HEPES pH 7.5, 150 mM NaCl, 1% Triton X-100, 0.1% sodium deoxycholate). HEK293T cells expressing Halo-V5 fusion proteins of EZH2, EED, or SUZ12 were lysed as in the CLAP procedure except that 1 mL of mammalian lysis buffer was used in place of standard lysis buffer. Beads were then incubated with HEK293T lysates expressing either Halo-EZH2-V5, Halo-EED-V5, or Halo-SUZ12-V5. The antibody-coupled beads were incubated with lysate overnight. After binding, the beads were washed three times with mammalian lysis buffer for 2 minutes with shaking (1500 rpm) on a ThermoMixer. After washes, the supernatant was removed and replaced with 18  $\mu$ L of mammalian lysis buffer + 2  $\mu$ L of a 1:60 dilution of diluted HaloTag Alexa Fluor 660 Ligand (Promega). This was incubated at room temperature in the dark for 20 minutes. The reaction was quenched by adding 4X LDS loading buffer and heating at 70°C for 7 minutes. After heating, the beads were placed on a magnet and the supernatant was loaded on a 3-8% Tris-Acetate gel. The gel was imaged directly on the LI-COR Odyssey.

### Measurement of *in vitro* binding of PRC2 components to RNA

Direct binding of RNA to covalently immobilized Halo-tagged fusion proteins was assayed on a Biacore T200 (GE Healthcare Life Sciences) and Series S Sensor Chip CM5 (GE Healthcare Life Sciences, BR100530). The Halo capture reagent (chloroalkane) was coupled to the chip by amine coupling according to the manufacturer's protocol (Promega, P6741) with the following deviations. The Halo capture reagent was resuspended in anhydrous DMSO (5 mg/mL) and diluted to 2.5 mg/mL in 1x HBS-N Buffer (GE Healthcare) and injected onto the chip until 300 resonance units (RU) of amine ligand was immobilized. Ethanolamine (1 M, pH 8.5) (Sigma-Aldrich, 15014) was injected for 7 minutes at 10  $\mu$ L/minute to block remaining active sites on the chip.

HEK293T cells transfected with DasherGFP (ATUM Biosciences, FPB-27-609), DasherGFP-3x- $\lambda$ N, EZH2, EED, or PTBP1 fused to a N-terminal HaloTag were prepared as described above. Cells were resuspended in 2 mL of 1x HBS-EP+ buffer (GE Healthcare, BR100669) supplemented with 1X Protease Inhibitor (Promega, G6521), 2.5 mM manganese chloride, 0.5 mM calcium chloride, 40 U of Turbo DNase (Ambion, AM2239), 40  $\mu$ g of RNase A and 100 U of RNase T1 mix (Ambion, EN0551), and incubated on ice for 10 minutes. Cells were then sonicated (Branson Ultrasonics) for 30 seconds at 5W (0.7 seconds on, 0.7 seconds off) then incubated at 37°C for 10 minutes at 1100 RPM on a ThermoMixer. Samples were then placed on ice for 2 minutes prior to centrifugation at 16000 x g for 2 minutes at 4°C. Clarified lysate was injected onto flow cells 2 and 4 of the chip for 60 seconds to allow Halo-tagged proteins to covalently bind the chip surface, followed by a 1 second injection of 50 mM NaOH to clean the chip surface and remove non-covalently bound RNA, DNA, and protein. Injections of lysate and NaOH pulses were continued until 10 RU of Halo-tagged protein was covalently immobilized on the chip surface. Flow cells 1 and 3 of the chip were left blank to be used as reference surfaces.

RNA derived from the Maltose Binding Protein (MBP, 1-240 nucleotides) fused to 5 copies of the BoxB aptamer (MBP-5x-BoxB), the A-repeat (260-1,002 nucleotides), or the E-repeat (11,963-12,705 nucleotides) of the Xist RNA were *in vitro* transcribed using the T7 RiboMAX Express Large Scale RNA Production System (Promega, P1320) after PCR amplification to incorporate a T7 promoter. *In vitro* transcribed RNA was diluted with water and 10X HBS-EP+ Buffer to a final concentration of 1.1  $\mu$ M prior to heat denaturation at 70°C for 2 minutes. 1M magnesium chloride was added to a final concentration of 3.25 mM and allowed to cool to room temperature. RNA was then stored on ice or at 4°C prior to injection over all four flow cells at 25°C at 100  $\mu$ L/min for 60 seconds. The different concentrations of RNA were injected by the instrument in a randomized order. After injection ended, dissociation was monitored in each flow cell for 500 seconds. Regeneration of the sensor chip surface was performed by injecting 50 mM NaOH at 100  $\mu$ L/min for 3 sec, waiting 180 seconds for the baseline to stabilize, then injecting a 1 second pulse of NaOH, waiting 240 second for the baseline to stabilize, and washing the injection needle.

Sensorgrams were processed with Biacore T200 Evaluation Software, (version 3.0). The y-axes were zeroed at the baseline for each cycle and x-axes were aligned at the injection start. We used the first 100 seconds of the dissociation curve for global fitting. Bulk refractive index changes and systematic deviations in sensorgrams were removed by subtracting the responses in reference flow cells (1 and 3) corresponding to the sample flow cells (2 and 4). The averaged sensorgrams for 0 nM RNA were then subtracted from sensorgrams for all other concentrations. After double referencing kinetic data and removing injection and pump spikes, the data were fit globally by non-linear regression to a simple 1:1 Langmuir binding model with a bulk refractive index term to determine association/dissociation rate constants ( $k_a$ ,  $k_d$ ), analyte binding capacity ( $R_{max}$ ), and the equilibrium dissociation constant ( $K_D$ ). Sensorgrams and 1:1 binding model curve fits were exported and plotted.

