The molecular basis of RNA recognition and transcriptional regulation by FUS and PRC2

by

Xueyin Wang

B.S., University of Toronto, 2011

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirement for the degree of

Doctor of Philosophy

Department of Chemistry and Biochemistry

2017

This thesis entitled:
The molecular basis of RNA recognition and transcriptional regulation by FUS and PRC2 written by Xueyin Wang
has been approved for the
Department of Chemistry and Biochemistry

_____
Thomas R. Cech

_____
Roy Parker

Date_____

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

Xueyin Wang (Ph.D. Biochemistry)

Title: The molecular basis of RNA recognition and transcriptional regulation by FUS and PRC2

Thesis directed by Prof. Thomas R. Cech

RNA and protein interact with each other in the cell nucleus to form RNP (ribonucleoprotein) complexes. These RNPs play key roles in various steps of gene expression. This thesis focuses on two nuclear RNA-binding proteins: Fused in Sarcoma (FUS) and Polycomb Repressive Complex 2 (PRC2).

The normal functions of FUS are not fully understood, but RNA binding has been suggested to be crucial for FUS function and many RNA targets have been identified. Yet the features of RNAs necessary for FUS binding had not been systematically characterized. FUS bound all five published RNA motifs and other RNAs tightly, including fragments of an *Escherichia coli* mRNA, suggesting that FUS has a wide range of nucleic-acid binding ability. Additional pull-down experiments reveal that FUS binds the C-terminal domain (CTD) of RNA polymerase II in an RNA-dependent manner, providing insight into its function.

PRC2 has been shown to interact with RNAs promiscuously *in vitro* and *in vivo*, as evidenced by the broad spectrum of transcripts that it binds. Detailed quantitative binding experiments with RNAs of defined sequence and length, coupled with analysis of fRIP-seq and ChIP-seq data, show that PRC2 preferentially binds G-tracts in RNA. The G-tracts can be either single-stranded or, preferably, folded G-quadruplexes, which are ubiquitous in the transcriptome. This explains the promiscuous nature of PRC2 targeting to RNA. Additional binding and histone methyltransferase assays showed that RNA is not

a methyltransferase inhibitor; rather it sequesters PRC2 away from nucleosomes. More surprisingly, human PRC2 has robust DNA-binding activity, and RNA inhibits PRC2 binding to nucleosome by competing PRC2 off the linker region of nucleosomes.

Collectively, findings presented in this thesis offer new insights into the mechanistic details underlying the recognition of RNAs by FUS protein and the PRC2 methyltransferase. It is clear that the dysregulation of FUS is linked to neurodegenerative diseases, and that PRC2 is dysregulated in some types of cancers. Thus, understanding the mechanism by which these nuclear RNA-binding proteins recognize RNAs helps pave new avenues towards developing novel therapeutics to treat RNA-dominant diseases.

Acknowledgement

My time spent as a graduate student in Boulder has been one of the most unforgettable experiences of my entire life to date. And joining Tom Cech's laboratory has been a highlight of my journey. As a mentor, Tom is a wonderful professional role model and I came to learn quickly that he is a very gracious man. I learned so much from him. Indeed, his enthusiasm for research and biology was contagious and inspiring. Even though Tom's days were always packed with other duties and obligations, I found during my graduate student years that he was always keen to take time out of his busy schedule to meet with me to discuss experiments, scientific literature, and life in general. Tom has provided me with substantial support, guidance, patience, and insight during my graduate training. His positive attitude is always uplifting, and during times when I felt down or had been experiencing hardship, his encouraging words would embolden me. His mentorship helped me develop into a more thoughtful person and I learned so much from him that I will carry for the rest of my life. I will never forget the time when another research group published data that was similar to mine before me, and I was completely devastated and felt hopeless. That day, Tom walked into the lab to find me and gave me a big hug and comforted me. I feel extremely lucky to have had the opportunity to be a part of the Cech lab and am really fortunate to have such a wonderful advisor. Many Thanks, Tom!

I would also like to gratefully thank everyone in Tom's lab for creating such a pleasant environment to work in. I would like to take the opportunity to thank all the team members who worked on FUS assembly, as well as the PRC2 dream team, for their enormous support, help, discussion, and encouragement over the years. Especially, I

want to thank Jacob Schwartz and Chen Davidovich for training me and teaching me numerous technical assays, and for providing useful advice on my work on FUS and PRC2. I would like to acknowledge Karen Goodrich, Anne Gooding, Richard Paucek, and Dan Youmans for their help, contribution and dedication to the project. In addition, I would like to thank Ciji Lim, Jens Schmidt, and Richard Paucek for providing insightful discussion on my project, being fun to be around in lab, and always willing to help me whenever I bothered them. Furthermore, I would like to thank all of the Cech lab graduate students who I encountered during my time in the lab, including Andrew Dalby, Kate Goldforb, Linghe Xi, and Dan Youmans. Finally, I would like to thank Art Zaug and Karen Goodrich for their enjoyable conversions and for their support. I will remember all of the fun times we have had together: lab meetings, birthday celebrations, lab hikes, and the Christmas parties. Thank you all!

I would like to thank my committee members and other members of the department for their years of support and guidance. And I have great appreciation for my friends at Boulder who spent a great deal of time with me and made my life more colorful.

Last but not the least, I would like to give my deepest appreciation to my family: my mom and my dad for their unconditional love, endless care, and support throughout my life; my grandparents for raising me when I was a baby. And my beloved grandpa, who I was not able to be with when he passed away, is always in my heart and his spirit is always with me. I love you! I could never forget about the time that I left home on a 20-hour flight from Beijing to Ottawa when I was only 18-year old. That was the first time that I saw my parents crying as I walked through the airport's security check. All these

memories and experiences are so meaningful and valuable to me. I would never be the same without them.

THANK YOU ALL!!!

Table of Contents

List of Tables

List of Figures

Chapter I: Background on RNA-protein interactions

In cells, RNAs generally interact with proteins and form complexes known as ribonucleoproteins (RNPs). These RNPs have important regulatory roles in coordinating gene expression. Pre-mRNA synthesis by RNA polymerases, mRNA splicing by spliceosome complexes, and protein synthesis by the ribosomes are only but a few examples of how prevalent and important RNA-protein interactions are in fundamental molecular operations. In addition, RNA-protein interactions also are pivotal to maintaining the structure and activity of crucial intracellular machinery, such as telomerase and the signal recognition particle. Dysregulation of the RNA-protein interactions in RNPs could lead to the malfunction of physiological processes and the onset of diseases. Therefore, understanding the molecular basis of how proteins recognize and interact with RNA is important for gaining insights into biological processes and causes of diseases.

In each eukaryotic cell, there are thousands of different RNAs existing at a given time. How do RNA-binding proteins distinguish among these different RNAs? And what is the molecular basis of how these proteins bind RNAs? These are key questions in the RNA field, and answers to these questions would contribute to understanding many fundamental molecular processes that depend on RNA molecules.

RNA BINDING DOMAINS

To act on RNAs, proteins typically contain canonical RNA binding domains (RBDs), which are divided into two categories. The first category includes proteins domains that not only bind to RNA, but also possess enzymatic activity towards the RNA it binds, as seen in GTPases and helicases (Lunde et al., 2007). The second category

consists of proteins that have a domain that only functions to bind RNAs, such as the RRM (RNA-recognition motif) domain (Lunde et al., 2007).

RNA-recognition motif (RRM)

In humans, ~1% of genes encode a protein with an RNA-recognition motif, which makes the RRM one of the most abundant protein domains (Cléry et al., 2008). Indeed, proteins with the RRM domain are involved in many essential biological processes, including mRNA and rRNA processing, RNA export and stability, etc.

The typical RRM contains 80-90 amino acids, which fold into a four-stranded anti-parallel β-sheet and two α-helices (Maris et al., 2005). This results in a αβ sandwich structure with a $\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$ topology. Many RRMs interact with RNA in a highly sequence–specific manner, and numerous structures of RRM-RNA complexes have been solved either by X-ray crystallography or NMR. Based on the available structures, the binding specificity for RNA comes from eight conserved residues (typically aromatic and positively charged amino acids), which are found in most RRMs (Alfano et al., 2004).

Most proteins with RRMs do not contain only a single copy of the RRM. Instead, it is common for proteins to have multiple domains. RRM domain duplication is thought to allow for the recognition of an even greater number of nucleotides in RNA, and thus leads to a higher binding affinity and specificity to RNA. One good example came from a co-crystal structure of HuD with RNAs (Wang and Tanaka Hall, 2001). However, other studies have reported that the RRM domain is also involved in mediating protein-protein interaction (Tripsianes et al., 2014). Analysis of published RRM structures either alone or together with its cognate RNA, reveal that this small but unique domain is highly diverse both in terms of structure and function.

Zinc-finger domain (Znf)

A zinc-finger domain, as the name suggests, coordinates zinc ions, which helps stabilize protein folding and maintains the integrity of binding site. Zinc-finger domains typically interact with a broad variety of partners, including DNA, RNA, and proteins. Therefore, it is not surprising that in eukaryotes, zinc finger proteins are a part of an extremely diverse and abundant family of proteins, and are involved in many processes, including DNA recognition, transcription, translation, etc. (Laity et al., 2001).

Most zinc-finger domains have a characteristic fold. Based on structural classifications, zinc-finger domains can be categorized into three major divisions, namely C2H2-like finger, treble clef finger, and the zinc ribbon (Kadrmas and Beckerle, 2004).

KH motif

The KH motif was first identified in hnRNP K. This motif typically consists of ~50 amino acids, which fold into a stable βααβ core (Du et al., 2005). In addition to the core, a conserved GxxG loop linking two α-helices is essential for the interaction with a nucleic acid backbone (Grishin, 2001). Deletion of this loop abolishes the interaction of nucleic with protein.

Multiple copies of the KH motif co-exist in one protein, and so far up to 15 copies have been found in the protein Vigillin (Musco et al., 1996). Multiple copies of the KH motif coordinate with each other to achieve better recognition and targeting to nucleic acid sequences.

SPECIFICITY AND NON-SPECIFICITY IN RNA-PROTEIN INTERACTIONS

Generally, a cellular protein-RNA interaction is determined by combinatorial factors, including the intrinsic affinity of protein to RNA, the concentration of RNA and protein, and the competition between other proteins and other RNAs for binding. The interaction between protein and RNA can be characterized as being either specific or nonspecific. Specific interactions involve protein recognition of cognate RNA sequences or structural motifs. Examples include the interaction of MS2 coat protein with ms2 bulged-hairpin RNA (LeCuyer et al., 1995); RNase P recognition of precursor tRNA (Frank and Pace, 1998). On the other hand, protein-RNA interactions characterized as nonspecific are those that lack a defined sequence or structural motif. Such examples could be found in protein-RNA involved in degradation (Parker and Song, 2004) and in translation initiation, such as eIF1A, which interacts with mRNAs (Aitken and Lorsch, 2012; Battiste et al., 2000). Substrate binding by these types of proteins is not well understood; especially to the extent of how the protein distinguishes between different RNA substrates. Perhaps nonspecific RNA binding proteins recognize abundant sequence or structural motifs, which could underlie their observed nonspecific binding properties. Also, it is possible that nonspecific RNA binding proteins require specific cooperating proteins to help coordinate specific targeting to RNA.

In most cases, the nonspecific interaction between protein and RNA is dependent on electrostatic forces (salt bridges) between the protein and the phosphate backbone of the RNA. However, specific interactions are mostly achieved by van der Waals interactions for shape complementarity or hydrogen bonding to specific bases.

MAJOR TYPES OF RNA

In cells, the three major types of RNA include mRNA, tRNA and rRNA (Figure 1.1).

mRNA (messenger RNA)

Messenger RNA (mRNA) is a subtype of RNA, which was first described by scientists Elliot Volkin and Lazarus Astrachan. In eukaryotes, mRNA is synthesized by RNAPII during transcription from one strand of dsDNA and this single-stranded RNA contains trinucleotide codons, each of which encodes a specific amino acid. Thus, an mRNA message contains the instructions for the formation of a protein. In prokaryotes, mRNA is the exact copy of the transcribed DNA sequence. Notably, in eukaryotes, typically separate blocks of DNA sequences called exons are spliced together to make the mRNA (Faustino and Cooper, 2003). In addition, eukaryotic mRNAs are decorated with a 5' cap-structure consisting of a 7-methylguanosine covalently linked to the first nucleotide through a 5'-5' triphosphate bridge, and a 3' string of adenosines called a poly(A) tail (Cowling, 2010). These 5' and 3' decorations are added enzymatically co-transcriptionally or post-transcriptionally (Mitchell et al., 2010). Unlike in prokaryotes where mRNA is degraded rapidly, the 5' and 3' modifications of mRNA in eukaryotes dramatically improve their stability.

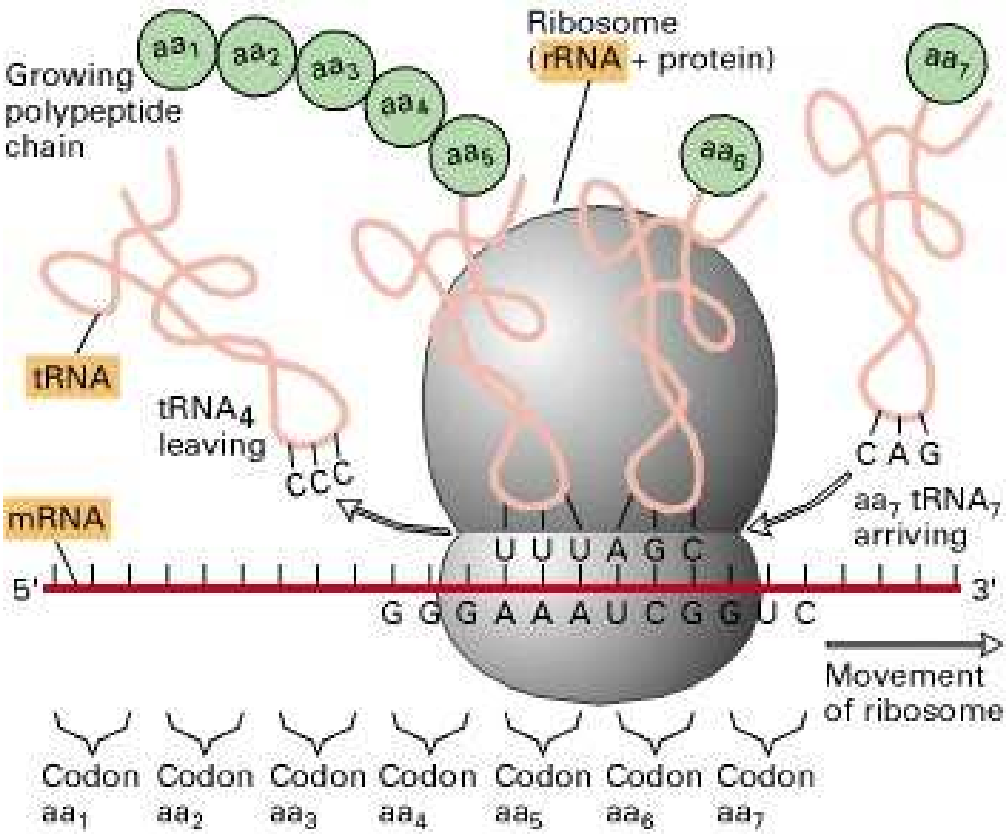


Figure 1.1. The three major types of RNA:mRNA, tRNA and rRNA; their roles in protein synthesis. ((Lodish et al., 2000)

tRNA (Transfer RNA)

Transfer RNA (tRNA) is a noncoding subtype of RNA molecule, which helps decode the mRNA into protein. tRNA contains three hairpin loops, which fold into a distinctive “L-shape” structure (Shi and Moore, 2000). One of the hairpin loops contains an anticodon, which recognizes and decodes an mRNA codon at specific sites in the ribosome during translation. Upon codon recognition, each tRNA with its own attached amino acid transfers that specific amino acid to the growing polypeptide chain. This addition of amino acids occurs repetitively until the mRNA is fully decoded.

In general, usually 40-to-60 different types of tRNA exist inside the cell, depending on the species, and these RNAs function to decode mRNAs. The trinucleotide nature of

the genetic code indicates that the four different nucleotides (A, U, G, C) in mRNA can yield a total of 64 codon combinations. Of these 64 codons, only 61 specify an amino acid, and the remaining 3 serve as stop signals halting protein synthesis. Despite 64 coding possibilities, there are only 20 naturally occurring amino acids. Thus, most amino acids are specified by multiple codons. This phenomenon related to the genetic code is known as degeneracy. Indeed, the implication of this degeneracy is that some tRNAs utilize “wobble” base pair strategies which allow tRNAs to base pair atypically in the third nucleotide position of a codon (Agris et al., 2007). For example, a G can form a base pair with a C or U in the third position of the codon. Via Wobble base pairing, coverage of all of the codons of the genetic code is ensured.

Each tRNA is charged with an amino acid by an aminoacyl-tRNA synthetase. This enzyme uses a two-step reaction (see the following) to catalyze the loading of amino acid to tRNA (Pang et al., 2014).

1. **amino acid + ATP -----> aminoacyl-AMP + PP_i.**
2. **aminoacyl-AMP + tRNA -----> aminoacyl-tRNA + AMP**

The specific recognition between tRNA synthetases and tRNA ensures that each tRNA is charged correctly with its corresponding amino acid.

rRNA (Ribosomal RNA)

Ribosomal RNAs (rRNA), with ribosomal proteins, form the ribosome, which is the cellular protein-synthesizing machinery. This important machinery translates mRNA into protein. More than half of the weight of a ribosome belongs to RNA.

rRNA are synthesized in the nucleolus, a dense region in the nucleus. rRNA differ in size, and combined with ribosomal protein, rRNA could form either large or small

subunits of the ribosome (Myasnikov et al., 2009). In bacteria, 50S and 30S (based on the rate of sedimentation) are referred to as large and small subunits, respectively. Compared to prokaryotic ribosomes, eukaryotic ribosomes are larger with 40S and 60S being the small and large subunits, respectively.

The rRNAs fold into defined structures, which are highly conserved between species, even with different sequences. Across all species, the function of rRNA is shared and the rRNAs play important roles in coordinating mRNA and tRNA (Jackson et al., 2010). In fact, all the key catalytic activity of the ribosome comes from the rRNA component. The catalytic activity of rRNA stands as a key piece of evidence to support the "RNA World Hypothesis."

NON-CODING RNA

Non-coding RNAs (ncRNAs) are a group of RNA molecules transcribed from the genome but are not translated into proteins. The tRNAs and rRNAs described above are abundant classes of ncRNAs. Due to advances in sequencing technologies, it has now been shown that about 90% of the eukaryotic genome is transcribed into RNA and among them, only ~2% encode for proteins (Rinn and Chang, 2012). In other words, ncRNAs account for the majority of the transcriptome.

Based on recent studies, ncRNA can be classified into two major types: housekeeping and regulatory ncRNAs. Housekeeping ncRNAs include RNAs which play maintenance roles, such as ribosomal, transfer, telomerase, and small nuclear RNAs. On the other hand, regulatory ncRNAs include RNAs such as microRNAs, piwi-interacting

RNAs and long ncRNAs that are involved in modifying the production, translation, or stability of other RNAs.

MicroRNA

MicroRNAs (miRNAs) are transcribed by RNAPII and III, and are typically single-stranded RNAs that are 20-24 nt in length (He and Hannon, 2004). Mature miRNAs are produced following a series of processing steps. Initially, miRNAs are transcribed into primary-microRNAs (pri-miRNA) (Figure 1.2A), which are then cleaved in the nucleus by the proteins DROSHA/DGCR8 (Winter et al., 2009). The hairpin structure formed by this cutting is referred to as a pre-miRNA. Afterward, the pre-miRNA is exported from the nucleus to the cytoplasm where it is further processed by DICER resulting in a miRNA duplex lacking a hairpin structure. This duplex miRNA is then bound by AGO2, which catalyzes the removal of the reverse complement of the microRNA (Winter et al., 2009). This produces a mature single-stranded miRNA that can be assembled into a RNA induced silencing (RISC) complex. miRNAs target RISC to mRNAs via the 3'UTR, which contain sequences complementary with the associated miRNAs (He and Hannon, 2004). The extent by which the complementarity matches between the miRNA and mRNA dictates the mechanisms of gene silencing, either via mRNA degradation or translation inhibition. In both cases, gene silencing is achieved (He and Hannon, 2004).

Piwi-interacting RNAs

Piwi-interacting RNAs (piRNAs) are small ncRNA (~24-31 nt in length). piRNAs are derived from distinctive transposons, and contain 5' end uridine and 3' end 2'-O-methyl modification (Iwasaki et al., 2015). piRNAs form complexes with Piwi proteins of the Argonaute family (Figure 1.2B). Similarly to miRNA, piRNAs will pair with

complementary RNA derived from transposable elements. Based on guidance by piRNAs, PIWI proteins target and cleave the transposon RNAs, resulting in silencing of transposable elements during germ line development (Ishizu et al., 2012).

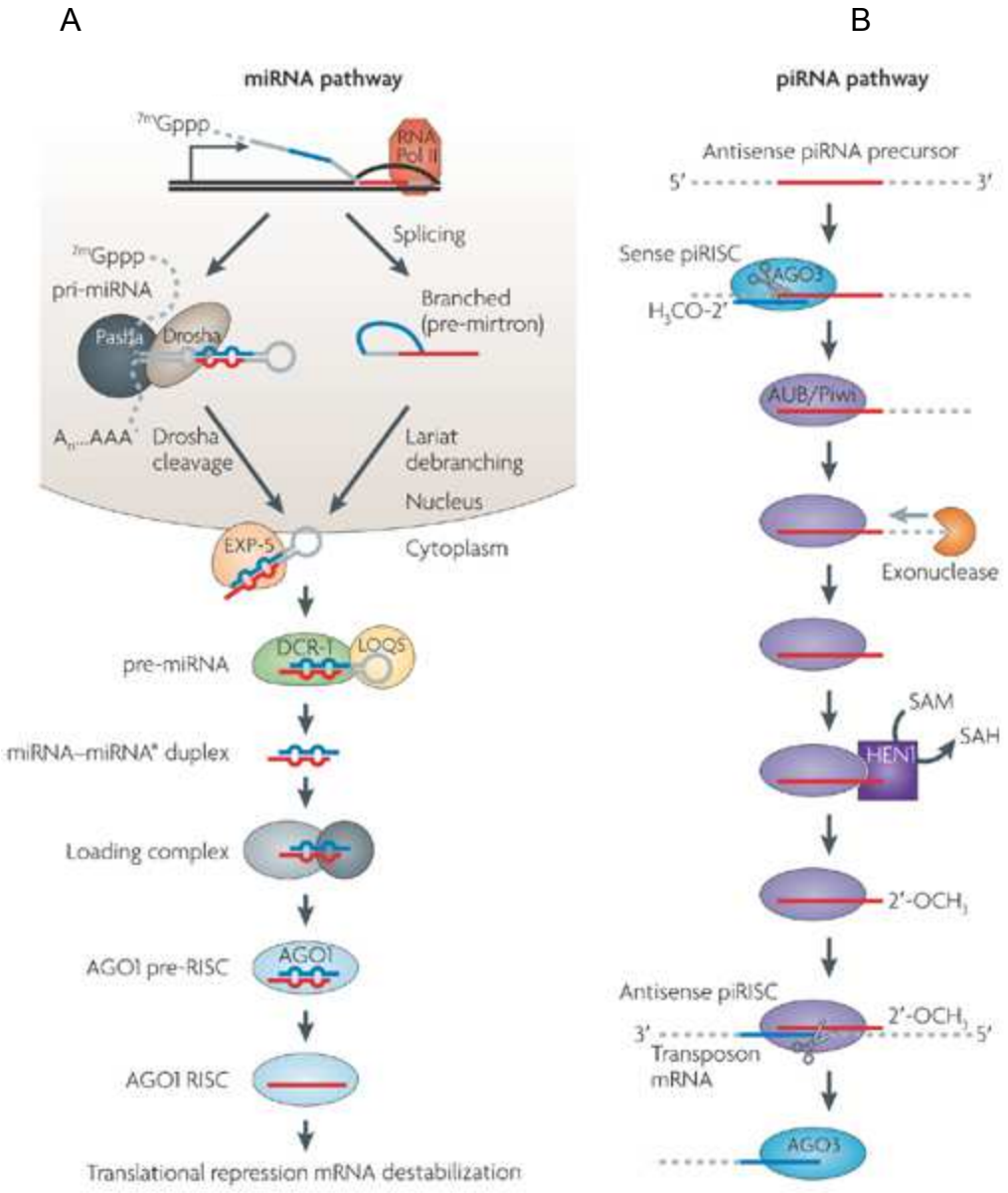


Figure 1.2. The miRNA and piRNA silencing pathways in *Drosophila*. A. miRNAs are transcribed to generate a primary miRNA(pri-miRNA), which produces a precursor miRNA (pre-miRNA) by Drosha cleavage. Pre-miRNA are exported to the cytoplasm and formed complexes with AGO1, which leads to translational repression. B. piRNAs are loaded to form piRISC complex, which targets transposon mRNA, cleaves transposon, and results in transposon silencing. (Ghildiyal and Zamore, 2009)

Long noncoding RNAs

Long ncRNAs (lncRNAs) are defined as non-protein coding RNAs with a length greater than 200 nucleotides. Even though the cut-off number for lncRNA classification is arbitrary, it helps separate lncRNAs from distinct small regulatory ncRNAs. Thousands of lncRNAs in the mammalian transcriptome have been predicted to exist by next-generation sequencing (Wilusz et al., 2009). Many lncRNAs are considered to be the key regulators in various cell processes, including transcription, chromatin remodeling, translation, etc.

lncRNA PLAYS REGULATORY ROLES WITH SOME CHROMATIN-MODIFIERS

In recent years, the theme of long noncoding RNAs controlling gene activity by mediating epigenetic mechanisms has become much more pervasive. Broadly, lncRNAs exercise the roles of recruiters, decoys, stimuli, and scaffolds, or some combination thereof. This section will provide examples of lncRNAs driving gene regulation through the activity of chromatin-modifiers.

G9a

G9a is a histone methyltransferase that deposits repressive methyl marks on H3K9. Its essential role is highlighted by lethality and severe growth defects observed in G9a-deficient embryos (Tachibana et al., 2002). It has been suggested that G9a targets transcriptionally active euchromatin regions, as opposed to repressive pericentric heterochromatin. G9a targeting has been associated with regulation of genes important to development (Zylicz et al., 2015).

Three lncRNAs, Kcnq1ot1, Air, and ROR, have been suggested to interact with G9a, either via the recruitment model or decoy model. Kcnq1ot1 is a 91 kb transcript,

transcribed from the antisense strand of the Kcnq1 gene by RNA pol II (Pandey et al., 2008). It is exclusively localized in the nuclear compartment with moderate stability. In a ChIRP (Chromatin Isolation by RNA Purification) study, Kcnq1ot1 was shown to interact with chromatin. Additional RIP experiments used antibodies raised against G9a to pull out Kcnq1ot1 in a lineage-specific manner. This is consistent with a lineage-specific difference in the H3K9me3 modification in the Kcnq1 gene. The G9a-Kcnq1ot1 interaction also contributes to imprinting in the mouse placenta. A similar recruiting mechanism has been observed with Air lncRNAs, which are largely unspliced and retained in the nucleus. It has been shown that Air is involved in silencing clusters of multiple imprinted genes in cis on chromosome 17 in mice (Wagschal et al., 2008). This silencing mechanism involves Air recruiting G9a to the paternal Slc22a3.

In the case of lncRNA ROR, it has been demonstrated that ROR evicts G9a (Figure 1.3). Human ROR (Regulator of Reprogramming) is 2.6 kb in length, and true to its name, it functions by reprograming human induced pluripotent stem cells (iPSCs) and shares regulatory miRNAs with the TFs OCT4, SOX2, and NANOG. Fan et al. show that ROR occupies and activates the TESC promoter by repelling the histone G9a methyltransferase and promoting the release of histone H3K9 methylation (Fan et al., 2015) . This decoying mechanism leads to a reduction in tumor growth and metastasis.

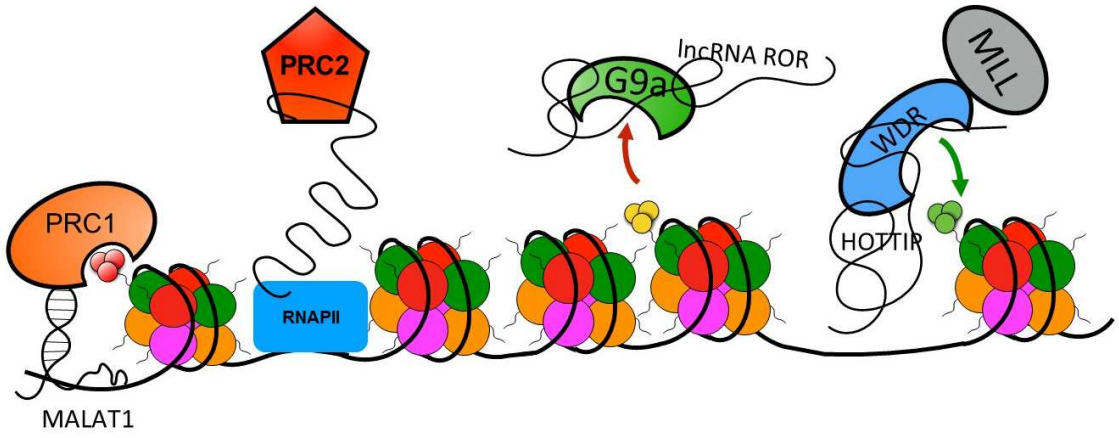


Figure 1.3 lncRNAs regulate transcription through histone modifiers. Polycomb repressive complex 1(PRC1) interacts with lncRNA, either TUG1 or MALAT1. These interactions regulate methylation status and localization of PRC1. Polycomb repressive complex 2 (PRC2) is inhibited by binding lncRNA or nascent pre-mRNA. lncRNAs Kcnq1ot1, Air, and ROR regulate the activity of G9a, which is enzyme that methylates H3K9. HOTTIP interacts with the WDR5-MLL complex and localizes the complex to the 5'HOXA locus.

MLL

MLL (Mixed Lineage Leukemia) is a protein first identified as a functional ortholog of the trithorax (trx) complex in *Drosophila* (Schuettengruber et al., 2011). The canonical role of the MLL protein is to methylate H3K4, a trigger of gene activation. And MLL has been shown to be required for the maintenance of activated genes during normal embryogenesis, hematopoiesis, and neurogenesis (Lim et al., 2009; McMahon et al., 2007; Morey et al., 2015). The essentiality of MLL is underscored by the embryonic lethality of MLL-knockout mice.

How MLL might be recruited to specific genomic loci still remains to be fully elucidated. However, one study has shed light on a possible mechanism involving a lncRNA called HOTTIP (Wang et al., 2011), which appears to play a role in the trafficking of MLL to specific HOXA genes. Specifically, Wang and colleagues show that HOTTIP interacts with the WDR5-MLL complex and localizes the complex to the 5'HOXA locus

(Figure 3). Quite remarkably, a follow-up investigation identified a single residue (F266) on WDR5 that is necessary for RNA-binding and indispensable for gene activation. A similar mechanism has been proposed with the lncRNA HoxBlinc (Deng et al., 2016), which recruits the Set1/MLL complex. This recruitment is followed by the transcriptional activation of the Hoxb gene, thus regulating cardiac/hematopoietic differentiation. These studies together highlight the profound role of lncRNA binding in the regulation of active chromatin states.

Another compelling aspect of the scaffold-like property of some lncRNAs is the ability of a single transcript to bind multiple chromatin-modifier complexes. For instance, the Fendrr lncRNA is specifically expressed in the nascent lateral mesoderm in a developing embryo and has been reported to interact with both MLL and PRC2 complexes (Grote et al., 2013). Fendrr targets complexes to specific promoters to alter the epigenetic landscape. Such epigenetic changes lead to attenuation of the expression of transcription factors, which are important in lateral mesoderm development. Therefore, as knowledge about the factors binding lncRNAs increases, it will become important to begin investigating how different factors engage in functional cross-talk.

HP1

Heterochromatin protein 1 (HP1) was first characterized in *Drosophila* as localizing to heterochromatin and being involved in position-effect variegation (Delattre et al., 2000; James and Elgin, 1986). HP1 binds methyl marks on H3K9 and elicits chromatin packaging and gene silencing. Early embryonic lethality in *Drosophila* is caused by the loss of HP1. And in humans, the loss of HP1 has been shown to correlate with metastatic breast cancer (Akhtar and Becker, 2000).

The HP1 protein contains a conserved N-terminal chromodomain, followed by a variable hinge region, and finally a conserved chromoshadow domain at the C-terminus. The chromodomain has been suggested to be an RNA-binding module; for example, the MOF histone acetyltransferase in *Drosophila* specifically interacts with roX RNA via its chromodomain (Akhtar and Becker, 2000), and these protein-RNA interactions are responsible for the recruitment of MOF to the X chromosome in male *Drosophila*. Evolutionarily, HP1's chromodomain shares homology to the MOF variant, therefore it has been speculated that RNA acts similarly to recruit HP1 to pericentromeric loci. Indeed, one early study found that the RNase treatment of cells induces dispersion of HP1 from the pericentomeric foci (Muchardt et al., 2002) . Furthermore, replenishing RNase-treated cells with purified nuclear RNA rescues the pericentric structures. Later it was shown that HP1 directly binds nuclear RNA using the electrophoretic mobility shift assay (EMSA). These early studies collectively suggested that the mechanism of HP1-mediated chromatin compaction in cells involves RNA actors. However, they did not offer a functional connection between specific RNAs and HP1.

In 2011, Maison and colleagues reported that strand-specific centromeric RNAs (transcribed in the forward direction) co-localize with HP1 in mouse cells (Maison et al., 2002). Using HP1 ChIP experiments, it was observed that HP1 is enriched at the genomic regions encoding the centromeric RNAs. This study helped confirm a particular link between the subnuclear localization of RNA transcripts and HP1 recruitment. Interestingly, Maison et al. also revealed that the post-translational SUMOylation of HP1 actively promotes binding of the protein to the purine-rich sense RNA transcripts, and the combinatorial effects of SUMO and RNA binding together initiate targeting of HP1 to

pericentric heterochromatin. This work reveals how post-translational modifications of chromatin-readers and remodelers might regulate their intrinsic binding properties and, subsequently, recruitment.

