Elucidating Nucleic Acid Binding Properties of Polycomb Repressive Complex 2

By Richard D. Paucek

Biochemistry Thesis Defense
March 31st, 2017

Cech Laboratory
Department of Chemistry and Biochemistry,
BioFrontiers Institute,
University of Colorado Boulder
Boulder, CO, 80309-0215, USA

Primary Thesis Advisor: Dr. Thomas R. Cech, Ph.D.

Defense Committee Members: Dr. Dylan Taatjes, Ph.D., Dr. Leslie Leinwand, Ph.D., and Dr. Thomas R. Cech, Ph.D.

TABLE OF CONTENTS

ABSTRACT III

ACKNOWLEDGMENTS IV

1. BACKGROUND AND SIGNIFICANCE

1.1. Chromatin-Modifiers: Regulators of Gene Expression……1

1.2. The Polycomb Repressive Complex 2 Silences Genes……2

1.3. Current Models for the Recruitment of PRC2 to Genes……5

1.4. Emerging Roles of RNA in PRC2 Regulation and Recruitment……8

1.5. Thesis Aims……14

2. PRC2 TARGETS GUANINE-RICH RNA

2.1. RNA-based Recruitment of PRC2……17

2.2. PRC2 Preferentially Associates with G-rich RNA in Vitro and in Vivo……18

2.3. PRC2 Binds Tightly to G-quadruplex forming RNAs……23

2.4. G-quadruplex Stabilizing Ligands Inhibit PRC2-RNA Interaction……27

2.5. Materials and Methods……30

3. NUCLEOSOME-FREE DNA DICTATE PRC2 BINDING IN VITRO

3.1. Histone-based Recruitment of PRC2……36

3.2. PRC2 Binding to Nucleosomes……38

3.3. PRC2 Interacts with Linker-Regions of Nucleosomes……44

3.4. The Effect of RNA on PRC2 Association with Nucleosomes……48

3.5. Materials and Methods……50

4. PRC2 IS GUIDED TO SITES OF DNA METHYLATION

4.1. DNA-based Recruitment of PRC2……..……..…… 54

4.2. Testing PRC2-DNA Binding Properties…….…… 55

4.3. PRC2 Recognizes 5-Methyl-CpG Dinucleotodies in DNA…….…… 58

4.4. Materials and Methods…… 60

5. DISCUSSION 62

6. REFERENCES 65

Abstract

Polycomb Repressive Complex 2 (*PRC2*) is a histone methyltransferase that specifically deposits mono-, di-, and tri-methylation marks onto chromatin. This activity triggers epigenetic silencing, a process critical for cell differentiation and maintenance of cellular identity. In mammalian cells, how PRC2 is recruited to target sites is unknown, but it is speculated that RNA, histone modifications, nucleosome architecture, and DNA elements all possess direct or indirect recruitment and regulatory roles. However, the relative binding affinity of PRC2 for these diverse biological substrates remains poorly understood. In the present study, the binding affinity of PRC2 for various RNAs, nucleosomes, and DNAs were tested using bulk biochemical and single-molecule approaches using purified recombinant human PRC2 complex. In addition, the effect of PRC2's association with nucleosomes in the presence of RNA was tested. Binding experiments revealed that PRC2 reads consecutive guanines in RNAs, and that RNA sequences with the propensity to form G-quadruplex structures are ideal ligands. Testing PRC2 interaction with nucleosomes demonstrated that nucleosome substrates with greater numbers of nucleosome repeats bound more tightly. Yet, it was subsequently determined that PRC2 binding to nucleosomes is in fact dictated by nucleosome-free linker DNA regions. Intriguingly, PRC2-nucleosome associations are weakened in the presence of competitor RNA. Finally, the DNA-binding ability of PRC2 was studied, where it was observed that PRC2 has the biophysical property of condensing DNA, and can preferentially recognize methylated DNA. Collectively, findings from this work offer new insights into the mechanistic details of PRC2 recruitment.

Acknowledgements

I would like to thank past and present members of the Cech Laboratory who have invited insightful discussion and offered outstanding technical support during the course of this independent project. Research in the lab was so exhilarating that it never crossed my mind that I might be working too feverishly. And it was not only surprising, but eye opening for me to discover a culture in the biological sciences that only seeks to explore and push the boundaries of human understanding, creativity, and curiosity, and it was a thrilling and rare opportunity to be a part of that. This background has contributed immensely to my growth as an independent thinker and voracious scientific investigator.

I would like to highlight the support that CJ, Xueyin, and Tom have provided. CJ has not only been a remarkable mentor, but someone who I can confidently say will be a lifelong friend. I am also grateful to Xueyin. It was outstanding to work alongside of Xueyin, who made the "PRC2 Dream Team" a reality. Her generosity and good humor was appreciated daily. Lastly, I want to thank Tom. Throughout my time in the lab, Tom has been consistently an absolute standout faculty mentor who has been unyielding in his support of my research interests. He wholeheartedly encourages the creative freedom of scientific exploration, and is superb when it comes to explaining the theories that drive experiments. Certainly, Tom offers a depth of insight not often seen from many people.

Section 1 — Background and Significance

1.1 — Chromatin-Modifiers: Regulators of Gene Expression

At a given time, a typical human cell may express only a small fraction of the more than 20,000 protein-coding genes present in chromosomes. **Chromatin structure**, the packaging of DNA by histone proteins, is a critical regulator of gene expression. Cellular differentiation, as well as cancer, are characterized by a re-organization of chromatin architecture that alters global gene expression patterns. **Chromatin-remodeling** is a cellular process that sculpts chromatin structure through covalent and non-covalent modifications. And it is now established that gene inactivation in cells rests on the recruitment of chromatin-modifying enzymes that function to compact chromatin, a hallmark of gene repression. When the gene repression is inherited through multiple cell divisions, or even in a new generation of organisms, it is called "***epigenetic silencing***." Among the class of proteins that carry out this regulatory role are **Polycomb-group** (**PcG**) proteins, a highly conserved family of silencing factors that exert gene silencing through the modification of histone proteins.[1,7] Moreover, it is known that mutations in PcG proteins have pleiotropic effects that can lead to the inadvertent activation of oncogenes and repression of tumor suppressor genes, key drivers of oncogenesis.[9,13,88] However, the mechanism by which mammalian PcG proteins are recruited to select genes to enforce silencing remains obscure, information which if known could lead to new avenues for cancer therapeutics.

1.2 — *The Polycomb Repressive Complex 2 Silences Genes*

Polycomb Repressive Complex 2 (**PRC2**) is one of the two major PcG protein complexes in human cells required for epigenetic silencing. Given this biological role, PRC2 is essential for embryonic development and cellular differentiation, as demonstrated by the embryonic lethality of deletion of several of its core subunits in mice.[55] And it is known that PRC2-mediated epigenetic silencing can promote cancer by repressing tumor-suppressor genes and its dysregulation also lead to oncogene de-repression.[9,51,75] Indeed, PRC2 has

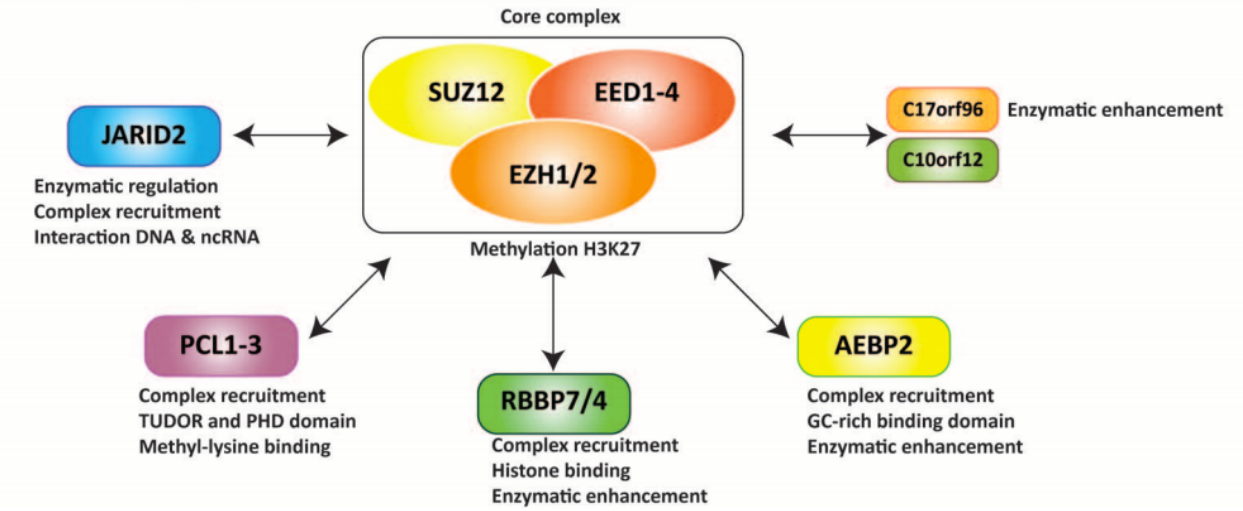


Figure 1 | Subunit composition of Polycomb Repressive Complex 2 (PRC2).[1] The polypeptides, EZH2, EED, and SUZ12 comprise the "*Catalytic Core*" complex of PRC2. Functionally, EZH2 contains a conserved SET domain with histone methyltransferase activity. And EZH1 is the catalytic component of a non-canonical PRC2 complex. Various proteins associate with the core complex to add additional functionality. These include JARID2, PCL1-3, RBBP4/7, AEBP2, and C17orf96/C10orf12. Notably, the accessory subunit AEBP2 is a DNA-binding transcriptional repressor, whereas RBBP4 is a histone-binding subunit. JARID2 on the other hand is thought to bind GC-rich DNA and thought to modulate PRC2 enzymatic activity.

gathered much interest in the development of cancer drugs that can inhibit its catalytic activity.[12]

PRC2 is a multi-subunit protein complex and histone methyltransferase that deposits a mono-, di-, and tri-methylation mark on lysine 27 of the H3 protein ("***H3K27me3***"). Genome-wide, H3K27 methylation is strongly correlated with condensed, or repressed, chromatin structure that promotes gene-inactivation.[55] Mechanistically, chromatin compaction is thought to silence genes by restricting gene promoter access to transcriptional machinery. Therefore, these histone methyl modifications ultimately result in the inactivation of genes. The "*catalytic core*" of the PRC2 complex comprises three essential subunits: **EZH2**, **EED**, and **SUZ12** (as depicted in the cartoon in **Figure 1**), which together are necessary for basal H3K27 methyltransferase activity. In particular, EZH2 is the catalytic moiety responsible for the addition of methyl groups to histone H3 at lysine 27. This is accomplished by the SET domain of EZH2 by using the methyl group donor S-adenosyl-L-methionine (**SAM**). EED on the other hand has a "reader" role by directly binding to pre-existing tri-methyl lysine residues on histone H3. H3K27me3 binding by EED has a positive role on gene silencing by stimulating PRC2 catalytic activity. The third component of the ternary complex is SUZ12, a zinc-finger protein which might be involved in RNA-binding. SUZ12 also has a reader role as the VEFS domains of SUZ12 is known to be involved in the recognition of "active" chromatin marks such as H3K36me3.[20] Notably, SUZ12 mutations are strong suppressors of position-effect variegation.[7] Accessory subunits that associate with the core complex, such as AEBP2, RBPP4, and JARID2, have been proposed to expand the basic functions of the core complex itself by altering biochemical properties such as catalytic

efficiency, or giving the complex entirely new functions like the ability to bind different histone post-translational modifications (**PTMs**).[20,39]

It is also important to underscore that each PRC2 component is not a simple static entity. Rather, allosteric regulation of the complex occurs through a crosstalk between subunits. For instance, as discussed above, the binding of methylation marks by EED

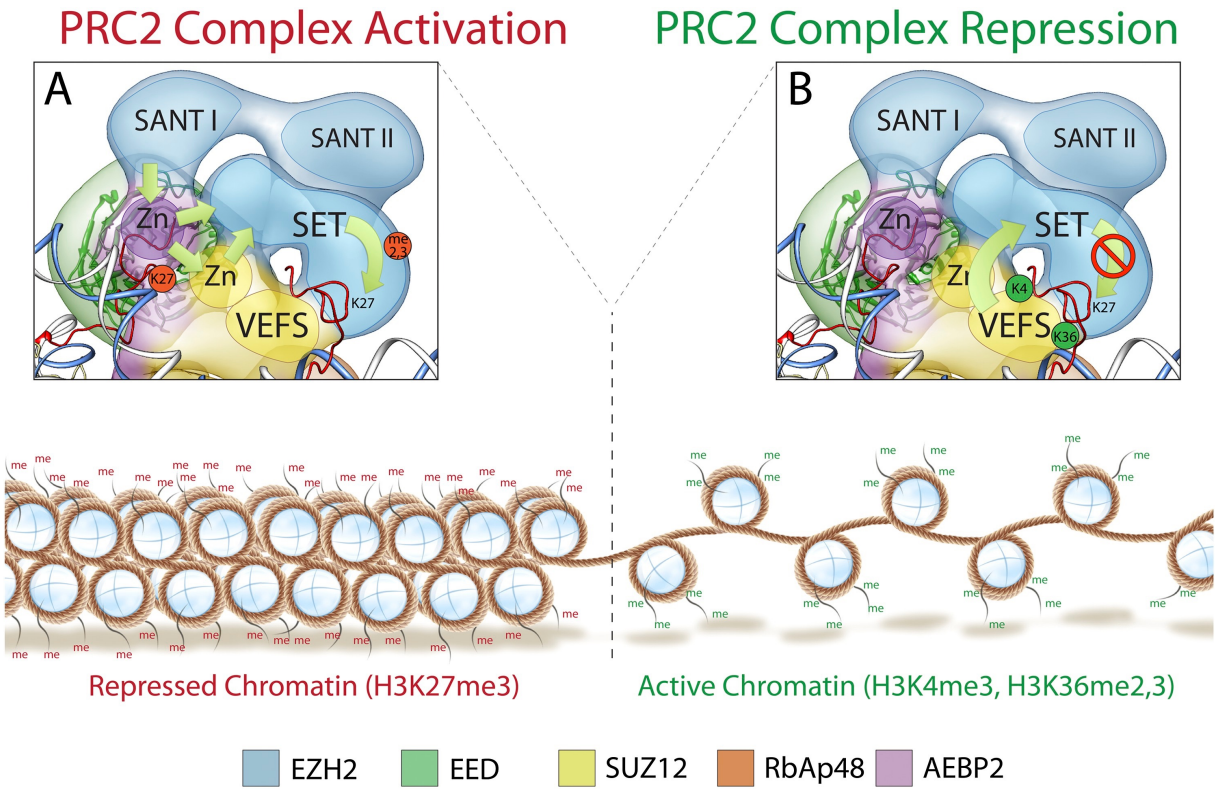


Figure 2 | Mechanism and Allosteric Regulation of PRC2 during gene silencing.[20]
A) At genomic regions of repressed chromatin, H3K27me3 marks are detected by the EED component. This recognition is transferred via SANT domains to the SET domain of EZH2, and this stimulates methyltransferase activity of EZH2. This is thought to reinforce gene silencing.
B) At active genomic regions, H3K4me3 and H3K36me3 active marks are recognized by the VEFS domain of SUZ12. This binding results in an allosteric change that inhibits the enzymatic activity of EZH2.

stimulates the catalytic activity of the PRC2 complex.[20] And H3K4me3 and H3K36me3 marks, which are normally associated with active chromatin, are also recognized by PRC2 and result in inhibition of PRC2 catalytic activity.[55] The roles of different subunits in the positive and negative regulation of the PRC2 complex are summarized in **Figure 2**.

1.3 — *Current Models for the Recruitment of PRC2 to Genes*

The precise mechanism of activity and regulation of the PRC2 complex remains an active area of research, and hinges on acquiring atomic-level structural information either by X-ray crystallography or single-particle cryo-electron microscopy (cryo-EM). However, perhaps the most outstanding question regarding PRC2 biology is that of its gene specificity. Precisely, what is the decision-maker that brings PRC2 to its target genes? During the process of stem-cell differentiation, whether a cell becomes a neuron, a myocyte, an adipose cell, and forth, is a matter of the correct combination of genes being switched on and off. Therefore, as an epigenetic regulator, clearly PRC2 must be able to discriminate between different sets of genes to exert silencing in a cell-type specific manner. And in the case of some types of cancer, where PRC2 is implicated in its role of repressing chromatin regions associated with tumor-suppressor genes, unknown factors must also somehow be able to hijack or dysregulate the recruitment mechanism of PRC2.

Historically, the recruitment mechanism of PRC2 has received thorough investigation in fruit flies, a model organism for the study of development biology and epigenetic regulation. Indeed, in *Drosophila*, PRC2 recruitment is rather well understood and is known to be dictated by the recognition of specific Polycomb Response Elements (**PRE**) in DNA.[25]

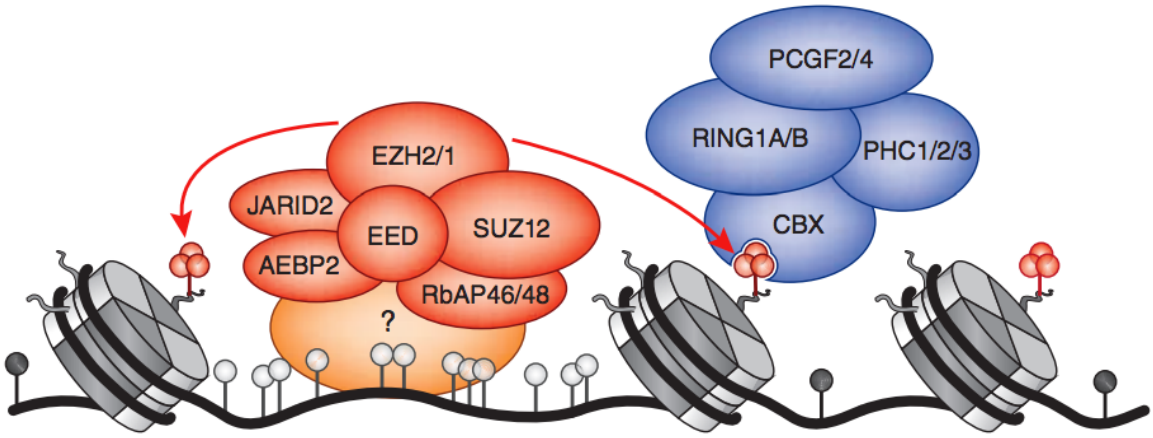


Figure 3 | A Model of PRC2 recruitment to target genes.[25] PRC2 complex with JARID2 and AEBP2 accessory subunits is proposed to be recruited to CpG island's through a mechanism that remains to be elucidated. Recruitment may involve other unknown factors (marked by "?"). PRC2 recruitment to genes leads to the deposition of H3K27me3 (red circles). The H3K27me3 marks can then recognized by other factors, such as PRC2 itself, or the CBX subunit of PRC1.