### Enhanced crosslinking and immunoprecipitation

eCLIP was performed as previously described,<sup>71</sup> with slight modifications. Cells were lysed in 1 mL lysis buffer (50mM Tris pH 7.5, 100mM NaCl, 1% NP-40, 0.5% sodium deoxycholate, 1X Promega PIC). RNA was digested with Ambion RNase I (1:3,000 dilution) to achieve a size range of 100-500 nucleotides in length. We used a fixed RNase condition for all experiments to enable comparison between proteins. Lysates were precleared by mixing with Protein G beads (Dynabeads) for 30 minutes at 4°C. Target proteins were immunoprecipitated overnight at 4°C from 10 million cells with 5  $\mu$ g of antibody coupled to 40  $\mu$ L of Protein G beads in 100  $\mu$ L lysis buffer. Antibodies were pre-coupled to beads for 1 hour at room temperature with mixing, followed by 3 washes with lysis buffer to remove unbound protein. After immunoprecipitation, beads were washed four times with high salt wash buffer (50 mM TrisHCl pH 7.4, 1 M NaCl, 1 mM EDTA, 1% NP-40, 0.1% SDS, 0.5% sodium deoxycholate) and four times with wash buffer

(20 mM Tris-HCl pH 7.4, 10 mM MgCl<sub>2</sub>, 0.2% Tween-20). RNA and protein were eluted by incubating in NLS elution buffer (20 mM Tris-HCl pH 7.5, 10 mM EDTA, 2% N-lauroylsarcosine, 2.5 mM TCEP) supplemented with 100 mM DTT at 50°C for 20 minutes. Samples were then run through an SDS-PAGE gel and transferred to a nitrocellulose membrane using the iBlot transfer system. The region 70 kDa above the molecular weight of the protein of interest was isolated and treated with Proteinase K (NEB), followed by purification with RNA Clean & Concentrate-5 (Zymo, >17 nucleotides protocol). Anti-V5 antibody (Bethyl, A190-120A) was used for all CLIP experiments.

### CLIP RNase titration

RNase titration was carried out to resolve RNA-protein complexes at RNase I (Ambion) dilutions of 1:50, 1:3,000, and 1:50,000. eCLIP was performed as previously described,<sup>71</sup> and radiolabeling of captured RNA-protein complexes was performed as detailed above.

Titration of RNase in a CLIP assay is expected to resolve RNA signal such that it co-migrates with the size of the protein on an SDS-PAGE gel, highlighting specificity for the RBP (i.e., PTBP1, Figure S3C). We found that CLIP of PRC2 shows generally low levels of RNA co-migration with scaling RNase concentrations, but slightly higher than those in the absence of UV or in untransfected controls (Figures S3A and S3B). Specifically, while the CLIP procedure works well for known RBPs such as PTBP1, there is a qualitatively different result for the PRC2 components which show much lower overall RNA intensity, the sizes do not shift dramatically with RNase, and the amount of RNA that co-migrates with the protein is a tiny fraction of the total RNA (Figure S3A). This same trend was observed in previously published PRC2 CLIP experiments, including a modest co-migration band with large amounts of background RNA signal.<sup>21,22</sup> This preferential RNA signal near the size of the protein itself has been shown to occur for other non-RBPs.<sup>177</sup>

Despite these clear differences relative to other RBPs like PTBP1, the difference in overall RNA amounts observed for the PRC2 component in the +UV samples is qualitatively higher than in the -UV samples or in the untransfected controls, which is likely caused by UV crosslinking-specific sources of background.

### Possible explanations for UV- and protein-dependent non-specific associations

We considered several possible explanations for why CLIP identifies RNA-protein interactions that do not occur *in vivo*.

- (i) The detected RNAs may be crosslinked to other non-specific proteins present after immunoprecipitation. Because CLIP relies on immunoprecipitation, the stringency of purification is limited to wash conditions (i.e., high salt followed by low salt) that maintain the antibody-protein interaction as well as the interaction between the Protein G bead and the antibody. These conditions have been shown to retain non-crosslinked protein-protein interactions in the case of PRC2 components<sup>22</sup> and may similarly retain other protein-protein interactions that form in solution. To specifically compare the limited stringency of CLIP-washes to denaturing CLAP-washes, we performed an experiment in which we could keep all other parameters identical (i.e., protein amount and capture) and simply vary the wash conditions. Because antibody-epitope interactions do not withstand the denaturing CLAP-wash conditions, we covalently conjugated two different proteins (Halo-PTBP1-V5 and Halo-EZH2-V5) onto HaloLink resin, split the resin, performed either CLIP-washes or CLAP-washes, ran samples on denaturing SDS-PAGE, and then visualized all proteins present after elution using a total protein stain. For both proteins, we identified non-specific background proteins purified in the CLIP-wash conditions that were not detected in the CLAP-wash conditions (Figure 2C). Further, many of these detected non-specific proteins are within the size range that is excised from the nitrocellulose membrane in a CLIP experiment for each target protein (Figure 2C, red line), and therefore would not be excluded by the gel separation and size-based extraction of RNA-protein complexes steps present in CLIP. These results may explain why the level of background RNAs is significantly lower when CLIP is performed in non-crosslinked lysates, with IgG controls, or in knockout cells that lack the target protein that is immunoprecipitated.<sup>20,22</sup>
- (ii) The detected RNAs may be free RNAs in solution that are not fully removed in the CLIP assay. It is commonly accepted that the nitrocellulose membrane transfer step after denaturing gel electrophoresis in CLIP enriches for crosslinked complexes because nitrocellulose is expected to only bind to proteins and not free RNA.<sup>67</sup> However, while this step does enrich for bound RNA, we found that there is still a significant amount of free RNA that is retained. Specifically, we tested the amount of non-crosslinked RNA retained after CLIP washes, gel separation, and nitrocellulose membrane transfer by measuring the amount of RNA recovered from UV-crosslinked cells and non-crosslinked cells (Figure S5D). Importantly, we recovered a large amount of RNA from the non-crosslinked samples; in fact, we observed only ~4-fold less RNA relative to the amount purified from UV-crosslinked samples (Figure S5D). Detection of non-crosslinked RNA-protein interactions may be due in part to the abundance of free RNA and protein from the low efficiency of RNA-protein crosslinking by UV light (~1-5%).<sup>63</sup>
- (iii) The specific protein that is purified may interact with RNAs in solution to form non-crosslinked RNA-protein complexes. This could lead to enrichment of an RNA that is either (i) crosslinked to a distinct protein *in vivo* (i.e., another RBP) or (ii) free RNA in solution; both would increase with increasing amounts of non-crosslinked RNA that remains after CLIP-washes. To directly compare the efficiency of CLIP-washes to CLAP-washes in removing non-crosslinked RNA, we took equivalent amounts of either UV-crosslinked or non-crosslinked (-UV) cell lysates and coupled them to NHS-activated magnetic beads, which bind to all proteins in the sample (Figure S5A). We then split these beads, performed either CLIP-washes or CLAP-washes, and eluted the RNAs using Proteinase K. We found that CLIP-washes retain a large amount of -UV RNAs relative to CLAP-washes (which leave a virtually undetectable amount of RNA) (Figure 2D). Because the amount of -UV RNA unambiguously