In addition to RNA having a recruitment role that captures free HP1 to pericentric heterochromatin, an antagonizing 'eviction' role of lncRNA has also been proposed. One such study has identified a lncRNA called BORDERLINE in *S. pombe* (Keller et al., 2013) which, when processed into short RNAs, evicts HP1 and prevents the spreading of HP1 and histone H3K9 methylation beyond the pericentromeric repeat region.

At pericentric and telomeric regions of chromosomes, heterochromatin formation is orchestrated by a series of interactions involving Suv39h, HP1, and Suv4-20h. Current literature shows that Suv39h methylates H3K9, which serves as a precursor to binding of HP1. Upon binding of methylated H3K9, HP1 recruits Suv4-20h by direct protein-protein interaction. Then Suv4-20h proceeds to establish H4K20me3 marks. Alternative ways of targeting Suv4-20h to H4K20 have been proposed. One study provides evidence implicating lncRNA in this mechanism. Specifically, pre-rRNA antisense transcripts (PAPAS) bind pre-rRNA coding regions and recruit Suv4-20h2 in quiescent cells (Bierhoff et al., 2014). This recruitment promotes H4K20me3-mediated transcriptional silencing of rDNA. In addition, the authors observed a similar scheme at retrotransposon elements, where lncRNA triggers H4K20me3 and transcriptional repression.

PRC1

The Polycomb repressive complex 1 (PRC1) has a core that consists of four proteins: Bmi1, HPH, Ring1, and CBX. The chromodomain of CBX binds to trimethylated histone H3 lysine 27 and initiates the direct catalysis of H2A119 ubiquitination. Such

ubiquitination marks have been thought for years to recruit PRC2 in hierarchical fashion and subsequently enforce gene silencing and chromatin compaction. However, recent studies have provided alternative models that reveal emerging roles for PRC1. Details can be found in the review (Gil and O'Loghlen, 2014).

Regarding PRC1, an early study by Bernstein and colleagues suggested that CBX proteins bind RNAs *in vitro* (Bernstein et al., 2006). Around 2010, two more studies provided mechanistic insights into the functional connections between lncRNA and CBX proteins. In particular, the antisense lncRNA, ANRIL, which is transcribed from the INK4b/ARF/INK4a tumor suppressor locus, recruits PRC1 to that specific locus for transcriptional repression via a direct interaction with the CBX7 subunit. Such repression regulates senescence and proliferation of prostate cancer cells. Interestingly, the authors observed a possible ternary complex consisting of H3K27me3-ANRIL-CBX7 (Yap et al., 2010).

Another CBX subunit of PRC1 that binds lncRNA is CBX4. Some lncRNA transcripts known to interact with CBX4 include TUG1 and MALAT1/NEAT2 (Figure 1.3). These CBX4-RNA interactions stimulate the SUMOylation of the E2F1 growth factor (Yang et al., 2011), a post-translation modification that results in increased cellular proliferation. Intriguingly, the methylation status of CBX4 appears to dictate whether CBX4 interacts with TUG1 or with MALAT1/NEAT2, with the unmethylated variant binding the latter. Given that TUG1 and MALAT1/NEAT1 exhibit differential subnuclear localization, with TUG1 being localized to Polycomb bodies (PcGs) and MALAT1/NEAT2 located in interchromatin granules (ICGs), the methylation status of CBX4 can therefore dictate where PRC1 traffics in the subnuclear environment. This work provides a clear

example of how lncRNAs can act as scaffolds to organize nuclear architecture and influence recruitment of chromatin-modifier complexes.

P300/CBP

P300 and CREB-binding protein (CBP) are two highly homologous and conserved proteins that have intrinsic histone acetyltransferase (HAT) activity, which plays a critical role in regulating gene expression through lysine acetylation of histone H3 (Vo and Goodman, 2001). These proteins act as transcriptional coactivators for a number of nuclear genes. Unlike other histone acetyltransferases, p300/CBP are able to acetylate all four histones both *in vitro* and *in vivo* (Bannister and Kouzarides, 1996). Therefore, p300/CBP are capable of coupling with a variety of transcription factors during chromatin remodeling. Not surprisingly, p300/CBP are involved in a wide array of basic cellular processes, such as DNA damage repair and cell proliferation, and are inherently crucial for embryonic development and cancer.

For the first time in 2015 (Postepska-Igielska et al., 2015), p300/CBP were implicated to interact with lncRNA. The study presents a fascinating model where the antisense lncRNA Khps1 forms a DNA/RNA triplex with the SPHK1 promoter and these triplexes recruit CBP/p300. This recruitment triggers an open chromatin structure, followed by binding of transcription factors, and eventually leads to the activation of SPHK1 transcription.

Recently, a genome-wide analysis of p300/CBP binding to RNA using PAR-CLIP was published (Bose et al., 2017). Bose and colleagues suggest that RNA transcribed locally directly interacts with CBP and stimulates catalytic HAT activity, thereby promoting

gene expression. They also suggest that eRNAs at enhancers may interact with p300/CBP and control transcription activation.

Lysine-specific demethylase 1 (LSD1) is a protein responsible for removing mono- and di-methyl modifications from H3K4 and H3K9 of histones, and it plays pivotal roles during embryonic development and cancer (Foster et al., 2010). This is highlighted by a variety of tumors that display LSD1 overexpression (Hino et al., 2016). LSD1 is the first identified histone demethylase and has been found to associate with a number of transcriptional corepressor complexes including CoREST, CtBP, and a subset of HDAC complexes (Hino et al., 2016). The lncRNA HOTAIR, transcribed from the HoxC locus, has been reported to interact with LSD1 and also with PRC2. This illustrates the scaffold-like function of some lncRNAs, which has been studied in detail in the case of yeast telomerase RNA (Tsai et al., 2010). Curiously, the TERRA RNA transcribed from telomeres has been shown to interact with LSD1. TERRA is bound by LSD1 at TRF2-depleted telomeres, and the RNA promotes the physical interaction between LSD1 and MRE11 (Zappulla and Cech, 2004). This physical interaction stimulates MRE11 nuclease activity and consequently stimulates removal of 3'G-strand overhangs at uncapped telomeres.

RNA-BINDING PROTEINS RELATED TO THIS THESIS

FUS

Amyotrophic lateral sclerosis (ALS) is a progressive motor neuron disease characterized by loss of the upper and lower motor neurons of the spinal cord; most patients die within 2 to 5 years after diagnosis (Da Cruz and Cleveland, 2011). The

pathogenesis of ALS remains unclear and satisfactory therapeutic targets have yet to be identified. The majority of ALS is sporadic; however, approximately 10% of ALS patients have a familial form (fALS) of the disease, exhibiting a classic Mendelian inheritance pattern (Fiesel and Kahle, 2011). One gene, accounting for ~5% of fALS, is *FUS.*

FUS, Fused in Sarcoma (also known as Translocated in liposarcoma, TLS), is a 53 kDa RNA-binding protein, mainly located in the nucleus and is expressed ubiquitously (Kiernan et al., 2011). The N-terminal region has a highly unusual low-complexity amino acid composition, enriched in Ser, Tyr, Gln and Gly residues, while the C-terminal region contains potential RNA-binding domains. These domains include a RNA-binding domain (RRM), and a C2/C2 zinc finger domain flanking by a long stretch of arginines and glycines (Da Cruz and Cleveland, 2011) (Figure 1.4A). A R/H/K$X_{2\text{-}5}$PY nuclear localization signal (NLS) resides at the extreme C-terminus and known pathological fALS-mutations are found in this region (Fiesel and Kahle, 2011). These mutations lead to redistribution of FUS to the cytoplasm and FUS has been detected in the cytoplasmic aggregates of both neurons and glial cells of fALS patients (Deng et al., 2010; Hewitt et al., 2010).

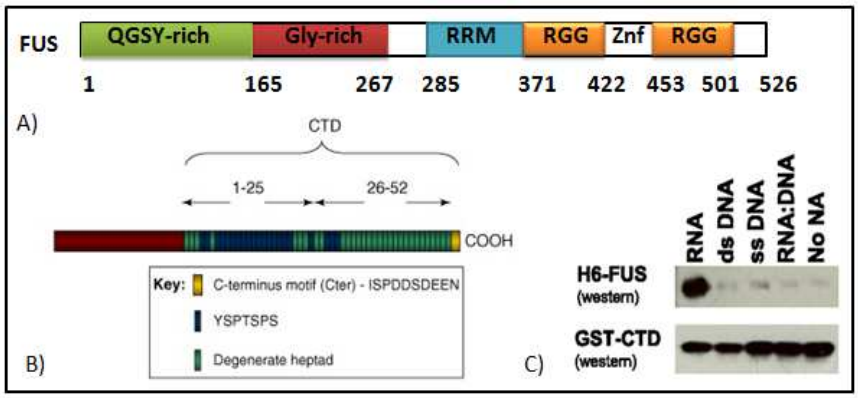


Figure 1.4. A) Schematic Representation of FUS B) Schematic Structure of the CTD of RNAPII (Chapman et al., 2008). C) FUS-CTD interaction in an RNA-dependent manner (Schwartz et al., 2012).

The normal functions of FUS are not fully understood but FUS has been implicated in many processes associated with RNA production and many RNA targets have been identified (Lagier-Tourenne et al., 2010); however, the feature in RNA that FUS recognizes has been the subject of disagreement. Which roles of FUS depend on its RNA-binding properties is also an open question. Characterizing RNA-binding specificities of FUS will help gain more insight into RNA-FUS interactions and FUS functions. One group has utilized an *in vitro* SELEX analysis and determined that the FUS-binding motif is GGUG (Lerga et al., 2001). However, some RNAs with no GGUG motif are able to specifically bind to FUS (Hoell et al., 2011). More recently, high throughput sequencing has discovered many RNA targets of FUS within the mammalian genome (Hoell et al., 2011; Ishigaki et al., 2012; Lagier-Tourenne et al., 2012). Based on these studies, FUS-binding regions of these RNAs have been reported to readily form secondary structures (Hoell et al., 2011; Ishigaki et al., 2012; Lagier-Tourenne et al., 2012) and to be enriched either in G/C nucleotides (Ishigaki et al., 2012; Lagier-Tourenne et al., 2012) or A/U nucleotides (Hoell et al., 2011). Nevertheless, these reported enrichments are present only in less than 10% of the FUS-binding sites. These studies suggest that FUS binding specificity is complicated and both sequence and structure specificities of RNAs may be required for FUS binding. In addition, FUS contains an RNA-recognition motif (RRM) and zinc-finger domain (Znf) that have each been reported to have some ability to bind RNA (Deal et al., 2010; Liu et al., 2013; Nguyen et al., 2011). Some residues in each of these domains have been suggested to potentially contribute to binding RNA (Iko et al., 2004; Liu et al., 2013). However, many questions remain concerning the amino acids within FUS involved in these interactions. Understanding how

these RNA targets bind FUS and defining the interface between these targets and FUS may shed light on understanding the RNA-binding properties of FUS.

FUS has been shown to be associated with RNA-polymerase II transcription complexes (Bertolotti et al., 1996) and been suggested to play roles in regulating transcription by RNAPII and RNAPIII (Tan and Manley, 2010; Tan et al., 2012). In addition, recent studies have indicated that RNA plays roles in FUS regulation of transcription. One study has reported that FUS inhibits the acetyltransferase activity of CREB-binding protein (CBP) and p300 on the cyclin D1 promoter (Wang et al., 2008). This inhibition of histone acetylation is dependent on the expression of noncoding RNAs and leads to reduced transcription of the cyclin D1 gene (Wang et al., 2008). Furthermore, studies from our lab have demonstrated that FUS is associated with the C-terminal domain (CTD) of RNAPII (Schwartz et al., 2012). The CTD is located in the largest subunit (RPB1) of RNAPII and is unique to RNAPII among DNA-dependent RNA polymerases (Chapman et al., 2008). In mammals, the CTD comprises 52 repetitive heptapeptides with the consensus sequence $Y_1S_2P_3T_4S_5P_6S_7$ (Figure 4B) (Muñoz et al., 2010). Phosphorylation of Ser2 and Ser5 are tightly connected to transcriptional elongation and also with pre-mRNA processing (Corden et al., 1985; Muñoz et al., 2010). FUS not only binds to the CTD, but also regulates phosphorylation of Ser2 in the CTD at thousands of human genes (Schwartz et al., 2012). Our lab has utilized an *in vitro* pull-down experiment and demonstrated that the FUS-CTD interaction is RNA-dependent (Figure 4C). However, how RNA is involved in this interaction and the mechanism by which RNA modulates this interaction remain to be further determined. Elucidating the mechanism will facilitate the understanding of the roles that FUS plays in transcription regulation.

PRC2

Polycomb repressive complex-2 (PRC2) is absolutely required for epigenetic processes during embryonic development and cancer. PRC2 is a multiprotein complex, which deposits methyl groups onto the histone H3 at position K27 (Cao et al., 2002; Müller et al., 2002a). Such methylations result in inactivation of genes and repressed chromatin. The crucial role of PRC2 is highlighted by early embryonic lethality in mice after the deletions of these genes (O'Carroll et al., 2001). Because of their significance, PRC2 has become a high-priority drug target. However, how PRC2 is targeted to specific loci of chromatin and how it alters gene expression during the developmental processes are still not fully understood.

PRC2 complex binds H3K27me3 marks of histones, stimulating the catalytic activity of PRC2 (Figure 1.5A) (Hansen et al., 2008; Margueron and Reinberg, 2011). H3K4me3 and H3K36me3 marks, normally associated with active chromatin, are also recognized by PRC2 (Figure 1.5B) (Schmitges et al., 2011a; Yuan et al., 2012a). These recognitions, on the other hand, result in inhibition of PRC2 catalytic activity. The current big question puzzling the field is how PRC2 is initially recruited to these chromatin loci (Figure 1.6).

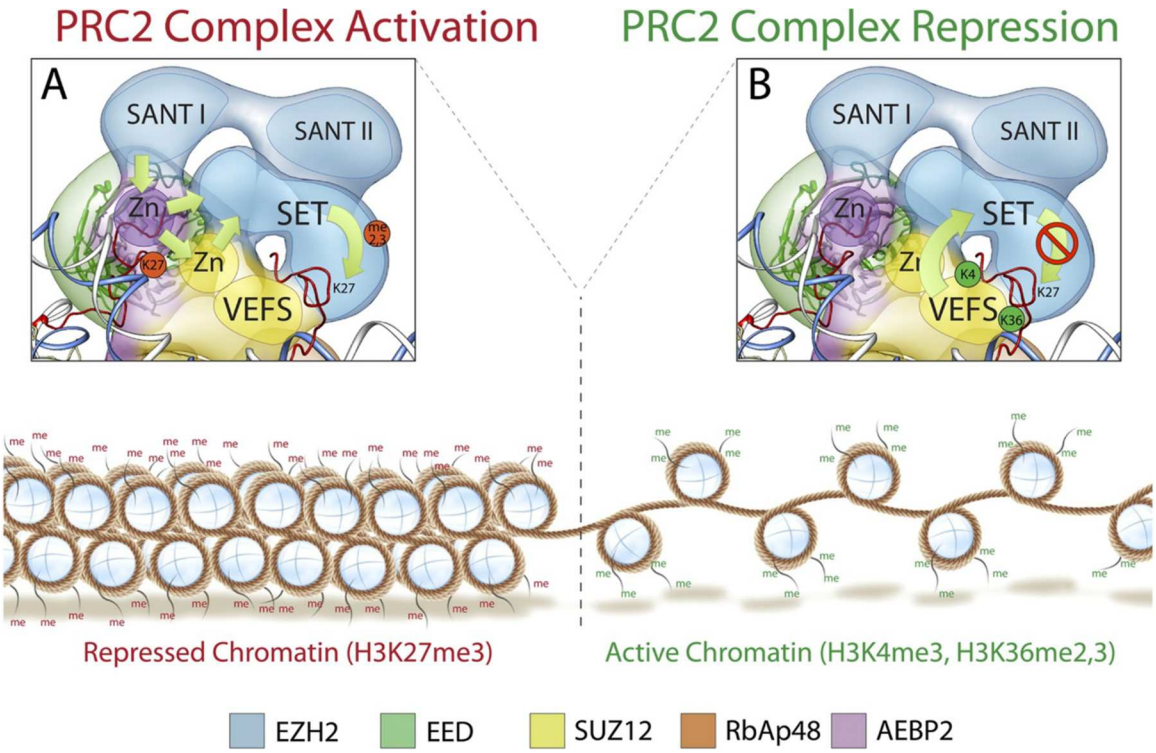


Figure 1.5. Allosteric regulation of PRC2 upon reading different histone marks. A. H3K27me3 marks are detected by the EED component, which triggers allosteric changes of subunit EZH2 and stimulates methyltransferase activity of PRC2. B. H3K4me3 and H3K36me3 active marks are recognized by the VEFS domain of SUZ12, which leads to an allosteric change and inhibitions of EZH2. (Ciferri et al., 2012)

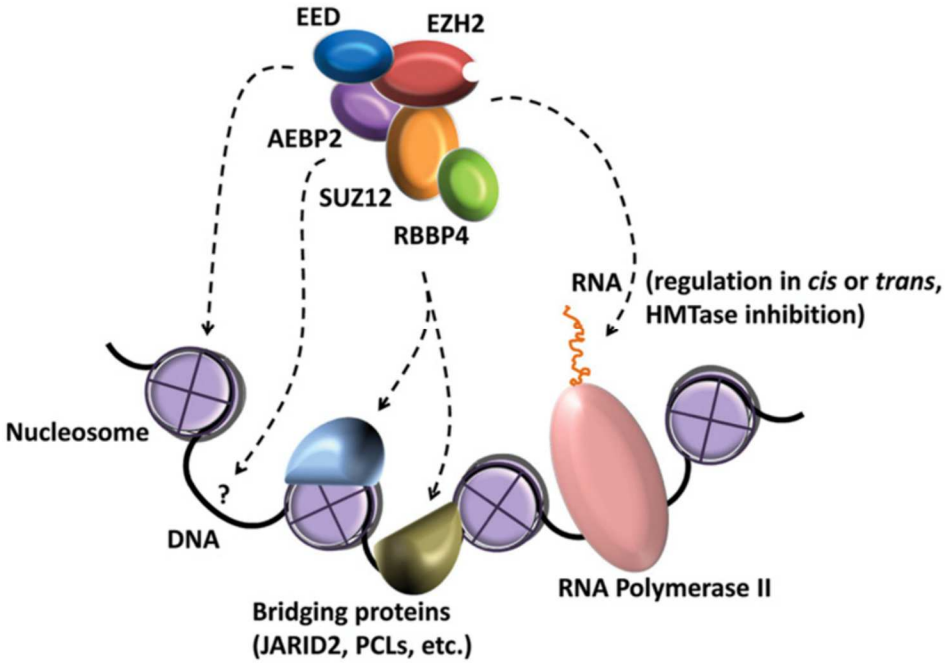


Figure 1.6 Multiple factors suggested to recruit PRC2 to chromatin in mammals. (Davidovich and Cech, 2015)

In the fruit flies, PRC2 recognizes specific Polycomb Response Element (PRE) and via these recognitions, PRC2 is recruited to chromatin (Ringrose and Paro, 2007). However, in mammals, information regarding of PRC2-specific recruiters is still missing. At 2007, a pioneer study has identified, one lncRNA (long-non coding), referred as *HOTAIR*, involved in recruiting PRC2 to the HOXD locus and other loci of chromatin (Rinn et al., 2007). Following that study, another lncRNA, being called *Xist*, has been shown to target PRC2 to the mammalian X chromosome (Zhao et al., 2008a). Based on these studies, a simple model, RNA recruiting PRC2 to the chromatin, has been proposed. In the model, PRC2 has been suggested to interact with RNA specifically and therefore been directly recruited to the chromatin loci. However, paradigm has been shifted when in our lab, purified human PRC2 was found to bind diverse lncRNAs and mRNAs *in vitro* with similar affinities (Davidovich et al., 2013a, 2015a). This “promiscuous” binding pattern is also consistent with thousands of RNAs associated with PRC2 *in vivo* (Kaneko et al., 2013a; Kanhere et al., 2010; Khalil et al., 2009; Zhao et al., 2010). A new model, referred as “scanning” has been proposed. In this model, promiscuous RNA binding by PRC2 allows PRC2 to scan the chromatin environment and then deposit methylation marks. However, recent studies have reported RNA inhibits the histone methyltransferase activity of PRC2 (Cifuentes-Rojas et al., 2014a; Kaneko et al., 2014a). Based on these studies, an “eviction” model has been suggested, where RNA acts as a decoy to strip PRC2 away from the chromatin. Researchers have also suggested coexistence of different models.

One important biological system requiring PRC2 and lncRNAs is X-chromosome inactivation (XCI). Female mammalian cells inactivate one of their two X chromosomes to equalize X-linked gene expression with that of males. One chief mediator of XCI is the

17 kb lncRNA Xist (X-inactive specific transcript), which is transcribed from the Xist gene on one X-chromosome (Augui et al., 2011). While the Xist RNA is capped, spliced, and polyadenylated, it remains exclusively retained in the nucleus and coats one X-chromosome in cis. This RNA coating initiates XCI, coinciding with the exclusion of RNA pol II and the eventual silencing of most gene expression on that X-chromosome. The possibility that Xist RNA might directly recruit PRC2 has been much discussed (Brockdorff, 2013; Davidovich and Cech, 2015).

Recent studies have provided intriguing evidence for indirect recruitment of PRC2. McHugh and colleagues identified RNA-binding proteins associated with Xist RNA by UV-crosslinking followed by purification and quantitative mass spectrometry (RAP-MS) (McHugh et al., 2015). Strikingly, the top hit was not PRC2, but instead three factors: SHARP/Spen, SAF-A and LBR. In brief, subsequent work discovered that SHARP/Spen binds directly to Xist RNA and recruits the SMRT-HDAC3 complex. This leads to deacetylation of histone H3 and facilitates the enrichment of PRC2. As validation, the knockdown of SHARP and HDAC3 reactivates gene expression on the silenced X chromosome and leads to the depletion of PRC2. Another study utilized a similar approach, but instead relied on formaldehyde-crosslinking under non-denaturing conditions (CHIRP-MS) (Chu et al., 2015). This study identified 81 proteins that directly or indirectly interact with Xist. Furthermore, the authors proposed that heterogeneous nuclear ribonucleoprotein K (hnRNP K) might have a direct influence on H3K27me3 and Polycomb recruitment.

What is consistent between the two studies is that neither found components of PRC2 in the Xist RNA interactome. But it is worthwhile to emphasize that another

proteomics paper used an approach similar to that of McHugh et al., but came to a different conclusion (Minajigi et al., 2015). Specifically, they identified RBBP4/7 as possibly interacting with Xist. While RBBP4 is a known subunit of the PRC2 complex, it is also a component of the NuRD and SIN3 histone deacetylase complexes (Clark et al., 2015; Millard et al., 2016). The study did not identify any of the core PRC2 subunits Ezh2, Suz12 and Eed.

Finally, two independent groups (Moindrot et al., 2015; Monfort et al., 2015) utilized elegant genetic screens to determine which genes are required for Xist-mediated XCI. Both groups identified the RNA-binding protein Spen as the top hit. In total, these studies discovered dozens of new factors, each of which warrants further exploration in relation to Xist RNA function.

In summary, this chapter has introduced various RNAs, their functions and interacting proteins. Some of the discussion regarding histone modifiers with lncRNAs will be published in my ensuing review article. The remaining chapters in this thesis will further explore the binding interactions in detail for FUS and PRC2. Chapter II will discuss the binding properties of FUS to RNA and how FUS-RNA interactions play roles in regulating the FUS-CTD interaction. Chapter III will investigate why PRC2 binds RNA promiscuously and elucidate the features of RNAs that are preferentially bound by PRC2. Chapter IV will explore the interactions between PRC2 and nucleosomes, and will focus on the interplay of PRC2 between nucleosome and RNA. Finally, the Appendix investigates interaction of FUS-RNA in different concentrations of salt and discusses automethylation of PRC2.

Chapter II: FUS binds RNA promiscuously *in vitro*

INTRODUCTION

The goal of the work presented in this chapter was to characterize the elements of RNA substrates that are sufficient for FUS binding. Elucidating RNA-binding properties of FUS leads to a better understanding of FUS function in living systems. Jacob Schwartz established a protocol for FUS purification and provided ideas for designing RNA mutants. I was involved in the design of all experiments, their execution, data analysis and manuscript preparation.

FUsed in Sarcoma (FUS, also known as Translocated in LipoSarcoma, TLS), is an abundant nuclear protein that has been implicated in transcription, mRNA splicing and mRNA transport (Da Cruz and Cleveland, 2011; Fiesel and Kahle, 2011; Lagier-Tourenne et al., 2010). Mutations in *FUS* are detected in ~5% of familial ALS (amyotrophic lateral sclerosis) patients as well as in sporadic ALS (Hewitt et al., 2010; Kiernan et al., 2011). ALS is a progressive motor neuron disease characterized by loss of the upper and lower motor neurons (Deng et al., 2010). Patients typically die within three to five years after onset of the disease. Dysregulation of RNA is emerging as a pathogenic mechanism in ALS. Therefore, understanding the biology and biochemistry of the FUS protein may provide insights into how this protein can potentially cause the onset of the disease.

FUS, together with EWS (Ewing's sarcoma) and TAF15 (TBP-associated factor 15) in vertebrates, belongs to the FET or TET family (Lagier-Tourenne et al., 2010). The FUS protein has 526 amino acids and is composed of a SYGQ (serine, tyrosine, glycine and glutamine)-rich region at its N-terminus, an RNA-recognition motif (RRM), multiple RGG (arginine, glycine and glycine)-repeat regions, a C2C2 zinc finger motif, and a

nuclear localization signal (NLS) at its extreme C-terminus. FUS recognition of RNA is mediated by both the RRM and the zinc-finger-containing RGG-Znf-RGG domain (Iko et al., 2004; Liu et al., 2013; Schwartz et al., 2013).

RNA binding has been suggested to be crucial for FUS function. FUS inhibits the acetyltransferase activity of CREB-binding protein (CBP) and p300 on the cyclin D1 promoter (Wang et al., 2008). This inhibition of histone acetylation is dependent on the expression of noncoding RNA *in cis*, and it leads to reduced transcription of the cyclin D1 gene. More generally, our previous work has shown that FUS binds the C-terminal domain (CTD) of RNA polymerase II (RNA Pol II) in an RNA-dependent manner and orchestrates phosphorylation at position Ser2 of the CTD hexapeptide motif (Schwartz et al., 2012, 2013).

Several groups have published RNA sequences that promote FUS binding (Hoell et al., 2011; Ishigaki et al., 2012; Lagier-Tourenne et al., 2012; Lerga et al., 2001; Ray et al., 2013; Schwartz et al., 2013; Takahama et al., 2013). One group has utilized *in vitro* SELEX analysis to identify GGUG as a preferred FUS-binding motif (Lerga et al., 2001). However, some RNAs with no GGUG motif are able to bind to FUS (Hoell et al., 2011). More recently, high throughput sequencing has discovered many RNA targets of FUS within the mammalian genome (Hoell et al., 2011; Ishigaki et al., 2012; Lagier-Tourenne et al., 2012). Based on these studies, FUS-binding regions of these RNAs have been reported to readily form secondary structures (Hoell et al., 2011; Ishigaki et al., 2012; Lagier-Tourenne et al., 2012) and to be enriched either in G/C nucleotides (Ishigaki et al., 2012; Lagier-Tourenne et al., 2012) or A/U nucleotides (Hoell et al., 2011). However,

these reported enrichments represent less than 10% of the FUS-binding regions. These studies suggest that FUS binding is complicated and that both sequence and structure of RNAs may recruit FUS.

Elucidating the nucleic acid targets of FUS is important for understanding its cellular roles. To characterize the features of RNA targets necessary for FUS binding, we have thoroughly evaluated the binding affinities of FUS with all five published RNA motifs and additional sequences, using electrophoretic mobility shift assays (EMSAs). We found that FUS is able to bind all published RNA sequences within a ten-fold range of binding affinities. In contrast to expectation, however, FUS bound other RNAs including fragments of an *Escherichia coli* mRNA with binding constants similar to those of the published motifs. Consistent with promiscuous binding, we demonstrated that FUS binds RNA in a length-dependent manner. Finally, using competition experiments, we found that FUS had only a modest preference for binding ssRNA relative to ssDNA of the same length and sequence. We conclude that FUS has a wide range of nucleic-acid binding ability.

RESULTS

FUS is able to bind many RNAs

Five different RNA sequences have been reported to be preferentially bound by FUS protein (Hoell et al., 2011; Ishigaki et al., 2012; Lagier-Tourenne et al., 2012; Lerga et al., 2001; Ray et al., 2013; Takahama et al., 2013). Among these, GGUG, CGCGC, and GUGGU are suggested to contain a specific sequence motif recognized by FUS (Lagier-Tourenne et al., 2012; Lerga et al., 2001; Ray et al., 2013). On the other hand, Stem-loop and TERRA form unique secondary and tertiary structures suggested to

promote FUS binding (Hoell et al., 2011; Takahama et al., 2013). We hypothesized that FUS may bind one of these RNAs with exceptionally higher affinity than the others. To test this hypothesis, we measured the binding of *E. coli*-expressed FUS protein to eight RNAs including the five published motifs and three negative control sequences (Table 2.1). We also tested prD RNA, one of many human ncRNAs that recruits FUS *in vivo* identified in our previous study (Schwartz et al., 2012). EMSA was performed with increasing concentrations of MBP-FUS protein and a trace amount of end-labeled RNA to measure binding affinities (Figure 2.1A).

Table 2.1. Sequences of RNAs used and the corresponding equilibrium dissociation constants measured by EMSA. Twelve sequences were used to test whether FUS possesses selectivity to bind nucleic acids. These include seven published sequences, four published negative control sequences and the MBP sequence that originates from *E. coli.* Among these sequences, stem-loop, Htelo and TERRA form unique secondary structures. The binding affinity between FUS and each sequence was measured. The calculated K_d values are listed and uncertainties represent the range of two or more replicates.

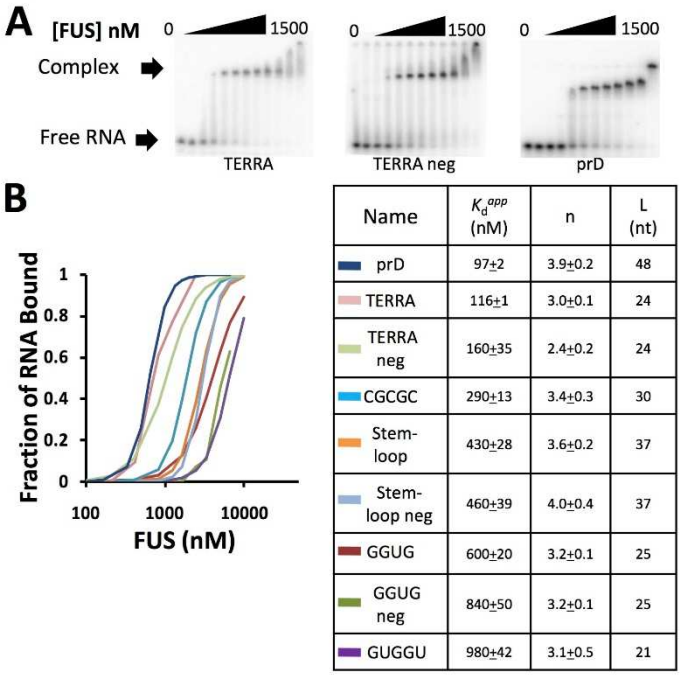
Name	Sequence	K_d^{app} (nM)	Structure formed	Putative negative control	K_d^{app} (nM)
GGUG	UUGUAUUUUGAGCUAGU UU**GGUG**AC	600±20	N.A.	UUGUAUUUUGAGCUAGU UU**CCUC**AC	840±50
GUGGU	CAACUUAGGU**GGUUG**AUUUG A	980±42	N.A.	N.A.	N.A.
Stem-loop	GAUUUAUCUUUAACUACUCAA GAUACUGAACAUGACA	430±28		GAUUUAUCUUUAACUACUCU AUCUUCUGAACAUGACA	460±39
CGCGC	AGGUCUCAGUUCAU**CGCGC**GA GGUUAUAGU	290±13	N.A.	N.A	N.A.
prD	AUUGAGGAGCAGCAGAGAAG UUGGAGUGAAGGCAGAGAGG GGUUAAGG	97±2	N.A.	N.A.	N.A.
MBP1-200	See materials and methods section	56±2	N.A.	N.A.	N.A.
TERRA	$(UUAGGG)_4$	116±1		$UUAGGG(UUAGUG)_2UUAGGG$	160±35
Htelo DNA	$AGGG(TTAGGG)_3$	730±38		$AGGG(TTAGTG)_2TTAGGG$	690±48

Discrete shifted bands were observed, indicating RNA-protein complexes of specific stoichiometry and absence of aggregation. All nine sequences tested were bound by FUS, each with a K_d^{app} in the range between 100 nM and 1000 nM (Figure 2.1B and Figure 2.2B). The similar binding affinities of very different RNAs (e.g., CGCGC, stem-loop, and GGUG) cast doubt on their specificity for binding to FUS. This skepticism was reinforced by the small differences in affinity between three of the proposed motifs and their mutated forms (cf. TERRA and TERRA neg, stem-loop and stem-loop neg, GGUG and GGUG neg). Furthermore, the prD RNA binds FUS as well as any of the other published RNAs but contains none of the motifs (Schwartz et al., 2013).

The EMSA patterns suggested positive cooperativity between FUS and RNA, as it took only two or three protein concentration points to proceed from unshifted RNA to the low-mobility completely shifted complex (Figure 2.1A). We quantified and fit the binding data with the Hill equation, which revealed that FUS bound each sequence with positive cooperativity (Figure 2.1B). The low-mobility complex is thought to contain at least four FUS protein (Schwartz et al., 2013), and the fact that intermediates with one, two or three bound proteins do not accumulate is expected for highly cooperative binding. At higher FUS concentrations, the FUS-RNA complexes shifted more towards the well of the gel. This suggests that additional FUS molecules are associated with the RNA in the highly-retarded species compared to the initial low-mobility FUS-RNA complex. Alternatively, some of these complexes may contain multiple FUS associated with multiple RNAs.