PREs are cis-regulatory DNA elements and binding through these response-elements recruits *Drosophila* PRC2 to chromatin, and therefore control transcription of genes in the vicinity nearby.[25] However, in mammals, such analogous PREs are not found. It appears that despite the good conservation between *Drosophila* and human PRC2 proteins, the molecular system that governs PRC2 targeting and regulation might be significantly different. Therefore, detailed information regarding PRC2-specific recruiters in mammals is still not understood. While the mechanism of PRC2 recruitment to genomic loci in humans remains to be elucidated, because none of the core components of PRC2 (EZH2, EED, SUZ12) possess an obvious DNA-binding domain,[55] it is thought that perhaps chromatin targeting must be

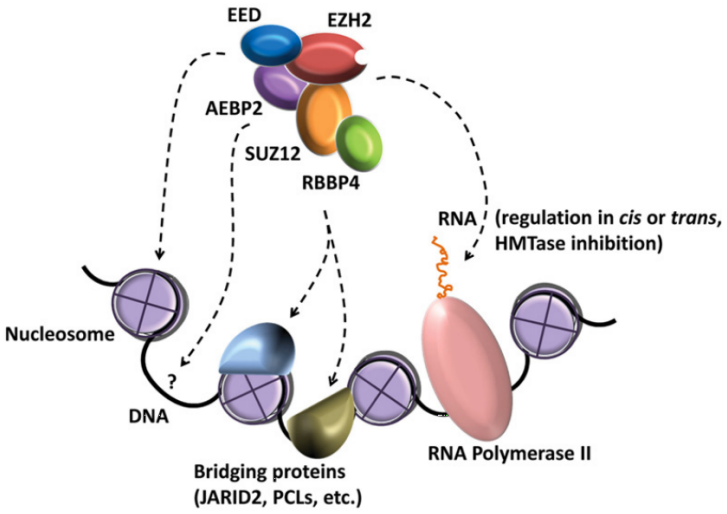


Figure 4 | Multiple modes of PRC2 recruitment to chromatin in mammals.[27] PRC2 is a methyltransferase complex involved in gene silencing. Direct and indirect interactions with nucleosomes, bridging proteins, RNA, and DNA may be responsible for the recruitment of PRC2 to genomic loci.

specified extrinsically from the core complex. To date, several mechanisms of recruitment of PRC2 to genomic loci have been proposed: First is a DNA-based mechanism that involves the association of the PRC2 core complex with co-factors that have DNA-binding ability, as illustrated in **Figure 3**. For instance, the proteins AEBP2 and JARID2 have been proposed to help mediate PRC2 targeting through preferential binding of GC-rich DNA regions.[49,62,70] A second mechanism posits that chemical modifications displayed on histone proteins are responsible for the positive and negative recruitment of PRC2. The premise of this model stems from the chromatin "reading" ability of PRC2, namely through the EED subunit which recognizes various histone PTMs.[20] Also influencing PRC2 targeting is the effect of chromatin density, which has been suggested to be a potent recruiting factor.[75] And yet a third

mechanism puts forth that nascent chromatin-associated long non-coding RNAs (***lncRNAs***) act as biological lassos which sequester PRC2 and enriches PRC2 complexes at particular target genomic loci.[30] These various modes of binding are illustrated by the cartoon model in **Figure 4**. Indeed, PRC2 targeting might involve a combination of some, if not all, of the aforementioned mechanisms. Given the decisive cellular role of PRC2 in maintaining the repressed states of different combinations of genes that vary by cell type, one might infer that PRC2 must be targeted to genes by a well-coordinated and intricate process.

1.4 — *Emerging Roles of RNA in PRC2 Regulation and Recruitment*

Female mammalian cells silence one of their two X chromosomes to equalize X-linked gene expression with male mammalian cells.[50] This is a cellular event that forms a unique condensed structure called the **Barr body**. By light microscopy of a DNA-stained nuclei, a Barr body appears as a dense and dark spot at the nuclear periphery of a human female cell. And X-chromosome inactivation (**XCI**) describes the molecular and cellular process responsible for equalizing gene expression between male and female mammals.[50]

The molecular basis of XCI has been the subject of decades of intense scientific research and controversy. It is now known that the chief mediator of XCI is in fact a 17-kb long noncoding RNA (lncRNA) called **XIST** (X-inactive specific transcript), which is transcribed from the *Xist* gene on the X-chromosome in female cells.[65] While the XIST RNA is post-transcriptionally processed as most RNAs, and therefore undergoes modifications

such as capping, splicing, and polyadenylation, unlike mRNAs, XIST is not exported to the cytoplasm.[18] Instead, XIST remains exclusively retained in the nucleus. Strikingly, when XIST RNA is transcribed from its gene on the X-chromosome, it coats the X-chromosome *in cis*. This coating of the X-chromosome by XIST RNA is well observed by RNA fluorescence *in situ* hybridization experiments, one example of which is shown in **Figure 5**. This RNA-coating is known to initiate XCI and result in the exclusion of RNA polymerase II (*RNAPII*) from promoter regions, lead to the deposition of H3K27me3 by PRC2, and culminating in the eventual silencing of all gene expression on the X-chromosome. [50]

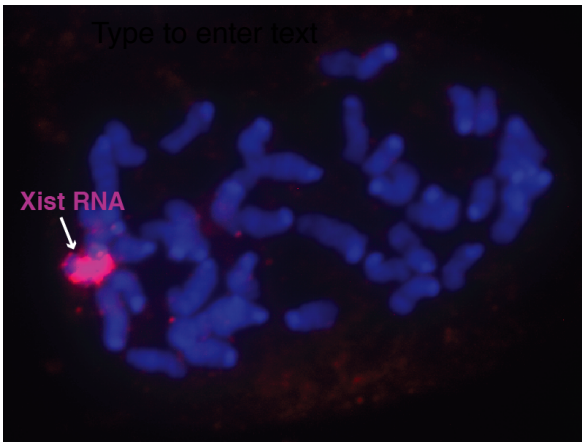


Figure 5 | XIST RNA coats the X-chromosome *in cis*.[50] XIST RNA (shown in red) shows co-localization with the DAPI-stained X-chromosome (DNA shown in blue).

Following its discovery, XIST RNA quickly shifted the paradigm in the field for the functional roles that lncRNAs play in gene regulation, and became a highly-cited example of how mammalian lncRNAs may serve as mediators of epigenetic regulation. This example of XIST-mediated X-chromosome inactivation, along with others, prompted researchers in the PcG-protein field to speculate that perhaps RNA is the "missing" recruitment factor for mammalian PRC2 that was being hunted after for so long.

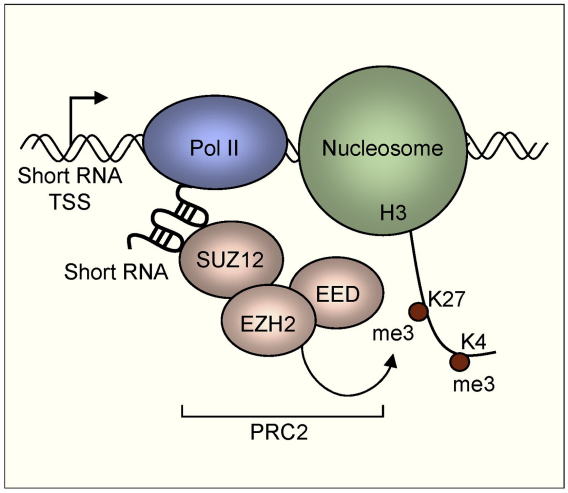


Figure 6 | Recruitment of PRC2 by specific interaction with nascent RNA.[45] A prior study had suggested that PRC2 bound directly to the XIST-RepA stem-loop motif.[95] Data from that study purported a model in which RNAs transcribed from the 5′ end of Polycomb regulated genes interact with PRC2, and thereby facilitates PRC2 interaction with chromatin and permits the repression of genes *in cis*.

The initial observation that H3K27me3 silencing marks are enriched on the X-chromosome during XCI is what sparked the general hypothesis that the PRC2 methyltransferase is involved, and that perhaps RNA directly recruits PRC2. Indeed, an early model of PRC2-mediated XCI proposed that the "**RepA**" stem-loop region of XIST RNA forms a structural motif that specifically recruits PRC2 to the X-chromosome.[45] A cartoon model of RNA-mediated recruitment by RepA is shown in **Figure 6**. As the model puts forth, PRC2 binding to RepA would in-turn result in the methylation of H3K27 of nucleosomes and cascade into the repression of genes.[45]

Despite numerous attempts to define the molecular mechanisms underlying XCI, it is worth noting that there still remains several experimental discrepancies that render a model of direct recruitment of PRC2 by XIST RNA problematic. For instance, although the initial PRC2 binding region on XIST that was defined *in vitro* is required for XCI,[45] it appears to be

dispensable for PRC2 recruitment *in vivo*.[19] And in addition to a model that proposes a direct interaction between PRC2 and XIST RNA, there are several studies that suggest an indirect model of PRC2 recruitment to the X-chromosome that involves other proteins.[18] Therefore the functional role that RNA molecules play in PRC2 recruitment and activity remains elusive not only in the case of XCI mechanism, but in the general scheme of polycomb-regulated genes. Perhaps part of the problem is that the RNA-binding activity of PRC2 is not well understood, and therefore, attempts to attribute functional roles to PRC2-RNA interaction have been difficult. Interestingly, while significant research effort has been invested into the development of small molecules that target the histone methyltransferase activity of PRC2 as an avenue for cancer therapy,[47,56] the RNA-binding property of PRC2 has received far less attention.

In the Cech laboratory, the physical association of RNA and PRC2 is of great interest. And while PRC2 is clearly an RNA-binding protein both *in vitro* and *in vivo*,[46,83,93] how and why PRC2 binds RNA have been at the center of several alternative and competing perspectives.[10,30] The dominant model for RNA-functionality is the RNA-recruitment model, where PRC2 has been suggested to interact with RNA specifically in a way that facilitates its recruitment to chromatin. The premise of the RNA recruitment model is that nascent, newly formed RNA transcribed at active genes contain specific motifs that are preferentially bound by PRC2. This then leads to an enrichment of PRC2 at the particular gene, and expedites silencing. However, this basic paradigm has been shifted in the Cech laboratory, where it has been found that purified human PRC2 is able to bind very diverse lncRNAs and mRNAs *in vitro* with quite similar affinities.[28,30] Moreover, RNA immuno-precipitation studies coupled

to sequencing analysis (**RIP-seq**) in human cells has demonstrated that PRC2 is able to bind thousands of RNA molecules *in vivo.* This is a finding that is consistent with *in vitro* binding observations. These results represent a stunning departure from the simple "RepA-binding" model (**Figure 6**), and have prompted researchers to declare PRC2 to be a "***promiscuous***" RNA-binding protein.[30] But how would promiscuous RNA-binding be possible, considering what is known about protein-RNA interactions? In many cases, protein-RNA interactions are specific and rely on a defined RNA sequence or structural motif. For instance, the RNase P enzyme is solely responsible for recognizing and processing all transfer RNA (tRNA) molecules through structural recognition, and does not process other RNA ligands lacking the appropriate tertiary shape.[68] Indeed, while it sounds contradictory for a promiscuous RNA-binding protein to have inherent specificity, these two are not necessarily mutually exclusive. Instead, it is possible that "hidden specificity" exists in the recognition of RNAs by the PRC2 complex.

How then might PRC2 be able to accomplish the herculean task of associating with most of the transcriptome? One obvious solution to this problem would be if the interaction with RNA involved the sugar-phosphate backbone and is largely driven by electrostatics with 2'-hydroxyl groups. However, such a binding-mode would be indicative of salt-dependence. And the very modest salt dependence of the PRC2-RNA interaction as reported by Davidovich and colleagues refutes this hypothesis.[30] Perhaps the best remaining explanation may be that the RNA-binding motif is so common that it is in fact ubiquitous among thousands of RNA molecules. This hypothesis remains to be explored.

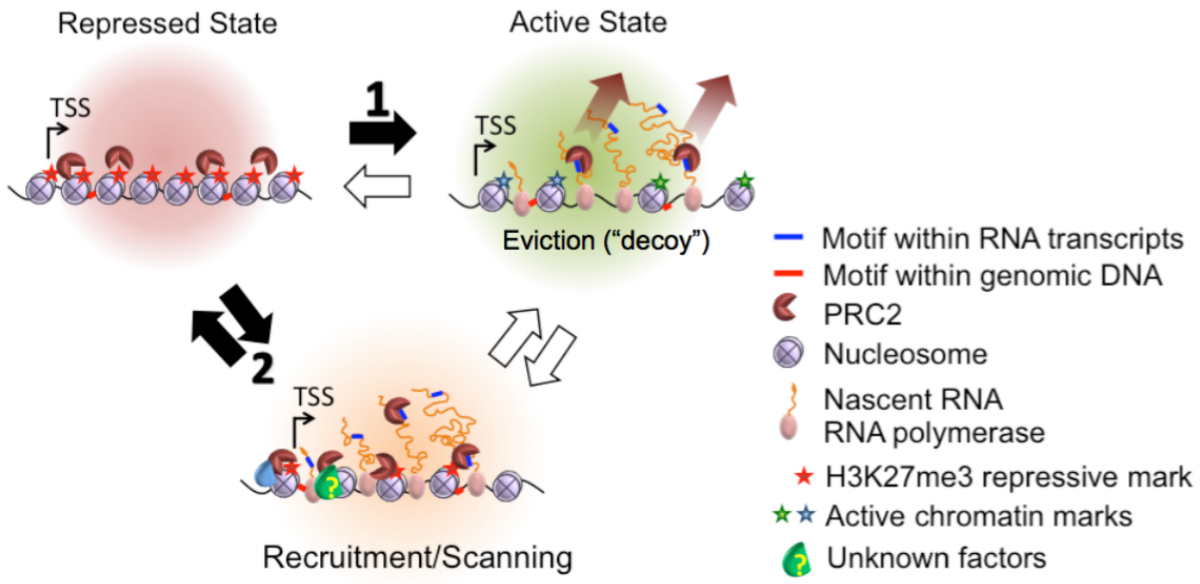


Figure 7 | Various models for RNA functionality in PRC2 recruitment and regulation in mammals.[89] RNA-functionality models for PRC2 that have been previously proposed include (1) eviction of PRC2 by RNA due to pioneer round of transcription following de-repression of a genomic loci, and (2) maintenance of epigenetic silencing marks (H3K27me3) by PRC2 recruitment via RNA to target genes. While these models require further research for validation, they need not be mutually exclusive.

Another "scanning" model had been proposed for PRC2 based on previous findings in the Cech laboratory on the promiscuous PRC2-RNA interaction.[30] In this model, promiscuous RNA-binding by PRC2 allows the complex to scan nascent RNAs more generally to locate possible sites of action. Interestingly, recent studies have reported that RNA inhibits the catalytic histone methyltransferase activity of PRC2.[21] Based on these studies, a "decoy" model has also been suggested, where RNA serve as a decoy to prevent

PRC2 from binding active genes and inhibits its action. Proposed models are illustrated and described in **Figure 7**.

Overall, whether the broad range of affinities for RNA-binding observed *in vitro* is at all physiologically relevant is not currently known. But the prevailing thought proposed for PRC2-RNA is that RNAs regulate gene expression both *in cis*[42,44,64,95] as well as *in trans*.[34,83] Perhaps the greatest obstacle in arriving towards a consensus on a recruitment mechanism is simply the great apparent complexity underlying PRC2 recruitment and function. The PRC2 field has expanded dramatically after lncRNAs have been recognized as important participants in PRC2 functions. Testing standing models would begin to illuminate detailed mechanistic features of the effect of RNAs on PRC2-chromatin interactions. And more broadly, unraveling these mechanisms would lead to new understanding of fundamental questions in biology, such as how gene expression is altered during cell differentiation.

1.5 — *Thesis Aims*

The recruitment of PRC2 to genes is thought of as an intricate and cell-type specific process that relies on a combination recruitment factors including: nucleic acids, accessory proteins, chromatin density, and histone modifications. At the onset of this project, there was little data that elucidated the nature of PRC2 interaction with RNA and chromatin, and how their relative affinities compared. Therefore, the principle aim here was to reveal the binding characteristics and relative affinities of the PRC2 complex for different biological substrates

in order to help build a simple mechanistic model for PRC2 recruitment that could inform standing *in vivo* observations.

One of the greatest obstacles in arriving towards a consensus on a recruitment mechanism is simply the great apparent complexity underlying PRC2 recruitment and function. *In vivo* studies, while powerful for placing observed experimental effects within an anatomical or physiological context, are often too obscure to deduce mechanistic details. In this respect, *in vitro* biochemical approaches to determining binding constants of PRC2–substrate interactions will inform the nature and extent of *in vivo* studies that may be required to assess potential interactions. Furthermore, *in vitro* studies are motivated by a reductionism that seeks to understand the whole by studying constituent elements of a mechanism. And the advantage of an *in vitro* system is that the complexity can be manipulated one variable at a time to resolve the intrinsic properties of a macromolecule. Therefore, this motivates *in vitro* approaches using recombinant purified PRC2 protein.

Given the recognized functional role of RNA in PRC2 biology, one goal was to define the RNA-binding motifs that might recruit PRC2. A two-hairpin motif found in XIST RNA has been suggested to be an RNA-binding motif for PRC2. However, in a study carried out by Davidovich *et al*., it was demonstrated that PRC2 binds natural RNAs promiscuously. This work hinted that there are undefined motifs in RNA that may direct PRC2 association. To probe the RNA motif requirements of PRC2 binding, *in vitro* binding using the electrophoretic mobility shift assay (**EMSA**) was used to determine the dissociation constant of different synthetic RNAs, and the effects of disrupting putative motifs was observed. Yet

another aspect of the PRC2 recruitment model is the affinity of PRC2 for nucleosomes. As of yet, PRC2 interaction with nucleosomes particles *in vitro* has remained unexplored. Therefore, characterizing PRC2 affinities to nucleosome substrates would be valuable to understanding PRC2 recruitment mechanisms. To accomplish this, nucleosome particles were assembled *in vitro* through salt gradient dialysis, a procedure that leads to the assembly of nucleosomes, and binding was tested through EMSA. Next, competition experiments between RNA and nucleosomes were pursued to reveal whether RNA has positive or negative effects on PRC2 association with nucleosomes. Lastly, PRC2 has been suggested to be recruited through DNA elements as well. In particular, the literature has hinted that CpG-islands, which characteristically feature 5-methyl-cytosine modifications and are often found in gene promoters, and can recruit PRC2. It has been speculated by some that PRC2 enrichment at CG-rich sites may occur through an unknown bridging protein. However, an alternative hypothesis that has remained untested is that PRC2 is able to directly read CpG sites. Therefore, another goal of this project was to determine the binding constants between PRC2 and CpG methylated DNA.

Thus, the aims of this project were to determine PRC2 binding constants with RNA, nucleosomes, and DNA. This of interest in part due to the value of PRC2 as a therapeutic target: It is known that mutations in PcG proteins have pleiotropic effects that can lead to the inadvertent activation of oncogenes and repression of tumor suppressor genes, key drivers of oncogenesis. Therefore, information gained from this study has the potential to contribute to on-going research in the development of cancer therapeutics. And collectively, findings from this

work will offer new insights into properties that are intrinsic to PRC2 as well as extrinsic factors that govern its recruitment to genomic loci.

Section 2 — PRC2 Targets Guanine-Rich RNA

2.1 — *RNA-based Recruitment of PRC2*

PRC2 has been defined as an RNA-binding protein complex.[30,43-45,92,94] Although, a specific motif in RNAs recruiting PRC2, as well as the functional relevance of RNA-binding, are still lacking substantial experimental support. Still, PRC2 has become a model system to investigate how RNA molecules recruit and regulate chromatin-modifiers.[46]

Earlier studies argue in favor of more specific RNA-binding to discrete RNA hairpin motifs. However, one recent study from the Cech lab has reported a physical association of PRC2 with thousands of both coding RNAs and ncRNAs *in vitro* and *in vivo*, indicating a possible promiscuity to RNA-binding by PRC2.[30] This view stands in stark contrast to other studies, where previous studies argue in favor of a more specific RNA binding to defined RNA hairpin motifs.[21,79,91] Attempts to decipher this apparent contradiction have been recently made, resulting in a model in which PRC2 binds RNAs with different affinities that range from low-to-mid nanomolar affinities *in vitro*.[30] Nevertheless, the question of RNA specificity still remains.

2.2 — *PRC2 Preferentially Associates with G-rich RNA in Vitro and in Vivo*

To quantitatively characterize the affinity of PRC2 for different substrates *in vitro*, nucleic acid-free human PRC2 was expressed and purified with five subunits (EZH2-SUZ12-EED-RBBP4-AEBP2) using the baculovirus system and a three-column purification scheme. Additional information on protein expression and purification is found in the methods section. **Figure 8A** shows a typical SDS-PAGE gel of recombinant protein stained with coomassie blue. Analysis by Fast Protein Liquid Chromatography (FPLC) shows that the PRC2 complex runs as a homogenous and mono-dispersed peak (**Figure 8B**), making the recombinant protein suitable for *in vitro* studies. Furthermore, the protein was shown to be active in methylating H3K27 (**Figure 8C**). The homogeneity of protein as well as particle integrity was also assessed by negative-stain single-particle election microscopy (**EM**), as presented in **Figure 8D & 8E**. From single-particle EM data, mostly intact particles of about ~17 microns in diameter were observed, which is consistent with monomeric protein complexes, as characterized in an EM study conducted by Ciferri and colleagues. A small percentage of the population consisted of small particles, which may be attributed to either perspectives of the complex oriented in a plane that is perpendicular to the carbon grid, or to the presence of sub-complexes. Notably, EM confirms that the protein preparation has no apparent protein aggregates and is otherwise homogenous.