represents background signal (i.e., non-crosslinked RNA-protein interactions), the -UV/-UV ratio of CLIP-washes to CLAP-washes (>200-fold) indicates the high signal-to-noise properties of CLAP washes. In addition, CLAP-washes enrich for +UV RNAs to a much greater degree than do CLIP-washes (23-fold versus 1.9-fold, respectively, Figure S5B). Because neither nitrocellulose transfer (see above) nor CLIP-washes can fully deplete non-crosslinked RNAs, strongly associated (but not crosslinked) RNA-protein interactions that form in solution may persist throughout the CLIP protocol. This direct association of RNA and protein in solution may explain why we observe strong binding sites in the -tag samples for PRC2 on the A-repeat of XIST (Figure 1C), which we and others show bind with high affinity *in vitro*<sup>5,45,58,59</sup> (Figure S2). Furthermore, the association of proteins with free RNAs would also lead to the resolution of a band at the precise molecular weight of the protein on SDS-PAGE, which would explain why we and others [S<sup>177,178</sup> observe this qualitatively with CLIP gels of PRC2 and other non-RBPs (Figures S3A and S19B).

All three of these issues arise in CLIP because the protein purification and denaturation steps must be decoupled to protect antibodies from denaturation. Despite this, we expect that the signal from *in vivo* crosslinked RNA-protein interactions will be strongly enriched over background for *bona fide* RBPs. However, in cases where a protein does not bind to RNA *in vivo* or binds to few (or low abundance) RNA targets, these issues may lead to high detection of non-specific RNA-protein interactions because they will be present at a significantly higher abundance relative to *bona fide* interactions.

### Comparison of CLIP-washes to CLAP-washes on NHS beads

Either UV-crosslinked or non-crosslinked HEK293T cells (20 million cells each) were harvested and lysed as previously described. Lysates were then coupled to NHS-activated magnetic beads (Pierce) overnight at 4°C rotating on a HulaMixer Sample Mixer (Thermo). After coupling overnight, beads were quenched by removing 500  $\mu$ L of flowthrough lysate and adding 500  $\mu$ L of 1M Tris pH 7.5 and incubated for an additional 30 minutes at room temperature. Flowthroughs were then removed, and samples were either subject to standard CLIP-washes or CLAP-washes as described above. Remaining RNAs were released by Proteinase K elution at 50°C for 30 minutes, followed by purification with RNA Clean & Concentrate-5 (Zymo), and then run on an Agilent TapeStation High Sensitivity RNA assay to measure RNA sizes and concentration.

### Covalent linkage and affinity purification

CLAP was performed on HEK293T and pSM33 cells as previously described.<sup>77</sup> Briefly, post-crosslinking, cells were resuspended in 1 mL of cold lysis buffer (50 mM HEPES pH 7.4, 100 mM NaCl, 1% NP-40, 0.1% SDS, 0.5% sodium deoxycholate) supplemented with 1X Protease Inhibitor Cocktail (Promega), 200 U of RiboLock (NEB), 20 U of TURBO DNase (Ambion), and 1X manganese/calcium mix (0.5mM CaCl<sub>2</sub>, 2.5 mM MnCl<sub>2</sub>). Samples were incubated on ice for 10 minutes and then at 37°C for 10 minutes at 1150 rpm shaking on a ThermoMixer (Eppendorf). Lysates were cleared by centrifugation at 15000g for 2 minutes and the supernatant was collected for capture to HaloLink Resin (Promega). For each CLAP capture, 200  $\mu$ L of 25% HaloLink Resin (50  $\mu$ L of HaloLink Resin total) was used per 10 million cells. Resin was washed three times with 2 mL of 1X TBS (50 mM Tris pH 7.5, 150 mM NaCl) and incubated in 1X Blocking Buffer (50 mM HEPES, pH 7.5, 10  $\mu$ g/mL Random 9-mer, 100  $\mu$ g/mL BSA) for 20 minutes at room temperature with continuous rotation. After the incubation, resin was washed three times with 1X TBS. Cleared lysate was mixed with 50  $\mu$ L of HaloLink Resin and incubated at 4°C overnight with continuous rotation. The captured protein bound to resin was washed three times with lysis buffer at room temperature and then three times at 90°C for 3 minutes while shaking at 1200 rpm with each of the following buffers: 1X ProK/NLS buffer (50 mM HEPES, pH 7.5, 2% NLS, 10 mM EDTA, 0.1% NP-40, 10 mM DTT), high salt buffer (50 mM HEPES, pH 7.5, 10 mM EDTA, 0.1% NP-40, 1M NaCl), 8M urea buffer (50 mM HEPES, pH 7.5, 10 mM EDTA, 0.1% NP-40, 8 M Urea), and Tween buffer (50 mM HEPES, pH 7.5, 0.1% Tween 20, 10 mM EDTA). After the last wash, samples were centrifuged at 7500g for 30 seconds and supernatant was discarded. For elution, HaloLink Resin was resuspended in 100  $\mu$ L of ProK/NLS buffer + 10  $\mu$ L of Proteinase K (NEB) and incubated at 50°C for 20 minutes while shaking at 1200 rpm. Elutions were then transferred to micro-spin cups (Pierce, Thermo Fisher), centrifuged at 2000g for 30 seconds, and purified with RNA Clean and Concentrate-5 (Zymo, >17 nucleotides protocol).

### Radiolabeling of captured RNA-protein complexes

After all the appropriate washes (in CLIP or CLAP buffers) were performed on captured RNA-protein complexes, they were buffer exchanged with 1X FastAP buffer and then 5' end dephosphorylated by incubating in 100  $\mu$ L FastAP mix (1X Fast AP buffer, 8  $\mu$ L FastAP Enzyme, 2  $\mu$ L Murine RNase Inhibitor, 5  $\mu$ L TURBO DNase) at 37°C for 15 min. RNA was then end repaired using T4 PNK. 300  $\mu$ L of the end repair reaction buffer (1X T4 PNK Buffer, 7  $\mu$ L T4 PNK Enzyme, 1  $\mu$ L Murine RNase Inhibitor, 1  $\mu$ L TURBO DNase) was added on top of the FastAP mix and incubated at 37°C for another 15 minutes. Samples were then washed twice each with high salt wash buffer, low salt wash buffer, and 1X PNK buffer. They were then resuspended in 1 mL PNK buffer until 5' end phosphorylation. We collected 200  $\mu$ L (20%) of beads or resin from the previous step and removed the supernatant. RNA was radiolabeled using 4  $\mu$ L of hot PNK mix (0.2  $\mu$ L T4 PNK, 0.4  $\mu$ L  $\gamma$ -<sup>32</sup>P-ATP, 0.4  $\mu$ L 10x PNK buffer, 3  $\mu$ L H<sub>2</sub>O) and incubated at 37°C in a ThermoMixer at 1100 rpm for 5 minutes. Once radiolabeling was complete, the supernatant was discarded and each reaction was washed twice with 100  $\mu$ L of high salt wash buffer, low salt wash buffer, and 1X PNK buffer. Samples were then either cleaved with tobacco etch virus (TEV) protease (CLAP samples) or directly resuspended in 40  $\mu$ L of 1X NuPAGE loading buffer by pipetting

