Defining the Universe of Functional RNA-Protein Interactions

Thesis by
Jimmy Kang Guo

In Partial Fulfillment of the Requirements for
the degree of
Doctorate of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2024
(Defended April 24, 2024)

Jimmy Kang Guo
ORCID: 0000-0002-7211-4117

ACKNOWLEDGEMENTS

I have had the incredible fortune of working with and learning from so many talented and wonderful people at Caltech.

First and foremost, I would like to thank my advisor, Mitch Guttman, for taking a chance on me all those years ago. I met Mitch during my first year of medical school and expressed my intent to work in his lab, even though I had no ties to the PhD program at the time. Without hesitation, he quickly set me up with a mentor and a project. Even though I was only part-time, Mitch always found time to meet with me for anything I needed and ended up as my biggest advocate for being admitted into the MD-PhD program. Mitch's unwavering support continued throughout my PhD, both on a professional and personal level. I think it is truly rare to see an advisor care so deeply for each of his trainees and their individual success. I wouldn't be half the scientist I am today without Mitch's mentorship, trust, and support.

Next, I would like to thank Mario Blanco, my mentor from day one. I first started off in the lab under Mario's guidance, as he patiently worked around my strange hours when I was a part-time medical student. I have much to thank him for – for teaching me practically everything I know about molecular biology, for the foundational work he put in for the CLAP and SPIDR projects, and for always keeping me extremely well-fed whenever he welcomed me to his home. Mario has been one of my greatest and most steadfast supporters. I wouldn't have gotten through my PhD without him, and I am honored to call him a close friend.

I would also like to acknowledge my thesis committee – David Baltimore, Shasha Chong, Ryan Flynn, and Rebecca Voorhees – who have provided valuable input throughout committee meetings and discussions. In particular, I'd like to thank Ryan not only for his expert insight on CLIP and all things RNA, but also for the advice he gave me many years ago on starting research during medical school and applying into the MD-PhD program.

My time in the Guttman lab has truly been special and I extend my thanks to all my fellow labmates. Special thanks to Carl Urbinati, who despite being a full-time professor at LMU,

made the long commute every week to help out on projects and accompanied me on many Red Door trips for Arnold Palmers. I am also grateful to Abhik Banerjee, who always looked out for me with his everlasting enthusiasm, positivity, and vested interest in my training and success. I thank my desk mate and friend, Olivia Ettlin, for all the laughs and antics we shared, even if they occasionally tested the patience of our labmates. Many thanks to Shawna Hiley for all her editing help on our manuscripts and grants. I also thank Inna-Marie Strazhnik for her wonderful illustrations seen in all chapters of this thesis, and for teaching me how to make figures myself.

For the SPIDR project, I wish to thank members of the Jovanovic lab for their help with the mass spectrometry work, the Querido lab for their extensive efforts in characterizing LARP1, showcased in Chapter 4, Andrew Perez for developing the bead-barcoding method, and Isabel Goronzy for computational help.

Looking back, I would also like to express my gratitude to my prior research mentor, Howard Chang, and members of his lab. My experience in Howard's lab solidified my interest in RNA, and I received much formative life advice there from countless Chang lab members. I especially want to thank Chun-Kan Chen, with whom I had the pleasure of briefly overlapping with and who ultimately recommended me to Mitch's lab in the first place.

On a more personal note, I would like to thank my friends, family, and partner, all of whom I could not imagine surviving graduate school without. To my close friends from medical school who have supported me since the start of this journey – Franklin, Emily, Aditi, and Zack – thank you for always lending an open ear to both my triumphs and struggles. To my family – my parents, Song and Chunfang, and my late cat, Mimi – for their endless love and encouragement. They have supported me since the very start of my scientific journey and made countless sacrifices to ensure my well-being and success. And finally, to Kat, for her enduring love, support, and understanding. Thank you for putting up with me all these years, and for all the fond memories we've shared.

ABSTRACT

RNA has been proposed to mediate many central mechanisms of cell biology, including protein recruitment to chromatin, genome structure organization, and gene expression. In most cases, these critical functions have been widely attributed to the proteins to which RNAs bind. One paradigm example of this is the Xist long non-coding RNA, which complexes with many distinct proteins to orchestrate X-chromosome inactivation. Beyond Xist, there are many critical non-coding RNAs (ncRNAs) that are not yet functionally characterized because we lack information on what proteins they bind to. In this thesis, Chapter 1 discusses the growing gap between the vast potential of ncRNA functions and what has been demonstrated to be functionally meaningful. We highlight critical discrepancies between biochemical evidence supporting specific RNA-protein interactions and genetic evidence demonstrating the same interactions are often dispensable for function. Chapter 2 explores previously reported RNA-protein interactions for many chromatin proteins (i.e., PRC2, CTCF, etc.), demonstrating that they do not represent *bona fide* interactions in cells. We present Covalent Linkage Affinity Purification (CLAP), a method that employs denaturing purification of RNA-protein complexes, showing that CLAP accurately removes false signals that do not occur *in vivo*, while retaining known RNA-protein interactions. Chapter 3 details a highly multiplexed method of mapping RBPs and their *in vivo* binding sites across dozens to hundreds of targets within a single experiment. We present Split and Pool Identification of RBP targets (SPIDR), which enables the rapid, *de novo* discovery of RNA-protein interactions at an unprecedented scale and separates *bona fide* RBPs from non-RBPs. Using SPIDR, we uncover a previously unknown LARP1 binding site on the 18S ribosomal RNA that is directly adjacent to the mRNA entry channel, which may explain how LARP1 achieves translational control of sequence-specific mRNAs. Finally, Chapter 4 proposes new experimental and analytical approaches to evaluate the potentially wide universe of ncRNA-protein functions at scale. Together, these results provide a comprehensive framework for evaluating RNA-protein interactions and underscore the growing importance of RNA-mediated functions in cell biology.

PUBLISHED CONTENT AND CONTRIBUTIONS

1. Guo, J.K., and Guttman, M. (2022). Regulatory non-coding RNAs: everything is possible, but what is important? Nat Methods 19, 1156–1159. 10.1038/s41592-022-01629-6.

   J.K.G. wrote the review with M.G.

2. Guo, J.K.*, Blanco, M.R.*, Walkup, W.G., Bonesteele, G., Urbinati, C.R., Banerjee, A.K., Chow, A., Ettlin, O., Strehle, M., Peyda, P., et al. (2024). Denaturing purifications demonstrate that PRC2 and other widely reported chromatin proteins do not appear to bind directly to RNA in vivo. Mol Cell 84, 1271-1289.e12. 10.1016/j.molcel.2024.01.026.
   *Equal contribution

   J.K.G. performed CLAP experiments, analyzed data, interpreted results, generated figures, and wrote the manuscript with M.R.B. and M.G.

3. 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. bioRxiv, 2023.06.05.543769. 10.1101/2023.06.05.543769.
   *Equal contribution

   J.K.G. developed the SPIDR method, performed SPIDR experiments, analyzed data, interpreted results, and wrote the manuscript with M.J. and M.G.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1
Number *Page*

Chapter 2
Number *Page*

Chapter 3
Number *Page*

Chapter 4
Number *Page*

C h a p t e r 1

REGULATORY NON-CODING RNAS: EVERYTHING IS POSSIBLE, BUT WHAT IS IMPORTANT?

J.K. Guo and M. Guttman

A modified version of this chapter was published as “Regulatory non-coding RNAs: everything is possible, but what is important?” in *Nature Methods* 19: 1156–1159. doi: 10.1038/s41592-022-01629-6.

1.1 ABSTRACT

In recent years, the number of annotated non-coding RNAs (ncRNAs) and RNA-binding proteins (RBPs) has increased dramatically. The wide range of RBPs identified highlights the enormous potential for RNA in virtually all aspects of cell biology, from transcriptional regulation to metabolic control. Yet, there is a growing gap between what is possible and what has been demonstrated to be functionally important. Here we highlight recent methodological developments in studying RNA-protein interactions, discuss the challenges and opportunities for exploring their functional roles, and provide our perspectives on what is needed to bridge the gap in this rapidly expanding field.

1.2 INTRODUCTION

Mammalian genomes encode thousands of non-coding RNAs (ncRNAs), with ~20,000 annotated long ncRNA genes, a number that rivals and may ultimately exceed the total number of protein-coding genes in the human genome[1]. Yet, most ncRNAs remain functionally uncharacterized and the diversity of biological roles they play are largely unexplored. Identifying what proteins interact with a ncRNA can provide critical insights into its possible functions and mechanisms, enabling formation of experimentally testable hypotheses. For example, mapping various protein interactions to ncRNAs have led to proposed models whereby specific ncRNAs can: **(i)** guide regulatory proteins to specific genomic DNA sites[2,3], **(ii)** tether multiple protein components into macromolecular complexes[4,5], **(iii)** mediate and stabilize three-dimensional chromatin loops[6–8], **(iv)** activate[9] or inhibit[10] specific enzymatic functions, and **(v)** compete proteins away from their mRNA[11] or genomic DNA targets[12–14].

Over the past decade, the development of global RNA-centric proteomics methods (**Figure 1a**), such as RBR-ID[15] and RBDmap[16], have enabled unprecedented exploration of which proteins bind to RNA. These efforts have vastly expanded the number of identified RBPs,

with >4,000 human proteins (~20% of the human proteome) currently annotated as "RNA-binding" by UniProt[17]. These RBPs include many that lack canonical RNA-binding domains, such as RRMs or KH domains, and encompass critical chromatin and transcriptional regulators, nuclear structure proteins, and metabolic enzymes[15,16]. The large number of putative RBPs representing such a diverse functional spectrum suggests a vast potential for regulatory ncRNA function.

Despite this, it remains mostly unknown how many of these RBPs interact with ncRNAs, and which specific ncRNAs they might associate with. Typically, defining the RNAs that proteins bind *in vivo* requires protein-centric approaches, such as CLIP[18] (**Figure 1a**). When paired with high-throughput sequencing[19,20], CLIP can comprehensively define specific sites on RNAs that interact with a protein of interest. This approach utilizes UV-light to create a covalent photo-crosslink between a protein and its bound RNAs, but not between pairs of proteins. Because these crosslinks can be formed in a living cell, a specific protein can be purified – usually via an antibody – under stringent washing conditions that disrupt non-crosslinked RNA-protein and protein-protein interactions. However, because CLIP can only map a single protein at a time, it is technically challenging to explore the thousands of annotated RBPs. Additionally, CLIP relies on high-quality antibodies, which are not always available – especially for non-canonical RBPs. Accordingly, there have been limited efforts to map most of these proteins to specific RNAs. Moreover, even in cases where interactions between specific non-canonical RBPs and RNAs have been identified, the functional relevance of these interactions have been questionable.

1.3 A CAUTIONARY TALE FROM XIST AND PRC2

The Xist lncRNA represents a valuable case study illustrating some of the practical challenges in deciphering ncRNA-protein interactions and function. Briefly, Xist is required for initiating chromosome-wide transcriptional silencing on the X chromosome to balance gene expression between male (XY) and female (XX) mammals[21]. Although Xist was first

identified in 1991[22], the molecular components required for initiation of chromosome-wide silencing were not identified until 2015[23–25].

In the intervening years, extensive characterization of Xist showed that: **(i)** Xist coats the inactive X[26], **(ii)** Xist is sufficient to initiate transcriptional silencing on the X[27], **(ii)** initiation of Xist corresponds to accumulation of PRC2 and its associated H3K27me3 repressive mark over the inactive X[28], **(iii)** the A-repeat region of Xist is required for transcriptional silencing[29], and **(iv)** the A-repeat of Xist interacts with PRC2[30]. Because PRC2 was known to be involved in transcriptional silencing in other contexts[31], this led to a model where Xist binds directly to PRC2 via the A-repeat to silence transcription (**Figure 1b**).

While this model seemingly explained these observations, there was a problem: deletion of PRC2 did not impact Xist-mediated transcriptional silencing[24,32,33] (**Figure 1b**). Because Xist-PRC2 interactions were identified using either *in vitro* measurements[30] or native RIP[34], they might represent in-solution associations rather than *bona fide* interactions that occur *in vivo*. In a classic experiment, Mili and Steitz showed that native IP methods can identify RNA-protein interactions that could not have occurred *in vivo*[35]. Similarly, mammalian PRC2 was shown to interact with bacterial RNAs with comparable affinity to that of other mammalian RNAs, including the A-repeat[14].

In fact, PRC2 has been reported to bind promiscuously to virtually all RNAs and the biological significance of this remains a topic of debate[36]. While most studies of RNA interactions with PRC2 rely on *in vitro* measurements and RIP, there have been recent attempts to map PRC2 using CLIP[12,13,37]; these reported further evidence of promiscuous PRC2 binding to RNA, including to Xist. However, while experimentally stringent, these studies highlight another critical issue related to the analysis of CLIP data. It is well-documented that the number of reads mapping to an RNA is proportional to its overall abundance and therefore simply identifying reads does not indicate binding[19,38]. For this reason, it is not possible to distinguish between promiscuous binding of a protein to all RNAs from the lack of binding to any RNAs. In addition**,** because of the low efficiency of UV-crosslinking and stringency of CLIP, the complexity of the underlying sequencing library is

often extremely low, leading to read pileups at specific locations due to PCR duplications. Indeed, many of the reported interactions between PRC2 and specific RNA regions in CLIP experiments appear to be PCR duplicates rather than enrichment of true binding events[37].

Consistent with the idea that the Xist-PRC2 association might not represent an *in vivo* binding event, several studies purified Xist using different *in vivo* crosslinking strategies coupled with high stringency washes and mass spectrometry[23–25]. None of these methods identified an association between Xist and any previously reported PRC2 components. In contrast, these studies all independently identified SPEN (also known as SHARP)[23–25], a transcriptional co-repressor. Follow-up studies have demonstrated that SPEN is required for Xist-mediated transcriptional silencing in cell-based models[23,24] and in early development[39]. SPEN has been shown to bind to the A-repeat region of Xist via CLIP[40,41], congruent with the finding that Xist lacking the A-repeat cannot silence transcription[29] (**Figure 1b**).

Importantly, the discrepancy between biochemical evidence supporting specific RNA-protein interactions and genetic evidence demonstrating that these same interactions are often dispensable for function is not limited to Xist and PRC2. For example, recent evidence indicates that PRC2 is dispensable for HOTAIR-mediated gene silencing[42], even though it was initially reported to bind to PRC2 using RIP[5,43]. Similarly, the YY1 transcriptional regulator was reported to bind to Xist to tether the RNA to chromatin[44], yet neither deletion of the YY1 protein[24] nor deletion of the reported YY1 binding site from Xist (F-repeat)[29] impacts the localization of Xist to chromatin or Xist-mediated transcriptional silencing. Consistent with this, stringent purification of Xist followed by mass spectrometry failed to identify YY1 as an Xist binding protein[23,24].

These examples highlight the practical issues associated with deleting an identified "binding site" as evidence supporting the functional role of an RNA-protein interaction (**Figure 1a**). Specifically, deletion of a binding site on an RNA may result in phenotypic effects due to disruption of a different protein (e.g. SPEN rather than PRC2 to the A-repeat). Similar issues may occur when disrupting the RNA-binding region of a protein, which could impact its overall structure and other essential functions. For example, deletion of the RNA-binding

region of CTCF impacts formation of chromatin loops[6,7], yet because it overlaps a zinc finger motif (a known DNA-binding motif), it is unclear if the observed impacts are solely due to RNA-binding. Because of these potential issues, alternative approaches that directly test the importance of the RNA-protein interaction are critical. One way to do this is by reconstituting the RNA-protein interaction via a synthetic fusion upon disruption of the RNA binding region and/or protein binding site and measuring whether this can rescue the expected phenotype (**Figure 1a**). For example, synthetically tethering an RNA-binding mutant of SPEN directly to Xist was shown to rescue transcriptional silencing on the X[39] (**Figure 1b**).

1.4 BRIDGING THE GAP BETWEEN DISCOVERY AND FUNCTION

Many additional non-canonical RBPs such as metabolic enzymes (e.g. ENO1[45]) and various chromatin complexes including DNA methylation enzymes (e.g. DNMT1[10], TET2[15]), repressive (e.g. PRC1[46]) and activating (e.g. WDR5[47]) chromatin modifiers, transcription factors (e.g. SOX2[48]), and three-dimensional DNA structure proteins (e.g. CTCF[6,7]) have been reported to bind to RNAs. Based on these observations, chromatin regulators have emerged as central players in the mechanisms by which ncRNAs regulate gene expression (**Figure 1c**). While an attractive model due to the intrinsic high local concentrations that ncRNAs can form in the nucleus[3], the functional importance of RNA binding in chromatin regulation remains untested in most cases. As the number of proteins reported to bind to RNA continues to increase, we are faced with a growing chasm between the potential for what ncRNAs can do and the reality of what functional roles they play.

Motivated by the lessons learned from the examples discussed above, we propose a comprehensive framework – including new experimental methods – that will be useful for bridging this gap. This framework consists of: **(i)** Stringent experimental methods to define high confidence RNA-protein interactions – including high stringency and/or denaturing purification for RNA-centric proteomic discovery and protein-centric RNA mapping. **(ii)** Scalable methods that can characterize the large numbers of putative RBPs. This will require

development of new tools that utilize the stringency and binding site precision of CLIP, but with dramatically improved throughput. Moreover, we anticipate needing additional affinity reagents or alternative purification strategies to map proteins that are currently inaccessible via existing antibodies. **(iii)** Rigorous computational and statistical methods to identify meaningful regions of RNA binding that account for abundance, complexity, and other sources of artifacts. **(iv)** Quantitative measurements of protein and RNA binding affinity[49] and occupancy frequency[50] in living cells. Such approaches will enable more precise characterization of true binding events through establishing quantitative criteria, including accurate measurements of potentially promiscuous RBP interactions. **(v)** Precise functional characterization of an RNA-protein interaction through targeted disruption of the interaction and rescue through reconstitution.

With a reliable framework such as this, we anticipate being able to define classes of ncRNA and protein functions to fully understand the scale and scope of ncRNA-mediated functions. This information will allow us to explore what intrinsic properties of RNA make it such a widespread and versatile molecular regulator. Moreover, it will allow us to begin to address more global questions, such as why a large fraction of the human proteome has evolved to bind to RNA, and why the genome encodes so many distinct ncRNA species.

1.5 FIGURES

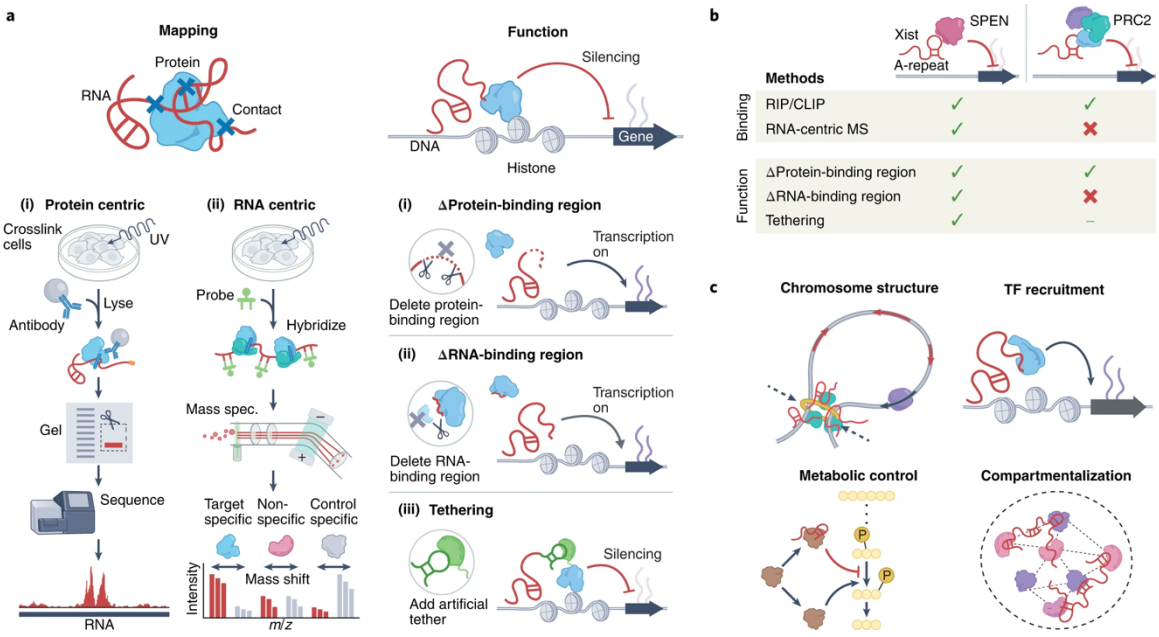


Figure 1: Identification and functional characterization of RNA-protein interactions.

(a) Left: methods for mapping RNA-protein interactions *in vivo* using UV crosslinking, including: (i) protein-centric approaches in which a specific protein is selectively purified and its associated RNAs are mapped using high-throughput sequencing; and (ii) RNA-centric approaches where a specific RNA or set of RNAs are selectively purified and the bound proteins are identified by mass spectrometry. Right: experimental framework for dissecting the functional relevance of an RNA-protein interaction illustrated by a ncRNA-protein complex that acts to repress transcription. Schematic of methods to disrupt an RNA-protein interaction via: (i) deletion (Δ) of a protein-binding region on the RNA; or (ii) deletion (Δ) of the RNA-binding region on the protein; and (iii) rescuing a phenotype through synthetically tethering the effector protein to the RNA. **(b)** Experimental evidence for (green tick) or against (red cross) the functional interaction between Xist and either SPEN (left) or PRC2 (right). **(c)** A range of proposed roles for RNA-mediated regulation of cellular processes, including mediation of three-dimensional DNA structure, recruitment of transcription factors (TFs) to genomic sites, feedback inhibition of metabolic pathways, and

subcellular compartmentalization of proteins and RNA. *m*/*z*, mass-to-charge ratio; P, phosphate group.

1.6 REFERENCES

1. Frankish, A. *et al.* GENCODE 2021. *Nucleic Acids Res* **49**, D916–D923 (2021).
2. Rinn, J. L. & Chang, H. Y. Genome regulation by long noncoding RNAs. *Annu Rev Biochem* **81**, 145–166 (2012).
3. Quinodoz, S. A. *et al.* RNA promotes the formation of spatial compartments in the nucleus. *Cell* **184**, 5775-5790.e30 (2021).
4. Kaneko, S. *et al.* Interactions between JARID2 and noncoding RNAs regulate PRC2 recruitment to chromatin. *Mol Cell* **53**, 290–300 (2014).
5. Tsai, M. C. *et al.* Long noncoding RNA as modular scaffold of histone modification complexes. *Science (1979)* **329**, 689–693 (2010).
6. Hansen, A. S. *et al.* Distinct Classes of Chromatin Loops Revealed by Deletion of an RNA-Binding Region in CTCF. *Mol Cell* **76**, 395-411.e13 (2019).
7. Saldaña-Meyer, R. *et al.* RNA Interactions Are Essential for CTCF-Mediated Genome Organization. *Mol Cell* **76**, 412-422.e5 (2019).
8. Mumbach, M. R. *et al.* HiChIRP reveals RNA-associated chromosome conformation. *Nat Methods* **16**, 489–492 (2019).
9. Zovoilis, A., Cifuentes-Rojas, C., Chu, H.-P., Hernandez, A. J. & Lee, J. T. Destabilization of B2 RNA by EZH2 Activates the Stress Response. *Cell* **167**, 1788-1802.e13 (2016).
10. di Ruscio, A. *et al.* DNMT1-interacting RNAs block gene-specific DNA methylation. *Nature* **503**, 371–376 (2013).
11. Lee, S. *et al.* Noncoding RNA NORAD Regulates Genomic Stability by Sequestering PUMILIO Proteins. *Cell* **164**, 69–80 (2016).
12. Kaneko, S., Son, J., Shen, S. S., Reinberg, D. & Bonasio, R. PRC2 binds active promoters and contacts nascent RNAs in embryonic stem cells. *Nat Struct Mol Biol* **20**, 1258–1264 (2013).
13. Beltran, M. *et al.* The interaction of PRC2 with RNA or chromatin is mutually antagonistic. *Genome Res* **26**, 896–907 (2016).

14. Davidovich, C., Zheng, L., Goodrich, K. J. & Cech, T. R. Promiscuous RNA binding by Polycomb repressive complex 2. *Nat Struct Mol Biol* **20**, 1250–1257 (2013).
15. He, C. *et al.* High-Resolution Mapping of RNA-Binding Regions in the Nuclear Proteome of Embryonic Stem Cells. *Mol Cell* **64**, 416–430 (2016).
16. Castello, A. *et al.* Comprehensive Identification of RNA-Binding Domains in Human Cells. *Mol Cell* **63**, 696–710 (2016).
17. Bateman, A. *et al.* UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res* **49**, D480–D489 (2021).
18. Ule, J. *et al.* CLIP Identifies Nova-Regulated RNA Networks in the Brain. *Science (1979)* **302**, 1212–1215 (2003).
19. van Nostrand, E. L. *et al.* Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat Methods* **13**, 508–514 (2016).
20. Zarnegar, B. J. *et al.* irCLIP platform for efficient characterization of protein–RNA interactions. *Nat Methods* **13**, 489–492 (2016).
21. Heard, E. & Disteche, C. M. Dosage compensation in mammals: fine-tuning the expression of the X chromosome. *Genes Dev* **20**, 1848–1867 (2006).
22. Brockdorff, N. *et al.* Conservation of position and exclusive expression of mouse Xist from the inactive X chromosome. *Nature* **351**, 329–331 (1991).
23. Chu, C. *et al.* Systematic Discovery of Xist RNA Binding Proteins. *Cell* **161**, 404–416 (2015).
24. McHugh, C. A. *et al.* The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature* **521**, 232–236 (2015).
25. Minajigi, A. *et al.* A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation. *Science (1979)* **349**, (2015).
26. Clemson, C. M., McNeil, J. A., Willard, H. F. & Lawrence, J. B. XIST RNA paints the inactive X chromosome at interphase: evidence for a novel RNA involved in nuclear/chromosome structure. *Journal of Cell Biology* **132**, 259–275 (1996).
27. Penny, G. D., Kay, G. F., Sheardown, S. A., Rastan, S. & Brockdorff, N. Requirement for Xist in X chromosome inactivation. *Nature* **379**, 131–137 (1996).

28. Plath, K. *et al.* Role of Histone H3 Lysine 27 Methylation in X Inactivation. *Science (1979)* **300**, 131–135 (2003).
29. Wutz, A., Rasmussen, T. P. & Jaenisch, R. Chromosomal silencing and localization are mediated by different domains of Xist RNA. *Nat Genet* **30**, 167–174 (2002).
30. Zhao, J., Sun, B. K., Erwin, J. A., Song, J.-J. & Lee, J. T. Polycomb Proteins Targeted by a Short Repeat RNA to the Mouse X Chromosome. *Science (1979)* **322**, 750–756 (2008).
31. Margueron, R. & Reinberg, D. The Polycomb complex PRC2 and its mark in life. *Nature* **469**, 343–349 (2011).
32. Kalantry, S. & Magnuson, T. The Polycomb Group Protein EED Is Dispensable for the Initiation of Random X-Chromosome Inactivation. *PLoS Genet* **2**, e66 (2006).
33. Schoeftner, S. *et al.* Recruitment of PRC1 function at the initiation of X inactivation independent of PRC2 and silencing. *EMBO J* **25**, 3110–3122 (2006).
34. Zhao, J. *et al.* Genome-wide Identification of Polycomb-Associated RNAs by RIP-seq. *Mol Cell* **40**, 939–953 (2010).
35. Mili, S. & Steitz, J. A. Evidence for reassociation of RNA-binding proteins after cell lysis: Implications for the interpretation of immunoprecipitation analyses. *RNA* **10**, 1692–1694 (2004).
36. Davidovich, C. *et al.* Toward a Consensus on the Binding Specificity and Promiscuity of PRC2 for RNA. *Mol Cell* **57**, 552–558 (2015).
37. Rosenberg, M. *et al.* Motif-driven interactions between RNA and PRC2 are rheostats that regulate transcription elongation. *Nat Struct Mol Biol* **28**, 103–117 (2021).
38. Hafner, M. *et al.* CLIP and complementary methods. *Nature Reviews Methods Primers* **1**, 20 (2021).
39. Dossin, F. *et al.* SPEN integrates transcriptional and epigenetic control of X-inactivation. *Nature* **578**, 455–460 (2020).
40. Chen, C.-K. *et al.* Xist recruits the X chromosome to the nuclear lamina to enable chromosome-wide silencing. *Science (1979)* **354**, 468–472 (2016).
41. Lu, Z. *et al.* Structural modularity of the XIST ribonucleoprotein complex. *Nat Commun* **11**, 6163 (2020).

42. Portoso, M. *et al.* PRC2 is dispensable for HOTAIR-mediated transcriptional repression. *EMBO J* **36**, 981–994 (2017).
43. Rinn, J. L. *et al.* Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* **129**, 1311–1323 (2007).
44. Jeon, Y. & Lee, J. T. YY1 Tethers Xist RNA to the Inactive X Nucleation Center. *Cell* **146**, 119–133 (2011).
45. Huppertz, I. *et al.* Riboregulation of Enolase 1 activity controls glycolysis and embryonic stem cell differentiation. *Mol Cell* **82**, 2666-2680.e11 (2022).
46. Guttman, M. *et al.* lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature* **477**, 295–300 (2011).
47. Wang, K. C. *et al.* A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* **472**, 120–124 (2011).
48. Holmes, Z. E. *et al.* The Sox2 transcription factor binds RNA. *Nat Commun* **11**, 1805 (2020).
49. Porter, D. F. *et al.* easyCLIP analysis of RNA-protein interactions incorporating absolute quantification. *Nat Commun* **12**, 1569 (2021).
50. Sharma, D. *et al.* The kinetic landscape of an RNA-binding protein in cells. *Nature* **591**, 152–156 (2021).

C h a p t e r 2

DENATURING PURIFICATIONS DEMONSTRATE THAT PRC2 AND OTHER WIDELY REPORTED CHROMATIN PROTEINS DO NOT APPEAR TO BIND DIRECTLY TO RNA IN VIVO

Jimmy K. Guo*, Mario R. Blanco*, Ward G. Walkup IV, Grant Bonesteele, Carl R. Urbinati, Abhik K. Banerjee, Amy Chow, Olivia Ettlin, Mackenzie Strehle, Parham Peyda, Enrique Amaya, Vickie Trinh, and Mitchell Guttman

*Equal contribution

A modified version of this chapter was published as "Denaturing purifications demonstrate that PRC2 and other widely reported chromatin proteins do not appear to bind directly to RNA *in vivo*" in *Molecular Cell*. doi: 10.1016/j.molcel.2024.01.026.

2.1 ABSTRACT

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.

2.2 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[1–4], 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[5–8] 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[9–17].

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[18,19]. PRC2 components have been reported to bind broadly to many RNAs, including lncRNAs and mRNAs[20–22]. These observations have led to the proposal that many canonical PRC2 functions are mediated by its interactions with RNA[5,20,23,24], including PRC2 recruitment to genomic DNA sites[11,23,25–28], tethering of PRC2 components[20], and activation of PRC2 enzymatic activity[29]. Additional proposals include that PRC2 binding to nascent pre-mRNAs precludes binding to DNA at active genes[5,21,22] and that PRC2 can act as a nuclease to degrade specific RNAs[30]. 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)[31–33], PRC1 components (e.g., RING1, CBX7)[16,34], trithorax components (e.g., WDR5),[32,35] transcription factors (e.g., SOX2)[36–38], and chromatin structure proteins (e.g., YY1 and CTCF)[6–8,39–41].

The paradigm example for the functional relevance of chromatin protein-RNA interactions is the Xist lncRNA[9,42]. 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[43,44]) and mediating chromosome-wide silencing[43,44]. PRC2 was reported to bind to the A-repeat of Xist[9,45], a region that is required for Xist-mediated silencing[46]. This led to a model whereby Xist directly binds to PRC2 and recruits this repressive chromatin complex to the X to mediate silencing[9,42,47]. Other lncRNAs, such as HOTAIR, were similarly reported to directly bind PRC2 to silence transcription[10,48,49].

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[50,51], and deletion of the A-repeat from Xist does not prevent PRC2 recruitment to, or H3K7me3 accumulation on, the X[52–55]. Similarly, deletion of PRC2 does not impact HOTAIR-mediated silencing[52,56,57]. 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[14], 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[46,55]. 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[5,58,59], and RNA immunoprecipitation (RIP), which utilizes native purification conditions in crosslinked or non-crosslinked cells to immunoprecipitate a protein and measure its associated RNAs[60,61]. 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*[62].

More recently, several of the interactions detected by RIP[5,15] have been confirmed using crosslinking and immunoprecipitation (CLIP) methods[7,8,20,21]. 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[63–67]. 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[63].

However, several observations suggest that many of the RNA-chromatin protein interactions reported, including by CLIP, may not represent interactions that occur *in vivo*: (i) 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[55,68,69]. (ii) 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[5,58,59]. (iii) 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[21,22]. (iv) 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[21,22,70]. (v) 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[7].

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*.

2.3 RESULTS

2.3.1 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[62] – 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 eCLIP (a specific implementation of CLIP[71]) 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[15–17,32,72], CLIP[21,22], and *in vitro*[5,45,59] experiments. We confirmed that the tagged proteins are (i) well-expressed (**Figure S1A, S1B**), (ii) specifically purified using a V5 antibody (**Figure S1C**), (iii) properly incorporated into the endogenous PRC2 complex (**Figure S1D**), and (iv) retain their RNA binding activity *in vitro* (**Figure S2**). Consistent with previous observations[21,22], we observed that radioactive isotope (^{32}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 to 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 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, $p<10^{-6}$, **Figure 1D, 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[73], and NEAT1[74] (**Figure 1B, S4C**). In addition, we observed strong binding to NORAD, a lncRNA that is predominantly localized in the cytoplasm (**Figure S4C**)[75,76].

Surprisingly, we also observed significant PRC2 binding in the -tag samples with >850 RNAs containing significantly enriched binding sites for all 3 PRC2 components (**Figure 1E, 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, 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[9,15] (**Figure 1B, 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 (**Figure 1C, 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*.

2.3.2 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 (i) the captured protein may associate with other proteins that are crosslinked to RNA or (ii) other proteins that are crosslinked to RNA may be retained after immunoprecipitation (**Figure 2A**). Additionally, we considered potential RNA-dependent sources of background where (iii) the captured protein may interact with RNA that is crosslinked to another RBP or (iv) 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)[77,78] (**Figure 2B**). CLAP, like other methods that utilize covalent linkage[79–81], integrates an epitope tag that enables covalent coupling to a resin (e.g., HaloTag[82], SpyTag[83]) 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 (i) protein purification and (ii) crosslinked RNA purification relative to CLIP-conditions (**Figure 2C, 2D, S5A, 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', **Figure 2D, 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 (**Figure 2E, 2F, S5C**). In fact, virtually all of the RNA regions that were significantly enriched in the -tag CLIP samples were depleted when measured by CLAP (**Figure 2E-2H**, **S5C**). These results demonstrate that CLAP accurately removes non-specific RNA-protein associations that do not occur *in vivo*.

2.3.3 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 RNA binding proteins (RBPs) that are known to interact with RNA through distinct binding modes: (i) PTBP1 is an RBP that contains multiple RNA Recognition Motif (RRM) domains, binds predominately within intronic regions, and has high selectivity towards a defined RNA sequence motif (HYUUUYU)[84] and (ii) SAF-A (also known as hnRNPU) is an RNA binding protein that contains tandem RGG (arginine-glycine-glycine) motifs and binds promiscuously to many nascent pre-mRNA with a broad localization profile[85].

We performed CLIP and CLAP on each protein across replicates (**Figure 3A**, **S6A, S6B**), sequenced the RNA, and observed the expected binding sites with PTBP1 binding primarily at intronic regions containing its known motif (**Figure 3B, 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 (**Figure 3D-3F, 3G-3I**) and highly reproducible between replicates (**Figure S6A, S6B**). Consistent with this, when we visualized ^{32}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**).

2.3.4 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}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 (**Figure 4A, S7A**), despite successfully 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 (**Figure 4A, S7A, 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 (**Figure 4B-4E, 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 by CLAP (**Figure 4B, 4C, S8C**). These results are strikingly different from the RNA binding profiles observed by CLAP for PTBP1 or SAF-A (**Figure 4B, 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 (**Figure 4C, S8B**), which does not bind RNA *in vitro* (**Figure S2, 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 (**Figure S9A-S9D**). In addition, we confirmed that the Halo-tagged PRC2 components (i) were successfully purified in all cases (**Figure S10**), (ii) could still be properly incorporated into the endogenous PRC2 complex (**Figure S1D**), (iii) were expressed at levels that exceed endogenous levels (**Figure S1A, S1B**), (iv) retain their reported RNA binding activity *in vitro* (**Figure S2**), and (v) 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 kDa vs 30 kDa) into both alleles of the endogenous *Eed*, *Ezh2*, or *Suz12* genes in mouse embryonic stem cells (mESCs) using CRISPR/Cas9 (**Figure 4G, 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 (**Figure S12C, 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.

2.3.5 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[4,32,39,86–89], including repressors and activators of transcription[16,36,37,90], DNA-binding proteins[6,38], and proteins involved in 3D chromatin organization[8,40,41,91]. 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.

(i) CTCF is a zinc finger domain-containing protein that directly binds DNA and is critical for the formation of topologically-associated domain structures[92,93]. Recently, CTCF has

been reported to bind RNA to guide CTCF to specific locations on genomic DNA[7], locally constrain CTCF mobility on chromatin[94], and shape higher-order chromatin structure[7,8,40,41]. The evidence for CTCF as an RNA-binding protein comes from CLIP data indicating that CTCF binds many different RNA targets including lncRNAs such as Tsix, Firre, and others[7]. Yet in contrast to what is observed for other well-defined RNA binding proteins, CLIP analysis of CTCF shows strong RNA association even in the absence of *in vivo* UV-crosslinking[7]. Moreover, disruption of a specific region of CTCF that was reported to ablate interaction with RNA *in vitro* (termed the 'RNA binding region[8]') still leads to detection of an appreciable amount of RNA by CLIP[41].

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 (**Figure 5A, S13B, S13E**). Because RNA binding to CTCF has been proposed to impact CTCF localization, we performed genome-wide mapping of Halo-tagged CTCF by 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.

(ii) YY1 is a DNA-binding protein that is thought to play an important role in mediating enhancer-promoter loop interactions[95]. 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[6,39]. 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 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 (**Figure S13D, 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.

(iii) WDR5 is a component of the MLL complex which deposits the H3K4me3 modification, a mark of transcriptionally active chromatin[96]. 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[35]. 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 observations 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 (**Figure 5E, S13B, S13E**). Given the proposed role for WDR5-RNA interactions in guiding WDR5 to chromatin, we performed ChIP-seq on Halo-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.

2.3.6 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 chromatin proteins might. To do this, we analyzed previously published data generated from RNA purification in UV-crosslinked cells followed by global protein identification using mass spectrometry[2,4,87,88,97–99] 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[100] (**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 (**Figure 6B, S15A**).

(i) SPEN (also called SHARP) interacts with the SMRT/NCOR2 corepressor to activate histone deacetylase activity of HDAC3 and recruit it to specific genomic sites[55,101]. 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[55,68,102,103]. While the interaction between Xist and SPEN was previously observed by CLIP experiments[55,104,105], 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[78,104–106] (**Figure 6D**). We also identified binding sites within several other lncRNAs, including Kcnq1ot1 and Chaserr (**Figure 6D**, **S16A**). Because Kcnq1ot1 and Chaserr have been shown to repress expression of their neighboring genes[101,107–109], 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.