Our MBP-FUS protein was purified from *E. coli* (Figure 2.2A), while one previous publication carried out EMSA with His_6-FUS purified from insect cells (Hoell et al., 2011).

To test for differences in FUS obtained from these expression systems, His_6-FUS purified from insect cells was compared with MBP-FUS purified from *E. coli* by EMSA (Figure 2.2C). The two proteins both formed discrete RNA-protein complexes and the protein concentration necessary to shift half of the radioactively labeled RNA was similar. In both cases, the observation of discrete complexes suggests well-folded protein. Therefore, we used MBP-FUS purified from *E. coli* for all remaining experiments.



Name	K_d^{app} (nM)	n	L (nt)
prD	97±2	3.9±0.2	48
TERRA	116±1	3.0±0.1	24
TERRA neg	160±35	2.4±0.2	24
CGCGC	290±13	3.4±0.3	30
Stem-loop	430±28	3.6±0.2	37
Stem-loop neg	460±39	4.0±0.4	37
GGUG	600±20	3.2±0.1	25
GGUG neg	840±50	3.2±0.1	25
GUGGU	980±42	3.1±0.5	21

Figure 2.1. FUS binds many RNAs. A. A trace amount of TERRA (left), or TERRA neg (middle), or prD (right) was incubated with increasing concentrations of MBP-FUS (0,15,31,62,125,188,250,375,500,750,1000 and 1500 nM). Bindings were analyzed by electrophoretic mobility shift assays (EMSA). **B.** Summary of RNA binding data for MBP-FUS with nine different RNAs. Left, Quantification of Fbound (RNA in complexes per total RNA in lane) as a function of MBP-FUS concentration. Right, the apparent dissociation constant was calculated for each RNA. N and L represents Hill coefficients and length of the RNAs, respectively. Uncertainties represent the range of two or more replicates. Length shows how many nucleotides are in each RNA sequence.

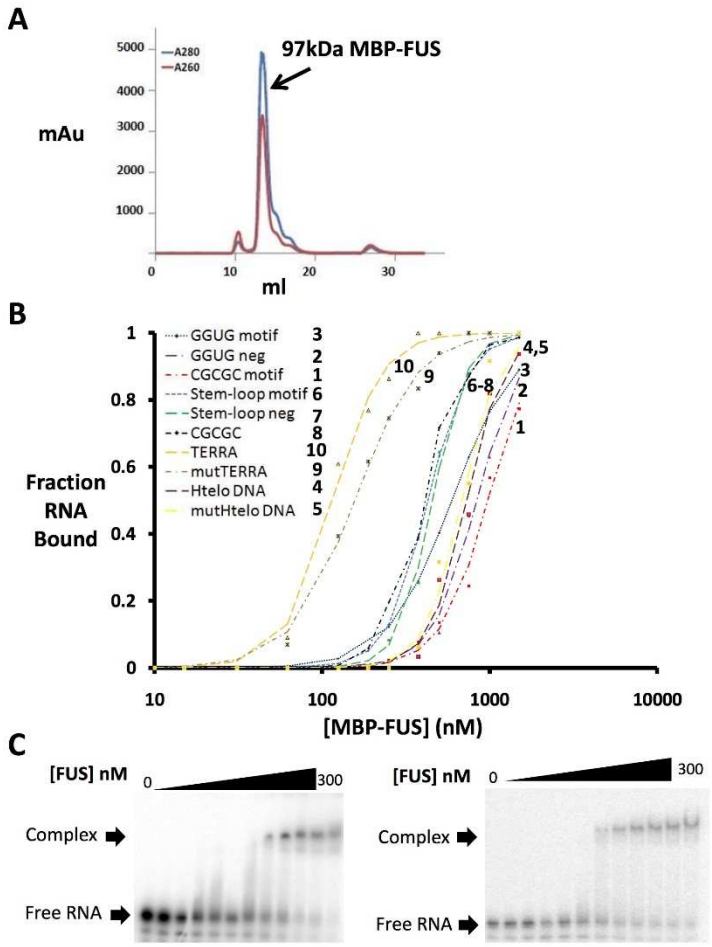


Figure 2.2. RNA binding curves for FUS with different RNAs. A. MBP-FUS purified by Superdex 200 size-exclusion chromatography has a size corresponding to a monomer, based on comparisons to five standard globular proteins (not shown). **B.** RNA binding curves for MBP-FUS with twelve different RNAs. Quantification of Fbound (RNA in complexes per total RNA in lane) as a function of MBP- FUS concentration. **C.** His6-FUS purified from insect cells (left) and MBP-FUS purified from *E. coli* (right) were tested for binding to prD RNA. His6-FUS was expressed in Hi5 cells [11]. The purification was done in the same way as for MBP-FUS.

FUS binds RNA in a length-dependent manner

To further test FUS's specificity for RNA binding, we performed EMSAs with portions of the mRNA for the maltose-binding protein from *E. coli,* an organism that does not possess FUS. Surprisingly, the first 200 nt of the *E.coli* MBP mRNA (MBP1-200) bound FUS with a reasonably high affinity (K_d^{app} =56 nM$\pm$2 nM; Figure 2.3A). The electrophoretic mobility of the RNA-protein complex decreased progressively as the FUS

concentration was increased, suggesting the loading of more and more FUS onto the mRNA and low binding specificity. The Hill coefficient was 4.8±0.1, indicating that multiple FUS proteins bound this non-human sequence in a positively cooperative manner.

Even though MBP-RNA originates from *E. coli*, it was still possible that some sequence or structure hidden in this RNA could have been responsible for promoting FUS binding. To test this possibility, we *in vitro* transcribed a series of MBP RNAs, including RNA containing the first 10 nt (MBP1-10), first 20 nt (MBP1-20), and so on, and then measured their binding to FUS. If MBP1-200 contained some sequence or structure necessary to bind FUS, then there should be a sudden increase in affinity at the length corresponding to the inclusion of the motif. If no such sequence or structure existed in MBP1-200, FUS might bind all the truncated sequences.

As shown in Figure 2.3B, there was no discrete length cut-off for FUS binding, but rather an incremental increase in affinity with increasing RNA length. FUS bound MBP1-20 but not MBP1-10, defining a minimum length for RNA binding. In addition, more FUS bound as the RNA-length increased, seen as the binding curves shifted from right to left, indicating an increase in binding affinity (Figure 2.3B). In other words, K_d^{app} decreased with increasing RNA length. Plotting log (K_d) versus log (RNA length) revealed a linear relationship between dissociation constant and RNA length with a slope of -1 (Figure 2.3C), consistent with promiscuous binding (Davidovich et al., 2013a).

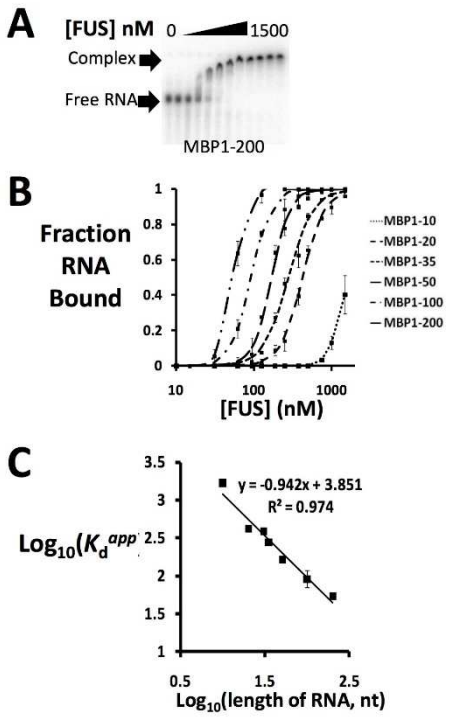


Figure 2.3. FUS binds RNA in a length-dependent manner. A. EMSA of FUS for RNA containing the first 200 bases from *E. coli* MBP mRNA reveals a tight binding affinity between FUS and this RNA. **B.** Binding curves were plotted for FUS and RNAs comprising of 10, 20, 35, 50, 100, 200 bases from *E. coli* MBP mRNA, respectively. Error bars represent the range of two or more replicates. **C.** Further analysis of binding curves in B. Plotting log (K_d^{app}) versus log (RNA length) revealed a linear relationship with a slope of -1.

A specific RNA-binding protein is able to recognize its specific RNA motif within a longer RNA

The conclusions above relied on the assumption that a sequence-specific RNA-binding protein can recognize and bind its specific motif hidden in a long sequence and that shorter RNAs without this motif will no longer recruit the protein to bind (**Figure 2.4A**). To validate this assumption, we substituted the MS2 recognition motif for a portion of the MBP1-200 sequence. The MS2 motif forms a stem-loop structure, recruiting specifically MS2 coat protein with a high affinity (K_d^{app}= 4nM) (Johansson et al., 1998). The position of substituting the MS2 motif was selected from locations where the MS2 RNA can still

be properly folded in the context of the MBP long sequence, using the mFold program to predict RNA secondary structures (**Figure 2.5**).

We *in vitro* transcribed two series of MBP mRNAs, each comprising the first 10 nt, the first 20 nt, and so on from the 5'-end of the mRNA. One series had the 21 nt MS2 motif substituted for nucleotides 71-92 and the other did not. RNA sequences in the series lacking the MS2motif bound MS2 coat protein weakly with a micromolar binding affinity (**Figure 2.4B**). In contrast, in the series containing the MS2 motif, the binding affinity increased substantially for the third, fourth and fifth RNAs, which contained the MS2 motif. Even though the binding affinity did not increase all the way down to 10nM, the K_d^{app} we measured for binding to an isolated 21 nt MS2 motif, the K_d^{app} still dropped dramatically from several micromolar into the 100 nM range (**Figure 2.4B**). The reduced affinity of the coat protein for the MS2 site in the context of long RNAs could be due to sampling of multiple RNA conformations, some of which disrupt the motif. Nevertheless, this control experiment supports our conclusion that if there were a specific motif embedded in a long RNA molecule, it could be found by testing the binding of a series of truncated versions of the long RNA.

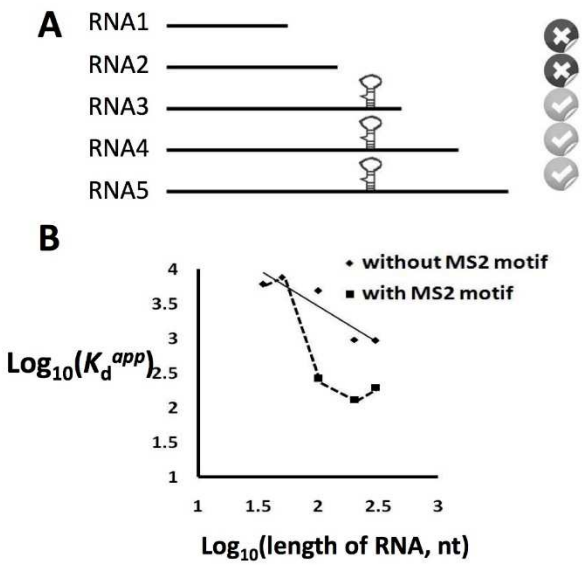


Figure 2.4. MS2 protein is able to find and recognize its specific MS2 motif with the system used. A. Schematic representation to show the system used to study a RNA-binding protein recognizes and binds a motif if such a motif is hidden in the longer RNA sequence. **B.** Plotting log (Kdapp) versus log (RNA length) for RNA sequences with or without the MS2 motif showed more than an order of magnitude drop in K_d^{app}. Dash line connecting the second point to the third point shows the difference between the K_d^{app} without and with MS2 motif. Each point represents the average of at least two or more replicates.

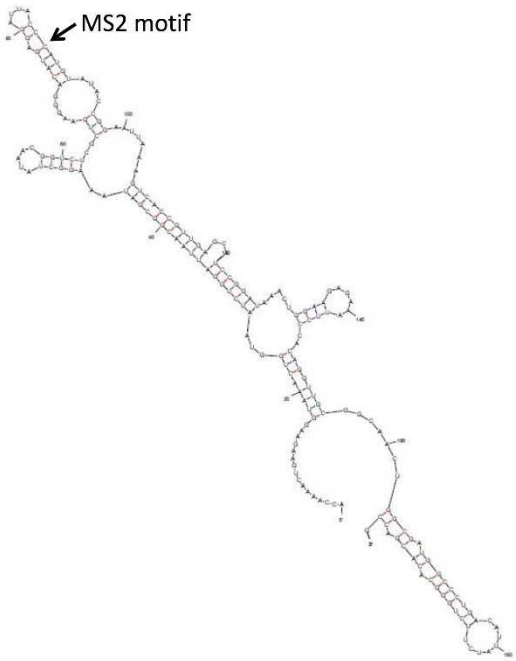


Figure 2.5. The secondary structure of MBP1-200 RNA with an MS2 motif. The motif was substituted for nucleotides 71-92 as predicted by the mFold program.

FUS binds RNA without requiring a specific sequence or structure

We have observed that FUS is capable of binding many RNAs *in vitro* without dramatic differences in binding affinity and that it binds RNAs such as prD that do not contain any of the published motifs. To further understand the binding of FUS to prD, we divided the 48mer prD RNA into two 24mer RNAs named 5'prD and 3'prD (Figure 2.6A). Both of the two RNAs bound FUS with similar affinities and the gel patterns of the two RNAs look identical (Figure 2.6A), indicating the two RNAs are able to recruit FUS similarly. The reduced affinity for these half molecules relative to prD is expected from the length dependence shown in Figure 2.3.

Next, we synthesized four additional RNAs by mutating six consecutive nucleotides of 5'prD into a stretch of adenines (Figure 2.6B). As expected, each of the mutants bound FUS, and the K_d^{app} values were in a narrow range between 400 and 800nM, suggesting neither sequence nor structure in the 5'prD was required for FUS binding. 24mer polyA also bound FUS with an affinity (846 nM) somewhat lower than those of the 5'prD mutants, further indicating FUS does not differentiate dramatically among RNA sequences or structures for binding. In conclusion, FUS binds RNAs with diverse sequences and structures, and when same-length RNAs are compared the binding affinities are within an order of magnitude.

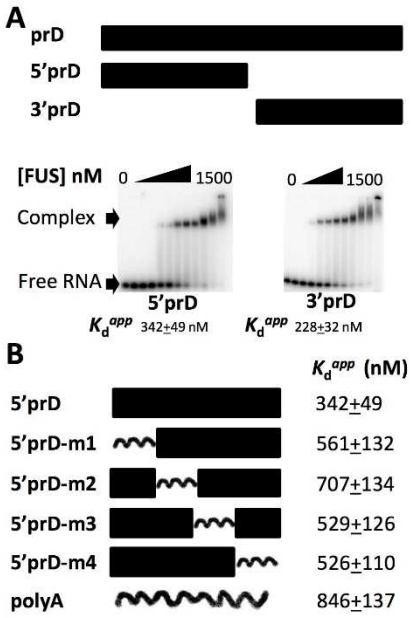


Figure 2.6. FUS binds the same-length RNA without dramatic difference in binding affinities. A. Top: landscape view of 5'prD and 3'prD. Bottom: EMSAs between 5'prD or 3'prD and FUS to show both RNA bind FUS equally well. **B.** FUS binds all the four mutates of 5'prD as well as the same length polyA. Four mutants of 5'prD were generated by sequentially substituting six nucleotides with six adenines. The binding between each mutate and MBP-FUS was measured and the dissociation constants were calculated based on three replicated.

Contributions of FUS Domains to RNA Binding

The FUS protein has an N-terminal low-complexity (LC) domain, capable of forming fibrous structures in isolation (Kato et al., 2012), followed by two RNA-binding domains – an RRM and a zinc finger (Figure 2.7A). RGG domains surround each RNA-binding domain. Both the RRM-containing and zinc finger-containing (RGG-Zn-RGG) domains bound prD RNA (Figure 2.7B), as reported previously (Iko et al., 2004). The LC domain did not bind RNA (data not shown). We found that the binding affinities of the two domains were very similar to each other and to that of the full-length protein; this may indicate that only one RNA-binding domain is available for binding in the full-length protein. Full binding cooperativity required both RNA-binding/RGG domains (see full-length FUS and del-LC);

RGG-Zn-RGG bound with intermediate cooperativity, and LC-RRM bound with full affinity but no cooperativity.

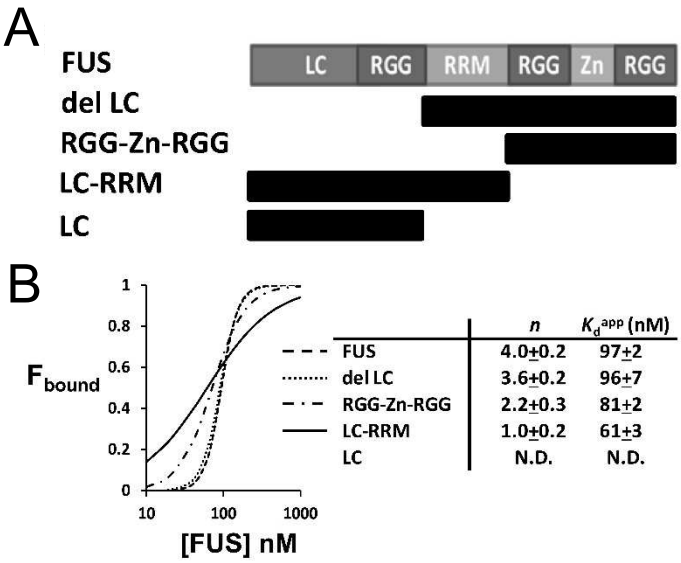


Figure 2.7. Contributions of FUS Domains to RNA Binding. A. Domain structure of FUS protein and truncated versions of the protein that were tested for RNA binding. **B.** Summary of RNA-binding data for truncation mutants of FUS. Uncertainties represent the range of 2 or more replicates. The LC domain bound RNA so weakly that binding parameters were not determined (N.D.).

FUS binds single-stranded DNA and double-stranded nucleic acids with reduced affinity

Previous literature has shown that FUS binds single-stranded DNA (ssDNA) and double-stranded DNA (dsDNA) (Baechtold et al., 1999; Bertolotti et al., 1996). To test how well FUS interacts with different forms of nucleic acid, we synthesized six forms of prD each having the same sequence and length (48 nt) as prD RNA. These forms included sense and anti-sense ssRNA, sense ssDNA, dsRNA, dsDNA, and DNA/RNA hybrid (Table 2.2). As expected, FUS bound anti-sense prD RNA. Sense ssDNA bound FUS but with a 3-fold weaker K_d^{app} compared to RNA. This is consistent with the observation of weaker affinities with Htelo DNA than Terra RNA (Table 2.1). Weaker

binding to DNA suggests that the 2' hydroxyl group may contribute to binding FUS. In addition, FUS bound dsDNA and DNA/RNA hybrid more weakly. Surprisingly, full-length FUS was capable of binding dsRNA with lower K_d^{app}, indicating FUS is not only a single-stranded nucleic acid binding protein.

Table 2.2. Equilibrium dissociation constants (K_d^{app}, nM) for different nucleic acid forms binding to full-length FUS as well as each truncation. Different nucleic acid forms of prD RNA were generated and the dissociation constants between each nucleic acid substrate and full length FUS were measured in EMSA. Each form had the same length as prD RNA. As-prD RNA was the anti-sense complementary strand of prd RNA and ssDNA had the same sequence as prD RNA. DNA/RNA was a hybrid of ssDNA and as-prD RNA. Two truncations of full length FUS, including LC-RRM (comprising of low-complexity domain and RNA recognition motif) and RGG-Znf-RGG (containing zinc finger regions flanked by arginine-glycine-glycine rich regions) were made fused with a N-terminal MBP tag. Binding between each truncation and each nucleic acid substrate was quantified. All the below numbers are in nanomolar units and uncertainties listed represent the range of two or more replicates.

Name	prD RNA	As-prD RNA	ssDNA	dsRNA	dsDNA	DNA/RNA
FUS	97±2	79±4	280±26	350±38	1200±110	980±100
LC-RRM	60±3	280±22	220±32	2100±60	1400±170	1200±250
RGG-Znf-RGG	81±2	260±32	290±29	1100±80	1000±81	1100±110

As expected, both FUS truncations also bound anti-sense prD RNA, although with 3-fold lower affinity compared to full-length FUS. Sense ssDNA bound both truncations with similar K_d^{app} as anti-sense prD RNA. In addition, each truncation bound dsDNA and DNA/RNA hybrid very weakly. The FUS truncations, however, bound dsRNA with three- to six-fold higher K_d^{app} compared to full-length FUS, suggesting dsRNA may require both RRM and RGG-Znf-RGG for optimal recognition.

If the same nucleic-acid binding site(s) on FUS bound both ssRNA and ssDNA, then ssDNA should compete for ssRNA binding. To test this, we performed competition assays with radioactively labeled RNA and increasing concentrations of unlabeled ssRNA

or ssDNA using EMSA. The radiolabeled prD RNA formed a slow-migrating complex with FUS (Figure 2.8.). A hundred-fold excess of cold ssRNA resulted in displacement of FUS-bound labeled ssRNA (Figure 2.8., lane 5). However, a ten times greater fold excess of cold ssDNA than cold ssRNA was required to displace FUS-bound labeled ssRNA, which is consistent with ssDNA having a weaker binding affinity to FUS. Thus, ssRNA and ssDNA bind to the same site(s) on FUS or they bind to mutually exclusive sites.

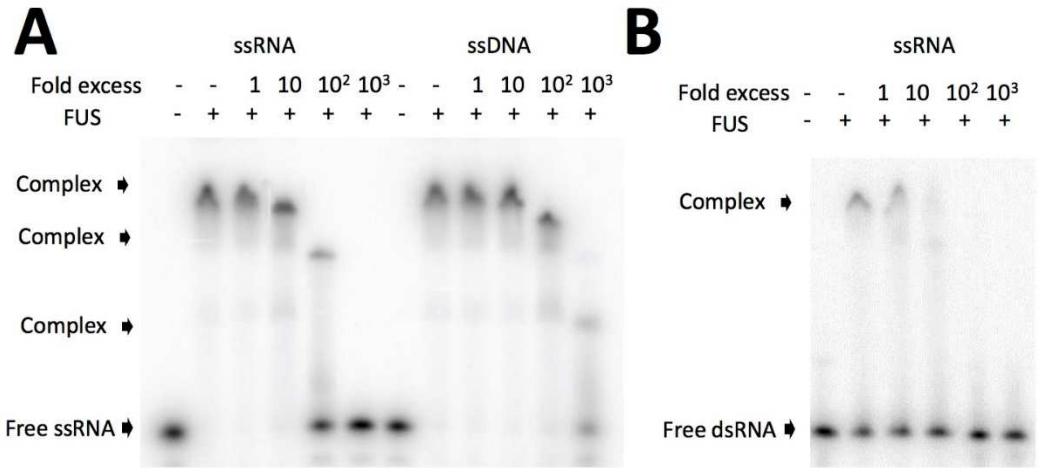


Figure 2.8. FUS has mutually exclusive binding sites for single- stranded RNA and DNA and dsRNA. A. ssRNA and ssDNA bind FUS mutually exclusively. Increasing amounts of unlabeled ssRNA or ssDNA were mixed with 20nM radiolabeled ssRNA and the mixture was added to samples containing 350 nM FUS. **B.** ssRNA and dsRNA compete for binding to FUS. The same binding competition assay was performed as in A except unlabeled ssRNA was mixed with 20nM radiolabeled dsRNA.

Interestingly, new discrete FUS-RNA complexes were observed with addition of cold ssRNA or ssDNA (Figure 2.8., lanes 5 and 12), which migrated more rapidly than the original FUS-RNA complexes. This is consistent with multiple FUS proteins being associated with one RNA molecule in the original complexes. The addition of the cold nucleic acid may strip one or two FUS molecules from the original complexes, resulting in the new discrete, fast-migrating complexes. These partially saturated RNAs appear

when the competitor nucleic acid concentration is similar to the FUS protein concentration, so there is little if any free FUS protein available to bind the RNA probe. Note that such partially saturated complexes did not appear in the binding experiments with trace amounts of RNA probe (Figure 2.2 and 2.8), where FUS is in excess and available for cooperative binding.

We performed the same competition assays between radioactively labeled dsRNA and cold ssRNA. The shifted FUS-dsRNA complexes migrated slightly slower than the FUS-ssRNA complexes. Excess amounts of cold ssRNA competed away the FUS-bound labeled dsRNA (Figure 2.8., lane 4), indicating binding of FUS to dsRNA and ssRNA is mutually exclusive. As dsRNA has a weaker binding affinity than ssRNA, about ten-fold excess of cold ssRNA was required to displace FUS-bound labeled dsRNA compared to the hundred-fold excess of cold ssRNA required to displace FUS-bound labeled ssRNA.

FUS-RNA interaction promotes FUS binding to the CTD of RNAPII

We have previously shown that FUS binds CTD of RNAPII, we next investigated whether FUS interacting with RNA is associated with FUS-CTD interaction. The prD RNA was titrated in FUS-CTD pull down assay and FUS pulled down CTD at an RNA concentration similar to the K_d^{app} between prD and FUS (Figure 2.9A). Other RNAs also potentiates FUS-CTD interaction (data not shown). One interpretation for the FUS-RNA-CTD interaction would be that RNA bridges between FUS and CTD. However, we tested binding between CTD and RNA with EMSA and we have observed no binding (Figure 2.9B). Additionally, in the presence of 1 μM CTD, K_d^{app} between FUS and prD did not change (Figure 2.9C), suggesting that CTD does not sandwich between FUS and RNA.

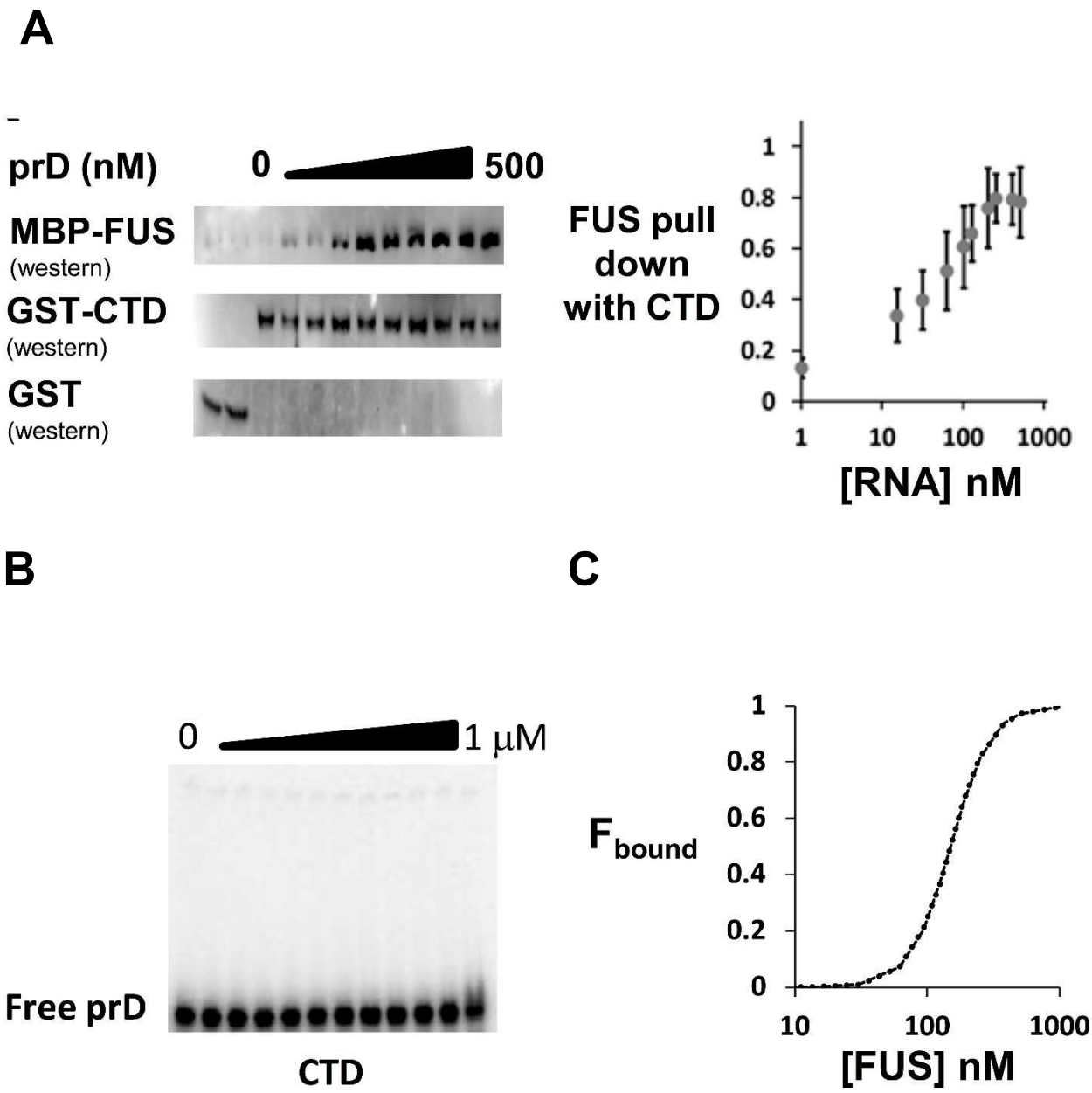


Figure 2.9 RNA Promotes FUS Binding to the CTD of RNAP. **A**. Left, a representative western from pulldown of MBP-FUS by GST-CTD, stimulated by titrating in the prD RNA. Right, MBP-FUS binding to GST-CTD as a function of prD RNA concentration averaged from 3 replicates (error bars represent standard error). **B.** Binding curve of GST-CTD for the prD RNA shows that CTD does not bind RNA. **C.** EMSA of FUS for RNA with (dashed line) or without (dotted curve) the presence of 1 μM GST-CTD reveals no change in the K_d^{app} of FUS for RNA.

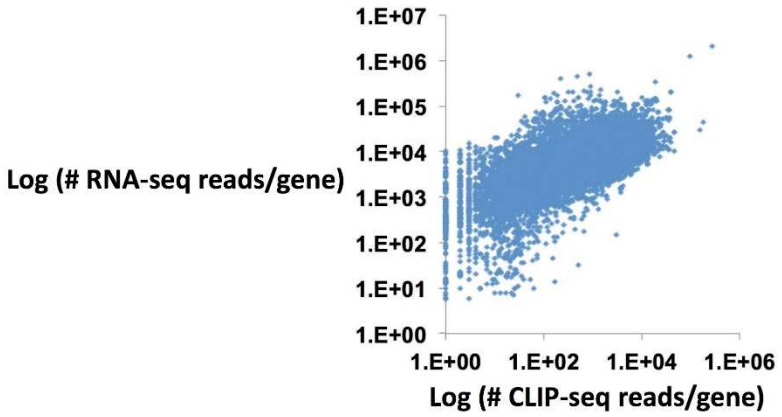


Figure 2.10. Comparing the RNAs bound to FUS *in vivo* (CLIP-seq) with relative RNA abundance (RNA-seq). Both CLIP-seq and RNA-seq were generated from whole cell extract of HEK 293T/17 cells (Schwartz et al., 2012). Data show a trend towards more abundant RNAs being more frequently bound by FUS (R=0.18). Note that FUS is a nuclear protein, so even in the extreme case that nuclear RNA abundance was the only determinant of FUS binding, that correlation would be imperfectly represented in RNA-seq data on whole- cell RNA.

DISCUSSION

The prevailing idea in the FUS field has been that a specific sequence or structure in RNA allows FUS to bind, which led to a simple model by which some motif in RNA recruits FUS functions. Several studies have published RNA motifs (Hoell et al., 2011; Lagier-Tourenne et al., 2012; Lerga et al., 2001; Ray et al., 2013; Takahama et al., 2013) that recruit FUS, but their binding has not been tested side-by-side with the same FUS protein preparation. Here, we synthesized all these motifs and measured their K_d^{app} for binding to FUS. We found that FUS is able to bind all of these published motifs, but it also binds to their respective negative controls that disrupt the motifs with only slightly reduced affinity. Furthermore, even *E. coli* MBP mRNA binds FUS, and we provide evidence that this is not due to a hidden motif. We conclude that FUS has the ability to bind many RNAs

in vitro with similar binding affinities and without requiring a well-defined sequence or structure.

Promiscuous binding of FUS to RNA is consistent with many observations *in vivo*. For example, FUS crosslinks to many thousands of RNAs in cells, including 5'UTRs, 3'UTRs and introns (Hoell et al., 2011; Ishigaki et al., 2012; Lagier-Tourenne et al., 2012; Rogelj et al., 2012; Schwartz et al., 2012). Comparison of the RNAs bound to FUS *in vivo* (CLIP-seq) with relative RNA abundance (RNA-seq) shows a definite trend towards FUS binding to abundant RNA (Figure 2.10), consistent with a promiscuous binding component. However, the correlation between FUS-binding and RNA abundance is weak; this is not unexpected, because our biochemical results in no way preclude specific binding of FUS to certain RNAs *in vivo*, *e.g.*, through cooperation with site-specific RNA-binding proteins.

The first published motif (GGUG) was determined based on a SELEX analysis (Lerga et al., 2001). The traditional SELEX technique depended on very low-throughput cloning and DNA sequencing technology, and insufficient coverage of all possible sequences or PCR amplification advantages of certain sequences may have contributed to the identification of the motif. Recently the high-throughput RNAcompete method (Ray et al., 2013) identified the CGCGC motif; we did not find preferential FUS binding to this motif, but this difference could be due to our testing full-length FUS whereas the RNAcompete study tested a truncated FUS (containing the RRM and additional 50 amino acids flanking the N- and C-terminus of the RRM). Two other published motifs were found by transcriptome-wide PAR-CLIP or CLIP-seq analysis (Hoell et al., 2011; Lagier-

Tourenne et al., 2012). *In vivo* conditions are different from *in vitro* conditions, and FUS binding to the identified motifs *in vivo* may be influenced by other RNA-binding proteins selectively bound to certain RNAs, either precluding or cooperating with FUS binding. Lagier-Tourenne *et al*. (Lagier-Tourenne et al., 2012) conclude that the presence of their GUGGU motif was neither necessary nor sufficient for FUS binding, in agreement with our *in vitro* analysis.