(CLIP samples). Samples were loaded on a 4–12% NuPAGE Bis-Tris gel and run at 180V for 50 minutes. Radiolabeled RNA-protein complexes were transferred to nitrocellulose membrane using the iBlot 1.0 system. After the transfer, the membrane was rinsed three times in 1X PBS, wrapped in Saran wrap, and exposed to a phosphor screen. Imaging of phosphor screen was performed on a Typhoon scanner.

Image files for all raw scans for  $^{32}\text{P}$  CLIP and CLAP experiments are deposited in Mendeley Data: <https://doi.org/10.17632/wmsbzbv6kg5.2>.

### Failure of CLAP to identify RNA-protein interactions identified by CLIP cannot be due to differences in assay sensitivity

We note that the failure of CLAP to detect interactions compared to those identified in CLIP cannot be attributed to differences in assay sensitivity. Because CLAP, like CLIP, utilizes UV-crosslinking to generate covalent crosslinks between interacting RNA and protein, there should be no intrinsic difference between the two methods in their ability to detect protein-RNA interactions of different affinities or stability. Once UV-crosslinked, each RNA-protein interaction would be covalently linked and therefore of identical strength. Instead, the differences in stability or affinity of a protein-RNA interaction would be reflected by the amount of RNA bound to a protein at a given time. As such, the only confounding factor in which CLIP may be more sensitive than CLAP would be if it captures more protein; however, we demonstrate that the HaloLink resin used for CLAP consistently recovers higher amounts of protein than Protein G beads used in CLIP, likely due to having higher protein binding capacity (Figure S9). Thus, if a real interaction occurs and could be detected by CLIP (e.g., can form a UV-induced crosslink) then it should also be detectable by CLAP.

### Protein quantification by Halo-ligand

Because of the irreversible, covalent nature of Halo-tagged protein capture, directly measuring captured protein requires elution with TEV protease. However, this cleavage is not 100% efficient and the efficiency can vary based on the exact fusion protein being captured. To quantitatively compare protein capture efficiency (bound protein) between CLIP and CLAP samples, the following were measured by Halo-ligand labeling: total protein (the amount of protein added to the conjugation reaction), unbound protein (the amount of protein present in the flowthrough), and protein loss (the amount of protein lost after all wash steps). The concentration of captured protein is equal to the total protein minus the unbound protein (Figure S9A).

Immediately following either CLIP or CLAP protein capture, flowthroughs were saved and kept on ice. All washes performed for either CLIP or CLAP were subsequently kept, pooled, and saved on ice separately. 20  $\mu\text{L}$  of each sample (input, flowthrough, and washes) was combined with 1.5  $\mu\text{L}$  of 1:60 diluted HaloTag Alexa Fluor 660 Ligand (Promega) and incubated at room temperature for 20 minutes in the dark. Reactions were stopped by adding LDS loading buffer to 1X final concentration (4  $\mu\text{L}$  10X Bolt reducing agent, 10  $\mu\text{L}$  4X NuPAGE LDS Sample Buffer, 4  $\mu\text{L}$  H<sub>2</sub>O), denatured at 80°C for 6 minutes and run on a Bolt 4–12% Bis-Tris Plus Gel (all products Thermo Fisher Scientific). Resolved gel was imaged directly on a Li-Cor Odyssey CLx and protein bands were quantified on Image Studio.

The higher capture efficiency demonstrated for CLAP (relative to CLIP, Figures S9B and S9C) across all tested proteins is consistent with the fact that the HaloLink resin (Promega) has a higher binding capacity compared to Protein G beads (Dynabeads). The amount of protein lost during the various wash steps was comparable in both approaches, indicating that the bound protein was not being somehow preferentially lost at subsequent steps of the process (Figure S9D).

### Elution and visualization of CLAP-purified proteins

To verify successful purification of Halo-tagged proteins after CLAP, we performed an additional three final washes on HaloLink Resin using TEV buffer (50 mM HEPES, pH 7.4, 1 mM EDTA, 0.1% NP-40). The resin was resuspended in 83  $\mu\text{L}$  of Elution Buffer and split into a 75  $\mu\text{L}$  (ProK elution) and 8  $\mu\text{L}$  (TEV elution) reaction. 25  $\mu\text{L}$  of 4X ProK/NLS Buffer and 10  $\mu\text{L}$  of ProK were added to the ProK elution tube and the sample was incubated at 50 °C for 30 minutes while shaking at 1200 rpm. 2.3  $\mu\text{L}$  of ProTEV Plus Protease (Promega) was added to the TEV Elution and the sample was incubated at 30°C for 30 minutes while shaking at 1200 rpm.

The TEV elution sample was mixed with 1X LDS Sample Buffer (Invitrogen) and 1X Reducing Agent (Invitrogen) and heated for 6 minutes at 70°C. The sample was run on a 3–8% Tris Acetate Gel (Invitrogen) for 1 hour at 150V. The gel was transferred to a nitrocellulose membrane using an iBlot Transfer Device (Invitrogen). The nitrocellulose membrane was blocked with Odyssey Blocking Buffer (LI-COR) for 30 minutes. We incubated the membrane in Anti-FLAG mouse monoclonal antibody (Sigma, F3165) and V5 rabbit polyclonal antibody (Santa Cruz, sc-83849-R) at a 1:2500 dilution for 2 hours at room temperature to detect the protein. We visualized the protein by incubating the membrane in 1:17500 dilution of both IRDye 800CW Goat anti-Rabbit IgG (LI-COR, 925-32211) and IRDye 680DR Goat anti-Mouse IgG (LI-COR, 925-68070) for 1 hour at room temperature followed by imaging on a LI-COR Odyssey.