(ii) EWSR1 is a multifunctional protein belonging to the FET (FUS, EWSR1, TAF15) family of proteins that bind DNA and regulate gene expression[110,111]. EWSR1 is known to associate with the basal transcriptional machinery[112] (transcription factor IID and RNA polymerase II[113]) and transcriptional coactivators (CREB-binding protein and histone acetyltransferase p300) to stimulate transcriptional activation[114]. EWSR1 functions have been attributed in part to its ability to undergo oligomerization through RNA-dependent phase separation in cells[113]. 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 (**Figure S15B**, **S15C**), consistent with previous reports by CLIP[115–117]. Because RNA is thought to seed the binding of FET proteins to RNA polymerase II[118,119] and depletion of EWSR1 leads to a global reduction in nascent transcription[120], 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 [121,122].

(iii) CHTOP is a component of an arginine methyltransferase complex[123] containing PRMT1 and PRMT5 that promotes methylation of arginine 3 of Histone H4 (H4R3), a histone modification associated with transcriptional activation[124]. CHTOP has also been reported to bind to DNA sequences containing 5-hydroxymethylcytosine (5hmC)[123,125,126], an intermediate of DNA methylation and an epigenetic mark that recruits DNA-binding proteins[127,128]. Beyond its reported DNA and chromatin roles, CHTOP was also recently identified as a novel component of the TREX mRNA export complex[125,126]. To explore CHTOP binding, we performed CLAP and identified >40,000 enriched sites within ~7,000 RNAs, including within the introns of ALYREF (**Figure 6E**, **6F**), a core factor of the TREX

complex[129,130]. Globally, CHTOP intronic binding sites were predominately located at the 5' end of the first intron (**Figure 6F**, **S16C**), possibly due to co-transcriptional recruitment to pre-mRNAs via splicing, similar to previous observations for TREX binding to mRNA[131]. The association between CHTOP and nascent pre-mRNAs may act to facilitate its recruitment to the genomic DNA region of actively transcribed genes[126].

(iv) 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[132]. In addition, PSPC1 has been reported to localize at promoters[120,133] and guide chromatin regulators[134], 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 (**Figure S15D**, **S16A**), a lncRNA that is essential for paraspeckle assembly[135]. Beyond NEAT1, we identified significant PSPC1 binding sites within the introns and 3'-UTRs of >10,000 mRNAs (**Figure S15B**, **S16B**); these binding preferences are consistent with those observed by previous CLIP experiments[74,136]. 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.

(v) TET2 is a dioxygenase that catalyzes stepwise DNA demethylation by converting 5-methyl cytosine (5mC) to 5-hydroxymethyl cytosine (5hmC) on DNA[137]. To explore TET2 binding to RNA, we performed CLAP and uncovered >7,000 binding sites within ~1,500 RNAs (**Figure 6G, 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[138] (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 (**Figure 6H, 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[139–143]. Because TET2 functions as a DNA demethylase, and because TET2 lacks a known DNA-binding domain[144], 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[134]. 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[134,145,146]. 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**).

2.4 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[147] 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")[148–150]. 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[58]; 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[4,21,22,70,151]. 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[59] 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[152], a RBP that binds many RNAs[153]. Moreover, previous reports of a global role for RNA in PRC2 localization to chromatin[23] were shown to be primarily due to a technical issue and that, once corrected, PRC2 localization is independent of RNA[154,155].

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[156]. 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[63–67,157,158]. Our results demonstrate that CLIP can quantitatively separate *bona fide* interactions that occur *in vivo* from non-specific associations for well-characterized RNA binding proteins (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[67] and for which it is most commonly used[159]. 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[117,160,161]. 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)[34,70], and those reported here. Accordingly, extreme care is needed especially when evaluating proteins that: (i) lack canonical RNA-binding domains, (ii) exhibit binding profiles that approximate those of input RNA, or (iii) 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: (i) it takes significantly less time to perform because it does not require gel extraction, (ii) it allows for fragmentation of RNA using heat and so eliminates structural biases associated with RNase digestion, and (iii) 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[162], proximity labeling[163–165], and reverse transcribe and tagment (RT&Tag)[166]). 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[167] 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.

2.5 MAIN FIGURES

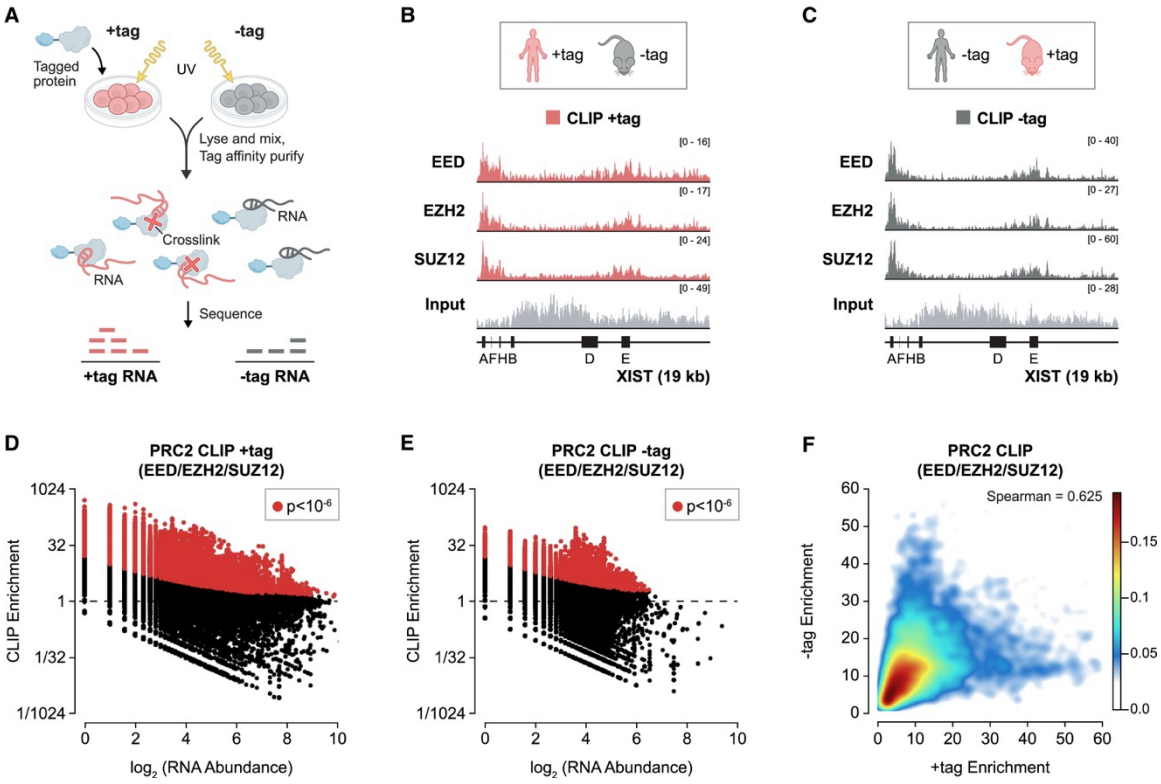


Figure 1: A method to identify RNA-protein associations that could not have formed *in vivo*.

(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, grey). 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, C)** CLIP enrichment profiles for each PRC2 protein (EED, EZH2, SUZ12) are plotted across XIST in the +tag (red) samples (b) and in the -tag (grey) samples (c). Input reads for the EZH2 samples are plotted in light grey. **(D, E)** Scatter plots of input RNA abundance (log scale, x-axis) compared to CLIP enrichment (log scale, y-axis) across 100-nucleotide 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 **Figure S4A** and **S4B**. **(F)** Density scatter plot comparing the levels of significant ($p<10^{-6}$) +tag CLIP enrichments (x-axis) to significant -tag CLIP enrichments (y-axis) for all 3 PRC2 components across all human RNAs.

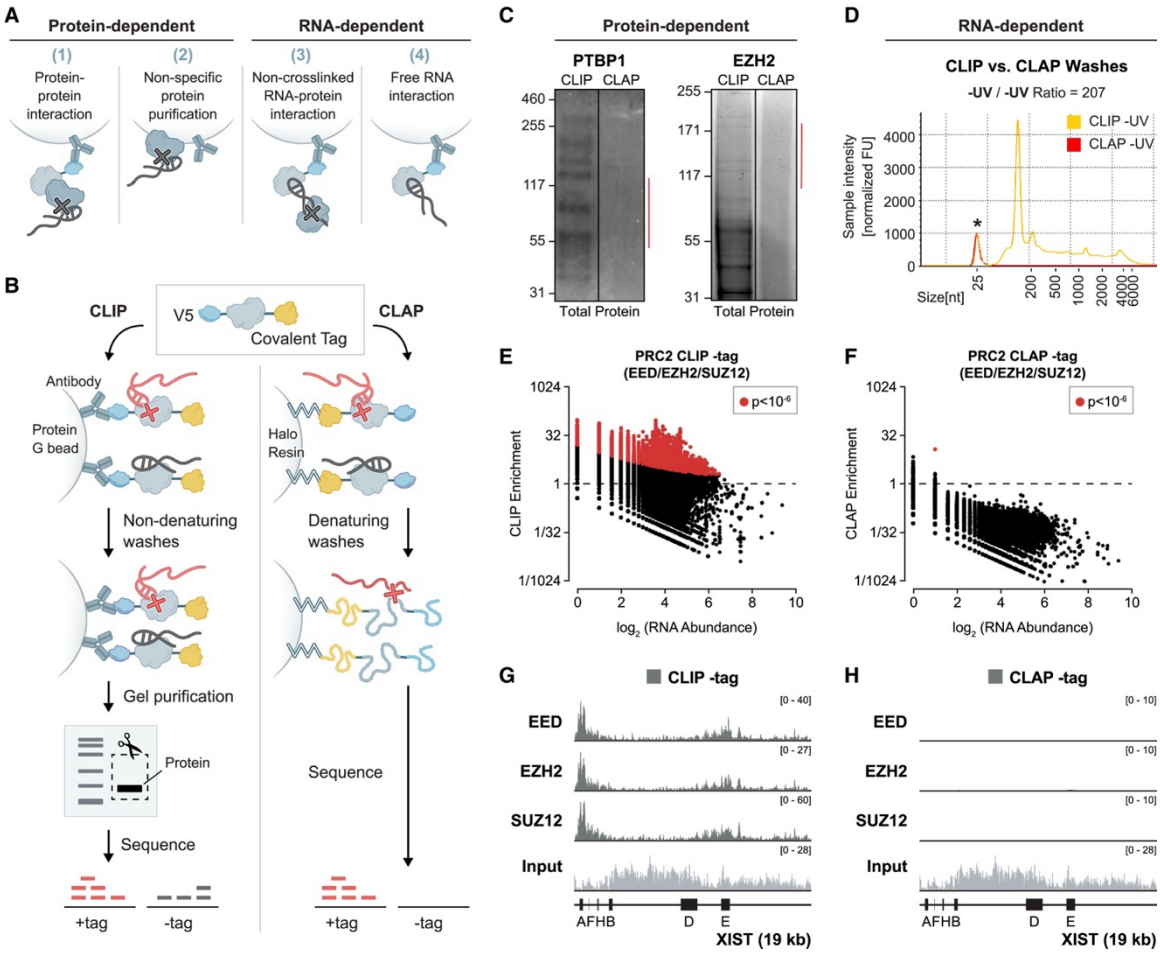


Figure 2: CLAP removes non-specific RNA-protein associations.

(A) Protein-dependent background (left) and RNA-dependent background (right) that could lead to detection of RNA not crosslinked to the purified protein (grey 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, F)** Scatter plots (left, CLIP; right, CLAP) of input RNA abundance compared to enrichment across 100-nucleotide windows of all human RNAs in the -tag experiments. Plots include all 3 PRC2 components; individual components are plotted in **Figure S4B** and **S5C**. **(G, H)** Enrichment profiles for each PRC2 protein (EED, EZH2, 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 grey (CLIP and CLAP input are identical because they come from the same lysate).

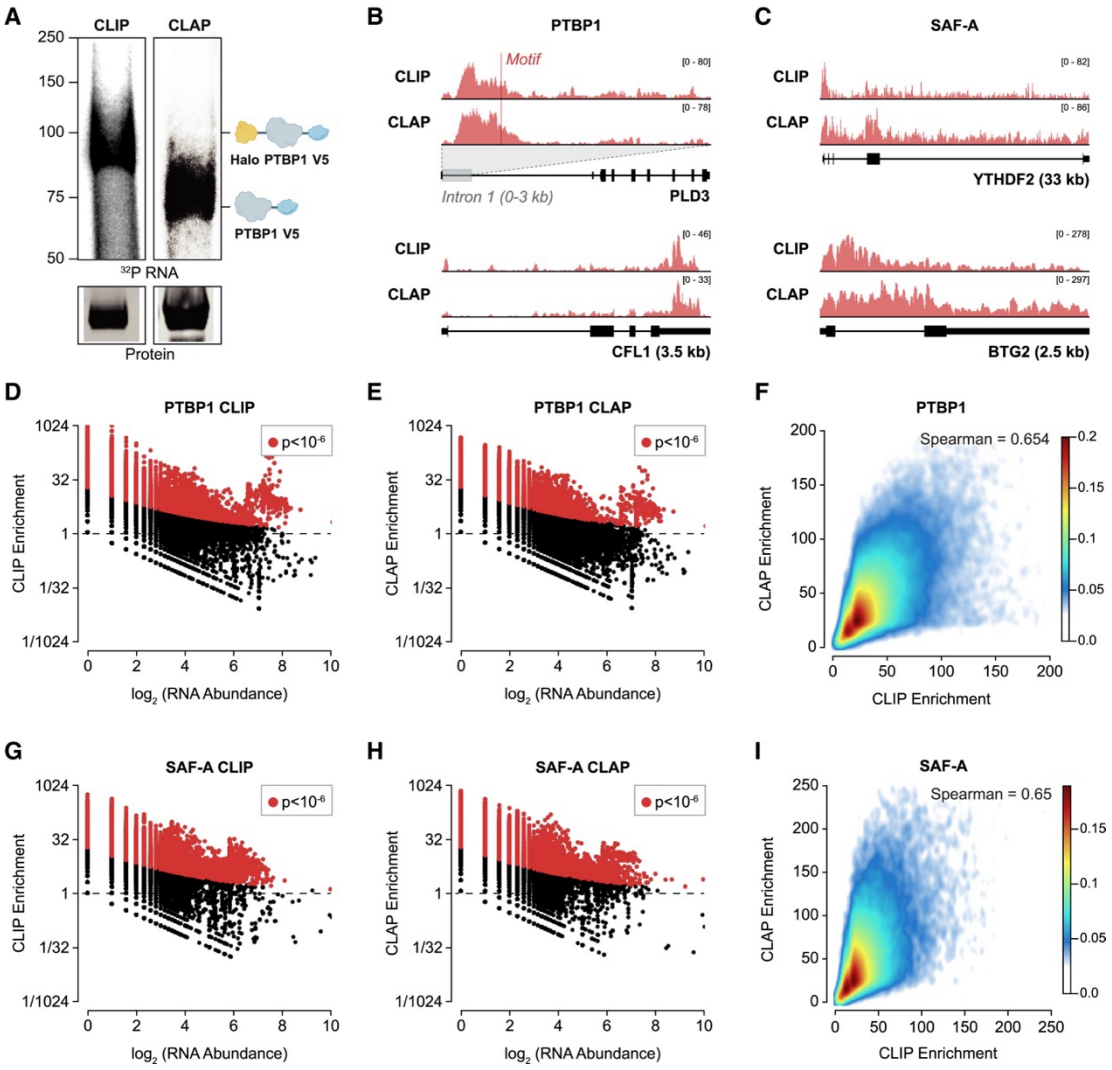


Figure 3: CLAP accurately maps *in vivo* crosslinked RNA-protein interactions.

(A) Visualization of radiolabeled RNA (^{32}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 0-3,000nt) 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, E)** Scatter plots of input RNA abundance compared to

enrichment across 100-nucleotide 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, H)** Scatter plots of input RNA abundance compared to enrichment across 100-nucleotide 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.

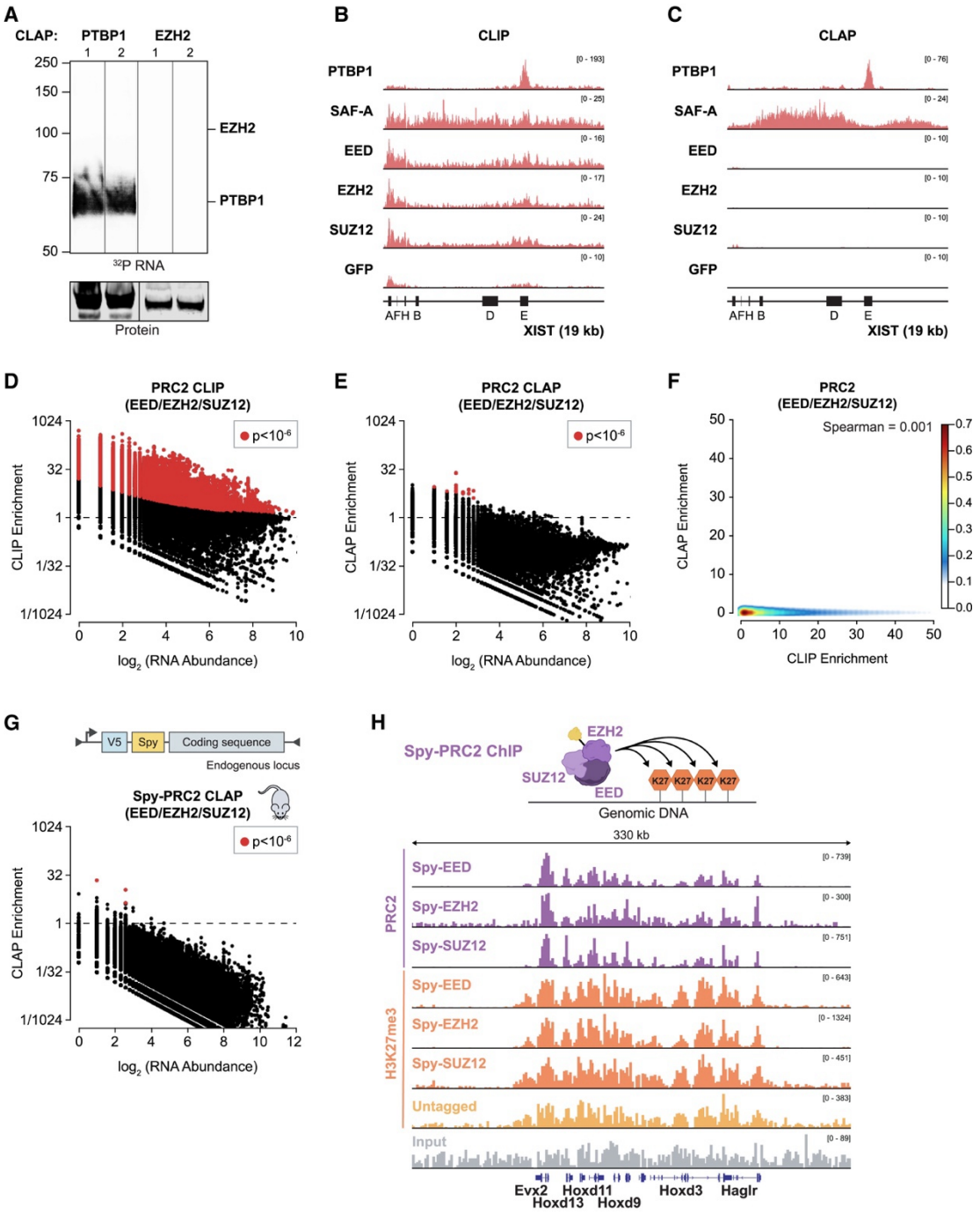


Figure 4: PRC2 components purified using denaturing conditions do not appear to bind RNA.

(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, 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, E)** Scatter plots of input RNA abundance compared to PRC2 enrichment across 100-nucleotide 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-nucleotide windows of all human RNAs. **(G)** Integration strategy at endogenous locus for generating V5-Spy-tagged proteins (top). Scatter plot of input RNA abundance compared to CLAP enrichment for all endogenous Spy-tagged PRC2 proteins across 100-nucleotide 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 grey).

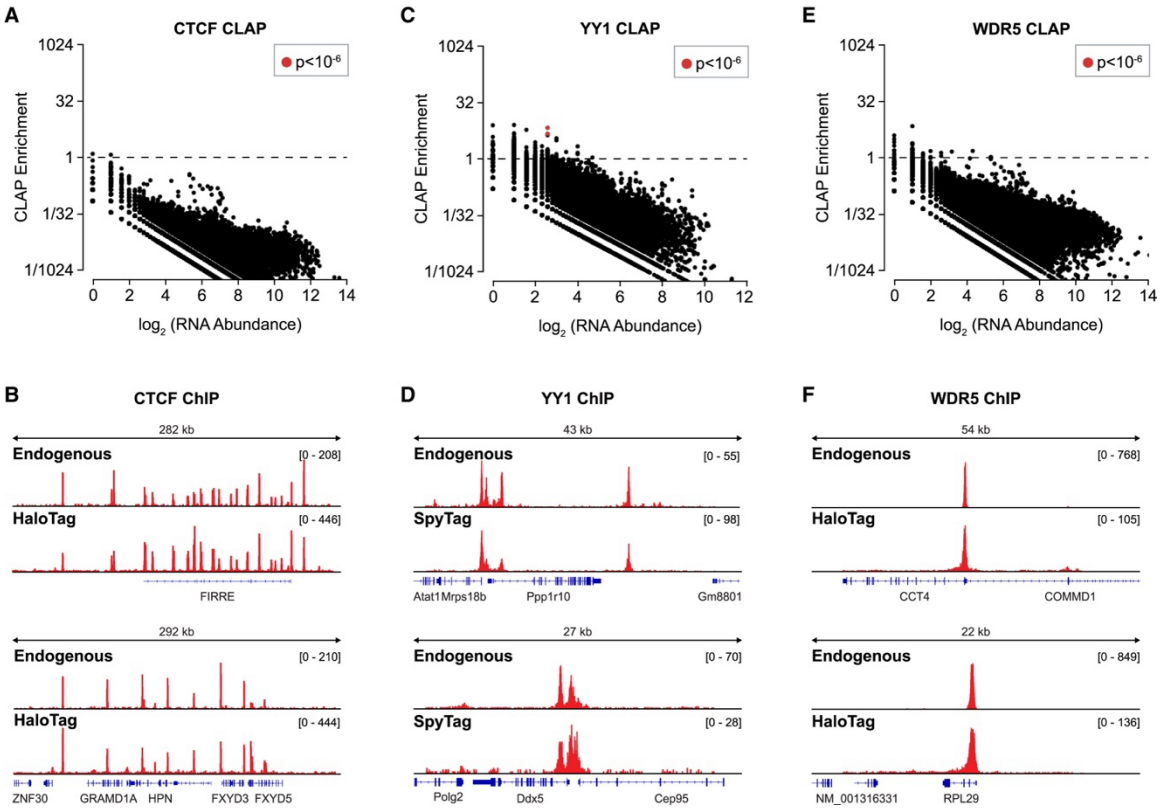


Figure 5: Several chromatin regulators reported to bind RNA do not appear to bind *in vivo*.

(A) Scatter plot of input RNA abundance compared to CLAP enrichment across 100-nucleotide 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 to CLAP enrichment across 100-nucleotide 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 to CLAP enrichment across 100-nucleotide windows of all human RNAs for Halo-WDR5-V5. **(F)** ChIP-seq of endogenous WDR5 versus Halo-V5-tagged WDR5.

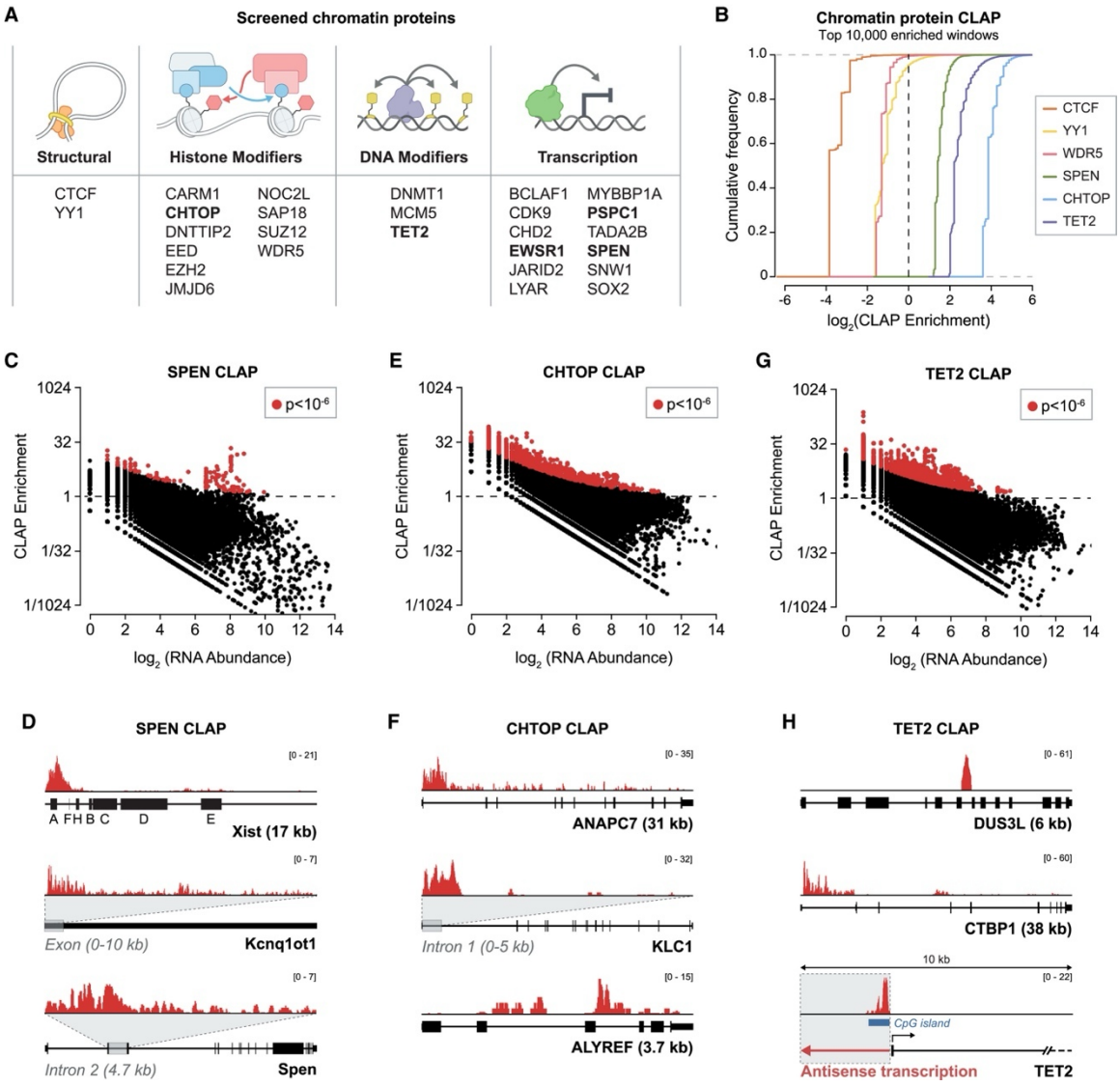


Figure 6: Denaturing purification identifies specific chromatin proteins that bind to RNA *in vivo*.

(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-nucleotide windows across all RNAs for chromatin proteins measured by CLAP. **(C, E, G)** Scatter plots of input RNA abundance compared to CLAP enrichment across 100-nucleotide windows of all RNAs for a set of Halo-V5-tagged chromatin proteins (SPEN, CHTOP, TET2). **(D)** Examples of CLAP enrichment

profiles for SPEN across Xist (top), Kcnq1ot1 (middle, 0-10 kb), and Spen 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, arrows indicate direction of transcription (black, sense; red, antisense).

2.6 SUPPLEMENTAL FIGURES

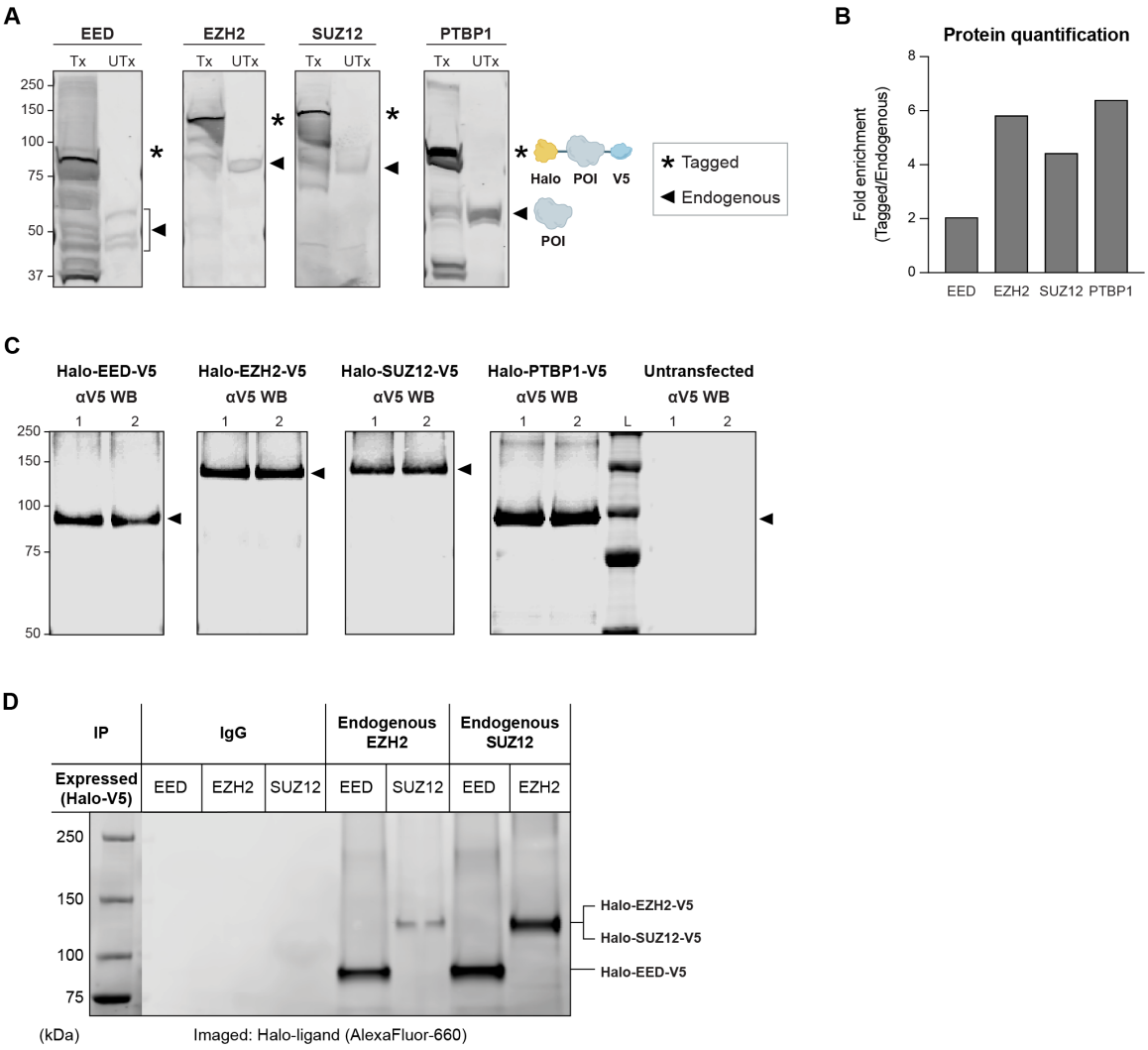


Figure S1: Controls for tagged PRC2 components (Related to Figure 1).

(A) Western blots performed against endogenous proteins within HEK293T cell lysates that were transfected with Halo-V5-tagged proteins (PTBP1, EED, EZH2, SUZ12). (*) denotes the molecular weight of the tagged version of the protein of interest (POI) (endogenous protein + ~33 kDa HaloTag + ~1.5 kDa V5 tag), black arrow depicts the molecular weight of the endogenous protein. "Tx" and "UTx" refer to transfected and untransfected cells, respectively. **(B)** Quantification of (A), depicting fold enrichment of tagged protein

expression relative to endogenous protein. **(C)** Western blots performed against V5 epitope after CLIP captures (via heat elution) of tagged PRC2 components (EED, EZH2, SUZ12, PTBP1) shown in replicate. Black arrows depict the molecular weight of the tagged proteins. An untransfected control was subject to V5 IP and run in replicate on the same blot as PTBP1 (and SUZ12, which is cropped) to demonstrate the specificity of the V5 antibody. **(D)** Endogenous PRC2 components (EZH2 or SUZ12) were immunoprecipitated from cell lysates expressing Halo-EED-V5, Halo-EZH2-V5, or Halo-SUZ12-V5 protein. The amount of tagged protein that was associated with the endogenous protein was visualized using a fluorescently labeled Halo-ligand (AlexaFluor-660) on a gel.

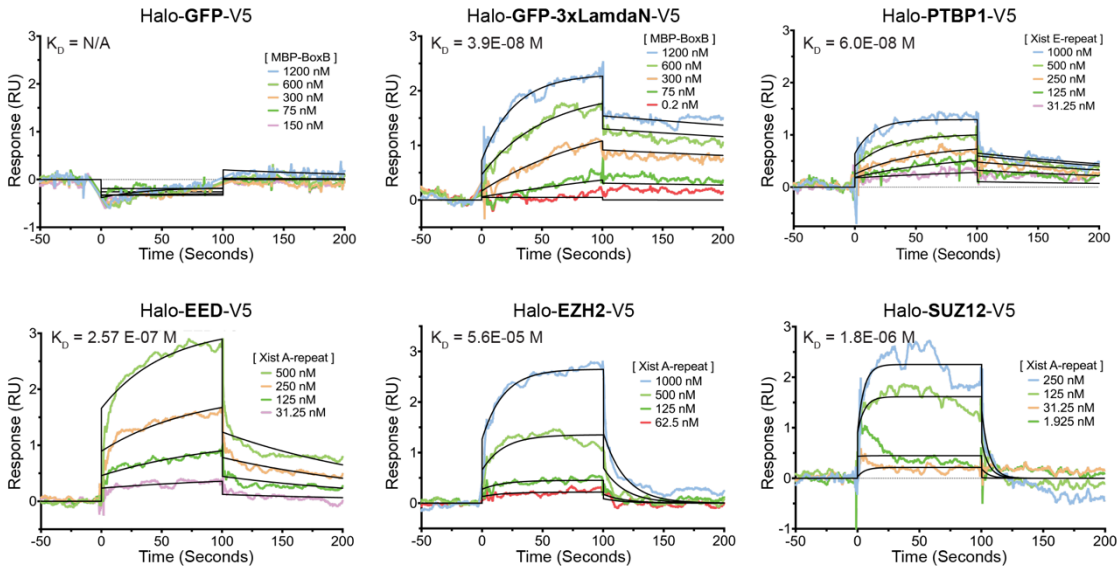


Figure S2: Tagged PRC2 components bind to RNA *in vitro* (Related to Figure 1).

HaloTag fusion proteins were immobilized on a Biacore chip functionalized with HaloTag ligand. In vitro transcribed RNA was injected and the affinity of RNA for HaloTag fusion proteins was measured. RNA substrates were the Maltose Binding Protein (MBP) fused with five tandem repeats of BoxB RNA hairpin or the Xist A-repeat. Sensorgrams for negative control (HaloTag-GFP-V5) and Halo-tagged proteins of interest (GFP-3xLambdaN, PTBP1, EED, EZH2, SUZ12) are shown. Panels show one data set (colored lines) fit globally to a 1:1 Langmuir binding interaction model with bulk refractive index (RI).

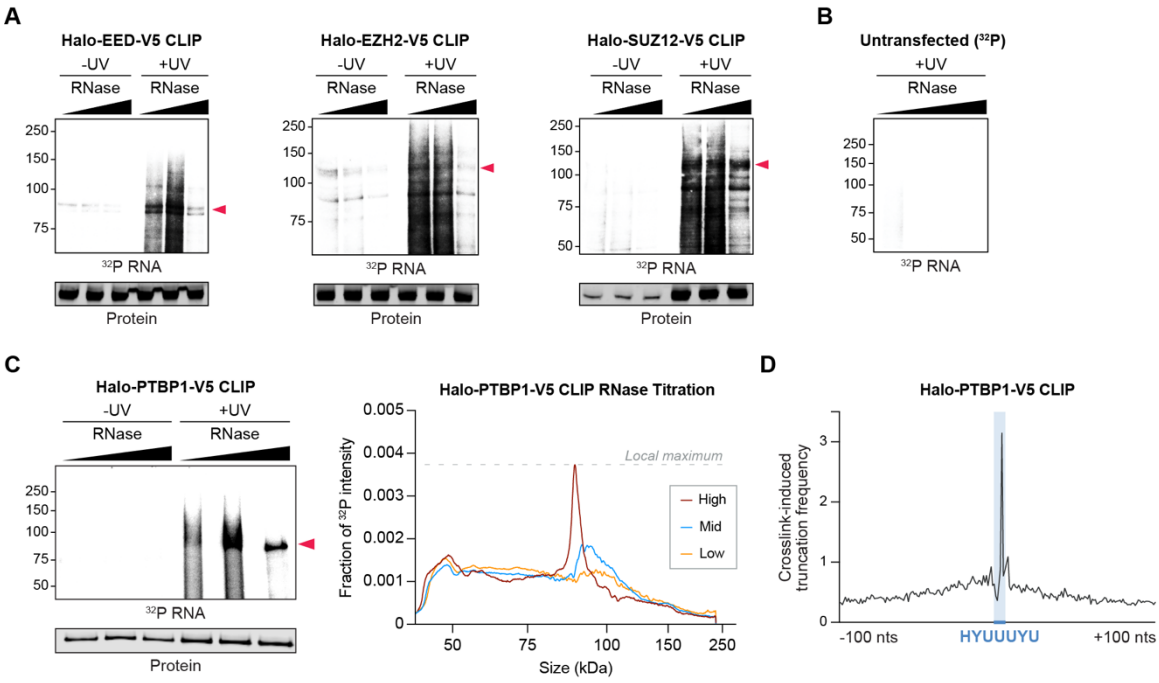


Figure S3: Visualization of RNA-protein complexes purified by CLIP at different RNase concentrations (Related to Figure 1).

(A, B) CLIP was performed from cells expressing proteins tagged with N-terminal Halo and C-terminal V5 tags with (+UV, right) and without UV-crosslinking (-UV, left). RNase I titration was carried out to resolve RNA-protein complexes at RNase I dilutions of 1:50, 1:3,000, and 1:50,000. (A) Tagged EED, EZH2, SUZ12, and (B) untransfected lysates were captured and visualized using the same CLIP conditions. Captured RNA was radiolabeled (^{32}P), run on an SDS-PAGE gel, and transferred to nitrocellulose membrane. Red arrow indicates the expected size of the immunoprecipitated protein. Protein inputs are shown for each lane by Halo-ligand labeling (below). **(C)** (Left) Same as in (A, B), but with tagged PTBP1. (Right) Quantification of ^{32}P Halo-PTBP1-V5 CLIP intensity for each RNase condition (High, Mid, Low, corresponding to 1:50, 1:3,000, and 1:50,000) measured by ImageJ (y-axis), integrated across the length of the gel which is scaled to expected molecular weights (x-axis). **(D)** Crosslink-induced truncation frequency (percentage) of Halo-PTBP1-V5 CLIP relative to known PTBP1 motif (HYUUUYU, shown in blue).