In vitro binding of PRC2 to RNAs was measured by an electrophoretic mobility shift assay (**EMSA**). In brief, a gel shift assay is performed by incubating the purified protein with a P^{32}-labeled RNA fragment containing the putative protein binding site. The binding reactions are then resolved on a non-denaturing agarose gel. By knowing the amount of

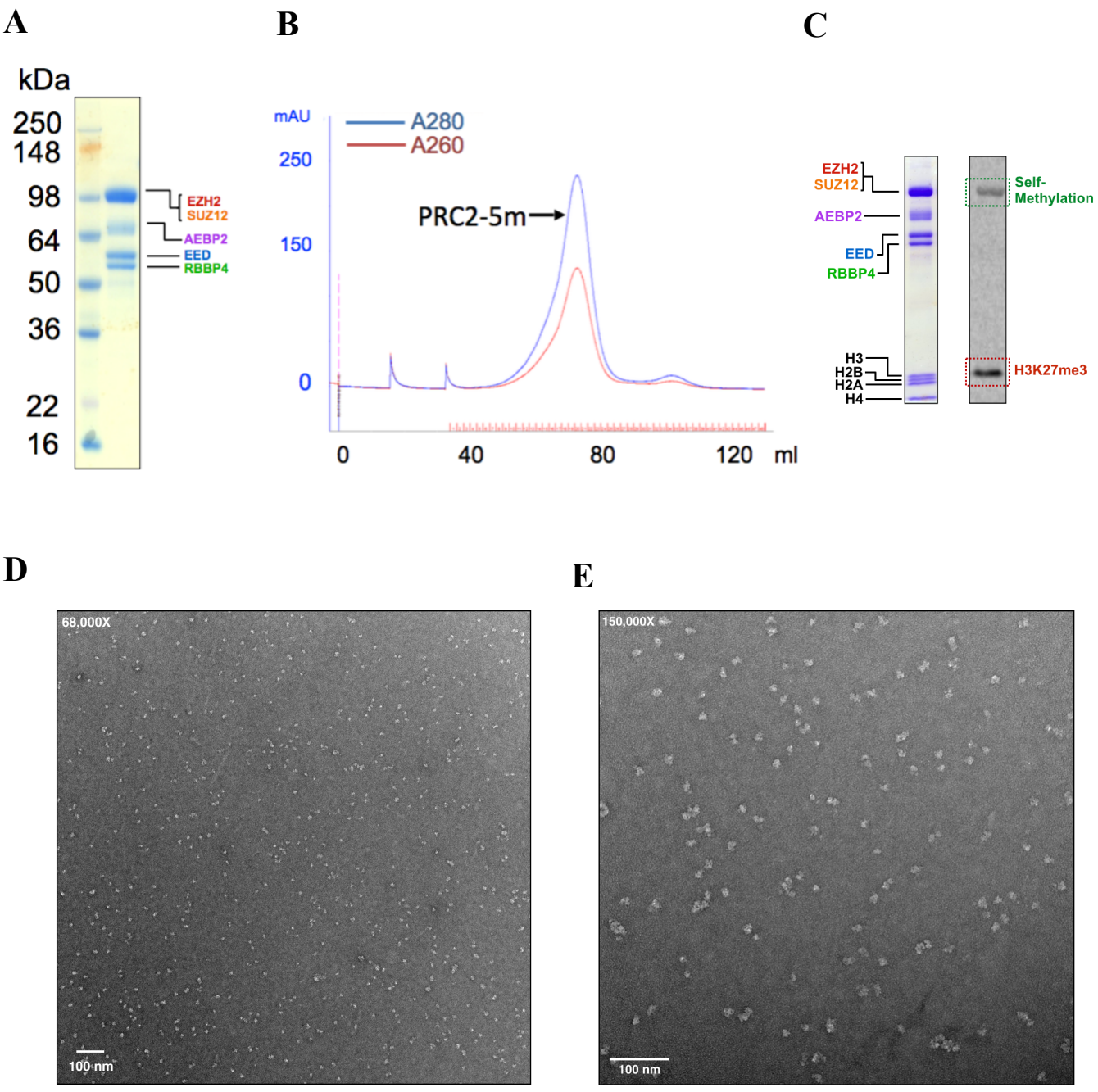


Figure 8 | Purified PRC2 5-mer complex exists as a stable complex *in vitro* and is homogenous and mono-dispersed.

A) Purity of a typical PRC2 complex assessed by SDS-PAGE and visualized by coomassie staining.[89]

B) PRC2 is mono-dispersed by size-exclusion chromatography.[89]

C) Histone methyltransferase assay shows that purified PRC2 is active in methylating H3K27.

D) Single particle negative-stain EM image of PRC2 deposited on a carbon grid. 68,000X magnification.

E) Single particle negative-stain EM image of PRC2 deposited on a carbon grid. 150,000X magnification.

protein titrated in the binding experiment, the apparent dissociation constant (K_d^{app}) of interaction can be calculated. Experimental details for PRC2-RNA binding experiments are provided in the methods section.

Prior to designing and testing different RNA sequences for their capacity to be bound by PRC2, several experimental considerations were taken into account. Davidovich *et al* (2013) have shown that the affinity of PRC2 to RNA is length-dependent. Thus, the greater the sequence length, the greater the affinity. And at a certain maximal RNA length (roughly ~200 nt), the dissociation constant (K_d^{app}) no longer decreases. For subsequent analysis, a "sweet-spot" RNA length was determined to be 40 nucleotides. At this length, quantitation by EMSA was reasonable, as binding was neither too tight nor too weak. Furthermore, 40 nucleotides is short enough that it is economical to synthesize. To be able to compare the binding of PRC2 to different sequences, all tested RNAs were therefore synthesized of a defined length of 40 nucleotides.

Given the promiscuous RNA-binding ability of PRC2, a first direction was to test whether PRC2 prefers binding to one of the four nucleotide-bases present within natural RNAs (Adenosine - A, Uracil - U, Guanosine - G, and Cytosine - C). Therefore four 40-mer homo-polymers were synthesized (A_{40}, U_{40}, G_{40}, C_{40}). EMSA gels and binding curves for PRC2 binding to the radio-labeled (P^{32}) homo-polymers are shown in **Figure 9A & B**. Surprisingly, the apparent dissociation constants (K_d^{app}) for these RNAs revealed that G_{40}, or "**poly (G)**", bound to PRC2 with the highest affinity at a K_d^{app} of 8.0 nM, while A_{40}, or "**poly (A)**" RNA showed very weak binding (K_d^{app} >10,000 nM), even at the maximum concentration point tested (5 uM of PRC2). Binding K_d values for the homo-polymer RNAs

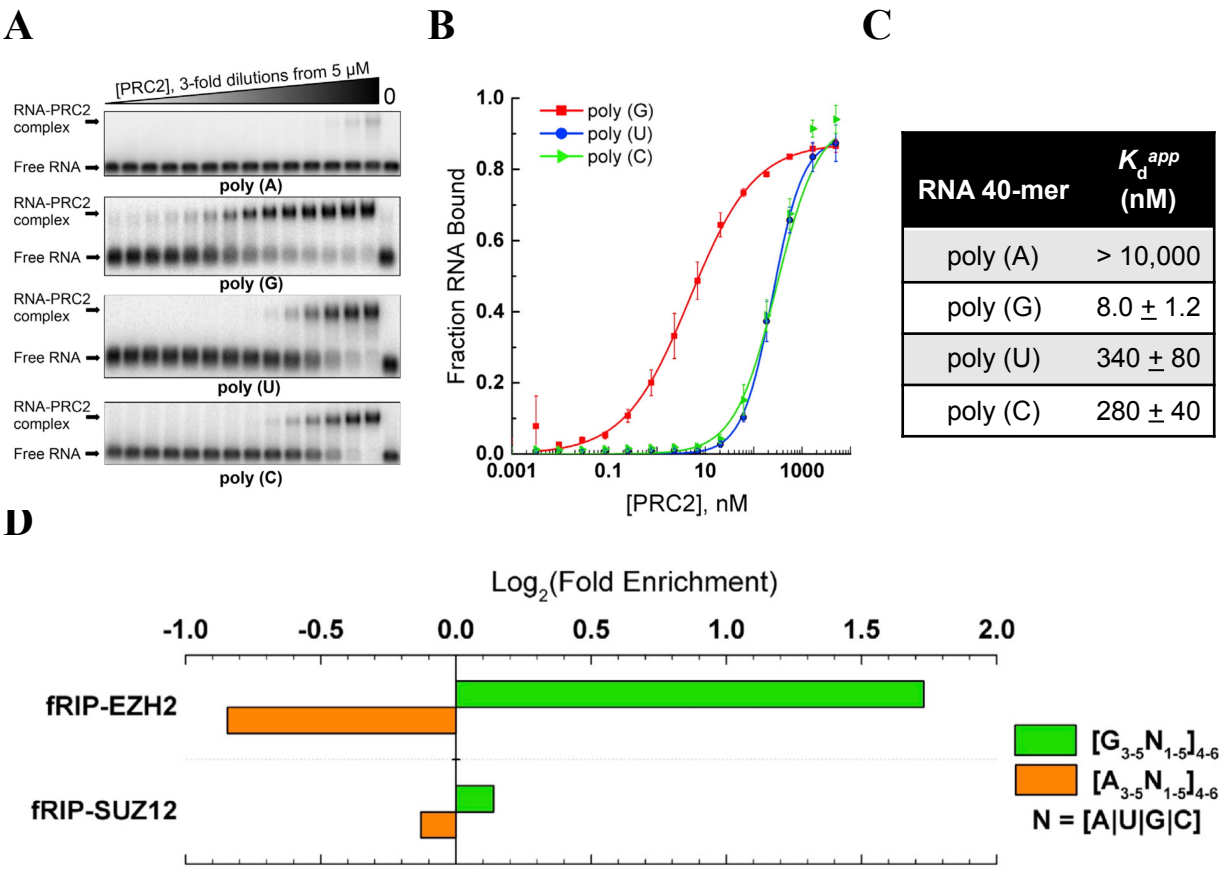


Figure 9 | PRC2 targets G-rich tracts *in vitro* and *in vivo*.
A) EMSA gels show PRC2 binding to poly(A), poly(G), poly(U), and poly(C) 40-mer homo-polymer RNAs.[89]
B) Binding curves for PRC2 binding to homo-polymers (results are n = 3).[89]
C) Table of K_d values for (A) (results are n = 3).[89]
D) Fold-enrichment of EZH2 and SUZ12 genome wide for G-rich and A-rich motifs using RNA immunoprecipitation-sequencing (RIP-seq) datasets.[89]
[Figure 9A-D are published in Wang *et al*., 2017. Gels in Figure 9A run by Karen Goodrich. Bioinformatics analysis in Figure 9D performed by collaborators in Davidovich lab.]

are listed in the table presented in **Figure 9C**. Quantitatively, a stunning >1000-fold difference in binding was observed between poly(G) and poly(A) RNAs. The "poly (U)" and "poly (C)" RNA sequences were also tested, and both had apparent K_d values of about 300 nM (**Figures 9C**), which are affinities that appear to be at an intermediate between those observed for poly (G) and poly (A) RNAs. Together, these results suggest that PRC2 prefers to bind guanine-rich RNAs *in vitro*.

From these findings, it was reasonable to suspect that perhaps G-tract motifs could guide the association between PRC2 and thousands of natural RNAs *in vivo.* One might infer that G-tracts are common enough in the transcriptome that this could perhaps be the underlying basis for promiscuous RNA-binding by PRC2. To address this, through collaboration with EMBL Australia group leader Dr. Chen Davidovich at Monash University, a bioinformatics study was carried out using sequencing data from RNA immuno-precipitation experiments using EZH2 and SUZ12 antibodies. And the association between PRC2 and G-tract-containing motifs of the sequence $[G_{3–5}N_{1–5}]_{4–6}$, where "G" symbolizes a guanine and "N" symbolizes any nucleotide, were quantified. Intriguingly, >9,000 association events between EZH2 and the $[G_{3–5}N_{1}–5]_{4–6}$ RNA motifs were identified in RIP-seq datasets published by Hendrickson *et al*. And statistical analysis revealed that this represented a significant association. Most strikingly, as a negative control for association, when the analysis was repeated where the guanine ("G") in the motif was replaced with an adenosine ("A") to yield the modified sequence motif, $[A_{3–5}N_{1–5}]_{4–6}$, the association between PRC2 and this A-tract motif was dramatically reduced. The fold enrichments of EZH2 for the

G-tract and A-tract motifs is portrayed in **Figure 9D**. Together, these data establish specificity for RNA-binding both *in vitro* and *in vivo*.

2.3 — *PRC2 Binds Tightly to G-quadruplex Forming RNAs*

In contrast to DNA, RNA has much greater structural and functional versatility. Indeed, the growing database of RNA structures has led to the characterization of numerous RNA secondary and tertiary structural motifs that span from basic hairpins to distinct folded structures as those seen in transfer RNAs. RNA is now known to be modular, with RNA polymers having the ability to fold into unique tertiary structures that in some cases eclipse proteins in terms of their structural complexity. In many ways, this has shifted the paradigm of the protein-nucleic acid recognition code. Indeed, it is now known that many proteins depend on the three-dimensional shapes of RNAs to promote recognition, as demonstrated by the tRNA-recognition mechanism of RNase P.[68]

One unique structural unit that can be adopted by RNA is the **G-quadruplex** ("**G4**"). A G-quadruplex is a structural conformation adopted by nucleic acids that is characterized by Hoogsteen base-paring.[24] And the formation of a very stable G-quartet arrangement via Hoogsteen base-pairs is essential to form G-quadruplexes, and has a characteristic planar array geometry, as shown by the cartoon representation in **Figure 10A**. In brief, simple G-quadruplex structures are assembled from a core of ≥ 2 π-π stacked G-quartets that are coordinated at their cores by essential alkali metal ions (namely $\mathbf{K^+}$). Moreover, the

intervening sequences between that run between G-quartets can vary in sequence identity and length, adding to the overall diversity that can be found in G-quadruplex structures.[24, 48]

Sequences with the potential to form G-quadruplex structures are broadly found in the genomes and transcriptomes of many organisms.[15] For example, the $(TTAGGG)_n$ repeat found in human telomeres (chromosome ends) are known to be able to form G-quadruplex structures.[24] From a therapeutics point-of-view, it is also interesting that sequences with the potential to form G-quadruplexes are also often found in promoter regions of various oncogenes. For instance, the NHE III(1) promoter region of the *c-MYC* oncogene can form G4s.[11]

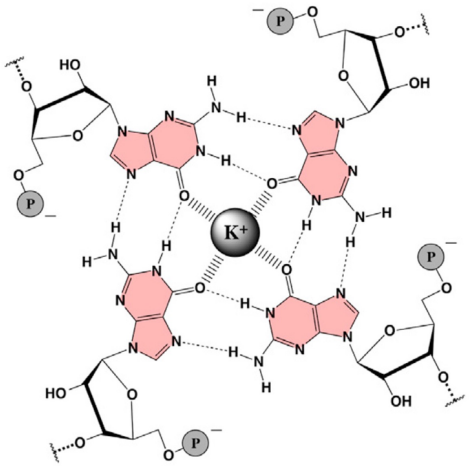
In addition to DNA G4s, RNA G4s have also been widely observed and are thought to play a role in translational regulation, among other functions. RNA sequences can form intra-molecular G4 structures,[14,24] which similarly to intramolecular DNA G-quadruplexes,[90] are also comprised of typical planar arrays of G-quartets. For example, the TERRA RNA transcribed from telomeric DNA repeats[3] can form a G4 RNA structure.[24] And recently, it was shown that RNA G-quadruplex structures play a role in various cellular functions including termination of transcription, regulation of telomerase activity, alternative splicing, and modulation of translation.[17] Lastly, it has been proposed that RNA quadruplex formation in triplet repeat disorders may be responsible for causing genetic instabilities. For example, the expansion of a $(CGG)_n$ repeat in the 5'-UTR of the FMR1 results in Fragile X Syndrome through an RNA-mediated mechanism.[23]

Given the propensity for PRC2 to bind G-rich RNA, it was reasoned that PRC2 may be able to recognize G-quadruplex motifs which are common in some G-rich sequences.

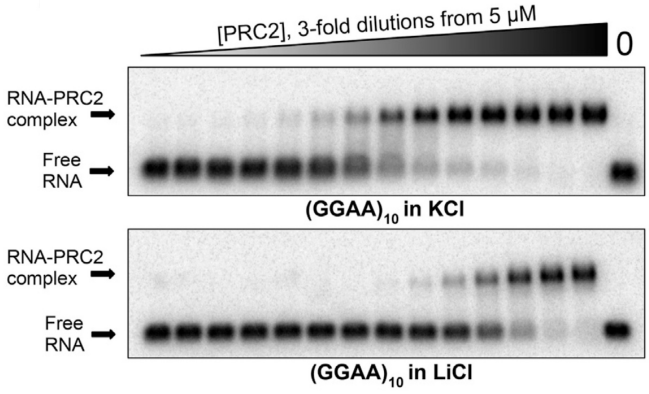
Therefore, binding of PRC2 to G-quadruplex-forming RNAs was tested and measured by EMSA. Several G4-forming RNAs were designed: "**TERRA40**" RNA is a G4 positive control based on a telomeric repeat sequence, and "**$(GGAA)_{10}$**" is another G4-forming RNA. As negative controls, a "**TERRA-mutant**" was designed that has mutations that disrupt G4-formation, along with another negative control for G4-formation, "**$(GA)_{20}$**." Notably, TERRA40/TERRA40-mutant and $(GGAA)_{10}$/$(GA)_{20}$ pairs have the same sequence identity, but differ only in the ability to form a G4 structure. It is also known that alkali metal ions can influence the propensity for G4s to fold, as G-quadruplex structure formation depends chiefly on K^+ ions to coordinate and stabilize G-tetrads. However, in the presence of Li^+ ions, these structures are not formed.[6] Therefore, K^+ ions help stabilize G4s, whereas Li^+ disrupts G4 formation.

This alkali metal ion-dependence for G4 formation was taken advantage of in order to test PRC2 binding specificity for G4 structures. As shown in **Figure 10B&C**, the K_d^{app} of PRC2 binding to the $(GGAA)_{10}$ G4-forming RNA in K^+ buffer was 7.7 nM. Strikingly, binding was attenuated significantly to 290 nM in Li^+ buffer. To ensure that the alkali metal ion-dependence came from the RNA rather than the protein, a control RNA $(GA)_{20}$ was tested. And as shown in **Figure 10C**, PRC2 binding to $(GA)_{20}$ did not exhibit an ion-dependence (390 nM in KCl; 320 nM in LiCl), which suggests that ion-dependence comes directly from the RNA sequence. Together, these results strongly indicate that PRC2 has a high affinity for RNA sequences with the potential to form G-quadruplex structures.

A



B



C

RNA 40-mer	G4-forming?	K_d^{app} in KCl	K_d^{app} in LiCl
$(GGAA)_{10}$	Yes	7.7 ± 2.4	290 ± 40
$(GA)_{20}$	No	390 ± 70	320 ± 13

Figure 10 | RNA G-quadruplex stabilization increases its affinity for PRC2
A) Depiction of a planar array of Hoogsteen-bonded guanines. Stacked G-quartets comprise a G4 RNA structure. Moreover, potassium ion coordinate between G-quartet layers and stabilize carbonyl oxygen atoms.[89]
B) EMSA shows $(GGAA)_{10}$ G4-forming RNA binding more tightly to PRC2 in G4-stabilizing conditions (KCl) than in G4-disrupting conditions (LiCl).[89]
C) Table of K_d values for $(GGAA)_{10}$ and $(GA)_{20}$ RNA. Results show substantial increase in binding in KCl relative to LiCl for $(GGAA)_{10}$ but not $(GA)_{20}$ (results are n = 3).[89]
[Figure 10A&B are published in Wang *et al*., 2017. Gels in Figure 10B run by Karen Goodrich.]