The human telomeric-repeat TERRA RNA (Azzalin et al., 2007) can fold into an intramolecular G-quadruplex structure (Williamson et al., 1989).In agreement with the report by Oyoshi *et al.* (Takahama et al., 2013), we found that FUS binds TERRA, and in fact it bound with a higher affinity than the other 24-mer RNAs we tested. However, we were unable to confirm the importance of the G-quadruplex structure of TERRA for FUS binding, because in our hands the authors' mutant "TERRA neg" RNA (containing mutations that would prevent G-quadruplex formation) bound with only slightly reduced affinity (table 2.1). This may perhaps be due to only one FUS concentration being selected for analysis (Takahama et al., 2013) rather than the full binding curves reported here.

We have demonstrated that length of RNA contributes to the binding affinities. However, it is not the sole factor to determine the binding affinities, as we observed 2- to 7-fold differences in binding affinities for same-length RNAs. For example, the 24-mers in Figure 2.6 had K_d^{app} values ranging from 228 nM to 846 nM, and TERRA bound even more tightly with K_d^{app} = 116 nM. This behavior fits the definition of "promiscuous binding": binding to many RNAs without the requirement for an obvious or well-defined protein-

binding motif and with affinities that are not enormously different (Davidovich et al., 2013a).

The Hill coefficients measured here were greater than 1.0, indicating positive cooperatively. Positive cooperativity typically occurs when one protein binding increases the binding affinity for the next protein-binding event via protein-protein interaction. Therefore, multiple FUS proteins appear to be associated with each RNA molecule. This is consistent with our previous conclusion that prD RNA binds at least four FUS molecules (Schwartz et al., 2013).

FUS contains two domains that contribute to binding to RNA, the RRM and RGG-Znf-RGG domains (Iko et al., 2004; Schwartz et al., 2013). The RRM in FUS is structurally similar to other RRMs, adapting a canonical β1 -α1-β2-β3-α2-β4 fold. However, based on the NMR structure of the RRM in FUS, two important aromatic amino acids are missing (Liu et al., 2013). These two aromatic amino acids normally stack with bases and contribute to specific RNA recognition in the canonical RRM, such as in hnRNPA1 (Ding et al., 1999). This suggests FUS may interact with RNA in a different way, e.g., through hydrogen bonding, and lacking these key amino acids may allow FUS to bind many RNAs. We have also demonstrated that both RRM and RGG-Znf-RGG domains bind RNA with similar affinity as full-length FUS. This suggests that FUS may have evolved to have two channels for selecting RNAs, resulting in targeting a much larger variety of RNA.

FUS binds single-stranded DNA. The binding competition results suggest that ssRNA and ssDNA interact with FUS at the same site or overlapping sites. Another possibility is that FUS binding one nucleic acid causes a FUS conformational change that

precludes the binding of the other nucleic acid. FUS binding single-stranded DNA is consistent with several observations in the literature. Similar *in vitro* competition results have been shown in other studies (Bertolotti et al., 1996; Tan et al., 2012). In fact, FUS protein has been isolated and purified by affinity chromatography on ssDNA (Calvio et al., 1995). The ability of FUS to bind DNA also fits well with its function associated with DNA damage repair. FUS, being one of the earliest proteins recruited to DNA lesions, interacts directly with the DNA repair factors, such as HDAC1 and DNA-PK (Mastrocola et al., 2013; Wang et al., 2013). Such interactions are required for successful DNA repair (Mastrocola et al., 2013; Wang et al., 2013).

FUS is reported to directly interact with the CTD of RNA polymerase II (Kwon et al., 2013; Schwartz et al., 2012). FUS-CTD interactions require RNA, either nascent transcript or noncoding RNA (Schwartz et al., 2012). Cooperative binding properties of FUS to RNA may facilitate the formation of higher-order assemblies. These assemblies orchestrate the phosphorylation status of C-terminal domain of RNAPII (Kwon et al., 2013; Schwartz et al., 2012, 2013). Being able to bind many RNAs may allow FUS to target a larger diversity of genes near their transcription start sites, which is consistent with our previous model (Schwartz et al., 2012). Also, FUS may bind to long introns and facilitate their splicing (Lagier-Tourenne et al., 2012).

FUS has the intrinsic ability to bind many RNAs without substantial differences in binding affinity, so what determines the FUS interactome *in vivo*? Many other proteins, including hnRNP proteins and splicing factors, are associated with nuclear RNAs, and these may preclude FUS binding. On the other hand, some FUS-partner proteins may

enhance FUS binding specificity. Similar partner-assisted specificity has been documented for protein-nucleic acid interactions including the homeobox (Hox) family transcription factors, which bind DNA rather nonspecifically by themselves but gain high DNA sequence specificity and enhanced affinity when paired with a Exd protein partner (Moretti et al., 2008). One can also speculate that post-translational modifications, such as arginine methylation (Dormann et al., 2012), may change the conformation of the protein, conveying preferences for certain RNAs. Future work is required to elucidate the requirements for FUS-RNA interaction in living systems.

MATERIALS AND METHODS

Protein expression and purification

The initial FUS expression plasmid was acquired as a gift from the M. G. Rosenfeld lab (UCSD). We added sequences encoding a His_6-MBP tag at the N-terminus of FUS, generating the His_6-MBP-FUS construct [9]. This expression plasmid was transformed into BL21 cells (Life Technologies) and grown in a 5ml LB-Amp culture overnight. Cultures (1L) were inoculated and grown at 37 ºC to OD600 > 0.8, followed by induction with 0.5 mM IPTG and growth for an additional 3-5 hours at 37 ºC. Bacterial cells were pelleted at 6000 rpm for 10min and lysed in lysis buffer (1 M KCl, 50 mM Tris pH 7.4, 10 mM imidazole, 1mM $CaCl_2$, 5% glycerol, 1% NP40, 1.5 mM β-mercaptoethanol, 1 M urea, micrococcal nuclease (New England Biolabs M02474; 1000 Kunitz Units per gram of cell pellet), followed by sonication (15sec on and 15sec off) for a total time of 1 min. Lysates were cleared by centrifugation at 17500 g for 20 min at 4 ºC and supernatants were incubated for 1 h with Ni-sepharose beads at 4 ºC. Beads were pelleted at 2000 rpm for 2 min and washed four times in wash buffer (1 M KCl, 50 mM Tris pH 7.4, 10 mM

imidazole, 1.5 mM β-mercaptoethanol, 1 M urea), followed by one time in wash buffer supplemented with 25mM imidazole. Protein (hereafter called MBP-FUS) was eluted in wash buffer supplemented with 250 mM imidazole. Highly concentrated FUS tends to form aggregates, but the MBP tag keeps FUS soluble. MBP tag itself does not bind RNA [18]. Thus, MBP tags were not cleaved after purification. Our purified MBP-FUS protein was analyzed by size exclusion chromatography, showing high purity and solubility (Suppl. Fig.1A). After purification, the A260/280 ratio was typically in the range 0.57-0.60, indicative of nucleic acid-free protein. The final purified protein (1mg) was treated with micrococcal nuclease (200 Kunitz Units) and 1.0 mM $CaCl_2$ to ensure the complete elimination of nucleic acid, and the nuclease was then inactivated by chelating the Ca^{2+} with 1.0mM EGTA; we determined that the residual inactivated micrococcal nuclease did not affect the measurement of FUS-RNA binding (data not shown). Protein was aliquoted with 10% glycerol, snap frozen in liquid nitrogen, and stored at -80 ºC.

The percent active protein was determined by titrating MBP-FUS into trace amounts of hot prD RNA and 200 nM cold prD RNA (48 nt) as the substrate. It typically required 6.7 proteins to bind each RNA molecule, which in the case of a 1:1 complex would mean that the protein was only 15% active. However, on the basis of our previous estimate of four FUS molecules per 48 nt RNA [9], the FUS preparation was 60% active (calculated as 4/6.7). Here we present K_d^{app} values based on active protein assuming a 1:1 complex so that they are directly comparable to those presented in our previous publication [9], understanding the real K_d values are likely to be four-fold higher. Other FUS publications do not report measuring or correcting for the percent active protein.

In vitro transcription of MBP RNA

For MBP1-10 and MBP1-20, DNA templates were synthesized by IDT. Complementary strands were annealed and used for *in vitro* transcription. The templates were as follows:

MBP 1–10 Forward, TAATACGACTCACTATAGGGAGACCAAAACTG

MBP 1–10 Reverse, CAGTTTTGGTCTCCCTATAGTGAGTCGTATTA

MBP 1–20 Forward, TAATACGACTCACTATAGGGAGACCAAAACTGAAGAAGGTAA

MBP 1–20 Reverse, TTACCTTCTTCAGTTTTGGTCTCCCTATAGTGAGTCGTATTA

For other longer MBP RNA constructs, DNA templates were amplified from plasmid pFastBac1 containing the maltose-binding protein (MBP) gene from *E. coli.* The primers used were as follows:

T7-Forward,

TAATACGACTCACTATAGGGAGACCAAAACTGAAGAAGGTAAACTGGTAATCTGG

MBP 1–50 Reverse, CCTTTATCGCCGTTAATCCAGATTAC

MBP 1–100 Reverse, TTCCGGTATCTTTCTCGAATTTCTTACCG

MBP 1–200 Reverse, CGGTCGTGTGCCCAGAAGATAATG

MBP 1–300 Reverse, GTAACGTACGGCATCCCAGGTAAAC

For the MBP RNA bearing the MS2 motif, only MBP1-100 Reverse was changed as follows:

MBP 1-100 MS2 Reverse: TTCCGGTATACATGGGTAATCCTC

DNA templates for transcription were generated by PCR with high-fidelity DNA polymerase (Phusion, NEB). The predicted size of PCR amplicons was confirmed by agarose gel electrophoresis with appropriate DNA size markers. The *in vitro* RNA transcription reactions were set up as described [19]. Briefly, the reactions were carried out with T7 RNA polymerase and were incubated at 37 ºC for 2 h, followed by inactivation at 65 °C for 20 min. A trace amount of radioactive CTP [α-^{32}P] was included in the reaction to body-label the transcripts. The reactions were spun down and supernatants were treated with RQ1 RNase-free DNase (M6101, Promega) to digest DNA template. The digestions were stopped by addition of 50mM EDTA. Unincorporated nucleotides were removed by a microspin G25 column (GE Healthcare 27-5325-01). Then, the reactions were mixed with formamide dye, incubated 5 min at 95 °C and loaded onto a 10% w/v 29:1 acrylamide:bis 7 M urea gel. The bands containing radiolabeled RNA were excised from the gel and the RNAs were eluted for 1 h at 4 °C by 0.3 M sodium acetate, pH 5.2. The eluant was precipitated with glycogen and ethanol at -80ºC overnight and the body-labeled RNAs were quantified by liquid scintillation counting.

End-radiolabeling RNA

prD RNA, GGUG RNA and other RNA oligos were synthesized by IDT and end-radiolabeled with γ-^{32}P-ATP using T4 polynucleotide kinase (NEB); incubation was at

37°C for 45 min, followed by inactivation with EDTA. Unincorporated nucleotides were removed and RNA was gel-purified as described for *in vitro* transcription.

Electrophoretic mobility shift assays

In a 20 µl binding reaction, a trace amount of ^{32}P-labeled RNA was incubated with MBP-FUS in binding buffer (50 mM Tris-HCl pH 7.4, 150 mM KCl, 2 mM $MgCl_2$, 2 mM DTT, 0.1mg/ml yeast tRNA, 0.1 mg/ml BSA and trace amount of orange dye) at room temperature for 30 min. A portion of each reaction was loaded onto a 4-20% TBE (Invitrogen EC62252BOX) gel and run at room temperature at 150V for 70 min. Gels were vacuum dried for 60 min at 80 °C and the [^{32}P] radioactive signal was detected by exposure to phosphorimager screens. The signals were acquired with a Typhoon Trio phosphorimager (GE Healthcare) and densitometry was quantified with ImageQuant software (GE Healthcare). Quantified data were fit to a sigmoidal binding curve with MATLAB (MathWorks), allowing calculation of both dissociation constants and Hill coefficients.

For competition assays, an appropriate concentration of unlabeled competitor RNA or DNA was mixed with 5000 cpm radiolabeled RNA of the same sequence in a 20 µl reaction. The binding reaction was performed as described above.

Chapter III: Targeting of Polycomb Repressive Complex 2 to RNA by Short Repeats of Consecutive Guanines

INTRODUCTION

The work presented in this chapter was designed to investigate the features of RNAs that are recognized by PRC2 binding, therefore helping identification of tight RNA binders *in vivo*. I took a leadership role in organizing the project, which involved a great deal of collaboration. All the EMSA experiments were done by Karen Goodrich. UV-crosslinking experiments and protein purifications were performed by Anne Gooding. Richard Paucek performed competition experiments between RNA and G-quadruplex drugs. Daniel Youmans carried out *in vivo* pull-down experiments. Haroon Naeem and Stuart Archer worked on bioinformatic analysis. Thomas Cech, Chen Davidovich, Richard Paucek, Daniel Youmans and I were involved in designing the experiments. I carried out the remaining experiments and analyzed the data, generated figures, and wrote the manuscript with Chen Davidovich and Thomas Cech.

PRC2 is the histone methyltransferase that deposits the H3K27me3 mark of silent chromatin. PRC2 promotes epigenetic silencing and is essential for embryonic development and differentiation, as shown by the embryonic lethality of deletion of several of its subunits in mice (reviewed in (Margueron and Reinberg, 2011)). PRC2-mediated epigenetic silencing can promote cancer by repressing tumor-suppressor genes and its dysregulation can lead to cancer through derepression of oncogenes (Bracken and Helin, 2009; Sauvageau and Sauvageau, 2010). Although substantial efforts have been made to develop anti-cancer therapeutic inhibitors to target its histone methyltransferase activity (Kim and Roberts, 2016; McCabe et al., 2012), the RNA-binding activity of PRC2 has remained poorly understood.

While PRC2 is clearly an RNA-binding protein both *in vitro* and *in vivo* (Khalil et al., 2009; Tsai et al., 2010; Zhao et al., 2008a, 2010), how and why it binds RNA have been the subjects of many alternative views (Brockdorff, 2013; Davidovich and Cech, 2015). According to proposed models, chromatin-bound RNA may recruit PRC2 specifically to sites of its action, or allow PRC2 to scan nascent RNAs more generally to locate possible sites of action, or serve as a decoy to prevent PRC2 from binding active genes and inhibiting its action; and these views are not necessarily mutually exclusive.

There is now general agreement that PRC2 binds natural RNAs promiscuously both *in vitro* and *in vivo* (Beltran et al., 2016a; Davidovich and Cech, 2015; Davidovich et al., 2013a; Kaneko et al., 2013a). We previously referred to “promiscuous” binding as the ability of PRC2 to interact with many RNAs that do not share a well-defined protein-binding motif and with affinities that are not enormously different (Davidovich et al., 2015a). But how would promiscuous binding be possible, considering what is known about protein-RNA interactions? One obvious solution would be if the interaction involved the RNA backbone and were largely driven by electrostatic interactions. But the very modest salt dependence of the affinity of PRC2 to RNA does not support this view (Davidovich et al., 2013a). The best remaining hypothesis is that the binding motif is so common that it occurs at similar frequency in many natural RNA molecules.

The idea that some binding specificity of PRC2 to HOTAIR could be obtained by G-quadruplex was previously considered (Wu et al., 2013), but the authors concluded that the quadruplex RNA was not likely to be significant for EZH2-EED binding and proposed a different secondary structure. A non-quadruplex structure was also proposed for this RNA in a subsequent study (Somarowthu et al., 2015). New information came

from the Reinberg lab, who used a pull-down assay to show that poly (G) but not poly (A) was bound by PRC2 *in vitro* (Kaneko et al., 2014a). This led the authors to speculate that Poly(G) could mimic elements of structured RNAs, including double-stranded segments and single-stranded loops, that could be involved in PRC2-RNA interactions. Despite being an important step forward, this observation was still too general and was therefore insufficient to explain the increased affinity observed between PRC2 to certain RNAs *in vitro* and some cohort of transcripts *in vivo*. Moreover, affinities were not measured, the potential formation of higher-order poly (G) species was not excluded and PRC2 binding to G-quadruplex (G4) RNA structures was postulated but not tested. The standard diagnostic tests for intramolecular G4 structures are cation-specific stabilization (K^+ versus Li^+, where only the former fits into the central cavity and stabilizes the structure) and sensitivity to mutations that break up blocks of G's (Williamson et al., 1989).

Here, we undertake quantitative binding studies with RNAs of defined length and sequence and show that human PRC2 binds G > C,U >> A. The distribution of G's in an RNA also affects binding affinity, and the tightest binders have four blocks of two or more G's; indeed, we find that PRC2 binds specifically to the folded form of G4 RNAs. Analysis of fRIP-seq and ChIP-seq data shows that *in vivo* PRC2 preferably binds to RNA motifs composed of short repeats of consecutive guanines and to DNA sequences coding for them. Double-stranded RNA is not a determinant for binding. Instead, we find that PRC2 preferentially binds G-tracts in RNA, either single-stranded or, preferably, folded G-quadruplex.

RESULTS

Examining the two-hairpin recruitment model of PRC2 *in vitro* and *in vivo*

Two-hairpin motifs in RNAs have suggested to bind PRC2 specifically *in vitro* and *in vivo*. To test this, we generated RNAs with the same length (434nt) but different sequence and we observed similar binding affinities among all of the RNAs (Figure 3.1A-D).

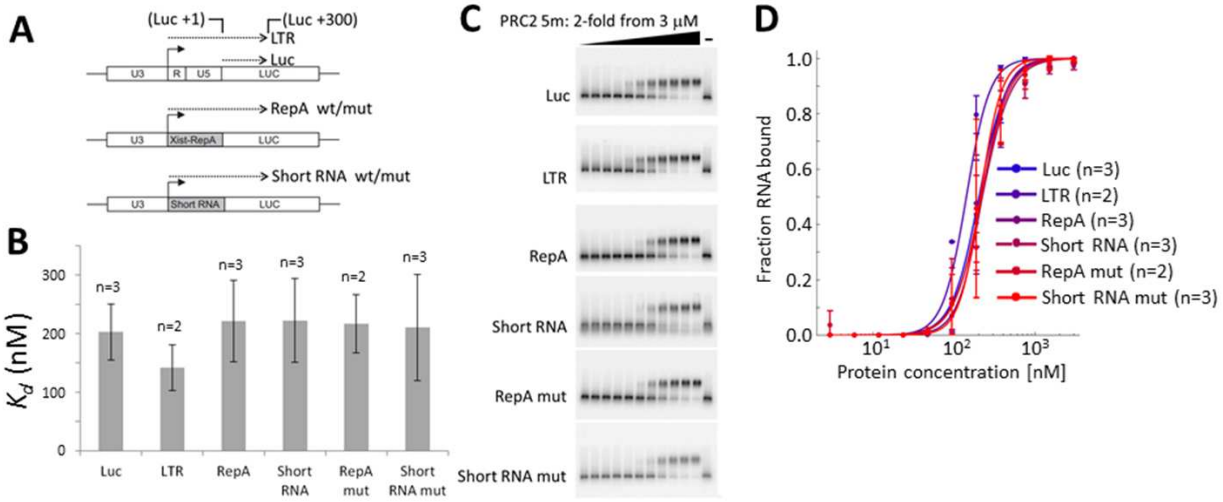


Figure 3.1. PRC2 binds RNA in the presence or absence of the two-hairpin motif that is repeated within RepA RNA. *In vitro* binding assays of recombinant PRC2 to in vitro transcribed RNAs generated from reporter vectors. **A.** *In vitro* transcribed RNAs were generated from sequences of reporter vectors (Kanhere et al., 2010) to include both the insert and 300 bases from the 5' end of the luciferase reporter mRNA. **B.** Dissociation constants were generated using EMSA of individual radiolabeled RNAs (<2 nM) in the presence of various concentration of human PRC2 5m. Error bars represent two to three independent experiments (n indicated), performed on different days. **C.** Representative gels are presented. **D.** Binding curves generated based on EMSA experiments in panel C and used to calculate dissociation constants (K_d's) that are presented in panel C**.** Binding curves of PRC2 to *in vitro* transcribed RNAs generated from reporter vectors. Indicated are the numbers of independent experiments performed on different days.

Kanhere et al. used luciferase reporter systems to test whether the two-hairpin motifs could recruit PRC2 and function in a repressive mechanism in cells. In their paper, they reported that addition of such motifs could repress gene expression. To test how PRC2 interacted with these RNAs, we first *in vitro* transcribed these five RNA sequences used in the original luciferase reporter systems. Two RNAs, RepA RNA and short RNA

(a short ncRNA), have the two-hairpin motifs. Two other RNAs (RepA mut and short RNA mut) contain mutations disrupting two-hairpin motifs. The fifth construct includes R and U5 region, originally used to estimate the basal expression level in the reporter system. The binding between each of these RNAs and PRC2 was tested. All these RNAs bound PRC2 with similar K_d values *in vitro* (Figure 3.1A-D).

We then repeated the experiment of Kanhere et al., testing whether these two-hairpin motifs could repress transcription of the luciferase system. Cells transfected with plasmids containing two-hairpin motifs showed low luciferase expression, while cells with plasmids lacking of two-hairpin motifs expressed higher luciferase signals (Figure 3.2A). These results were consistent with literature. Then, we investigated whether the repression was directed by PRC2 by knocking down SUZ12, followed by measurements of luciferase levels. By western blot, substantial SUZ12 and H3K27me3 levels were reduced after 48 hours of transfection (Figure 3.2B). However, expression levels of each vector were not changed (Figure 3.2A). Therefore, we concluded the repressive effects from the two-hairpin motifs were not dependent on PRC2.

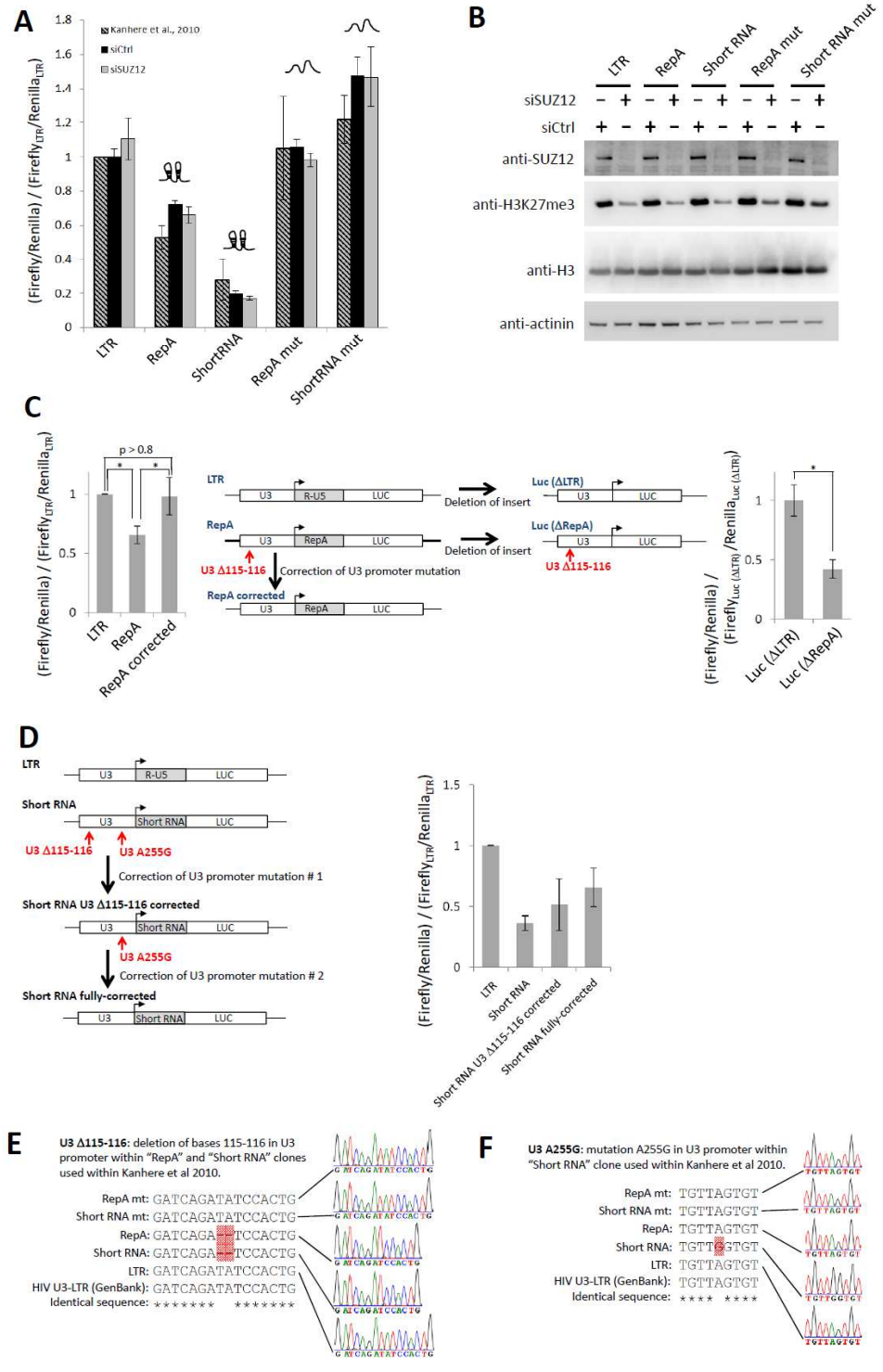


Figure 3.2. Effects of short hairpin RNAs on luciferase reporter expression are SUZ12-independent and can be tracked to promoter mutations in the reporter plasmid. A. Luciferase assays in the presence and absence of SUZ12 knockdown show that differences in expression level are PRC2-independent. Reporter vectors (Kanhere et al., 2010) were transfected

into HEK293T/17 cells 48 h post-transfection with siRNAs as indicated. Cells harvested 24 h later for luciferase reporter assay and **B.** immunoblotting, to confirm siRNA knockdown of the essential PRC2 subunit SUZ12. **C.** Reduced expression level of RepA reporter plasmid tracked to a mutation in an un-transcribed promoter region (see panel E for sequencing data). Left bar graph: Luciferase reporter assay performed before and after correcting a mutation that we identified in the U3 promoter of the RepA construct (U3 □115-116). Right bar graph: Luciferase reporter assay performed after deletion of inserts from the LTR plasmid (Luc (□LTR)) and RepA plasmid (Luc (□RepA)), demonstrating that differences in expression levels between these reporters, as previously observed (Kanhere et al., 2010), are due to promoter mutation and independent of the non-coding RNA sequence that was originally inserted upstream of the luciferase open reading frame. Error bars represent standard deviations based on four independent biological replicates, performed on different days. Asterisks represent $p<0.05$ (paired Student's t-test, two tailed). (Kanhere et al., 2010)**. D.** Two mutations in U3 promoter that were identified within Short RNA plasmid (see panels E and F) were corrected sequentially. Luciferase reporter assay performed for each of the resulting plasmids showed partial restoration of expression after sequence correction. Error bars represents standard deviations based on three to four independent biological replicates, performed on different days. **E.** Deletion of two bases (□115-116) within the U3 promoter, observed only in the vectors carrying wild-type two-hairpin motifs. These promoter mutations were not present in the vectors carrying the mutants designed to abolish the binding motif. (F) Mutation A255G within the U3 promoter region identified in the plasmid carrying the two-hairpin motif of the Short RNA, but not in any of the other reporters.

We then sequenced the U3 promoters in the five plasmids that were used in the original paper (Kanhere et al., 2010). A promoter mutation (U3 D115–116) was found only in the two plasmids carrying the wild-type sequences of the two-hairpin motifs (RepA and Short RNA; Figure 3.2E), which had lower luciferase expression levels. To test whether this mutation may affect luciferase expression, we corrected the mutation and tested the luciferase activity. We found the expression level of the corrected reporter (RepA corrected) increased back to the basal level (Figure 3.2C, left bar plot), which indicates that the expression difference observed between the RepA and RepA mut reporters (Kanhere et al., 2010) resulted from the U3 D115–116 promoter mutation, rather than differences in efficiency of recruitment of PRC2 by the transcribed RNA.

In addition, we corrected the two mutations that we identified within the U3 promoter of the Short RNA construct (Figure 3.2F), but not present in Short RNA mut. Correction of each mutation resulted in partially increased expression, rising to about one-

half the basal level (Figure 3.2D). Because the reporters were designed with the two-hairpin motif within the same transcript as the luciferase coding sequence, it could be that structure in the leader RNA affects the initiation or elongation of transcription, the initiation or efficiency of translation, or RNA stability, any of which is presumptively PRC2 independent. Additional tests (Figure 3.2C, right bar graph) showed that the promoter mutation U3 D115–116 gave a 2-fold expression decrease even in the absence of any leader RNA. These results demonstrate that the two-hairpin form of repeat A is not sufficient to achieve silencing in the reporter based system.

PRC2 Binds G > C,U >> A

To quantify the specificity of human PRC2 for oligonucleotides of different base composition, we purified recombinant holo-PRC2 5-mer complex (EZH2, EED, SUZ12, RBBP4 and AEBP2) that was monodisperse by size exclusion chromatography (Figure 3.3A,B), free of nucleic acid contamination, and active as a histone methyltransferase (Davidovich et al., 2013a, 2014). Binding of PRC2 to RNA was measured by an Electrophoretic Mobility Shift Assay (EMSA), and the shifted bands were shown to contain protein as well as RNA (Figure 3.3C). Because PRC2 affinity for RNA increases with RNA length (Davidovich et al., 2013a), RNAs of the same length were compared; usually these were 40-mers, long enough to allow for quantification of affinity while still within the range of chemical synthesis. G_{40}, or poly (G), bound to PRC2 with high affinity, while poly (A) showed little binding even at 5 μM PRC2 (Figure 3.4A). Quantitatively, this binding specificity of >> 1000x for poly (G) over poly (A) was in stark contrast to previous studies of natural RNAs, which showed modest specificities of < 10x. Poly (U) and poly (C) had

K_d^{app} values of ~300 nM (Figure 3.4A,B and Table 3.1), intermediate between those of poly (G) and poly (A).

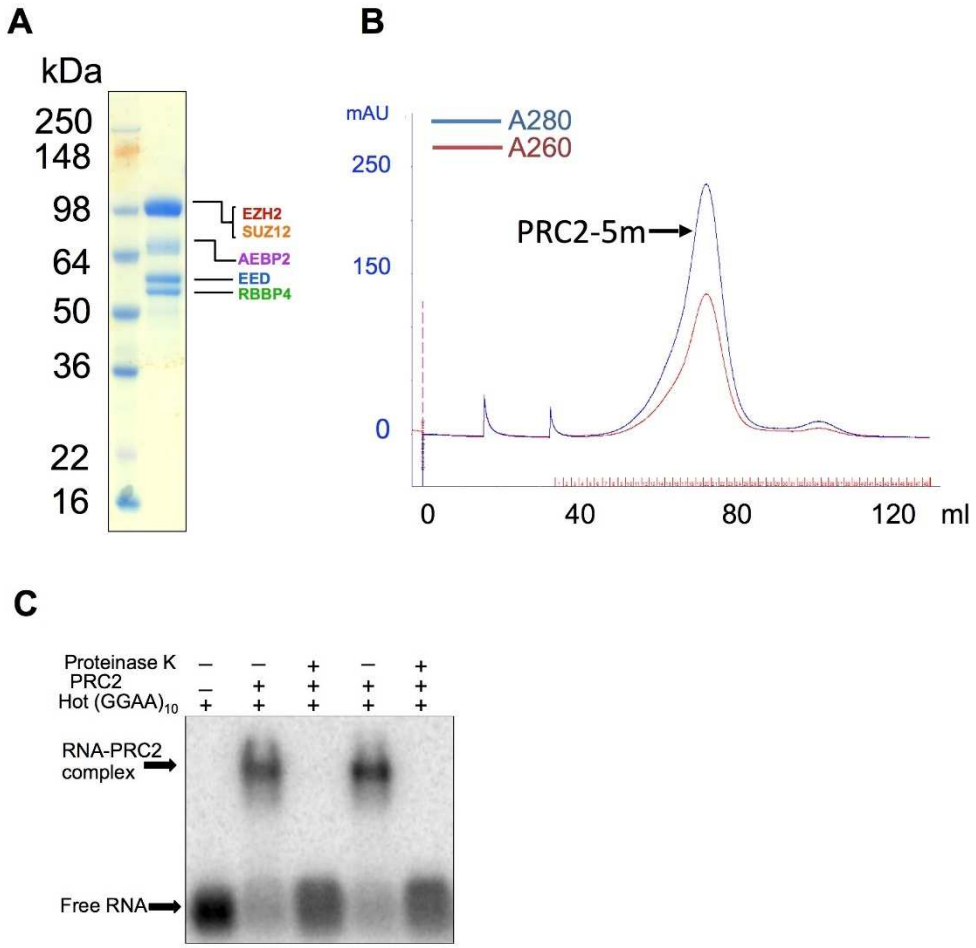


Figure 3.3. Purified PRC2 5-mer complex and its interaction with RNA. A. Purity of a typical PRC2 complex assessed by SDS-PAGE. **B**. PRC2 is monodisperse by size-exclusion chromatography on a Sephacryl 400 column. The lack of nucleic acid contamination is indicated by the A260/A280 ratio < 0.7 and by the absence of staining by ethidium bromide (the latter not shown here). **C.** The slow-migrating EMSA band is disrupted by proteinase K treatment, confirming that it is a protein-RNA complex rather than some oligomerization of the RNA component catalyzed by PRC2.