### Halo-tagged PRC2 components and purified PRC2 complexes associate and UV-crosslink with RNA *in vitro*

Given the evidence that PRC2 interacts with RNA *in vitro* and existing evidence in support of *in vivo* binding is based on RIP and CLIP, we wanted to ensure that failing to detect these interactions by CLAP was not due to technical limitations of our method. Accordingly, we considered several ways in which CLAP might fail to detect real interactions that occur *in vivo*.

First, we considered that integration of the covalent epitope tags (i.e., HaloTag) into the PRC2 components could disrupt their ability to bind to RNA. To ensure this was not the case, we measured *in vitro* binding affinities for each of the Halo-tagged PRC2



components (Figure S2). Specifically, we immobilized Halo-tagged PRC2 components on a Biacore chip and flowed in different concentrations of *in vitro* transcribed RNA (A-repeat of Xist, a previously reported PRC2 binding site) and measured surface plasmon resonance (SPR) responses. We observed binding affinities (reported as an equilibrium dissociation constant ( $K_D$ ), between each of the PRC2 components (EED, SUZ12, and EZH2) with the A-repeat of Xist that range from  $10^{-7}$  M to  $10^{-5}$  M. This was significantly higher than what was observed for GFP and a control RNA, for which we were unable to detect association; yet tagged PRC2 binding affinity was lower than that observed between  $\lambda$ N protein and BoxB RNA ( $K_D$  of  $3.9 \times 10^{-8}$  M) and PTBP1 with its known Xist binding site on the E-repeat ( $K_D$  of  $6.00 \times 10^{-8}$  M).

Next, we confirmed that Halo-tagged PRC2 components are capable of forming UV-induced crosslinks to RNA. To do this, we exploited the fact that PRC2 associates with RNA *in vitro* to assemble a PRC2-RNA complex and then crosslinked it with UV light. Specifically, we lysed uncrosslinked cells (allowing for post-lysis RNA-protein interactions), purified tagged PRC2 components on HaloLink resin, incubated them with *in vitro* transcribed RNA (A-repeat of XIST), and then crosslinked them with 254 nm UV light (using the same conditions for *in vivo* crosslinking). We then measured the amount of crosslinked RNA by washing away uncrosslinked RNA with CLAP washes and eluting the remainder with Proteinase K, followed by reverse transcription and PCR to read out the samples on an Agilent TapeStation assay (Figure S11A). We observed >5-fold enrichment of crosslinked RNA in our +UV samples relative to our -UV controls for each PRC2 component, in contrast to our negative control GFP which showed no enrichment (Figure S11A).

Finally, we and others<sup>59</sup> show that purified PRC2 complexes (with no exogenous tags) are capable of crosslinking to UV *in vitro*. Briefly, we incubated purified PRC2 complexes (Active Motif) with short biotinylated RNA sequences (GGAA, CCUU, mixture of GGAA+CCUU), crosslinked them with 254 nm UV light, ran samples on SDS-PAGE, transferred to nitrocellulose membrane, and visualized the RNAs with a streptavidin-conjugated infrared dye (Figure S11B). We observed high amounts of RNA signal specifically at the expected protein sizes of each individual PRC2 component.

### ***In vitro* RNA crosslinking on HaloLink resin**

Non-crosslinked HEK293T cells transfected with Halo-tagged versions of GFP, PTBP1, EED, EZH2, and SUZ12 were harvested and lysed as previously described, except the lysis buffer was replaced with a native lysis buffer (M-PER Mammalian Protein Extraction Reagent). Lysates were then sonicated (Branson Ultrasonics) for 1 minute at 4W (0.7 seconds on, 1 second off) to aid release of proteins from chromatin, then incubated at 37°C for 10 minutes at 1200 RPM on a ThermoMixer. Lysates were cleared by centrifugation at 15000g for 2 minutes, then incubated with RNase If (NEB) at a 1:500 effective dilution for 10 minutes at 37°C at 1200 RPM on a ThermoMixer. RNase reaction was quenched with an addition of 500  $\mu$ L ice cold lysis buffer supplemented with 20  $\mu$ L Protease Inhibitor (Promega PIC) and 5  $\mu$ L of RiboLock RNase inhibitor, followed by incubation on ice for 3 minutes. Each lysate was then bound to 50  $\mu$ L of HaloLink Resin and incubated at 4°C overnight. To remove background proteins, the resin was washed 3X with native lysis buffer. The resin was then mixed with 200 ng of denatured, *in vitro* transcribed XIST A-repeat RNA in 50  $\mu$ L binding buffer as previously described,<sup>5,58</sup> with slight modifications (50 mM Tris-HCl pH 7.5, 100 mM KCl, 5 mM MgCl<sub>2</sub>, 0.1 mM CaCl<sub>2</sub>). *In vitro* binding was allowed to proceed for 1 hour at 30°C. The captured proteins were then split into two conditions: -UV and +UV. For the +UV condition, the RNA-protein mixture was crosslinked on ice using 0.25 J cm<sup>-2</sup> (UV2.5k) of UV at 254 nm in a Spectrolinker UV Crosslinker (the same amount used for *in vivo* crosslinking). To remove uncrosslinked RNA, each sample was then washed three times each with native lysis buffer, RIPA buffer, high salt buffer, 8M urea buffer, and low salt buffer. Any remaining bound RNA was then released by digesting the resin with Proteinase K for 30 minutes at 50°C, followed by standard CLAP RNA library preparation. The molarity of each sample was then measured by Agilent TapeStation High Sensitivity DNA Assay.

### ***In vitro* RNA crosslinking and streptavidin-IR western blot**

12  $\mu$ g of recombinant PRC2 complex (Active Motif, 31337) was resuspended in 50  $\mu$ L of native lysis buffer (1X HBS-EP+ Buffer supplemented with 100 mM KCl), then incubated with 25  $\mu$ L of a 50 mM mix of 40-mer RNAs (chemically synthesized by IDT) of either GGAA sequences, CUCU sequences, or a mix of both that were denatured and rapidly annealed together (95°C for 1 minute and held at 4°C). Samples were incubated at 30°C for 20 minutes and then crosslinked on ice using 0.25 J cm<sup>-2</sup> (UV2.5k) of UV at 254 nm in a Spectrolinker UV Crosslinker (the same amount used for *in vivo* crosslinking). Protein-RNA complexes were then denatured at 75°C on a ThermoMixer for 6 minutes, run on an SDS-PAGE gel, transferred to a nitrocellulose membrane using the iBlot transfer system, and imaged using a streptavidin-conjugated infrared dye (IRDye 800CW Streptavidin, LI-COR) at a 1:2000 dilution in blocking buffer (Intercept, LI-COR).