2.4 — *G-quadruplex Stabilizing Ligands Inhibit PRC2-RNA Interaction*

The shape of the G-quadruplex has been the subject of intense study and has led to the rational design of small molecules that can bind these unique structures. Accordingly, the use of G-quadruplex-binding ligands which can bind to the G-quartet, or to intervening regions have been well explored in cancer therapeutics. Perhaps one of the most well-studied and characterized G-quadruplex stabilizing ligands is TMPyP4, a cationic porphyrin analogue, which was first described as a G4-ligand capable of regulating telomerase activity by stabilizing G4 formation in telomeric DNA sequences.[33,61,69,72,77] Given the ability for many sequences to form G4 structures in the transcriptome, one might reason that G4-ligands may have some novelty and use in RNA-targeted therapeutics.[15,59] In fact, credence has been lent to an approach of targeting RNA G4 structures using small molecules by a study where the authors targeted an RNA G4 in the 5′-untranslated region (5'-UTR) of the small GTPase *NRAS* proto-oncogene.[14] This G4-stabilization via the small molecule has been shown to decrease translation efficiency of the NRAS mRNA transcript, and thereby decrease cellular protein levels. And a number of other studies have shown the potential use of G4-targeting ligands to alter the mRNA levels of genes that contain G4 motifs in their promoters, such as the proto-oncogene *KRAS*.[22] Therefore, the potential to transcriptionally regulate cancer genes via G4-forming sequences has sparked the development of a variety of G4-targeting small molecules.

Experiments in the prior section revealed that PRC2 has a binding preference for G4 structures. Yet it was unknown how a G4-stabilizing ligand might influence PRC2 binding. It was reasoned that a G4-ligand might be able to modulate the binding of PRC2 to RNA

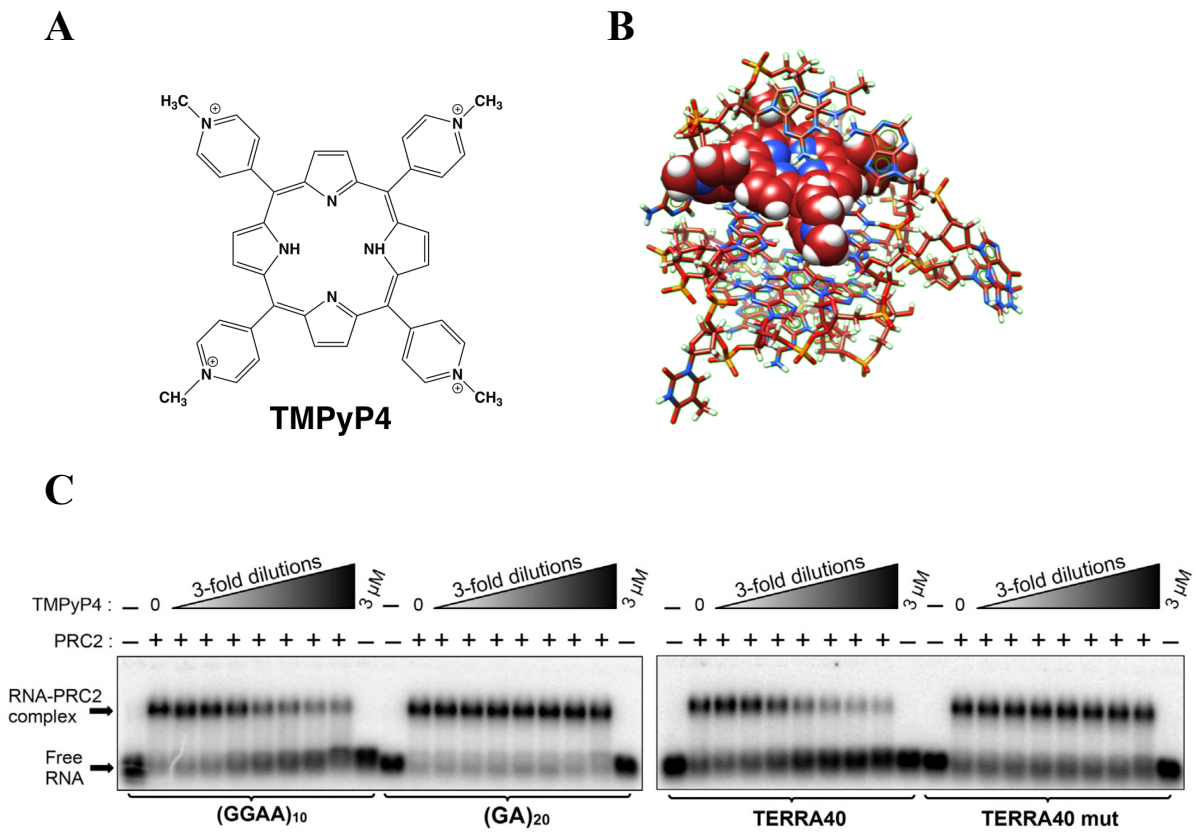


Figure 11 | The G4-stabilizing ligand TMPyP4 inhibits PRC2 binding to G4-forming RNAs

A) Chemical structure of meso-Tetra (N-methyl-4-pyridyl) porphine tetra tosylate (TMPyP4) G4-stabilizing small-molecule.

B) NMR structure of TMPyP4 bound to G-quadruplex DNA.[61]

C) Pre-formed G4 RNA-PRC2 complexes are disrupted by the TMPyP4 ligand, whereas non-G4 RNA-PRC2 complexes are unaffected.

[Figure 11C is published in Wang *et al.*, 2017]

positively or negatively. One hypothesis is that a G4-ligand might "lock" a particular RNA sequence into a G-quadruplex conformation by lowering the free energy associated with folding,[58] and therefore might stimulate association with PRC2. Alternatively, one could speculate that if the G4-ligand occludes a binding interface important for PRC2 recognition of an RNA, then this would in-turn decrease binding. To test the effects of G4-ligands, competition experiments were conducted between PRC2 bound to various RNAs and G4-ligands. RNAs tested include those capable of forming G4 structures, such as TERRA40 and $(GGAA)_{10}$, and as negative controls, non-G4-forming sequences TERRA40-mutant and $(GA)_{20}$ were used. Ligands tested include TMPyP4 (chemical structure provided in **Figure 11A**) and PDS (chemical structure shown in **Figure 12A**). The experimental results revealed that G4-ligands disrupt PRC2-G4 RNA interaction. TMPyP4 was able to compete off PRC2 association with the G4-forming RNAs $(GGAA)_{10}$ and TERRA40, but not the non-G4 RNAs $(GA)_{20}$ and TERRA40-mutant (**Figure 11C**). A similar trend was observed using the PDS G4-ligand (**Figure 12B**). Interestingly, a smaller concentration of the PDS ligand was required to compete off PRC2 from RNA than the TMPyP4 ligand, perhaps suggesting that PDS is able to bind the tested G4 RNAs more readily than TMPyP4.

To speculate on the mechanism of ligand-mediated RNA-binding interference, one can consider how TMPyP4 binds RNAs. NMR structural analysis of a G4 in complex with the TMPyP4 ligand, as illustrated in **Figure 11B**, shows that the ligand binds the face of the G4 through a hydrophobic base-stacking interaction. Perhaps then, PRC2 binding to G4 RNA relies on a similar binding mode. An alternative explanation for the binding interference is that PRC2 slightly unwinds or distorts G4s to achieve optimal binding, and that the G4 ligand

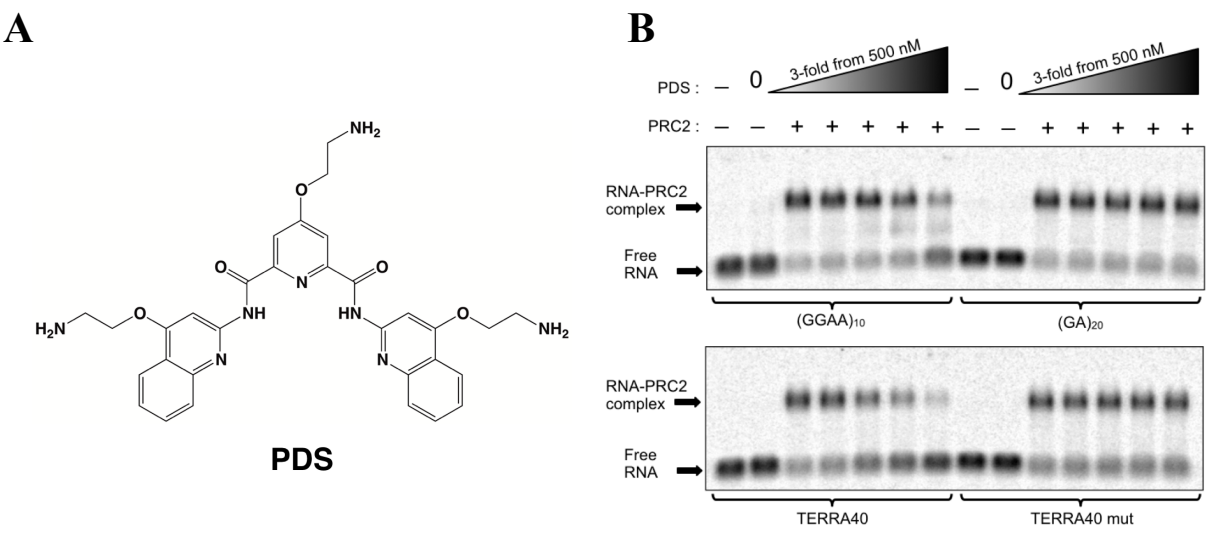


Figure 12 | The G4-stabilizing ligand PDS inhibits PRC2 binding to G4-forming RNAs
A) Chemical structure of pyridostatin (PDS) G4-stabilizing small-molecule.
B) Pre-formed G4 RNA-PRC2 complexes are disrupted by the PDS ligand, whereas non-G4 RNA complexes are unaffected.
[Figure 12B is published in Wang *et al*., 2017.]

might stabilize the G4 such that it can no longer be effectively distorted by PRC2. Taken together, these data demonstrate an approach to regulate PRC2 binding to RNA G-quadruplex structures using small molecules.

2.5 — *Materials and Methods*

Protein Expression and Purification: To purify PRC2, pFastBac1 expression constructs carrying gene sequences for Human EZH2, SUZ12, EED, RBBP4, and AEBP2 (UniProt: Q15910-2, Q15022, Q09028-1, O75530-1, and Q6ZN18-1) were used to generate

baculovirus stocks with the Bac-to-Bac Baculovirus Expression System (Life Technologies) according to manufacturer protocol. PRC2 subunits were expressed with N-terminal MBP tag fusions (harboring PreScission protease sites for down-stream removal). Appropriate amounts of baculovirus stock solutions carrying genes for the PRC2 subunits were used to infect a culture of Sf9 cells at a density of approximately ~2 million cells/mL in Sf-900 III SFM media. Infected insect cells were incubated for 72-h at 27°C in a rotary flask shaker at 130 rpm. Cells were harvested, and protein was purified as previously reported.[30] However, an additional Heparin column purification step was added to the purification scheme in order to extract cleaved MBP tags away from PRC2 complexes. To describe the purification procedure in brief, recombinant MBP-tagged PRC2 complex in the protein lysate was bound to amylose resin and washed extensively. MBP-tagged PRC2 was then eluted with 10 mM maltose. The eluate was concentrated to ~15 mg/ml using Amicon spin columns (30 kDa MWCO). The MBP tags fused to the PRC2 subunits were cleaved by digesting with PreScission protease at a mass ratio of 1:50 protease:protein. Following a 16 to 20-h incubation at 4°C, the efficiency of PreScission protease cleavage was assessed by running the protein on SDS-PAGE. Next, the protein was injected into a 5-mL Hi-Trap Heparin column and run with a gradient over 35 column volumes from Buffer A (10 mM Tris-pH 7.5 at RT, 150 mM NaCl, and 1 mM TCEP) to Buffer B (10 mM Tris-pH 7.5 at RT, 2 M NaCl, and 1 mM TCEP) with a flow rate of 1.5 ml/min. The fractions containing PRC2 were identified using SDS-PAGE, pooled, and concentrated with Amicon spin columns to ~15 mg/ml. The concentrated protein was then injected into a HiPrep 16/60 Sephacryl S-400 HR column with running buffer (250 mM NaCl, 10 mM Tris-HCl, pH 7.5 at RT, 1 mM TCEP-pH

7) at a flow rate of 0.5 ml/min. As before, PRC2 fractions were identified using SDS-PAGE, pooled, and concentrated. The final protein concentration was determined by absorbance at 280 nm. Finally, protein preparations were aliquoted, flash-frozen using liquid nitrogen, and stored at -80^{o}C until required for further use.

Negative-Stain Electron Microscopy: Negatively-stained protein samples were prepared as described.[21] In brief, a ~1:5000 dilution of a 20 μM aliquot of purified recombinant PRC2 was deposited onto a carbon grid that had been plasma cleaned in a 75% Ar/25% O_2 atmosphere for 20 seconds in a EMITEC glow discharge cleaning system. After 1 minute of incubation on the carbon grid at room temperature, the sample was negatively-stained with a solution of 2% uranyl formate (depleted uranium) and blotted dry. Raw images were acquired at a nominal magnification of 68,000× and 150,000× on a 1,024 × 1,024 pixel (1.52 Å/pixel at the detector-level) charge-coupled device (CCD) camera, using a FEI Tecnai F30 electron microscope operating at 300 keV.

In Vitro Histone Methyltransferase (HMTase) assay: The enzymatic activity of PRC2 complex was assessed through a histone methyltransferase (HMT) assay. To perform the assay, a 10-uL reaction was prepared which included 0.1 mg/ml nucleosomes, 0.4 μCi S-[methyl-3H]adenosyl methionine, 50 mM Tris-HCl pH 7.5, 100 mM KCl, 5 mM $MgCl_2$, 0.1 mM $CaCl_2$, 0.5 mM $ZnCl_2$, 2 mM 2-mercaptoethanol (BME), and 500 nM PRC2. Samples were then incubated for 30 minutes at 30°C and reactions were quenched by the addition of 10-μl of 2X stop buffer. Samples were then run on a native PAGE gel, dried for 1-h using a

vacuum pump, and exposed to a phosphor imaging plate. Following about a one-week exposure, the radioactive signal was scanned using a Typhoon Trio phosphorimager (GE Healthcare).

Electrophoretic Mobility Shift Assay (EMSA) for PRC2-RNA Binding: All 40-mer RNAs were synthesized by GE-Dharmacon. RNA was radiolabeled using [γ-^{32}P]-ATP as previously described by Davidovich *et al*., (2013), except RNA was dissolved in TE pH 7.5 instead of water. The counts of purified RNAs were determined by liquid scintillation counting. Radiolabeled RNA was adjusted to a volume of 60-µl with TE pH 7.5. The diluted RNA was heated for 10 minutes at 95°C to ensure melting of possible folded secondary and tertiary structures and snap-cooled on ice for ≥2 minutes. RNA was then incubated for 30 min at 37°C in RNA-binding buffer (50 mM Tris-HCl pH 7.5, 100 mM KCl, 2.5 mM $MgCl_2$, 0.1 mM $ZnCl_2$, 2 mM 2-mercaptoethanol, 0.1 mg/ml bovine serum albumin, 0.1 mg/ml fragmented yeast tRNA, 5% v/v glycerol). Next, stock PRC2 protein was diluted with RNA-binding buffer, added to RNA, and allowed to bind the RNA for 30 minutes at 30°C. Samples were then promptly loaded on a non-denaturing 0.7% agarose gel (SeaKem® GTG® Agarose) buffered with 1X TBE at 4°C. Gel electrophoresis was carried out for 90 min at 66 V in a 4°C cold room. Gels were then removed and vacuum dried for 80 minutes at 80°C on a Hybond N+ membrane and two sheets of Whatman 3 mm chromatography paper. Dried gels were exposed to phosphor imaging plates and signal acquisition was performed using a Typhoon Trio phosphorimager (GE Healthcare). For PRC2-RNA binding experiments in

LiCl, all procedures were identical, except 100 mM KCl was substituted with 100 mM LiCl in the RNA-binding buffer.

G4-ligand Competition Assay: Competition reactions were prepared identically to PRC2-RNA binding reactions (see above), except a constant concentration of PRC2 was used and variable concentrations of competitor G4-ligands (TMPyP4 or PDS) were used. The amount of protein used in each assay was determined based on the concentration of protein giving 80% maximal binding of the labeled RNA in the absence of any competitor G4-ligands.

Quantification of Genome- and Transcriptome-wide Association between Proteins and Putative Motifs: (methods from Davidovich lab[89]) Binding sites of proteins on RNA *in vivo* were identified from published RIP-seq data.[37] Short reads were downloaded from the NCBI Gene Expression Omnibus (GEO) repository and converted to fastq files using the SRA Toolkit. Data were mapped to the reference genome (hg19) using Bowtie2 with the option -q. In order to identify protein binding sites (PBS), output sam files were converted to bam files (bedtools[65]) and peaks were called using MAC2,[95] with the option –broad and the default false discovery rate (FDR) of 0.05, against experimental input.

For each putative binding motif that was tested, the start and end positions were identified within the human genome (hg19) on either the positive or negative strands and assigned as individual features into a single bed file (motif_search.pl, GitHub https://github.com/cdavidov/motif_search_v1-8). Overlapped motifs were merged by bedtools

merge[66] using the -s option. Next, the number of PBS-motif association events were determined by counting overlaps between PBS and motifs using bedtools closest[66] with the option -d. Since the number of PBS-motif association events would be expected to vary with dependence on the number and size of motifs and the number and size of PBS, the number of these association events that could be expected to be obtained by chance were empirically quantified. To this end, motifs were randomly shuffled using bedtools shuffle[66] under the default settings and the same reference genome, as used for the procedures above, and the number of PBS-motif association events was quantified again, as above. This process was repeated 1000 times, in order to empirically determine the probability density function of the number of PBS-motifs association events that were obtained by chance. The empirically obtained probability density function was then modeled to a Poisson distribution. And the modeled Poisson probability density function was used to calculate the probability of obtaining the number of PBS-motifs association events, as observed without shuffling the motif. Fold Enrichment (FE) was calculated by dividing the number of PBS-motif association events by the mean obtained from the modeled Poisson probability density function.

Section 3 — Nucleosome-Free DNA Dictate PRC2 Binding *In Vitro*

3.1 — *Histone-based Recruitment of PRC2*

A hallmark of eukaryotic genomes is the organization of about 2-meters total of DNA into nucleosomes, and the folding of nucleosome arrays to enable packaging within a cell nucleus that is only about ~10 microns in diameter.[8] The fundamental structural unit of DNA packaging is the nucleosome, which consists of an histone octamer core (two molecules each of four core histones proteins: H2A, H2B, H3, and H4), around which 147 base-pairs of DNA are wrapped in ~1.7 superhelical turns.[2,54] This association between DNA and histones packages and condenses the DNA.

Over the last few years, technological advancements, including the development of "chromosome conformation capture" assays such as Hi-C,[5] have revealed that the chromatin packaging in the nucleus is hierarchically structured into dynamically regulated subnuclear domains, as opposed to stochastic organization. Accordingly, histone modifications and chromatin density have been implicated in mediating recruitment of chromatin-modifiers to their target sites of action.