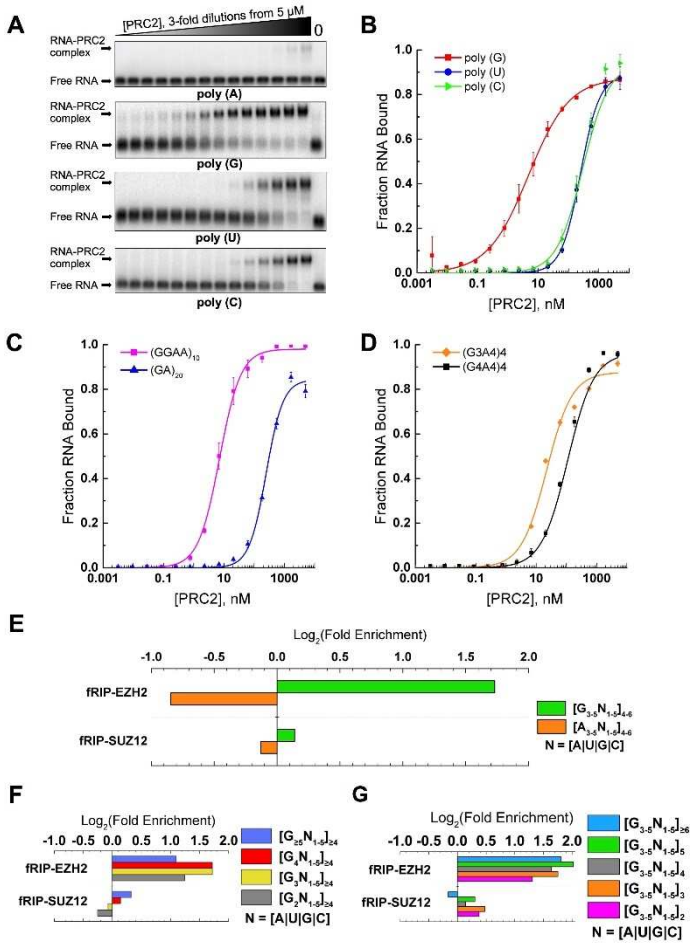


Figure 3.4. PRC2 referentially Binds to RNA-Containing G-Tracts *In Vitro* and *In Vivo*. A. Typical EMSA data for RNA 40-mers of homopolymeric sequence binding to PRC2 in 100 mM KCl. **B.** Data fit with equilibrium-binding curves and error bars give SD (n = 3). In this and other figures, many of the error bars are within the size of the symbol and, therefore, not visible. **C.** PRC2 binding to two oligonucleotides with the same base composition but a different grouping of Gs. **D.** PRC2 binding to RNAs containing blocks of Gs (sequences in Table 1). **E.** Enrichment of G-tracts, but not A-tracts, in RNA immunoprecipitated with EZH2, but not with SUZ12. Data of Hendrickson et al. (2016) were analyzed for the indicated motifs. The analysis cannot distinguish G from C or A from U. **F.** Association between PRC2 binding (fRIP-seq data) and putative motifs as a function of G-tract length. **G.** Association between PRC2 binding (fRIP-seq data) and putative motifs as a function of the number of G-tracts in proximity.

It seemed likely that guanines promoted PRC2 binding, but there was the formal possibility that adenines prevented binding. Thus, we asked whether the failure of poly (A) to bind could be rescued by insertion of G's into the sequence. 40-mers containing 50% G, $(GA)_{20}$ and $(GGAA)_{10}$, regained binding to PRC2 (Figure 3.4C). Other sequences containing 16 or 12 G's (40% or 30% G) also bound well (Figure 3.4D, Table 3.1). Thus, it appeared that PRC2 prefers binding G-tract RNAs.

Table 3.1. Equilibrium Constants for Binding of Various RNAs to PRC2. All RNAs are 40-mers. K_d^{app} values are mean ± SD (n = 3). Binding conditions were designed to approximate intracellular conditions (100 mM KCl) and to test for cation-dependent RNA structure formation (100 mM LiCl). N/A, not applicable.

RNA	Sequence	K_d^{app} in KCl (nM)	K_d^{app} in LiCl (nM)	Fold change of K_d^{app} in LiCl / K_d^{app} in KCl
poly (G)	G_{40}	8.0 ± 1.2	58 ± 11	7
poly (C)	C_{40}	280 ± 40	180 ± 30	0.6
poly (U)	U_{40}	340 ± 80	450 ±70	1
poly (A)	A_{40}	>10000	>10000	N/A
$(GA)_{20}$	$(GA)_{20}$	390 ± 70	320 ± 13	0.8
$(GGAA)_{10}$	$(GGAA)_{10}$	7.7 ± 2.4	290 ± 40	40
$(GU)_{20}$	$(GU)_{20}$	30 ± 10	160 ± 20	8
TERRA40	$(GA)_4(UUAGGG)_4(AG)_4$	80 ± 10	650 ± 50	8
TERRA40 mut	$(GA)_4(UGAGUG)_4(AG)_4$	490 ± 20	550 ± 30	1
(G3A4)4	$A_6G_3A_4G_3A_4G_3A_4G_3A_{10}$	39 ± 4.0	970 ± 110	20
(G3A4)3	$A_6G_3A_4G_3A_4G_3A_4GAGAGA_8$	980 ± 60	1500 ± 460	2
(G4A4)4	$A_6G_4A_4G_4A_4G_4A_4G_4A_6$	98 ± 6.0	670 ± 120	7
(G4A4)3	$A_6G_4A_4G_4A_4G_4A_{14}$	1300 ± 210	2100 ± 20	2
GC stem-loop	$(GC)_9(GA)_2(GC)_9$	3800 ± 350	4600 ± 690	1

Association between EZH2 and G-tract Motifs *in Vivo*

Given the high specificity that PRC2 demonstrated *in vitro* to RNAs containing multiple G-tracts (Figure 3.4A-D), and because such low complexity motifs are rather abundant genome-wide (Data not shown), we reasoned that these motifs could direct the association between PRC2 and thousands of its target transcripts *in vivo*. To this end, we quantified the association between PRC2 and G-tract-containing motifs of the sequence $[G_{3\text{-}5}N_{1\text{-}5}]_{4\text{-}6}$, where G is a guanine and N can be any nucleotide. Strikingly, we identified 9,239 association events between EZH2 protein binding sites (PBS) and $[G_{3\text{-}5}N_{1\text{-}5}]_{4\text{-}6}$ motifs in recently published RIP-seq data (G Hendrickson et al., 2016); this represented a significant association ($p<10^{-308}$) with Fold Enrichment (FE) of 3.32 over the number of association events that could be expected by chance (Figure 3.4E). Importantly, when we replaced the G in the motif to A, as a negative control, the association between PRC2 and the sequence motif $[A_{3\text{-}5}N_{1\text{-}5}]_{4\text{-}6}$ was dramatically reduced (FE=0.557). Interestingly, little to no enrichment was observed when the same analysis was performed for SUZ12-bound transcripts (FE of 1.101 and 0.914 for $[G_{3\text{-}5}N_{1\text{-}5}]_{4\text{-}6}$ and $[A_{3\text{-}5}N_{1\text{-}5}]_{4\text{-}6}$, respectively). This observation is in agreement with the different crosslinking sites that were observed for PRC2 subunits EZH2 and SUZ12 transcriptome-wide in two independent studies (Beltran et al., 2016; Hendrickson et al., 2016). Increasing the length of each G-tract from two to three or four G's (Figure 3.4F) or the number of tracts in the motif from two to three or more (Figure 3.4G) led to increased association with EZH2, therefore implicating short repeats of consecutive guanines as a determinant for RNA binding by PRC2 *in vivo*.

In accord with our *in vitro* data (Table 3.1), sequences of 8, 16 or 24 consecutive guanines, but not adenines, showed significant association with EZH2-binding sites on

RNA transcripts (fold enrichment of 2.6, 3.0 and 2.6, and p<10^-308, p=2.4*10^-5 and p=0.056, respectively; Figure 3.5A). Long G-tracts in DNA also showed strong association with PRC2 binding and with H3K27me3 marks (EZH2 or H3K27me3 ChIP-seq; Figure 3.5B). For all tested long G-tracts, the association with EZH2 and the repressive H3K27me3 mark was greater than that of the active H3K36me3 mark. Yet, unlike $[G_{3-5}N_{1-5}]_{4-6}$ sequences, very long G tracts are extremely infrequent in the human genome (489 G≥16 and 33 G≥24 sequences); thus it would not be safe to conclude that the latter play a major role in PRC2 regulation genome wide.

As another test for *in vivo* RNA binding, we co-expressed 3xFlag-EZH2 together with several size-matched RNAs in HEK293T cells and assayed for protein-RNA interactions using RNA immunoprecipitation (Figure 3.5D-H). The four test RNAs that showed binding *in vitro* were also bound *in vivo*. $(GGAA)_{10}$ RNA, which included multiple G-tracts, immunoprecipitated with EZH2 more efficiently than $(GA)_{20}$, in good agreement with our *in vitro* data. In another set of experiments, (G3A4)3 and (G3A4)4 showed similar binding to EZH2; it is not surprising to us that *in vivo* binding can in some cases be influenced by factors not present in the *in vitro* experiments. These differences require additional study.

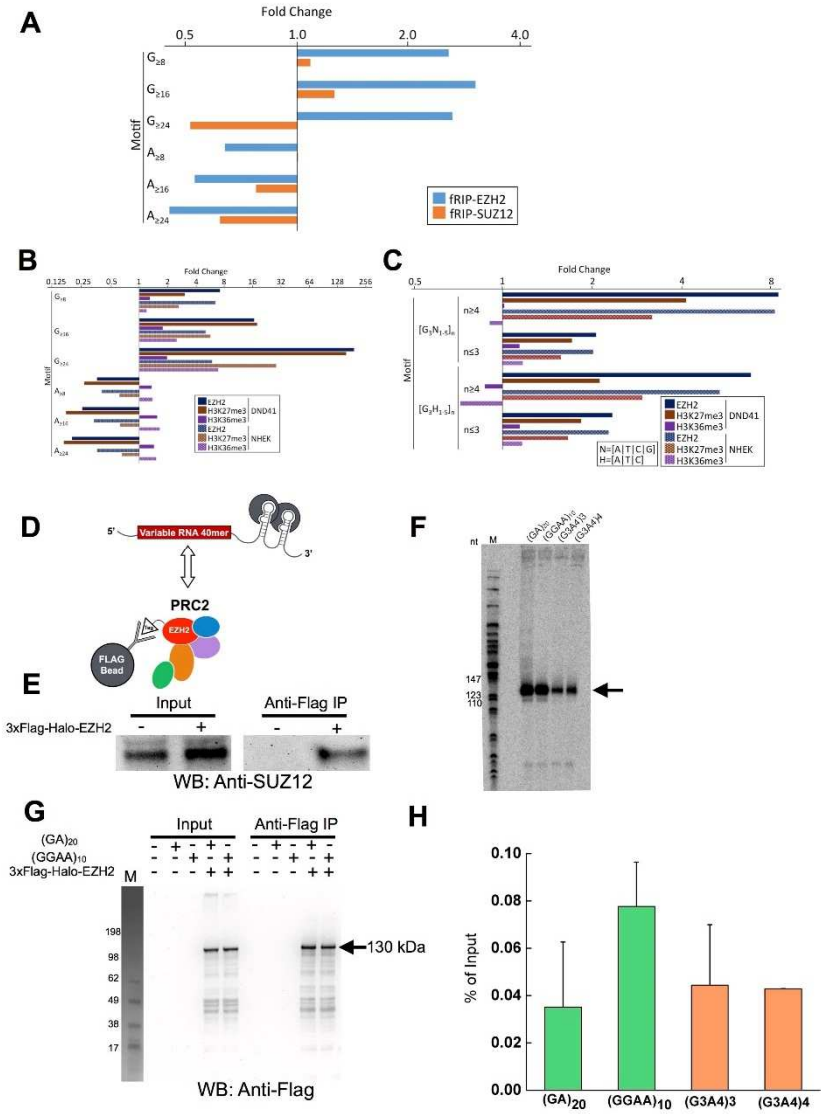


Figure 3.5. Transcriptome- and Genome-Wide Association Between PRC2 and G-rich Motifs. A. Transcriptome-wide association analysis of PRC2 subunits EZH2 and SUZ12 binding to motifs of long G or A tracts (fRIP-data from Hendrickson et al. (2016)). **B.** Association analysis between PRC2 subunits genome wide on motifs of long G or A tracts (ChIPseq data from the ENCODE project). **C.** Association analysis between PRC2 subunits genome wide to G-tract motifs of n≥4 or n≤3 repeats (ChIP-seq data from the ENCODE project). **D.** System for measuring PRC2-RNA interaction in cells. HEK293T cells were transfected with a plasmid with an RNA polymerase III promoter driving the expression of the test RNA sequence (red) followed by two MS2 stem-loops. The MS2 stem-loops bind MS2-eGFP protein, thereby stabilizing the RNA, and they also provide a primer-binding site for RT-qPCR. **E.** Western blot showing N-terminal 3x FLAG-Halo-tagged EZH2 is assembled into PRC2, as evidenced by co-IP with the SUZ12 subunit. **F.** Northern blot shows that the various test RNAs are expressed and stable when the cells are cotransfected with MS2-eGFP. **G.** Western blot shows IP of tagged EZH2 on Anti-FLAG beads, relative to 12 % of the input. **H.** RNA co-IP'd with the tagged EZH2, after subtraction of background, as a % of the RNA in the input; RT-qPCR measurements. Average of two or three biological replicates, +/- range of values obtained. Different color bars represent different sets of experiments. See METHODS for details.

PRC2 Prefers G4 Structures *in Vitro*

RNA sequences having four blocks of G's can form intramolecular G4 structures (Bugaut and Balasubramanian, 2012; Collie et al., 2010) which, like intramolecular DNA G-quadruplexes (Williamson et al., 1989), are comprised of planar arrays of Hoogsteen-bonded G-quartets. For example, TERRA is transcribed from telomeric DNA repeats (Azzalin et al., 2007), and it folds into a parallel G4 RNA structure (Collie et al., 2010). To test if PRC2's preference for blocks of G's were due to G4 structures, we designed TERRA40 and TERRA40mut; these have the same base composition, but only the former has four G_3 blocks enabling it to form G4 RNA. As shown in Table 3.1, PRC2 bound TERRA40 with six-fold higher affinity than TERRA40mut. As a further test, we designed (G3A4)4 and (G3A4)3, two 40-mers each containing 12 guanines, but only the former having G4-forming potential; PRC2 bound (G3A4)4 with affinity 25x higher than (G3A4)3, despite their high sequence identity (Table 3.1). A similar binding difference was found for (G4A4)4 and (G4A4)3, again with the sequence containing four blocks of G's binding much tighter (Table 3.1). Collectively, these results strongly indicate the high affinity of PRC2 for RNA sequences with the potential for G-quadruplex formation.

Whether each of these RNAs formed G4 structures was tested by circular dichroism (CD) spectroscopy (Tang and Shafer, 2006; Vorlíčková et al., 2012). Poly (G) formed intramolecular G4 structures in both K^+ and Li^+; the structures were so stable that they could not be melted at 90 °C (Figure 3.6A,B). Poly (U) is essentially unstructured, while poly (C) and poly (A) form their own characteristic base-stacked structures (Seol et al., 2007) (Figure 3.6C,D and 3.7A). G4 structure formation was also confirmed by CD for TERRA40, (G3A4)4, (G4A4)4 and – somewhat surprisingly – $(GGAA)_{10}$, for which the G4

structure would be comprised of only two G-quartets (Figure 3.6E,F and 3.7B). Native gel electrophoresis in KCl (Figure 3.6G; compare with Figure 3.7C,D) confirmed compact folded structures for poly (G) (heterogeneous due to multiple conformers) and for (G4A4)4, (G3A3)4, $(GGAA)_{10}$ and TERRA40, whereas each of their mutant versions ran as expected for an unfolded 40-mer, indicating a major structural disruption upon mutagenesis.

PRC2 has Elevated Affinity to G-rich RNA Even in the Absence of G4-Forming Sequences

Given the propensity of G-rich RNAs to fold into G4 structures, we asked if PRC2 also preferred to bind guanines in single-stranded RNA. We selected three of the sequences that do not fold into G4 structures and changed all of their G's into C's, allowing direct comparison of PRC2's affinity for G vs. C.

Their binding affinities were measured as $(GA)_{20}$ > $(CA)_{20}$, (G3A4)3 > (C3A4)3, and TERRA40mut > TERRA40mut-C (Table 3.2). Thus, PRC2 prefers G > C even when the G's are not participating in G4 structures.

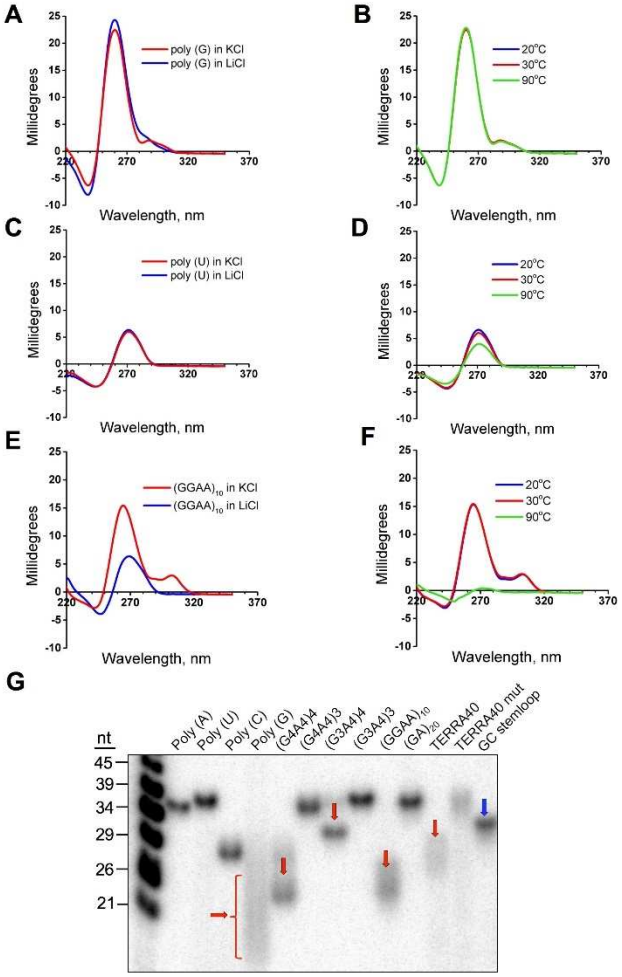


Figure 3.6. Validation of RNA Sequences that Fold into G4 Structures G4 structure formation assessed by CD (positive peak at 260–263 nm and negative peak at 237– 241 nm). All spectra were determined at least twice in 100 mM KCl and in 100 mM LiCl at three temperatures, with representative data shown here. Unless stated otherwise, CD was at 30ºC. **A.** Poly (G) in 100 mM KCl and in 100 mM LiCl forms such a stable set of structures that it does not require K+ to fold. **B.** Poly (G) in 100 mM KCl, structure does not melt even at 90ºC. **C. and D.** Poly (U) provides a negative control for G4 structures and shows little evidence of any structure (C) in the presence or absence of K+ or (D) at different temperatures. **E. and F.** (GGAA)10 undergoes (E) K+ -dependent folding into a G4 structure, (F) which is stable at 20ºC and 30ºC but melted at 90ºC. **G.** Native polyacrylamide gel electrophoresis at 30ºC of radiolabelled 40-mers. Poly (G) formed a heterogeneous array of intramolecularly folded structures (vertical red line) in this experiment and in multiple repeat experiments with different RNA heating-cooling protocols. Left lane, size markers; next four lanes, homopolymers; red arrows, folded G4 RNA structures; blue arrow, hairpin structure.

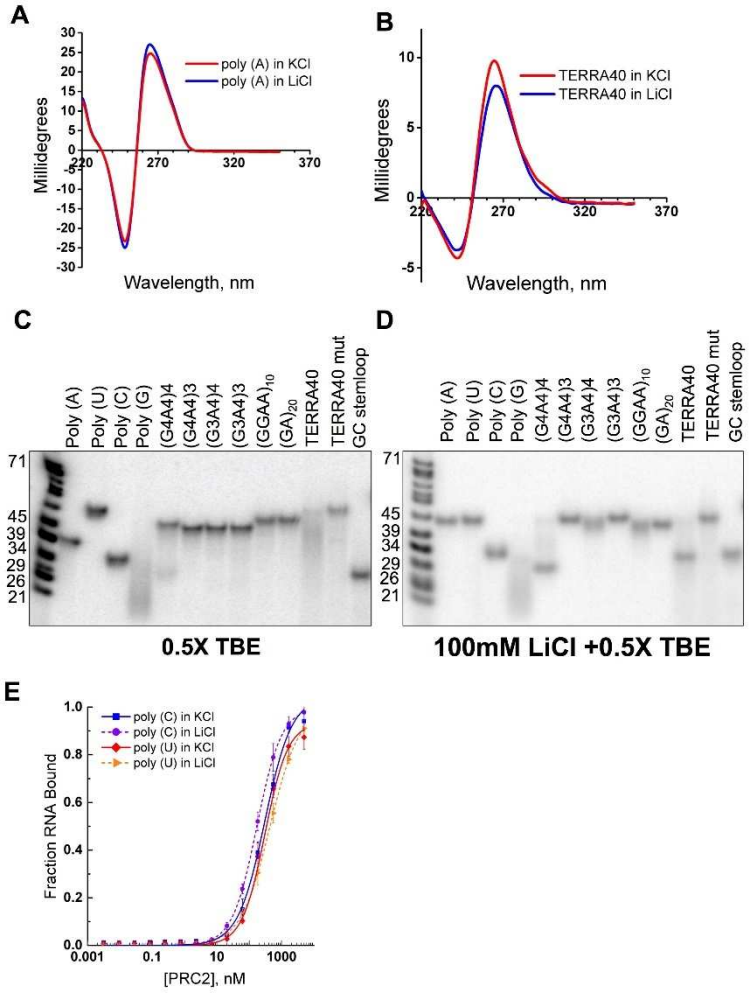


Figure 3.7. Characterization of Oligonucleotides. A. CD spectra of poly (A) in LiCl and KCl at 30°C. **B.** CD spectra of TERRA40 RNA in KCl and LiCl at 30 °C. **C.** Native gel electrophoresis of RNA 40-mers in 0.5X TBE without added KCl or LiCl. Most are unfolded, but Poly (G) forms a smear of various intramolecular G4 structures even without salt, while (G4A4)4 and TERRA40 exhibit partial folding. GC stemloop folds into a hairpin secondary structure under these conditions. Left lane, molecular weight markers, with some of their sizes listed at left. **D.** Native gel electrophoresis in LiCl of the same RNA 40-mers as in C. The results are similar to panel C, except (G4A4)4 and TERRA40 are more stably folded under these conditions. **E.** Poly (C) and poly (U) bind PRC2 equivalently in KCl and LiCl.

Table 3.2. PRC2 Binds G > C Even in the Absence of G4 Structure Formation. 40-mer RNAs were incubated with PRC2 in the 100 mM KCl condition. Kd app values are mean ± SD (n = 3).

RNA	Sequence	K_d^{app} (nM)	RNA	Sequence	K_d^{app} (nM)
$(GA)_{20}$	$(GA)_{20}$	390 ± 70	$(CA)_{20}$	$(CA)_{20}$	2000 ± 840
TERRA40 mut	$(GA)_4(UGAGUG)_4$-$(AG)_4$	490 ± 20	TERRA40 mut-C	$(CA)_4(UCACUC)_4$-$(AC)_4$	870 ± 60
(G3A4)3	$A_6G_3A_4G_3A_4G_3$-$A_4GAGAGA_8$	980 ± 60	(C3A4)3	$A_6C_3A_4C_3A_4C_3$-$A_4CACACA_8$	3400 ± 850

PRC2 Avoids Binding Double-Stranded RNA (dsRNA)

The binding of PRC2 to perfect dsRNA had not been previously tested. Two 40-mer hairpin RNAs were designed, each predicted to form 18 base pairs connected by a tetraloop. “GC stem-loop” (Table 3.1), which had exclusively G-C base pairs and a base composition of 50% G, did not show any binding (estimated K_d >> 2500 nM), in contrast with $(GA)_{20}$ and $(GGAA)_{10}$ which are also 50% G. “Hairpin 40,” which had a mixed sequence, showed no binding, but its size-matched control (Control 40, Figure 3.8A,B) bound PRC2 well (K_d^{app} = 13 nM). The lack of binding of these highly base-paired RNAs stands in sharp contrast to the preferential binding to folded G4 structures. However, a few “helix defects” were sufficient to restore binding to dsRNA: Looped Hairpin 64 (with two small internal loops, a bulge and a 3’-tail, giving 30% single-stranded nucleotides) bound PRC2 well, although not as well as the size-matched Control 64 (Figure 3.8C,D).

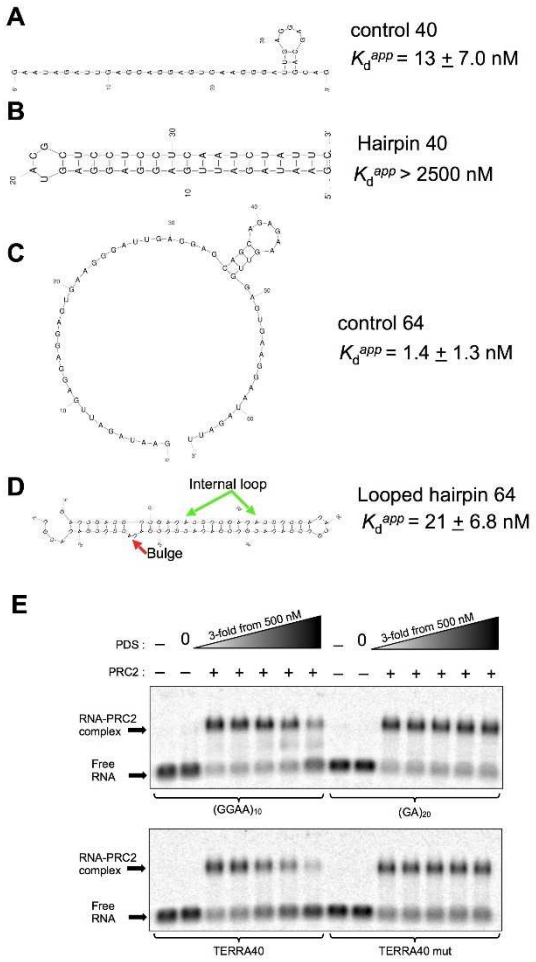


Figure 3.8. PRC2 Does Not Bind to a Long Hairpin RNA and its Binding to G4 RNA is Competed by the G4-ligand PDS. A – D. RNAs designed to form base-paired secondary structures and their mutated controls. The most stable secondary structure predicted by M-Fold is shown in each case. Although the control RNAs show little potential to form base-paired structure by M-Fold, they do have potential to form G4 structures. Binding to PRC2 was measured in 100 mM KCl in all cases. **E.** Pre-formed G4-RNA-PRC2 complexes are disrupted by the quadruplex-binding PDS compound (Rodriguez et al., 2008), whereas non-G4-RNA complexes are unaffected.

PRC2 Binds the Folded Form of G4 RNA Sequences

PRC2 might bind folded G4 RNA structures (Figure 3.9A), or it might bind to an unfolded or partially unfolded form, conceivably by facilitating G4 unfolding. To begin to address this question, we compared PRC2 binding in 100 mM KCl versus 100 mM LiCl.

The expectation was that if PRC2 binds to folded G4 RNA, then the presence of K^+ in the binding buffer should stabilize protein binding. On the other hand, if PRC2 were able to melt G4 structures and bind to the unfolded form, then K^+ would make it more energetically costly to unfold G4 RNA and would destabilize binding. As shown in Figure 3.9B,C and Table 3.1, PRC2 bound all G4-forming RNAs ~10x more tightly in K^+ than in Li^+. This supports the conclusion that it is the folded G4 RNA structure that is bound.

However, it seemed possible that the non-physiological LiCl conditions might perturb PRC2's structure and ability to bind RNA. We therefore measured PRC2 binding to $(GA)_{20}$, poly (C), and poly (U), none of which undergoes a cation-specific folding transition. Strikingly, the binding of these RNAs to PRC2 was virtually identical in 100 mM LiCl versus KCl (Figure 3.9D, 3.7E). Thus, we conclude that the differential effect of Li^+ vs K^+ on PRC2 binding to G4 RNAs is due to a structural change in the RNA, not the protein.

As four repeats of consecutive guanines are typically considered as the minimal requirement for the formation of a stable G-quadruplex structure, we wished to determine whether there is larger association *in vivo* between PRC2 binding to motifs of the form $[G_3N_{1\text{-}5}]_n$ when $n \geq 4$, compared to $n \leq 3$. Given the large abundance of motifs of the form $[G_3N_{1\text{-}5}]_{\leq 3}$ genome wide, this analysis was restricted to chromosome 12. In agreement with the G4 hypothesis, strong association between fRIP-EZH2 to the motif $[G_3N_{1\text{-}5}]_n$ (or $[G_3H_{1\text{-}5}]_n$, where H is any base except for G) was observed only where $n \geq 4$, not $n \leq 3$ (Figure 3.9F). Note that the analysis of $n \leq 3$ is dominated by $n=1$, given its relative abundance. The same observation persisted also when these motifs were tested for

association against binding sites for EZH2 on DNA (ChIP-seq) or the repressive chromatin mark H3K27me3, but not the active chromatin mark H3K36me3 (Figure 3.5C).

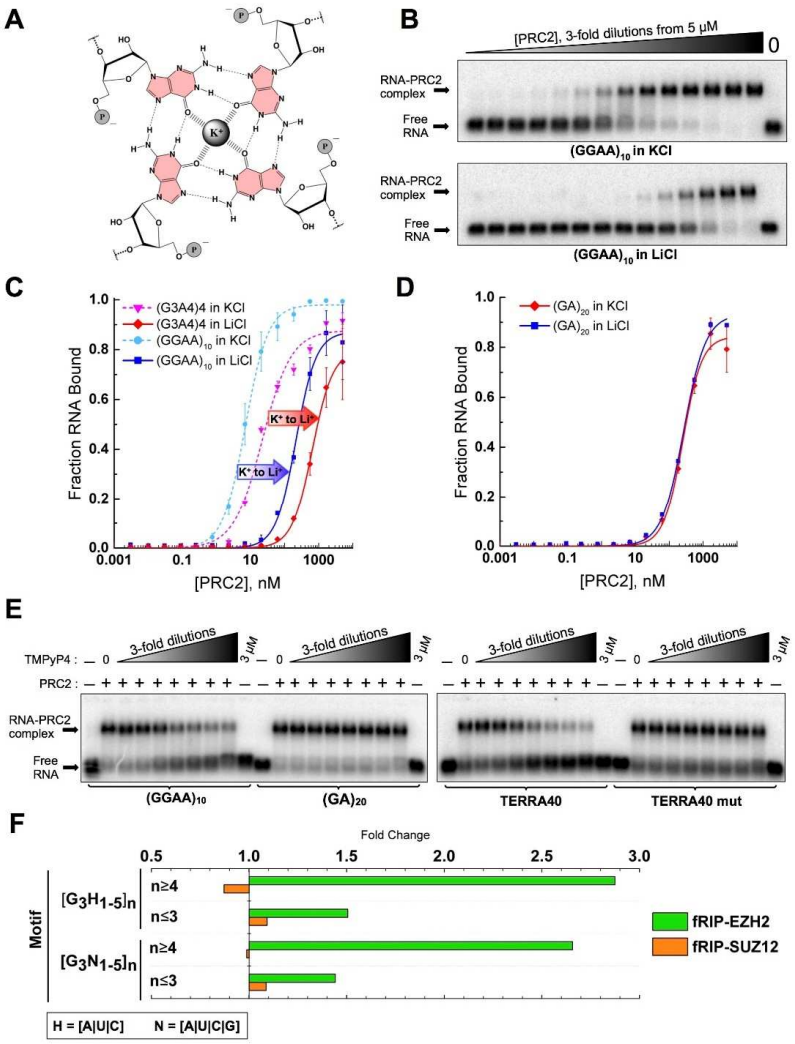


Figure 3.9. G-Quadruplex Stabilization Increases Its Affinity for PRC2. A. Planar array of Hoogsteen-bonded guanines, or G-quartet, forms one layer of a G4 RNA structure. Potassium ion binds between layers and stabilizes the electronegative carbonyl oxygen atoms. **B.** EMSA shows (GGAA)10 binding at lower PRC2 concentrations in 100 mM KCl than in 100 mM LiCl. **C.** Two of the oligonucleotides that show substantial increases in binding in KCl relative to LiCl. Error bars show SD (n = 3) in this and subsequent panels. **D.** The non-G4 oligonucleotide (GA)20 binds identically in KCl and LiCl. **E.** Pre-formed G4-RNA-PRC2 complexes are disrupted by the cationic porphyrin TMPyP4, whereas non-G4-RNA complexes are unaffected. **F.** *In vivo* fRIP-seq data show stronger association between EZH2 and larger numbers of repeats (n≥4) of three consecutive guanines, compared to shorter numbers of repeats (n≤3). Because of the enormous number of such motifs across the genome, this specific association analysis was restricted to human chromosome 12.

RNA Crosslinks to Four Subunits in the Assembled PRC2 Complex

Previous studies concluded that the RNA-binding subunit(s) of PRC2 are EZH2 (Kaneko et al., 2010; Zhao et al., 2008a), EZH2-EED (Wu et al., 2013) or SUZ12 (Kanhere et al., 2010). In another study EED was suggested to temper the binding interactions (Cifuentes-Rojas et al., 2014a). AEBP2 has been shown to slightly increase the affinity of PRC2 to RNA, by approximately two fold (Davidovich et al., 2013a). One potential challenge with assessing RNA binding to isolated PRC2 subunits is that these proteins may not be correctly folded in the absence of their binding partners. Indeed, recent crystal structures of the core complex of PRC2, including EZH2, EED and VEFS domain from SUZ12, show large and complex interfaces between these subunits (Brooun et al., 2016; Jiao and Liu, 2015a). To rule out such concern, we used fully assembled and monodisperse holo-PRC2 5-mer complex for protein-RNA photocrosslinking to assess which subunits are in proximity to bound RNA.