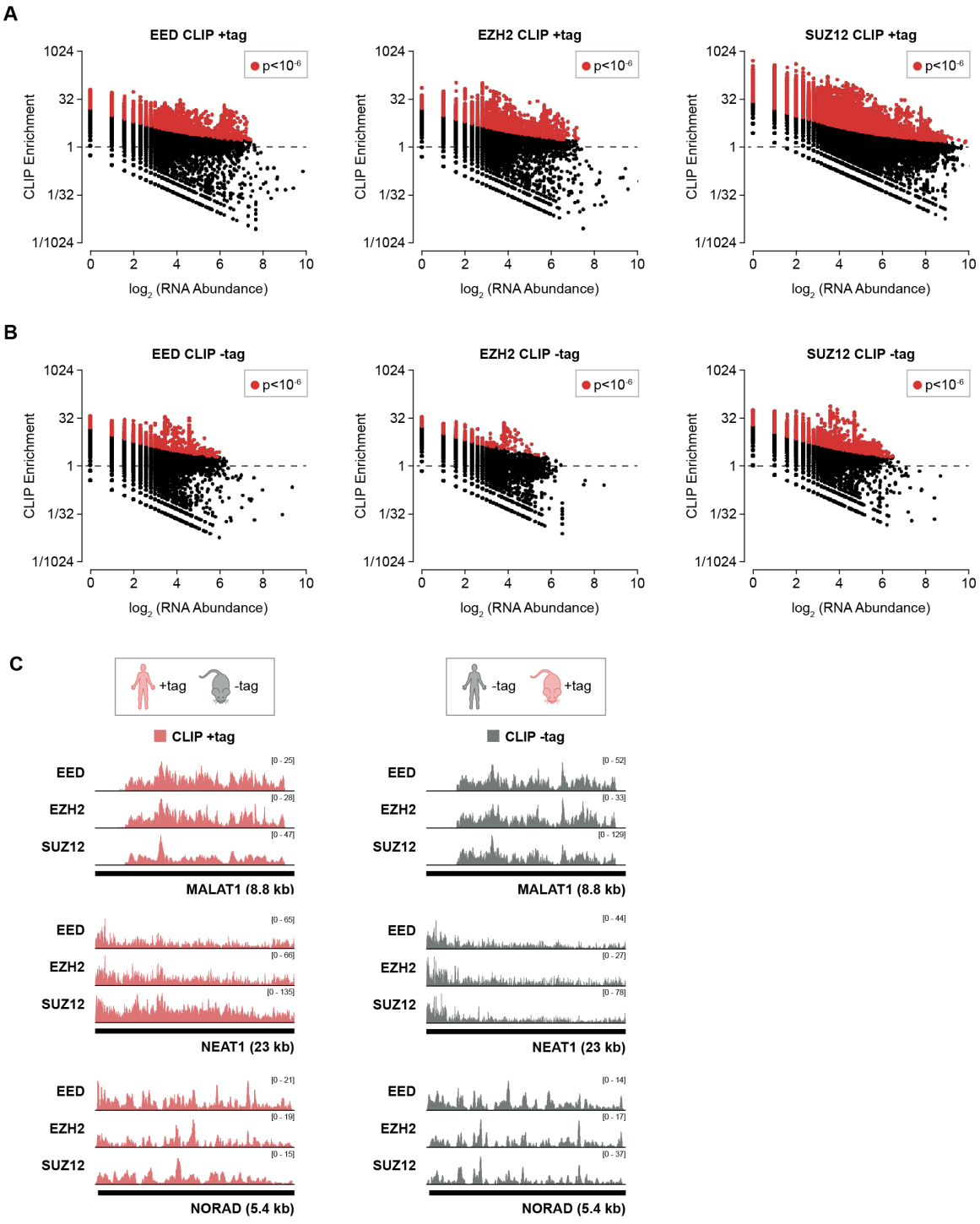


Figure S4: CLIP identifies many PRC2-RNA interactions that could not have occurred *in vivo* (Related to Figure 1).

(A) Scatter plots of input RNA abundance (log scale, x-axis) compared to CLIP enrichment (log scale, y-axis) across 100-nucleotide windows of all annotated human RNAs identified for each individual PRC2 component (EED, EZH2, SUZ12) in the +tag experiments. Windows with significant enrichment (binomial $p<10^{-6}$) are shown in red. **(B)** Same as in (a), but for -tag experiments. **(C)** CLIP enrichment profiles for each PRC2 component in the +tag (red, left) or -tag (grey, right) samples plotted across several human lncRNAs (MALAT1, NEAT1, NORAD).

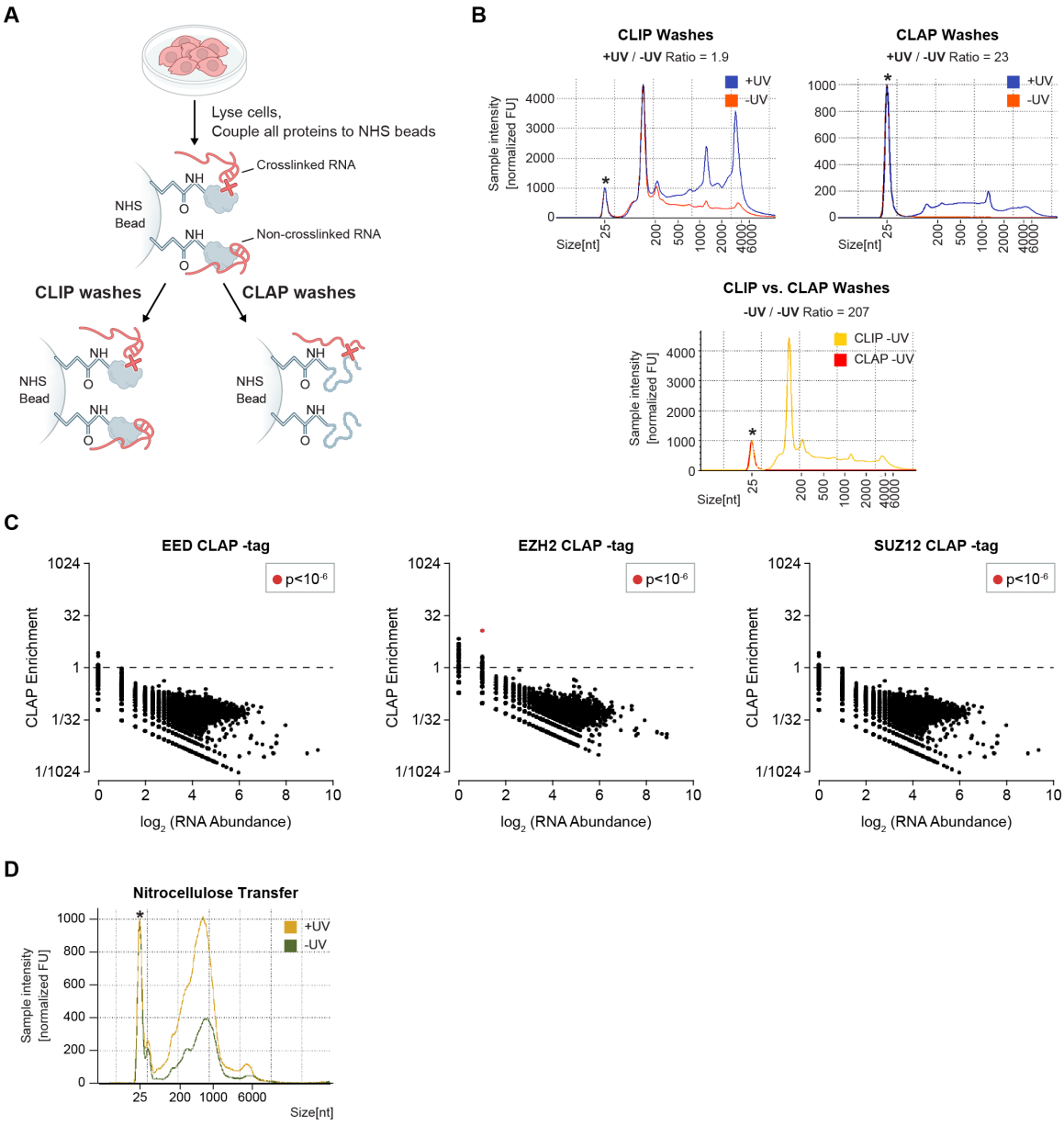


Figure S5: Sources of UV- and protein-dependent non-specific RNA associations (Related to Figure 2).

(A) Schematic overview of experiment comparing stringency of CLIP-washes to CLAP-washes. Equivalent amounts of crosslinked (+UV) or non-crosslinked HEK293T (-UV) whole cell lysate were coupled to amine-reactive beads and washed with either CLIP or CLAP wash buffers. **(B)** Remaining bound RNA-protein complexes were eluted using

Proteinase K and associated RNAs were purified and measured using the Agilent TapeStation High Sensitivity RNA assay. The ratio of +UV to -UV RNAs measured in CLIP conditions (top, left) was determined to be 1.9. The ratio of +UV to -UV RNAs measured in CLAP conditions (top, right) was determined to be 23. Because +UV samples contain both RNA that is crosslinked and non-crosslinked, to directly compare reduction in non-crosslinked we compared the -UV levels in CLIP and CLAP. We found that the ratio of -UV RNA levels measured in CLIP conditions relative to CLAP conditions (bottom, center) was ~207. (*) denotes the lower marker peak for sizing in all graphs. The -UV to -UV comparison plot is reproduced from Figure 2D. **(C)** Scatter plots of input RNA abundance (log scale, x-axis) compared to enrichment (log scale, y-axis) across 100-nucleotide windows of all annotated human RNAs identified for all 3 PRC2 components (EED, EZH2, SUZ12) in the -tag CLAP experiments. Windows with significant enrichment (binomial $p<10^{-6}$) are shown in red. **(D)** To determine how much non-crosslinked RNA is removed by separation on a nitrocellulose membrane, HEK293T whole cell lysate was run through an SDS-PAGE gel and transferred to a nitrocellulose membrane. RNA was eluted from the membrane using Proteinase K and quantified using the Agilent TapeStation High Sensitivity RNA assay. (*) denotes the lower marker peak for sizing.

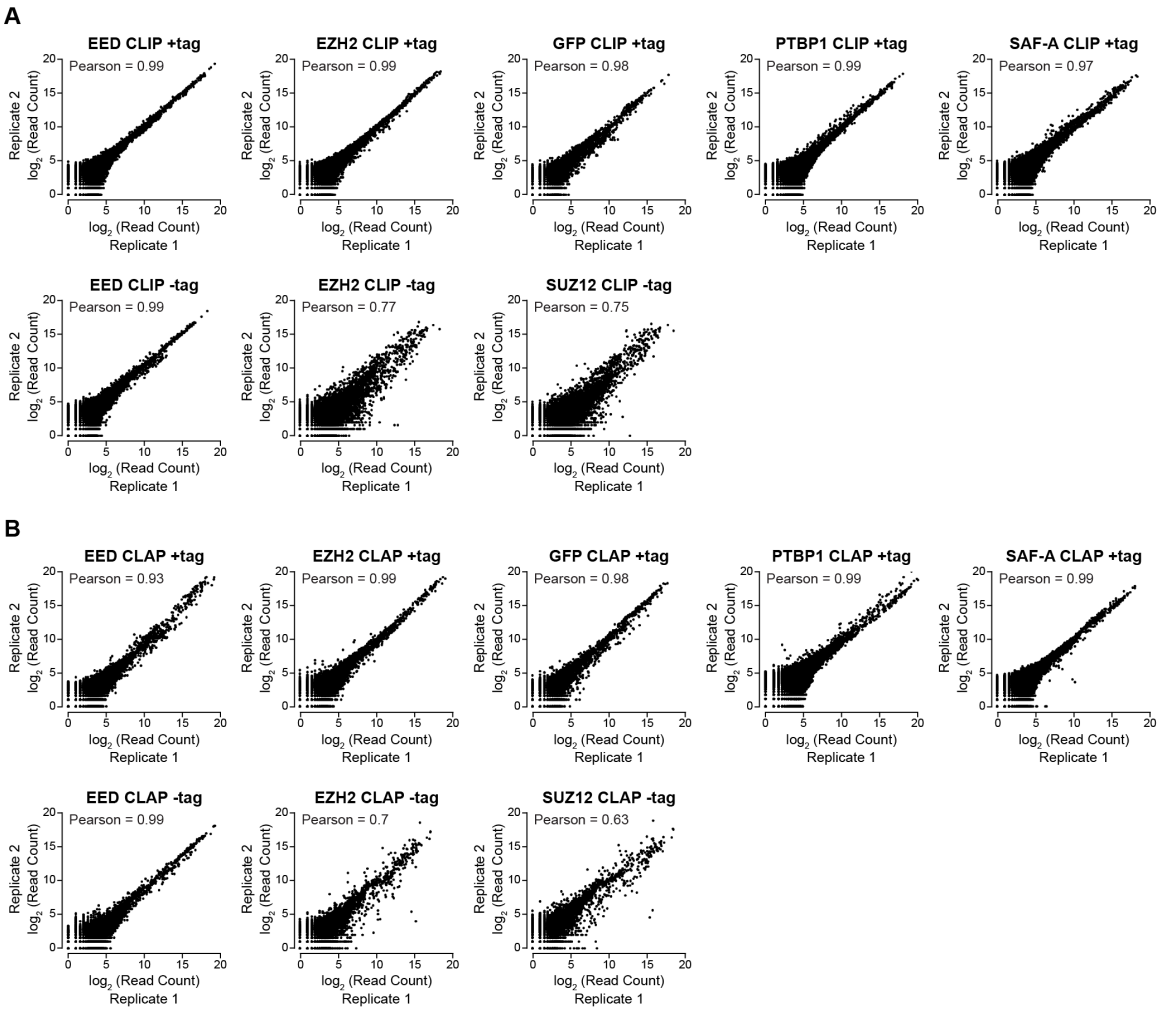


Figure S6: CLIP and CLAP results are robust across independent replicate experiments (Related to Figure 3).

(A) Genome-wide pairwise comparisons of reads across 100-nucleotide windows of all mapped reads (i.e., all annotated human, mouse, rRNA, and repetitive RNAs) for all independent biological replicates of tagged CLIP experiments. Pearson's correlation coefficients (r) are shown. **(B)** Genome-wide pairwise comparisons of reads across 100-nucleotide windows of all mapped reads (i.e., all annotated human, mouse, rRNA, and repetitive RNAs) for all independent biological replicates of tagged CLAP experiments. Pearson's correlation coefficients (r) are shown.

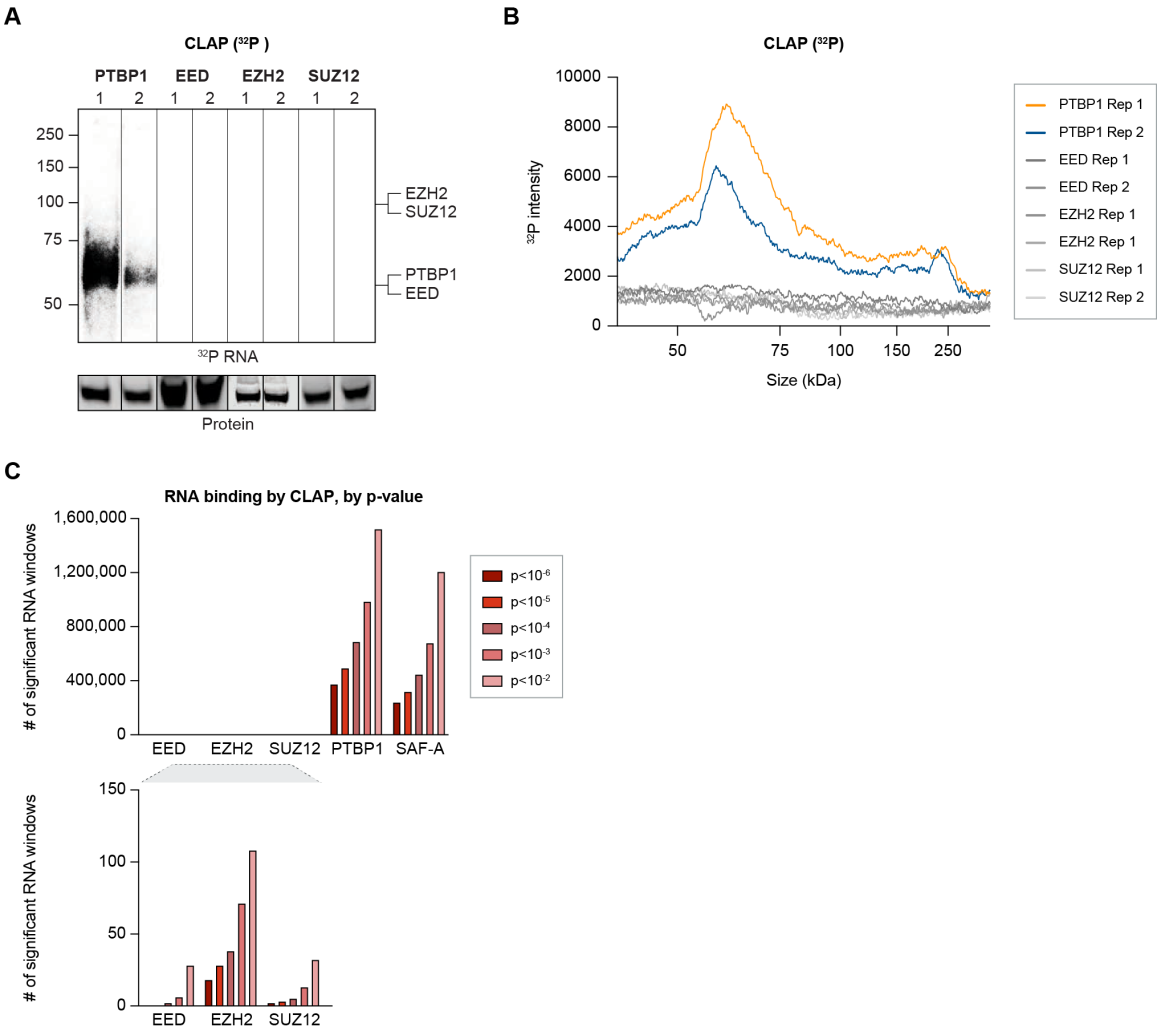


Figure S7: CLAP of PRC2 components does not recover detectable amounts of RNA (Related to Figure 4).

(A) Visualization of radiolabeled RNA (^{32}P) purified by CLAP from tagged versions of PTBP1, EED, EZH2, and SUZ12 after SDS-PAGE and transfer to a nitrocellulose membrane (performed in independent biological replicate). Protein capture was verified for each by western blot (below). **(B)** Quantification of ^{32}P CLAP intensities for all proteins and all replicates depicted in (A) measured by ImageJ (y-axis), integrated across the length of the gel which is scaled to expected molecular weights (x-axis). PTBP1 replicates show high concordance in terms of overall size distribution, despite slightly lower intensity in the second replicate. **(C)** (Top) Bar plot illustrating number of significant RNA windows

detected by CLAP (y-axis) across a range of *p*-values (10^{-2} to 10^{-6}) for all PRC2 components, PTBP1, and SAF-A (x-axis). (Bottom) Zoom-in of the bar plot to visualize number of significant windows for PRC2 components.

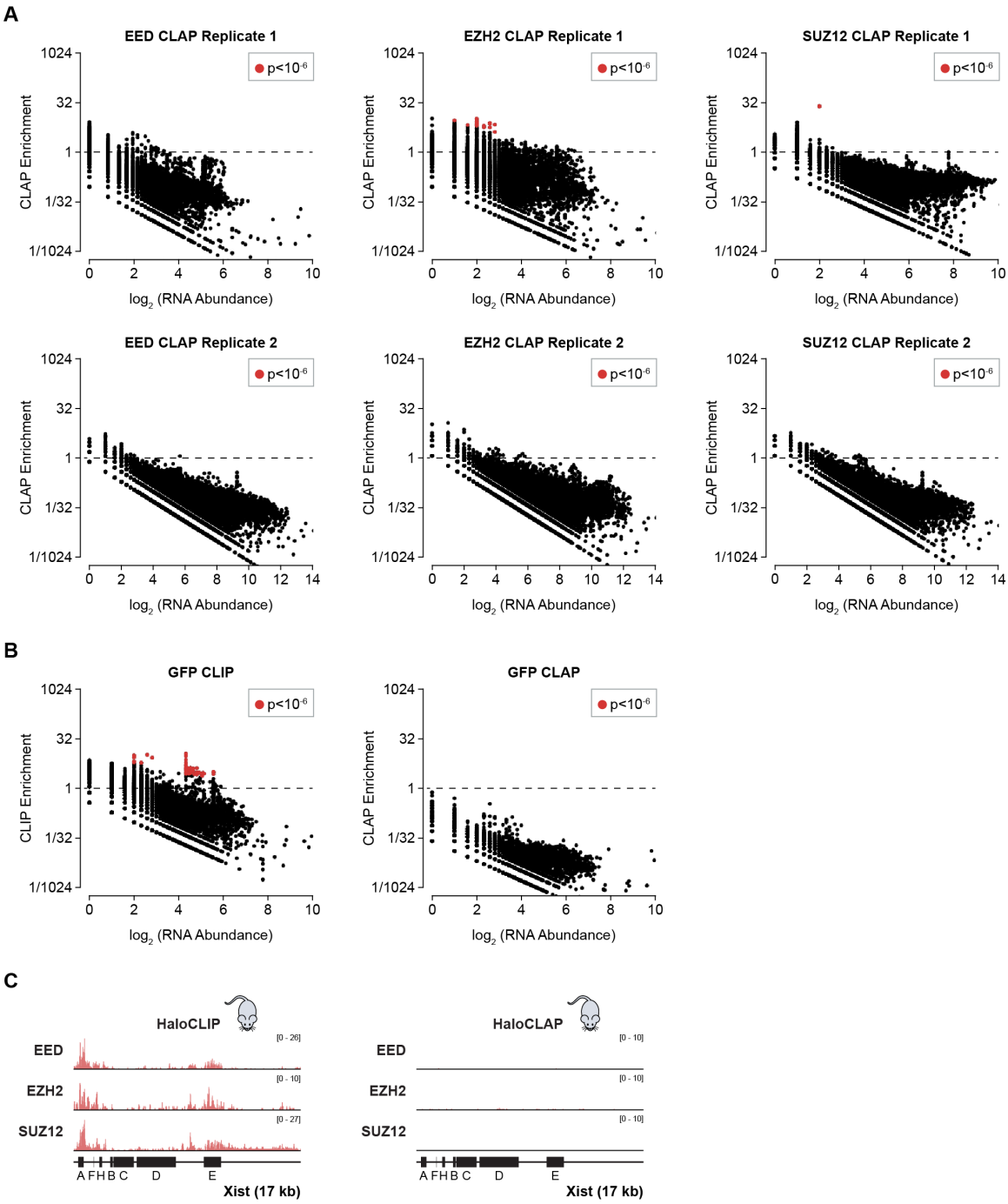


Figure S8: PRC2 components do not appear to bind RNA *in vivo* (Related to Figure 4).

(A) Scatter plots of input RNA abundance (log scale, x-axis) compared to enrichment (log scale, y-axis) across 100-nucleotide windows of all annotated human RNAs identified for

each individual PRC2 component (EED, EZH2, SUZ12) in the +tag CLAP experiments (top, Replicate 1) and independent biological replicates without the +tag/-tag system (bottom, Replicate 2). Windows with significant enrichment (binomial $p<10^{-6}$) are shown in red. (We found one gene, DNAJB1, that consistently showed higher levels in the EZH2 CLAP samples relative to input across the entire gene. Because this was uniformly higher across the entire gene and did not show any specific enriched regions, we suspect it is an artifact and manually masked it in our plots to avoid confusion). **(B)** Scatter plots (left, CLIP; right, CLAP) of input RNA abundance (log scale, x-axis) compared to enrichment (log scale, y-axis) across 100-nucleotide windows of all annotated human RNAs identified for Halo-GFP-V5 in the +tag experiments. Windows with significant enrichment (binomial $p<10^{-6}$) are shown in red. **(C)** CLIP (left) and CLAP (right) enrichments for EED, EZH2, and SUZ12 plotted across the mouse Xist lncRNA in the -tag experiments.

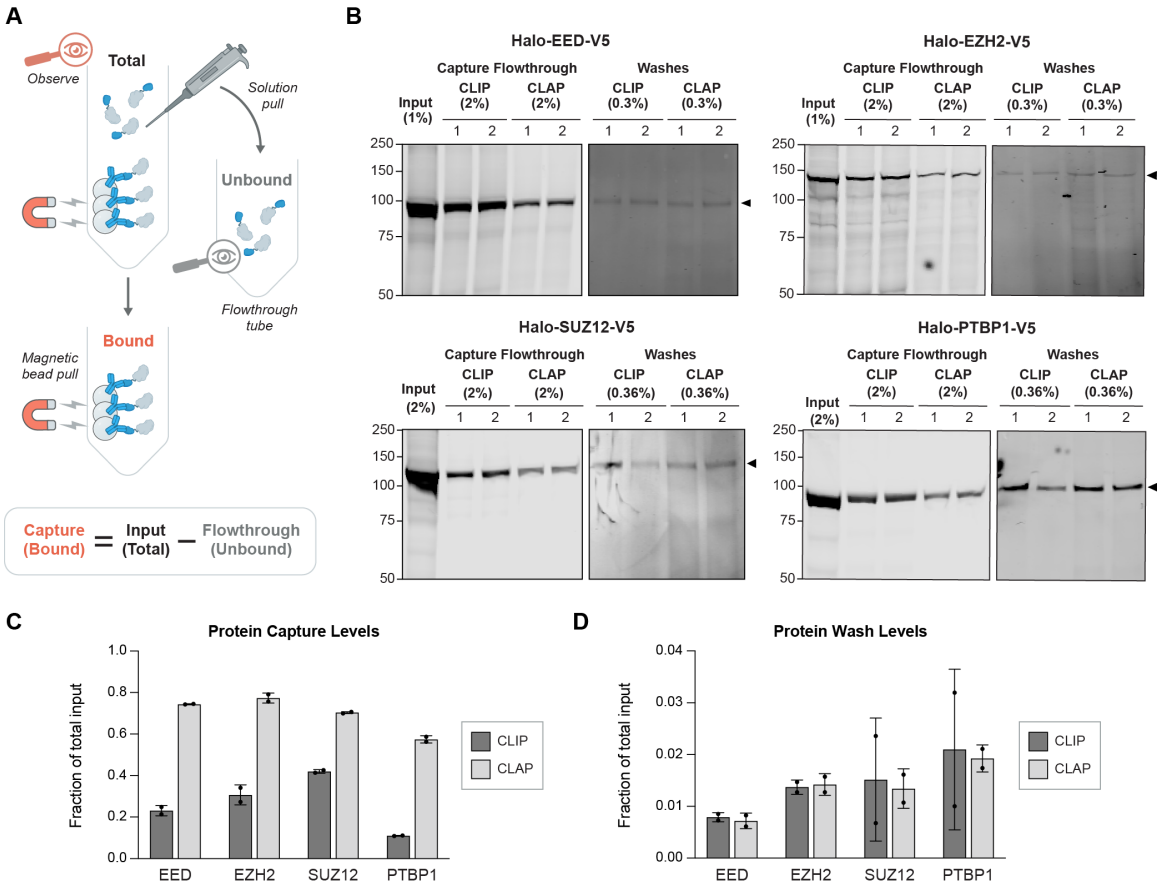


Figure S9: Quantification of protein yields for CLIP and CLAP experiments (Related to Figure 4).

(A) Schematic overview of method to quantitatively measure protein capture efficiency between CLIP and CLAP methods. The amount of tagged protein within the Input (Total) and Flowthrough (Unbound) samples can be directly measured using a fluorescently labeled Halo-ligand (AlexaFluor-660) on a gel. Thus, the Capture (Bound) amount can be computed by subtracting the Input (Total) and Flowthrough (Unbound) measurements (Capture = Input – Flowthrough). **(B)** Protein gels quantifying the amounts of Halo-V5-tagged proteins (by Halo-ligand AlexaFluor-660) within the Input, Capture Flowthrough, and Washes fractions of CLIP and CLAP for each protein (EED, EZH2, SUZ12, PTBP1) in replicate. Capture Flowthrough represents the flowthrough immediately saved after affinity purification. Washes (imaged at a higher contrast to show protein bands) represent the sum of all CLIP-

or CLAP-washes performed throughout the protocols. Black arrow depicts the expected molecular weight of the tagged protein. **(C)** Bar plots of protein capture levels (determined by Capture = Input – Flowthrough) for either CLIP or CLAP, for all proteins assayed (EED, EZH2, SUZ12, PTBP1) (error bars = S.D.). In all cases, CLAP captured more tagged protein than CLIP. **(D)** Bar plots of protein wash levels for either CLIP or CLAP, for all proteins assayed (EED, EZH2, SUZ12, PTBP1) (error bars = S.D.). In all cases, CLIP and CLAP wash levels were comparable.

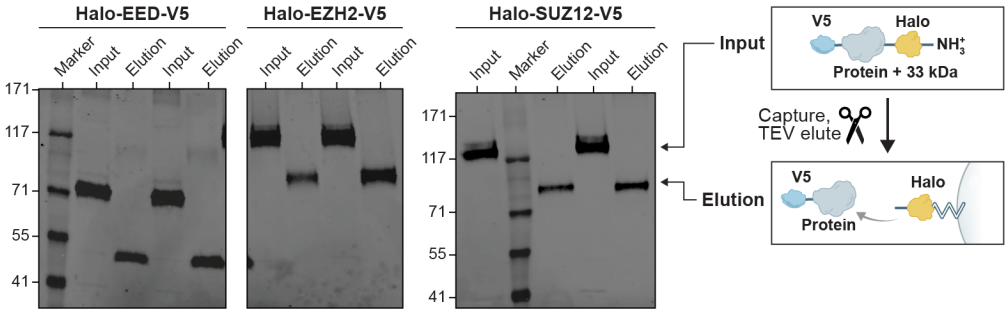


Figure S10: Tagged PRC2 components are successfully purified after CLAP (Related to Figure 4).

Western blots performed against V5 epitope after CLAP captures (via TEV protease elution) of tagged PRC2 components (EED, EZH2, SUZ12) shown in replicate. The TEV cleavage releases the remainder of the HaloTag-bound portion of the protein (Elution) from the resin, which migrates ~33 kDa lower than the uncleaved protein (Input).

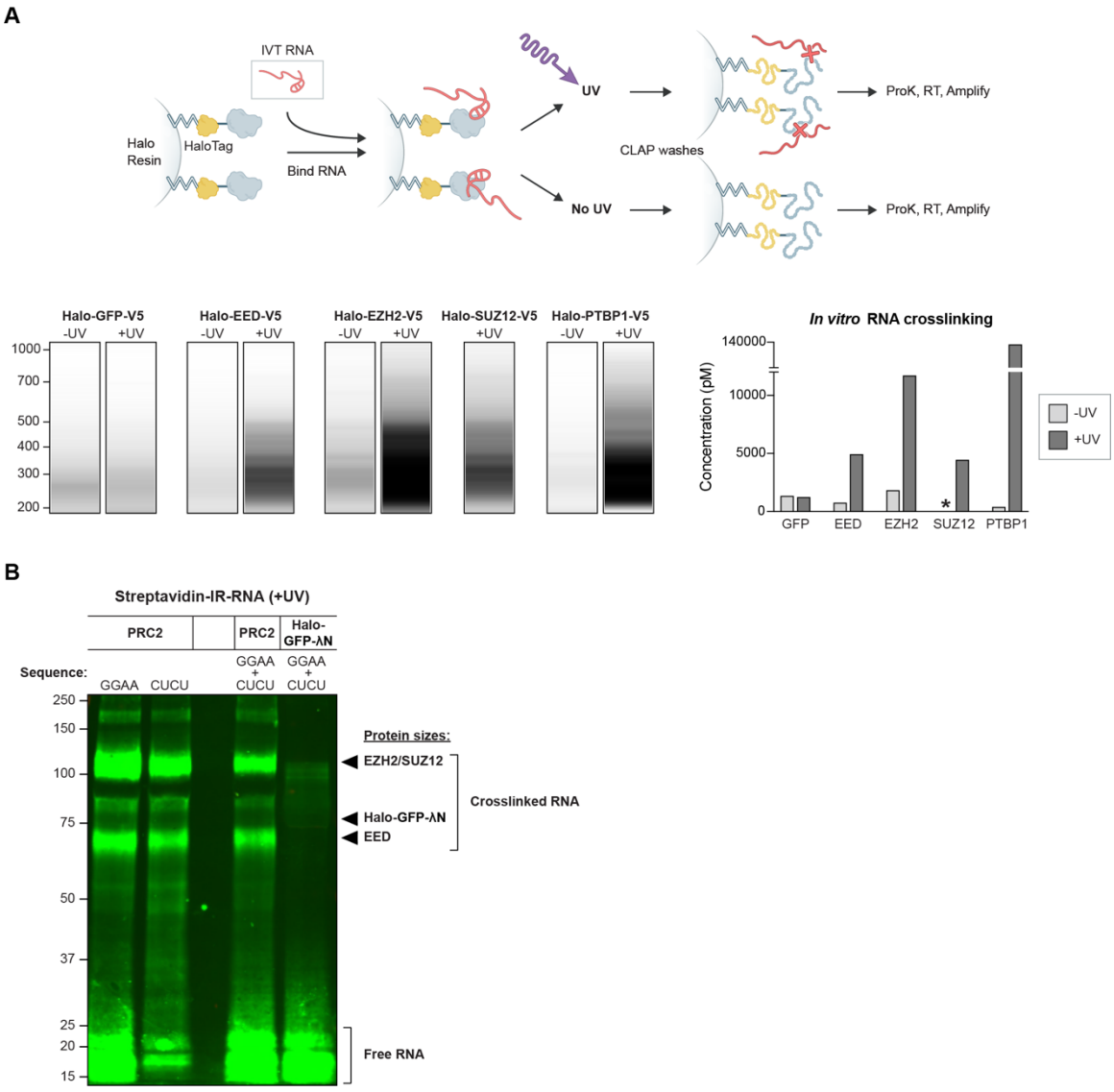


Figure S11: Tagged PRC2 components UV-crosslink to RNA when assembled *in vitro* (Related to Figure 4).

(A) (Top) Schematic overview of strategy to assess degree of *in vitro* UV-crosslinking for Halo-tagged proteins. Halo-tagged proteins were expressed in HEK293T cells, lysed, and Halo-tagged proteins were captured on HaloLink resin under native conditions, incubated with in vitro transcribed RNA (XIST A-repeat), and then either crosslinked (UV) or not crosslinked (No UV) in solution with 254 nm UV light. Samples were then washed with standard CLAP denaturing washes to remove non-crosslinked material. The remaining

protein-bound RNA was released from resin by Proteinase K, reverse transcribed, amplified by PCR, and read out on an Agilent TapeStation High Sensitivity DNA assay (see **STAR Methods**). (Bottom, left) Agilent TapeStation size profiles shown for each Halo-tagged protein assayed (GFP, EED, EZH2, SUZ12, PTBP1). (Bottom, right) Quantification (in pM) of each sample from the in vitro crosslinking assay. (*) indicates sample was undetectable on TapeStation. **(B)** Purified PRC2 complex (Active Motif) was incubated with short biotinylated RNA sequences (GGAA, CCUU, combination of annealed GGAA+CCUU), crosslinked with 254 nm UV light, run on SDS-PAGE, transferred to nitrocellulose membrane, and visualized with a streptavidin-conjugated infrared dye (see **STAR Methods**). Halo-GFP-λN was used as a positive control for RNA-binding and crosslinking. Crosslinked RNAs are indicated at the expected sizes of each protein (i.e., EED, EZH2, SUZ12, Halo-GFP-λN). Free RNAs are indicated at the bottom of the gel, which separate from crosslinked RNA-protein complexes in the denaturing gel.

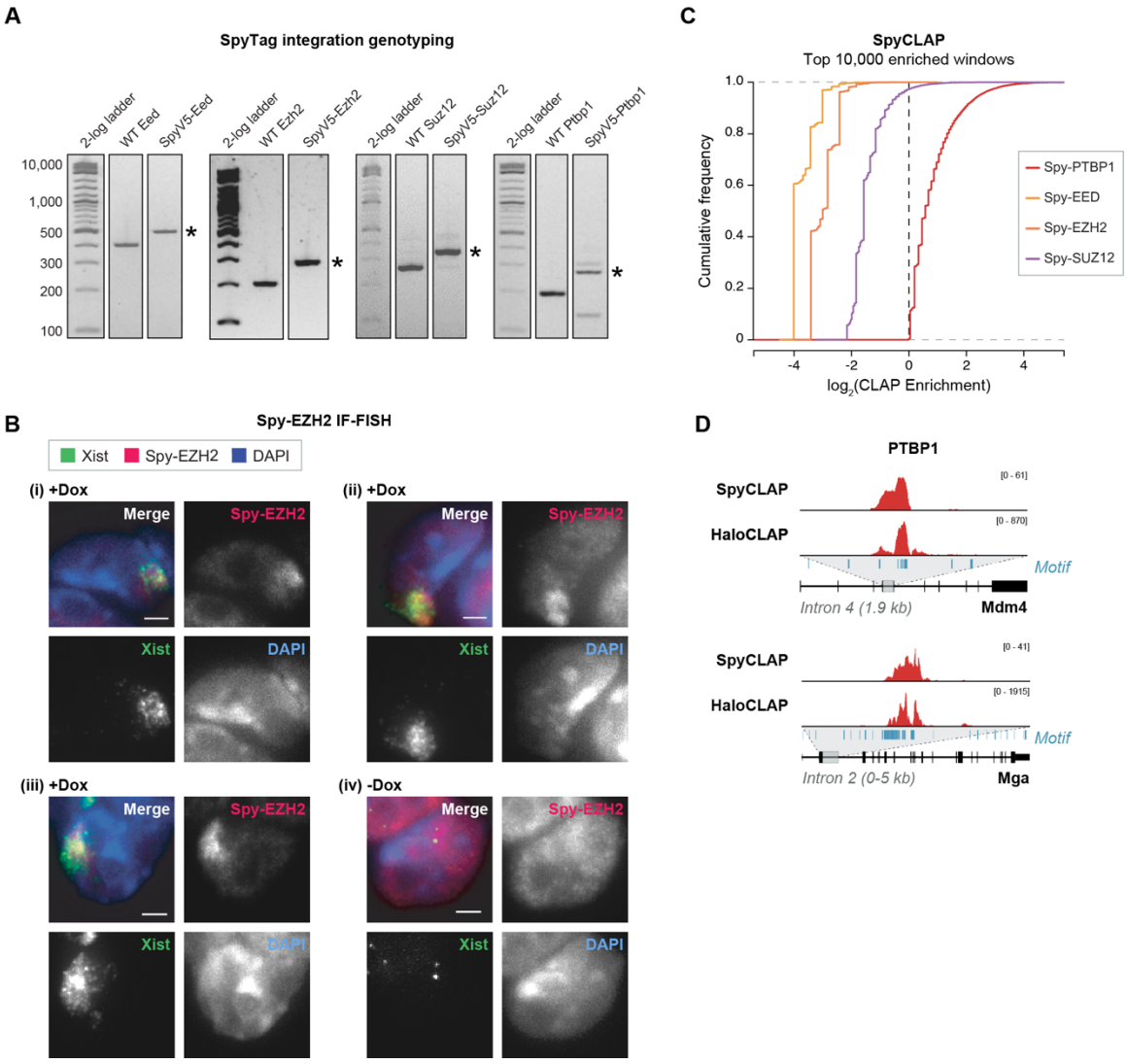


Figure S12: Controls for Spy-tagged PRC2 experiments (Related to Figure 4).

(A) Genotyping PCR for the endogenous, homozygous V5-Spy-tagged proteins, including PRC2 components (*Eed*, *Ezh2*, *Suz12*) and *Ptbp1*. Asterisk (*) indicates the expected size shift corresponding to the combined length of V5 and SpyTag sequences (81 nucleotides). **(B)** RNA-FISH-Immunostaining performed against a cell line expressing endogenous Spy-tagged EZH2 (red) under inducible expression of Xist RNA (green, driven by Dox-inducible promoter). Panels i-iii are from Dox-induced cells for Xist expression; panel iv is from non-induced cells. Individual fluorescent channels are shown in grey-scale. Scale bars, 2 μm. **(C)**

Cumulative Distribution Function (CDF) plot for the top 10,000 enriched 100-nucleotide windows across annotated mouse RNAs for Spy-tagged PTBP1, EED, EZH2, and SUZ12 measured by CLAP. **(D)** CLAP enrichment profiles for Spy-tagged PTBP1 (top) and Halo-tagged PTBP1 (bottom) plotted across two mRNAs (Mdm4, Mga). PTBP1 recognition motifs (HYUUUYU, blue lines) are shown.