The literature proposes a view where nucleosomes can serve as dynamic recruitment factors for PRC2 proteins, through a combination of lysine tail PTMs and local nucleosome density.[20,25,39,41] It is already known that different histone marks can either stimulate or dampen PRC2 catalytic activity, as discussed in Section 1. And experiments have shown that these histone marks can affect PRC2 binding.[84] For instance, the affinity pull-down from

Drosophila embryo nuclear extracts using H2A nucleosome arrays modified with ubiquitin retrieved several drosophila PRC2 subunits, including JARID2 and AEBP2.[41] In another independent study, a link between H3K9me3 repressive marks and PRC2 was established using tethering assays performed in embryonic stem cells (ESCs). From these tethering experiments, it was shown that a H3K9 methyltransferase can recruit PRC2 to target genes.[1,60] And that both methyltransferases co-occupy a set of genes in ESCs. Furthermore, the observation that EED can also bind methylated H3K9 nucleosomes suggests that H3K9me3 marks might also serve as a recruitment site for the PRC2 complex.[60] In addition to being methylated, H3K27 can also be acetylated by the acetyltransferases p300/CBP.[16] Acetylation and methylation at the same residue are known to be mutually exclusive and correspond to opposite transcriptional states.[36] Some literature suggests that genome regions enriched in H3K27ac active marks are non-permissive for PRC2 occupancy.[84] And that a prior deacetylation by histone deacetylases (HDACs) is necessary for PRC2 recruitment to these genomic loci in ESCs.[84] These findings suggest a histone-based model in which the transition between the active (H3K27ac) and repressed (H3K27me3) states is mediated by the activities of H3K27 erasers (HDACs) and writers (PRC2, p300/CBP).

It has also been reported that PRC2 activity and occupancy is regulated by the density of its substrate nucleosome arrays, and that neighboring nucleosomes activate the PRC2 complex with a portion of H3 histones.[55] Perhaps nucleosome density serves as a proxy for transcriptional state. Thus it is possible that PRC2 can sense the chromatin environment via nucleosome architecture to exert its role in the maintenance of transcriptional states.

3.2 — *PRC2 Binding to Nucleosomes*

The affinity of PRC2 for nucleosome substrates with different numbers of nucleosome repeats is of interest. Provided that it is thought that nucleosome substrates can regulate PRC2 activity, one imagines that they must regulate PRC2 binding events as well. Therefore, there is a unique opportunity to study PRC2 binding affinity to different nucleosomal substrates. For *in vitro* analysis, nucleosome particles can be assembled by a method called **salt gradient dialysis**,[63] wherein a mixture of histone octamers and DNA are slowly dialyzed from high-to-low salt buffer. This slow dialysis-based mixing protocol is necessary to prevent rapid aggregation and precipitation of histone protein, which is typically observed from the direct mixing of DNA and histone octamer in low salt buffer. Structurally, the nucleosome core particles formed *in vitro* are identical to nucleosomes formed *in vivo*, and this method has been historically employed in structure-function studies of nucleosomes.

To prepare nucleosome substrates, first DNA templates containing "**Widom 601**" nucleosome-positioning sequences were prepared. The Widom clone-601 nucleosome positioning sequence[53] is a 147 base-pair nucleic acid sequence identified through *in vitro* selection experiments (SELEX), and was shown to promote nucleosome assembly at the Widom 601 position. In total, 3 linear DNA templates containing 1-, 3-, and 12-, copies of the Widom 601 sequence were prepared by digesting from plasmid DNA and purified to homogeneity via column chromatography. An example trace chromatogram showing the purification of the DNA template from its vector backbone is shown in **Figure 13A**. Only fractions containing pure linker DNA fragments of the expected size were pooled and concentrated for later use in nucleosome reconstitution experiments. Additional experimental

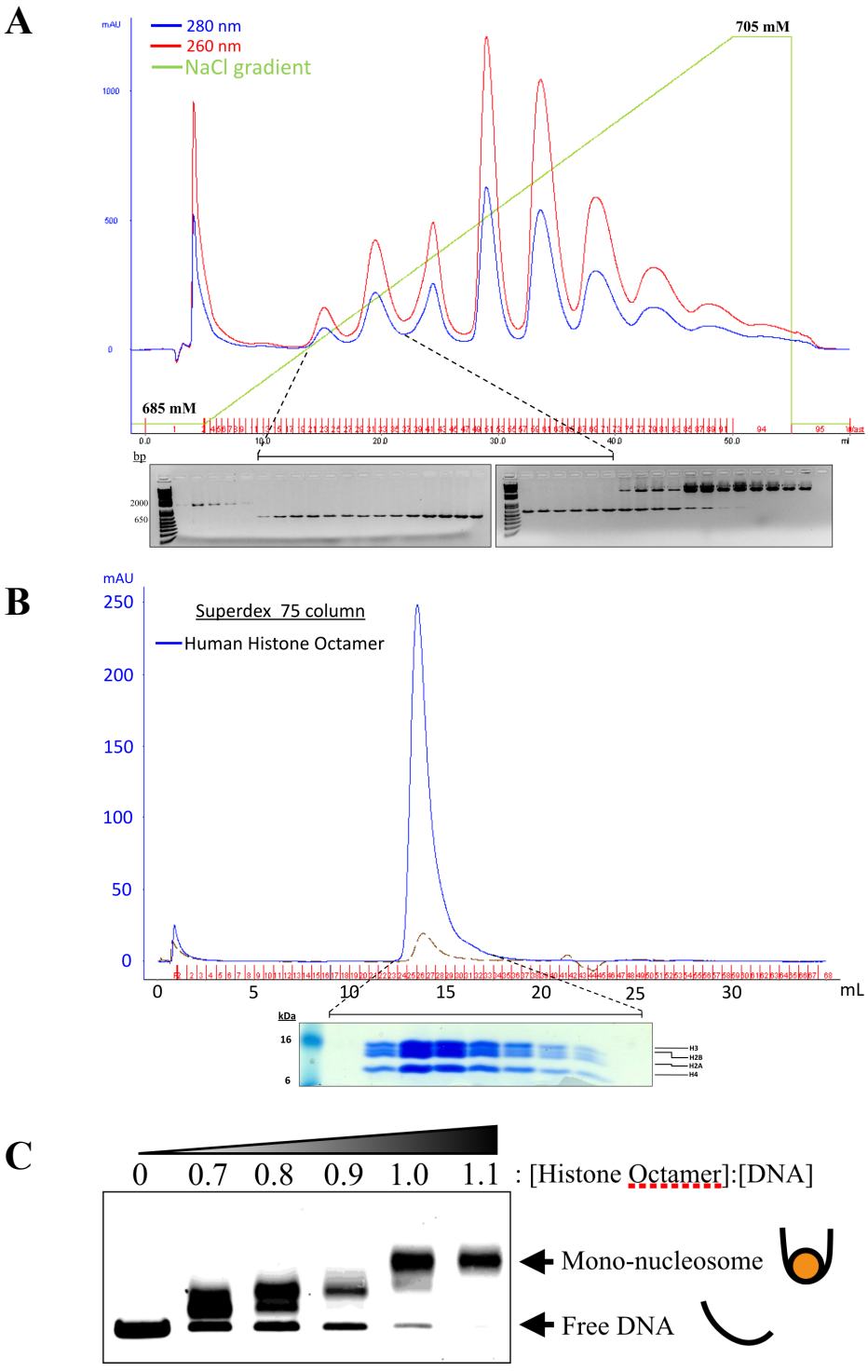


Figure 13 | Reconstitution of Nucleosomes *In Vitro* by Salt-Gradient Dialysis
A) A typical chromatograph showing the fractionation of DNA fragments.
B) An FPLC trace profile of recombinant histone octamer showing that the octameric protein runs as a homogenous, mono-dispersed peak.
C) Agarose gel electrophoresis experiment showing the formation of mono-nucleosomes by salt-gradient dialysis as the ratio of octamer to DNA is steadily increased.

details on DNA preparations are provided in the methods section.

The second component needed for assembling nucleosomes are histone proteins. Each individual histone (H2A, H2B, H3, and H4) was expressed and purified from bacteria, and assembled into octamer complexes. Then, octamers were purified by FPLC, as described in the methods. A sample chromatogram representing the homogeneity of protein is shown in **Figure 13B**. Only fractions of equal intensity of H2A, H2B, H3, and H4 were pooled and concentrated.

To reconstitute nucleosomes, purified histone octamers and DNA were mixed at varying ratios and subjected to salt gradient dialysis over a 24-hour period. A typical gel electrophoresis experiment showing the step-wise assembly of a mono-nucleosome is shown in **Figure 13C**. Notably, complete nucleosome assembly is achieved at an octamer:DNA ratio of 1.1, as indicated by the complete super-shift of the template "Free DNA" band.

In this work, nucleosome particles and arrays were analyzed using a combination of biochemical techniques such as EMSA and single-molecules techniques such as **Atomic Force Microscopy** (**AFM**). AFM has been frequently applied to study conventional nucleosome arrays,[52,74,85] and one goal is to extend this type of analysis to assess the quality *in vitro* reconstituted nucleosomes. In brief, AFM is able to "image" individual molecules by monitoring the flexing of a micro-machined silicon nitride cantilever as it rasters over a surface, contacting individual molecules deposited on a surface. For biological samples, mica silicate sheet mineral is typically used as a surface, as cleaving a single sheet of mica leaves an atomically-flat surface which yields the best signal-to-noise for imaging single biological molecules via AFM. The deflection of the AFM tip as it responds to surface features is

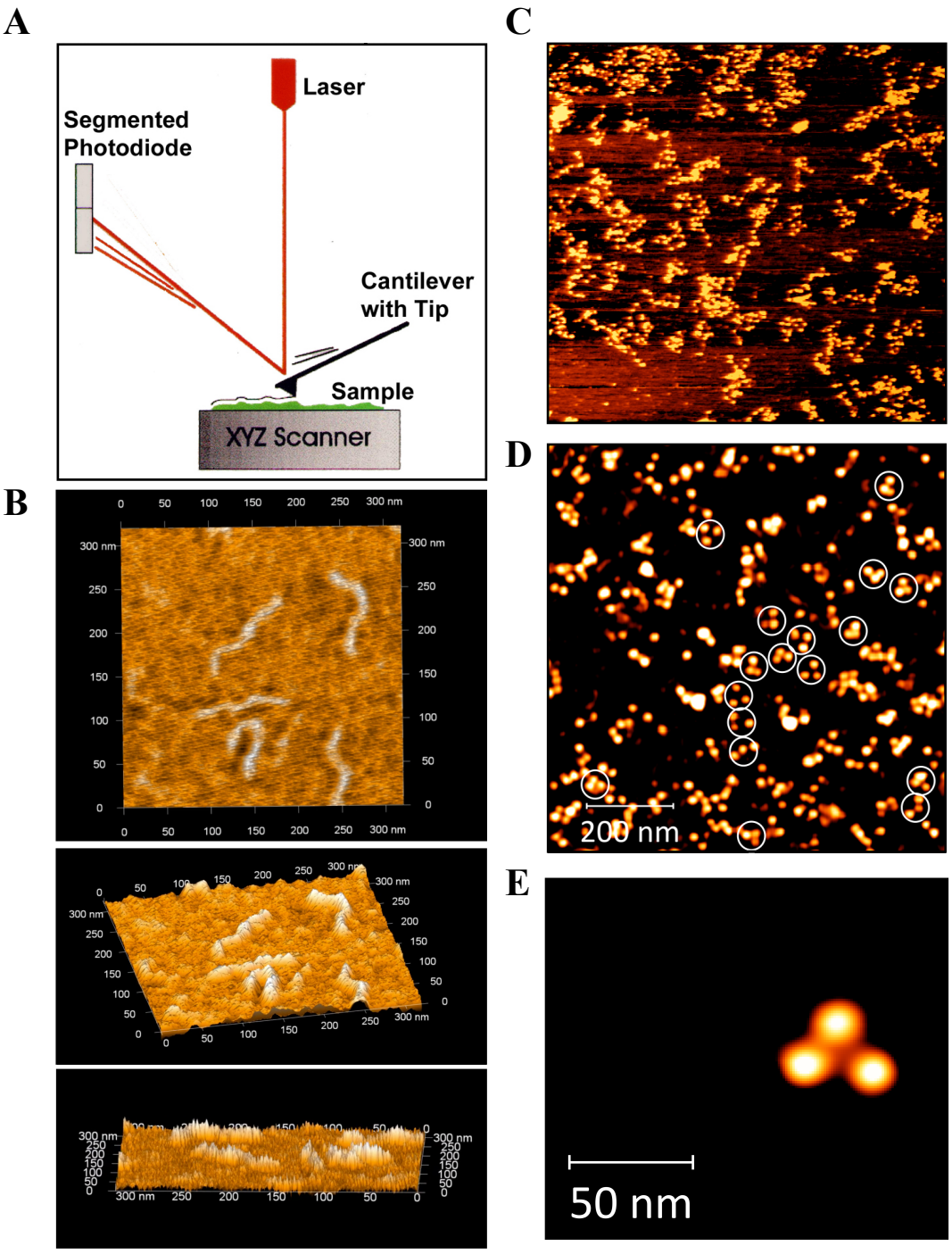
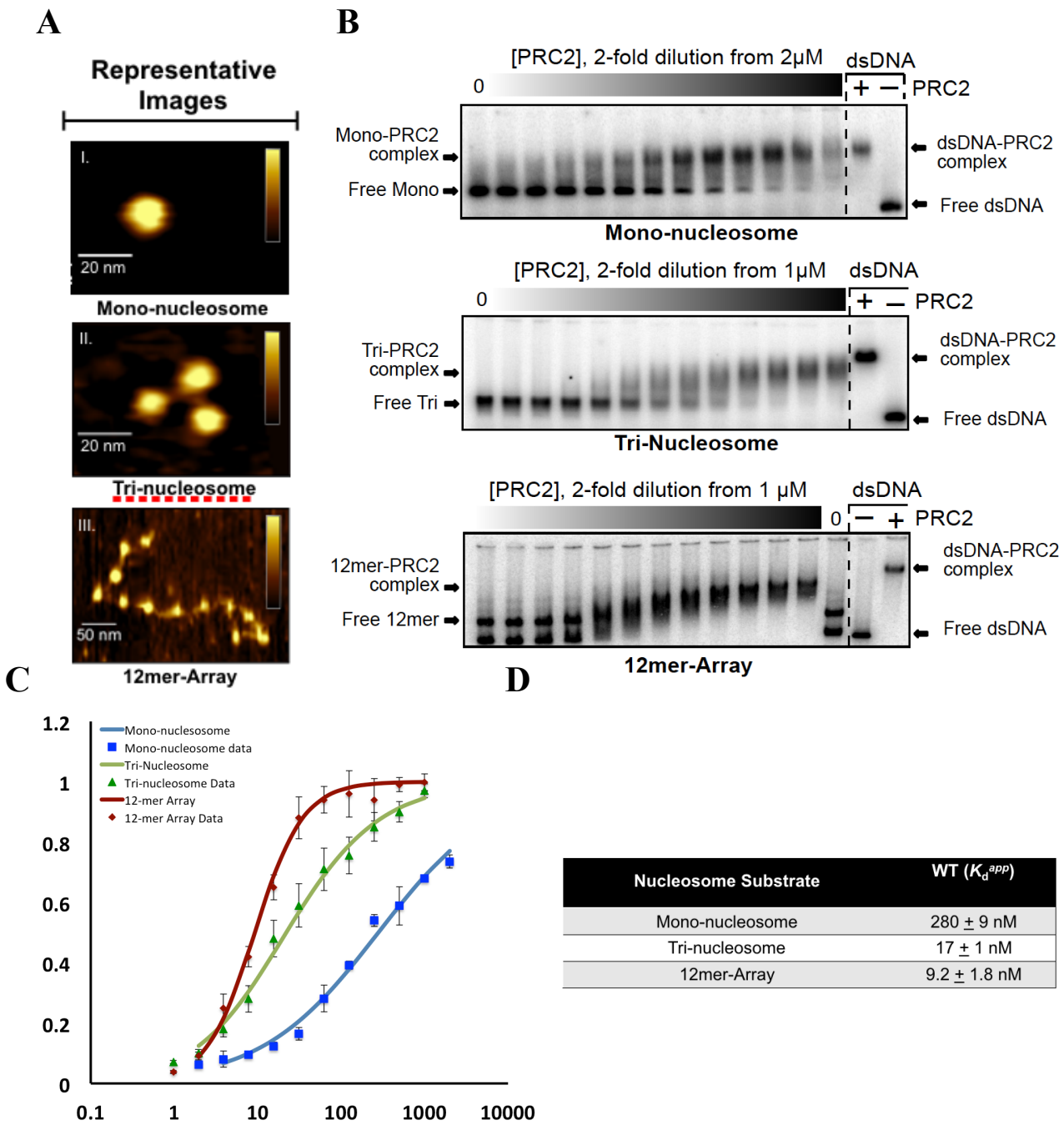


Figure 14 | Atomic-Force Microscopy imaging of biological samples
A) Schematic of typical AFM setup. [Figure from physics.UCSB]
B) AFM images of individual 400 base-pair DNA molecules on a freshly-cleaved mica surface. 3D data presented at different orientations shown to emphasize height information.
C) Large image field of nucleosome arrays on APTES-treated surface.
D) Large image field of tri-nucleosomes particles on APTES-treated surface.
E) Zoomed-in image of tri-nucleosome particle.

registered by reflected laser light and converted to a digital image map of molecule locations on the surface, plus their height profiles. These height data profiles are like mountains that reveal features of single molecules. A schematic of a basic AFM set-up is illustrated in **Figure 14A**.

Here, AFM is used to directly image single DNAs, and nucleosome particles in air and liquid. And AFM provides a powerful tool for obtaining unique insights into the basic biophysical properties of DNA polymers and nucleosomes, as well as the functional processes in which they are implicated. For surface preparation, one of two approaches were taken. One method of surface preparation is APTES-functionalization of the mica surface. This type of APTES monolayer has been demonstrated to capture molecules on the surface in a 2D configuration that is the projection of their 3D solution configuration.[78] This adsorption protocol has two main advantages: First, it preserves the structure of the molecule as it was in solution, and second it does not require the use of a cross-linking agents such as glutaraldehyde. A second approach to surface preparation is through divalent cation tethering of the biomolecule to the negatively-charged mica surface. Typical AFM images obtained for 400 base-pair long DNA polymers are presented in **Figure 14B**. Note, the different orientations of the 3D data illustrate the height information acquired through AFM. Sample images of reconstituted nucleosome arrays and particles are shown in **Figures 14C, D & E.**

To test the effects of the number of nucleosome repeats on PRC2 binding affinity, EMSA experiments were performed using mono-, tri-, and 12-, nucleosome particles. Representative AFM images of these nucleosome substrates are shown in **Figure 15A**. Representative gels of EMSA binding experiments using mono-, tri-, and 12mer radioactive-



Nucleosome Substrate	WT (K_d^{app})
Mono-nucleosome	280 ± 9 nM
Tri-nucleosome	17 ± 1 nM
12mer-Array	9.2 ± 1.8 nM

Figure 15 | PRC2 affinity for nucleosomes increases with increasing tandem nucleosome repeat length.

A) Representative AFM images of mono-, tri-, and 12mer-nucleosome substrates. Height ranges from 0 to 3 nM.

B) EMSA gels of PRC2 binding to radiolabeled nucleosomes.

C) Binding curves for the association of PRC2 with nucleosomes

D) Table of K_d of interaction of PRC2 with nucleosome substrates

labeled nucleosome substrates are shown in **Figure 15B**. The K_d^{app} of PRC2 association with mono-nucleosomes is 280 nM. This compares with a K_d of 17 nM for binding to tri-nucleosomes, and a K_d of 9.2 nM for 12mer-nucleosome arrays (Binding curves shown in **Figure 15C**; Table of affinities shown in **Figure 15D**). Binding results reveal that PRC2 exhibits a higher affinity towards nucleosome particle with greater numbers of tandem nucleosome repeats. This perhaps suggests that PRC2 binding to nucleosomes occurs in a cooperative manner.

3.3 — *PRC2 Interacts with Linker-Regions of Nucleosomes*

How then is the PRC2 complex recruited to most, if not all, promoters in the human genome to catalyze histone tri-methylation? One might predict that universal chromatin architecture genome-wide should play a significant role. A feature of chromatin architecture that could affect PRC2 recruitment and occupancy at genomic loci are nucleosome-free regions (**NFRs**). Previous molecular genetics studies have demonstrated that PRC2 populates active promoter regions.[30,43] Given that promoter regions are traditionally viewed as nucleosome-free, the observation that PRC2 is enriched at active promoter is somewhat problematic for a strict histone-based model of recruitment. However, the hypothesis remains that PRC2 is able to sense regions in the genome depleted of nucleosomes. Such a mechanism could indicate that PRC2 is in fact a sensor of transcriptional activity. Yet, thus far there is very little evidence to suggest that NFRs can directly drive the recruitment of PRC2.