One series of RNA 24-mers was based on a G4 structure, with 4-thio-uridine (4S-U) substituted near the 5' end, near the 3' end, or in an internal loop between G-quartets. These RNAs underwent crosslinking with similar efficiencies in the range of 9 – 14% (Figure 3.10A). To unambiguously identify the crosslinked protein subunits, the crosslinking was repeated with PRC2 complexes in which a single subunit was MBP-tagged and therefore had retarded electrophoretic mobility (Figure 3.11A, 3.10B). The results indicated crosslinking of RNA primarily to EZH2 (in agreement with Zhao et al., 2008 and Kaneko et al., 2010) and EED (in agreement with Wu et al., 2013 and Cifuentes-Rojas et al., 2014), and also to SUZ12 (in agreement with Kanhere et al., 2010) and AEBP2 (in agreement with our previous observations; Davidovich et al., 2013), but not

RBBP4 (Figure 3.11B). For comparison, a non-G4-forming 24-mer gave a very similar crosslinking pattern (Figure 3.11A,B). As an example of a longer RNA containing a variety of structural elements, 4S-U-substituted HOTAIR 400 RNA was tested; the same four PRC2 subunits were found to be crosslinked, although in this case with more equal efficiency (Figure 3.11A,B). Because the presence of an N-terminal tag could perturb RNA binding, we compared the crosslinking efficiencies of tagged and untagged subunits. As shown in Figures 3.11B and 3.10C-E, EZH2 and EED crosslinked with better efficiency than other subunits independent of MBP tagging. Importantly, although multiple observations have linked these four subunits, independently, with the RNA-binding activity of PRC2, to our knowledge this is the first evidence for function of all these proteins in RNA binding while assembled into a holo-PRC2 complex.

It seemed possible that one of these subunits provided the primary binding site or “anchor” for RNA binding, and perhaps the others were not real binding sites but simply regions where the RNA was held into proximity of the protein. To try to reveal a primary binding site, we conducted crosslinking experiments as a function of decreasing UV exposure, increasing salt concentration and increasing competitor RNA concentration (e.g., Figure 3.10F,G). In all cases, the crosslinking to each of the subunits decreased to a similar extent. Thus, these results do not support an “anchor site” hypothesis, but rather indicate a genuine function in RNA binding for all four protein subunits.

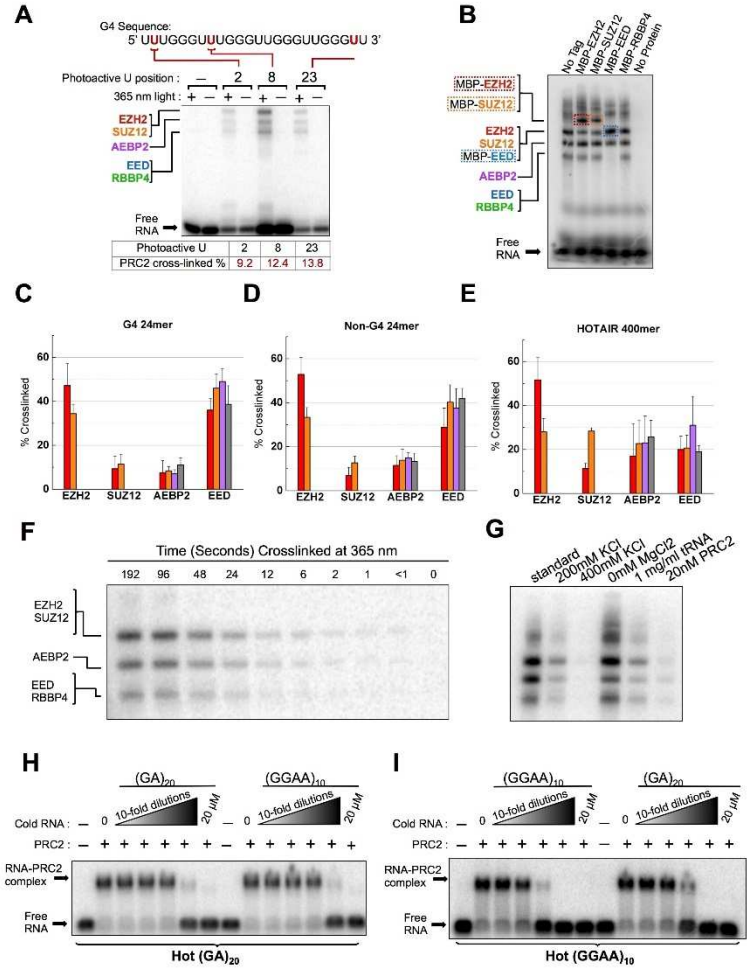


Figure 3.10. HOTAIR lncRNA Crosslinks to Four PRC2 Subunits, and Binding of a G4-structured RNA Competes with Binding of a Single-stranded RNA. A. G4-forming 24mer RNAs with 4S-U incorporated during chemical synthesis cross-link to PRC2 with similar efficiency. **B.** The complete gel for HOTAIR400 crosslinking from Figure 4A, right-most panel. This shows the free (uncrosslinked) RNA and fuzzy bands above MBP-EZH2 that may represent inter-subunit crosslinks and were not included in the quantification. **C,D,E.** Different methods of calculating the % crosslinking to the PRC2 subunits yield similar results, even though they rely on different tagged proteins or use data from untagged proteins only. Blue bars, MBP-EZH2. Orange bars, MBP-SUZ12. Yellow bars, MBP-RBBP4. Grey bars, no tagged subunits. Error bars = SD. G4 24mer, n = 5. Non-G4 24mer, n = 5 – 6. HOTAIR 400mer, n = 3 for orange and yellow bars, n = 6 for blue and grey bars. **F.** Crosslinking of HOTAIR400 to PRC2 subunits as a function of irradiation time. **G.** Crosslinking of HOTAIR400 to PRC2 subunits is diminished by higher salt, tRNA competitor, and reduction of the PRC2 concentration to 20 nM. Standard conditions defined in Methods. **H**. Pre-formed (GA)20-PRC2 complex is competed by unlabeled (GA)20 and by (GGAA)10. **I.** Pre-formed (GGAA)10-PRC2 complex is competed by the same unlabeled RNA and by (GA)20.

G-quadruplex RNA and Single-stranded RNA Bind to the Same Site(s) on PRC2

Finding that four PRC2 subunits were in proximity to bound RNA seemed to provide a possible explanation for our unexpected finding that PRC2 could bind two completely different types of RNA: highly folded G4 RNA and ssRNA containing G, C and U nucleotides. One could hypothesize that these different types of RNA might bind to different sites on PRC2. To test this idea, we competed the binding of radiolabeled TERRA40 (which has a G4 structure) with unlabeled TERRA40mut (which does not form G4), and vice versa. The two RNAs competed for binding (Figure 3.11C), indicating identical or mutually exclusive site(s). Similarly, $(GGAA)_{10}$ competed for binding with $(GA)_{20}$ (Figure 3.10H,I).

Thus, both the crosslinking results (G4 and non-G4 RNAs crosslink to the same four subunits) and the competition results (G4 and non-G4 compete for binding) lead to the model of Figure 3.11D. Our data cannot exclude a complex model involving multiple binding sites that are all capable of binding both G4 and non-G4 RNAs, but postulating a single RNA-binding site for PRC2, involving multiple subunits, is the simplest model.

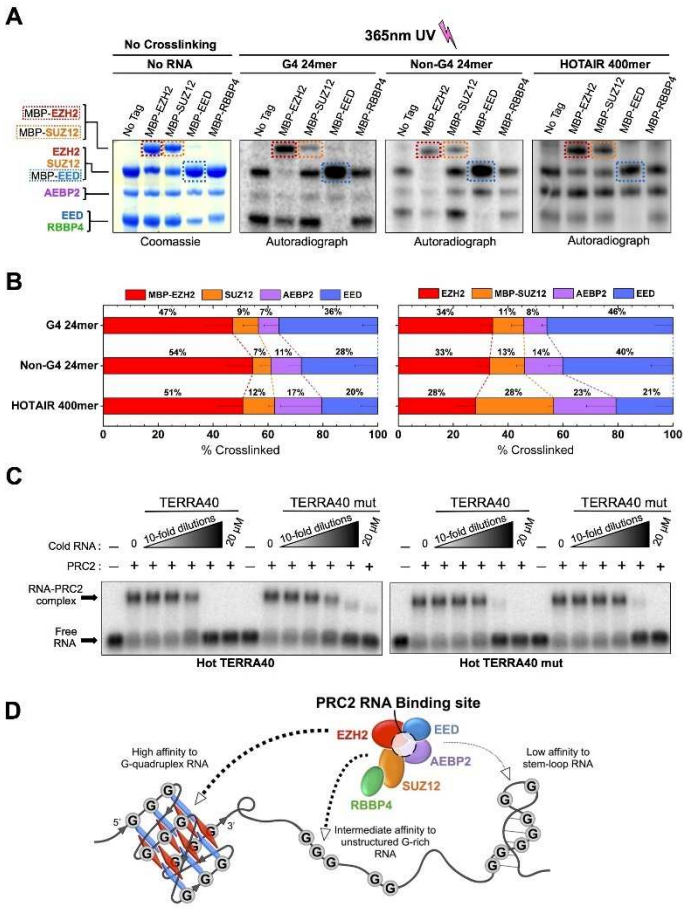


Figure 3.11. Crosslinking Maps the PRC2 Subunits in Proximity to Bound RNA. **A.** Coomassie, SDS-PAGE of purified PRC2 (no tag), and four singly MBP-tagged PRC2 complexes (e.g., MBP-EZH2 indicates a purified PRC2 5-mer complex in which only the EZH2 subunit was N-terminally tagged with MBP). For the next three gels, the indicated 4S-U-substituted RNA was radiolabeled and photocrosslinked to the untagged and MBP-tagged PRC2 complexes. The order of the lanes is the same as for the Coomassie gel. Colored boxes are MBP-tagged subunits, which, in the case of MBP-EED, overlap with other untagged subunits. **B.** Quantification of crosslinking efficiency between PRC2 subunits and various RNAs. There was no detectable crosslinking to RBBP4. Differences between left half (MBP-tagged EZH2) and right half (MBP-tagged SUZ12) could represent effect of the tag on RNA crosslinking or systematic error due to subtraction of background in gel. Error bars, SD; five to six independent experiments. **C.** Left: pre-formed TERRA40-PRC2 complex is competed by unlabeled TERRA40 RNA and by its non-G4 mutant. Right: pre-formed TERRA40mut-PRC2 complex is competed by the same unlabeled RNA and by TERRA40. These experiments were not designed to measure K_i but rather to assess mutually exclusive versus independent binding. **D.** Model for a single RNA-binding site at the conjunction of four PRC2 subunits, and a summary of its relative binding to various RNA structures/sequences.

PRC2 Binding Sites on Chromatin Genome-wide Are Associated with G-tract Motifs

PRC2 binding sites on chromatin in mammalian ES cells are known to associate with GC-rich DNA sequences (Mendenhall et al., 2010a). Given the strong association between PRC2 and G-tract RNAs *in vitro* and *in vivo*, it was important to determine whether motifs composed of short repeats of consecutive deoxyG's are enriched at PRC2 target genes genome wide. This is particularly important because RNA-mediated regulation of PRC2 has been previously proposed to take place in cis (Kanhere et al., 2010; Zhao et al., 2008a) and in trans (Rinn et al., 2007). Hence, it is not a foregone conclusion that PRC2 binding to RNA will track with binding to cognate DNA, from where PRC2-binding transcripts are being transcribed. Accordingly, EZH2 ChIP-seq peaks in the ENCODE data (ENCODE Project Consortium, 2011) were shown to associate with $[G_{3\text{-}5}N_{1\text{-}5}]_{4\text{-}6}$, with FE >8 ($p < 10^{-308}$) in both cell lines that were tested (Figure 3.12A). The negative control motif $[A_{3\text{-}5}N_{1\text{-}5}]_{4\text{-}6}$ was not enriched at EZH2-binding sites (FE<0.4 in both cell lines). In accord with these observations, H3K27me3 was associated with $[G_{3\text{-}5}N_{1\text{-}5}]_{4\text{-}6}$, albeit with slightly reduced Fold Enrichment (FE >3 for $[G_{3\text{-}5}N_{1\text{-}5}]_{4\text{-}6}$ and FE<0.6 for $[A_{3\text{-}5}N_{1\text{-}5}]_{4\text{-}6}$ in both cell lines). As a negative control, we repeated this analysis for the active chromatin mark H3K36me3, which did not associate with any of these motifs (1.0 >FE >1.2 for both motifs and both cell lines). Increasing the number of consecutive guanines within a given DNA repeat (Figure 3.13A) or the total number of G-tracts in the motif

(Figure 3.13B) revealed increased association with EZH2 genome-wide, implicating PRC2 enrichment at genes encoding G-tract RNAs.

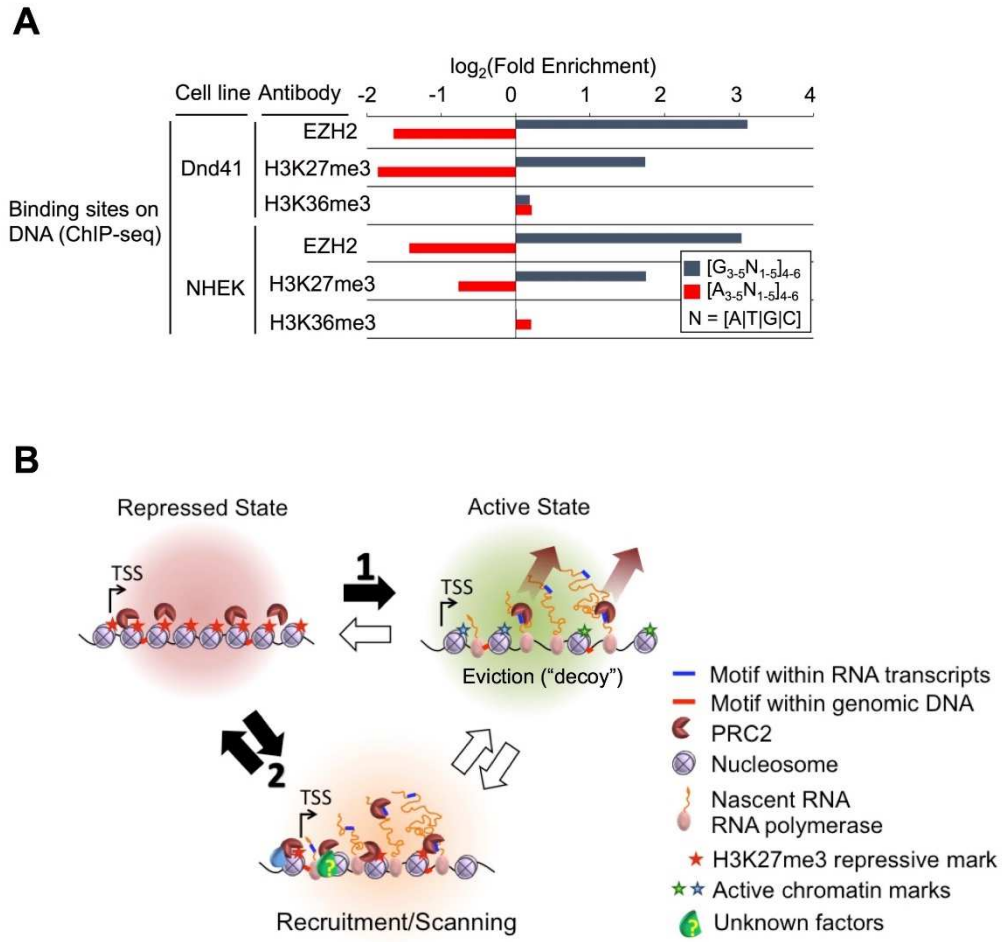


Figure 3.12. RNAs with Multiple G-Tracts Are Associated with PRC2-Binding Sites on Chromatin and with the H3K27me3 Mark. A. Enrichment of G-tracts, but not A-tracts, around EZH2-binding sites and chromatin mark H3K27me3, but not H3K36me3, in two cell lines (ENCODE data; Table S1). **B.** Enrichment of DNA sequences coding for PRC2-binding RNA motifs at Polycomb target genes provides a means for RNA-mediated regulation in cis. Such models that have been previously proposed include (1) facilitation of derepression processes and maintenance of the active state of Polycomb target genes by eviction of PRC2 from transcribing genes that should not be subjected to H3K27me3 repression, where the RNA acts as a decoy to evict PRC2 from chromatin; and (2) maintenance of epigenetic silencing by RNA recruitment of PRC2 to target genes or scanning for other recruitment factors (reviewed in Brockdorff, 2013 and Davidovich and Cech, 2015). While all these models are still pending validation, they are not mutually exclusive, and any of them, if valid, will likely rely on local enrichment of such PRC2-binding RNA motifs.

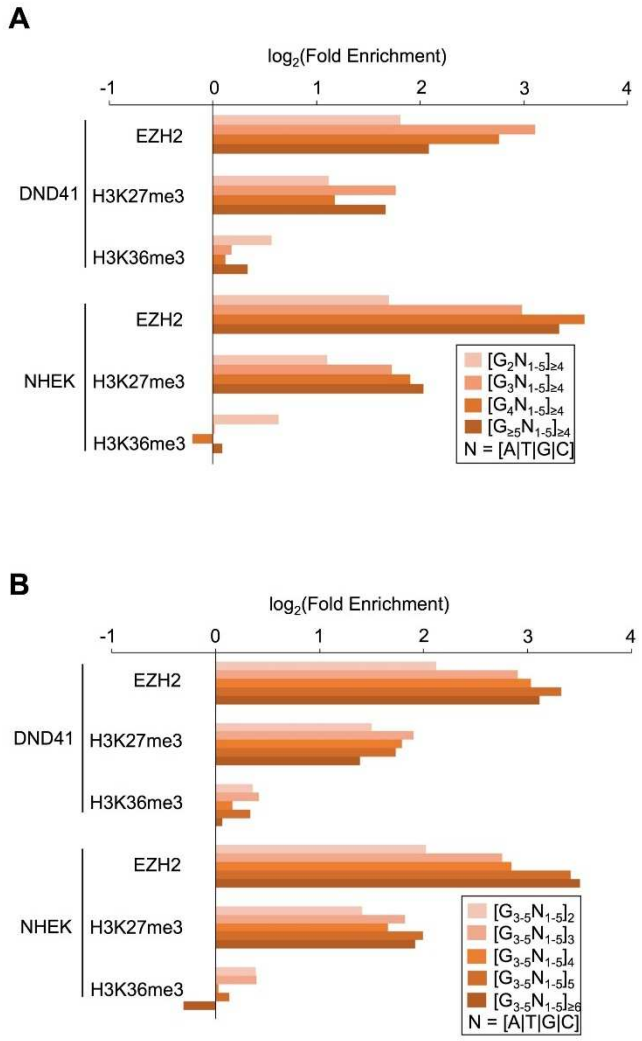


Figure 3.13. Enrichment of G-track Sequences in DNA Associated with EZH2 and H3K27me3 Chromatin in Two Human Cell Lines. A. Association between ENCODE ChIP-seq data and putative motifs as a function of G-tract length, in bases. **B.** Association between ENCODE ChIP-seq data and putative motifs as a function of the number of G-tracts in a motif.

DISCUSSION

The present work provides a quantitative basis for understanding the promiscuous binding of PRC2 to RNA. First, PRC2 binds single-stranded RNAs, especially those containing runs of G's, with moderate affinities, ~300 nM. [These affinities are for 40-mers; PRC2 binding affinity would increase with RNA length (Davidovich et al., 2013a). Only stretches of A's are excluded from binding, in agreement with Kaneko et al. (2014). Second, G-quadruplex structures have enhanced affinity, typically a full order of magnitude greater than those of closely matched sequences that cannot fold into G4 structures [Table 3.1, KCl data, compare $(GGAA)_{10}$ with $(GA)_{20}$, (G3A4)4 with (G3A4)3, and (G4A4)4 with (G4A4)3]. Remarkably, it takes multiple repeats of only two consecutive G's to form G4 in RNA, as shown by the CD spectrum and K^+-dependent native gel electrophoretic mobility of $(GGAA)_{10}$. Even $(GU)_{20}$, originally designed as an unfolded control RNA, clearly undergoes K^+-dependent folding (by native gel electrophoresis) into some structure(s) that are preferentially bound by PRC2 (Table 3.1). There are short G-tracts in almost any RNA that was ever tested *in vitro* for binding PRC2 (Cifuentes-Rojas et al., 2014a; Davidovich et al., 2013a, 2015a; Kanhere et al., 2010; Tsai et al., 2010; Wu et al., 2013; Zhao et al., 2008a); for example, HOTAIR 400 has eight tracts of GGG or longer. Thus, PRC2 recognition of this very common multiple-G-tract motif leads to its promiscuous binding to natural lncRNAs and pre-mRNAs.

Our analysis shows preferential binding of PRC2 to RNAs containing G-tracts *in vivo* (Figure 3.4E,F). Furthermore, we find stronger association of EZH2 with four or more repeats of three consecutive G's compared to short repeats (n≤3) (Figure 3.9F), consistent with binding to G4-forming sequences, although clearly four G-tracts are not

required for binding *in vivo* (Figure 3.4G and 3.5H). Transcriptome-wide, RNA secondary structures appear to be less folded *in vivo* than *in vitro*, which is attributed to the action of helicases and single-stranded RNA-binding proteins *in vivo* (Rouskin et al., 2014; Spitale et al., 2015). The same appears to be true of G4 RNA structures in mammalian cells (Guo and Bartel, 2016). It is possible that PRC2 gets a "fair chance" to compete with such helicases and other G-tract-binding factors for newly transcribed G4 structures in RNA, as they are emerging from the polymerase through transcription of active genes. Thus, we propose that, even if G4 structures are labile *in vivo*, the folded form can be trapped and stabilized by PRC2 binding, and that such structures provide high-affinity targets for PRC2. Promiscuous binding *in vivo* is then due to the ubiquity of G4-forming RNA sequences and to non-G4 tracts of G's, U's and C's that also bind PRC2.

In general agreement with previous hypotheses (Kaneko et al., 2014a), we found that the structure of a given RNA can indeed be a determinant for PRC2 binding. Yet, our results support a different specific mechanism, where structured RNA per se is insufficient to serve as a preferred target for PRC2: single-stranded RNAs containing G-tracts are good ligands for PRC2, and double-stranded RNA segments are not ligands at all (Figure 3.11D). However, purely double-stranded RNA is rare *in vivo*. We found that introduction of a few helix defects (internal loops and bulges) in an RNA hairpin made it a good PRC2 binder. Thus, *in vivo*, although double-stranded RNA per se is not expected to be a PRC2 target, it should not inhibit PRC2 binding to adjacent single-stranded nucleotides, such as bulged bases or internal loops.

A recent study suggested a major role for SUZ12 in the association between PRC2 and RNA in cells (Beltran et al., 2016a). Our analysis of SUZ12 fRIP-seq data indeed

confirmed genome-wide association between this protein and RNA (Table 3.2), even if based on this particular data no association was observed with G4-forming sequences. Our *in vitro* crosslinking data also confirmed contacts between RNA and SUZ12 (Figure 3.11A,B). The discordance between EZH2 and SUZ12 RNA sequence preferences observed here and by others (Beltran et al., 2016a; G Hendrickson et al., 2016) remains unexplained, but the different extent of RNA crosslinking by these two subunits (Figure 3.11B) is consistent with them binding RNA differently. Future studies will be required to elucidate the precise manner in which each PRC2 subunit functions in RNA binding and RNA-mediated regulation of this epigenetic modifier.

Importantly, the significant enrichment of DNA sequences coding for PRC2-binding RNA motifs around PRC2-binding sites on chromatin (Figure 3.12A and Figure 3.13) implies that RNA transcripts emerging from these loci, either coding or non-coding, will have better chance to compete for PRC2 binding than other transcripts. While the actual function of RNA binding by PRC2 is still under investigation by multiple independent groups, the strong association between PRC2 and low complexity motifs of short G-tract repeats emerging from Polycomb target genes provides a means for RNA-mediated regulation of PRC2 in cis (Figure 3.12B).

MATERIALS AND METHODS

Key resources table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-SUZ12	Santa Cruz	sc67105
Anti-Flag M2-HRP	Sigma	A8592

Goat Anti-Rabbit IgG-HRP	Santa Cruz	sc2301
Chemicals, Peptides, and Recombinant Proteins		
SuperSignal West Pico Chemiluminescent substrate	ThermofisherScientific	Cat#34080
Lipofectamin 2000 transfection reagent	ThermofisherScientific	11668019
SuperScript III Reverse Transcriptase	Life Technologies	180800-044
4-thio-UTP	TriLink Biotechnologies	N-1025
Critical Commercial Assays		
IQ SYBR Green Supermix	Biorad	1708880
T4 Polynucleotide Kinase	New England Biosciences	M0201
Experimental Models: Cell Lines		
HEK293T cells	ATCC	N/A
Recombinant DNA		
pFastBac1	Life technologies	10360014
pMS2-eGFP	Addgene	#27121
pcDNA3.1	Addgene	V790-20
pmU6-gRNA	Addgene	#53187
Sequence-Based Reagents		

Random Hexamer	ThermoFisherScientific	N8080127
$(GA)_{20}$-2xMS2 GAGAGAGAGAGAGAGAGAGA GAGAGAGAGAGAGAGAGAGA CTTGTTGGATGATCATGGACA TGAGGATTACCCATGTACTAG TAGATCTCGTACACCATCAGG GTACGTTCCAGCTAGCTTTTT TT	This Paper	N/A
$(GGAA)_{10}$-2xMS2 GGAAGGAAGGAAGGAAGGAA GGAAGGAAGGAAGGAAGGAA CTTGTTGGATGATCATGGACA TGAGGATTACCCATGTACTAG TAGATCTCGTACACCATCAGG GTACGTTCCAGCTAGCTTTTT TT	This Paper	N/A

(G3A4)3-2xMS2 GAAAAAAGGGAAAAGGGAAA AGGGAAAAGAGAGAAAAAAA CTTGTTGGATGATCATGGACA TGAGGATTACCCATGTACTAG TAGATCTCGTACACCATCAGG GTACGTTCCAGCTAGCTTTTT TT	This Paper	N/A
(G3A4)4-2xMS2 GAAAAAAGGGAAAAGGGAAA AGGGAAAAGGGAAAAAAAAA CTTGTTGGATGATCATGGACA TGAGGATTACCCATGTACTAG TAGATCTCGTACACCATCAGG GTACGTTCCAGCTAGCTTTTT TT	This Paper	N/A
qPCR_MS2tag1_F ATGATCATGGACATGAGGATT ACCC	This Paper	N/A
qPCR_MS2tag1_R and Northern Probe AGCTAGCTGGAACGTACCCT GATGG	This Paper	N/A

Software and Algorithms		
SRA Toolkit	Kodama et al., 2012	https://www.ncbi.nlm.nih.gov/sra
Bowtie2	Langmead and Salzberg, 2012	http://bowtie-bio.sourceforge.net/bowtie2
bedtools	Quinlan and Hall, 2010	http://bedtools.readthedocs.io/
MACS2	Zhang et al., 2008	https://github.com/taoliu/MACS
motif_search_v1-10.pl	This paper	https://github.com/cdavidov/motif_search_v1-8/blob/master/motif_search_v1-10.pl
MATLAB	Mathworks	www.mathworks.com
Other		
NuPage NOVEX 4-12% Bis-Tris Protein Gels	ThermofisherScientific	Cat# NP0321BOX
NOVEX WedgeWell 10-20% Tris-Glycine Mini Gels	ThermofisherScientific	Cat#XP10202BOX
ANTI-FLAG M2 Affinity	Sigma	A2220
Protease inhibitor Cocktail	ThermoFisher-Pierce	88266
Amersham Hybond N+	FisherScientific	RPN303B

Cell Lines

HEK293T cells were grown in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% fetal bovine serum (FBS), 2 mM GlutaMAX-I, 100 units/ml penicillin and 100 µg/ml streptomycin.

Protein Expression and Purification

Human PRC2-5m complexes (EZH2, EED, SUZ12, RBBP4 and AEBP2) were expressed in insect cells and purified as previously described (Davidovich et al., 2014) with adding additional Heparin column. Briefly, protein complex was bound to the amylose resin and washed thoroughly, followed by elution with 10 mM maltose. The elution was concentrated to ~15 mg/ml (Amicon Ultra-15 Centrifugal Filter Unit, 30 kDa MWCO, Millipore cat # UFC903024). Protein complex was digested by PreScission protease at a mass ratio of 1:50 protease:protein and the salt concentration was adjusted to ~250 mM NaCl. After 16-20 h incubation at 4°C, protein complex was run on SDS-PAGE to check for cleavage efficiency. The protein was injected into 5 ml HiTrap Heparin column (GE, 17-0407-03) with a gradient over 35 column volumes from Buffer A (10 mM Tris-HCl pH 7.5 at RT, 150 mM NaCl, and 1 mM TCEP) to Buffer B (10 mM Tris-HCl pH 7.5 at RT, 2 M NaCl, and 1 mM TCEP), with a 1.5 ml/min flow rate. The PRC2-peak fractions were identified using SDS-PAGE, pooled and concentrated as above. The concentrated protein was injected into HiPrep 16/60 Sephacryl S-400 HR with running buffer (250 mM NaCl, 10 mM Tris-HCl pH 7.5 at RT, 1 mM TCEP-pH 7). Protein fraction was collected with a flow rate of 0.5 ml/min. PRC2-peak fractions were identified using SDS-PAGE, pooled and concentrated as above. Final protein concentration was determined by absorbance

at 280 nm and the ratio of absorbance at 260 nm/280 nm < 0.7 was observed, an indication of no nucleic acid contamination.

Electrophoretic Mobility Shift Assay (EMSA) for PRC2-RNA Binding

All 40-mer RNAs were synthesized by GE-Dharmacon. RNA was radiolabeled as previously described (Davidovich et al., 2014) except RNA was dissolved in TE, pH 7.5 after PAGE purification. (Previously, we dissolved RNA in Milli-Q pure water, but we observed RNA degradation upon prolonged boiling in water.) The counts of purified RNAs were determined by liquid scintillation counting. Radiolabeled RNA, with specific activity no less than 100 000 cpm/pmol, was adjusted to a volume of 60 µl with TE, pH 7.5. The diluted RNA was heated for 10 min at 95°C to ensure melting of folded structure and snap-cooled on ice for 3-5 min. RNA was then allowed to fold for 30 min at 37°C in 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, 0.1 mg/ml fragmented yeast tRNA (Sigma cat # R5636), 5% v/v glycerol). The success of this method in breaking up intermolecular complexes was confirmed by native gel electrophoresis (see below). Next, stock protein was diluted with binding buffer, added and allowed to bind with radiolabeled RNA (1000cpm/lane) at 30°C for 30 min. Sample was loaded to non-denaturing 0.7% agarose gel (SeaKem® GTG® Agarose, Fisher Scientific cat # BMA 50070) buffered with 1XTBE at 4°C. Gel electrophoresis was carried out for 90 min at 66 V in an ice box within a 4°C cold room. Gels were vacuum dried for 60 min at 80°C on a Hybond N+membrane (Amersham, Fisher Scientific 45-000-927) and two sheets of Whatman 3 mm chromatography paper. Dried gels were exposed to phosphorimaging plates and signal acquisition was performed using a Typhoon Trio phosphorimager (GE Healthcare). For

PRC2-RNA binding in LiCl, 100 mM KCl was substituted with 100 mM LiCl in binding buffer.

Circular Dichroism Spectroscopy

RNA (10 μM) was folded in the same way as in PRC2-RNA binding condition (100 mM KCl or 100 mM LiCl). CD spectra of RNAs at a volume of 170 μl and a concentration of 10 μM were recorded in a 0.1 cm path- length quartz cell using a Chirascan-plus qCD spectropolarimeter (Applied Photophysics). The CD spectra were obtained at 0.5 nm intervals from 220 to 350 nm. The temperature of the cell holder was regulated with a temperature controller, and the cuvette chamber was flushed with a constant stream of dry N_2 gas to prevent the condensation of water on the cuvette exterior. CD spectra were measured at three different temperatures: 20 °C, 30 °C and 90 °C.

Native Gel Electrophoresis

10% polyacrylamide gels (29:1 ratio of acrylamide to bisacrylamide) were poured (0.75 mm thickness) using 0.5X TBE buffer. KCl or LiCl gels contained an additional 100 mM KCl or LiCl, respectively. These gels were run at 30 °C, the same temperature as the protein-binding studies. The running apparatus was completely bundled with a heat-exchanger connected to a circulating warm water bath (temperature set at 30 °C). The gels were run at ~166 V (~10 V/cm) at a maximum power of 10 W. The temperature of the gel was monitored and was around 31-32 °C during the entire running. Salt-containing gels also have salt in the running buffer. RNA samples were prepared in the same way as in PRC2-RNA binding (see above). 10 μl (1000 cpm) of each RNA samples were mixed with 1.0 μl of 60% glycerol containing marker dyes and 0.5X TBE or 0.5XTBE-100 mM KCl or 0.5X TBE-100 mM LiCl. Gels were pre-run for 1 h at 30 °C and samples were

loaded onto the gel. The gels were stopped running until the bromophenol blue dye hit 13 cm measured from the bottom of the well (~6-12 h).

RNA or G4-ligand Competition Assay

Competition reactions were set up identical to PRC2-RNA binding reactions (see above) except a constant concentration of PRC2 was included and variable concentrations of competitor RNAs or G4-ligands (TMPyP4 or PDS) 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 RNA in the absence of any competitor RNAs.

UV Crosslinking of RNA-PRC2 Complexes

RNA samples were either transcribed with 4-thio-UTP (TriLink Biotechnologies) (for HOTAIR400), or synthesized (GE-Dharmacon) with 4-thio-U incorporated into the oligo (16-40-mers). HOTAIR400 transcription included alpha-^{32}P-CTP to body label the RNA, and the short oligos (16-40-mers) were end- labeled with gamma-^{32}P-ATP as described above. Trace amounts of hot RNA were refolded (as above, except in some instances BSA and tRNA were omitted from the refold buffer) or used without refolding and incubated with 200 nM-1 μM PRC2 protein. PRC2 or PRC2 that contained an uncleavable MBP tag on one of the subunits was incubated with RNA for 30 min at 30 °C. Samples were moved to siliconized glass coverslips on ice, placed in a Stratalinker with 365 nm bulbs, and exposed to light for 10-30 min. Sample were diluted with SDS loading dye and loaded onto a Nupage 4-12% Bis-Tris gel (Life Technologies). Gel electrophoresis was carried out for 60 min at 180 volts. Gels were vacuum dried at 80 °C

for 30 min on Whatman 3mm chromatography paper. Dried gels were exposed to phosphorimaging plates and signal acquisition was performed using a Typhoon Trio phosphorimager (GE Healthcare).