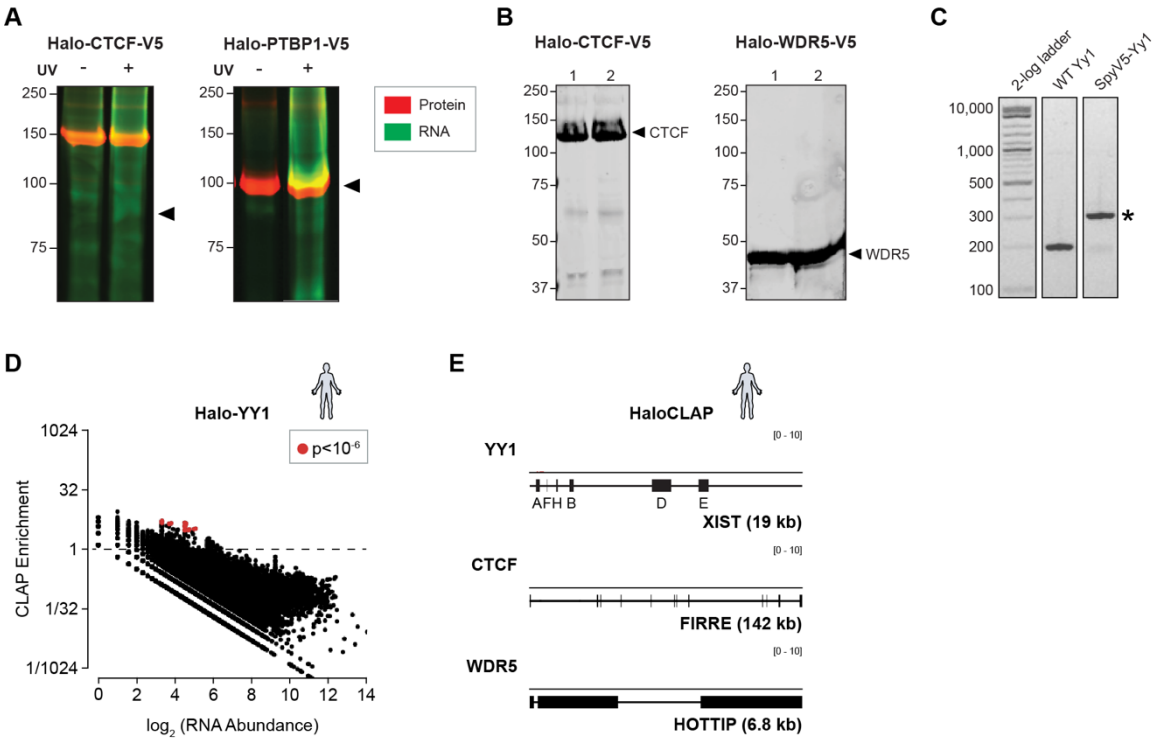


Figure S13: Other chromatin regulators do not appear to directly bind RNA *in vivo* (Related to Figure 5).

(A) HEK293T cells expressing Halo-CTCF-V5 (left) or Halo-PTBP1-V5 (right) were either crosslinked via UV irradiation (+UV) or not crosslinked (-UV) and subject to immunoprecipitation with V5 antibody and subsequent labeling of captured protein with a fluorescent Halo-ligand (AlexaFluor-660). Immunoprecipitated RNA was labeled through T4 RNA ligation of an IRDye-800 labeled RNA adapter. Immunoprecipitated complexes were run on an SDS-PAGE gel, transferred to a nitrocellulose membrane, and imaged on the LI-COR imaging system. Proteins (red) and RNA (green) were visualized simultaneously. **(B)** Western blots performed against V5 epitope after CLAP captures (via TEV protease elution) of Halo-V5-tagged chromatin proteins (CTCF, WDR5) shown in replicate. Black arrows indicate the expected molecular weight of the protein after TEV cleavage. **(C)** Genotyping PCR for the endogenous, heterozygous V5-Spy-tagged *Yy1* cell line. Asterisk (*) indicates the expected size shift corresponding to the combined length of V5 and SpyTag sequences (81 nucleotides). **(D)** Scatter plot of input RNA abundance (log scale, x-axis)

compared to CLAP enrichment (log scale, y-axis) across 100-nucleotide windows of all annotated human RNAs for Halo-tagged YY1. Windows with significant enrichment (binomial $p<10^{-6}$) are shown in red. **(E)** CLAP enrichment profiles for Halo-tagged versions of YY1, CTCF, and WDR5 plotted over respective lncRNAs (XIST, FIRRE, HOTTIP) that each protein was previously reported to interact with. In these plots, there are no significant enrichments.

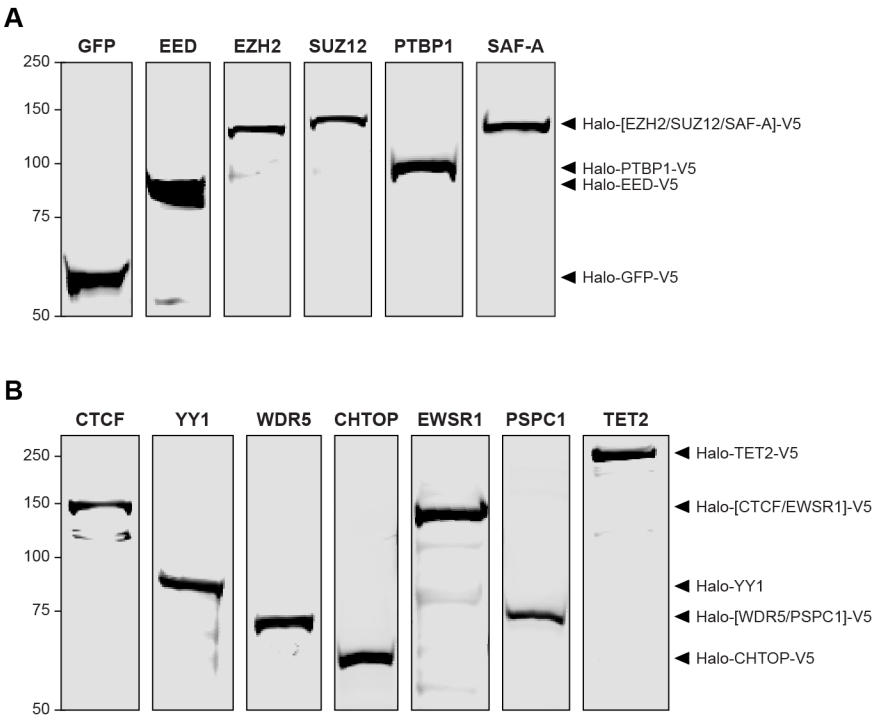


Figure S14: Confirmation of Halo-V5-tagged protein expression (Related to Figure 6).

(A) Protein expression gels of Halo-V5-tagged proteins corresponding to the proteins used in the +tag/-tag experiments. Expression was visualized via AlexaFluor-660 conjugated Halo-ligand. Black arrows indicate expected tagged protein molecular weights. **(B)** Same as in (A), but for the screened chromatin proteins by CLAP. Black arrows indicate expected tagged protein molecular weights.

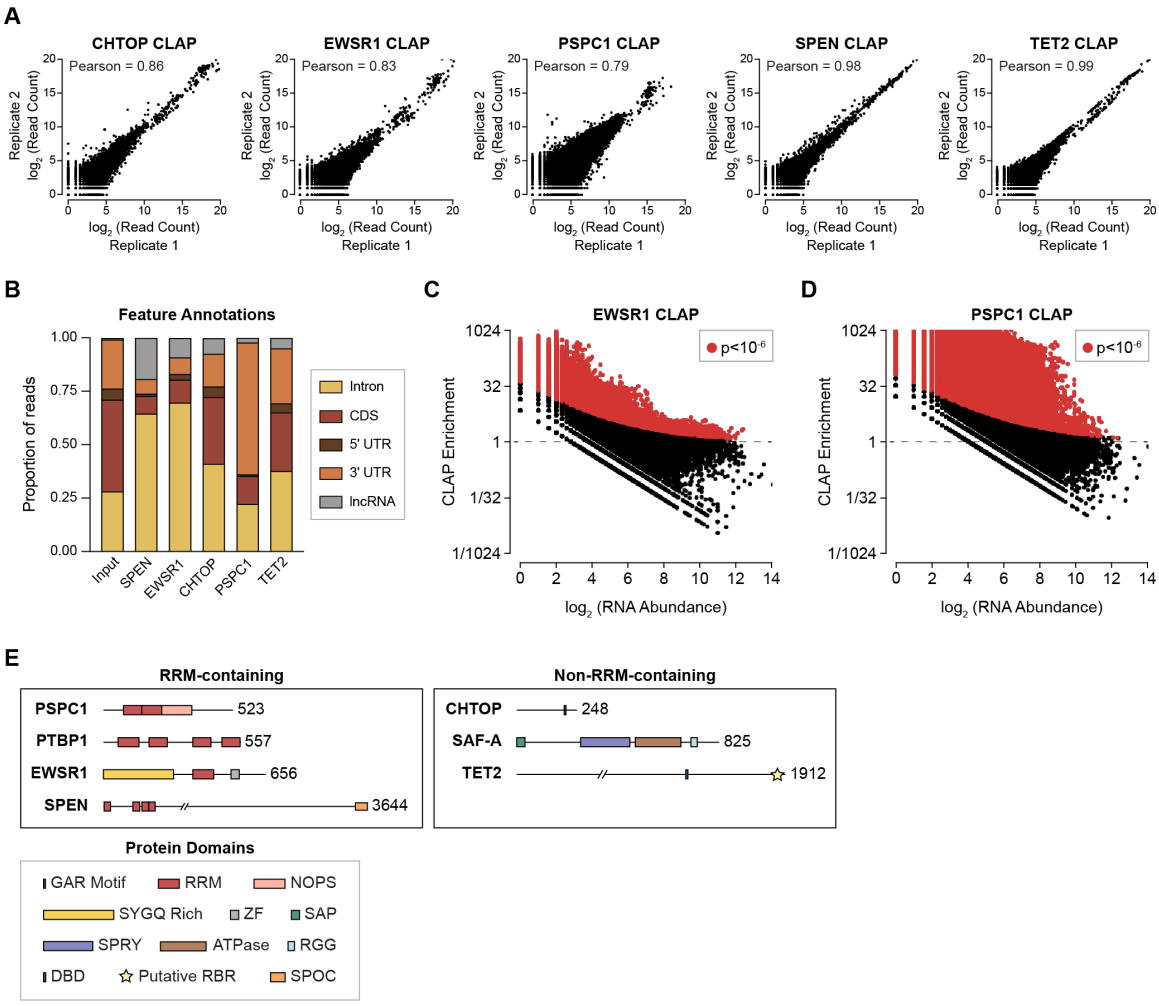


Figure S15: CLAP resolves RNA-binding capacity of proteins across a wide range of functional domains (Related to Figure 6).

(A) Genome-wide pairwise comparisons between independent biological replicate experiments for set of chromatin RNA-binding proteins identified by CLAP (CHTOP, EWSR1, PSPC1, SPEN, TET2). Pearson's correlation coefficients (r) are shown. **(B)** Distribution of input and CLAP reads mapping to distinct transcript features (introns, coding sequences, 5' UTRs, 3' UTRs, lncRNAs) for each chromatin protein (SPEN, EWSR1, CHTOP, PSPC1, TET2). **(C, D)** Scatter plots of input RNA abundance (log scale, x-axis) compared to CLAP enrichment (log scale, y-axis) across 100-nucleotide windows of all annotated human RNAs for Halo-V5-tagged EWSR1 (left) and Halo-V5-tagged PSPC1

(right). Windows with significant enrichment (binomial $p<10^{-6}$) are shown in red. **(E)** Diagram depicting the functional protein domains (from UniProt) of all proteins that we successfully mapped to RNA (scaled to protein length) in this study. Proteins are separated based on whether they contain an annotated RRM (RNA recognition motif) or not. "Putative RBR" refers to the RNA-binding region previously reported for TET2[4].

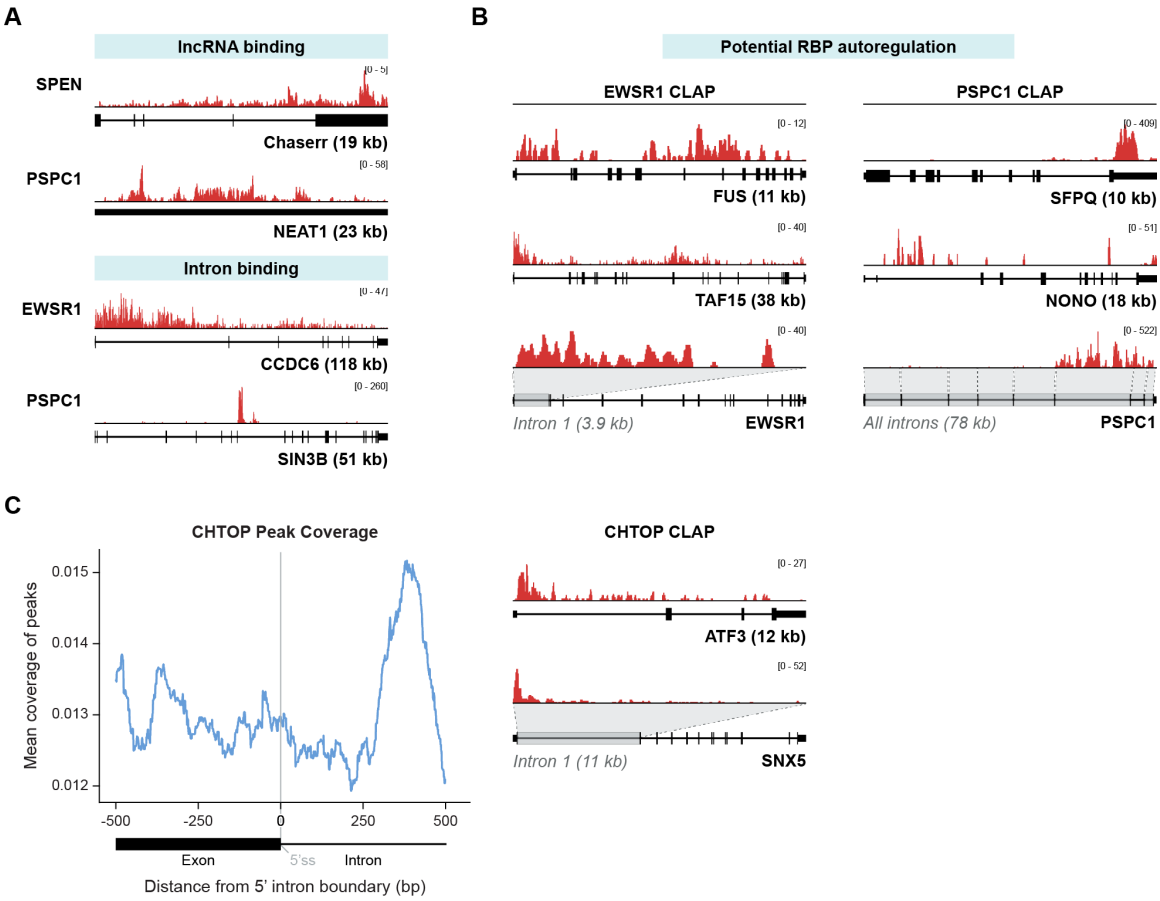


Figure S16: Specific chromatin-associated proteins bind to RNA *in vivo* (Related to Figure 6).

(A) Additional examples of binding events detected by CLAP (top, lncRNA; bottom, introns). CLAP enrichment profiles are plotted for SPEN across the Chaserr lncRNA, PSPC1 across the NEAT1 lncRNA, EWSR1 across the CCDC6 pre-mRNA, and PSPC1 across the SIN3B pre-mRNA. **(B)** Examples of potential autoregulatory binding of EWSR1 and PSPC1 to their own mRNAs in addition to mRNAs of closely related family members. CLAP enrichment profiles are shown for EWSR1 across the FUS, EWSR1 (first intron, ~3.9 kb), and TAF15 mRNAs (FET family), and PSPC1 across the SFPQ, NONO, and PSPC1 (all introns, spanning ~78 kb) mRNAs (DBHS family). **(C)** (Left) Metagene plot displaying mean coverage of CHTOP peaks (y-axis) across all exon-intron boundaries (x-axis), spanning +/- 500 nucleotides flanking the 5' splice site centered at 0 (5'ss, grey line). (Right)

CLAP enrichment profiles for CHTOP across the ATF3 pre-mRNA (top) and SNX5 pre-mRNA (bottom, first intron, ~11 kb).

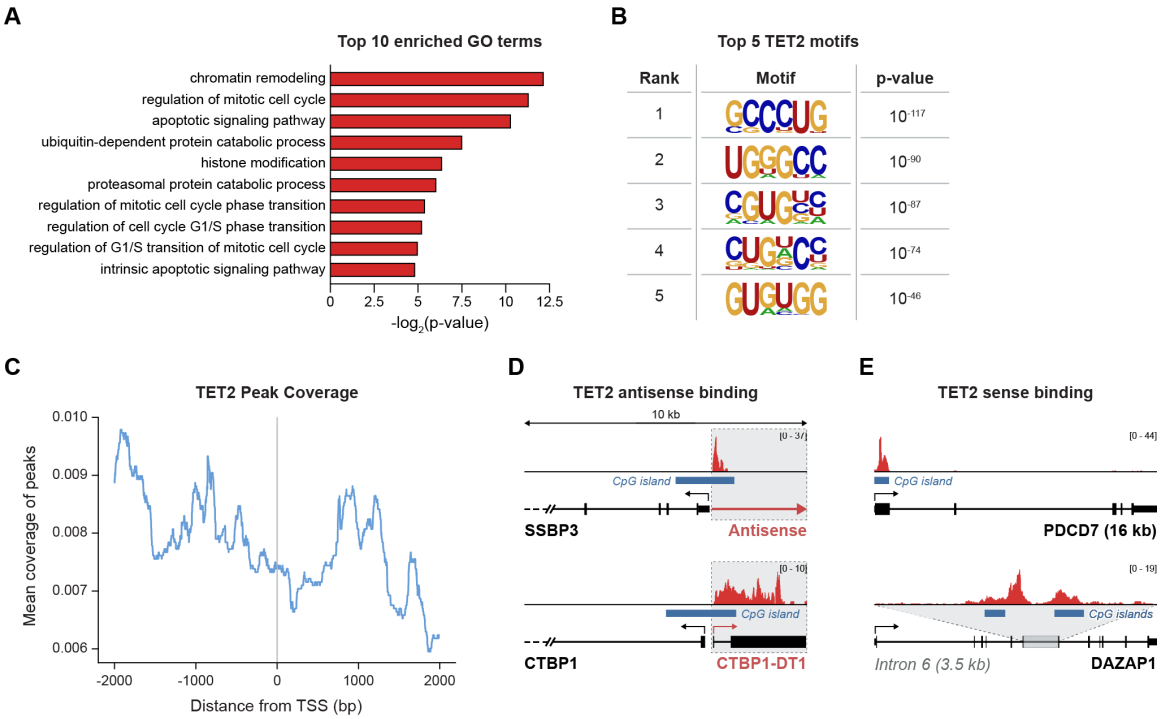


Figure S17: TET2 is a chromatin-associated RNA-binding protein (Related to Figure 6).

(A) Top 10 enriched GO terms (y-axis, biological process) for the mRNAs containing TET2 binding sites identified within RefSeq genes. *p*-values are from the binomial distribution (log scale, x-axis). **(B)** Top 5 significant motifs identified within TET2 CLAP peaks (right, *p*-value threshold < 10^{-40}). **(C)** Metagene plot displaying mean coverage of TET2 peaks (y-axis) across +/- 2,000 nucleotides flanking all transcription start sites (TSS) centered at 0 (grey line). **(D)** Examples of TET2 antisense RNA binding. CLAP enrichment profiles are shown for TET2 across the antisense RNA transcribed from the SSBP3 promoter (top) and CTBP1 Divergent Transcript (bottom, CTBP1-DT). Blue boxes indicate CpG islands, arrows indicate direction of transcription (black, sense; red, antisense). **(E)** CLAP enrichment profiles for TET2 across the PDCD7 mRNA (top) and DAZAP1 pre-mRNA (bottom, intron 6, ~3.5 kb). Blue boxes indicate CpG islands, black arrows indicate direction of transcription.

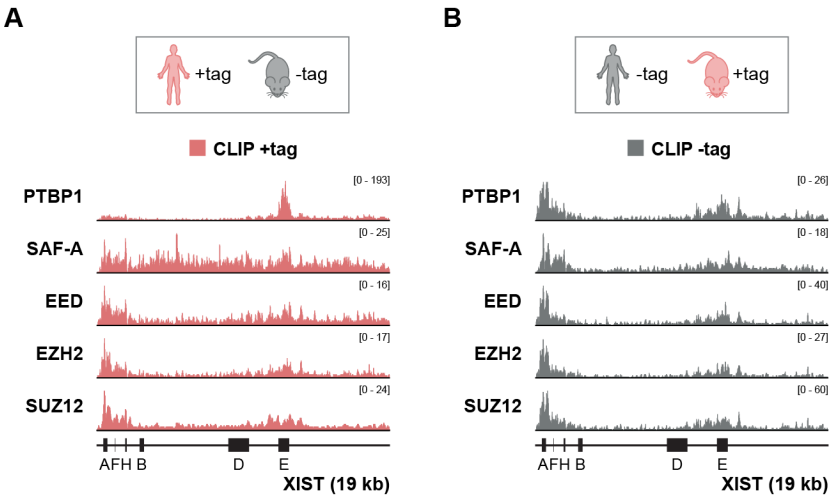


Figure S18: Background CLIP signal is non-random and comparable across different proteins (Related to Discussion and Figure 1).

(A) CLIP enrichments for PTBP1, SAF-A, EED, EZH2, and SUZ12 plotted across the human XIST lncRNA in the +tag experiments (red). Panel is reproduced from Figure 4B. **(B)** Same as in (A), but for -tag experiments (grey). EED, EZH2, and SUZ12 panels are reproduced from Figure 1C.

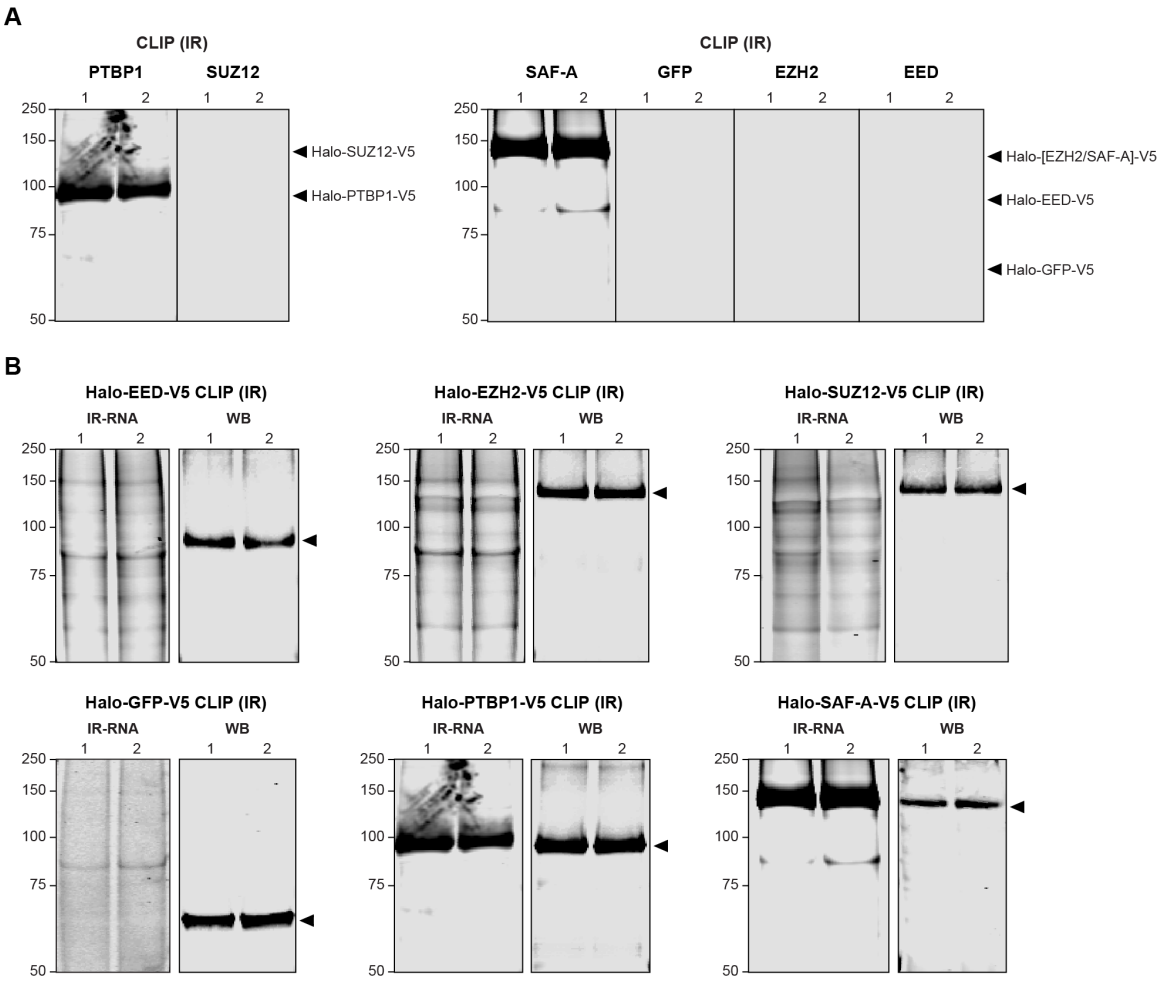


Figure S19: CLIP signal for PRC2 components is dramatically lower relative to *bona fide* RBPs (Related to Discussion and Figure 1).

(A) CLIP was performed on lysates from UV-crosslinked cells expressing proteins tagged with an N-terminal Halo and C-terminal V5. RNase I (NEB, RNase If) was used at a fixed dilution of 1:500 to digest RNA. Captured RNA from protein complexes was end-labeled with a phosphorylated cytidine conjugated to an infrared dye (see **STAR Methods**), run on an SDS-PAGE gel, and imaged directly on a LI-COR Odyssey. Black arrows indicate the expected size of the immunoprecipitated protein. (Left) Halo-V5-tagged PTBP1 and SUZ12 were run on the same gel and imaged at identical contrast. (Right) Halo-V5-tagged SAF-A, GFP, EZH2, and EED were run on the same gel and imaged at identical contrast. (B) Same

gels as in (A), except individually imaged at high contrast in all cases in order to visualize low amounts of infrared RNA (IR-RNA) signal. Corresponding western blots are shown alongside each IR-RNA gel depicting successful protein capture. Western blots are reproduced from Figure S1C.

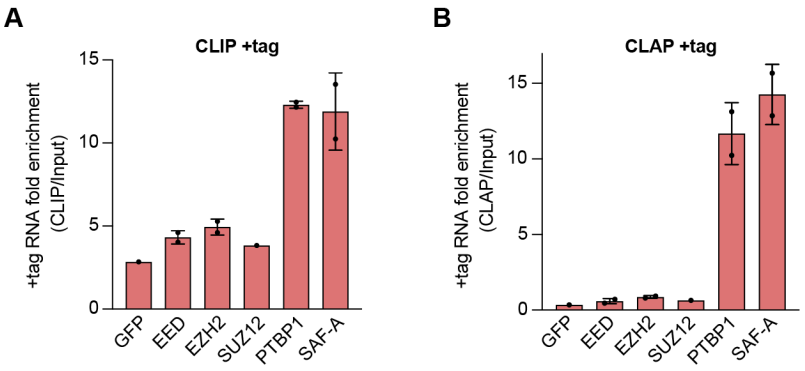


Figure S20: Both CLIP and CLAP quantitatively separate PRC2 from *bona fide* RBPs (Related to Discussion, Figure 1, and Figure 4).

(A) Fold enrichment of +tag RNAs in CLIP (relative to +tag RNA proportions in the input) for each tagged protein (GFP, EED, EZH2, SUZ12, PTBP1, SAF-A) where the protein was expressed in human cells (error bars = S.D.). **(B)** Same as in (A), but for CLAP (error bars = S.D.).

2.7 METHODS

2.7.1 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[168].

2.7.2 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)[159]. 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[77].

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[97] 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[169] (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 GGGGACCACTTTGTACAAGAAAGCTGGGTTtacaaatgtgttgtaaggccc; template: pcDNA3-Tet2, a gift from Yi Zhang[170] (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[77]. 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 µg of DNA. The mixture was transfected with the following settings using 100µL 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 µg/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^{-2} (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 µL of lysate was diluted to 1X final concentration of LDS loading buffer (4 µL 10X Bolt reducing agent, 10 µL 4X NuPAGE LDS Sample Buffer, 4 µL H2O), 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 µg of Rabbit IgG (CST, 2729S), anti-EZH2 (Active Motif, 39933), or anti-SUZ12 (Active Motif, 39357) antibodies were coupled to 50 µL of Protein G beads (Dynabeads) at room temperature for 30 minutes. Beads were washed three times with 300 µL 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 1mL 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 µL of mammalian lysis buffer + 2 µL 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 70ºC 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 µL/minute to block remaining active sites on the chip.

HEK293T cells transfected with DasherGFP (ATUM Biosciences, FPB-27-609), DasherGFP-3x-λN, 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 µg 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 μM 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 μL/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 μL/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)

eCLIP was performed as previously described[71], 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 µg of antibody coupled to 40 µL of Protein G beads in 100 µL 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 MgCl2, 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-lauroylsacrosine, 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[71], 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 (**Figure S3A, 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[21,22]. This preferential RNA signal near the size of the protein itself has been shown to occur for other non-RBPs[171].

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[22] 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[20,22].

(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[67]. 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%)[63].

(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*[5,45,58,59] (**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[171,172] observe this qualitatively with CLIP gels of PRC2 and other non-RBPs (**Figure S3A**, **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 µL of flowthrough lysate and adding 500 µL 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)

CLAP was performed on HEK293T and pSM33 cells as previously described[77]. 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 $CaCl_2$, 2.5 mM $MnCl_2$). 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 µL of 25% HaloLink Resin (50 µL 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 µg/mL Random 9-mer, 100 µg/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 µL 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 µL of ProK/NLS buffer + 10 µL of Proteinase K (NEB) and incubated at 50°C for 20 minutes while shaking at 1200 rpm. Elutions were then transferred to microspin 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 µL FastAP mix (1X Fast AP buffer, 8 µL FastAP Enzyme, 2 µL Murine RNase Inhibitor, 5 µL TURBO DNase) at 37°C for 15 min. RNA was then end repaired using T4 PNK. 300 µL of the end repair reaction buffer (1X T4 PNK Buffer, 7 µL T4 PNK Enzyme, 1 µL Murine RNase Inhibitor, 1 µL 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 µL (20%) of beads or resin from the previous step and removed the supernatant. RNA was radiolabeled using 4 µL of hot PNK mix (0.2 µL T4 PNK, 0.4 µL γ-^{32}P-ATP, 0.4 µL

10x PNK buffer, 3 µL H2O) 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 µL of high salt wash buffer, low salt wash buffer, and 1X PNK buffer. Samples were then either cleaved with TEV protease (CLAP samples) or directly resuspended in 40 µL 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}P CLIP and CLAP experiments are deposited in Mendeley Data (doi: 10.17632/wmsbzv6kg5.2).

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 µL of each sample (input, flowthrough, and washes) was combined with 1.5 µL 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 µL 10X Bolt reducing agent, 10 µL 4X NuPAGE LDS Sample Buffer, 4 µL H2O), 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, 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 µL of Elution Buffer and split into a 75 µL (ProK elution) and 8 µL (TEV elution) reaction. 25 µL of 4X ProK/NLS Buffer and 10 µL 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 µL 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 λN protein and BoxB RNA (K_D of 3.9 x 10^{-8} M) and PTBP1 with its known Xist binding site on the E-repeat (K_D of 6.00 x 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 [59] 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 µL ice cold lysis buffer supplemented with 20 µL Protease Inhibitor (Promega PIC) and 5 µL of RiboLock RNase inhibitor, followed by incubation on ice for 3 minutes. Each lysate was then bound to 50 µL 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 µL binding buffer as previously described[5,58], with slight modifications (50 mM Tris-HCl pH 7.5, 100 mM KCl, 5 mM $MgCl_2$, 0.1 mM $CaCl_2$). *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^{-2} (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 µg of recombinant PRC2 complex (Active Motif, 31337) was resuspended in 50 µL of native lysis buffer (1X HBS-EP+ Buffer supplemented with 100 mM KCl), then incubated with 25 µL 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^{-2} (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 [55,106,168] (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 (SpyCLAP)

CLAP on endogenous Spy-tagged proteins was performed previously described[77] 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[77]. Per capture, 250 µg of HaloTag-SpyCatcher was incubated with 50 µL 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 µL 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[173] + 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 µL 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 µL HBSS (Thermo Scientific) + 1X PIC and 2400 µL 2X RIPA + 1X PIC. Lysate was incubated overnight with Invitrogen M-280 Sheep Anti-Rabbit IgG Dynabeads coupled with 5 µg of H3K27me3 antibody (Active Motif, 39155) or 5 µg 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, 2µg/mL). After overnight Xist induction with 2 µg/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 µm; step size, 0.2 µm).

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 μg/mL.

Expression confirmation

Expression testing of Halo-tagged constructs was performed as previously described[77]. 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[55,68,78,105]. 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[78].

CLAP on Halo-tagged SPEN

CLAP on pSM33 cells expressing Halo-V5-tagged SPEN was performed as previously described[77] with slight modifications. After overnight Xist induction with 2 μg/mL doxycycline and UV-crosslinking, cell pellets were resuspended in 1 mL ice-cold Gagnon HLB buffer[173] 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[71], 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 µg of antibody coupled to 40 µL 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 µL of 20 µM 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 were treated as previously described[71,174]. 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 µL 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[101]. 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 ≥ 35 x 35 nucleotides or Illumina NextSeq 2000 with read length ≥ 50 x 50 nucleotides. Sequencing depth for each sample is reported in terms of raw read counts in **Table S3.**

2.7.3 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[175]. 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[176] 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 $< 10^{-6}$ 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 was visualized using the Integrative Genomics Visualizer[177] (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 ^{32}P gels

Full-length ^{32}P gel images were imported into ImageJ[178] for quantification. For each gel lane, the straight-line tool was used to measure the ^{32}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[179], 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 (doi: 10.17632/wmsbzv6kg5.2).

Mean coverage of binding sites

Visualization of binding site coverage across transcript features was performed using the RCAS tool[180] (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.

2.8 SUPPLEMENTAL INFORMATION

Table S1. ORFeome entry clones of screened chromatin proteins, related to Figure 6 and STAR Methods; can be found online[181].

Table S2. Sequences of reagents used for CRISPR-targeting, related to STAR Methods; can be found online[181].

Table S3. Read count statistics for all CLIP and CLAP samples, related to STAR Methods; can be found online[181].

2.9 REFERENCES

1. Kwon, S.C., Yi, H., Eichelbaum, K., Fohr, S., Fischer, B., You, K.T., Castello, A., Krijgsveld, 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. 10.1038/nsmb.2638.
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. 10.1016/j.cell.2012.04.031.
3. Castello, A., Hentze, M.W., and Preiss, T. (2015). Metabolic Enzymes Enjoying New Partnerships as RNA-Binding Proteins. Trends in Endocrinology & Metabolism *26*, 746–757. 10.1016/j.tem.2015.09.012.
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. 10.1016/j.molcel.2016.09.034.
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. 10.1038/nsmb.2679.
6. Sigova, A.A., Abraham, B.J., Ji, X., Molinie, 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 (1979) *350*, 978–981. 10.1126/science.aad3346.
7. Kung, J.T., Kesner, B., An, J.Y., Ahn, J.Y., Cifuentes-Rojas, C., Colognori, 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. 10.1016/j.molcel.2014.12.006.
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. 10.1101/gad.236869.113.

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 (1979) *322*, 750–756. 10.1126/science.1163045.

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., et al. (2007). Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell *129*, 1311–1323. S0092-8674(07)00659-9 [pii] 10.1016/j.cell.2007.05.022.

11. Rinn, J.L., and Chang, H.Y. (2012). Genome regulation by long noncoding RNAs. Annu Rev Biochem *81*, 145–166. 10.1146/annurev-biochem-051410-092902.

12. Pandey, R.R.R.R., Mondal, T., Mohammad, F., Enroth, S., Redrup, L., Komorowski, J., Nagano, T., Mancini-Dinardo, D., Kanduri, C., Amaral, P.P., et al. (2008). Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. Mol Cell *32*, 232–246. 10.1016/j.molcel.2008.08.022.

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. 10.1101/gad.1983810.

14. Jeon, Y., and Lee, J.T. (2011). YY1 Tethers Xist RNA to the Inactive X Nucleation Center. Cell *146*, 119–133. S0092-8674(11)00668-4 [pii] 10.1016/j.cell.2011.06.026.

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. 10.1016/j.molcel.2010.12.011.

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. 10.1038/nature10398.

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. Proceedings of the National Academy of Sciences *106*, 11667–11672. 10.1073/pnas.0904715106.

18. Margueron, R., and Reinberg, D. (2011). The Polycomb complex PRC2 and its mark in life. Nature *469*, 343–349. 10.1038/nature09784.

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. 10.1038/nrm2763.

20. Kaneko, S., Bonasio, R., Saldana-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. 10.1016/j.molcel.2013.11.012.

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. 10.1038/nsmb.2700.

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. 10.1101/gr.197632.115.

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. 10.1038/s41588-020-0662-x.

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 (1979) *381*, 1331–1337. 10.1126/science.adh0059.

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. 23/16/1831 [pii] 10.1101/gad.1811209.

27. Guttman, M., and Rinn, J.L.L. (2012). Modular regulatory principles of large non-coding RNAs. Nature *482*, 339–346. 10.1038/nature10887 nature10887 [pii].

28. Koziol, M.J., and Rinn, J.L. (2010). RNA traffic control of chromatin complexes. Curr Opin Genet Dev *20*, 142–148. 10.1016/j.gde.2010.03.003.

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. 10.18632/oncotarget.5728.

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. 10.1016/j.cell.2016.11.041.

31. Di Ruscio, A., Ebralidze, 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. 10.1038/nature12598.

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. 10.1186/s13059-016-0878-3.

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, rna.079479.122. 10.1261/rna.079479.122.

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. 10.1016/j.cels.2017.09.014.

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. 10.7554/eLife.02046.

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. 10.1038/s41467-020-15571-8.

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. 10.1093/nar/gkaa067.

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. 10.1016/j.molcel.2023.06.012.

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. 10.1016/j.cell.2019.06.001.

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. 10.1016/j.molcel.2019.07.039.

41. Saldaña-Meyer, R., Rodriguez-Hernaez, J., Escobar, T., Nishana, M., Jácome-López, K., Nora, E.P., Bruneau, B.G., Tsirigos, A., Furlan-Magaril, M., Skok, J., et al. (2019). RNA Interactions Are Essential for CTCF-Mediated Genome Organization. Mol Cell *76*, 412-422.e5. 10.1016/j.molcel.2019.08.015.

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. 10.1016/j.jmb.2013.06.031.

43. Wutz, A. (2011). Gene silencing in X-chromosome inactivation: advances in understanding facultative heterochromatin formation. Nat Rev Genet *12*, 542–553. 10.1038/nrg3035.