### **Cell line generation**

Murine embryonic stem cells (mES) containing a dox-inducible Xist<sup>55,106,168</sup> (bs/ps pSM33, cells were kindly provided by K. Plath) were CRISPR-targeted to endogenously tag selected proteins with SpyTag-V5 (see Table S2 for gRNA used for targeting and ultramer sequences used for insertion templates). N-terminal V5-Spy targeting was performed for *Eed*, *Ezh2*, and *Ptbp1* alleles, and C-terminal V5-Spy targeting was performed for *Suz12* and *Yy1* alleles.

In brief, cells were co-transfected with plasmids expressing sgRNAs, wtCas9, and ultramers (IDT) and selected on antibiotics for which resistance was conferred following successful plasmid transfection. Single colonies were picked and screened by gDNA isolation and PCR confirmation for the tagged version of the protein.



### CLAP on Spy-tagged proteins

CLAP on endogenous Spy-tagged proteins was performed previously described<sup>77</sup> with slight modifications. A histidine-tagged HaloTag-SpyCatcher fusion protein was first expressed in BL21 DE3 *E. coli* bacterial cells and purified by IMAC as previously described.<sup>77</sup> Per capture, 250  $\mu$ g of HaloTag-SpyCatcher was incubated with 50  $\mu$ L of HaloLink resin, bound for 30 minutes at room temperature with continuous rotation, then washed three times with lysis buffer at room temperature. Clarified lysate from Spy-tagged cell lines was then mixed with 50  $\mu$ L of HaloLink resin pre-coupled to SpyCatcher and incubated at 4°C overnight with continuous rotation. CLAP was then performed according to standard protocol.

### Crosslinking for ChIP

Mouse embryonic stem cell lines (pSM33) were crosslinked in either 1% formaldehyde or 1% formaldehyde and 2 mM DSG (Thermo #2059). Briefly, cells were washed with room temperature 1X PBS and then incubated in crosslinker solution in PBS at room temperature with gentle rocking for 45 minutes. Cells were then washed with room temperature 1X PBS. After washing, a 1% formaldehyde solution was then added on top of cells and further incubated at room temperature for 10 minutes. To quench the formaldehyde crosslinking, 2 mL of a 2.5M Glycine solution was added to each plate and incubated for an additional 5 minutes. After quenching, cells were washed three times in 1X cold PBS. After the last wash, 7.5 mL of scraping buffer (1X PBS + 0.5% BSA) was added to cells. Cells were then scraped using a rubber policeman, aliquoted into 10 million cell aliquots, and flash frozen until ChIP was performed.

To solubilize chromatin and fragment DNA, cells were lysed and then sonicated. To begin nuclear fractionation, cells were resuspended in 1 mL of Gagnon HLB Buffer<sup>179</sup> + 1X PIC and incubated on ice for 10 minutes. Samples were briefly vortexed and centrifuged at 1250g for 3 minutes at 4°C. Samples were then resuspended in 600  $\mu$ L of mammalian lysis buffer (1% TritonX-100, 0.1% sodium deoxycholate, 0.1% SDS, 150mM NaCl, and 50mM HEPES pH 7.5) + PIC and transferred to 15 mL Diagenode conical tubes. Following transfer, cells were sonicated using a Bioruptor waterbath sonicator (in 15 mL tubes with adapters) at max intensity for 30 seconds, followed by either 30 seconds of rest for 27 cycles (1% formaldehyde samples) or 30 seconds of rest for 36 cycles (1% formaldehyde + DSG samples). Samples were then transferred to 1.5 mL tubes and cleared of insoluble material by pelleting at 13000 rpm for 10 minutes at 4°C. Supernatants were mixed with 1800  $\mu$ L HBSS (Thermo Scientific) + 1X PIC and 2400  $\mu$ L 2X RIPA + 1X PIC. Lysate was incubated overnight with Invitrogen M-280 Sheep Anti-Rabbit IgG Dynabeads coupled with 5  $\mu$ g of H3K27me3 antibody (Active Motif, 39155) or 5  $\mu$ g of Goat anti-V5 antibody (Bethyl, A190-119A). Samples were washed with 1 mL of low salt buffer, high salt buffer, LiCl wash buffer, and finally TE buffers. Samples were reverse-crosslinked and Proteinase K-digested overnight at 65°C, and DNA was subsequently purified with Zymo Clean and Concentrate.

### RNA-FISH + immunostaining

Spy-tagged pSM33 cells were seeded at low density onto poly-D-lysine coated coverslips (Neuvitro H-12-1.5-pdl) for 6 hours prior to doxycycline administration (Sigma D9891-1G, 2 $\mu$ g/mL). After overnight Xist induction with 2  $\mu$ g/mL doxycycline, cells were fixed with 4% formaldehyde in PBS for 15 minutes at room temperature and the ViewRNA Cell Plus (Thermo Fisher Scientific, 88-19000-99) kit was used for immunofluorescence (IF) combined with *in situ* RNA visualization per the manufacturer's protocol. Stained coverslips were mounted onto slides using ProLong™ Gold Antifade Mountant with DAPI (Thermo Fisher Scientific, P36935). Imaging was performed using a Leica DMI 6000 Deconvolution Microscope with the Leica HC PL APO 63x/1.30 GLYC CORR CS2 objective. Images were projected with maximum projection (3  $\mu$ m; step size, 0.2  $\mu$ m).

Primary antibodies and the dilutions used are as follows: anti-V5 (Bethyl, A190-120A; 1:100 dilution), Goat anti-Rabbit IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor™ 568 (Invitrogen; Catalog #A-11011). For Xist labeling, Thermo Fisher Scientific FISH probe design ID: VB4-19746 was used.

### Dox expression of Halo-tagged proteins in HEK293T cells for ChIP

A doxycycline (Dox) inducible mammalian protein expression destination vector with an in-frame fusion of an N-terminal Halo tag and C-terminal V5 was used to generate expression clones for CTCF, YY1, and WDR5 from human cDNA clones. These plasmids were co-transfected with a plasmid expressing a reverse tetracycline-controlled transactivator and grown for 16 hours prior to induction with Dox. Protein expression was titrated with Dox concentration and pellets were collected after 6 hours of induction with either no Dox (leaky expression) or at a Dox concentration of 2  $\mu$ g/mL.

### Expression confirmation

Expression testing of Halo-tagged constructs was performed as previously described.<sup>77</sup> The same method of labeling and visualization was utilized to measure input, capture flowthrough, and protein loss from either CLIP- or CLAP-washes.