RNA Immunoprecipitation

HEK293T cells were grown in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% fetal bovine serum (FBS), 2 mM GlutaMAX-I, 100 units/ml penicillin and 100 µg/ml streptomycin. $5.5x10^6$ HEK293T cells were seeded in a Corning 10 cm^2 tissue culture dish 24 hours before transfection. The next day, HEK293T cells were transfected with Lipofectamine 2000 (Life Technologies) using 5.5 µg of DNA per plasmid. Three plasmids were used: PCDNA3.1-3xFlag-Halo-EZH2, P_MU6-test RNA-MS2, and pMS2-eGFP (Addgene: #27121). Transfection was carried out in media without antibiotics (DMEM, 10% FBS, and 2 mM GlutaMax-I). Cells were incubated at 37°C/5% CO_2 for 24 hours and then harvested by trypsin and collected by centrifuging at 100xg for 5 min. Cell pellet was washed once with 1 mL cold PBS, transferred to an Eppendorf tube and re-centrifuged for 5 min at 100xg.

All of the PBS was removed and the cell pellet was resuspended in 1 mL cold lysis buffer (25 mM Tris-HCl pH 7.5, 5% glycerol, 150 mM NaCl, 2.5 mM $MgCl_2$, 2 mM TCEP pH 7.0, 1x Protease Inhibitor Cocktail (ThermoFisher-Pierce: 88266), 1% NP-40 Substitute (Sigma: 74385)) and rotated at 4°C for 30 min. Lysate was cleared by centrifuging at 16,100xg for 10 min. 12.5 µL of packed anti-flag M2 affinity beads (Sigma: A2220) were used per IP and pre-equilibrated by washing 3 times with 1 mL of cold lysis buffer per wash. The final wash was removed and beads were re-suspended in 50 µL of

lysis buffer per IP. 50 µL of the anti-flag bead resuspension was applied to 150 µL of cleared lysate and rotated at 4°C for 2 hours.

Anti-Flag beads were centrifuged at 100 rcf and washed 4x with 1mL of room temperature wash buffer (25 mM Tris-HCl pH 7.5, 5% glycerol, 500 mM NaCl, 2.5 mM $MgCl_2$, 2 mM TCEP pH 7.0, 0.1% TritonX-100). The final wash was removed and beads were prepared for RNA or protein analysis.

For RNA analysis, beads were re-suspended in 1 mL Trizol containing 2 µg of glycogen. Trizol extraction was carried out according to the manufacturer's protocol and RNA extracts were treated with RQ1 DNase (Promega: M6101). DNase treatment was terminated by performing a phenol:chloroform:isoamyl alcohol purification. Purified RNA was reverse transcribed using SuperScript III (Life Technologies: 180800-044) with random hexamer (ThermoFisherScientific, N8080127) and oligo $d(T)_{18}$ (ThermoFisherScientific, 18418012). qPCR was performed using a LightCycler 480, iQ SYBR Green Supermix (Bio-Rad: 1708882), and primers to the MS2 constant region.

RNA sequences:

$(GA)_{20}$-2xMS2

GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGACTTGTTGGATGA

TCATGGACATGAGGATTACCCATGTACTAGTAGATCTCGTACACCATCAGGGTA

CGTTCCAGCTAGCTTTTTTT

$(GGAA)_{10}$-2xMS2

GGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAACTTGTTGGATGA

TCATGGACATGAGGATTACCCATGTACTAGTAGATCTCGTACACCATCAGGGTA

CGTTCCAGCTAGCTTTTTTT

(G3A4)3-2xMS2

GAAAAAAGGGAAAAGGGAAAAGGGAAAAGAGAGAAAAAAACTTGTTGGATGATC

ATGGACATGAGGATTACCCATGTACTAGTAGATCTCGTACACCATCAGGGTACG

TTCCAGCTAGCTTTTTTT

(G3A4)4-2xMS2

GAAAAAAGGGAAAAGGGAAAAGGGAAAAGGGAAAAAAAAACTTGTTGGATGATC

ATGGACATGAGGATTACCCATGTACTAGTAGATCTCGTACACCATCAGGGTACG

TTCCAGCTAGCTTTTTTT

Northern Blot

Trizol (1 mL) containing 2 µg of glycogen was added to 100 µL of cell lysate that was prepared as described in the RNA Immunoprecipitation section. Trizol extraction was carried out according to manufacturer's protocol and RNA pellets were re-suspended in 15 µL of TE pH 7.0 (10 mM Tris-HCl pH 7.0, 1 mM EDTA). 1 µg of RNA extract was heated at 92°C for 5 min in 51% formamide, 16 mM EDTA, .5x TBE, .025% Xylene Cyanol. A 6% polyacrylamide gel containing 7 M urea and 1x TBE was pre-run for one hour at 30 watts prior to loading boiled samples. The samples were run at 30 watts for one hour and transferred onto a Hybond N+ membrane (GE Healthcare) in 1xTBE for 1 hour at room temperature. The membrane was crosslinked at 1200 mJ/cm^2 and pre-hybridized for 30 min at 50°C in Church buffer (0.5M sodium phosphate, 1mM EDTA, 7% SDS, 1% BSA). MS2_tag1 R oligo was ^{32}P end-labeled using T4 Polynucleotide Kinase (New England Biosciences: M0201) and hybridized in Church buffer at 50°C overnight.

Western Blot

HEK293T cells were lysated as described in the RNA immunoprecipitation section. 50 µL of lysate was saved for Western blot analysis. 950 µL of lysate was used for pull-down with Anti-flag beads. Standard Western blot protocol was used and antibodies include: Anti-FLAG M2-HRP (Sigma A8592, 1:2500), Anti-SUZ12 (Santa Cruz sc67105, 1:200), and Goat Anti-Rabbit IgG-HRP (Santa Cruz sc2301, 1:5000).

QUANTIFICATION AND STATISTICAL ANALYSIS

Quantification of K_d

Equilibrium dissociation constant (K_d) was quantified from binding curves based on standard quantifications of EMSA (Goodrich and Kugel, 2007). Unless stated otherwise, reported values represent the average and standard deviation of three independent experiments performed on different days.

RNA-Immunoprecipitation

Data was analyzed as percent to input = $(1/(2^{Ct(IP)-Ct(Input)}))$*dilution_factor*100%. Background (expression and IP of RNA sequences in the absence of 3xFlag-Halo-EZH2) was subtracted and biological replicates were averaged.

Quantification of Genome- and Transcriptome-wide Association between Proteins and Putative Motifs

Binding sites of proteins on RNA *in vivo* were identified from published fRIP-seq data (Hendrickson et al., 2016), and ENCODE ChIP-seq data were used for protein-binding sites on DNA (See Supplemental Table S1 for NCBI GSE accession numbers). Short reads were downloaded from the NCBI Gene Expression Omnibus (GEO) repository and converted to fastq files using the SRA Toolkit (Kodama et al., 2012). Data were mapped to the reference genome (hg19) using Bowtie2 (Langmead and Salzberg,

2012) with the option -q. In order to identify protein binding sites (PBS), output sam files were converted to bam files (bedtools; Quinlan and Hall, 2010) and peaks were called using MACS2 (Zhang et al., 2008), 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/blob/master/motif_search_v1-10.pl). Overlapped motifs were merged by bedtools merge (Quinlan and Hall, 2010) using the -s option, to force merging of motifs from the same strand. Next, the number of PBS-motif association events determined by counting overlaps between PBS and motifs using bedtools closest (Quinlan and Hall, 2010) 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, we empirically quantified the number of these association events that could be expected to be obtained by chance. To this end, motifs were randomly shuffled using bedtools shuffle (Quinlan and Hall, 2010) 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 (PDF) of the number of PBS-motifs association events that were obtained by chance. The empirically obtained PDF was then modelled to a Poisson distribution (MATLAB). The modelled Poisson PDF 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 modelled Poisson PDF.

Data quantifying the association between PRC2 and motifs of the form $[G_nN_m]_k$, with different lengths or G-tracts (n) or different number of repeats (k), were generated using the same pipeline, with the following modification: considering a given motif of the form $[G_nN_m]_k$, motifs represented specific n or k were retained only if they were not contained also within motifs that were identified for n+1 or k+1, respectively. Specifically, retention of motifs that are characterized by exactly n or k was done using bedtools intersectBed (Quinlan and Hall, 2010) using the option -v (exclusion), with -a as the motif file representing a given n or k and -b as the motif file for the corresponding n+1 or k+1.

Chapter IV: PRC2 recruitment to DNA in chromatin and its inhibition by RNA reveal molecular mechanisms of epigenetic control

INTRODUCTION

In the preceding chapter, we revealed hidden specificity of RNA binding to PRC2 in great detail. However, the interplay of PRC2 between RNA and chromatin remained poorly understood. Thus, I became interested in comparing the relative affinity of PRC for RNA and nucleosomes. The experiments presented in this chapter were the result of collaboration with Richard Paucek. Richard and I contributed equally to the design of all experiments, their execution, data analysis and authorship of the paper.

Polycomb Repressive Complex 2 (PRC2) is a histone methyltransferase that specifically deposits methyl groups onto lysine 27 of histone H3. PRC2 is absolutely required for epigenetic silencing during embryonic development and cancer, and the importance of PRC2 for stem cell renewal and pluripotency is highlighted by the early embryonic lethality of mice with deletions of these genes(Margueron and Reinberg, 2011). In *Drosophila*, recruitment of Polycomb group proteins to target genes involves DNA sequences called Polycomb Responsive Elements (PREs)(Müller and Bienz, 1991), (Mihaly et al., 1998). However, how PRC2 is recruited to sites of action in mammalian cells has remained poorly understood.

Earlier work suggested that RNA could contribute to targeting PRC2 to specific genomic loci(Zhao et al., 2008b) or holding it “poised but in check”(Kaneko et al., 2014b). PRC2 clearly interacts with RNA *in vitro* and *in vivo*, in both cases binding so many RNAs that its binding could be characterized as promiscuous (Davidovich et al., 2013b), (Davidovich et al., 2015b). Recently, our group demonstrated that PRC2 reads RNA motifs consisting of short repeats of consecutive guanines, which are ubiquitous in

the human transcriptome (Wang et al., 2017). While the hidden specificity in promiscuous RNA-binding has been revealed, the question of RNA functionality still remains: what role do RNA molecules play in the regulation of PRC2?

Two studies set forth the idea that RNA inhibits PRC2 catalytic activity on histone H3K27 (Kaneko et al., 2014b), (Cifuentes-Rojas et al., 2014b), although mechanistic details were not provided. New insight came from the Jenner lab, who concluded via pull-down experiments that purified nuclear RNA antagonizes PRC2 binding to chromatin (Beltran et al., 2016b). Despite being an important initial observation, chromatin includes nucleosomes with different histone modifications, and whether such modifications would counteract RNA inhibition was not explored.

Pre-existing histone modifications have been proposed to be important determinants of PRC2 recruitment to genomic loci. Prior studies have shown that the PRC2 complex reads H3K27me3 marks of repressive chromatin and that binding to these histone modifications stimulates PRC2 enzymatic activity (Margueron et al., 2009). On the other hand, H3K4me3 and H3K36me3 marks of active chromatin are also recognized by PRC2, and such binding events inhibit PRC2 catalytic activity (Schmitges et al., 2011b). A histone-based perspective of PRC2 recruitment is epitomized by studies of pediatric high-grade gliomas, where it was suggested that a H3K27M missense mutation is able to sequester PRC2 and suppress its enzymatic function (Lewis et al., 2013). Although the equilibrium dissociation constants (K_d) for interaction of various modified histone peptides with PRC2 have been measured (Xu et al., 2010) and the binding of modified histone peptides even visualized at atomic resolution (Jiao and Liu, 2015b), the nucleosome-binding activity of PRC2 has remained thus far understudied (Nekrasov et

al., 2005). In addition, the affinity of PRC2 for nucleosomes *vis-a-vis* RNAs is unknown, and exploring this gap in knowledge is imperative for clarifying standing models for PRC2-RNA functionality in mammals.

Here we undertook quantitative binding studies by taking advantage of a completely recombinant reconstituted chromatin system, and we demonstrated that human PRC2 has a higher affinity for nucleosome arrays. Through competition experiments, we observed that RNA is able to passively evict PRC2 from nucleosomes and thereby suppress H3K27 methylation; RNA is not an active-site inhibitor of PRC2. Binding analysis of PRC2 with nucleosomes incorporating different histone modifications (H3K27me3, H3K4me3, and H3K27M) revealed that these modifications surprisingly do not contribute much to PRC2 affinity for nucleosome substrates. Instead, we found that human PRC2 has robust DNA-binding activity, and – contrary to the common view in the field – it is the protein-free linker DNA that dictates PRC2 affinity to nucleosomes. Furthermore, RNA antagonizes PRC2 interaction with DNA.

In exploring specificity of PRC2 DNA-binding activity, previous bioinformatics analyses showed an enrichment of PRC2 peaks at CG-rich DNA regions (Wang et al., 2017), (Mendenhall et al., 2010b). In agreement with these past observations, we found that PRC2 complexes have enhanced binding affinity for CG-rich DNA *in vitro*. Strikingly, we also found that CpG-methylated DNAs are preferentially bound by PRC2, and that it is the AEBP2 subunit that discriminates between unmethylated and methylated DNAs. Furthermore, purified PRC2 complexes incorporating JARID2 or EZH1 subunits were found to alter the enzymatic activity of PRC2, but otherwise exhibited similar binding affinity and trends for RNA, DNA, and nucleosomes. Remarkably, these intrinsic

properties of the interaction of PRC2 with nucleosomes provide a straightforward explanation for previously perplexing observations of PRC2 *in vivo*.

RESULTS

PRC2 binds longer nucleosome arrays with increased affinity and cooperativity

To quantify the binding affinities of human PRC2 with nucleosomes, we assembled 601-positioned mononucleosomes by standard salt gradient dialysis (Figure 4.1a. b). The 601-Widom sequence positions the nucleosome symmetrically with a DNA linker at each end (Dyer et al., 2004). Our nucleosome assemblies underwent quality control using EcoRI-digestion and atomic-force microscopy (AFM) to verify that nucleosomes were neither under- nor over-saturated with histones (Figure 4.2a and Figure 4.1c-e).

Recombinant human PRC2 5-mer complex (EZH2, EED, SUZ12, RBBP4, and AEBP2) was purified as previously described (Davidovich et al., 2013b)(Figure 4.1f-g). Binding of PRC2 to nucleosomes was measured by an Electrophoretic Mobility Shift Assay (EMSA), and the shifted bands were shown to contain PRC2 as well as nucleosomes (Figure 4.1h). As determined by EMSA, the apparent dissociation constant (K_d^{app}) for mononucleosomes was 280 $\pm$ 19 nM (throughout the chapter, errors = SD, n = 3; Figure 4.2b-c). Notably, this binding

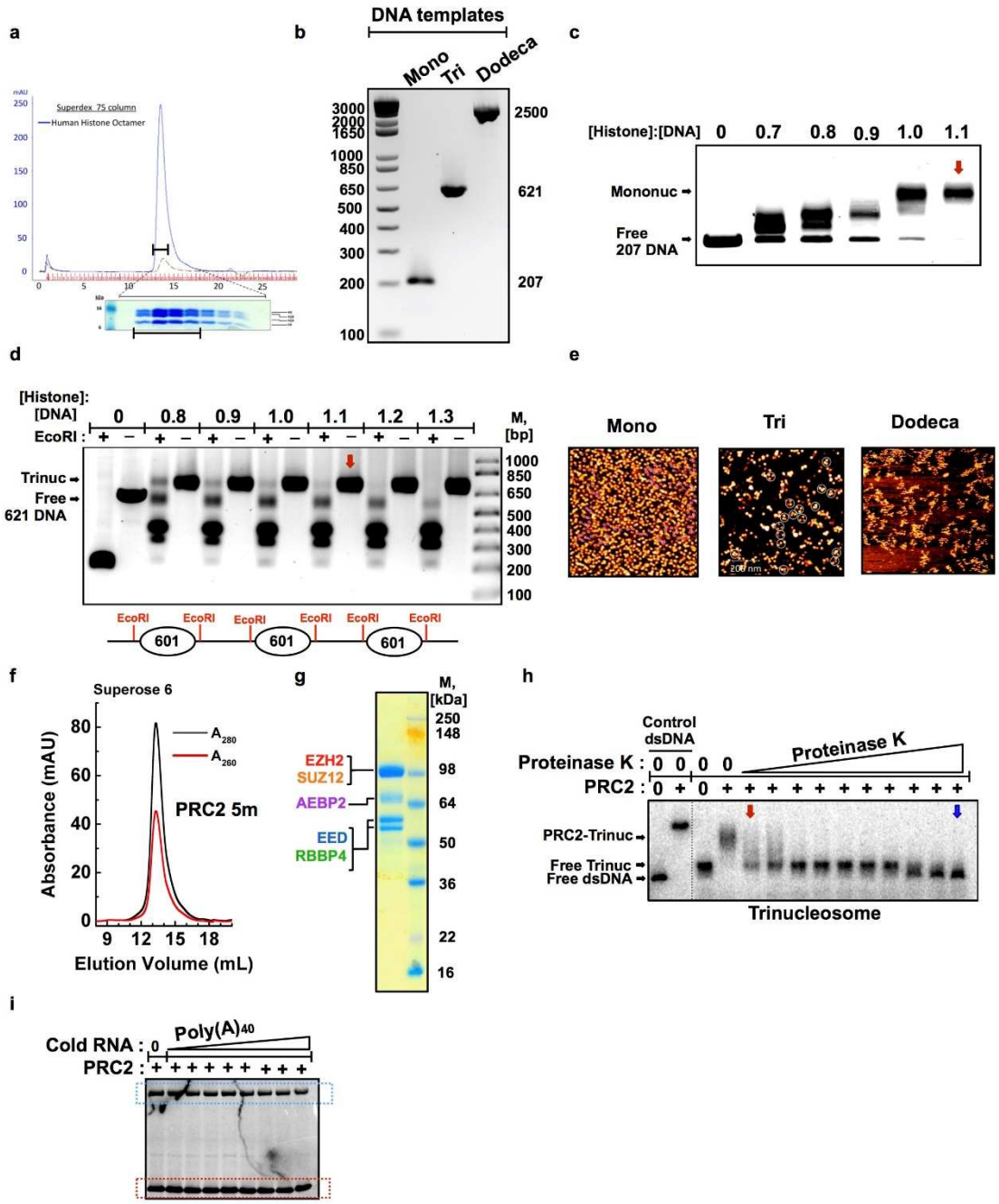


Figure 4.1. Purification of human PRC2 5-mer complex and its interaction with nucleosomes. A. Elution profile of human histone octamer shows that the octamer is monodisperse by size-exclusion chromatography on a Superdex 75 column. Only fractions that had stoichiometric amounts of each histone subunit (1 [H2A]:1 [H2B]:1 [H3]:1 [H4]), as assessed by SDS-PAGE, were pooled, concentrated, and used for nucleosome reconstitution. **B**. DNA agarose gel showing the purity of tandemly repeated nucleosome positioning DNA used in the reconstitution of mononucleosomes (207 bp), trinucleosomes (621 bp), and dodecanucleosomes (2500 bp). **C.** Representative agarose gel showing the assembly of mononucleosomes as the ratio of histone octamer-to-DNA is increased. In this example, the red arrow indicates a ratio at which mononucleosomes were fully assembled. **D.** To ensure that histone octamers were properly assembled on the DNA, preparative gel electrophoresis was performed on trinucleosomes treated with EcoRI. The accessibility of the EcoRI cleavage sites show how saturated the nucleosomes are on the DNA. The red arrow indicates an ideal trinucleosome sample with accessible EcoRI

restriction sites. **E**. Representative AFM images of mono-, tri-, and dodecanucleosomes shows homogeneity in assembled nucleosomes over a large field of view. **F**. PRC2 is monodisperse by size-exclusion chromatography on a Superose 6 Increase column. The lack of nucleic acid contamination is indicated by the A260/A280 ratio < 0.7 and by the absence of staining by ethidium bromide (the latter is not shown here). **G.** Purity of a typical PRC2 5-mer complex is assessed by SDS-PAGE. **H**. The slow-migrating EMSA band is disrupted by proteinase K treatment (red arrow), confirming that it is a PRC2-nucleosome complex rather than an oligomerization of the nucleosome component stimulated by PRC2. At the highest proteinase K concentration, the trinucleosome band shifted to a free dsDNA band (blue arrow). **I.** Histone methyltransferase (HMTase) assays show that Poly(A)$_{40}$ RNA does not affect H3K27 methylation. EZH2 automethylation signal is indicated by the blue-dashed box, and H3K27 methylation signal is indicated by the red-dashed box. ^{14}C-SAM was used as a cofactor in the HMTase reactions. Data points represent 2-fold titrations ranging from 0–60 μM RNA.

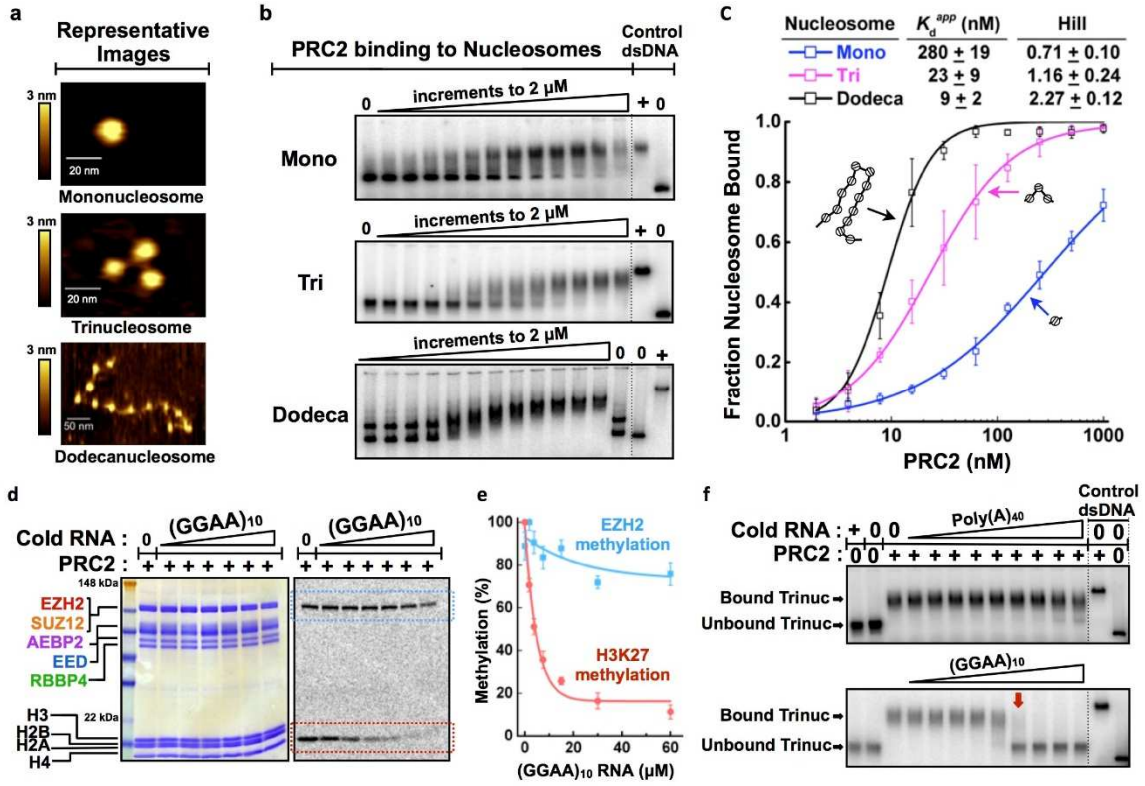


Figure 4.2. RNA controls H3K27 methylation status by preventing PRC2 binding to nucleosomes. A. Representative topographic images of mono-, tri-, and dodeca-nucleosomes, as visualized by atomic-force microscopy, show the success of the 601-sequence in positioning nucleosomes. **B**. PRC2 binding to ^{32}P-radiolabeled mono-, tri- and dodeca-nucleosome substrates is examined by EMSA. Data points represent 2-fold titrations ranging from 0–2 μM PRC2. **C**. EMSA data fit with equilibrium binding curves shows that PRC2 binds to nucleosome arrays with increased affinity and cooperativity. Error bars give SD, n = 3. In this figure and others, several of the error bars are not visible because they are hidden by the size of the data point symbol. **D**. Histone methyltransferase (HMTase) assays show that (GGAA)$_{10}$ RNA inhibits H3K27 methylation. The persistence of EZH2 automethylation signal (blue-dashed box), but not H3K27 methylation signal (red-dashed box) indicates that RNA is not an active site inhibitor. ^{14}C-SAM was used as a cofactor in the HMTase reactions. PRC2 and histone proteins were visualized by

Coomassie-staining (left-hand gel) and methylation levels were determined by ^{14}C-autoradiography (right-hand gel). Data points represent 2-fold titrations ranging from 0–60 μM RNA. **E**. Quantification of autoradiography gel in (d). Data fit with exponential decay curves. Error bars give SD, n=3. **F**. Pre-formed PRC2-trinucleosome complexes are disrupted by $(GGAA)_{10}$ RNA, but not by negative control $Poly(A)_{40}$ RNA. PRC2-trinucleosome complexes were fully disrupted (lane with dashed red box) at a PRC2:RNA molar ratio of ~2:1. Data points represent 3-fold titrations ranging from 0–10 μM RNA.

is dramatically stronger than the 40–400 μM K_d^{app} range measured for PRC2 binding to various H3 peptide substrates (Margueron et al., 2009), (Xu et al., 2010).

Yuan *et al.* (Yuan et al., 2012b) reported that neighboring nucleosomes activate PRC2; therefore, we hypothesized that PRC2 might bind head-to-tail nucleosome arrays with a lower K_d^{app} than mononucleosomes. We assembled tri- and dodeca-nucleosome arrays, both with a linker DNA length of 60 base pairs. As shown in Figure 4.2b and c, the K_d^{app} values of PRC2 for tri- and dodeca-nucleosomes were 23 $\pm$ 9 nM and 9 $\pm$ 2 nM, respectively. Notably, we did not adjust these K_d^{app} values for the number of binding sites on tri- and dodeca-nucleosomes, so the affinity per nucleosome increases from mono- to tri-nucleosomes but does not increase further for dodecanucleosomes. In addition, cooperativity increased substantially for PRC2 binding to arrays in comparison to mononucleosomes (Hill coefficient n = 2.27 $\pm$ 0.12 and 0.71 $\pm$ 0.10, respectively, Figure 4.2c). Thus, PRC2 prefers binding to tandem nucleosome repeats over mononucleosomes, with enhanced affinity largely achieved with trinucleosomes, and the dodecanucleosomes binding most cooperatively.

For our EMSA experiments, controls for protein-free DNA and the PRC2-DNA complex are shown in the right-hand lanes of each gel in Figure 4.2b. These controls address some potential concerns about the binding studies. Namely, the nucleosomes do not dissociate at the subnanomolar concentrations used in the binding reaction, because protein-free DNA runs as a distinguishably separate species than nucleosomes

on the agarose gel, and no free DNA is observed in the experimental lanes. Alternatively, if nucleosomes unraveled and the released free DNA were then bound by PRC2, the resulting PRC2-DNA complex would have lower mobility than the PRC2-nucleosome complexes; no such PRC2-DNA species was observed in the experimental lanes. The exception is the dodecanucleosomes (bottom panel of Figure 4.2b), where half of the DNA is fully assembled and the other half runs as under-saturated arrays. In this case, both the fully assembled and under-saturated arrays are bound by PRC2.

RNA is not an active site inhibitor of PRC2 but instead inhibits histone methylation by sequestering PRC2 from nucleosomes

RNA has been previously shown to inhibit PRC2 catalytic activity (Kaneko et al., 2014b), (Cifuentes-Rojas et al., 2014b). To quantitatively measure RNA-mediated enzymatic inhibition, PRC2 and *in vitro* reconstituted mononucleosomes were incubated with radiolabeled S-adenosylmethionine (^{14}C-SAM) methyl donor, and RNA was titrated into the reaction. For this analysis, $(GGAA)_{10}$ RNA was used due to its optimal binding, and $Poly(A)_{40}$ was used as a negative control RNA that does not bind PRC2(Kaneko et al., 2014b), (Wang et al., 2017).

In the absence of RNA, our histone methyltransferase assays revealed the expected methylation of histone H3 (dashed red box, Figure 4.2d). We also observed automethylation of the EZH2 subunit, as has been previously reported by other groups (Müller et al., 2002b), (Sanulli et al., 2015) (dashed blue box, Figure 4.2d). As seen in Figure 4.2e, the presence of $(GGAA)_{10}$ RNA in the HMT assay dramatically inhibited H3K27 methylation but not EZH2 automethylation. $Poly(A)_{40}$ RNA, which does not bind to PRC2, had no observable inhibitory effects (Figure 4.2i).

It is striking that RNA had only a small effect on EZH2 automethylation, even at the highest RNA concentration tested (60 μM). This automethylation signal provides the basis for the key conclusion that RNA does not inhibit the active site of EZH2. While the detailed molecular analysis of EZH2 automethylation is beyond the scope of this paper (manuscript in preparation), it is useful here for the reader to know that an active-site mutation in EZH2 abolishes both automethylation and H3K27 methylation, so the methylation of EZH2 is intrinsic and not due to some contaminating protein. Thus, the persistence of automethylation in the presence of RNA indicates that the RNA is not itself an active-site inhibitor, but rather must inhibit H3K27 methylation by other means.

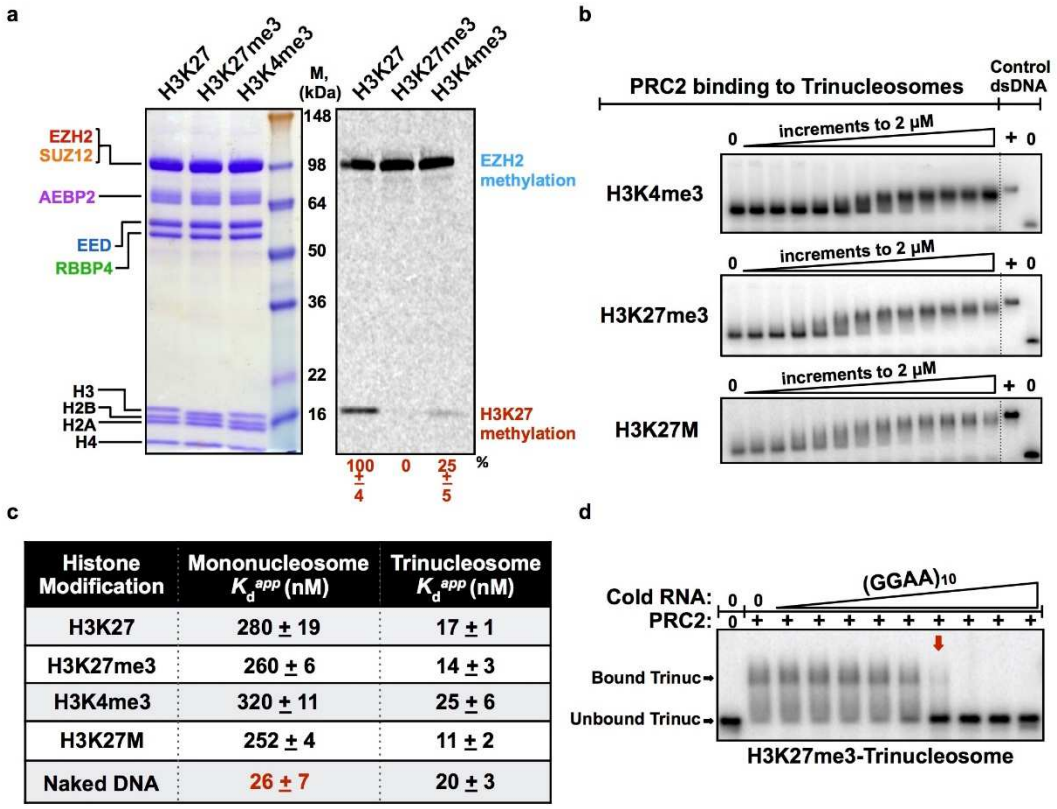
One obvious hypothesis for the mechanism of RNA inhibition is that RNA simply disrupts the association of PRC2 with nucleosomes. Therefore, we titrated unlabeled RNA with pre-formed complexes of PRC2 and radiolabeled trinucleosomes. As shown in Figure 4.2f, $(GGAA)_{10}$ RNA stripped PRC2 from nucleosomes. Dissociation from nucleosomes was substantially complete at a sub-stoichiometric RNA concentration, consistent with each RNA molecule possibly interacting with two PRC2 complexes. In contrast, $Poly(A)_{40}$ failed to compete (top panel, Figure 4.2f). Together, these data support the conclusion that RNA and nucleosomes share the same or mutually exclusive binding sites for PRC2. These findings are in agreement with a recent study suggesting that the interaction of PRC2 with RNA and chromatin is mutually antagonistic (Beltran et al., 2016b).

Histone modifications have small effects on PRC2 affinity for nucleosomes *in vitro* and RNA antagonizes PRC2 binding to modified nucleosomes

PRC2 is known to bind the N-terminal tail of H3 where it recognizes unmodified H3K27 via the EZH2 subunit, and it binds H3K4me3 and H3K27me3 at allosteric regulatory sites. Accordingly, we hypothesized that the binding of PRC2 to nucleosomes would be significantly affected if the nucleosomes carried these covalent modifications. More specifically, perhaps H3K27me3 histone modification would counteract the inhibitory effects of RNA. We therefore reconstituted mono- and tri-nucleosomes containing unmodified H3, H3K4me3, or H3K27me3 modification (see Methods) (Figure 4.4, 4.5,4.6a)**.** We validated each of the modified nucleosomes with HMT assays; PRC2 methylated the unmodified H3 nucleosomes, but not H3K27me3 nucleosomes (Figure 4.3a). Furthermore, as expected, PRC2's catalytic activity was diminished four-fold when nucleosomes contained H3K4me3 active marks (Figure 4.3a). We also prepared nucleosomes harboring a H3K27M cancer mutation, which is associated with pediatric glioma and thought to sequester PRC2 *in vivo* due to a tight binding interaction[13].