44. Galupa, R., and Heard, E. (2015). X-chromosome inactivation: new insights into cis and trans regulation. Curr Opin Genet Dev *31*, 57–66. 10.1016/j.gde.2015.04.002.

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. 10.1016/j.molcel.2014.05.009.

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. 10.1038/ng820.

47. Lee, J.T. (2012). Epigenetic regulation by long noncoding RNAs. Science (1979) *338*, 1435–1439. 10.1126/science.1231776 338/6113/1435 [pii].

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. nature08975 [pii] 10.1038/nature08975.

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. 10.1016/j.devcel.2008.08.015.

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. 10.1371/journal.pgen.0020066.

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. 10.1038/sj.emboj.7601187.

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 (1979) *300*, 131–135. 10.1126/science.1084274 1084274 [pii].

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. 10.1016/j.molcel.2014.01.002.

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. 10.1371/journal.pbio.0020171.

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. 10.1038/nature14443.

56. Brockdorff, N. (2013). Noncoding RNA and Polycomb recruitment. RNA *19*, 429–442. 10.1261/rna.037598.112.

57. Portoso, M., Ragazzini, R., Brenčič, 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. 10.15252/embj.201695335.

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. 10.1016/j.molcel.2014.12.017.

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. 10.1016/j.molcel.2017.02.003.

60. Peritz, T., Zeng, F., Kannanayakal, T.J., Kilk, K., Eir?ksd?ttir, E., Langel, U., and Eberwine, J. (2006). Immunoprecipitation of mRNA-protein complexes. Nat Protoc *1*, 577–580. 10.1038/nprot.2006.82.

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. 10.1038/nprot.2006.47.

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. 10.1261/rna.7151404 rna.7151404 [pii].

63. Darnell, R.B. (2010). HITS-CLIP: panoramic views of protein-RNA regulation in living cells. Wiley Interdiscip Rev RNA *1*, 266–286. 10.1002/wrna.31.

64. Konig, 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. 10.1038/nsmb.1838.

65. Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano Jr., M., 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. 10.1016/j.cell.2010.03.009.

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. 10.1038/nature07488.

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 (1979) *302*, 1212–1215. 10.1126/science.1090095.

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. 10.1016/j.cell.2015.03.025.

69. Minajigi, A., Froberg, J.E., Wei, C., Sunwoo, H., Kesner, B., Colognori, D., Lessing, D., Payer, B., Boukhali, M., Haas, W., et al. (2015). A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation. Science (1979) *349*. 10.1126/science.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. 10.1038/s41594-020-00535-9.

71. Van Nostrand, E. V, Pratt, G.A., Shishkin, A.A., Gelboin-Burkhart, C., Fang, M.Y., Sundararaman, 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. 10.1038/nmeth.3810.

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. 10.1016/j.molcel.2010.03.019.

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. 10.1158/0008-5472.CAN-14-2931.

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. 10.1016/j.celrep.2022.110928.

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. 10.1038/ncomms12209.

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. 10.1016/j.cell.2015.12.017.

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. 10.1016/j.cell.2020.10.004.

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. 10.1038/s41594-022-00739-1.

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. 10.1016/j.gpb.2018.04.003.

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. 10.1016/j.molcel.2020.05.009.

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. 10.1038/s41467-018-03575-4.

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. 10.1021/cb800025k.

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. Proceedings of the National Academy of Sciences *109*, E690–E697. 10.1073/pnas.1115485109.

84. Ray, D., Kazan, H., Chan, E.T., 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. 10.1038/nbt.1550.

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. 10.1093/nar/gkx460.

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. 10.1038/ncomms11212.

87. Castello, A., Fischer, B., Frese, C.K., Horos, R., Alleaume, A.-M., Foehr, S., Curk, T., Krijgsveld, J., and Hentze, M.W. (2016). Comprehensive Identification of RNA-Binding Domains in Human Cells. Mol Cell *63*, 696–710. 10.1016/j.molcel.2016.06.029.

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., Schwinn, M.K., et al. (2016). SONAR Discovers RNA-Binding Proteins from Analysis of Large-Scale Protein-Protein Interactomes. Mol Cell *64*, 282–293. 10.1016/j.molcel.2016.09.003.

89. Trendel, J., Schwarzl, T., Horos, R., Prakash, A., Bateman, A., Hentze, M.W., and Krijgsveld, J. (2019). The Human RNA-Binding Proteome and Its Dynamics during Translational Arrest. Cell *176*, 391-403.e19. 10.1016/j.cell.2018.11.004.

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. 10.1038/nature09819.

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. 10.1038/s41592-019-0407-x.

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 U S A *112*, E6456–E6465. 10.1073/pnas.1518552112.

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. 10.1016/j.celrep.2016.04.085.

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. 10.1016/j.cell.2021.11.012.

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. 10.1016/j.cell.2017.11.008.

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. 10.1038/nature16952.

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. 10.1093/nar/gkaa1040.

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. 10.1016/j.molcel.2019.04.018.

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., et al. (2019). Systematic Discovery of Endogenous Human Ribonucleoprotein Complexes. Cell Rep *29*, 1351-1368.e5. 10.1016/j.celrep.2019.09.060.

100. Gene Ontology Consortium (2004). The Gene Ontology (GO) database and informatics resource. Nucleic Acids Res *32*, 258D – 261. 10.1093/nar/gkh036.

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. 10.1016/j.cell.2021.10.014.

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. 10.1016/j.celrep.2015.06.053.

103. Monfort, A., Minin, G. Di, Postlmayr, A., Arieti, F., and Wutz, A. (2015). Identification of Spen as a Crucial Factor for Xist Function through Forward Genetic Screening in Haploid Embryonic Stem. Cell Rep, 1–8. 10.1016/j.celrep.2015.06.067.

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. 10.1038/s41467-020-20040-3.

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*. 10.7554/eLife.54508.

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 (1979) *354*, 468–472. 10.1126/science.aae0047.

107. Kanduri, C. (2011). Kcnq1ot1: A chromatin regulatory RNA. Semin Cell Dev Biol *22*, 343–350. 10.1016/j.semcdb.2011.02.020.

108. Nagano, T., and Fraser, P. (2009). Emerging similarities in epigenetic gene silencing by long noncoding RNAs. Mammalian Genome *20*, 557. 10.1007/s00335-009-9218-1.

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. 10.1038/s41467-019-13075-8.

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. 10.1146/annurev-biochem-060614-034325.

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. 10.1128/MCB.18.3.1489.

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. 10.1261/rna.078827.121.

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. 10.1093/nar/gkx912.

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. 10.1038/s41586-020-2077-3.

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. 10.1186/s13059-020-01982-9.

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. 10.1016/j.celrep.2013.11.017.

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 Fibrous Polymers of Low-Complexity Domains. Cell *155*, 1049–1060. 10.1016/j.cell.2013.10.033.

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. 10.1038/s41589-021-00904-5.

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. 10.1002/gcc.20858.

122. Luo, Y., Blechingberg, 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. 10.1186/s12864-015-2125-9.

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. 10.1016/j.celrep.2014.08.071.

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. 10.1038/nsmb.1568.

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. 10.1016/j.molcel.2019.04.034.

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. 10.1038/emboj.2012.342.

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., et al. (2009). Conversion of 5-

Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. Science (1979) *324*, 930–935. 10.1126/science.1170116.

128. Yildirim, O., Li, R., Hung, J.-H., Chen, P.B., Dong, X., Ee, L.-S., Weng, Z., Rando, O.J., and Fazzio, T.G. (2011). Mbd3/NURD Complex Regulates Expression of 5-Hydroxymethylcytosine Marked Genes in Embryonic Stem Cells. Cell *147*, 1498–1510. 10.1016/j.cell.2011.11.054.

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*. 10.7554/eLife.61503.

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. 10.1093/nar/gkx597.

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. 10.1016/j.cell.2006.10.044.

132. Hirose, T., Virnicchi, G., Tanigawa, A., Naganuma, T., Li, R., Kimura, H., Yokoi, T., Nakagawa, S., Bénard, M., Fox, A.H., et al. (2014). NEAT1 long noncoding RNA regulates transcription via protein sequestration within subnuclear bodies. Mol Biol Cell *25*, 169–183. 10.1091/mbc.e13-09-0558.

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. 10.1093/nar/gkw271.

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. 10.1038/s41588-018-0060-9.

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. 10.1016/j.molcel.2009.01.026.

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. 10.1172/JCI89484.

137. Tan, L., and Shi, Y.G. (2012). Tet family proteins and 5-hydroxymethylcytosine in development and disease. Development *139*, 1895–1902. 10.1242/dev.070771.

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 (1979) *333*, 1300–1303. 10.1126/science.1210597.

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. 10.1038/nature11742.

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. 10.1038/emboj.2012.357.

141. Peng, L., Li, Y., Xi, Y., Li, W., Li, J., Lv, R., Zhang, L., Zou, Q., Dong, S., Luo, H., et al. (2016). Methyl-CpG-binding domain protein 3-like 2 (MBD3L2) promotes Tet2 enzymatic activity for mediating 5mC oxidation. J Cell Sci. 10.1242/jcs.179044.

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*. 10.1126/sciadv.aau6986.

143. Wang, X., Rosikiewicz, W., Sedkov, Y., Martinez, T., Hansen, B.S., Schreiner, P., Christensen, J., Xu, B., Pruett-Miller, S.M., Helin, K., et al. (2022). PROSER1 mediates TET2 O-GlcNAcylation to regulate DNA demethylation on UTX-dependent enhancers and CpG islands. Life Sci Alliance *5*, e202101228. 10.26508/lsa.202101228.

144. Ko, M., An, J., Bandukwala, H.S., Chavez, L., Äijö, 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. 10.1038/nature12052.

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. 10.1038/s41594-020-00526-w.
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. 10.1038/s41467-020-18729-6.
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. 10.1016/j.cell.2018.05.024.
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. 10.1016/j.molcel.2017.01.009.
149. Cheutin, T., and Cavalli, G. (2014). Polycomb silencing: from linear chromatin domains to 3D chromosome folding. Curr Opin Genet Dev *25C*, 30–37. 10.1016/j.gde.2013.11.016.
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. 10.1016/j.cell.2010.12.026.
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. 10.1038/s41594-019-0197-y.

152. Cherney, R.E., Mills, C.A., Herring, L.E., Braceros, 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*. 10.1242/bio.059955.

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. 10.1016/j.molcel.2019.10.001.

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. bioRxiv, 2023.08.16.553488. 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. bioRxiv, 2023.08.16.553503. 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. 10.1038/s41592-022-01629-6.

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. 10.1038/nsmb.1545.

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. 10.1016/j.molcel.2016.02.012.

159. Hafner, M., Katsantoni, M., Köster, T., Marks, J., Mukherjee, J., Staiger, D., Ule, J., and Zavolan, M. (2021). CLIP and complementary methods. Nature Reviews Methods Primers *1*, 20. 10.1038/s43586-021-00018-1.

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. 10.1186/gb-2014-15-1-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. bioRxiv, 2023.06.05.543769. 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. 10.1016/j.cell.2016.03.007.

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. 10.1016/j.cell.2019.05.027.

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. 10.1016/j.molcel.2019.07.030.

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*. 10.7554/eLife.29224.

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. 10.1038/s41592-022-01618-9.

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. WIREs RNA *9*. 10.1002/wrna.1436.

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 (1979) *341*. 10.1126/science.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. 10.1038/nmeth.1638.

170. Wang, Y., and Zhang, Y. (2014). Regulation of TET Protein Stability by Calpains. Cell Rep *6*, 278–284. 10.1016/j.celrep.2013.12.031.

171. 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*. 10.1002/cpz1.659.

172. 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. 10.1038/s41467-021-21623-4.

173. 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. 10.1038/nprot.2014.135.

174. 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. 10.1038/nmeth.4100.

175. 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. 10.1093/bioinformatics/bts635.

176. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods *9*, 357–359. 10.1038/nmeth.1923.

177. 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. 10.1038/nbt.1754.

178. 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. 10.1186/s12859-017-1934-z.

179. Grant, C.E., Bailey, T.L., and Noble, W.S. (2011). FIMO: scanning for occurrences of a given motif. Bioinformatics *27*, 1017–1018. 10.1093/bioinformatics/btr064.

180. 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–e91. 10.1093/nar/gkx120.

181. Guo, J.K., Blanco, M.R., Walkup, W.G., Bonesteele, G., Urbinati, C.R., Banerjee, A.K., Chow, A., Ettlin, O., Strehle, M., Peyda, P., et al. (2024). Denaturing purifications demonstrate that PRC2 and other widely reported chromatin proteins do not appear to bind directly to RNA in vivo. Mol Cell 84, 1271-1289.e12. 10.1016/j.molcel.2024.01.026.

C h a p t e r 3

SPIDR: A HIGHLY MULTIPLEXED METHOD FOR MAPPING RNA-PROTEIN INTERACTIONS UNCOVERS A POTENTIAL MECHANISM FOR SELECTIVE TRANSLATIONAL SUPPRESSION UPON CELLULAR STRESS

Erica Wolin*, Jimmy K. Guo*, Mario R. Blanco, Andrew A. Perez, Isabel N. Goronzy, Ahmed A. Abdou, Darvesh Gorhe, Mitchell Guttman, Marko Jovanovic

*Equal contribution

A modified version of this chapter was published as “SPIDR: a highly multiplexed method for mapping RNA-protein interactions uncovers a potential mechanism for selective translational suppression upon cellular stress” in *bioRxiv*. doi: 10.1101/2023.06.05.543769.

3.1 ABSTRACT

RNA binding proteins (RBPs) play crucial roles in regulating every stage of the mRNA life cycle and mediating non-coding RNA functions. Despite their importance, the specific roles of most RBPs remain unexplored because we do not know what specific RNAs most RBPs bind. Current methods, such as crosslinking and immunoprecipitation followed by sequencing (CLIP-seq), have expanded our knowledge of RBP-RNA interactions but are generally limited by their ability to map only one RBP at a time. To address this limitation, we developed SPIDR (Split and Pool Identification of RBP targets), a massively multiplexed method to simultaneously profile global RNA binding sites of dozens to hundreds of RBPs in a single experiment. SPIDR employs split-pool barcoding coupled with antibody-bead barcoding to increase the throughput of current CLIP methods by two orders of magnitude. SPIDR reliably identifies precise, single-nucleotide RNA binding sites for diverse classes of RBPs simultaneously. Using SPIDR, we explored changes in RBP binding upon mTOR inhibition and identified that 4EBP1 acts as a dynamic RBP that selectively binds to 5'-untranslated regions of specific translationally repressed mRNAs only upon mTOR inhibition. This observation provides a potential mechanism to explain the specificity of translational regulation controlled by mTOR signaling. SPIDR has the potential to revolutionize our understanding of RNA biology and both transcriptional and post-transcriptional gene regulation by enabling rapid, *de novo* discovery of RNA-protein interactions at an unprecedented scale.

3.2 INTRODUCTION

RNA binding proteins (RBPs) play key roles in controlling all stages of the mRNA life cycle, including transcription, processing, nuclear export, translation, and degradation[1–5]. Recent estimates suggest that up to 30% of all human proteins (several thousand in total) bind to RNA[6–10], indicative of their broad activity and central importance in cell biology. Moreover, mutations in RBPs have been causally linked to various human diseases, including

immunoregulatory and neurological disorders as well as cancer[2–4,11]. Yet, we still do not know what specific roles most of these RBPs play because the RNAs they bind remain mostly unknown.

In addition, there are many thousands of regulatory non-coding RNAs (ncRNAs) whose functional roles remain largely unknown[12,13]; understanding how they work requires defining the proteins to which they bind[13–15]. For example, uncovering the mechanism by which the Xist long noncoding RNA (lncRNA) silences the inactive X chromosome required identification of the SPEN/SHARP RBP that binds to Xist[16–20] – a process that took >25 years after the lncRNA was discovered[14]. Given the large discrepancy between the number of ncRNAs and putative RBPs identified, and the number of RNA-protein interactions demonstrated to be functionally relevant, there is an urgent need to generate high-resolution binding maps to enable functional characterization[14].

Currently, the most rigorous and widely utilized method to characterize RBP-RNA interactions is crosslinking and immunoprecipitation followed by next generation sequencing (CLIP-seq)[21–26]. Briefly, CLIP works by utilizing UV light to covalently crosslink RNA and directly interacting proteins, followed by cell lysis, immunoprecipitation under stringent conditions (e.g., 1M salt) to purify a protein of interest followed by gel electrophoresis, transfer to a nitrocellulose membrane, and excision of the protein-RNA complex prior to sequencing and identification of the bound RNAs. CLIP and its related variants have greatly expanded our knowledge of RNA-RBP interactions and our understanding of gene expression from mRNA splicing to microRNA targeting[21–26].

Yet, CLIP and all of its variants (with one recent exception[27] which we discuss in more detail below; see **Note 1**) are limited to mapping a single RBP at a time. As such, efforts to generate reference maps for hundreds of RBPs in even a limited number of cell types have required major financial investment and the work of large teams working in international consortiums (e.g., ENCODE)[23,28,29]. Despite these herculean efforts and the important advances they have enabled, there are critical limitations: (i) Only a small fraction of the total number of predicted RBPs have been successfully mapped using genome-wide methods (ENCODE has

so far characterized the binding patterns of < 10% of known RBPs); (ii) Of these, most have been mapped in only a small number of cell lines (mainly K562 and HepG2); (iii) Because each protein map is generated from an individual experiment, a large number of cells is required to map dozens, let alone hundreds, of RBPs – this is particularly challenging for studying primary cells, disease models, or other populations of rare cells. Further, because these datasets are highly cell type-specific, the generated maps are not likely to be directly useful for studying these RBPs within other cell-types or model systems (e.g., patient samples, animal models, or perturbations). Thus, it is critically important to enable the generation of comprehensive RBP binding for any cell type of interest in a manner that is accessible to any individual lab.

To overcome these challenges, we developed SPIDR (Split and Pool Identification of RBP targets), a massively multiplexed method to simultaneously profile the global RNA binding sites of dozens to hundreds of RBPs in a single experiment. SPIDR is based on our split-pool barcoding strategy that maps multiway nucleic acid interactions using high throughput sequencing[30–32]; the vastly simplified version of split-pool barcoding we present here, when combined with antibody-bead barcoding, increases the throughput of current CLIP methods by two orders of magnitude. Using this approach, we can reliably identify the precise, single nucleotide RNA binding sites of dozens of RBPs simultaneously and can detect changes in RBP binding upon perturbation. Using this approach, we uncovered a mechanism driven by dynamic RBP binding to mRNA that may explain the specificity of translational regulation controlled by mTOR signaling. Thus, SPIDR enables rapid, *de novo* discovery of RNA-protein interactions at an unprecedented scale and has the potential to transform our understanding of RNA biology and both transcriptional and post-transcriptional gene regulation.

3.3 RESULTS

3.3.1 SPIDR: A highly multiplexed method for mapping RBP-RNA interactions

We developed SPIDR to enable highly multiplexed mapping of RBPs to individual RNAs transcriptome-wide. Briefly, SPIDR involves: (i) generating highly multiplexed antibody-bead pools by tagging individual antibody-bead conjugates with a specific oligonucleotide (tagged bead pools), (ii) performing RBP purification using these tagged antibody-bead pools in UV-crosslinked cell lysates, and (iii) linking individual antibodies to their associated RNAs using split-and-pool barcoding (**Figure 1A** and **Supplemental Figure 1**).

We first devised a highly modular scheme to generate hundreds of tagged antibody-beads such that each unique bead population is labeled with a specific oligonucleotide tag and all bead populations are combined to generate an antibody-bead pool (**Figure 1A** and **Supplemental Figure 1**). Because this approach does not require direct chemical modification of the antibody, we can utilize any antibody (in any storage buffer) and rapidly link it to a defined sequence on a bead at high efficiency using the same coupling procedure utilized in traditional CLIP-based approaches (see **Methods**). Using this pool, we perform on-bead immunopurification (IP) of RBPs in UV-crosslinked lysates using standard conditions and assign individual protein identities to their associated RNAs using split-and-pool barcoding, where the same barcode strings are added to both the oligonucleotide bead tag and immunopurified RNA (**Figure 1A**). We dramatically simplified our split-and-pool tagging method such that the entire protocol can be performed without the need for specialized equipment in ~1 hour (see **Methods**).

After split-and-pool tagging and subsequent library preparation, we sequenced all barcoded DNA molecules (antibody-bead tags and the converted cDNA of RNAs bound to corresponding RBPs). We then matched all antibody-bead tags and RNA reads by their shared barcodes; we refer to these as SPIDR clusters (**Figure 1A**). We merged all SPIDR clusters by protein identity (specified by the antibody-bead tag) to generate a high-depth binding map for each protein. The resulting datasets are analogous to those generated by traditional individual CLIP approaches.

To ensure that IP using a pool containing multiple antibodies can successfully and specifically purify each of the individual proteins, we performed an IP in K562 cells using a

pool of antibodies against 39 RBPs and measured the purified proteins by liquid chromatography tandem mass spectrometry (LC-MS/MS). We confirmed that 35 of the 39 targeted RBPs enriched at least 2-fold relative to a negative control, showing that multiplexed enrichment of several RBPs simultaneously is possible (**Supplemental Figure 2**). The few exceptions were RBPs that were simply not detected (neither in the pooled IP nor under control conditions) and likely reflect either a poor antibody or lack of RBP expression in this cell line.

3.3.2 SPIDR accurately maps dozens of RBPs within a single experiment

To test whether SPIDR accurately maps RBPs to RNA, we performed SPIDR in two widely studied human cell lines (K562 and HEK293T cells). Specifically, we generated antibody bead pools containing 68 uniquely tagged antibody-beads targeting 62 distinct RBPs across the RNA life cycle, including splicing, processing, and translation factors (**Figure 1B**, **Supplemental Tables 1, 2**). As negative controls, we included antibodies against epitopes not present in endogenous human cells (GFP and V5), antibodies that lack affinity to any epitope (mouse IgG), and oligonucleotide-labeled beads lacking any antibody (empty beads).

Using these pools, we performed SPIDR on 10 million UV-crosslinked cells. Focusing on the K562 data (which were sequenced at greater depth), we generated a median of 4 oligonucleotide tags per SPIDR cluster with the majority of clusters (>80%) containing tags representing only a single antibody type (**Supplemental Figure 3**), indicating that there is minimal 'crosstalk' between beads in a SPIDR experiment. This specificity enables us to uniquely assign RNA molecules to their corresponding RBPs. After removing PCR duplicates, we assigned each sequenced RNA read to its associated RBP and identified high confidence binding sites by comparing read coverage across an RNA to the coverage in all other targets in the pooled IP (**Supplemental Figure 4, Supplemental Figure 5**, **Supplemental Table 1**; see **Methods** for details). Using this approach, we detected the precise binding sites for SAF-A, PTBP1, SPEN, and HNRNPK on the *XIST* RNA[17,20,23]

(**Figure 1C**). Although most proteins (38/53 RBPs in K562) contained more than 2 million mapped RNA reads (**Supplemental Figure 4**), we observed specific binding to known target sites even for RBPs with lower numbers of reads. For example, SLBP (Stem Loop Binding Protein) had only 1.5 million mapped reads yet displayed strong enrichment specifically at the 3' ends of histone mRNAs as expected[29] (**Figure 1D**).

To systematically assess the quality, accuracy, and resolution of our SPIDR binding maps and the scope of the SPIDR method, we explored several key features:

(i) Accurate mapping of classical RNPs. We targeted RBPs of diverse functionality, such as those which bind preferentially to RNAs coding for proteins and/or lncRNAs, to introns, exons, miRNAs, etc., as well as more "classical" ribonuclear protein (RNP) complexes, such as the ribosome or spliceosome (**Figure 2A**). We observed precise binding to the expected RNAs and binding sites. For example, we observed binding of:

- LSM11 to the U7 small nuclear RNA (snRNA)[33] and the telomerase RNA component (TERC)[34] (**Figure 2A and 2B**).
- WDR43, a protein that is involved in ribosomal RNA (rRNA) processing, to the 45S pre-rRNA and the U3 small nucleolar RNA (snoRNA), which is involved in rRNA modification[35] (**Figure 2A and 2C**).
- LIN28B to a distinct region of the 45S pre-rRNA, consistent with recent reports of its role in ribosomal RNA biogenesis in the nucleolus[36] (**Figure 2A and 2C**).
- NOLC1 (also known as NOPP140), a protein that localizes within the nucleolus and Cajal bodies[37,38], to both the 45S pre-rRNA (enriched within the nucleolus) and various small Cajal-body associated RNAs (scaRNAs) (**Figure 2A**).
- DDX52, a DEAD-box protein that is predicted to be involved in the maturation of the small ribosomal subunit[39,40] and RPS3, a structural protein contained within the small ribosomal RNA subunit, to distinct sites on the 18S rRNA (**Figure 2A**).
- FUS and TAF15 to distinct locations on the U1 snRNA[41,42] (**Figure 2A**).
- SMNDC1 specifically to the U2 snRNA[43] (**Figure 2A**).

- SSB (also known as La protein) binding to tRNA precursors consistent with its known role in the biogenesis of RNA Polymerase III transcripts[44,45] (**Figure 2A**).
- LIN28B to the let-7 miRNA[46–50] (**Figure 2D)**.
- LARP7 binding to 7SK[51] (**Figure 2A**, **Supplemental Figure 5**).

(ii) Many RBPs bind their own mRNAs to autoregulate expression levels. Many RBPs have been reported to bind their own mRNAs to control their overall protein levels through post-transcriptional regulatory feedback[52–54]. For example, SPEN protein binds its own mRNA to suppress its transcription[55], UPF1 binds its mRNA to target it for Nonsense Mediated Decay[56], TARDBP binds its own 3'-UTR to trigger an alternative splicing event that results in degradation of its own mRNA[57,58], and DGCR8, which together with DROSHA forms the known microprocessor complex, binds a hairpin structure in *DGCR8* mRNA to induce cleavage and destabilization of the mRNA[59] (**Figure 2E**). In addition to these cases, we observed autoregulatory binding of proteins to their own mRNAs for nearly a third of our targeted RBPs (15 proteins) (**Supplemental Figure 6**).

(iii) Different antibodies that capture the same protein or multiple proteins within the same complex show similar binding. We considered the possibility that including antibodies against multiple proteins contained within the same complex, or that otherwise bind to the same RNA, within the same pooled sample could compete against each other and therefore limit the utility of large-scale multiplexing. However, we did not observe this to be the case; in fact, antibodies against different proteins known to occupy the same complex displayed highly comparable binding sites on the same RNAs. For example, DROSHA and DGCR8, two proteins that bind as part of the microprocessor complex, showed highly consistent binding patterns across known miRNA precursors with significant overlap in their binding sites (odds-ratio of 316-fold, hypergeometric p-value < 10^{-100}). Similarly, when we included two distinct antibodies targeting the same protein, HNRNPL, we observed highly comparable binding profiles for both antibodies (**Figure 2F**) and significant overlap in defined binding sites (odds-ratio of 15-fold, hypergeometric p-value < 10^{-100}). Taken together, our results indicate that SPIDR can be used to map different RBPs that bind to the

same RNA targets and can successfully map multiple antibodies targeting the same protein. As such, SPIDR may be a particularly useful tool for directly screening multiple antibodies targeting the same protein to evaluate utility for use in CLIP-like studies.

(iv) Transcriptome-wide SPIDR maps are highly comparable with CLIP. Because K562 represents the ENCODE-mapped cell line with the largest number of eCLIP datasets, we were able to benchmark our SPIDR results directly to those generated by ENCODE. To do this, we compared the profiles for each of the 33 RBPs that overlap between SPIDR and ENCODE datasets in K562 cells[23,28,29] (see **Methods**). We observed highly overlapping binding patterns for most RBPs, including HNRNPK binding to *POLR2A* (**Figure 3A**), PTBP1 binding to *AGO1* (**Figure 3B**), RBFOX2 to *NDEL1* (**Figure 3C**), and the binding of several known nuclear RBPs to *XIST* (**Figure 3D**). To explore this data on a global scale, we compared RNA binding sites for each RBP and observed significant overlap between SPIDR- and ENCODE-derived binding sites for the vast majority of proteins (29/33, $p<0.01$, **Figure 3E**). Moreover, we observed that in virtually all cases each RBP preferentially binds to the same RNA features (e.g., introns, exons, CDS, miRNAs, 5' and 3'UTRs) in both datasets (**Figure 3F**, **Supplemental Figure 7**). Finally, the binding motifs identified within the significant SPIDR-defined binding sites match those defined by CLIP and *in vitro* binding assays[29] (e.g., RNA Bind-N-Seq, **Figure 3G**).

(v) SPIDR enables high-resolution RBP mapping at single nucleotide resolution. Next, we explored whether SPIDR can provide single nucleotide resolution maps of precise RBP-RNA binding sites, as is the case for some current CLIP-seq approaches. Specifically, UV crosslinking creates a covalent adduct at the site of RBP-RNA crosslinking, which leads to a preferential drop-off of the reverse transcriptase at these sites (**Figure 4A**). To explore this, we computed the number of reads that end at each position of an RNA (truncations) and compared these counts to those expected by chance. We observed strong positional enrichments at known protein binding sites. For example, we observe strong enrichment for RPS2 and RPS6 – two distinct structural components of the small ribosomal RNA subunit – at the precise locations where these proteins are known to contact the 18S rRNA in the

resolved ribosome structure (**Figure 4B**). Moreover, examining individual mRNAs bound by HNRNPC (**Figure 4C**) or PTBP1 (**Figure 4D**) showed that the precise binding site corresponds to the known motif sequence. When we computed this enrichment more globally, we observed that HNRNPC (**Figure 4E**) and PTBP1 (**Figure 4F**) reads tend to terminate immediately proximal to these well-known binding sequences[29].

Taken together, our data demonstrate that SPIDR generates highly accurate single nucleotide RBP-binding maps for dozens of RBPs within a single experiment. Moreover, SPIDR can simultaneously map RBPs representing diverse functions and binding modalities, including RBPs that bind within thousands of RNAs (e.g. CPSF6), RBPs that bind only a few very specific RNAs (e.g. SLBP), as well as RBPs that bind primarily within intronic regions within the nucleus (e.g. PTBP1) and RBPs that bind primarily to exonic regions within the cytoplasm (e.g. UPF1).

3.3.3 LARP1 binds to the 40S ribosome and mRNAs encoding translation-associated proteins

In addition to the three known structural components of the small ribosomal subunit (RPS2, RPS3, and RPS6), we noticed that LARP1 also showed strong binding to the 18S ribosomal RNA (**Figures 2A, 5A**). LARP1 is an RNA binding protein that has been linked to translational initiation of specific mRNAs. It is known to bind to the 5' end of specific mRNAs, primarily those encoding critical translation proteins such as ribosomal proteins and initiation and elongation factors, via recognition of a terminal oligopyrimidine (TOP) sequence in the 5' UTR of these transcripts[60]. The exact role of LARP1 in translation has been debated because it has been reported to both promote and repress translation of mRNAs containing a TOP-motif[60–66].

Although LARP1 is known to bind TOP-motif containing mRNAs, how it might promote translation initiation of these mRNAs is mostly unknown. Because we identified a strong

binding interaction between LARP1 and the 18S ribosomal RNA, we explored where in the initiating ribosome this interaction occurs. Interestingly, the LARP1 binding site on the 18S ribosomal RNA (1698-1702 nts) is at a distinct location relative to all other 18S binding proteins that we explored and corresponds to a position within the 48S structure that is directly adjacent to the mRNA entry channel (**Figure 5B**). More generally, we observed strong binding of LARP1 at the TOP-motif sequence within the 5' UTR of translation-associated mRNAs (**Figure 5C**).

These results suggest that LARP1 may act to promote increased translational initiation of TOP-motif containing mRNAs by directly binding to the 43S pre-initiation complex and recruiting this complex specifically to mRNAs containing a TOP-motif. Because LARP1 is positioned immediately adjacent to the mRNA in this structure, this 43S+LARP1 complex would be ideally positioned to access and bind the TOP motif to facilitate efficient ribosome assembly and translational initiation at these mRNAs. This mechanism of direct ribosome recruitment to TOP-motif containing mRNAs through LARP1 binding to the 43S ribosome and the mRNA would explain why the TOP-motif must be contained within a fixed distance from the 5' cap to promote translational initiation[67] (**Figure 5D**).

3.3.4 4EBP1 binds specifically to LARP1-bound mRNAs upon mTOR inhibition

Translation of TOP motif-containing mRNAs is selectively repressed upon inhibition of the mTOR kinase, which occurs in conditions of physiological stress[68–71]. Recent studies have shown that under these conditions, LARP1 binds the 5'-UTR of TOP-containing mRNAs, and it has been postulated that this binding activity is responsible for the specific translational repression of these mRNAs[60,72]. Yet, the mechanism by which LARP1 binding might repress translation remains unknown.

The canonical model for how mTOR inhibition leads to translational suppression is through the selective phosphorylation of 4EBP1[69,71]. Specifically, when phosphorylated, 4EBP1

cannot bind to EIF4E, which is the critical initiation factor that binds to the 5' mRNA cap and recruits the remaining initiation factors through direct binding with EIF4G[73,74]. When 4EBP1 is not phosphorylated (i.e., in the absence of mTOR), it binds to EIF4E and prevents it from binding to EIF4G and initiating translation. While this differential binding of 4EBP1 to EIF4E upon mTOR modulation is well-established and is central to translational suppression, precisely how it leads to selective modulation of TOP mRNA translation has remained unclear. Specifically, direct competition between 4EBP1 and EIF4G for binding to EIF4E should impact translation of all EIF4E-dependent mRNAs, yet the observed translational downregulation is specific to TOP-containing mRNAs[69,71,75] and this specificity is dependent on LARP1 binding[60].

To explore the mechanism of translational suppression of TOP-containing mRNAs upon mTOR inhibition, we treated HEK293T cells with torin, a drug that inhibits mTOR kinase. We adapted SPIDR to map multiple independent samples within a single split-and-pool barcoding experiment (**Figure 6A, Supplemental Table 2,** see **Methods**) and used this approach to perform SPIDR on >50 distinct RBPs, including LARP1, numerous translational initiation factors, and 4 negative controls in both torin-treated and untreated conditions.

To ensure that mTOR inhibition robustly leads to translational suppression of TOP-containing mRNAs, we quantified global protein levels in torin-treated and untreated cells using quantitative mass spectrometry (see **Methods**) to determine protein level changes globally. Although the level of most proteins does not change upon torin-treatment, we observed a striking reduction of proteins encoded from TOP motif-containing mRNAs. Indeed, this translational suppression was directly proportional to the strength of the TOP-motif contained within the 5'-UTR of each mRNA (**Figure 6B, Supplemental Table 3**).

Next, we explored changes in RBP binding upon mTOR inhibition. We measured the number of RNA reads observed for each protein upon torin treatment relative to control. While the majority of proteins showed no change in the number of RNA reads, the sole exception was 4EBP1, which showed a dramatic increase (>20-fold) in the overall number of RNA reads produced upon mTOR inhibition (**Figure 6C**). Interestingly, this increase corresponded to

increased binding specifically at mRNAs containing a TOP-motif (p-value < 8 x 10^{-10}, Mann-Whitney, **Figure 6D** and **6E**). Notably, this did not simply reflect an increased level of 4EBP1 binding at the same sites, but instead corresponded to the detection of many statistically significant binding sites only upon mTOR inhibition that were not observed in the presence of mTOR activity (control samples). Consistent with these observations, a previous study observed that 4EBP1 can be in proximity to translationally suppressed mRNAs upon mTOR inhibition[76].

In contrast to 4EBP1, which showed a dramatic transition in binding activity to mRNA upon mTOR inhibition, we did not observe a global change in the number of RNA reads purified by LARP1 upon mTOR inhibition (**Figure 6C**). Indeed, in both torin-treated and untreated samples we observed strong binding of LARP1 to TOP motif mRNAs as well as to the 18S ribosomal RNA suggesting that this interaction with the 40S ribosome and TOP mRNAs occurs independently of mTOR activity. However, we did observe a 1.7-fold increase in levels of binding of LARP1 at TOP mRNAs upon mTOR inhibition (p-value < 5.4 x 10^{-16}, Mann-Whitney, **Figure 6D** and **6F**). This increased enrichment at TOP mRNAs could reflect more LARP1 binding at these specific mRNAs or could reflect the fact that the LARP1 complex might be more stably associated with each mRNA due to translational repression.

Together, our results suggest a model that may reconcile the apparently divergent perspectives about the role of LARP1 as both an activator and repressor of translational initiation and explains how selective mTOR-dependent translational repression is achieved (**Figure 6G**). Specifically, LARP1 binds to the 40S ribosome and 5' untranslated region of mRNAs containing a TOP motif regardless of mTOR activity. In the presence of mTOR (**Fig 6G, right side**), this dual binding modality can act to promote ribosome recruitment specifically to TOP-containing mRNAs and promote translation of these mRNAs. In the absence of mTOR (**Fig 6G, left side**), 4EBP1 can bind to TOP-containing mRNAs, potentially via the LARP1 protein already bound to these mRNAs. Indeed, most of the significant 4EBP1 binding sites are also bound by LARP1 under Torin treatment (60% overlap, odds-ratio of 12-fold, hypergeometric p-value < 10^{-100}). By binding selectively to

these TOP-containing mRNAs, 4EBP1 can bind to EIF4E and prevent binding between EIF4E and EIF4G, a necessary requirement for initiation of translation. In this way, LARP1/4EBP1 binding to specific mRNAs would enable sequence-specific repression of mRNA translation. This model would explain the apparently divergent roles of LARP1 as both an activator and repressor of translation as it indicates that LARP1 may act as a selective recruitment platform that can either activate or repress translation through the distinct factors that co-bind in the presence or absence of mTOR activity.

3.4 DISCUSSION

Here we present SPIDR, a massively multiplexed method to generate high-quality, high-resolution, transcriptome-wide maps of RBP-RNA interactions. SPIDR can map RBPs with a wide-range of RNA binding characteristics and functions (e.g., mRNAs, lncRNAs, rRNAs, small RNAs, etc.) and will enable the study of diverse RNA processes (e.g., splicing, translation, miRNA processing, etc.) within a single experiment and at an unprecedented scale.

While we show that SPIDR can accurately map dozens of RBPs within a single experiment, the numbers used mostly reflect the availability of high-quality antibodies. As such, we expect that this approach can readily be applied to even larger pool sizes for hundreds or thousands of proteins simultaneously. As such, we expect that this approach can readily be applied to even larger pool sizes for hundreds or thousands of proteins simultaneously. Because of this, we expect that SPIDR will represent a critical technology for exploring the many thousands of human proteins that have been reported as putative RNA binding proteins but that remain largely uncharacterized[6–10]. Similarly, we expect that this technology will be crucial for assessing the putative functions of the >20,000 annotated ncRNAs which have remained largely uncharacterized.

Because the number of cells required to perform SPIDR is comparable to that of a traditional CLIP experiment, yet a single SPIDR experiment reports on the binding behavior of dozens (and likely hundreds) of RBPs, this approach dramatically reduces the number of cells required to map an individual RBP. Accordingly, SPIDR will be a valuable tool for studying RBP-RNA interactions in many different contexts, including within rare cell types and patient samples where large numbers of cells may be difficult to obtain.

We showed that SPIDR generates single nucleotide contact maps that accurately recapitulate the RNA-protein contacts observed within structural models. This suggests that SPIDR will also be well-suited to add high-resolution binding information for entire RNP complexes in a single experiment, as it will allow simultaneous targeting of all proteins within a complex. We envision that, in conjunction with more traditional structural biology methods, this approach will help elucidate the precise structure of various RNP complexes, including for mapping proteins that are not currently resolved within these structures (e.g. LARP1 binding within the 48S ribosome).