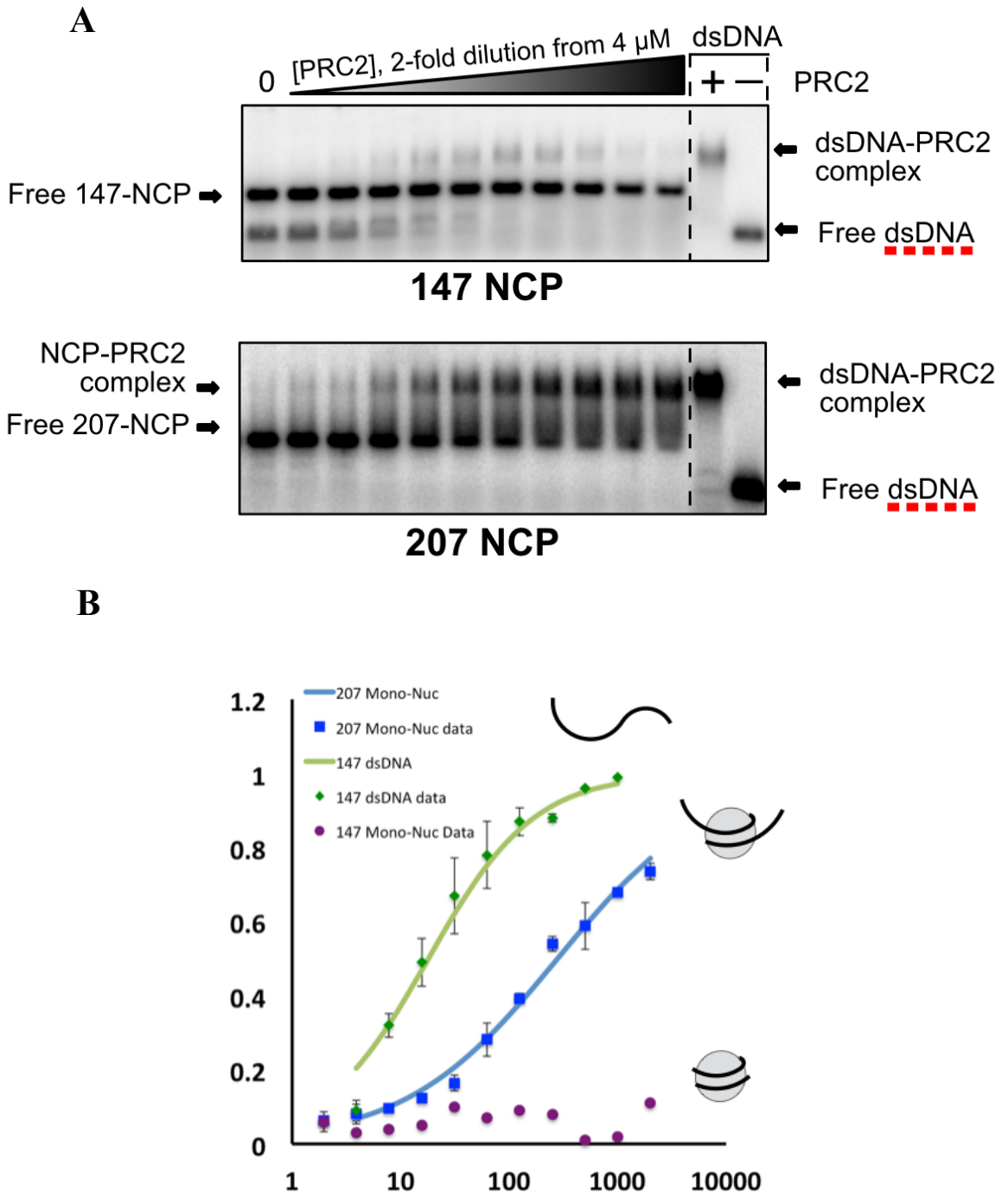


Figure 16 | PRC2 Binding to nucleosomes requires DNA linker regions
A) Representative EMSA gels of PRC2 binding to 147-Nucleosomes lacking a DNA linker, and 207-Nucleosomes containing linker DNA.
B) Binding curves for PRC2 binding to nucleosome substrates (results are n=3).

To test the hypothesis that nucleosome-free linker DNA is important for PRC2 binding to nucleosomes, several mono-nucleosome substrates were reconstituted *in vitro* harboring varying lengths of linker DNA. Specifically, substrates tested include a 147-bp "core" nucleosomes which contains no linker regions and a 207-bp nucleosome which contains a sum of 60-bp in linker DNA. EMSA binding experiments performing using PRC2 and the mono-nucleosome variants revealed, quite stunningly, that a 147-core nucleosome particle attenuated PRC2 binding (**Figure 16A and B**). The 207-nucleosomes increased PRC2-binding dramatically. These *in vitro* results suggest that linker DNA may in fact dictate PRC2 targeting to nucleosomes and challenges several prevailing histone-based models of PRC2 recruitment to genomic loci.

In order to test PRC2 binding to linker-DNA regions of nucleosomes using an orthogonal experimental method, a micrococcal nuclease-protection assay was designed. In brief, micrococcal nuclease (**MNase**) is a unique nuclease with the ability to induce double-stranded breaks within nucleosome linker regions, but only single-stranded nicks within the nucleosome itself. Because of this property, MNase can be used to determine the approximate positions of nucleosomes in a region of DNA, given that the nucleosomes are consistently positioned in the sample. Thus, the nucleosome structure protects DNA from MNase-mediated digestion and leaves a footprint on DNA that marks the position of nucleosomes. MNase-treatment of isolated chromatin, followed by DNA gel electrophoresis typically yields a "ladder" corresponding to histone-protected DNA fragments. It was reasoned that if PRC2 indeed binds linker-DNA regions of nucleosome-arrays, then PRC2 should be able to protect against MNase-mediated digestion of linker DNA. This would

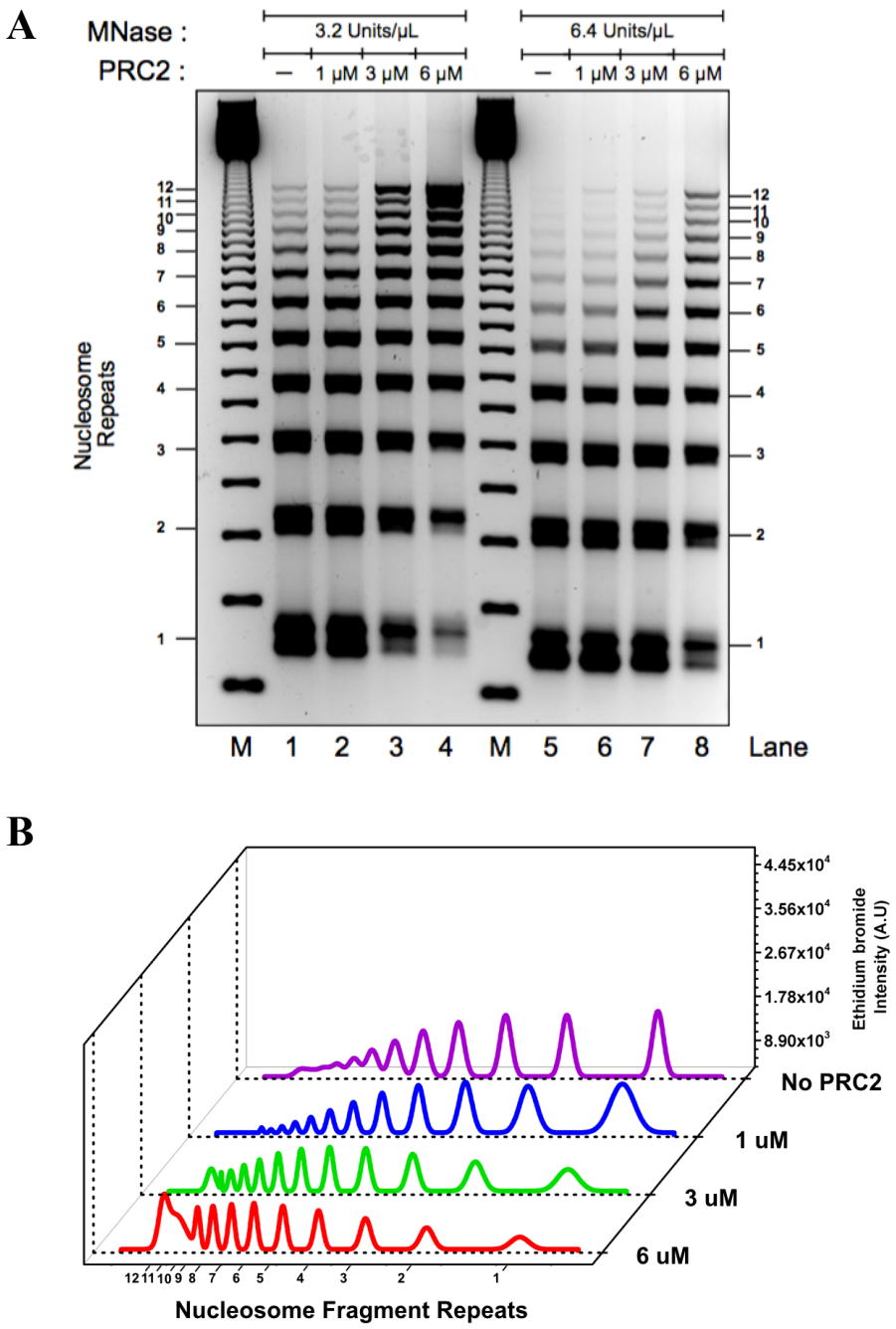


Figure 17 | PRC2 binding to nucleosome-linkers protects against micrococcal nuclease digestion

A) PRC2 protects nucleosome-linker regions from micrococcal nuclease digestion, as demonstrated by the increase in the number of large protected DNA fragments in proportion to increasing PRC2 concentration.

B) Quantitation of (A) lanes 1-4, showing a population shift in protected fragment lengths.

manifest as a decrease in the population of lower-molecular weight DNA bands in favor of higher-molecular weight bands. To conduct this analysis, a PRC2 binding experiment was carried out using 12-mer nucleosome array substrates, followed by MNase treatment (experimental details provided in methods section).

Strikingly, it was observed that with increasing PRC2 concentration, protection against MNase-digestion increased (**Figure 17A**). As predicted, this protection-ability was seen by the shift in the population of DNA-protected fragments from a low-to-high molecular weight. The waterfall plot in **Figure 17B** illustrates this population shift. These data suggests that PRC2 indeed is binding linker-DNA regions under these experimental conditions.

3.4 — *The Effect of RNA on PRC2 Association with Nucleosomes*

The association of PRC2 with chromatin is thought to be regulated by interaction with RNA.[10,21,26,27,38] As demonstrated in previous sections, recombinant PRC2 exhibits robust RNA-binding activity, especially to G-rich RNA sequences. And an extensive body of literature propose functional roles of RNA in cells. For example, one research group used RNA immunoprecipitation (RIP) to demonstrate that PRC2 interacts with the long non-coding RNA HOTAIR, and that this interaction modulates H3K27me3 deposition *in trans.*[71] Others have shown that XIST RNA may be important in potentially recruiting PRC2 to the X chromosome during XCI.[95] Other groups have performed knockdown experiments of RNA transcripts and have shown that this leads to changes in PRC2-chromatin association.

Together, these data have led to models in which specific RNAs recruit PRC2 to chromatin *in cis* or *in trans*.

However, PRC2 binding to nascent RNAs at almost all active promoters is difficult to reconcile with models in which RNA-binding plays only a role for PRC2-mediated silencing of genes. Therefore, it was asked whether PRC2-RNA binding and chromatin association might be either cooperative and result in the formation of a ternary (PRC2-RNA-Nucleosome) complex, or antagonistic, with RNA and chromatin competing for PRC2 interaction. To test this question, a competition experiment was performed between PRC2 bound to tri-nucleosomes and RNA. By titrating RNA, the effects on PRC2 association with nucleosomes were observed. Surprisingly, the results showed that RNA competed off PRC2 association with nucleosomes (**Figure 18**), a finding that suggests that PRC2 binding to RNA and nucleosomes is mutually exclusive. Additionally, RNAs that were previously characterized to bind well to PRC2, such as $(GGAA)_{10}$, were able to more readily compete off PRC2-nucleosome interaction than weak-binding RNAs such as poly (A).

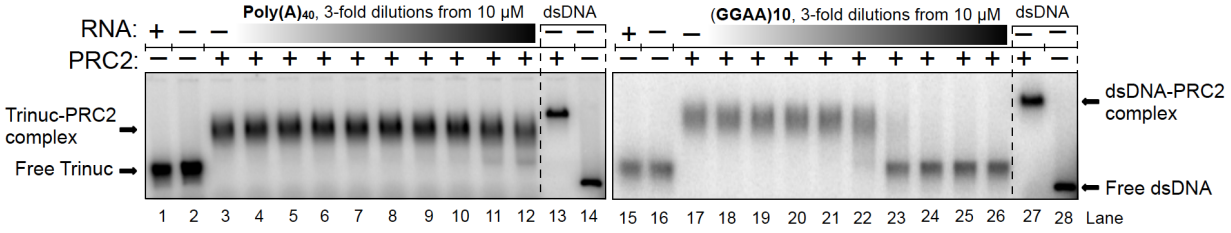


Figure 18 | RNA disrupts the association of PRC2 with nucleosomes. PRC2 complex that is bound to labeled tri-nucleosomes is competed off by titrating RNA. Titration using the weak binding RNA $Poly(A)_{40}$ shows little competition, whereas titrating using a tight binding RNA $(GGAA)_{10}$ shows strong competition effects.

3.5 — *Materials and Methods*

Histone Octamer Refolding: In a octamer refolding reaction, each of the pure human histone components (H2A, H2B, H3, H4) were prepared in unfolding buffer (6 M GdmCl, 10 mM Tris-HCl pH 7.5, 5 mM DTT). Equimolar amounts of H3 and H4 were mixed with 1.05 equivalents of H2A and H2B at 1 mg/ml and octamers were refolded by dialysis against refolding buffer (2 M NaCl, 10 mM Tris-HCl pH 7.5, 1 mM EDTA, 5 mM DTT). The refolded octamers were purified by gel filtration on a Superdex S200 10/300GL column. Octamer-peak fractions were analyzed by SDS-PAGE, and octamer-containing fractions were pooled, and concentrated. Octamer stocks were stored at 4 °C until needed for nucleosome reconstitution experiments.

Generation of template DNA for Nucleosome Reconstitution: Plasmids containing a 1, 3, and 12 × repeat of the Widom 601 nucleosome positioning sequence separated by 30-bp linker DNA segments were prepared recombinantly, and purified from cells using Qiagen Plasmid Giga kits. The plasmid DNA portions containing the nucleosome positioning sequence(s) were released from the vector backbone by restriction enzyme digestion. The backbone was removed by PEG precipitation with 9% PEG 6000. This was followed by purification by size-exclusion chromatography using a Mono Q 10/100 column. Following chromatography, fractions were assessed by DNA gel electrophoresis, and pure fractions were pooled, and concentrated by ethanol precipitation. The DNA pellet was dissolved in TE buffer pH 7.5, and the concentration was determined using NanoDrop.

In Vitro Nucleosome Reconstitution: Nucleosomes particles and arrays were reconstituted at a small scale from 1 to 6 uM in a volume of 25–50 µl. Widom 601 DNA (1-, 3-, or 12- repeats) were mixed with histone octamers at appropriate ratios and assembled by dialysis from RB-High buffer (10 mM Tris-HCl pH 7.5, 0.1 mM EDTA, 2 M KCl, 1 mM DTT) to RB-Low buffer (0.25 M KCl, 10 mM Tris-HCl pH 7.5, 1 mM EDTA, 1 mM DTT). The reactions were dialyzed gradually from RB-High buffer to RB-Low buffer at a flow-rate of 1.45 mL/min using a peristaltic pump (BioRad Model EP-1 Econo Pump). After dialysis for about 24 hours, buffer was exchanged into TCS buffer (20 mM Tris-HCl pH 7.5, 1 mM EDTA, 1 mM DTT) and allowed to dialyze for at least 1 hour. Nucleosomes were then transferred to LoBind tubes and stored at 4^{o}C. Quality of nucleosomes was assessed by restriction enzyme digestion, native PAGE, and AFM.

Micrococcal Nuclease Digestion analysis of Chromatin: Micrococcal nuclease (MNase) preferentially cleaves DNA in inter-nucleosomal (linker) regions of chromatin. Therefore, the partial digestion of chromatin with MNase reveals the periodic spacing of nucleosome fragments. To test the ability of PRC2 to protect linker regions of chromatin, an MNase-protection assay was set up. Micrococcal nuclease was diluted to concentrations of 3.2 Units/uL and 0.8 Units/uL in MNase reaction buffer (10 mM potassium HEPES pH 7.5, 10 mM KCl, 1.5 mM $MgCl_2$, 0.5 mM EGTA and 10% w/v glycerol). A PRC2-12mer-nucleosome binding reaction was prepared in a final volume of 45-uL, and incubated at 30^{o}C for 30 minutes. Following incubation, $CaCl_2$ is added to a final concentration of 5 mM, and 5-uL of diluted MNase is added to the mixture. The MNase reaction was allowed to proceed

for approximately 10 minutes before quenching with 12.5-uL of 500 mM EDTA. Then, 2.5-uL of a 10 mg/ml stock solution of proteinase K was added to digest PRC2 proteins and histones in the mixture by incubating for 30 minutes at 50^{o}C. The DNA was purified by phenol-chloroform extraction. Finally, the DNA was loaded onto a native 1.35% agarose gel (BP160) and run in 1X TBE buffer at 160 V for 3 hours. DNA bands were visualized by ethidium bromide staining. A 123-bp DNA ladder (1 ug per lane) was used as a size marker.

Atomic-Force Microscopy: For single-visualizing visualization, Atomic-Force microscopy (AFM) was utilized. For surface deposition, samples were diluted in deposition buffer (10 mM Tris-HCl pH 7.5, 3 mM $NiCl_{2)}$ to a final DNA concentration of 0.2 nM, and 50 uL were pipetted onto freshly-cleaved ultra-clean mica (Grade V, SPI). Samples were then incubated at room temperature for about 10 min. The mica discs were then rinsed with purified 18.2-MΩ deionized water. For samples intended for AFM imaging in air, the washed disks were then dried using a gentle N_2 gas flow, perpendicular to the mica surface. For samples intended for AFM imaging in liquid, the rinsed mica disks were quickly exchanged into imaging buffer (10 mM Tris-HCl pH 7.5, 3mM $NiCl_2$). All samples were imaged by AFM immediately following preparation. Data acquisition was performed using a Cypher ES Environmental Atomic-Force Microscope (Asylum Research). The samples were imaged in tapping mode using a commercial silicon cantilever with a spring constant of 46 N/m. For air imaging, Asylum AC240 cantilevers were used; for liquid imaging, Bruker SNL-10 cantilevers were used. In both cases, the cantilevers were manually tuned using thermal profiles. Images were captured at 512 x 512 pixels in the trace direction, at a scan size of 2

μm and a scan rate of 1.0 Hz. Throughout the experiments, deflection was maintained at zero. Image processing and data analysis were carried out using Matlab (MathWorks) with custom-written code, as well as publicly available software (Gwyddion).

RNA-Nucleosome Competition Assay: Competition reactions were set up identically to PRC2-RNA binding reactions except a constant concentration of PRC2 was included and variable concentrations of competitor RNAs were used. Competitor RNAs were folded in the same way as radioactively labeled RNAs. The amount of protein used in each assay was determined based on the concentration of protein giving 80% maximal binding of the labeled nucleosomes.