### Note on SPEN

SPEN is a 450 kilodalton protein that is challenging to resolve on a protein gel, a problem that we and others have previously observed.<sup>55,68,78,105</sup> Despite this, there are several lines of orthogonal evidence indicating that this Halo-SPEN fusion protein is correct. These include the fact that Halo-SPEN properly localizes to the inactive X upon induction of Xist, that the observed RNA binding profiles are comparable to those of endogenous HaloTag-integrated SPEN, and that this fusion protein can functionally compensate for loss of the endogenous protein in cell-based functional experiments.<sup>78</sup>

### CLAP on Halo-tagged SPEN

CLAP on pSM33 cells expressing Halo-V5-tagged SPEN was performed as previously described<sup>77</sup> with slight modifications. After overnight Xist induction with 2  $\mu$ g/mL doxycycline and UV-crosslinking, cell pellets were resuspended in 1 mL ice-cold Gagnon HLB buffer<sup>179</sup> with 1X PIC and RiboLock, mixed by pipetting, and incubated on ice for 10 minutes. Cells were then briefly vortexed and then centrifuged at 800g for 8 minutes at 4°C. Supernatant was removed (cytoplasmic fraction) and then 1 mL standard lysis buffer (50 mM HEPES pH 7.4, 100 mM NaCl, 1% NP-40, 0.1% SDS, 0.5% sodium deoxycholate) supplemented with 1X PIC and RiboLock was added. Nuclear pellet was then resuspended and incubated on ice for 10 minutes, followed by sonication (Branson Ultrasonics) for 30 seconds at 4W (0.7 seconds on, 2.3 seconds off). CLAP was then performed according to standard protocol.

### IR-CLIP

eCLIP was performed as previously described,<sup>71</sup> with slight modifications. 1 mL of HEK293T lysate (20 million cells) expressing tagged protein was digested with RNase If (NEB) at a fixed dilution of 1:500 and then immunoprecipitated overnight at 4°C with 5  $\mu$ g of antibody coupled to 40  $\mu$ L of Protein G beads. Rabbit anti-V5 antibody (Bethyl, A190-120A) was used for immunoprecipitation in all cases. After CLIP-washes and end-repair, RNAs were end-labeled using 1.5  $\mu$ L of 20  $\mu$ M pCp-IR680LT (Jena Bioscience) and ligated with High Concentration T4 RNA ligase I (NEB). Samples were then washed with additional CLIP-washes, then eluted and run on SDS-PAGE. Gels were then imaged directly on a LI-COR Odyssey. Corresponding western blots for each CLIP experiment were performed using a Goat anti-V5 (Bethyl, A190-119A), with the exception of Halo-tagged SAF-A, which was not recognized by the antibody. Instead, anti-HaloTag (Promega, G9211) was used for the SAF-A samples.

### Library construction and sequencing

#### CLIP samples

CLIP samples were treated as previously described.<sup>71,180</sup> Briefly, after immunoprecipitation and wash steps, RNA was dephosphorylated (FastAP), cyclic phosphates were removed (T4 PNK), and RNA was ligated on Protein G beads to an RNA adapter containing a RT primer binding site. The ligated protein-bound RNA was then run through a denaturing PAGE gel and transferred to nitrocellulose membrane (as described above). RNA was then extracted by Proteinase K and purified using a spin column (Zymo). RNA was then reverse transcribed into single stranded cDNA and subsequently degraded with NaOH. Following RT, a second adapter was ligated to the single stranded DNA. PCR amplification was achieved using primers that targeted the 3' and 5' ligated adapters.

#### CLAP samples and input RNA samples

50  $\mu$ L of lysate was taken prior to immunoprecipitation for input processing. Libraries were constructed using the same steps as outlined above, except the dephosphorylation, cyclic phosphate removal, and ligation were performed in solution rather than on Protein G beads. In the case of CLAP samples, all steps were performed on purified RNA from Proteinase K elution.

#### ChIP samples

Post reverse-crosslinking ChIP libraries were constructed as previously described.<sup>101</sup> Briefly, purified DNA was end repaired and dA-tailed using 1X NEBNext Ultra II End Repair/dA-Tailing Module (NEB, E7546L). DNA adaptors were then ligated to each sample and cleaned up using 0.7X SPRI (AMPure XP) followed by a repeat clean-up with 1X SPRI. PCR amplification was achieved using primers that add the indexed full Illumina adaptor sequences.

The molarity of PCR amplified libraries was measured by Agilent TapeStation High Sensitivity DNA Screentape, and all samples were pooled at equal molarity. The pool was then size-selected on a 2% agarose gel, cut between 150-700 nucleotides (CLIP/CLAP) or 280-1300 nucleotides (ChIP), and purified with Zymo Clean and Concentrate. The final libraries were measured by Agilent Bioanalyzer and Qubit dsDNA High Sensitivity assay (Thermo Fisher) to determine the loading density of the final pooled sample. Pooled samples were paired-end sequenced on either an Illumina HiSeq 2500 with read length  $\geq$  35 x 35 nucleotides or Illumina NextSeq 2000 with read length  $\geq$  50 x 50 nucleotides. Sequencing depth for each sample is reported in terms of raw read counts in Table S3.

### QUANTIFICATION AND STATISTICAL ANALYSIS

#### Read processing and alignment

##### CLIP/CLAP samples

Paired-end RNA sequencing reads were trimmed to remove adaptor sequences using Trim Galore! v0.6.2 and assessed with FastQC v0.11.8. Read pairs were then aligned to a combined genome reference containing the sequences of repetitive and structural RNAs (ribosomal RNAs, snRNAs, snoRNAs, 45S pre-rRNAs, tRNAs) using Bowtie2. The remaining reads were then aligned to a combined genome reference containing the mouse (mm10) and human (hg38) genomes using STAR aligner.<sup>172</sup> PCR duplicates were removed using the Picard MarkDuplicates function. For mixing experiments, only reads that mapped uniquely in the genome and unambiguously to the human or mouse genomes were kept for further analysis. For experiments done in a single species, the appropriate reference genome and alignments were used (mm10 for mouse and hg38 for human).

### ChIP samples

Paired-end DNA sequencing reads were aligned to the appropriate reference genome (mm9 for mouse and hg19 for human) using Bowtie2 v2.3.1<sup>173</sup> with the default parameters and with the following deviations. We used a local alignment search (`-local`) and disabled searching for discordant alignments (`-no-discordant`).