In addition to accurately measuring multiple proteins simultaneously, because of the nature of the split-and-pool barcoding strategy used, this approach also allows for multiple samples to be pooled within a single experiment. This ability to simultaneously map multiple proteins across different samples and conditions will enable exploration of RBP binding patterns and their changes across diverse biological processes and disease states. Until now, systematic comparative studies of RBP-RNA interaction changes at scale have been impossible, even for large consortia (e.g., ENCODE), which have invested massive amounts of time and effort to generate CLIP-seq data for only two cell lines. Our 4EBP1 results highlight the critical value of SPIDR for enabling exploration of RBP dynamics across samples. Specifically, 4EBP1 was not commonly thought to directly bind to mRNA, nonetheless, including 4EBP1 within our larger pool of target proteins allowed us to uncover changes across two different experimental conditions that may explain how specificity of mTOR-mediated translational suppression is achieved.

Although we focused on the differential RNA binding properties of 4EBP1/LARP1, there are many additional insights into RBP biology that we expect can be uncovered from exploration of this dataset. For example, we observe that TARDBP (TDP43) shows strong binding to U6 snRNA and to multiple scaRNAs, a class of ncRNAs that play critical roles in spliceosome-associated snRNA biogenesis[77]. TDP43 is an RBP of great interest because of its well-known genetic link to various neurodegenerative disorders, such as amyotrophic lateral sclerosis (ALS)[78–81]. These observations could provide new mechanistic insights into how disruption of this RBP impacts splicing changes and pathogenesis in neurodegeneration.

Thus, we expect that SPIDR will enable a fundamental shift for studying mechanisms of transcriptional and post-transcriptional regulation. Rather than depending on large consortium efforts to generate reference maps within selected cell-types, SPIDR enables any standard molecular biology lab to rapidly generate a comprehensive and high-resolution genome-wide map within any cell-type or experimental system of interest without the need for specialized training or equipment.

3.5 SUPPLEMENTAL NOTES

Note 1: Comparison to a previous multiplexed CLIP method

A recent study reported a variant of CLIP called Antibody-Bead eCLIP (ABC) that utilizes direct chemical conjugation of an oligo sequence to an antibody followed by proximity ligation between the antibody-oligo and RNA to enable multiplexed mapping of 10 proteins simultaneously[27]. Our approach differs from this strategy in several key practical and conceptual ways.

***Antibody labeling*:** The ABC method utilizes direct chemical modification of each antibody. First, this requires large excess of each antibody and selective purification of each conjugate to generate each labeled reagent. This necessitates a more elaborate, multi-step chemical modification and purification procedure for each antibody and therefore is not readily

accessible for labeling large numbers of distinct antibodies. Second, because ABC utilizes chemical modification of the antibody using NHS chemistry, the precise site of oligo conjugation on each antibody is random. This could impact both epitope recognition (when modified within the recognition site) and protein-G binding to the FC region of the antibody; both will decrease the efficiency of IP. In contrast, SPIDR utilizes labeling of the protein G bead instead of direct modification of the antibody. As such, SPIDR is a rapid, efficient, and highly modular strategy already utilized in standard IP strategies to couple antibodies to beads. Because of this distinction, the SPIDR approach can work with the same amounts of antibody used in standard approaches and with antibodies produced and stored in any buffer condition, without the need for specific purification or chemical modification.

***Proximity-ligation versus split-pool detection*:** The ABC method utilizes proximity-ligation to link an antibody sequence to its RNA target. There are several conceptual limitations to this strategy. First, the efficiency of proximity-ligation is limited as ligation must occur between each oligo and RNA end at 1:1 stoichiometry, resulting in many failed ligation events. This will decrease the efficiency of the overall RNA detection rate, an issue that will primarily impact RNAs of low abundance or in low cell numbers. Second, proximity-ligation methods are highly sensitive to distance constraints between the two ligating components. Accordingly, the success of this approach will depend on the distance between the RBP and the RNA and where on the RBP the specific antibody binds. Therefore, there are likely to be antibodies for which this approach will not produce comparable results to standard CLIP. Moreover, this approach is highly sensitive to the RNA fragment size generated. If fragments are too long, this will be problematic because there might be multiple proteins that can ligate; if the fragments are too short, the ends might not be capable of ligation. In contrast, the SPIDR method utilizes split-pool barcoding which is not dependent on the distances between the antibody, RBP, and RNA and therefore not susceptible to these distance constraints. Because of this, SPIDR could be used for analyzing RBP-RNA interactions within higher-order assemblies and in the presence of additional crosslinkers beyond UV. Finally, because SPIDR utilizes barcodes on the beads and because split-pool barcoding is not limited to

pairwise contacts, a single barcode can provide information on the identity of multiple RNAs simultaneously thereby increasing the resolution detected per sequenced read.

For these reasons, SPIDR enables a greater level of multiplexing of proteins, is more broadly applicable to wide-range of antibodies, and is readily accessible to any molecular biology lab.

Note 2: Features and Limitations of the Method

First, similar to CLIP and other immunoprecipitation methods, SPIDR is constrained by the availability of antibodies that have been validated to specifically enrich for RBPs of interest. We note that SPIDR may offer the opportunity to partly alleviate this problem as its multiplexing capability allows for the inclusion of several distinct antibodies, including those that may not have been previously validated, against an RBP of interest without increasing the experimental burden.

Second, the SPIDR protocol requires that each experiment is performed under the same IP conditions for all RBPs. Although we show that standard conditions work for many diverse proteins, they may not be suitable for all RBPs. One possible solution is to match antibodies (and target RBPs) by similar IP conditions.

Finally, in the current protocol, we used the same antibody amount for each RBP of interest, which may in part explain the uneven coverage of RNA reads measured for each RBP. Although we do identify well-known binding sites for nearly all RBPs targeted, higher sequencing coverage might be needed for RBPs at the lower end of this distribution. An alternative solution would be to adapt the antibody amount to equalize coverage for each RBP after performing an initial pre-screen and low-depth sequencing run.

3.6 MAIN FIGURES

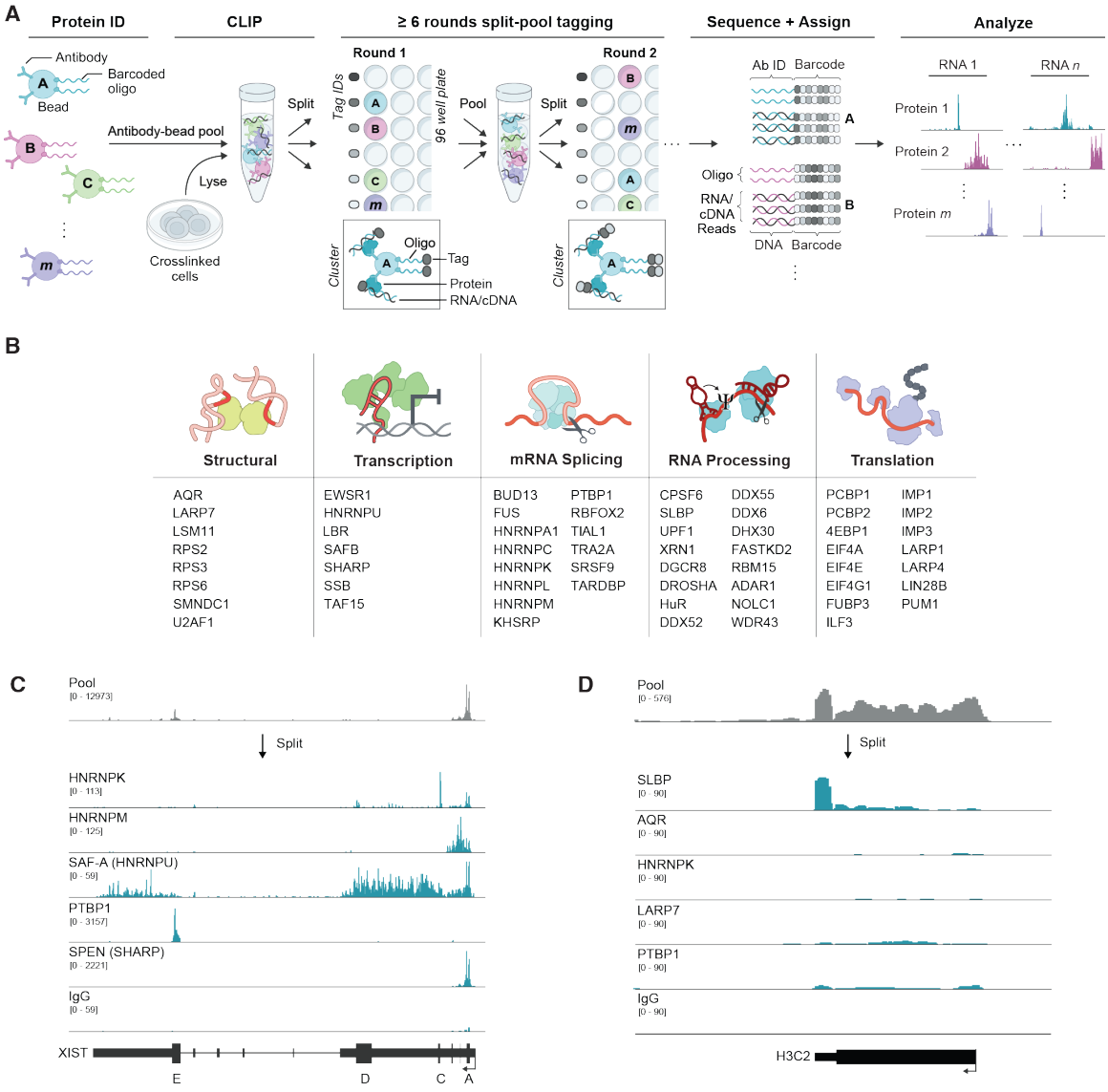


Figure 1: SPIDR (Split and Pool Identification of RBP targets) – a highly multiplexed method to map protein-RNA interactions.

(A) Schematic overview of the SPIDR method. The bead pool is incubated with UV crosslinked lysate in a single tube. After immunopurification, each bead is uniquely labeled by split-and-pool barcoding. The complexity of the barcode generated depends on the number of individual tags used in each split-pool round and the number of split-pool rounds.

For example, after 8 rounds of split and pool barcoding, using 12 barcodes in each round, the likelihood that two beads will end up with same barcode is ~ 1 in 430 million ($1/12^8$). Oligos and RNA molecules and their linked barcodes are sequenced and RNAs are matched to proteins based on their shared barcodes. (The bead labeling strategy was adapted from ChIP-DIP, a Guttman lab protocol used for multiplexed mapping of hundreds of proteins the DNA, https://github.com/GuttmanLab/chipdip-pipeline). **(B)** Schematic list of the different RBPs mapped by SPIDR in K562 and/or HEK293T cells, functional assignments based on literature review. **(C)** An example of the raw alignment data for the pool (all reads before splitting by bead identities) and for specific RBPs (all reads assigned to specific RBP beads) across the *XIST* RNA. Blocks represent exons, lines introns, and thick blocks are the annotated *XIST* repeat regions (A-E). **(D)** Raw alignment data for SLBP across the *H3C2* histone mRNA. Top track is pooled alignment data; tracks below are reads assigned to SLBP or other RBPs and controls.

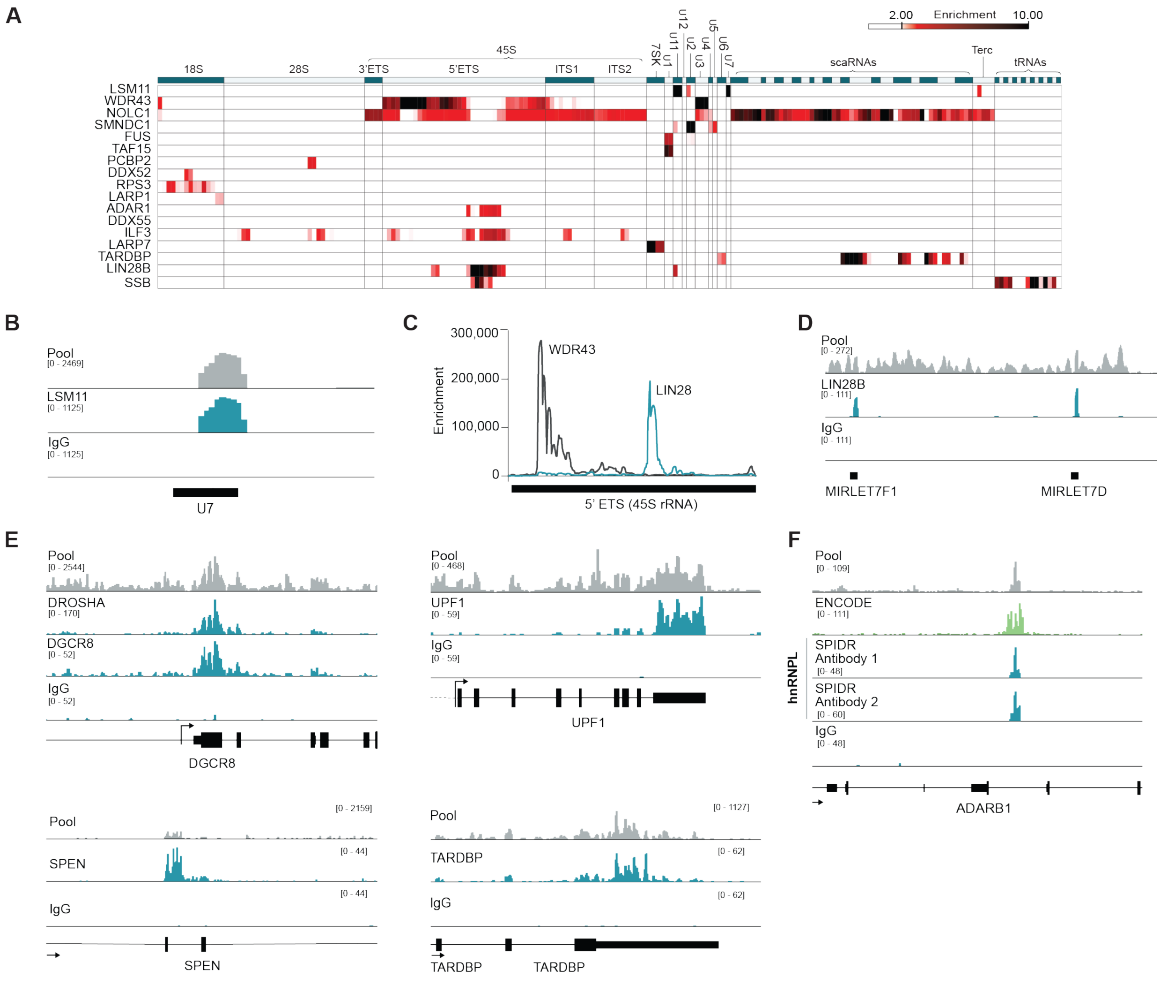


Figure 2: SPIDR accurately maps binding of a diverse set of RBPs.

(A) RNA binding patterns of selected RBPs (rows) relative to 100nt windows across each classical non-coding RNA (columns). Each bin is colored based on the enrichment of read coverage per RBP relative to background. **(B)** Sequence read coverage for LSM11 binding to U7 snRNA. For all tracks, "pool" refers to all reads prior to splitting them by paired barcodes (shown in gray), and individual tracks (shown in teal) reflect reads after assignment to specific antibodies. **(C)** Enrichment of read coverage relative to background for WDR43 and LIN28B over the 5' ETS region of 45S RNA. **(D)** Sequence reads coverage for LIN28B binding to let-7 miRNAs. **(E)** Sequence reads coverage for DROSHA/DGCR8, UPF1, SPEN, and TARDBP to their respective mRNAs. **(F)** Sequence reads coverage for two

distinct antibodies to HNRNPL in a single SPIDR experiment. For comparison, HNRNPL coverage from the ENCODE-generated eCLIP data is shown (bright green).

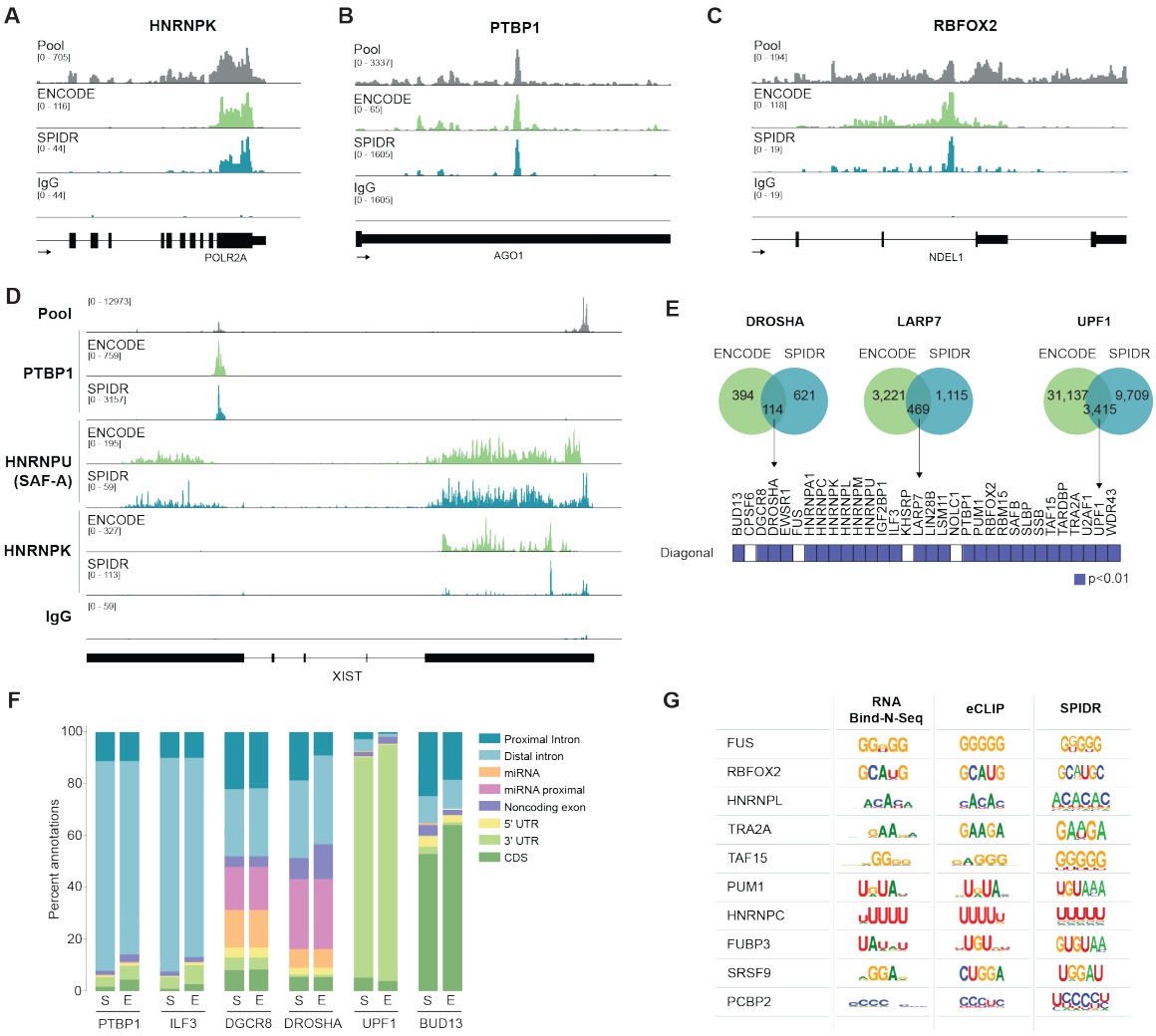


Figure 3: SPIDR data is highly comparable to previous eCLIP datasets.

Examples of concordant binding identified by eCLIP (ENCODE consortium) and SPIDR. Sequence reads coverage is shown for individual proteins measured by ENCODE (green) and SPIDR (teal) along with a negative control (IgG). **(A)** HNRNPK, **(B)** PTBP1, and **(C)** RBFOX2. **(D)** Comparison of ENCODE and SPIDR data for multiple proteins bound to the *XIST* lncRNA. Sequence reads coverage for PTBP1, HNRNPU (SAF-A), and HNRNPK are shown. **(E)** Significance of overlap between binding sites detected by SPIDR and those identified within paired proteins in the ENCODE data. Each bin represents the paired protein between both experiments; blue represents a hypergeometric p-value of less than 0.01. **(F)** Peak annotation in matched SPIDR and ENCODE data. Stacked bar plot showing the

percentage of peaks detected in the SPIDR (S) or ENCODE (E) datasets in various annotation categories. **(G)** Comparison of significant motifs identified within SPIDR peaks (right, p-value threshold < $1e^{-40}$) to those reported for RNA Bind-n-Seq (left) or eCLIP (middle)[29].

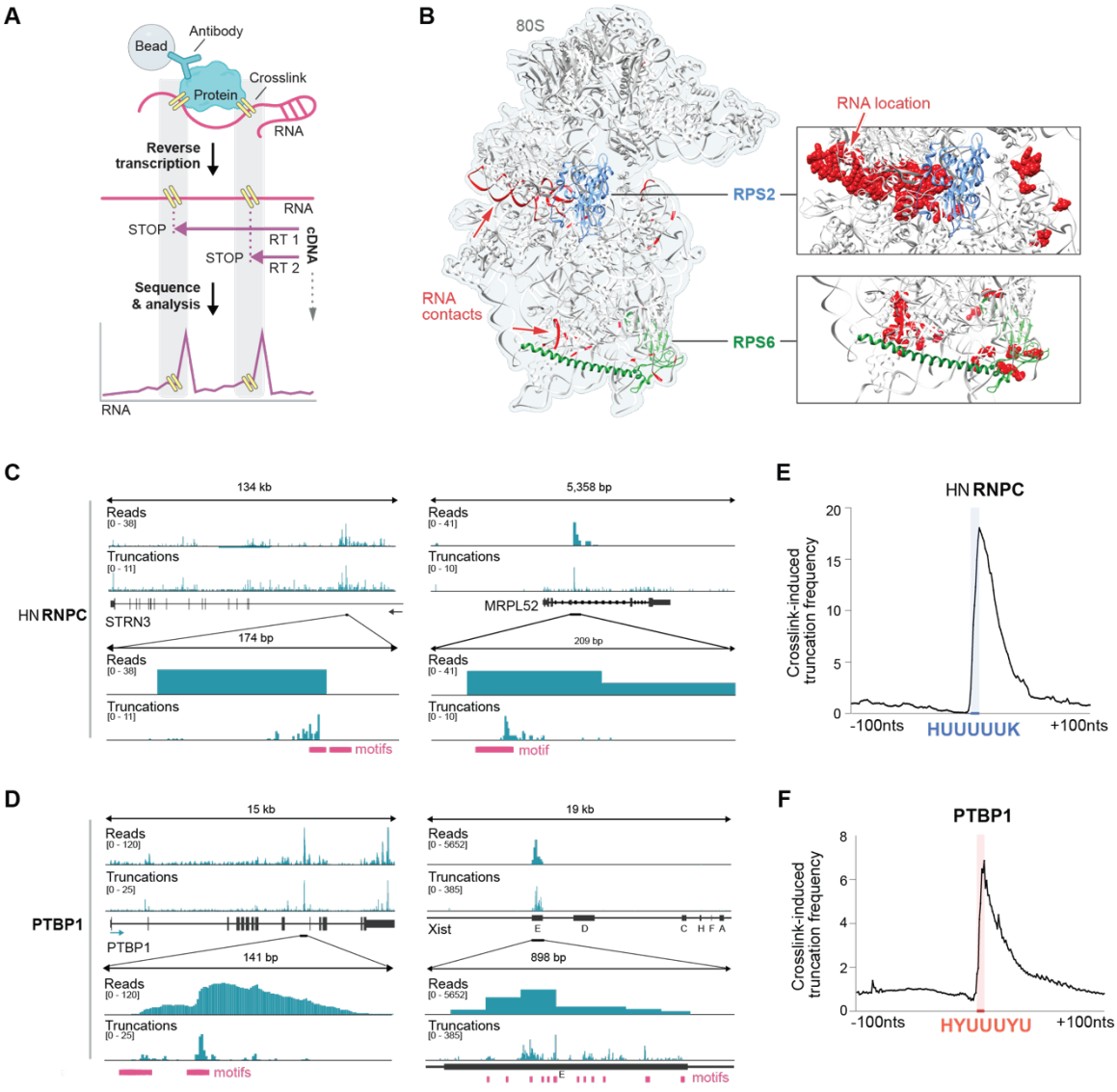


Figure 4: SPIDR enables high-resolution RBP mapping at single nucleotide resolution.

(A) Schematic showing how reverse transcription pause sites can be used to map RBP-RNA interactions at single nucleotide resolution. UV light crosslinks the RBP to the target RNA at points of direct contact. During reverse transcription, the enzyme preferentially stalls at the crosslinking site, leading to termination of cDNA synthesis (STOP). Mapping the 3'-end of the cDNA (truncations) may identify the RBP binding site at single nucleotide resolution. **(B)** The SPIDR determined binding sites of RPS2 and RPS6 overlayed on the known 80S ribosome structure. RPS2 protein is shown in light blue and RPS6 protein in green; direct

RNA contacts of each of these two ribosomal proteins detected by SPIDR are shown in red. SPIDR data shown is from HEK293T cells. **(C)** HNRNPC binding sites for *STRN3* (left) and *MRPL52* (right). Both raw read alignments ("Reads", top) and 3'-end truncations of the cDNA ("Truncations", bottom) are shown. The upper two panels show the mapped reads and truncations for the whole gene, the lower two panels are zoomed-in on the indicated region. The known binding motifs for HNRNPC are depicted in magenta. **(D)** Examples of PTBP1 binding sites for *PTBP1* itself (left) and *XIST* (right). Both raw read alignments and 3'-end truncations of the cDNA reads are shown. Known binding motifs for PTBP1 are in magenta. **(E)** Truncation frequency (3' ends of the mapped cDNA reads) over all significantly enriched HNRNPC peaks is shown centered on the motif position. The region of the steep frequency rise of truncations is shown by the blue line on the y-axis and corresponds to the sequence shown in blue. **(F)** The truncation frequency (3' ends of the mapped cDNA reads) over all significantly enriched PTBP1 peaks is shown relative to the motif position within each peak. The region of the steep frequency rise of truncations is shown by the orange line on the y-axis and corresponds to the sequence shown in orange.

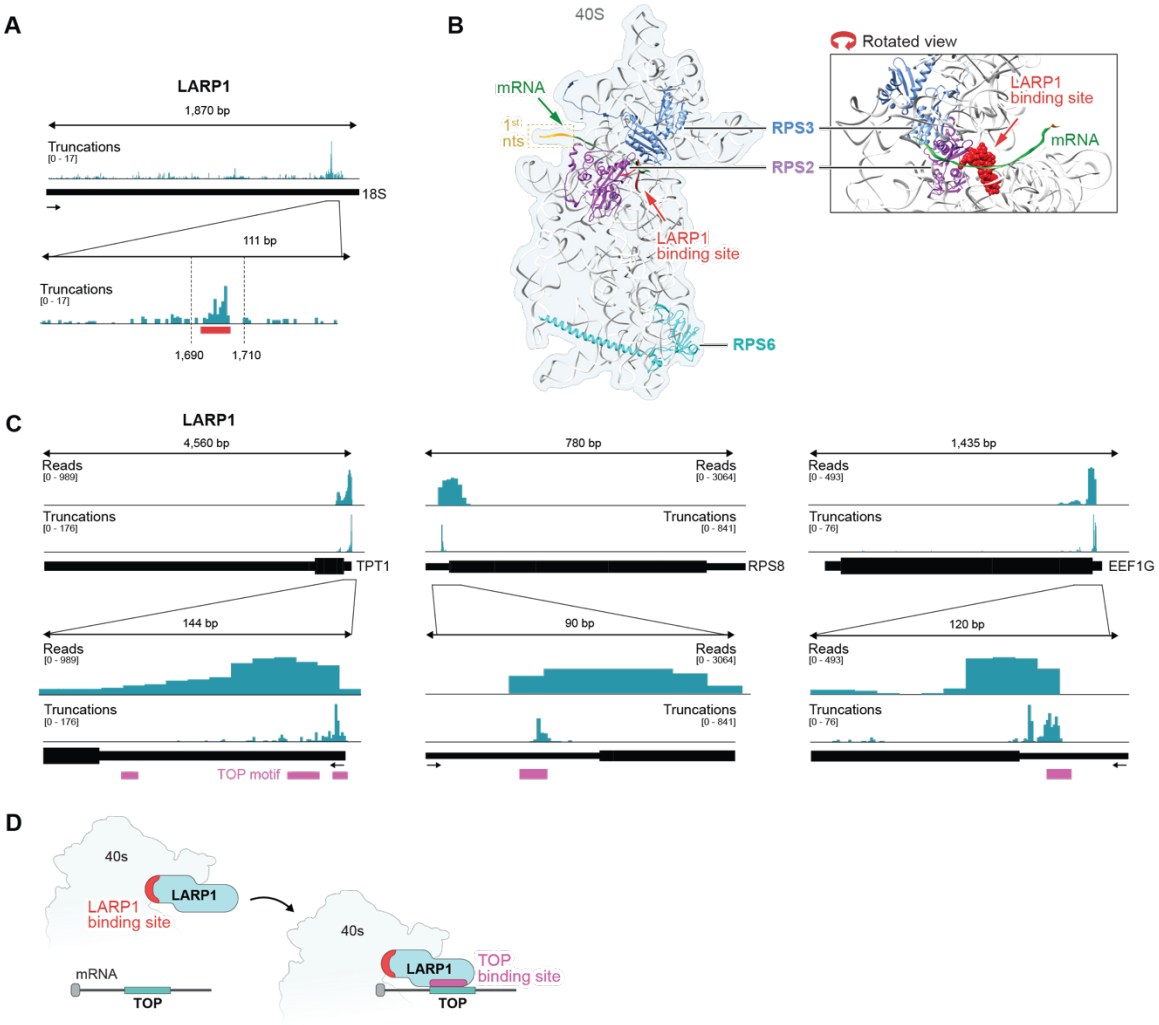


Figure 5: LARP1 binds to 18S rRNA near the mRNA entry channel and at TOP-motifs contained within the 5'UTRs of mRNAs.

(A) Frequency of 3'-end truncations of LARP1 reads plotted across the 18S rRNA. Zoom-in shows accumulation near nucleotide position 1700 (indicated by the red bar). Data shown is from HEK293T cells. **(B)** The structure of the 40S ribosomal subunit bound to the 5'-end of an mRNA molecule (green). The first nucleotide of the mRNA is indicated in orange. RPS2 (purple) and RPS3 (blue) are indicated for orientation and the mRNA is shown in green. The LARP1 binding site detected on the 18S rRNA is indicated in red. **(C)** Examples of LARP1 binding for three different mRNAs containing TOP motifs in their 5'UTRs: *TPT1* (left), *RPS8* (middle), and *EEF1G* (right). Both read alignments ("Reads", top) and 3'end

truncations of the cDNA reads (“Truncations”, bottom) are shown. TOP motifs within the 5’-UTRs are depicted in magenta. **(D)** Model of LARP1 interactions based on SPIDR data. LARP1 shows preferential binding to both the 18S rRNA (close to the mRNA entry channel of the 40S subunit) and the TOP motifs within the 5’UTR of specific mRNAs. In this way, LARP1 could facilitate recruitment of the 40S subunit to TOP motif-containing mRNAs.

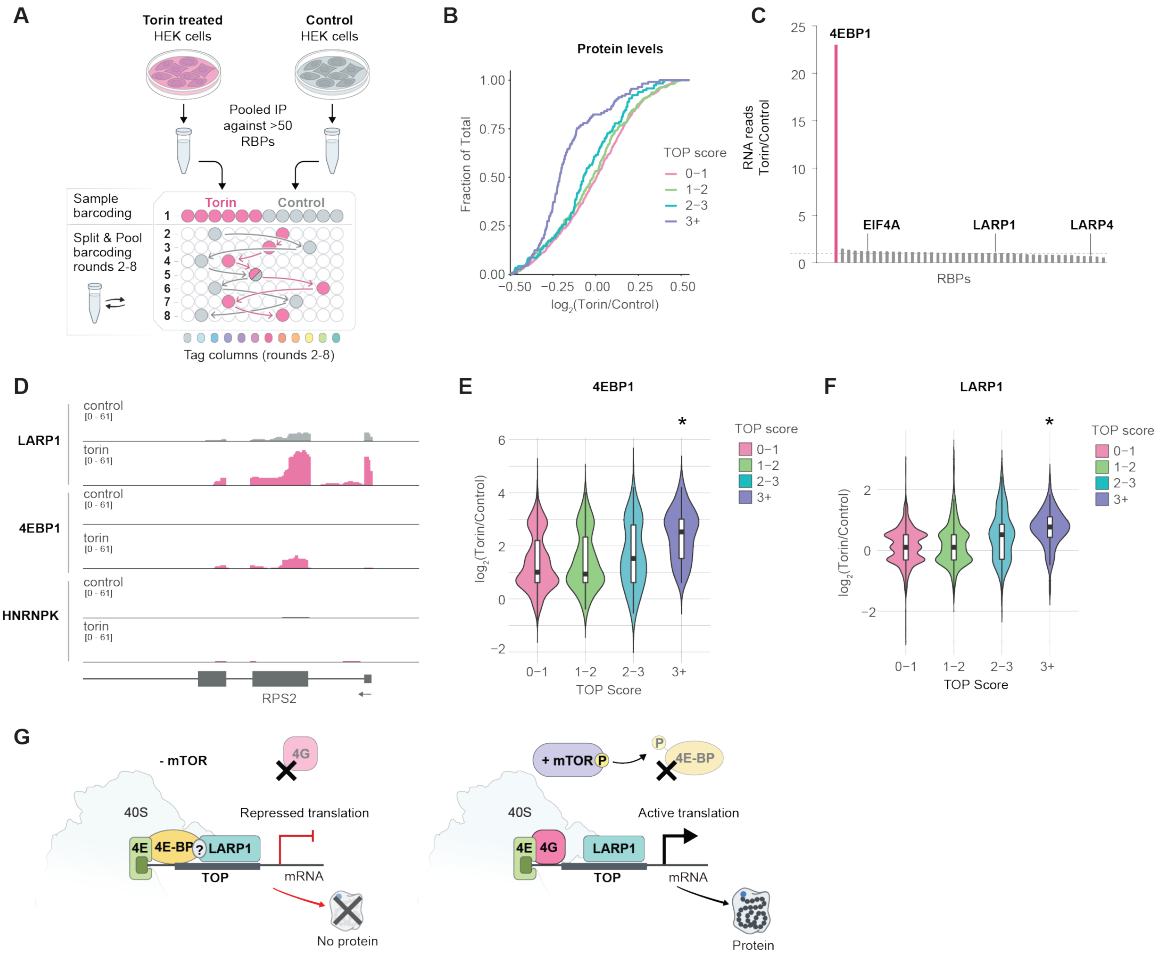
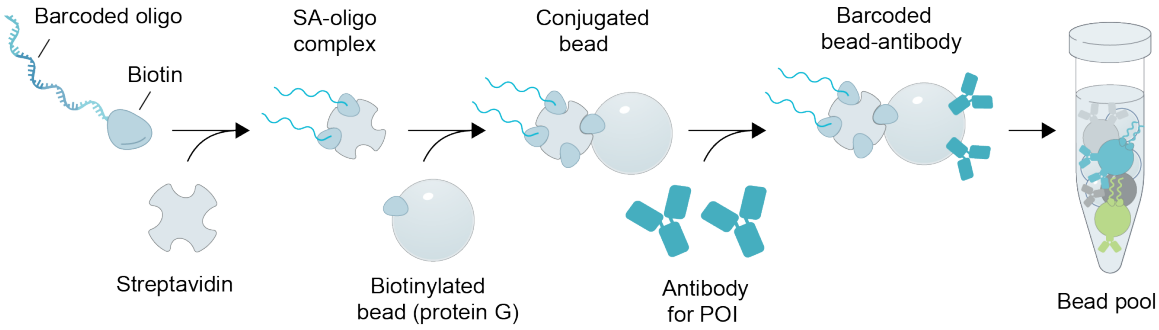


Figure 6: 4EBP1 binds specifically to LARP1-bound mRNAs upon mTOR inhibition.

(A) Schematic of experimental approach for the mTOR perturbation experiment. HEK293T cells were treated with either 250nM torin or control (solvent only) for 18 hours. SPIDR was performed on both samples. The multiplexed IP was performed separately, and the samples were mixed after the first round of barcoding (see **Methods** for details). **(B)** Cumulative Distribution Function (CDF) plots of protein changes in torin versus control treated samples as determined by LC-MS/MS. $\log_2$ ratios (Torin/Control) are shown on the x-axis and fraction of total (from 0 to 1) is shown on the y-axis. Proteins were grouped into four categories based on their TOP motif score as previously published[60]. The analysis was performed on the 2000 most highly expressed genes (based on RNA expression, see **Methods**). **(C)** Number of SPIDR reads assigned to each RBP in the torin-treated samples

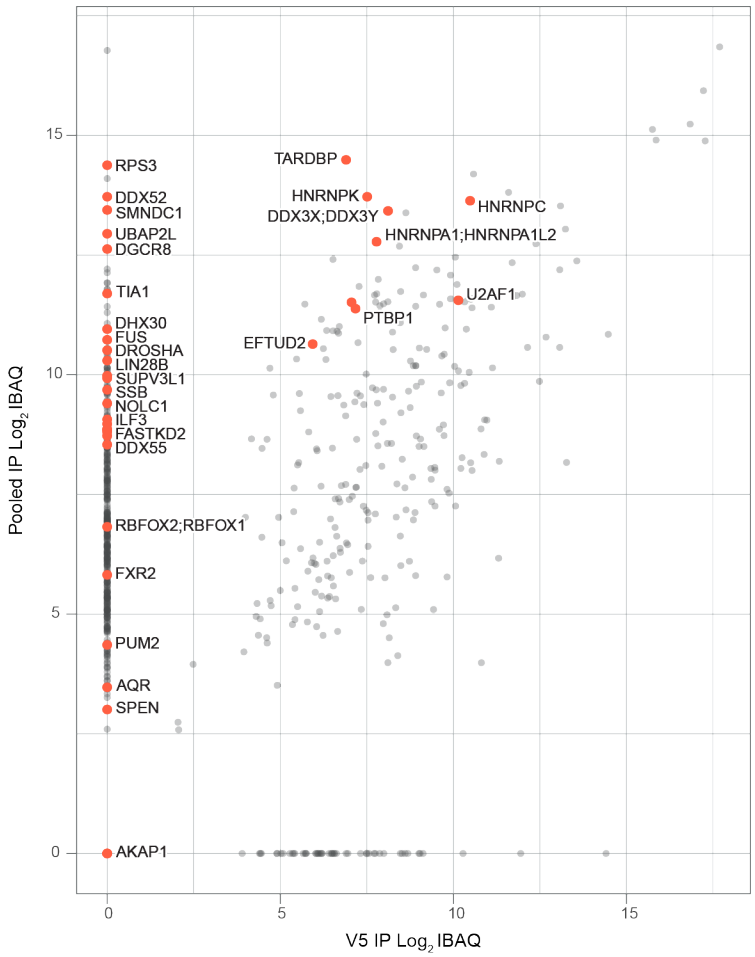
versus control samples. 4EBP1 (pink line), EIF4A, LARP1, and LARP4 are also indicated. Dashed line corresponds to enrichment of 1. **(D)** Raw alignment data for selected RBPs across *RPS2*, an mRNA with a strong TOP motif. For each protein "control" and "torin" treatment tracks are shown. **(E)** Violin plots of the $\log_2$ ratios (torin/control) of significant binding sites for 4EBP1 are shown. The RNA targets are grouped based on their TOP motif score as published in (Philippe et al., 2020)[60]. **(F)** Violin plots of the $\log_2$ ratios (Torin/Control) of significant binding sites for LARP1 are shown. The RNA targets were grouped based on their TOP motif score as published in (Philippe et al., 2020)[60]. For (E) and (F) the asterisks indicate statistical significance (p-value < 0.00001, Mann-Whitney). **(G)** Model of mTOR-dependent repression of mRNA translation. LARP1 binds to the 40S ribosome and to 5' untranslated region of TOP-containing mRNAs independent of mTOR activity. When mTOR is active (i.e., in the absence of torin; right side), this dual binding modality can recruit the ribosome specifically to TOP-containing mRNAs and promote their translation. When mTOR is inactive (i.e., in the presence of torin), 4EBP1 can bind to TOP-containing mRNAs (possibly through an interaction with LARP1) and to EIF4E. The interaction between 4EBP1 and EIF4E prevents binding between EIF4E and EIF4G, which is required to initiate translation. In this way, LARP1/4EBP1 binding specifically to TOP-containing mRNAs would enable sequence-specific repression of translation.