Section 4 — PRC2 is Guided to Sites of DNA Methylation

4.1 — *DNA-based Recruitment of PRC2*

The finding that RNA antagonizes PRC2 binding to nucleosomes *in vitro* appears contradictory to a model where RNA has specific roles for targeting PRC2 to chromatin. After all, how might PRC2 be able to be loaded onto nucleosomes if RNA can strip the complex away? While a functional recruiting role for RNA cannot be ruled out from the findings presented, it is clear that at least *in vitro*, RNA antagonizes PRC2 binding to nucleosomes. On the other hand, the role of DNA in driving the recruitment of PRC2 to genes has been broadly observed and warrants additional investigation. For example, subunits of PRC2, such as AEBP2, have been shown to have DNA-binding activity and therefore have been suggested to direct the binding of PRC2 to specific genomic loci.

Intriguingly, findings from several genomic and functional experiments have shown that there is a strong correlation between sites of PRC2 enrichment and CpG islands (**CGIs**) that often populate gene promoter regions.[70] In fact, genome-wide analysis of EZH2 and SUZ12 occupancy shows a remarkable association of PRC2 to a particular subset of CGIs.[70] And several studies have provided evidence for *de novo* recruitment of PRC2 to CpG islands in pluripotent cells. For example, in experiments performed by Mendenhall *et al*. using engineered ESCs that harbor artificial bacterial chromosomes (BACs) containing CGI motifs, it was demonstrated that these CGI motifs are sufficient to induce recruitment of catalytically active PRC2. These finding appear to argue in favor of a model of PRC2 recruitment which relies on the local architecture of chromatin, and that the presence of permissive CG-rich

chromatin sites are enough to lead to an enrichment in PRC2. Complementary to these findings, Riising *et al*. have shown that the drug-mediated inhibition of RNAPII is followed by *de novo* recruitment of PRC2 to these sites in embryonic stem cells. The recruitment of PRC2 following drug treatment was shown to be non-random and specifically targeting nucleosome-free CGIs. In many other cases, PRC2 targeting to CGI in DNA appears to be a recurrent theme in studies of various tissues and cell types.

4.2 — *Testing PRC2-DNA Binding Properties*

In *Drosophila*, polycomb response DNA elements are responsible for the initial recruitment of PRC2 to DNA.[25] However, a DNA-based mechanism remains unclear in mammals. Some clues have emerged regarding PRC2 targeting using a DNA-based mechanism. For instance, PRC2 associated proteins JARID2, AEBP2, and PCL1-3 appear to have some role in targeting the complex to unmethyated CpG islands and other regions.[62] And interestingly, the recombinant PRC2 5mer (EZH2-EED-SUZ12-RBBP4-AEBP2) complex used throughout this study was found to have the ability to bind DNA quite well. And indeed, this DNA-binding activity was found in previous sections to have the important function of dictating the affinity of PRC2 complexes for nucleosomes.

One intriguing question about the DNA-binding activity of PRC2 is whether PRC2 is able induce changes in the local conformation of DNA through DNA looping. It is known that DNA looping is deeply involved in many cellular processes such as transcription.[86] And DNA looping is known to be especially important in the regulation of gene expression, where

proteins bridge distal genes.[40] It is possible that PRC2 might be able to bridge chromatin in order to expedite H3K27me3 deposition and gene silencing. To address this question, Atomic-Force Microscopy imaging was performed using a 2.5 kb long linear DNA to visualize single molecules. A typical image of a naked DNA molecule is shown in **Figure 19A, I**. Interestingly, it was observed that at low concentrations of PRC2, the protein complex is in fact able to condense DNA through the formation of DNA loops (**Figure 19A, II**). And at higher concentrations of PRC2, it was observed that the population of molecules shifts to higher-order condensed DNA structures as shown in **Figure 19A, III**. While this DNA looping ability is an interesting biophysical property of PRC2, it is not known what the biological significance of this function might be. To speculate, perhaps DNA condensation helps create a local 3D chromatin architecture that is non-permissive for transcriptional machinery. A basic and speculative model of silencing by DNA condensation is proposed in **Figure 19B**.

During the study of PRC2 DNA-binding properties, a single-molecule fluorescence imaging system was also established to visualize PRC2 interaction with a doubly-tethered lambda DNA as shown in **Figure 19C**. Although, the system was not fully explored, the imaging reveals that PRC2 is perhaps able to "sample" DNA (**Figure 19D**). Although, this is only speculative and these experiments as of this time were mainly proof-of-concept. Certainly, this single-molecule fluorescence imaging system would be an interesting avenue for future research on PRC2 recruitment mechanisms.

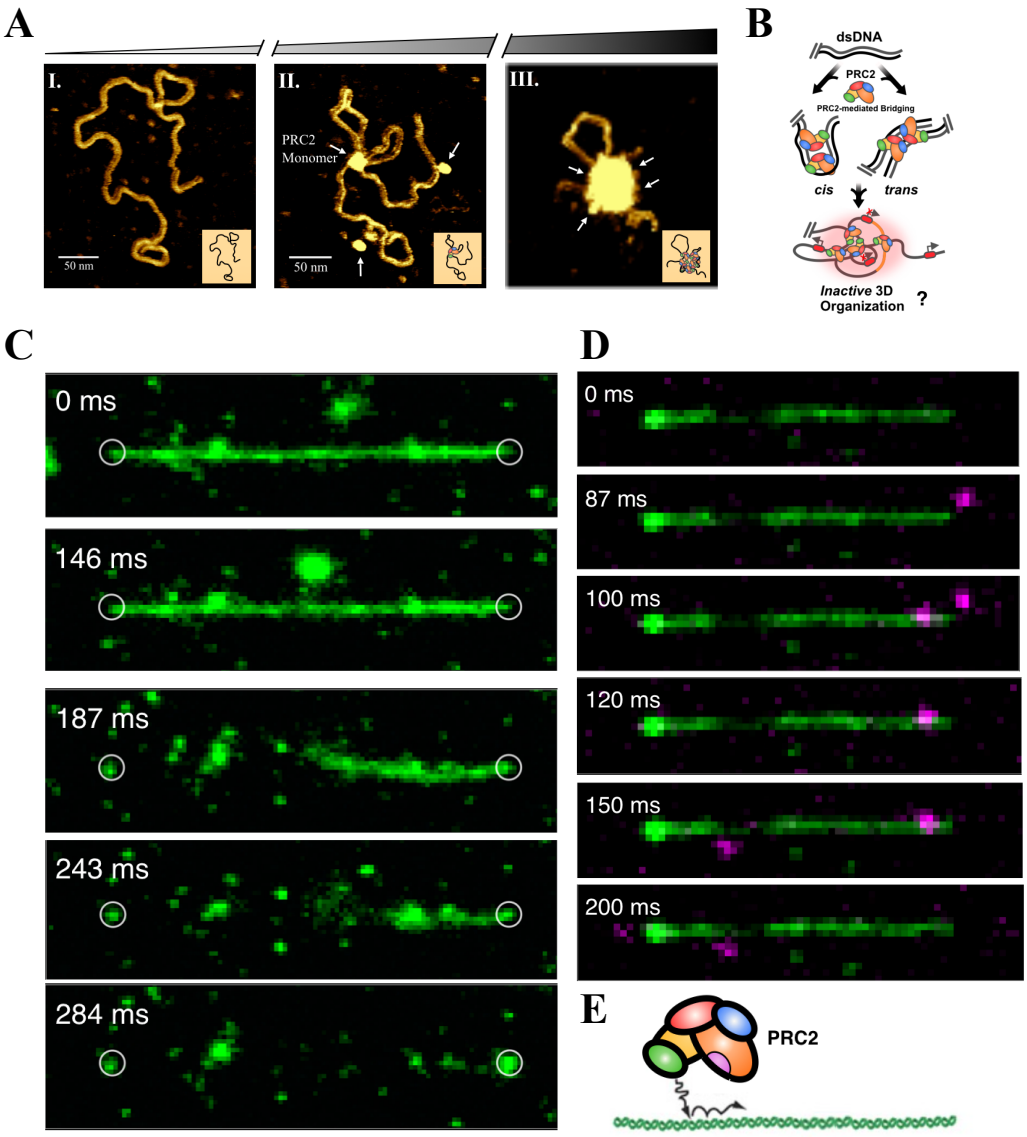


Figure 19 | PRC2 condenses DNA

A) AFM images of 2.5-kb linear DNA incubated with varying concentration of PRC2; I. No protein; II. 50 nM PRC2; III. 500 nM PRC2.

B) Possible model for the functional role of PRC2-mediated condensation of DNA.

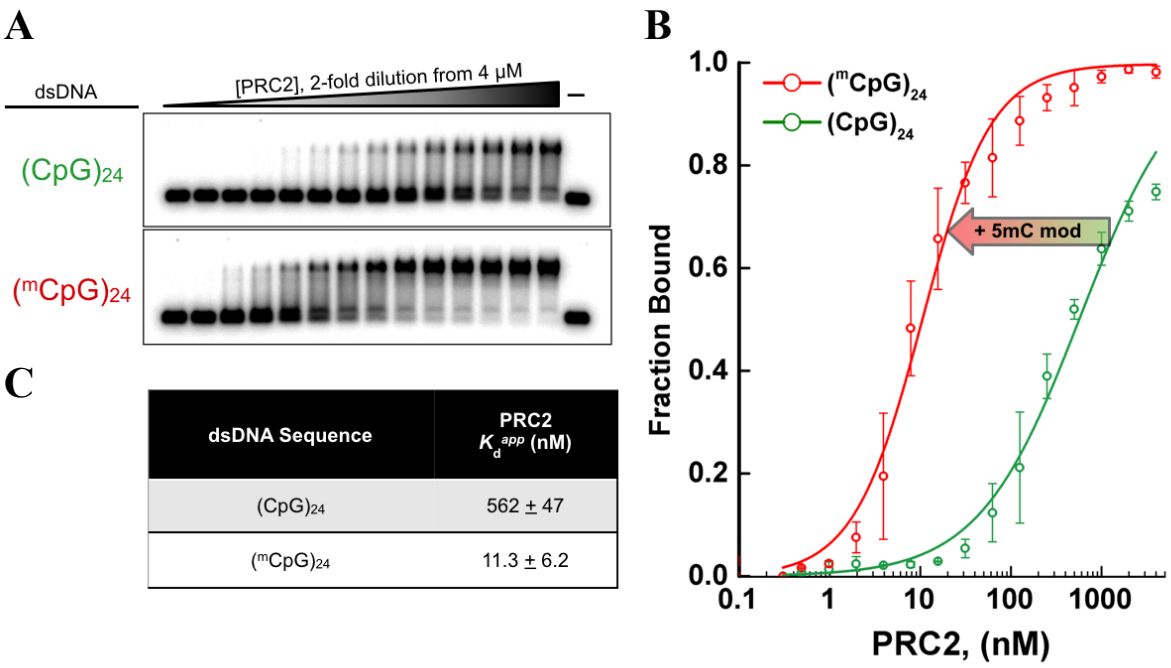
C) Photo-cleavage shows that tethered lambda DNA (shown by green color) is not non-specifically sticking to the glass surface.

D) PRC2 (shown by magenta color) binding event to lambda DNA shows that PRC2 does not non-specifically stick to DNA and perhaps randomly samples DNA.

4.3 — *PRC2 Recognizes 5-Methyl-CpG Dinucleotides in DNA*

Given the ability of PRC2 to bind DNA, an important question still stands: How is PRC2 recruited to accessible CGIs? Several possibilities exist that could explain the molecular events that underlie PRC2 recruitment to CGIs. Perhaps DNA methylation at CGIs leads to the local depletion of nucleosomes. From *in vitro* nucleosome binding experiments conducted and presented in previous sections, depletion of nucleosomes might correlate with PRC2 enrichment. Another possibility is that DNA methylation could evict transcription factors and other DNA binding proteins that might otherwise compete with PRC2 for DNA binding. Both scenarios would make DNA more accessible for PRC2. Moreover, *in vitro* and *in vivo* experiments have suggested that the AEBP2 PRC2 subunit appears to have affinity for GC-rich DNA.[57] Specifically, analysis of DNA at PRC2-bound loci revealed an enrichment in CpG dense genomic regions.[49] Accordingly, the fact that CpG dinucleotide density is a good predictor of mammalian PRC2 recruitment suggests that PRC2 relies in part on a DNA-based recruitment mechanism.[57] However, the mechanism by which CpG islands may recruit PRC2 remains unclear. One might speculate that perhaps an intermediate bridging protein that recognize 5-mC modifications is able to recruit PRC2 via protein-protein interaction. However, such a putative factor remains unidentified. Given that it is known that the PRC2 subunits are enriched at GC-rich DNA sites, another hypothesis is that PRC2 might be able to recognize 5-mC through a direct interaction. To test this hypothesis, methylated CpG-rich DNA was prepared *in vitro* using the CpG methyltransferase M.SssI as described in the methods section. PRC2 binding to methylated and un-methylated dsDNA was tested through EMSA experiments (**Figure 20A**). Binding to the un-methylated DNA

yielded a modest K_d^{app} of 562 nM. Strikingly, PRC2 binding to the methylated DNA was enhanced dramatically with a corresponding K_d^{app} measurement of 11.3 nM (**Figure 20B & C**). This difference constitutes a drastic ~50-fold increase in binding affinity for the CpG methylated DNA sequence. These results suggest that PRC2 is able to directly read 5-mC DNA modifications.



C

dsDNA Sequence	PRC2 K_d^{app} (nM)
$(CpG)_{24}$	562 ± 47
$(^mCpG)_{24}$	11.3 ± 6.2

Figure 20 | PRC2 exhibits a binding preference for CpG methylated DNA.
A) EMSA gels of PRC2 binding to labeled double-stranded DNA probes.
B) Binding curves for (A)
C) Table of Kd of interaction between PRC2 and methylated and non-methylated DNA.

4.4 — *Materials and Methods*

Electrophoretic Mobility Shift Assay (EMSA) for PRC2-dsDNA Binding: Double-stranded DNA (dsDNA) substrates were made by annealing single-stranded complementary oligonucleotides synthesized by Integrated DNA Technologies (IDT). To prepare the double-stranded DNA substrates, the single-stranded DNA oligonucleotides were mixed in a 1:1 molar ratio and boiled at 95^{o}C for 10 minutes to melt possible secondary structures. The single-stranded oligonucleotides were then annealed by slowly cooling to room temperature over the course of 24 hours. The annealed DNA was radiolabeled using [γ-^{32}P]-ATP and T4 Polynucleotide Kinase. Excess [γ-^{32}P]-ATP was removed using G-50 spin columns (GE illustra MicroSpin G-50 Columns for DNA purification). Radiolabeled DNA samples were loaded on a native 8% polyacrylamide gel and ran at 18 Watts for 2 hours. DNA was gel extracted and dissolved in TE buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA) and the counts of the purified DNA were determined using liquid scintillation counting. For binding reactions, stock DNA was diluted to working cpm in TE buffer, and the PRC2 protein stock was diluted in 2X DNA-binding buffer (50 mM Tris-HCl pH 7.5 at 25°C, 100 mM KCl, 2.5 mM $MgCl_2$, 0.1 mM $ZnCl_2$, 2 mM 2-mercaptoethanol, 0.1 mg/ml bovine serum albumin, 5% v/v glycerol). Protein and DNA were mixed 1:1 for a final volume of 10-uL and then incubated at 30^{o}C for 30 minutes. Following incubation, samples were loaded on a 1.0% agarose gel (BP160, Thermo Fisher Scientific) and run at 66 V for 90 min. The gel was then dried, exposed to a phosphor imaging screen, and scanned.

Single-molecule DNA-curtain experiments: Glass coverslips and microscopy slides with holes drilled using diamond drill bits were cleaned by sonication in acetone and ethanol. This was then followed by treatment with a mixture of concentrated sulfuric acid to 30% hydrogen peroxide (3:1 ratio). The coverslips and slides were rinsed in water and dried. Afterwards, the slides and coverslips were silanized using (3-aminopropyl)triethoxysilane (APTES) in acetone, and then passivated with a solution of 100 mg/ml mPEG-5000 and mPEG-5000 containing 1% biotin-mPEG by reacting overnight in the dark at room temperature. Slides and coverslips were then assembled into flow cells, containing four channels each separated by double-sided adhesive tape. Pipette tips were inserted into the drilled holes and the channels were sealed with glue. A peristaltic pump was used to generate laminar flow. The channels were hydrated and rinsed using T50 buffer (10 mM Tris-HCl pH 7.5, 50 mM KCl) before proceeding to experiments. For DNA immobilization, 0.2 mg/ml neutravidin solution was infused and incubated for 10 min, and followed by extensive washes with T50 buffer. Then, 500 pM of Cy5-labeled λ-DNA in T50 buffer was injected into the neutravidin treated flow chamber for 30 min, followed by a wash with T50 and imaging buffer (50 mM HEPES, 130 mM KCl, 10% glycerol, 2 mM TROLOX, 0.005% tween-20, 3.2% glucose, glucose oxidase). λ-DNA was anchored to the surface using complementary oligonucleotides containing biotin molecules. DNA was visualized through TIRF-microscopy (Nikon STORM) by fluorescent emission in the far-red channel. Dynamic experiments were initiated by flushing in 1 nM Cy3-labelled PRC2 in imaging buffer. All proteins were freshly diluted from a 100 nM stock and immediately injected. PRC2 dynamics were observed the Cy3 channel.

Section 5 — Discussion

To highlight the current understanding of the molecular mechanisms underlying PRC2 recruitment, PRC2 complexes are thought to be recruited to genes by three broad modes that include (I) association with RNA (II) targeting to PTMs on histone lysine tails and sensing nucleosome dense regions (III) binding DNA elements directly or indirectly through association with bridging proteins.[25] In this work, components of each of these recruitment mechanisms was tested via *in vitro* approaches using recombinant human PRC2 purified from baculovirus systems. And the present work here provides a quantitative and qualitative basis for understanding the promiscuous binding of PRC2 to RNA, as well as interactions with nucleosomes and DNA.

PRC2 has been well characterized as an RNA-binding protein complex,[27-30] and it has been proposed to interact with RNAs promiscuously *in vitro* and *in vivo* due to the broad spectrum of transcripts that it binds.[30] And through this work described here, it was revealed that PRC2 in fact reads RNA motifs consisting of short repeats of consecutive guanines. This mode of binding to guanine-rich tracts that are ubiquitous in the total RNA transcriptome clarifies the promiscuous nature of PRC2 targeting to RNA. Additionally, it was revealed that RNA structure can be a predictor for PRC2 binding, with unstructured single-stranded RNAs containing G-tracts being good ligands, and RNAs capable of forming G4s being even more superb substrates. While the actual function of RNA binding by PRC2 is still under investigation, the strong association between PRC2 and RNAs with G-tract motifs from RIP-seq data provides a means for further investigating RNA-mediated regulation of genes by

PRC2. Intriguingly, it was observed that G4-stabilizing ligands can inhibit the interaction between PRC2-RNA. This hints at the possibility of modulating gene-regulation by PRC2 through the use of small compounds that specifically target G-quadruplex forming RNA sequences.

The finding that PRC2 binding to nucleosomes is dictated by nucleosome-free regions *in vitro* challenges much of what is thought about histone-based PRC2 recruitment. Yet, this result might be in agreement with the previously perplexing observation that PRC2 is enriched at active promoters genome-wide. Given that active promoters are expected to be relatively nucleosome-free, this provides a means for PRC2 to sense local transcriptional states of promoter regions. These data then suggest that DNA plays an even more crucial role in PRC2 targeting than previously thought. Indeed, the subsequent finding that methylated DNA serves as a better ligand for PRC2 than un-methylated DNA is exciting as it suggests that epigenetic modifications at the level of DNA are also important determinants of PRC2 binding. Lastly, the finding that PRC2 binding to nucleosomes and RNA is mutually antagonistic might challenge prevailing models that purport that RNA directs PRC2 to genes. The findings reported here are summarized by a proposed interaction model for PRC2, as shown by the cartoon in **Figure 21**. In this model, PRC2 engages in DNA-based chromatin scanning. And CpG sites will selectively enrich for PRC2. Lastly, nascent RNA might either act to recruit PRC2 through G-rich targeting, or alternatively, evict it from promoter regions.