### Gene window enrichment calculations

All human (hg38) and mouse (mm10) annotated genes (RefSeq, downloaded from UCSC GRCh38 and GRCm38, respectively) were used as a reference set except for the genes encoding the transfected proteins. We treated exonic regions and intronic regions of each annotated gene as separate reference genes for computing enrichment. For each reference gene, we enumerated 100 nucleotide windows that span across the gene; for each window, we calculated: (i) the number of reads overlapping the window in the protein elution sample (e.g., CLIP or CLAP) and (ii) the maximum of either the number of observed reads over the window or the median read count over all windows within the gene in the input sample. Because all windows overlapping a gene should have the same expression level in the input sample, this approach provides a conservative estimation of the input coverage because it prevents windows from being scored as enriched if the input values over a given window are artificially low due to stochastic fluctuation, while at the same time accounting for any non-random issues that lead to increases in read counts over a given window (i.e., alignment artifacts leading to non-random assignment or pileups).

To directly compare the number of reads within each window between sample and input, we normalized each window count by the total number of reads sequenced and the overall complexity within each sample. For example, if one sample was sequenced twice as deeply as another, then we would expect to observe – on average – twice as many reads over a given window for that sample. To account for the differences in the number of mapped (aligned) reads between samples, we scaled the total number of sequenced reads by the proportion of aligned reads within each sample.

For each window, we computed enrichment by dividing the normalized sample counts by the normalized input counts. Nominal *p*-values were calculated for each window using a binomial test where *k* (number of successes) is defined as the number of reads in the protein elution samples within the window, *N* (number of trials) is the sum of the number of reads in the protein elution and input samples, and *p* (probability of success) is the expected number of reads in the elution sample divided by the sum of the expected number of reads per window in elution and input samples. (The expected number of reads is defined as the total number of reads scaled by the proportion of aligned reads within each sample). For plotting and reporting purposes, we considered all regions with a nominal binomial *p*-value < 10<sup>-6</sup> as significant. However, the overall results reported are robust to the precise *p*-value cutoff used.

### Analysis of +tag and -tag samples

We generally observed a higher level of detection within the human RNAs than mouse RNAs, which may reflect difference in UV-crosslinking efficiency between HEK293T cells and pSM33 cells or potential biases towards mapping to the human genome. Because of these biases, we performed the reciprocal experiments to ensure that the general trends for the specific RBPs in question are the same. To account for this bias and to compare the same RNAs, we focused our analysis on the human RNAs for the +/-tag experiments.

### Plotting and visualization

Data were visualized using the Integrative Genomics Visualizer<sup>171</sup> (IGV). IGV plots for specific RNAs were generated by computing enrichments (as described above) across 1 nucleotide windows and the enrichment value was plotted at the midpoint of each window. Single reference transcripts were chosen for display based on the “knownCanonical” annotation in GENCODE V40. All transcripts are shown left-to-right as 5' to 3' unless otherwise noted. Scatter plots depicting CLIP or CLAP enrichments were merged by replicates to ensure that lack of detection was not due to low sequencing coverage. Aggregate scatter plots were shown for all proteins within the PRC2 complex as well as individual scatter plots for each individual PRC2 component (EED, EZH2, SUZ12). Transcript feature plots were generated by counting all reads that mapped within exons, introns, 5' untranslated, or 3' untranslated regions defined by RefSeq. lncRNA annotation files were derived from GENCODE vM10 (mouse) and v44 (human).

### Reproducibility of CLIP/CLAP replicates

For all indicated CLIP or CLAP replicates, experiments were performed independently from beginning to end, and thus represent independent biological replicates.

To measure the reproducibility between CLIP and CLAP replicate samples, we computed the number of reads within each 100-nucleotide window within each sample and plotted these counts between individual replicates. Pearson's correlation coefficients were computed for each set of replicate experiments.

### Quantification of <sup>32</sup>P gels

Full-length <sup>32</sup>P gel images were imported into ImageJ<sup>176</sup> for quantification. For each gel lane, the straight-line tool was used to measure the <sup>32</sup>P intensity of the entirety of the lane, which was then converted to an ROI. The “Plot Profile” tool was used to retrieve x,y coordinates where the x-axis represented distance and the y-axis represented intensity values. The x-axis values were then scaled to expected molecular weights based on the sizing of the protein ladder on the same gel.

### Crosslink-induced truncation analysis

We computed all significant PTBP1 binding sites from our CLIP data. We randomly sampled 65,000 regions from this list and defined the location of the known PTBP1 motif (HYUUUYU) within each region using the “Find Individual Motif Occurrences” (FIMO,<sup>175</sup> MEME suite). We used the position of each identified motif occurrence to center the peak at the motif location and computed the number of crosslink-induced truncation sites along the peak from 100 nucleotides downstream to 100 nucleotides upstream. We then plotted the positional counts of these crosslink-induced truncations at each position.

### Definitions

Reads mapping to nascent pre-mRNAs are defined as paired sequencing reads with at least one read in the pair aligning in part within an intron (i.e., between an exon and intron, or exclusively within an intron). “Focal binding sites” are defined as enrichment observed within a small window (100 nucleotides) relative to the remainder of the RNA. We use this term to contrast with “broad” or “promiscuous” binding for RNA-binding profiles such as those of SAF-A, which are often enriched over the entire length of a transcript.

### Peak calling

Significant peaks for downstream analysis were computed from windows (as described above) and filtered on 100-nt windows based on meeting all of the following criteria: (i) containing at least 5 reads in the elution sample, (ii)  $p$ -value  $< 10^{-3}$ , and (iii) minimum enrichment of 3-fold above the input sample. The full tables for all CLIP or CLAP experiments, including all gene windows, enrichments, sample counts, input counts, and  $p$ -values are available for download at Mendeley Data: <https://doi.org/10.17632/wmsbzy6kg5.2>.

### Mean coverage of binding sites

Visualization of binding site coverage across transcript features was performed using the RCAS tool<sup>174</sup> (version 1.26.0) using R software (version 4.3.1).

### Gene Ontology enrichment

Genes overlapping filtered CLAP peaks were used as input IDs for GO annotation analysis (biological process complete). All human genes in database were used for the reference list. Tables containing enrichment values and statistics were downloaded from the PANTHER classification system (18.0) website directly. Binomial  $p$ -values are reported with Bonferroni correction for multiple testing.

### Motif enrichment analysis

Filtered CLAP peaks were used as input for de novo motif analysis by HOMER (<http://homer.ucsd.edu/homer/>) using options -rna -len 6. Motifs with a reported  $p$ -value  $< 10^{-40}$  were considered significant. Searching background (-bg) was set as all human transcripts.