3.7 SUPPLEMENTAL FIGURES



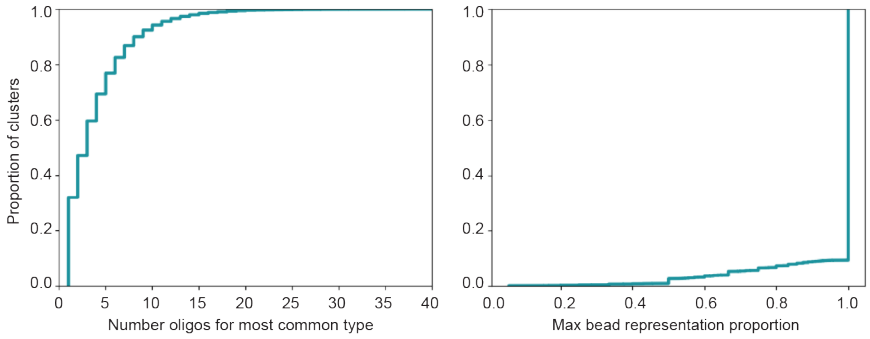
Supplemental Figure 1: Schematic of our multiplexed antibody-bead labeling strategy.

Populations of biotinylated protein G beads are incubated with a streptavidin-biotin oligo complex. Each population of beads is labeled with an oligo with a specific sequence and then incubated with one type of capture antibody such that each population has a unique capture antibody and a corresponding oligo tag that can be recognized after sequencing. Populations are combined to create the bead pool.



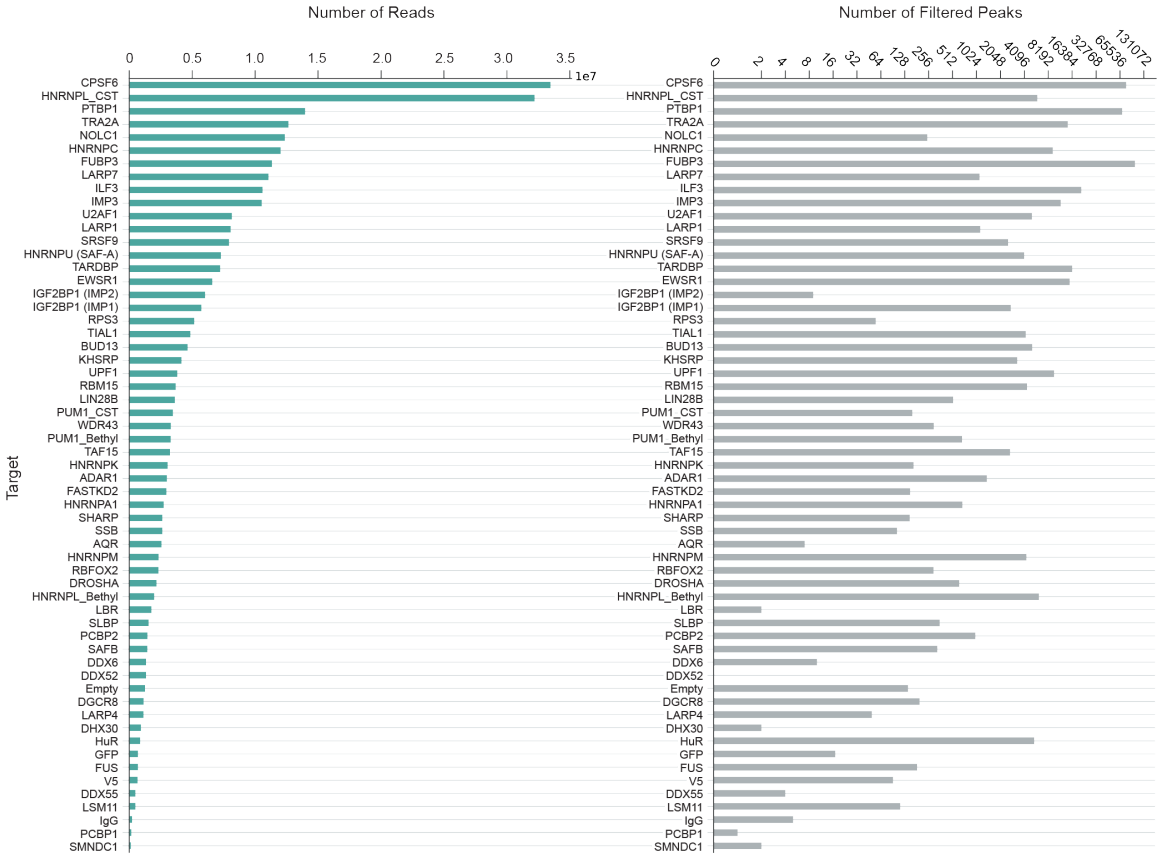
Supplemental Figure 2: Multiplexed IP of dozens of RBPs accurately recovers targeted proteins.

Scatter plot showing $\log_2$ transformed IBAQ (intensity based absolute quantification)[82] values for all identified proteins in either the pooled IP with 39 targets (y-axis) versus those detected with a V5 negative control IP (x-axis) by LC-MS/MS. Target proteins that should be detected by the antibodies included in the pool of 39 used are marked in red.



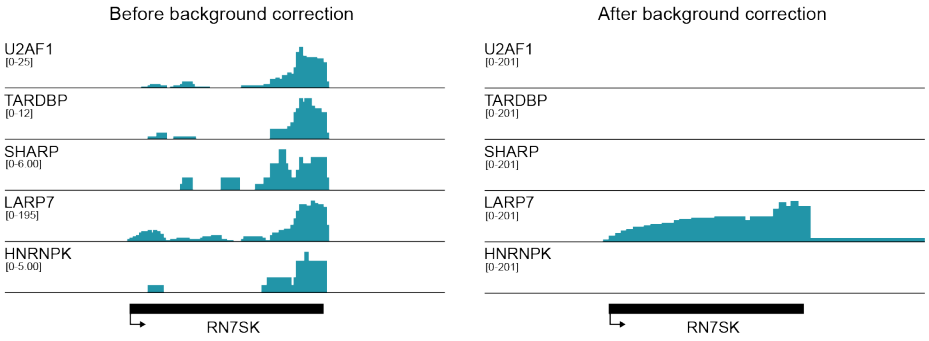
Supplemental Figure 3: Uniqueness of beads and number of oligos per bead of the experiment.

Observed distributions of labeled beads after sequencing. Each bead is defined in sequencing by a particular, unique combinatorial barcode acquired during split-pool. A SPIDR cluster represents any set of molecules, oligo or RNA, that share the same bead combinatorial barcode. Left: CDF plot showing the number of independent oligos matched within an individual SPIDR cluster. Right: CDF plot describing the degree of heterogeneity of these detected oligos within each SPIDR cluster, as determined by oligos with a shared combinatorial barcode. X-axis represents the homogeneity of the oligo types with 1 indicating that all oligos are of the same type.



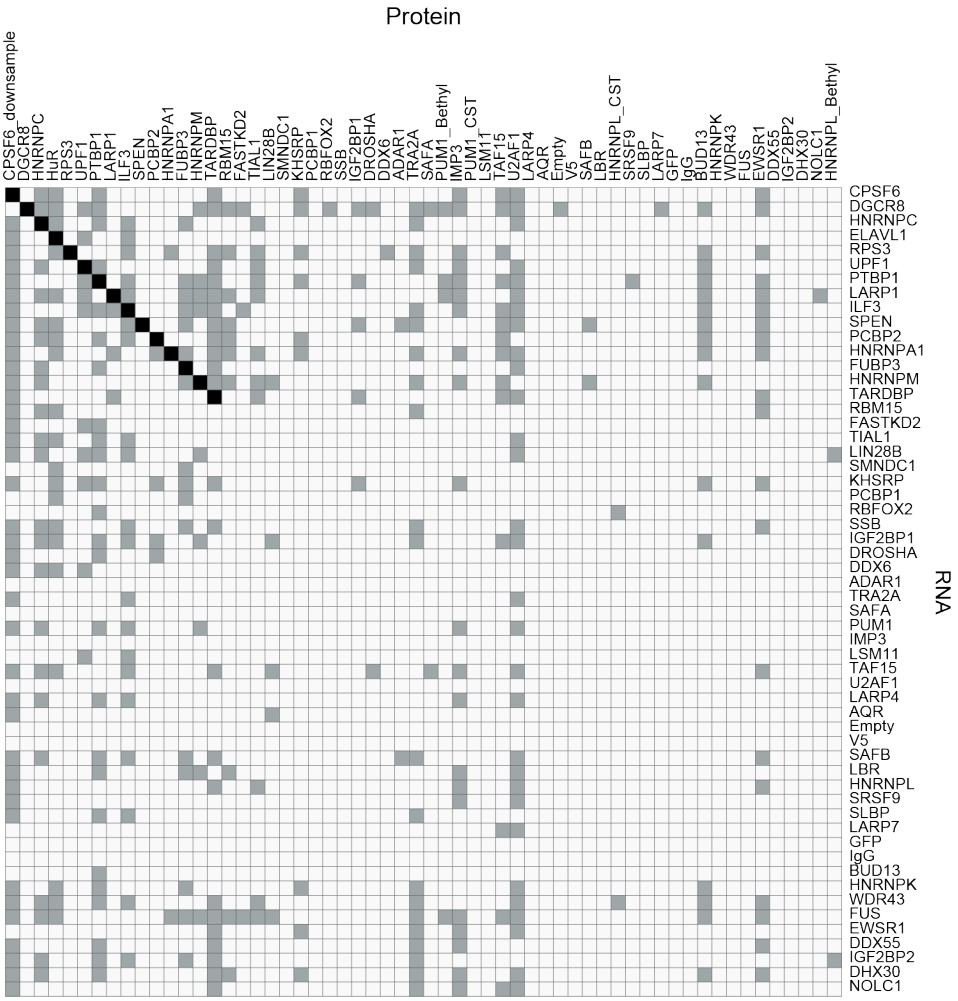
Supplemental Figure 4: Mapped unique reads per RBP and significant binding sites identified per RBP.

Number of deduplicated mapped reads and number of significant binding sites within uniquely mapped genomic regions per IP. The order is determined by the number of unique mapped reads in both plots.



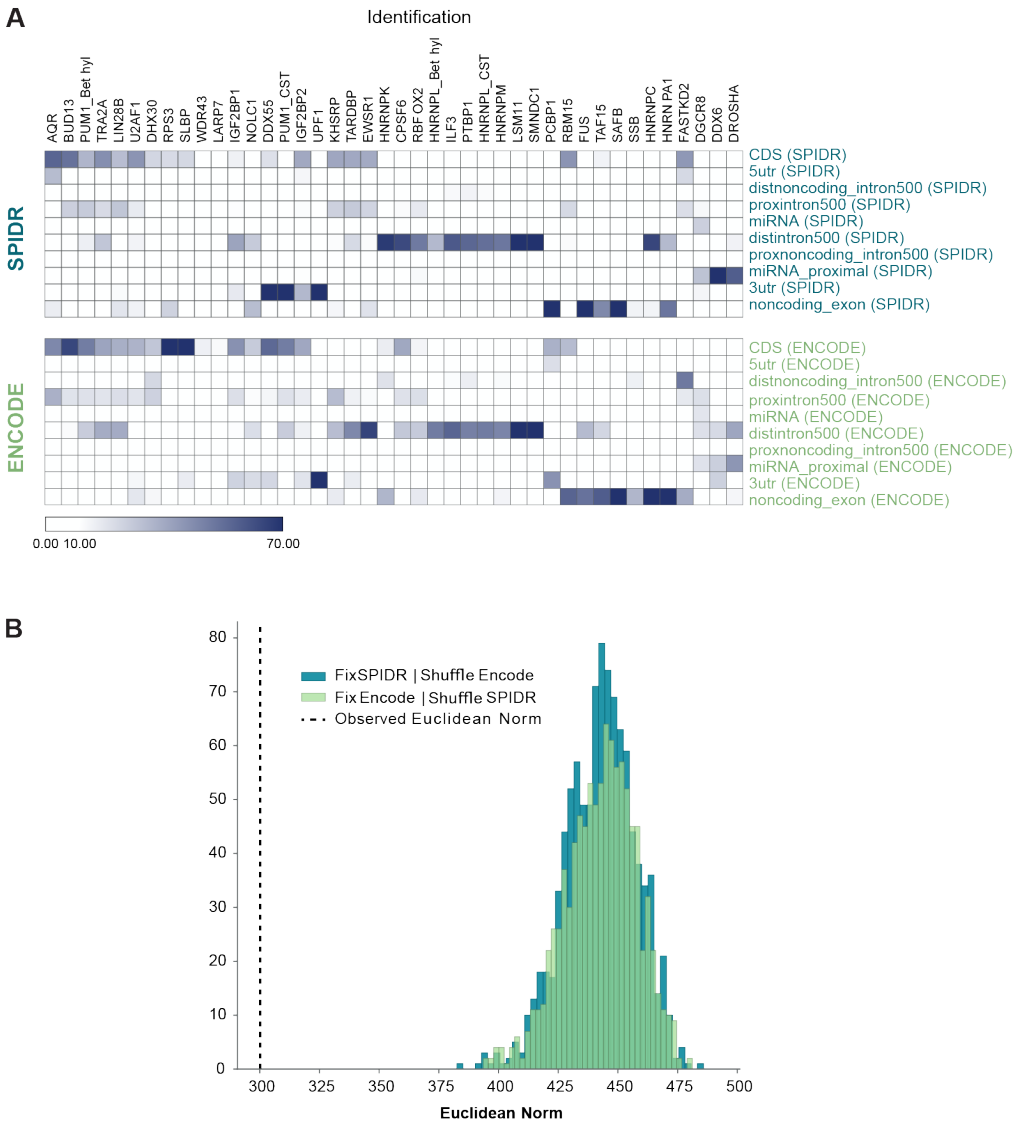
Supplemental Figure 5: Background correction.

An example of our background correction method that utilizes the total read coverage across all proteins to normalize each individual protein. Shown are example tracks on *RN7SK* before and after background correction. Left: Raw alignment data for the entire pooled dataset (top track) and for representative antibodies against U2AF1, TARDBP, SHARP, LARP7, and HNRNPK on *RN7SK*. Right: Background corrected data for the same set of antibodies. Signal that was not antibody-specific has been normalized out. The reads in the right are binned in 5 nucleotide windows. *RN7SK* is known to be bound by LARP7[51].



Supplemental Figure 6: Autoregulatory binding between RBPs targeted by SPIDR and their RNAs.

Auto-regulatory binding matrix with protein (x-axis) binding to each mRNA (y-axis) shown. Each target protein included in SPIDR performed in K562 cells marked by whether it has significantly enriched binding within its own RNA, or in any of the other SPIDR target RNAs. Proteins that bind their own RNA are marked in black, and instances of binding to genes of other SPIDR targets are marked in gray.



Supplemental Figure 7: Global comparison of annotations (intron, exon, etc.) of binding sites per RBP as called by ENCODE versus SPIDR.

(A) Heatmaps showing the percentage of significant binding sites in each of the annotation categories for SPIDR performed in K562 cells and ENCODE (see **Methods** for details). **(B)** Quantitative assessment of the similarity of heatmaps between SPIDR and ENCODE. The Euclidean distance (L2 norm) between the ENCODE and SPIDR percentage tables/heatmaps was calculated. The calculated distance is indicated by the dashed line. The statistical significance was calculated by randomly shuffling the columns of either the SPIDR

percentage table and keeping the original ENCODE table or vice versa, meaning shuffling the columns of the ENCODE table and keeping the original SPIDR table. This was done 1000 times in each direction and every time the Euclidean distance was calculated. The values are represented by the two histograms. The Euclidean distance of all of the randomly shuffled 2000 comparison was always larger than of the true pair, which shows that the two original annotation tables from SPIDR and ENCODE are highly significantly similar (p-value < 0.0005).

3.8 METHODS

3.8.1 EXPERIMENTAL CONDITIONS

Experimental conditions

Cell culture

K562 cells (ATCC, CCL-243) and HEK293T cells (ATCC, CRL-3216) were purchased from ATCC and cultured under standard conditions. K562 cells were cultured in K562 media consisting of 1X DMEM (Gibco), 1 mM Sodium Pyruvate (Gibco), 2 mM L-Glutamine (Gibco), 1X FBS (Seradigm), 100 U/mL Penicillin-Streptomycin (Life Technologies). 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), and 1X FBS (Seradigm).

UV-crosslinking

Crosslinking was performed as previously described[23]. Briefly, K562 cells were washed once with 1X PBS and diluted to a density of ~10 million cells/mL in 1X PBS for plating onto culture dishes. HEK293T cells were washed once with 1X PBS and crosslinked directly on culture dishes. RNA-protein interactions were crosslinked on ice using 0.25 J cm^{-2} (UV 2.5k) of UV at 254 nm in a Spectrolinker UV Crosslinker. Cells were then scraped from culture dishes, washed once with 1X PBS, pelleted by centrifugation at 330 x g for 3 minutes, and flash-frozen in liquid nitrogen for storage at -80°C.

Torin-1 treatment

HEK293T cells were treated at a final concentration of 250 nM Torin-1 (Cell Signaling Technology, #14379) in standard HEK293T media for 18 hours prior to UV-crosslinking and harvesting.

Bead biotinylation

The bead labeling strategy was adapted from ChIP-DIP, a Guttman lab protocol used for multiplexed mapping of hundreds of proteins the DNA (https://github.com/GuttmanLab/chipdip-pipeline). Briefly, 1 mL of Protein G Dynabeads (ThermoFisher, #10003D) were washed once with 1X PBST (1X PBS + 0.1% Tween-20) and resuspended in 1mL PBST. Beads were then incubated with 20 µL of 5 mM EZ-Link Sulfo-NHS-Biotin (Thermo, #21217) on a HulaMixer for 30 minutes at room temperature. Following NHS reaction, beads were placed on a magnet and 500 µL of buffer was removed and replaced with 500 µL of 1M Tris pH 7.4 to quench the reaction for an additional 30 minutes at room temperature. Beads were then washed twice with 1 mL PBST and resuspended in their original storage buffer until use.

Labeling biotinylated beads with oligonucleotide tags

Unique biotinylated oligonucleotides were first coupled to streptavidin (BioLegend, #280302) in a 96-well PCR plate. In each well, 20 µL of 10 µM oligo was added to 75 µL 1X PBS and 5 µL 1 mg/mL streptavidin. The 96-well plate was then incubated with shaking at 1600 rpm on a ThermoMixer for 30 minutes at room temperature. Each well was then diluted 1:4 in 1X PBS for a final concentration of 227 nM.

For each experiment, the appropriate amount of biotinylated Protein G beads (10 µL beads per capture antibody) was washed once in 1X PBST. Beads were then resuspended in oligo binding buffer (0.5X PBST, 5 mM Tris pH 8.0, 0.5 mM EDTA, 1M NaCl). 200 µL of the bead suspension was aliquoted into individual wells of a 96-well plate, followed by addition of 4 µL of 227nM streptavidin-coupled oligo to each well. The 96-well plate was then incubated with shaking at 1200 rpm on a ThermoMixer for 30 minutes at room temperature. Beads were then washed twice with M2 buffer (20 mM Tris 7.5, 50 mM NaCl, 0.2% Triton X-100, 0.2% Na-Deoxycholate, 0.2% NP-40), twice with 1X PBST, and resuspended in 200 µL of 1X PBST.

Binding antibody to labeled Protein G beads

2.5 µg of each capture antibody was added to each well of the 96-well plate containing labeled beads in 1X PBST. The plate was incubated with shaking at 1200 rpm on a ThermoMixer for 30 minutes at room temperature. After incubation, beads were washed twice with 1X PBST + 2 mM biotin (Sigma, #B4639-5G), resuspended in 200 µL of 1x PBST + 2mM biotin, and left shaking at 1200 rpm for 10 minutes at room temperature. All wells containing beads were then pooled together and washed twice with 1 mL 1X PBST + 2 mM biotin. At this stage, each bead in the bead pool contains a single type of capture antibody with a corresponding unique oligonucleotide tag.

Pooled immunoprecipitation

For each experiment, 10 million cells were lysed in 1 mL RIPA buffer (50mM HEPES pH 7.4, 100mM NaCl, 1% NP-40, 0.5% Na-Deoxycholate, 0.1% SDS) supplemented with 20 µL Protease Inhibitor Cocktail (Sigma, #P8340-5mL), 10 µL of Turbo DNase (Invitrogen, #AM2238), 1X Manganese/Calcium mix (2.5 mM $MnCl_2$, 0.5 mM $CaCl_2$), and 5 µL of RiboLock RNase Inhibitor (Thermo Fisher, #EO0382). Samples were incubated on ice for 10 minutes to allow lysis to proceed. After lysis, cells were sonicated at 3-4 W of power for 3 minutes (pulses 0.7 s on, 3.3 s off) using the Branson sonicator and then incubated at 37°C for 10 minutes to allow for DNase digestion. DNase reaction was quenched with addition of 0.25 M EDTA/EGTA mix for a final concentration of 10 mM EDTA/EGTA. RNase If (NEB, #M0243L) was then added at a 1:500 dilution and samples were incubated at 37°C for 10 minutes to allow partial fragmentation of RNA to obtain RNAs of approximately ~300-400 bp in length. RNase reaction was quenched with addition of 500 µL ice cold RIPA buffer supplemented with 20 µL Protease Inhibitor Cocktail and 5 µL of RiboLock RNase Inhibitor, followed by incubation on ice for 3 minutes. Lysates were then cleared by centrifugation at 15000 x g at 4°C for 2 minutes. The supernatant was transferred to new tubes and diluted in additional RIPA buffer such that the final volume corresponded to 1 mL lysate for every 100 µL of Protein G beads used. Lysate was then combined with the labeled antibody-bead pool and 1 M biotin was added to a final concentration of 10 mM as to quench any disassociated streptavidin-coupled oligos. Beads were left rotating overnight at 4°C on a HulaMixer.

Following immunoprecipitation, beads were washed twice with RIPA buffer, twice with high salt wash buffer (50 mM HEPES pH 7.4, 1 M NaCl, 1% NP-40, 0.5% Na-Deoxycholate, 0.1% SDS), and twice with Tween buffer (50 mM HEPES pH 7.4, 0.1% Tween-20).

Ligation of the RNA Phosphate Modified ("RPM") tag

After immunoprecipitation, 3' ends of RNA were modified to have 3' OH groups compatible for ligation using T4 Polynucleotide Kinase (NEB, #M0201L). Beads were incubated at 37°C for 10 minutes with shaking at 1200 rpm on a ThermoMixer. Following end repair, beads were buffer exchanged by washing twice with high salt wash buffer and twice with Tween buffer. RNA is subsequently ligated with an "RNA Phosphate Modified" (RPM) adaptor (Quinodoz et al 2021) using High ConcentrationT4 RNA Ligase I (NEB, M0437M). Beads were incubated at 24°C for 1 hour 15 minutes with shaking at 1400 rpm, followed by three washes in Tween buffer. After RPM ligation, RNA was converted to cDNA using SuperScript III (Invitrogen, #18080093) at 42°C for 20 minutes using the "RPM Bottom" RT primer to facilitate on-bead library construction and a 5' sticky end to ligate tags during split-and-pool barcoding. Excess primer is digested with Exonuclease I (NEB, #M0293L) at 37°C for 15 minutes.

Split-and-pool barcoding to identify RNA-protein interactions

Split-and-pool barcoding was performed as previously described[31] with minor modifications. Specifically, beads were split-and-pool ligated over ≥ 6 rounds with a set of "Odd," "Even," and "Terminal" tags. The number of barcoding rounds performed for each SPIDR experiment was determined based on the complexity of the given bead pool. All split-and-pool ligation steps were performed for 5 minutes at room temperature and supplemented with 2 mM biotin and 1:40 RiboLock RNase Inhibitor to prevent RNA degradation. We ensured that virtually all barcode clusters (>95%) represented molecules belonging to unique, individual beads. See Supplemental Table 4 for a template to calculate the number of barcoding rounds required to resolve a given number of starting beads.

Compared to previously published approaches, we reduced the number of barcodes per round, but increased the rounds of split and pool barcoding as we optimized the ligation step. Therefore, the barcoding procedure was significantly simplified in contrast to previous versions. For example, for the K562 cells pooled experiment, 6 rounds of 24 barcodes were used for combinatorial barcoding (with a scheme of Odd, Even, Odd, Even, Odd, Terminal tag). For the HEK293T cells mTOR inhibition experiment, 6 rounds of 36 barcodes were used for combinatorial barcoding to achieve sufficient barcode complexity. Of the 36 barcodes used in round one of the ligations, 18 were used to label the control condition and the remaining 18 were used to label the torin treated condition. The samples were then pooled together for the remaining 5 rounds of ligation.

Library preparation

After split-and-pool barcoding, beads were aliquoted into 5% aliquots for library preparation and sequencing. RNA in each aliquot was degraded by incubating with RNase H (NEB, #M0297L) and RNase cocktail (Invitrogen, #AM2286) at 37°C for 20 minutes. 3' ends of the resulting cDNA were ligated to attach dsDNA oligos containing library amplification sequences using a "splint" ligation as previously described (Quinodoz et al 2021)[31]. The "splint" ligation reaction was performed with 1X Instant Sticky End Master Mix (NEB #M0370) at 24°C for 1 hour with shaking at 1400 rpm on a ThermoMixer. Barcoded cDNA and biotinylated oligo tags were then eluted from beads by boiling in NLS elution buffer (20 mM Tris-HCl pH 7.5, 10 mM EDTA, 2% N-lauroylsarcosine, 2.5 mM TCEP) for 6 minutes at 91°C, with shaking at 1350 rpm.

Biotinylated oligo tags were first captured by diluting the eluant in 1X oligo binding buffer (0.5X PBST, 5 mM Tris pH 8.0, 0.5 mM EDTA, 1M NaCl) and subsequently binding to MyOne Streptavidin C1 Dynabeads (Invitrogen, #65001) at room temperature for 30 minutes. Beads were placed on a magnet and the supernatant, containing cDNA, was moved to a separate tube. Biotinylated oligo tags were amplified on-bead using 2X Q5 Hot-Start Mastermix (NEB #M0494) with primers that add the indexed full Illumina adaptor sequences.

To isolate barcoded cDNA, the supernatant was first incubated with a biotinylated antisense ssDNA ("anti-RPM") probe that hybridizes to the junction between the reverse transcription primer and splint sequences to reduce empty insertion products. This mixture was then bound to MyOne Streptavidin C1 Dynabeads at room temperature for 30 minutes. Beads were placed on a magnet and the supernatant, containing the remaining cDNA products, was cleaned up on Silane beads (Invitrogen, #37002D) as previously described[83]. Finally, cDNA was amplified using 2X Q5 Hot-Start Mastermix (NEB #M0494) with primers that add the indexed full Illumina adaptor sequences.

After amplification, libraries were cleaned up using 1X SPRI (AMPure XP), size-selected on a 2% agarose gel, and cut at either ~300 nt (barcoded oligo tag) or between 300-1000 nt (barcoded cDNA). Libraries were subsequently purified with Zymoclean Gel DNA Recovery Kit (Zymo Research, #4007).

Sequencing

Paired-end sequencing was performed on either an Illumina NovaSeq 6000 (S4 flowcell), NextSeq 550, or NextSeq 2000 with read lengths ≥ 100 x 200 nucleotides. For the K562 data, 37 SPIDR aliquots were generated and sequenced from two technical replicate experiments. The two experiments were generated using the same batch of UV-crosslinked lysate processed on the same day. For the HEK293T data, 9 SPIDR aliquots were generated from a single technical replicate. Each SPIDR library corresponds to a distinct aliquot that was separately amplified with different indexed primers, providing an additional round of barcoding as previously described[31]. Minimum required sequencing depth for each experiment was determined by the estimated number of beads and unique molecules in each aliquot. For oligo tag libraries, each library was sequenced to a depth of observing ~5 unique oligo tags per bead on average. For cDNA libraries, each library was sequenced with at least 2x coverage of the total estimated library complexity.

3.8.2 ANALYSIS AND PROCESSING PIPELINE

Read processing and alignment

Paired-end RNA sequencing reads were trimmed to remove adaptor sequences using Trim Galore! v0.6.2 and assessed with FastQC v0.11.8. Subsequently, the RPM (ATCAGCACTTA) sequence was trimmed using Cutadapt v3.4 from both 5' and 3' read ends. The barcodes of trimmed reads were identified with Barcode ID v1.2.0 (https://github.com/GuttmanLab/sprite2.0-pipeline) and the ligation efficiency was assessed. Reads with or without an RPM sequence were split into two separate files to process RNA and oligo tag reads individually downstream, respectively.

RNA 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 the human (hg38) genome using STAR aligner. Only reads that mapped uniquely to the genome were kept for further analysis.

Barcode matching and filtering

Mapped RNA and oligo tag reads were merged, and a cluster file was generated for all downstream analysis as previously described. MultiQC v1.6 was used to aggregate all reports. To unambiguously exclude ligation events that could not have occurred sequentially, we utilized unique sets of barcodes for each round of split-and-pool. All clusters containing barcode strings that were out-of-order or contained identical repeats of barcodes were filtered from the merged cluster file. To determine the amount of unique oligo tags present in each cluster, sequences sharing the same Unique Molecular Identifier (UMI) were removed and the remaining occurrences were counted. To remove PCR duplication events within the RNA library, sequences sharing identical start and stop genomic positions were removed.

Splitting alignment files by protein identity

Barcode strings from filtered cluster files were then used to assign protein identities to the alignment file containing all mapped RNA reads. Because each cluster represents an individual bead, the frequency of oligo tags (each representing unique protein type) was used to determine protein assignments. Specifically, for each cluster we required ≥ 3 observed oligo tags and that the most common protein type represented ≥ 80% of all observed tags. RNA reads were then split into separate alignment files by barcode strings corresponding to protein type.

Background correction and peak calling

In order to determine what portion of the observed signal is specific to a particular capture antibody, rather than common pileups regardless of the protein captured, we normalized coverage for each protein relative to the coverage detected for all other proteins. Specifically, for a protein of interest, we computed the number of reads that were mapped to that protein. We then randomly downsampled all reads not assigned to that protein such that it had a comparable number of reads as the protein of interest. To measure the expected variance in the control sample, we repeated this downsampling procedure at least 100 independent times. We then computed read counts per window across the transcriptome (either 10nts or 100nts) for the protein of interest and each of the randomized control samples. We computed a normalized enrichment as the number of observed reads within the window (observed) divided by the average of the read counts overt that window across the >=100 permutations (expected). To assess the significance of this enrichment score, we measured how often the observed score was seen in the >=100 permutations. A p-value was assigned as the number of random scores greater than or equal to the observed scores divided by the number of random permutations used (we included the actual observed score in the numerator and denominator). All windows that had at least 10 observed reads and a p-value less than 0.05 were considered significantly enriched.

Peak annotation

Enriched windows were first filtered to only include regions resulting from reads that could be uniquely mapped in the second STAR alignment, and then poor alignments to rRNA regions (chr21: 88206400-8449330) were removed. These filtered peaks were then annotated based on overlap with GENCODE v41 transcripts. In the case of overlapping annotations, the final assigned annotation was chosen based on the following priority list: miRNA, CDS, 5'UTR, 3'UTR, proximal intron (within 500 nt of the splice site region), distal intron (further than 500 nt of the splice site region), non-coding exon, and finally non-coding intron. Windows for which the primary gene annotation was a miRNA host gene were marked as miRNA proximal.

SPIDR comparison to ENCODE

ENCODE datasets

43 of the proteins included in SPIDR also had a matched K562 ENCODE eCLIP experiment with paired-end sequencing data. The raw FASTQ files for these datasets were downloaded from the ENCODE website (https://www.encodeproject.org/) and aligned to the genome using the same parameters as in the SPIDR dataset.

For comparison of matched SPIDR and ENCODE datasets, the larger of the pair of alignment files was downsampled to the depth of the smaller alignment file. Windows of enrichment in ENCODE datasets were then determined using the same background correction strategy and thresholding as in SPIDR (minimum read count of 10, p-value < 0.05). As was done in the SPIDR data, all ENCODE datasets were used as negative controls for one for one another when determining background correction factors and calling windows of enrichment.

Motif enrichment analysis

Filtered SPIDR peaks were used to subset the corresponding SPIDR alignment files, such that only reads that fell within enriched windows were kept. These reads were then used as

input for *de novo* motif analysis by HOMER (http://homer.ucsd.edu/homer/). Motifs with a reported p-value < 10^{-40} were considered significant.

Comparison of bound RNA features

Enriched windows for both SPIDR and ENCODE, as determined using the SPIDR workflow of background correction and thresholding, were annotated based on overlap with GENCODE v41 transcripts. Peaks annotated as intergenic were removed, and then both the SPIDR and ENCODE datasets were filtered to include only proteins that had greater than 100 peaks.

The likelihood of seeing a similarity between SPIDR and ENCODE in the region annotations is visualized by comparing the observed values to randomly shuffled values. The inputs for this method are two matrices, one for SPIDR and one for ENCODE, with the percentage of annotations observed for a given region type for a given RBP. Shuffling is performed by randomly switching percentages across RBPs, keeping the relative values between regions constant. This can be thought of as randomly shuffling the columns of one of the input matrices. A distance is calculated by flattening the two input matrices into vectors, taking the difference between the two vectors, and calculating an L2-norm on that difference. In Supplemental Figure 7 the histogram of L2-norms shows the distribution we would expect if RBPs had no effect on the L2-norm between SPIDR and ENCODE. The dashed vertical line represents the L2-norm when the input matrices were flattened but not shuffled.

The basic algorithm is as follows:

1. Calculate the true L2-norm between SPIDR and ENCODE
2. Keeping SPIDR constant, randomly switch probabilities between RBPs while keeping the percentages within an RBP the same for ENCODE
3. Repeat step 2, shuffling SPIDR and keeping ENCODE constant
4. Repeat steps 2 and 3 for 1000 samples

Single nucleotide resolution analysis

We computed the frequency of reads ending at the 3' end of the cDNA. We computed enrichment for each of these counts by randomly downsampling all reads not assigned to the specific protein and computing the same 3' end coverage. Enrichments and p-values were computed as described above and as previously reported in (Banerjee et al. 2020)[84].

mTOR analysis

Background corrected bedgraphs were generated from control and +Torin conditions for each RBP in each condition. These bedgraph values were then mapped on to Refseq genes using the bedtools map command (arguments: -c 4 –o absmax). Where multiple isoforms were present for the same gene, the isoform with the highest map count was used. To normalize for possible detection bias due to fewer antibody beads in one condition versus the other we adjusted the map value by the ratio of antibody beads as determined by number of bead clusters corresponding to each antibody in each respective condition. Number of antibody (bead clusters) were defined and calculated using the same values used to generate the split bam files for each protein (options: minimum number of oligos=3,fraction unique=0.8, max number of RNAs in clusters=100). The ratio of cluster-corrected values for each gene across the two conditions was then compared per gene and separated based on TOP score. Published TOP scores[60] were used to generate categories for violin plots.

For the protein changes CDF plots, we first selected for the 2000 highest expressed genes based on previous RNA-seq data[84]. Input TPM values for HEK293 cells were taken from input CLAP (sub_input.merged.bam) data from HEK293T cell in (Banerjee et al. 2020)[84].The input samples were downsampled to 20M reads prior to TPM calculation. Featurecounts was used to calculate read overlaps with hg38 protein coding refseq genes and further converted to TPM values. The top 2000 expressed genes (based on HEK293 input TPM) were used to plot the average protein log2 fold changes (Torin versus control) vs TOP score. Published TOP scores[60] were used to plot CDF values.

Mass spectrometry

Multiplexed Immunopurification (IP) for mass spectrometry

10 million K562 cells were lysed in 4mL of RIPA on ice for 10 minutes. The lysate was clarified by centrifugation at 15000g for 2 minutes, and then split in half for either the pooled IP with 39 antibodies or the negative control IP with an anti-V5 antibody. Each half of the lysate was combined with 10ug total antibody (0.25ug per each antibody for the pooled IP) and 100uL of Protein G beads and left rotating at 4C overnight. The beads were then washed twice with RIPA, twice with High Salt Wash Buffer, twice with Clap-Tween, and finally three times with Mass Spec IP Wash Buffer (150mM NaCl, 50mM Tris-HCl pH 7.5, 5% Glycerol). Each sample was then reduced, alkylated, Trypsin digested, and desalted as described in (Parnas et al, 2015)[85]. Peptides were reconstituted in 12uL 3% acetonitrile/0.1% formic acid.

mTOR proteomics

5 million cells each of control and 250nM Torin-1 treated HEK cells were lysed in 250uL Mass Spec Lysis Buffer (8M urea, 75mM NaCL, 50mM Tris pH 8.0, 1mM EDTA) for 30min at room temperature. Samples were then clarified by centrifugation at 23000g for 5 minutes, and the protein content in the supernatant was measured by BCA assay (ThermoFisher, #PI23227). 40ug of protein for each sample was reduced with 5mM final dithriothreitol (DTT) for 45 minutes at room temperature and subsequently alkylated with 10mM final iodoacetamide (IAA) for 45 minutes in the dark at room temperature. 50mM Tris (pH 8.0) was then added to each sample such that the final concentration of urea was less than 2M. Samples were digested overnight with 0.4ug Trypsin (Promega, #V5113) for a 1:100 enzyme to protein ratio. Peptides were desalted on C18 StageTips according to (Rappsilber et al., 2007)[86].

LC-MS/MS

LC-MS/MS analysis was performed on a Q-Exactive HF. 5uL of total peptides were analyzed on a Waters M-Class UPLC using a C18 25cm Thermo EASY-Spray column (2um, 100A, 75um x 25cm) or IonOpticks Aurora ultimate column (1.7um, 75um x 25cm) coupled to a benchtop ThermoFisher Scientific Orbitrap Q Exactive HF mass spectrometer. Peptides were separated at a flow rate of 400 nL/min with a linear 95 min gradient from 5% to 22% solvent B (100% acetonitrile, 0.1% formic acid), followed by a linear 30 min gradient from 22 to 90% solvent B. Each sample was run for 160 min, including sample loading and column equilibration times. Data was acquired using Xcalibur 4.1 software.

The IP samples were measured in a Data Dependent Acquisition (DDA) mode. MS1 Spectra were measured with a resolution of 120,000, an AGC target of 3e6 and a mass range from 300 to 1800 m/z. Up to 12 MS2 spectra per duty cycle were triggered at a resolution of 15,000, an AGC target of 1e5, an isolation window of 1.6 m/z and a normalized collision energy of 28.

The Torin treated and control total lysate samples were measured in a Data Independent Acquisition (DIA) mode. MS1 Spectra were measured with a resolution of 120,000, an AGC target of 5e6 and a mass range from 350 to 1650 m/z. 47 isolation windows of 28 m/z were measured at a resolution of 30,000, an AGC target of 3e6, normalized collision energies of 22.5, 25, 27.5, and a fixed first mass of 200 m/z.

Database searching of the proteomics raw files

Proteomics raw files were analyzed using the directDIA method on SpectroNaut v16.0 for DIA runs or SpectroMine (3.2.220222.52329) for DDA runs (Biognosys) using a human UniProt database (Homo sapiens, UP000005640), under BSG factory settings, with automatic cross-run median normalization and imputation. Protein group data were exported for subsequent analysis.