Despite having clarified a few details regarding PRC2 recruitment mechanism in this work, elucidating the full nature of this biological phenomenon still faces some major hurdles. First, there is a lack of atomic-level structures of PRC2 with cognate substrates such

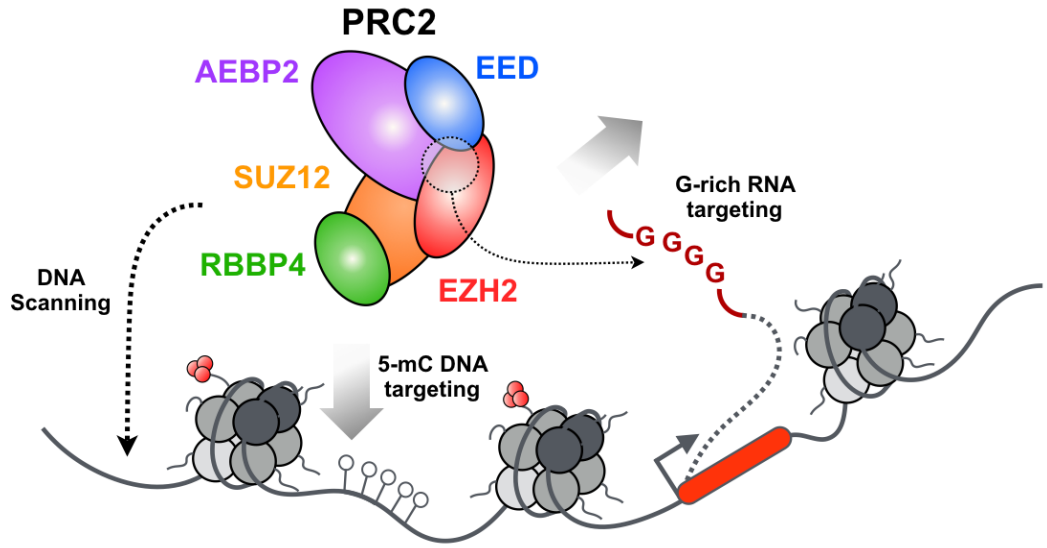


Figure 21 | Revised Model for the Recruitment of PRC2 to chromatin.
PRC2 senses chromatin state through DNA scanning, with nucleosome-free regions marking transcriptionally active genes. At sites of active transcription, PRC2 may be either evicted from the chromatin through G-rich RNA targeting or perhaps recruited to active genes through RNA. CpG islands are targeted preferentially by PRC2 and could mark promoter regions.

as nucleosomes, DNA, and RNA. Acquiring such structural information is critical because it will provide insight into which amino acids of PRC2 are key for their interaction. Structure-function studies would spur the discovery of separation-of-function mutations that will then allow researchers to fully analyze the biological significance of RNA-binding by PRC2. A second challenge will be to conduct parallel sequencing experiments to define the genomic occupancy of different components of PRC2 in relation to other transcription factors and nucleosome positions. Ideally, such experiments should be carried out in multiple cell lines that span from pluripotent cells, to differentiated cells of varying cell types, to disease models. Hopefully such an analysis will reveal how PRC2 is regulated across different tissue types and how it is dysregulated during tumorigenesis.

Section 6 — References

1. Aranda S, Mas G, Di croce L. Regulation of gene transcription by Polycomb proteins. Sci Adv. 2015;1(11):e1500737.

2. Arents G, Burlingame RW, Wang BC, Love WE, Moudrianakis EN. The nucleosomal core histone octamer at 3.1 A resolution: a tripartite protein assembly and a left-handed superhelix. Proc Natl Acad Sci USA. 1991;88(22):10148-52.

3. Azzalin, C.M., Reichenbach, P., Khoriauli, L., Giulotto, E., and Lingner, J. (2007). Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends. Science 318, 798-801.

4. 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.

5. Belton JM, Mccord RP, Gibcus JH, Naumova N, Zhan Y, Dekker J. Hi-C: a comprehensive technique to capture the conformation of genomes. Methods. 2012;58(3): 268-76.

6. Bhattacharyya D, Mirihana arachchilage G, Basu S. Metal Cations in G-Quadruplex Folding and Stability. Front Chem. 2016;4:38.

7. Birve A, Sengupta AK, Beuchle D, et al. Su(z)12, a novel Drosophila Polycomb group gene that is conserved in vertebrates and plants. Development. 2001;128(17):3371-9.

8. Bloom K. A close look at wiggly chromosomes. Dev Cell. 2013;25(4):330-2.

9. Bracken, A.P., and Helin, K. (2009). Polycomb group proteins: navigators of lineage pathways led astray in cancer. Nat Rev Cancer 9, 773-784.

10. Brockdorff, N. (2013). Noncoding RNA and Polycomb recruitment. RNA 19, 429-442.

11. Brooks TA, Hurley LH. Targeting MYC Expression through G-Quadruplexes. Genes Cancer. 2010;1(6):641-649.

12. Brooun, A., Gajiwala, K.S., Deng, Y.L., Liu, W., Bolanos, B., Bingham, P., He, Y.A., Diehl, W., Grable, N., Kung, P.P., et al. (2016). Polycomb repressive complex 2 structure

with inhibitor reveals a mechanism of activation and drug resistance. Nat Commun 7, 11384.

13. Boulay G, Dubuissez M, Van rechem C, et al. Hypermethylated in cancer 1 (HIC1) recruits polycomb repressive complex 2 (PRC2) to a subset of its target genes through interaction with human polycomb-like (hPCL) proteins. J Biol Chem. 2012;287(13): 10509-24.

14. Bugaut, A., and Balasubramanian, S. (2012). 5'-UTR RNA G-quadruplexes: translation regulation and targeting. Nucleic Acids Res 40, 4727-4741.

15. Bugaut A, Rodriguez R, Kumari S, Hsu ST, Balasubramanian S. Small molecule-mediated inhibition of translation by targeting a native RNA G-quadruplex. Org Biomol Chem. 2010;8(12):2771-6.

16. Calo E, Wysocka J. Modification of enhancer chromatin: what, how, and why?. Mol Cell. 2013;49(5):825-37.

17. Cammas A, Millevoi S. RNA G-quadruplexes: emerging mechanisms in disease. Nucleic Acids Res. 2016;

18. Cerase A, Pintacuda G, Tattermusch A, Avner P. Xist localization and function: new insights from multiple levels. Genome Biol. 2015;16:166.

19. Cerase A, Smeets D, Tang YA, et al. Spatial separation of Xist RNA and polycomb proteins revealed by superresolution microscopy. Proc Natl Acad Sci USA. 2014;111(6): 2235-40.

20. Ciferri C, Lander GC, Maiolica A, Herzog F, Aebersold R, Nogales E. Molecular architecture of human polycomb repressive complex 2. Elife. 2012;1:e00005.

21. 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.

22. Cogoi S, Paramasivam M, Filichev V, Géci I, Pedersen EB, Xodo LE. Identification of a new G-quadruplex motif in the KRAS promoter and design of pyrene-modified G4-decoys with antiproliferative activity in pancreatic cancer cells. J Med Chem. 2009;52(2):564-8.

23. Colak D, Zaninovic N, Cohen MS, et al. Promoter-bound trinucleotide repeat mRNA drives epigenetic silencing in fragile X syndrome. Science. 2014;343(6174):1002-5.

24. Collie, G.W., Haider, S.M., Neidle, S., and Parkinson, G.N. (2010). A crystallographic and modelling study of a human telomeric RNA (TERRA) quadruplex. Nucleic Acids Res 38, 5569-5580.

25. Comet I, Helin K. Revolution in the Polycomb hierarchy. Nat Struct Mol Biol. 2014;21(7):573-5.

26. Da rocha ST, Boeva V, Escamilla-del-arenal M, et al. Jarid2 Is Implicated in the Initial Xist-Induced Targeting of PRC2 to the Inactive X Chromosome. Mol Cell. 2014;53(2): 301-16.

27. Davidovich, C., and Cech, T.R. (2015). The recruitment of chromatin modifiers by long noncoding RNAs: lessons from PRC2. RNA 21, 2007-2022.

28. 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.

29. Davidovich C, Goodrich KJ, Gooding AR, Cech TR. A dimeric state for PRC2. Nucleic Acids Res. 2014;42(14):9236-48.

30. 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.

31. Di croce L, Helin K. Transcriptional regulation by Polycomb group proteins. Nat Struct Mol Biol. 2013;20(10):1147-55.

32. ENCODE Project Consortium. (2011). A user's guide to the encyclopedia of DNA elements (ENCODE). PLoS Biol. 9:e1001046.

33. Grand CL, Han H, Muñoz RM, et al. The cationic porphyrin TMPyP4 down-regulates c-MYC and human telomerase reverse transcriptase expression and inhibits tumor growth in vivo. Mol Cancer Ther. 2002;1(8):565-73.

34. Gupta R. A., Shah N., Wang K. C., et al. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature 464, 1071–1076 (2010).

35. Guo, J.U. and Bartel, D.P. (2016). RNA G-quadruplexes are globally unfolded in eukaryotic cells and depleted in bacteria. Science 353, aaf5371.

36. Hayashi A, Yamauchi N, Shibahara J, et al. Concurrent activation of acetylation and tri-methylation of H3K27 in a subset of hepatocellular carcinoma with aggressive behavior. PLoS ONE. 2014;9(3):e91330.

37. Hendrickson, D.G., Kelley, D.R., Tenen, D., Bernstein, B., and Rinn, J.L. (2016). Widespread RNA binding by chromatin- associated proteins. Genome Biol 17, 28.

38. Holoch D, Moazed D. RNA-mediated epigenetic regulation of gene expression. Nat Rev Genet. 2015;16(2):71-84.

39. Jiao, L., and Liu, X. (2015). Structural basis of histone H3K27 trimethylation by an active polycomb repressive complex 2. Science 350, aac4383.

40. Kadauke S, Blobel GA. Chromatin loops in gene regulation. Biochim Biophys Acta. 2009;1789(1):17-25.

41. Kalb R, Latwiel S, Baymaz HI, et al. Histone H2A monoubiquitination promotes histone H3 methylation in Polycomb repression. Nat Struct Mol Biol. 2014;21(6):569-71.

42. 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.

43. Kaneko, S., Son, J., Bonasio, R., Shen, S.S., and Reinberg, D. (2014). Nascent RNA interaction keeps PRC2 activity poised and in check. Genes Dev 28, 1983-1988.

44. 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.

45. Kanhere, A., Viiri, K., Araujo, 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.

46. Khalil, A.M., Guttman, M., Huarte, M., Garber, M., Raj, A., Rivea Morales, D., Thomas, K., Presser, A., Bernstein, B.E., van Oudenaarden, A., et al. (2009). Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. Proc Natl Acad Sci U S A 106, 11667-11672.

47. Kim, K.H., and Roberts, C.W. (2016). Targeting EZH2 in cancer. Nat Med 22, 128-134. Kodama, Y., Shumway, M., Leinonen, R., and International Nucleotide Sequence Database, C. (2012). The Sequence

48. Le, D.D., Di Antonio, M., Chan, L.K., and Balasubramanian, S. (2015). G-quadruplex ligands exhibit differential G-tetrad selectivity. Chem Commun (Camb) 51, 8048-8050.

49. Lee TI, Jenner RG, Boyer LA, et al. Control of developmental regulators by Polycomb in human embryonic stem cells. Cell. 2006;125(2):301-13.

50. Lee JT. Gracefully ageing at 50, X-chromosome inactivation becomes a paradigm for RNA and chromatin control. Nat Rev Mol Cell Biol. 2011;12(12):815-26.

51. Lewis PW, Müller MM, Koletsky MS, et al. Inhibition of PRC2 activity by a gain-of-function H3 mutation found in pediatric glioblastoma. Science. 2013;340(6134):857-61.

52. Lohr D, Bash R, Wang H, Yodh J, Lindsay S. Using atomic force microscopy to study chromatin structure and nucleosome remodeling. Methods. 2007;41(3):333-41.

53. Lowary PT, Widom J. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. J Mol Biol. 1998;276(1):19-42.

54. Luger K, Mäder AW, Richmond RK, Sargent DF, Richmond TJ. Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature. 1997;389(6648):251-60.

55. Margueron R, Reinberg D. The Polycomb complex PRC2 and its mark in life. Nature. 2011;469(7330):343-9.

56. McCabe, M.T., Ott, H.M., Ganji, G., Korenchuk, S., Thompson, C., Van Aller, G.S., Liu, Y., Graves, A.P., Della Pietra, A., 3rd, Diaz, E., et al. (2012). EZH2 inhibition as a therapeutic strategy for lymphoma with EZH2-activating mutations. Nature 492, 108-112.

57. Mendenhall, E.M., Koche, R.P., Truong, T., Zhou, V.W., Issac, B., Chi, A.S., Ku, M., and Bernstein, B.E. (2010). GC-rich sequence elements recruit PRC2 in mammalian ES cells. PLoS Genet 6, e1001244.

58. Moraca F, Amato J, Ortuso F, et al. Ligand binding to telomeric G-quadruplex DNA investigated by funnel-metadynamics simulations. Proc Natl Acad Sci USA. 2017;114(11):E2136-E2145.

59. Morris, M.J., Wingate, K.L., Silwal, J., Leeper, T.C., and Basu, S. (2012). The porphyrin TmPyP4 unfolds the extremely stable G-quadruplex in MT3-MMP mRNA and alleviates its repressive effect to enhance translation in eukaryotic cells. Nucleic Acids Res 40, 4137-4145.

60. Mozzetta C, Pontis J, Ait-si-ali S. Functional Crosstalk Between Lysine Methyltransferases on Histone Substrates: The Case of G9A/GLP and Polycomb Repressive Complex 2. Antioxid Redox Signal. 2015;22(16):1365-81.

61. Parkinson GN, Ghosh R, Neidle S. Structural basis for binding of porphyrin to human telomeres. Biochemistry. 2007;46(9):2390-7.

62. Pasini D, Cloos PA, Walfridsson J, et al. JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells. Nature. 2010;464(7286):306-10.

63. Peterson CL. Salt gradient dialysis reconstitution of nucleosomes. CSH Protoc. 2008;2008:pdb.prot5113.

64. Plath K., Fang J., Mlynarczyk-Evans S. K., Cao R., at al. Role of histone H3 lysine 27 methylation in X inactivation. Science 300, 131–135 (2003).

65. Plath K, Mlynarczyk-evans S, Nusinow DA, Panning B. Xist RNA and the mechanism of X chromosome inactivation. Annu Rev Genet. 2002;36:233-78.

66. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841-842.

67. Read Archive: explosive growth of sequencing data. Nucleic Acids Res 40, D54-56. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods 9, 357-359.

68. Reiter NJ, Osterman A, Torres-larios A, Swinger KK, Pan T, Mondragón A. Structure of a bacterial ribonuclease P holoenzyme in complex with tRNA. Nature. 2010;468(7325): 784-9.

69. Rhodes D, Lipps HJ. G-quadruplexes and their regulatory roles in biology. Nucleic Acids Res. 2015;43(18):8627-37.

70. Riising EM, Comet I, Leblanc B, Wu X, Johansen JV, Helin K. Gene silencing triggers polycomb repressive complex 2 recruitment to CpG islands genome wide. Mol Cell. 2014;55(3):347-60.

71. 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.

72. Rodriguez, R., Muller, S., Yeoman, J.A., Trentesaux, C., Riou, J.F., and Balasubramanian, S. (2008). A novel small molecule that alters shelterin integrity and triggers a DNA-damage response at telomeres. J Am Chem Soc 130, 15758-15759.

73. Rouskin, S., Zubradt, M., Washietl, S., Kellis, M., and Weissman, J.S. (2014). Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. Nature 505, 701-705.

74. Sato MH, Ura K, Hohmura KI, et al. Atomic force microscopy sees nucleosome positioning and histone H1-induced compaction in reconstituted chromatin. FEBS Lett. 1999;452(3):267-71.

75. Sauvageau, M., and Sauvageau, G. (2010). Polycomb group proteins: multi-faceted regulators of somatic stem cells and cancer. Cell Stem Cell 7, 299-313.

76. Seol, Y., Skinner, G.M., Visscher, K., Buhot, A., and Halperin, A. (2007). Stretching of homopolymeric RNA reveals single-stranded helices and base-stacking. Phys Rev Lett 98, 158103.

77. Shi, D.F., Wheelhouse, R.T., Sun, D., and Hurley, L.H. (2001). Quadruplex-interactive agents as telomerase inhibitors: synthesis of porphyrins and structure-activity relationship for the inhibition of telomerase. J Med Chem 44, 4509-4523.

78. Shlyakhtenko LS, Gall AA, Lyubchenko YL. Mica functionalization for imaging of DNA and protein-DNA complexes with atomic force microscopy. Methods Mol Biol. 2013;931:295-312.

79. Somarowthu, S., Legiewicz, M., Chillon, I., Marcia, M., Liu, F., and Pyle, A.M. (2015). HOTAIR forms an intricate and modular secondary structure. Mol Cell 58, 353-361.

80. Son J, Shen SS, Margueron R, Reinberg D. Nucleosome-binding activities within JARID2 and EZH1 regulate the function of PRC2 on chromatin. Genes Dev. 2013;27(24):2663-77.

81. Spitale, R.C., Flynn, R.A., Zhang, Q.C., Crisalli, P., Lee, B., Jung, J.W., Kuchelmeister, H.Y., Batista, P.J., Torre, E.A., Kool, E.T., et al. (2015). Structural imprints in vivo decode RNA regulatory mechanisms. Nature 519, 486-490.

82. Tang, C.F., and Shafer, R.H. (2006). Engineering the quadruplex fold: nucleoside conformation determines both folding topology and molecularity in guanine quadruplexes. J Am Chem Soc 128, 5966-5973.

83. Tsai, M.C., Manor, O., Wan, Y., Mosammaparast, N., Wang, J.K., Lan, F., Shi, Y., Segal, E., and Chang, H.Y. (2010). Long noncoding RNA as modular scaffold of histone modification complexes. Science 329, 689-693.

84. Van kruijsbergen I, Hontelez S, Veenstra GJ. Recruiting polycomb to chromatin. Int J Biochem Cell Biol. 2015;67:177-87.

85. Vahabi S, Nazemi salman B, Javanmard A. Atomic force microscopy application in biological research: a review study. Iran J Med Sci. 2013;38(2):76-83.

86. Vilar JM, Saiz L. DNA looping in gene regulation: from the assembly of macromolecular complexes to the control of transcriptional noise. Curr Opin Genet Dev. 2005;15(2): 136-44.

87. Vorlickova, M., Kejnovska, I., Sagi, J., Renciuk, D., Bednarova, K., Motlova, J., and Kypr, J. (2012). Circular dichroism and guanine quadruplexes. Methods 57, 64-75.

88. Wang W, Qin JJ, Voruganti S, Nag S, Zhou J, Zhang R. Polycomb Group (PcG) Proteins and Human Cancers: Multifaceted Functions and Therapeutic Implications. Med Res Rev. 2015;35(6):1220-67.

89. Wang X, Goodrich KJ, Gooding AR, et al. Targeting of Polycomb Repressive Complex 2 to RNA by Short Repeats of Consecutive Guanines. Mol Cell. 2017;65(6):1056-1067.e5.

90. Williamson, J.R., Raghuraman, M.K., and Cech, T.R. (1989). Monovalent cation-induced structure of telomeric DNA: the G-quartet model. Cell 59, 871-880.

91. Wu, L., Murat, P., Matak-Vinkovic, D., Murrell, A., and Balasubramanian, S. (2013). Binding interactions between long noncoding RNA HOTAIR and PRC2 proteins. Biochemistry 52, 9519-9527.

92. Yap K. L., Li S., Muñoz-Cabello A. M., Raguz S., Zeng L., et al. Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by Polycomb CBX7 in transcriptional silencing of INK4a. Mol. Cell 38, 662–674 (2010).

93. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol 9, R137.

94. 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.

95. Zhao, J., Sun, B.K., Erwin, J.A., Song, J.J., and Lee, J.T. (2008). Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. Science 322, 750-756.