3.9 SUPPLEMENTAL INFORMATION

Supplemental Table 1: Overview of the SPIDR experiment in K562 cells; can be found online[87]. The protein targets are listed, as well as the vendors and product numbers of the corresponding antibodies. The bead oligo used to represent each protein in sequencing is also included, along with the number of deduplicated mapped reads and number of significant binding sites within uniquely mapped genomic regions.

Supplemental Table 2: Overview of the SPIDR experiment HEK293T cells treated with Torin or Control (solvent only); can be found online[87]. The protein targets are listed, as well as the vendors and product numbers of the corresponding antibodies. The bead oligo used to represent each protein in sequencing is also included, along with the number of deduplicated mapped reads and number of significant binding sites within uniquely mapped genomic regions. Read counts and peaks are provided for "Torin" and "Control" treatments separately.

Supplemental Table 3: Protein measurements of HEK 293 HEK293T cells treated with Torin or Control (solvent only); can be found online[87]. The average protein log2 ratios (Torin versus Control) are listed for the 2000 highest expressed genes as based on previous RNA-seq data[84] (see **Methods**). As a gene selected from the RNA-seq data could be present in several protein groups, all protein log2 ratios for the protein groups where this gene was present are listed. Proteins were grouped into four categories based on their TOP motif score as previously published in (Philippe et al., 2020)[60]. The TOP motif score for each gene is also listed.

Supplemental Table 4: Barcode Calculator; can be found online[87]. A template to calculate the number of split-and-pool barcoding rounds required to resolve a given number of starting beads.

3.10 REFERENCES

1. Sonenberg, N. & Hinnebusch, A. G. Regulation of Translation Initiation in Eukaryotes: Mechanisms and Biological Targets. *Cell* **136**, 731–745 (2009).
2. Castello, A., Fischer, B., Hentze, M. W. & Preiss, T. RNA-binding proteins in Mendelian disease. *Trends Genet* **29**, 318–327 (2013).
3. Gao, F.-B. & Taylor, J. P. RNA-binding proteins in neurological disease. *Brain Res* **1462**, 1–2 (2012).
4. Gebauer, F., Schwarzl, T., Valcárcel, J. & Hentze, M. W. RNA-binding proteins in human genetic disease. *Nat Rev Genet* **22**, 185–198 (2021).
5. Gerstberger, S., Hafner, M. & Tuschl, T. A census of human RNA-binding proteins. *Nat Rev Genet* **15**, 829–845 (2014).
6. Caudron-Herger, M., Jansen, R. E., Wassmer, E. & Diederichs, S. RBP2GO: a comprehensive pan-species database on RNA-binding proteins, their interactions and functions. *Nucleic Acids Res* **49**, D425–D436 (2021).
7. Trendel, J. *et al.* The Human RNA-Binding Proteome and Its Dynamics during Translational Arrest. *Cell* **176**, 391-403.e19 (2019).
8. Queiroz, R. M. L. *et al.* Comprehensive identification of RNA–protein interactions in any organism using orthogonal organic phase separation (OOPS). *Nature Biotechnology 2019 37:2* **37**, 169–178 (2019).
9. Urdaneta, E. C. *et al.* Purification of cross-linked RNA-protein complexes by phenol-toluol extraction. *Nat Commun* **10**, (2019).
10. Perez-Perri, J. I. *et al.* Discovery of RNA-binding proteins and characterization of their dynamic responses by enhanced RNA interactome capture. *Nat Commun* **9**, (2018).
11. King, O. D., Gitler, A. D. & Shorter, J. The tip of the iceberg: RNA-binding proteins with prion-like domains in neurodegenerative disease. *Brain Res* **1462**, 61–80 (2012).
12. Guttman, M. *et al.* Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* **458**, 223–227 (2009).

13. Statello, L., Guo, C. J., Chen, L. L. & Huarte, M. Gene regulation by long non-coding RNAs and its biological functions. *Nature Reviews Molecular Cell Biology 2020 22:2* **22**, 96–118 (2020).
14. Guo, J. K. & Guttman, M. Regulatory non-coding RNAs: everything is possible, but what is important? *Nat Methods* **19**, 1156–1159 (2022).
15. Guttman, M. & Rinn, J. L. Modular regulatory principles of large non-coding RNAs. *Nature* **482**, 339–346 (2012).
16. Chu, C. *et al.* Systematic Discovery of Xist RNA Binding Proteins. *Cell* **161**, 404–416 (2015).
17. McHugh, C. A. *et al.* The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature* **521**, 232–236 (2015).
18. Robert-Finestra, T. *et al.* SPEN is required for Xist upregulation during initiation of X chromosome inactivation. *Nature Communications 2021 12:1* **12**, 1–13 (2021).
19. Engreitz, J. M. *et al.* The Xist lncRNA Exploits Three-Dimensional Genome Architecture to Spread Across the X Chromosome. *Science (1979)* (2013) doi:10.1126/science.1237973.
20. Jachowicz, J. W. *et al.* Xist spatially amplifies SHARP/SPEN recruitment to balance chromosome-wide silencing and specificity to the X chromosome. *Nat Struct Mol Biol* **29**, 239–249 (2022).
21. Hafner, M. *et al.* Transcriptome-wide Identification of RNA-Binding Protein and MicroRNA Target Sites by PAR-CLIP. *Cell* **141**, 129–141 (2010).
22. Chi, S. W., Zang, J. B., Mele, A. & Darnell, R. B. Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature* (2009) doi:10.1038/nature08170.
23. Van Nostrand, E. L. *et al.* Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat Methods* **13**, 508–514 (2016).
24. Zhang, C. & Darnell, R. B. Mapping in vivo protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data. *Nat Biotechnol* (2011) doi:10.1038/nbt.1873.
25. Ule, J. *et al.* CLIP identifies Nova-regulated RNA networks in the brain. *Science* **302**, 1212–1215 (2003).

26. Ramanathan, M., Porter, D. F. & Khavari, P. A. Methods to study RNA–protein interactions. *Nat Methods* **16**, 225 (2019).

27. Lorenz, D. A. *et al.* Multiplexed transcriptome discovery of RNA-binding protein binding sites by antibody-barcode eCLIP. *Nat Methods* 1–5 (2022) doi:10.1038/s41592-022-01708-8.

28. van Nostrand, E. L. *et al.* Principles of RNA processing from analysis of enhanced CLIP maps for 150 RNA binding proteins. *Genome Biol* **21**, 90 (2020).

29. Van Nostrand, E. L. *et al.* A large-scale binding and functional map of human RNA-binding proteins. *Nature* **583**, 711–719 (2020).

30. Quinodoz, S. A. *et al.* SPRITE: a genome-wide method for mapping higher-order 3D interactions in the nucleus using combinatorial split-and-pool barcoding. *Nature Protocols 2022 17:1* **17**, 36–75 (2022).

31. Quinodoz, S. A. *et al.* RNA promotes the formation of spatial compartments in the nucleus. *Cell* **184**, 5775-5790.e30 (2021).

32. Quinodoz, S. A. *et al.* Higher-Order Inter-chromosomal Hubs Shape 3D Genome Organization in the Nucleus. *Cell* **174**, 744-757.e24 (2018).

33. Pillai, R. S. *et al.* Unique Sm core structure of U7 snRNPs: assembly by a specialized SMN complex and the role of a new component, Lsm11, in histone RNA processing. *Genes Dev* **17**, 2321–2333 (2003).

34. Singh, R. N., Howell, M. D., Ottesen, E. W. & Singh, N. N. Diverse role of survival motor neuron protein. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **1860**, 299–315 (2017).

35. Bi, X. *et al.* RNA Targets Ribogenesis Factor WDR43 to Chromatin for Transcription and Pluripotency Control. *Mol Cell* **75**, 102-116.e9 (2019).

36. Sun, Z. *et al.* LIN28 coordinately promotes nucleolar/ribosomal functions and represses the 2C-like transcriptional program in pluripotent stem cells. *Protein Cell* **13**, 490–512 (2022).

37. Chen, H.-K., Pai, C.-Y., Huang, J.-Y. & Yeh, N.-H. Human Nopp140, which interacts with RNA polymerase I: implications for rRNA gene transcription and nucleolar structural organization. *Mol Cell Biol* **19**, 8536–8546 (1999).

38. Bizarro, J., Bhardwaj, A., Smith, S. & Thomas Meiera, U. Nopp140-mediated concentration of telomerase in Cajal bodies regulates telomere length. *Mol Biol Cell* **30**, 3136–3150 (2019).

39. Tseng, T. L. *et al.* The RNA helicase Ddx52 functions as a growth switch in juvenile zebrafish. *Development* **148**, (2021).

40. Tafforeau, L. *et al.* The Complexity of Human Ribosome Biogenesis Revealed by Systematic Nucleolar Screening of Pre-rRNA Processing Factors. *Mol Cell* **51**, 539–551 (2013).

41. Jutzi, D. *et al.* Aberrant interaction of FUS with the U1 snRNA provides a molecular mechanism of FUS induced amyotrophic lateral sclerosis. *Nature Communications 2020 11:1* **11**, 1–14 (2020).

42. Jobert, L. *et al.* Human U1 snRNA forms a new chromatin-associated snRNP with TAF15. *EMBO Rep* **10**, 494–500 (2009).

43. Rappsilber, J., Ajuh, P., Lamond, A. I. & Mann, M. SPF30 Is an Essential Human Splicing Factor Required for Assembly of the U4/U5/U6 Tri-small Nuclear Ribonucleoprotein into the Spliceosome. *Journal of Biological Chemistry* **276**, 31142–31150 (2001).

44. Bayfield, M. A. & Maraia, R. J. Precursor-product discrimination by La protein during tRNA metabolism. *Nat Struct Mol Biol* **16**, 430–437 (2009).

45. Blewett, N. H. & Maraia, R. J. La involvement in tRNA and other RNA processing events including differences among yeast and other eukaryotes. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **1861**, 361–372 (2018).

46. Piskounova, E. *et al.* Lin28A and Lin28B Inhibit let-7 MicroRNA Biogenesis by Distinct Mechanisms. *Cell* **147**, 1066–1079 (2011).

47. Nam, Y., Chen, C., Gregory, R. I., Chou, J. J. & Sliz, P. Molecular basis for interaction of let-7 MicroRNAs with Lin28. *Cell* **147**, 1080–1091 (2011).

48. Mayr, F., Schütz, A., Döge, N. & Heinemann, U. The Lin28 cold-shock domain remodels pre-let-7 microRNA. *Nucleic Acids Res* **40**, 7492–7506 (2012).

49. Heo, I. *et al.* TUT4 in Concert with Lin28 Suppresses MicroRNA Biogenesis through Pre-MicroRNA Uridylation. *Cell* **138**, 696–708 (2009).

50. Piskounova, E. *et al.* Determinants of MicroRNA Processing Inhibition by the Developmentally Regulated RNA-binding Protein Lin28. *Journal of Biological Chemistry* **283**, 21310–21314 (2008).

51. Markert, A. *et al.* The La-related protein LARP7 is a component of the 7SK ribonucleoprotein and affects transcription of cellular and viral polymerase II genes. *EMBO Rep* **9**, 569–575 (2008).

52. Hogan, D. J., Riordan, D. P., Gerber, A. P., Herschlag, D. & Brown, P. O. Diverse RNA-Binding Proteins Interact with Functionally Related Sets of RNAs, Suggesting an Extensive Regulatory System. *PLoS Biol* **6**, e255 (2008).

53. Mittal, N., Scherrer, T., Gerber, A. P. & Janga, S. C. Interplay between Posttranscriptional and Posttranslational Interactions of RNA-Binding Proteins. *J Mol Biol* **409**, 466–479 (2011).

54. Müller-Mcnicoll, M., Rossbach, O., Hui, J. & Medenbach, J. Auto-regulatory feedback by RNA-binding proteins. *J Mol Cell Biol* **11**, 930–939 (2019).

55. Carter, A. C. *et al.* Spen links rna-mediated endogenous retrovirus silencing and x chromosome inactivation. *Elife* **9**, 1–58 (2020).

56. Yepiskoposyan, H., Aeschimann, F., Nilsson, D., Okoniewski, M. & Mühlemann, O. Autoregulation of the nonsense-mediated mRNA decay pathway in human cells. *RNA* **17**, 2108 (2011).

57. White, M. A. *et al.* TDP-43 gains function due to perturbed autoregulation in a Tardbp knock-in mouse model of ALS-FTD. *Nat Neurosci* **21**, 552–563 (2018).

58. Ayala, Y. M. *et al.* TDP-43 regulates its mRNA levels through a negative feedback loop. *EMBO J* **30**, 277–288 (2011).

59. Han, J. *et al.* Posttranscriptional Crossregulation between Drosha and DGCR8. *Cell* **136**, 75–84 (2009).

60. Philippe, L., van den Elzen, A. M. G., Watson, M. J. & Thoreen, C. C. Global analysis of LARP1 translation targets reveals tunable and dynamic features of 5′ TOP motifs. *Proc Natl Acad Sci U S A* **117**, 5319–5328 (2020).

61. Ogami, K. *et al.* mTOR- and LARP1-dependent regulation of TOP mRNA poly(A) tail and ribosome loading. *Cell Rep* **41**, (2022).

62. Berman, A. J. *et al.* Controversies around the function of LARP1. *RNA Biol* **18**, 207–217 (2021).

63. Fuentes, P. *et al.* The 40S-LARP1 complex reprograms the cellular translatome upon mTOR inhibition to preserve the protein synthetic capacity. *Sci Adv* **7**, eabg9275 (2021).

64. Tcherkezian, J. *et al.* Proteomic analysis of cap-dependent translation identifies LARP1 as a key regulator of 5′TOP mRNA translation. *Genes Dev* **28**, 357–371 (2014).

65. Lahr, R. M. *et al.* La-related protein 1 (LARP1) binds the mRNA cap, blocking eIF4F assembly on TOP mRNAs. *Elife* **6**, (2017).

66. Fonseca, B. D. *et al.* La-related protein 1 (LARP1) represses terminal oligopyrimidine (TOP) mRNA translation downstream of mTOR complex 1 (mTORC1). *Journal of Biological Chemistry* **290**, 15996–16020 (2015).

67. Meyuhas, O. & Kahan, T. The race to decipher the top secrets of TOP mRNAs. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **1849**, 801–811 (2015).

68. Ma, X. M. & Blenis, J. Molecular mechanisms of mTOR-mediated translational control. *Nat Rev Mol Cell Biol* **10**, 307–318 (2009).

69. Thoreen, C. C. *et al.* A unifying model for mTORC1-mediated regulation of mRNA translation. *Nature* **485**, 109–113 (2012).

70. Yang, M., Lu, Y., Piao, W. & Jin, H. The Translational Regulation in mTOR Pathway. *Biomolecules* **12**, 802 (2022).

71. Liu, G. Y. & Sabatini, D. M. mTOR at the nexus of nutrition, growth, ageing and disease. *Nat Rev Mol Cell Biol* **21**, 183–203 (2020).

72. Hong, S. *et al.* LARP1 functions as a molecular switch for mTORC1-mediated translation of an essential class of mRNAs. *Elife* **6**, (2017).

73. Qin, X., Jiang, B. & Zhang, Y. 4E-BP1, a multifactor regulated multifunctional protein. *Cell Cycle* **15**, 781–786 (2016).

74. Gingras, A.-C. *et al.* Regulation of 4E-BP1 phosphorylation: a novel two-step mechanism. *Genes Dev* **13**, 1422–1437 (1999).

75. Hsieh, A. C. *et al.* The translational landscape of mTOR signalling steers cancer initiation and metastasis. *Nature* **485**, 55–61 (2012).

76. Jin, H. *et al.* TRIBE editing reveals specific mRNA targets of eIF4E-BP in Drosophila and in mammals. *Sci Adv* **6**, 8771–8783 (2020).

77. Darzacq, X. *et al.* Cajal body-specific small nuclear RNAs: a novel class of 2′-O-methylation and pseudouridylation guide RNAs. *EMBO J* **21**, 2746–2756 (2002).

78. De Boer, E. M. J. *et al.* TDP-43 proteinopathies: a new wave of neurodegenerative diseases. *J Neurol Neurosurg Psychiatry* **92**, 86–95 (2021).

79. Yahara, M., Kitamura, A. & Kinjo, M. U6 snRNA expression prevents toxicity in TDP-43-knockdown cells. *PLoS One* **12**, e0187813 (2017).

80. Izumikawa, K. *et al.* TDP-43 regulates site-specific 2′-O-methylation of U1 and U2 snRNAs via controlling the Cajal body localization of a subset of C/D scaRNAs. *Nucleic Acids Res* **47**, 2487–2505 (2019).

81. Asakawa, K., Handa, H. & Kawakami, K. Multi-phaseted problems of TDP-43 in selective neuronal vulnerability in ALS. *Cellular and Molecular Life Sciences 2021 78:10* **78**, 4453–4465 (2021).

82. Schwanhausser, B. *et al.* Global quantification of mammalian gene expression control. *Nature* **473**, 337–342 (2011).

83. Shishkin, A. A. *et al.* Simultaneous generation of many RNA-seq libraries in a single reaction. *Nature Methods 2015 12:4* **12**, 323–325 (2015).

84. Banerjee, A. K. *et al.* SARS-CoV-2 Disrupts Splicing, Translation, and Protein Trafficking to Suppress Host Defenses. *Cell* **183**, 1325-1339.e21 (2020).

85. Parnas, O. *et al.* A Genome-wide CRISPR Screen in Primary Immune Cells to Dissect Regulatory Networks. *Cell* **162**, 675–686 (2015).

86. Rappsilber, J., Mann, M. & Ishihama, Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat Protoc* **2**, 1896–1906 (2007).

87. Wolin, E. *et al.* SPIDR: a highly multiplexed method for mapping RNA-protein interactions uncovers a potential mechanism for selective translational suppression

upon cellular stress. *bioRxiv* 2023.06.05.543769 (2023) doi:10.1101/2023.06.05.543769.

Chapter 4

CONCLUSION AND FUTURE DIRECTIONS

J.K. Guo and M. Guttman

Credit: Inna-Marie Strazhnik

4.1 CONCLUSION

RNA-protein interactions are crucial for many cellular processes, extending beyond well-characterized roles in post-transcriptional gene regulation to influencing chromatin states and transcription itself. In the context of the nucleus, RNA has been demonstrated to seed local high-concentration gradients to facilitate the binding and recruitment of regulatory proteins to specific spatial territories[1]. With the emerging focus on RNA as a central player in these mechanisms, it is imperative to systematically define the universe of RBPs and their RNA interactions in order to begin dissecting these functions. Existing methods to globally explore these interactions include various RNA-centric proteomic approaches that have led to the identification of thousands of putative human RBPs[2]. Typically, these candidate RBPs are validated by protein-centric approaches to map their specific binding sites genome-wide, such as CLIP[3]. Yet despite the vast strides made in profiling these proteins, including valiant efforts by international consortiums (i.e., ENCODE), only a small fraction (<10%) of all putative targets has been validated to-date[4]. Furthermore, these experiments are often limited to 1-2 cell types and in some specific cases, the functional relevance of these reported interactions has been questionable. Thus, there is a pressing need for new experimental frameworks that are 1) highly specific and stringent to confidently define *bona fide* RBPs, and 2) higher-throughput to profile RBPs at scale across multiple cell lines and cell states.

In this thesis, we showed that many widely reported RNA-protein interactions for chromatin proteins such as PRC2 and CTCF do not occur *in vivo*. Using fully denaturing purifications by CLAP, we abrogated all non-specifically enriched RNAs detected by CLIP while retaining true interactions for known RBPs. Importantly, our findings are consistent with emerging orthogonal evidence that RNA is not essential for PRC2 localization on chromatin[5,6]. We then expanded the utility of the CLAP method to additionally screen for *bona fide* chromatin RBPs, identifying several non-canonical RBPs with annotated RNA-dependent functions in chromatin organization and gene regulation. Despite our findings, we emphasize that the issue does not lie inherently with the CLIP method, but rather how it is often utilized and interpreted. Specifically, CLIP data is standardly interpreted in isolation

(i.e., within its own experiment, with normalization to an input control[7]). We showed that CLIP, like any other molecular biology technique, is susceptible to background signal that is commonly shared across different proteins (i.e., potentially due to abundant spurious RNAs or RNAs crosslinked to abundant non-specific proteins). This can lead to the identification of non-specific RNAs as being specifically enriched, especially when proteins do not bind to RNA *in vivo*. However, when explored relative to true RBPs (i.e., by an RNA gel), we observed dramatically lower overall CLIP signal for non-RBPs like PRC2 and GFP. This suggested to us that direct, quantitative comparisons between proteins would allow us to resolve issues of CLIP specificity.

This is best highlighted with our SPIDR method, which effectively multiplexes CLIP experiments by several orders of magnitude. With each SPIDR experiment, we observed a moderate degree of shared background RNAs across different proteins within our targeted pools. This is likely due to the same reasons that individual CLIP is prone to detecting non-specific signal, which we fully explored in **Chapter 2** (i.e., from necessitated non-denaturing washes). There may additionally be off-target effects due to variable antibody quality, which is inherent to all immunoprecipitation (IP)-based methods. Regardless of the exact source of background, we normalized for this by leveraging the fact that each SPIDR experiment is inherently performed in pools of various different proteins. Specifically, we employed a background correction method that allowed us to perform quantitative comparisons across each protein in the pool, accounting for variables such as target read depth. Indeed, it was only with this approach that we uncovered a novel LARP1 binding site on the 18S ribosomal RNA with single-nucleotide resolution, which was otherwise obscured when analyzing the raw data (**Figure 1**). Because each protein within a set is subject to identical IP and wash conditions, this normalization strategy is more appropriately applicable to SPIDR than individual CLIP experiments which are subject to high batch variation.

Another point of emphasis we highlighted in **Chapter 1** was the importance of validating RNA-protein interactions using orthogonal approaches, including precise functional domain perturbations. To explore the specific LARP1-40S binding identified by SPIDR, we purified

recombinant human LARP1 and assembled it with the 40S ribosome from HEK293T cell lysate. We then used single-particle cryo-electron microscopy (cryo-EM) to determine the structure of the human LARP1-40S complex and resolved it to an overall resolution of 2.7 Å (**Figure 2**). We found that the recently identified Ribosome Binding Region (RBR) of LARP1[8] bound the mRNA entry channel in the 40S structure at a site corresponding to the exact binding site identified by SPIDR (**Figure 2**). This corroborated our unique model in which the 40S-bound LARP1 complex is capable of binding simultaneously to TOP-motif containing mRNAs to regulate their translation, as we explored in detail in **Chapter 3**. Furthermore, we found that deletion of the RBR, which is distinct from the domain responsible for binding TOP mRNAs (DM15), abrogates LARP1 binding to the ribosome (**Figure 3**). Together, this highlights the power of the SPIDR method to identify novel RBP binding sites with functionally relevant implications, as LARP1 is widely known to modulate translation yet has not been directly linked to ribosome binding before.

More generally, we expect methods such as SPIDR and CLAP to define a new framework for evaluating *in vivo* RNA-protein interactions. On one hand, SPIDR enables rapid identification of novel RBPs and RBP-binding sites for hundreds of targets within a single experiment, allowing for broad characterization of RBPs at an unprecedented scale. In addition, SPIDR can measure global changes in RBP binding across different cell states and treatment conditions, such as the mTOR inhibition experiment described in **Chapter 3** in which we uncovered a potential translational suppression mechanism of sequence-specific mRNAs. On the other hand, CLAP serves as an invaluable method for targeted validation of putative RBPs, especially RBPs with reported non-canonical RNA-binding domains (RBDs). CLAP has additional utility for profiling RBPs that lack existing high-quality antibodies or affinity reagents, by virtue of fusing a covalent tag to each protein of interest. We anticipate the combined synergy of these tools to pave the way for defining new classes of RBPs and, by proxy, the multitude of ncRNA functions.

4.2 FUTURE DIRECTIONS

As alluded to in **Chapter 1**, one key open question is why such a large fraction of the human proteome has evolved to bind RNAs. Of particular interest is the potential of RNA to specifically recruit regulatory proteins to chromatin to modulate gene expression. It is now clear that many noncoding RNAs achieve this by seeding the formation of spatial compartments in the nucleus to bind and recruit proteins into them[1]. RNA molecules are uniquely suited to this role as they accumulate at high concentrations at their transcriptional loci and adopt structural conformations to bind specific proteins. One prime example of this is the mechanism by which SPEN, an RBP, directly interacts with the Xist lncRNA to specifically recruit HDAC3 to the inactive X chromosome (Xi)[9]. This discovery highlights how chromatin regulators (SPEN) and chromatin-modifying complexes (HDAC3) can be recruited to specific DNA regions through interactions with RNA. SPEN has additionally been shown to localize to hundreds of other nuclear compartments in an RNA-dependent manner[1]. In **Chapter 2**, we demonstrated that SPEN binds to hundreds of RNAs beyond Xist, including lncRNAs and notably many nascent pre-mRNAs that remain enriched on chromatin near their genomic locus. Because the silencing of the Xi is dependent on HDAC3 and the interaction between SPEN and Xist, other SPEN-bound RNAs may be regulated in the same manner. TET2, which we uncovered as a *bona fide* chromatin RBP in **Chapter 2**, is another protein that appears to form nuclear compartments in a similar manner as SPEN (**Figure 4**). It has also been shown to localize to chromatin in a transcription-dependent manner[10], though the RNA-dependence of this compartment formation has yet to be explored. While TET2 itself functions as a DNA demethylase, it has also been reported to bind to HDAC1/2, which may represent another example in which HDAC complexes (which lack DNA-binding domains) are indirectly recruited to chromatin by RNA.

The specific examples highlighted here with SPEN and TET2 may represent a more widespread mechanism of gene regulation by which the cell achieves precise modulation of gene targets through RNA-protein interactions (**Figure 5**). Further validation of these mechanisms may involve precise genetic perturbations of SPEN or TET2 RNA-binding sites (i.e., by CRISPR-Cas9) followed by measurements of gene expression or chromatin states (i.e., H3K27ac for SPEN/HDAC3; 5hmC for TET2). Large-scale functional screens to

identify additional chromatin proteins with RNA-dependent functions (i.e., by proteomics) will also be crucial for understanding how generalizable this principle of RNA-guided gene regulation is.

Analogous to the dismantling of old structures to lay foundations for lasting strength, we expect a paradigm shift in the near future for understanding what constitutes a biologically meaningful RNA-protein interaction. The continued innovation of high-stringency and high-throughput technologies to profile RBPs, coupled with careful dissection of their functions, will pave the way for more accurate and reproducible findings. Together, this will be critical for exploring the universe of all things RNA.

4.3 FIGURES

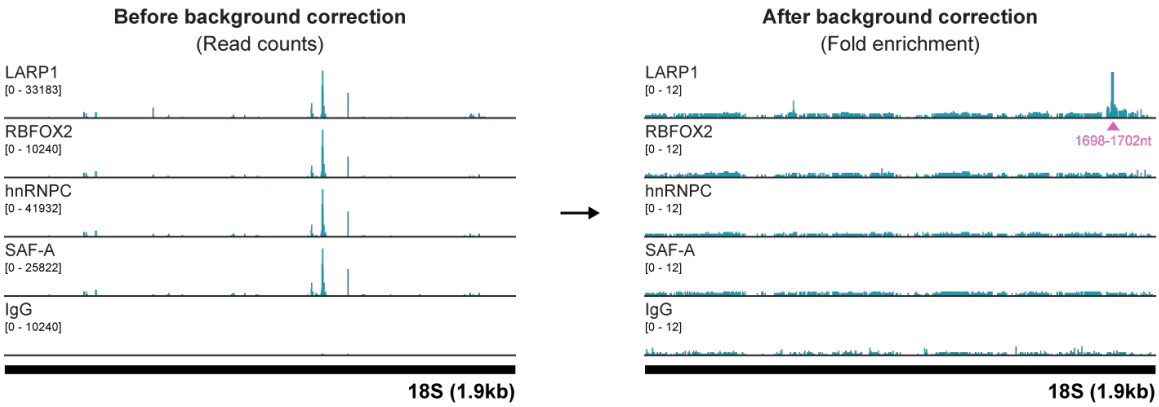


Figure 1: SPIDR enables identification of novel RNA-binding sites when interpreted relative to other RBPs.

(Left) SPIDR RNA read coverage of LARP1, RBFOX2, hnRNPC, SAF-A, and IgG on the 18S ribosomal RNA before background correction. (Right) Background-corrected data for the same set of proteins. Pink arrow denotes the previously reported LARP1 binding site between the 1,698-1,702 nucleotide position.

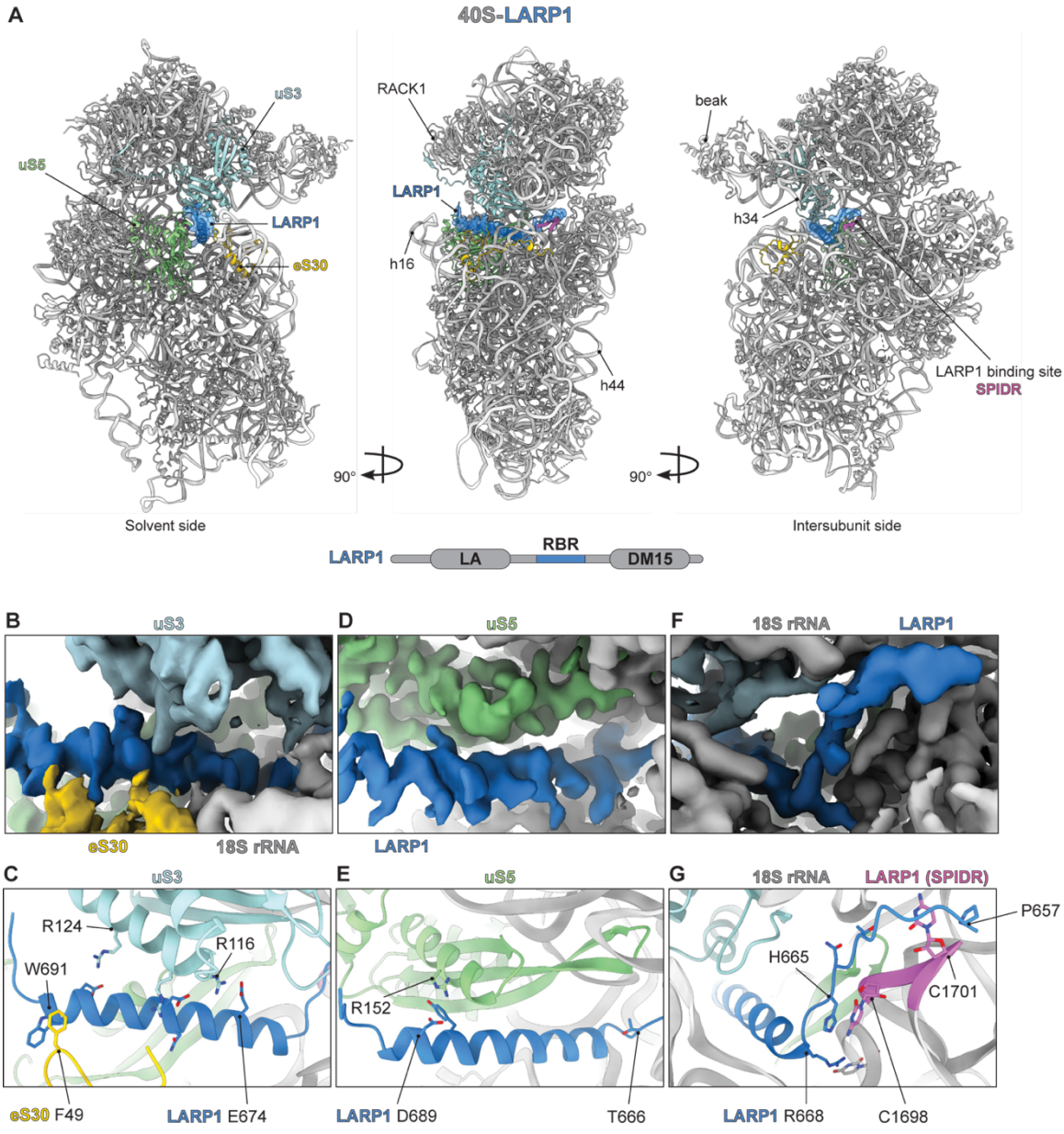


Figure 2: Cryo-EM structure of human LARP1 bound to the 40S subunit.

(A) The overall view of the 40S-LARP1 structure is shown in three different orientations. RBR of LARP1 (blue) fitted into the cryo-EM map (blue) to highlight its position within the mRNA channel in the 40S. **(B-C)** Cryo-EM map and atomic model to highlight the interaction between LARP1 and ribosomal proteins uS3 and eS30 located at the mRNA entry site. **(D-E)** LARP1 interacts with the ribosomal protein uS5 within the mRNA channel. **(F-**

G) Close-up to highlight the interaction between LARP1 and the universal conserved decoding bases in the P site and the A site, as identified by SPIDR.

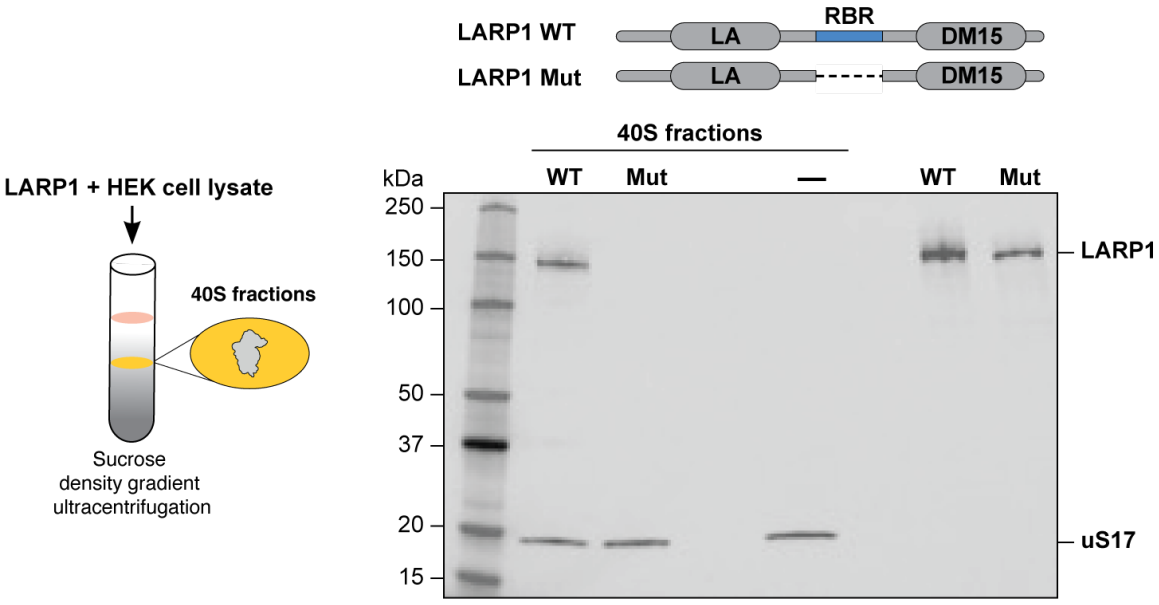


Figure 3: RBR of LARP1 is required for binding to 40S.

Recombinant wild-type (WT) LARP1 and a mutant LARP1 containing the RBR deletion (Mut) (top, right) were expressed, purified, and incubated with HEK293T whole cell lysate. 40S complexes were isolated by sucrose density gradient ultracentrifugation, then subject to western blot of both LARP1 and the control ribosomal protein uS17 (40S fractions). "–" denotes the 40S fraction from a lysate without LARP1. Rightmost two lanes depict positive controls probed against the recombinant proteins, demonstrating that the LARP1 antibody recognizes both forms of LARP1.

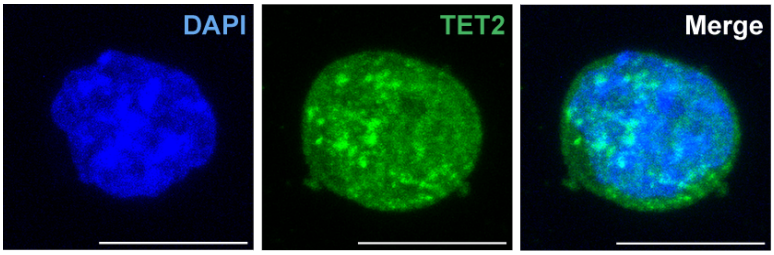


Figure 4: TET2 forms spatial compartments in the nucleus.

Immunofluorescence (IF) staining performed against TET2 (Abcam, anti-Tet2) in pSM33 mouse embryonic stem cells (green). Scale bars, 10 μM.

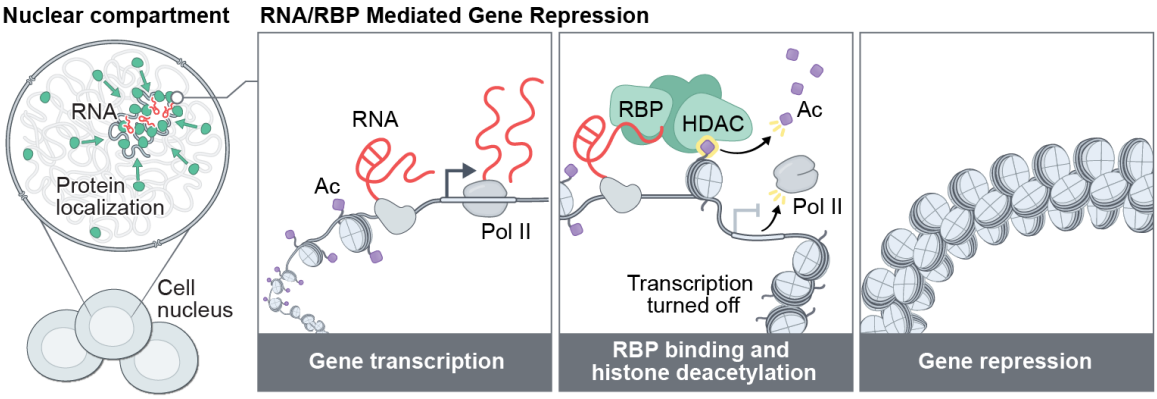


Figure 5: RBPs and chromatin regulators are recruited by RNA to specific genomic sites.

Model for how RBPs and effector chromatin complexes (HDAC shown here as an example) are recruited to specific genome territories within RNA-mediated spatial compartments to enact their functions.

4.4 REFERENCES

1. 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. 10.1016/j.cell.2021.10.014.
2. 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. 10.1093/nar/gkaa1040.
3. 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 (1979) *302*, 1212–1215. 10.1126/science.1090095.
4. 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. 10.1038/s41586-020-2077-3.
5. 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. (2024). The apparent loss of PRC2 chromatin occupancy as an artifact of RNA depletion. Cell Rep *43*, 113858. 10.1016/j.celrep.2024.113858.
6. Hall Hickman, A., and Jenner, R.G. (2024). Apparent RNA bridging between PRC2 and chromatin is an artifact of non-specific chromatin precipitation upon RNA degradation. Cell Rep *43*, 113856. 10.1016/j.celrep.2024.113856.
7. Hafner, M., Katsantoni, M., Köster, T., Marks, J., Mukherjee, J., Staiger, D., Ule, J., and Zavolan, M. (2021). CLIP and complementary methods. Nature Reviews Methods Primers *1*, 20. 10.1038/s43586-021-00018-1.
8. Saba, J.A., Huang, Z., Schole, K.L., Ye, X., Bhatt, S.D., Li, Y., Timp, W., Cheng, J., and Green, R. (2023). LARP1 senses free ribosomes to coordinate supply and demand of ribosomal proteins. bioRxiv, 2023.11.01.565189. 10.1101/2023.11.01.565189.

9. 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. 10.1038/nature14443.

10. 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. 10.1038/s41588-018-0060-9.