Unexpectedly, PRC2 bound all of these nucleosome variants with similar affinity (Figure 4.3b,c). Unmodified mononucleosomes were bound with K_d^{app} = 280 $\pm$ 19 nM, while H3K27me3-nucleosomes and H3K27M-nucleosomes were bound only slightly more tightly. H3K4me3-nucleosomes bound to PRC2 with an affinity that was only marginally weakened (Figure 4.2c and 4.6b). Together, these data lead to the

unanticipated conclusion that PRC2 binds nucleosomes containing different marks with similar nanomolar affinity.



Histone Modification	Mononucleosome K_d^{app} (nM)	Trinucleosome K_d^{app} (nM)
H3K27	280 ± 19	17 ± 1
H3K27me3	260 ± 6	14 ± 3
H3K4me3	320 ± 11	25 ± 6
H3K27M	252 ± 4	11 ± 2
Naked DNA	26 ± 7	20 ± 3

Figure 4.3. Histone modifications have small effects on PRC2 affinity for nucleosome substrates *in vitro*. A. Validation of unmodified and modified (H3K27me3 and H3K4me3) nucleosome variants using HMTase assays. PRC2 and histone proteins were visualized by Coomassie-staining (left-hand gel); methylation levels were visualized by ^{14}C-autoradiography (right-hand gel). Quantification of methylation signal showed that in comparison to unmodified nucleosomes (100% PRC2 activity), the catalytic activity of PRC2 was 4-fold decrease for H3K4me3-nucleosomes (25% PRC2 activity). And no methylation signal was detected for H3K27me3-nucleosomes (0% PRC2 activity). **B.** Representative EMSA gels for PRC2 binding to ^{32}P-radiolabeled modified trinucleosomes. Data points represent 2-fold titrations ranging from 0–2 μM. **C**. Table of quantifications of PRC2 binding affinity to mono- and tri-nucleosomes containing different modifications. Error bars give SD, n = 3. Quantification of EMSA data reveals that PRC2 binds to nucleosomes with different modifications with similar nanomolar affinity. Bottom row of the table shows K_d values for PRC2 binding to protein-free DNA controls (207 bp mononucleosome template DNA and 621 bp trinucleosome template DNA). **D**. $(GGAA)_{10}$ RNA disrupts pre-formed complexes of PRC2-H3K27me3-trinucleosome, indicating that histone marks do not impact the regulatory role of RNA. PRC2-H3K27me3-trinucleosome complexes were fully disrupted (dashed red box) at a PRC2:RNA molar ratio of ~2:1. Data points represent 3-fold titrations ranging from 0–10 μM RNA.

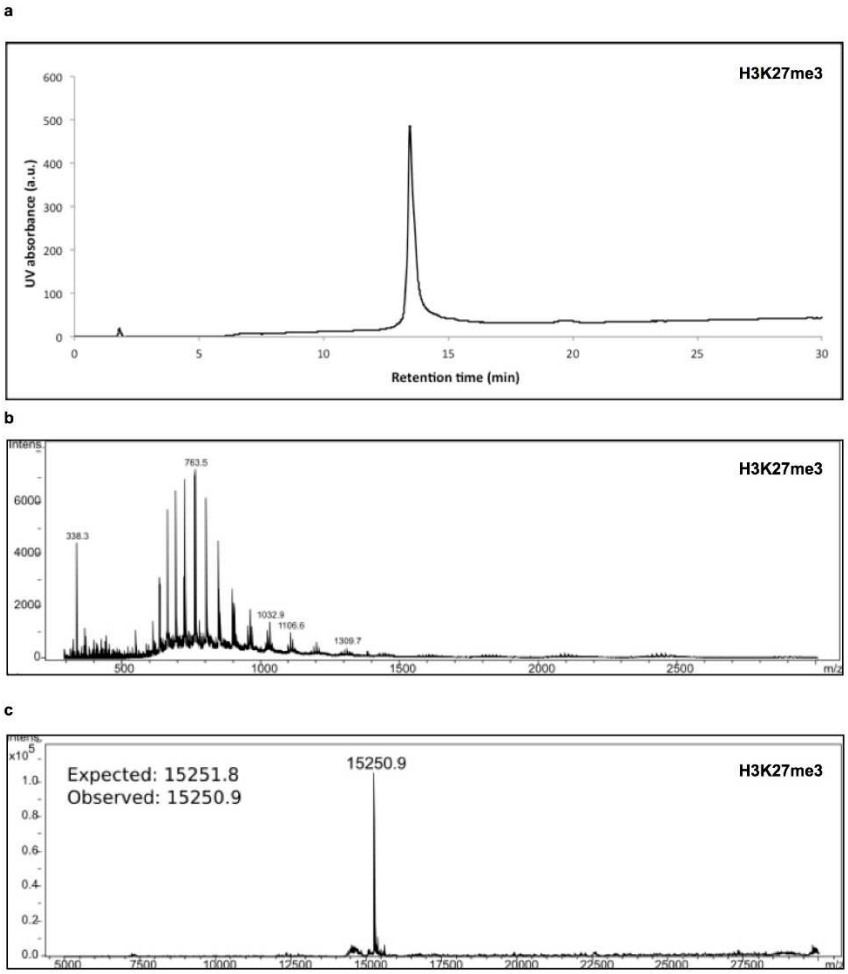


Figure 4.4. Analysis of semi-synthetic H3K27me3 histone. A. C18 analytical RP-HPLC chromatogram (gradient 0-50% HPLC solvent B) of semi-synthetic of H3K27me3 histone. **B.** Corresponding ESI-MS spectrum. **C.** Deconvolution of ESI-MS spectrum.

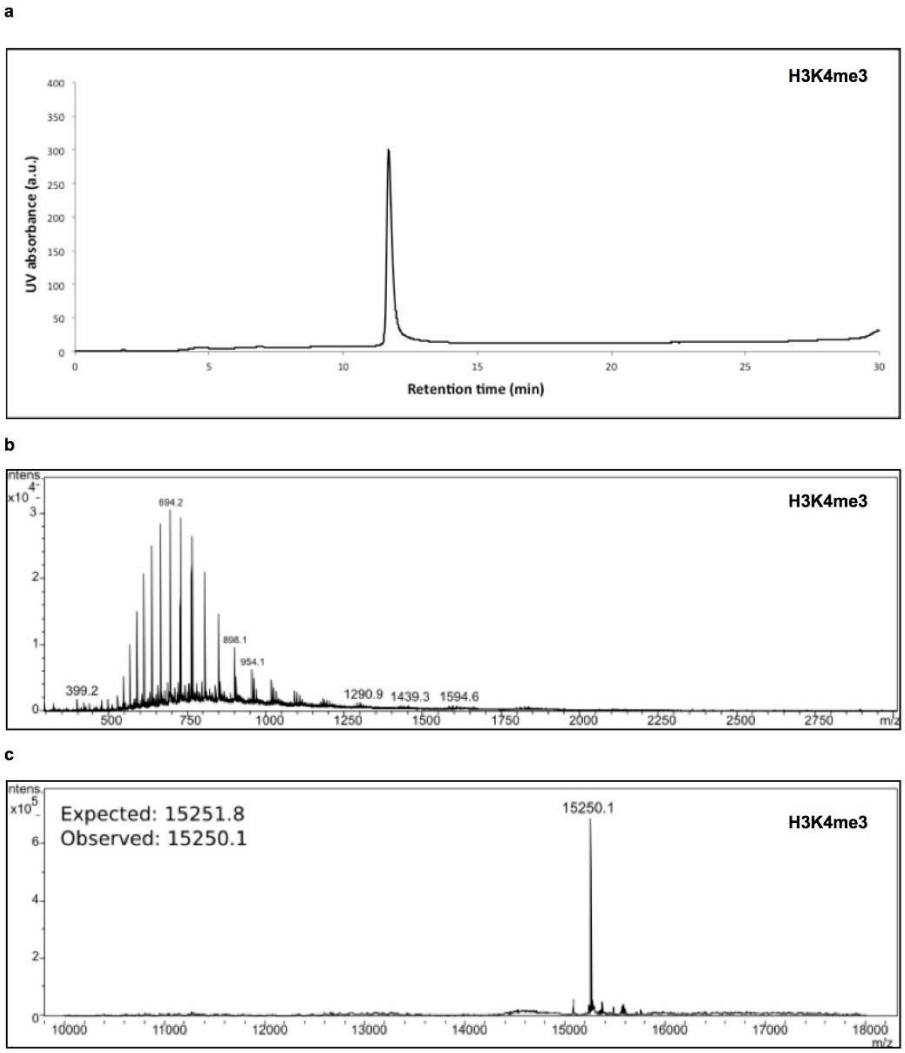


Figure 4.5. Analysis of semi-synthetic H3K4me3 histone. A. C18 analytical RP-HPLC chromatogram (gradient 0-50% HPLC solvent B) of semi-synthetic H3K4me3 histone. **B.** Corresponding ESI-MS spectrum. **C.** Deconvolution of ESI-MS spectrum.

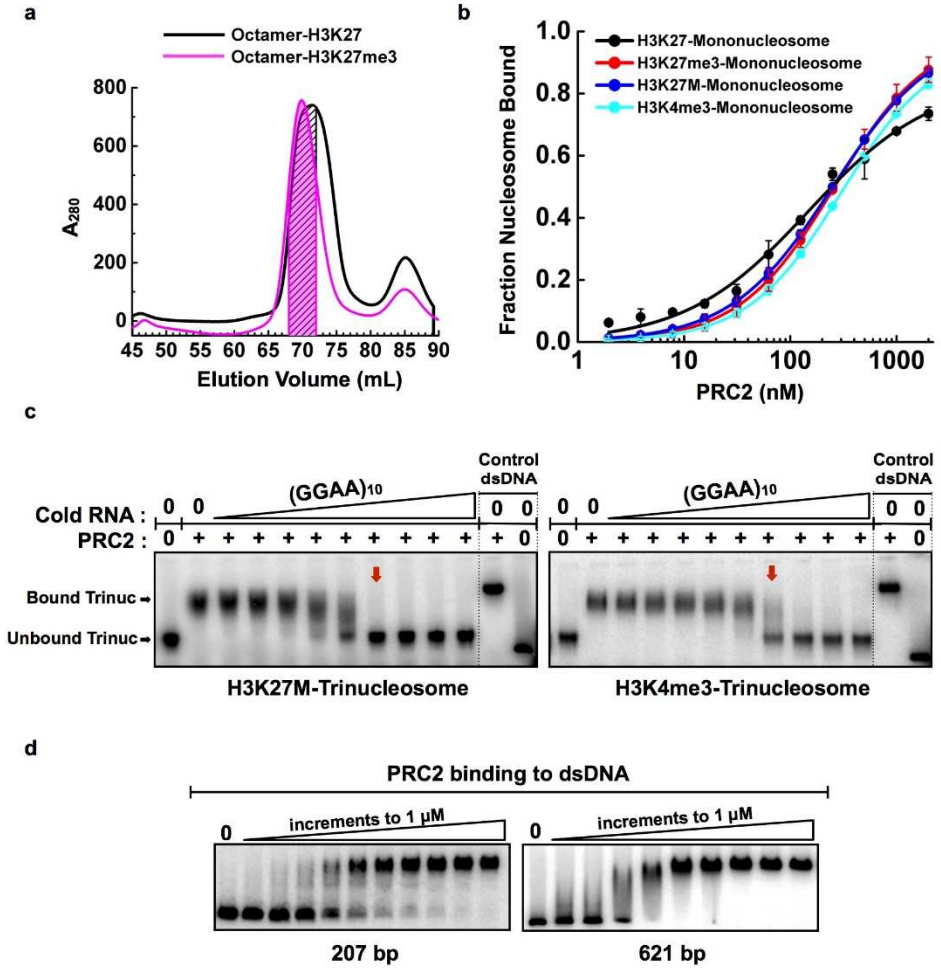


Figure 4.6. RNA strips PRC2 away from modified nucleosomes. A. The elution profile of H3K27me3-histone octamer is similar to unmodified histone octamer. Only fractions ranging from 68–72 mL (highlighted), where histone subunits were stoichiometric with one another (1:1:1:1), as determined by SDS-PAGE, were collected, pooled, and used for reconstitution experiments. Elution profiles for other modified-histone octamers were similar (data not shown). **B**. Equilibrium binding curves for EMSA data of PRC2 binding to modified mononucleosomes. **C**. $(GGAA)_{10}$ RNA disrupts pre-formed complexes of PRC2-H3K27M-trinucleosome and PRC2-H3K4me3-trinucleosome, indicating that neither histone marks nor a cancer mutation impact the regulatory role of RNA *in vitro*. All PRC2-trinucleosome complexes were fully disrupted (indicated by a red arrow) at a PRC2:RNA molar ratio of ~2:1. Data points represent 2-fold titrations ranging from 0–10 μM RNA. **D.** EMSA gels of PRC2 binding to protein-free 207 bp and 621 bp dsDNA templates. Data points represent 2-fold titrations ranging from 0–1 μM PRC2.

Given the similar binding by PRC2 measured with modified nucleosomes, we anticipated that RNA would inhibit PRC2 binding to all of these modified nucleosomes. Indeed, we observed that RNA disrupted complexes of PRC2 and H3K27me3-,

H3K4me3-, and H3K27M-nucleosomes at similar stoichiometry as observed previously for unmodified nucleosomes (Figure 4.2d and Figure 4.6c). Therefore, we conclude that histone marks contribute little to the targeting of PRC2 to nucleosomes and do not impact the ability of RNA to sequester PRC2 away from nucleosomes.

Nucleosome-free linker DNA dictates PRC2 binding to nucleosomes

While investigating nucleosome-binding properties, we tested PRC2 binding to control naked DNA templates, expecting the affinities to the DNAs to be markedly weaker than their histone-coated counterparts. Surprisingly, as shown in the table in Figure 4.2c and Figure 4.6d, the K_d^{app} values for PRC2 binding to the various naked DNA templates were all in the nanomolar range. More remarkably, PRC2's affinity for assembled mononucleosomes (K_d^{app} = 280 $\pm$ 19 nM) was much weaker than for the corresponding protein-free DNA (K_d^{app} = 26 $\pm$ 7 nM).

The simplest explanation for this unusual finding is that the histone octamer shields DNA surfaces that PRC2 would otherwise bind. In other words, we surmised that the protein-free linker regions of nucleosomes dictate PRC2 binding. To test this hypothesis, we reconstituted H2B-Cy5 labeled nucleosome core particles (NCPs) using 147 base pair template DNA and mononucleosomes using 207 base pair DNA (Figure 4.7a). Nucleosomes were purified by FPLC to assure homogeneity (Figure 4.7b). Astonishingly, once the linker DNA was removed, PRC2 binding to nucleosomes was attenuated to a micromolar affinity of K_d^{app} = 40 $\pm$ 16 μM (Figure 4.8a,b), consistent with the weak binding of PRC2 to histone tails(Xu et al., 2010). Yet, robust PRC2 binding to nucleosomes was restored with the addition of linker DNA (Figure 4.8a,b). And the strongest binding was observed for the protein-free 147 bp DNA (Figure 4.8b and Figure 4.7c). These results

with H2B-Cy5 fluorescently labeled nucleosomes were confirmed for nucleosomes labeled with ^{32}P-DNA (Figure 4.7d).

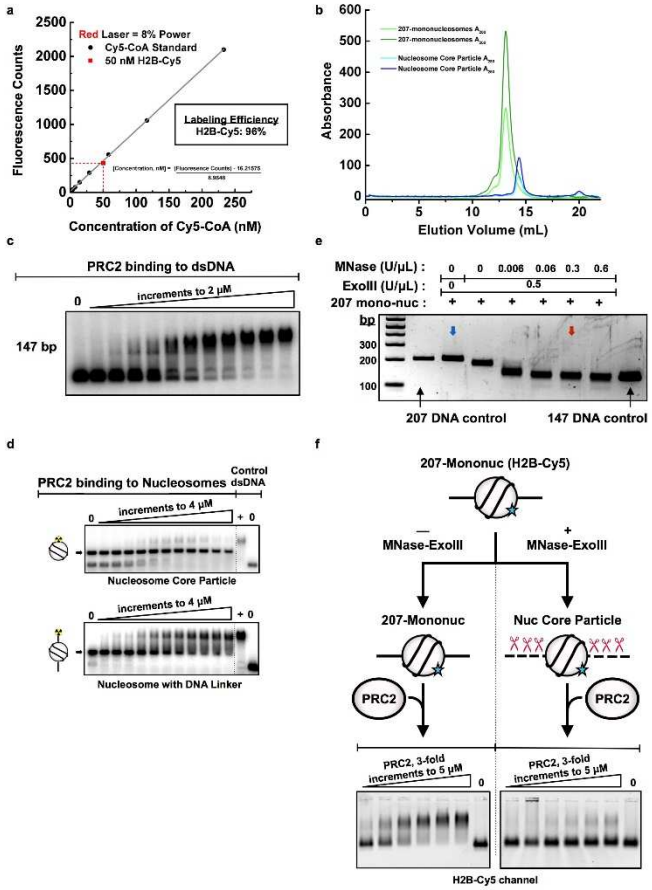


Figure 4.7. Removing linker DNA from nucleosomes attenuates binding affinity. A. Labeling efficiency of H2B-Cy5 is 96%, as determined using a Cy5-CoA calibration curve. Fluorescence signal was counted using an MST instrument (NanoTemper). **B.** Fractionation of mononucleosome and nucleosome core particles on a Superose 6 Increase column show that reconstituted nucleosomes are homogenous and monodisperse and lack contaminating free DNA. **C.** EMSA gel of PRC2 binding to protein-free 147 bp dsDNA template. Data points represent 2-fold titrations ranging from 0–2 μM PRC2. **D.** EMSA of PRC2 binding to 32P-radiolabeled nucleosome substrates. Nucleosome core particles lacking an accessible DNA linker attenuated PRC2 binding. Nanomolar binding affinity to nucleosomes was restored by the addition of a linker DNA. Data points represent 2-fold titrations ranging from 0–4 μM. **E.** Trimming of 207 bp mononucleosomes to nucleosome core particles was performed by MNase-ExoIII digestion. Following treatment of 207 mononucleosomes with MNase-ExoIII, aliquots were taken from each sample and proteinase K-treated to remove histone proteins. DNA were then purified and run on agarose gel. DNA was visualized by EtBr staining. Blue arrow indicates untrimmed-207 mononucleosome. Red arrow indicates fully trimmed-207 mononucleosome DNA. **F.** PRC2 binding to mononucleosomes with and without MNase-ExoIII treatment shows that PRC2 nanomolar binding affinity to nucleosomes is due to linker DNA binding.

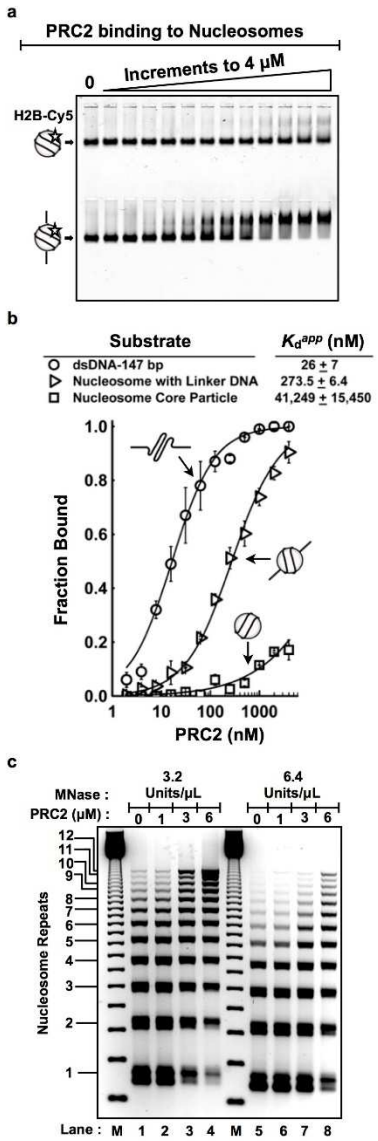


Figure 4.8. Nucleosome-free linker DNA dictates PRC2 binding to nucleosomes. A. EMSA of PRC2 binding to H2B-Cy5 labeled nucleosome substrates. Nucleosome core particles lacking an accessible DNA linker largely attenuate PRC2 binding (top gel). Nanomolar binding affinity to nucleosomes was restored by the addition of a 60 bp linker DNA (bottom gel). 2.5 nM of each nucleosome was used in binding reactions. Data points represent 2-fold titrations ranging from 0–4 μM. **B**. PRC2 binds protein-free 147 bp dsDNA > mononucleosome with linker DNA >> nucleosome core particle. Error bars give SD, n = 3. **C**. PRC2 protects linker regions of dodecanucleosome arrays from MNase-mediated digestion.

Because the lack of PRC2 binding to NCPs was so unanticipated, we performed an additional test. We generated NCPs by treating 207 bp mononucleosomes with MNase-ExoIII, which allowed the same nucleosome preparation to be tested with and

without linker DNA. As shown in Figure 4.7e, MNase-ExoIII treatment successfully trimmed mononucleosomes to core particles. We found that binding to these trimmed NCPs was again attenuated to micromolar affinity (Figure 4.7f).

Given that PRC2 will rarely encounter isolated mononucleosomes *in vivo*, we proceeded to test whether PRC2 binds linker DNA in a chromatin array. Therefore, we subjected dodecanucleosome arrays to limited MNase digestion with or without PRC2. Typically, limited MNase digestion of nucleosome arrays produces double-stranded breaks within nucleosome linker regions, which mark the position of nucleosomes. As shown in Figure 4.8c, in the absence of PRC2, a ladder corresponding to the approximate positions of nucleosomes was revealed. However, as PRC2 was titrated into the reaction, PRC2 binding to DNA linker regions protected the dodecanucleosome arrays against MNase digestion. This is seen by the accumulation of higher molecular weight DNA fragments (lanes 4 and 8 of Figure 4.8c and Figure 4.9). Such a protection pattern has been observed for H1, a known linker histone (Widlak and Garrard, 2006). Therefore, these data support the hypothesis that PRC2 binds linker regions of nucleosome arrays.

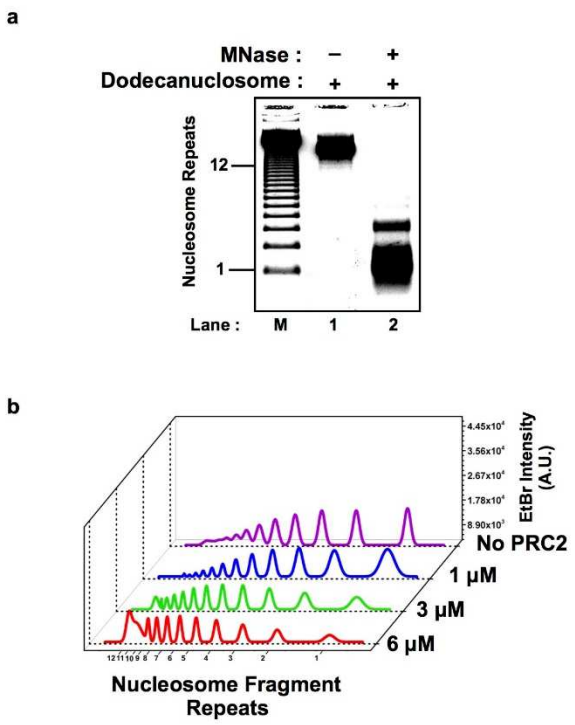


Figure 4.9. PRC2 binds accessible DNA regions of nucleosome arrays. A. MNase limit digestion shows that prolonged MNase treatment of dodecanucleosomes results in a mostly mononucleosome species. **B.** Quantifications of Fig. 4.8.C (lanes 1–4).

PRC2 binds a dinucleosome mimetic of promoter chromatin with optimal binding to long-linker dinucleosomes, and nucleosome spacing affects H3K27 methylation

PRC2 has been reported to occupy both promoter regions and inactive heterochromatin regions. However, these chromatin states may differ in nucleosome spacing and accessible linker DNA. Thus, we were interested in whether the spacing between two adjacent nucleosomes is an important factor for PRC2 binding and activity. We reconstituted a fluorescent dinucleosome mimetic from a human CpG island promoter of a PRC2-regulated gene[17] (Figure 4.10a) by inserting a 100 bp linker between two 601-positioned nucleosomes. For comparison, we also reconstituted dinucleosomes with a 50 or 10 bp linker, with the latter mimicking a gene-body dinucleosome (Schones et al., 2008), (Radman-Livaja and Rando, 2010). Details of dinucleosome design are shown in

Figure 4.10b,c. As shown by EMSA, PRC2 preferentially bound to long-linker dinucleosomes (Figure 4.11a,b). The K_d^{app} to dinucleosomes with a 100 bp linker was 13 ± 2 nM, and for a 50 bp linker, the K_d^{app} was slightly weaker at 24 ± 3 nM. Interestingly, binding to dinucleosomes with a 10 bp linker was substantially diminished, yet PRC2 still bound with a nanomolar affinity of 276 ± 16 nM. Therefore, while PRC2 prefers binding to longer stretches of protein-free DNA, it is clear that PRC2 can still be targeted to more closely packed nucleosomes. Nevertheless, our conclusion still holds that a stretch of accessible DNA allows PRC2 to bind nucleosome substrates optimally.

From this binding analysis, we hypothesized that dinucleosomes with a 100 bp linker length would have the greatest stimulation of histone methylation due to optimal binding to PRC2. As shown in Figure 4.11c and d, PRC2 activity does not increase linearly with increasing dinucleosome spacing. Rather, there appears to be a “Goldilocks effect” with the 50 bp linker-dinucleosome stimulating PRC2 to a greater extent than either the 10 or 100 bp linker-dinucleosomes. Perhaps very long linker DNAs can position PRC2 too far from its H3 substrate for catalysis to occur.

PRC2 has a binding preference for CG-rich DNA and recognizes methylated CpG dinucleotides in DNA through its AEBP2 subunit

PRC2 has been broadly observed to localize with CpG islands (CGIs) *in vivo*, and it has been shown that there is a strong correlation between PRC2 occupancy and GC-rich sequences genome-wide(Wang et al., 2017),(Mendenhall et al., 2010b),(Ku et al., 2008). It seemed likely to us that such targeting occurs via PRC2’s intrinsic DNA-binding activity. As shown in Figure 4.12a, we tested PRC2 binding to $(CG)_{30}$ and $(TA)_{30}$ dsDNA

substrates via fluorescence polarization. PRC2 bound the CG-rich substrate with a substantially lower K_d^{app}, in agreement with the previous *in vivo* observations.

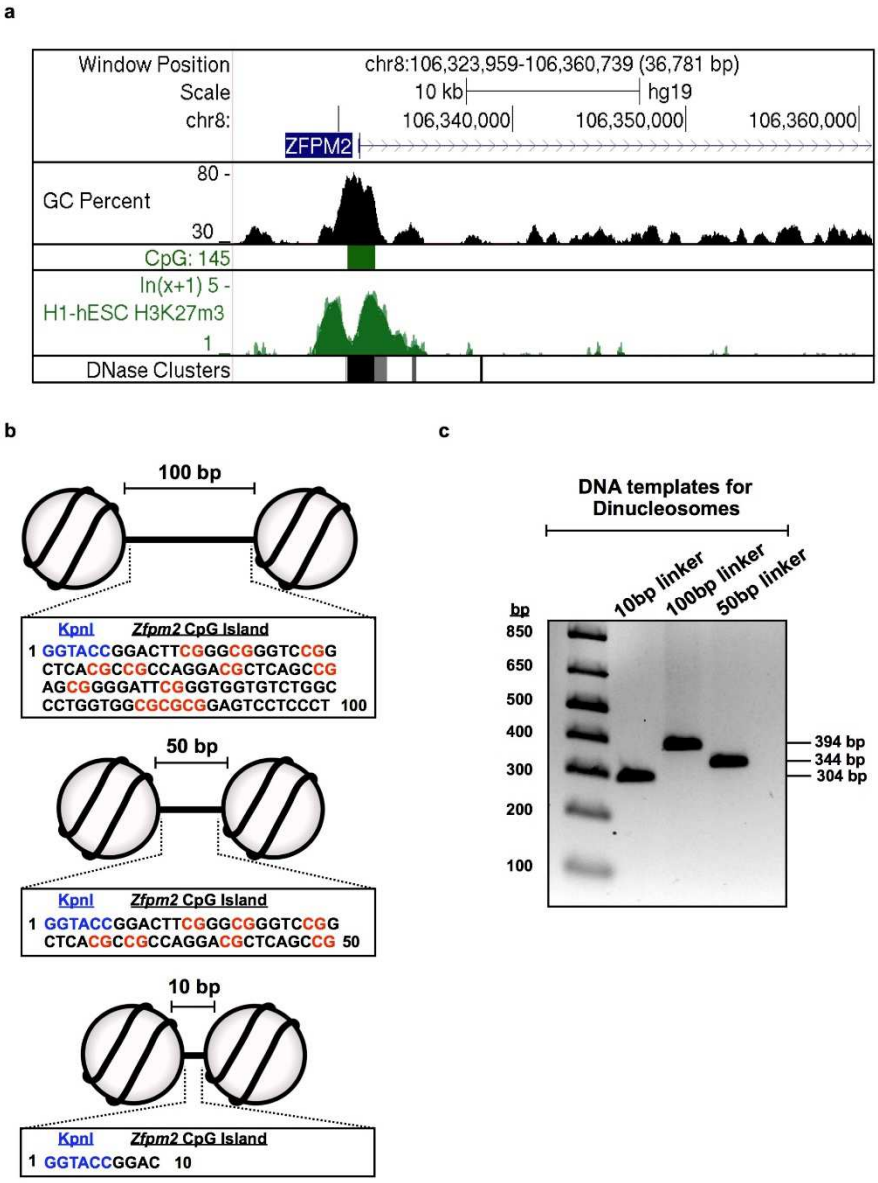


Figure 4.10. Design of a promoter dinucleosome mimic. A. Genome browser snapshot of the human *ZFPM2* gene showing the regional GC percent, the log2 ratio enrichment for H3K27me3 in H1-hESC cell line, and regional DNase hypersensitive clusters. **B.** Design of dinucleosome mimics from the *ZFPM2* promoter. The linker sequence adjoining the two nucleosomes is from the promoter CpG island of *ZFPM2*, which shows H3K27me3 enrichment in ES cells. **C.** DNA agarose gel showing the purity of dinucleosome positioning DNA with varying DNA linker (10, 50, and 100 bp) used for the reconstitution of dinucleosomes.

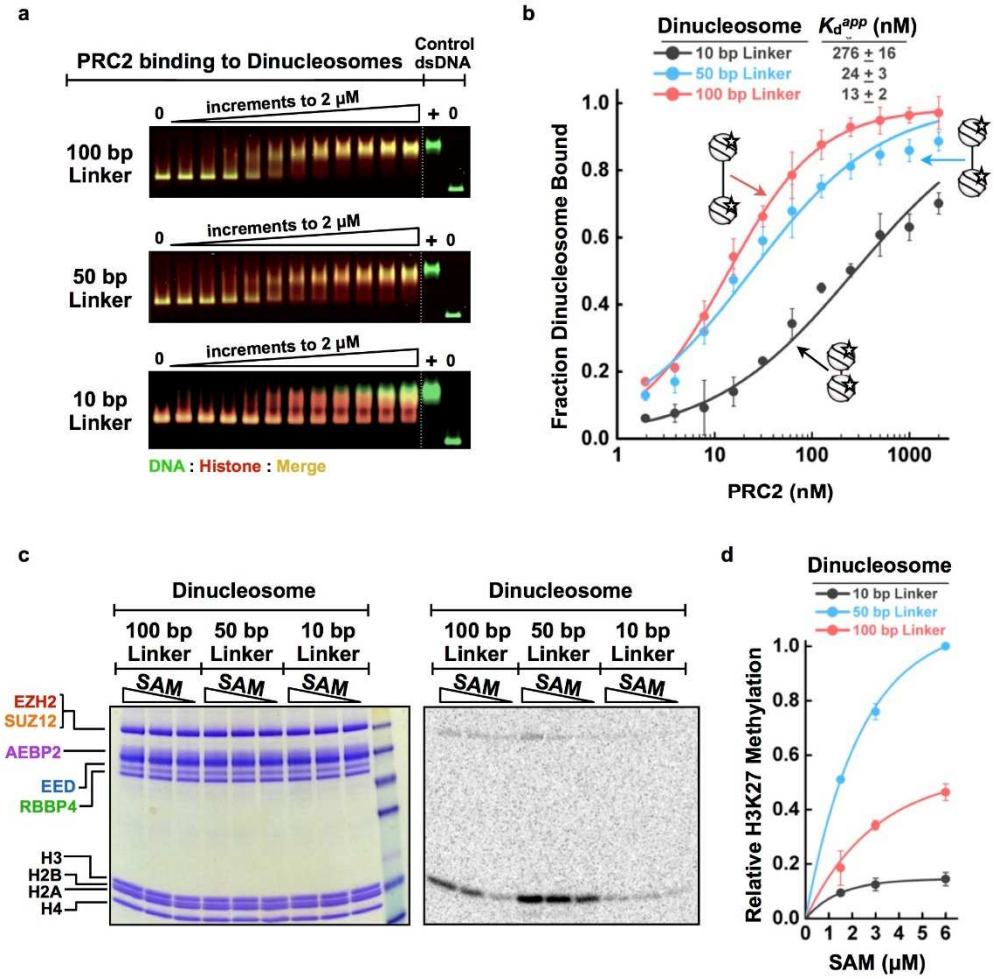


Figure 4.11. PRC2 binds and methylates long-linker dinucleosomes preferentially. A. EMSA gels for PRC2 binding to fluorescent dinucleosomes with 10, 50, or 100 bp linker DNA joining the two nucleosomes. Histone proteins are site-specifically labeled with dye Cy5 on H2B (red channel); DNA was visualized by SYBR Green staining (green channel); assembled nucleosomes show an expected overlap of DNA and histones (yellow, merge). 2.5 nM of each nucleosome was used in binding reactions. Data points represent 2-fold titrations ranging from 0–2 μM PRC2. **B**. EMSA data fit to binding curves show that PRC2 preferentially binds to long-linker dinucleosomes. Error bars give SD, n = 3. **C**. HMTase assays showing PRC2 enzymatic activity to dinucleosomes with 10, 50, or 100 bp linker DNA. Data points represent 6, 3, and 1.5 μM SAM. **D**. Quantification of PRC2 activity with dinucleosome substrates.

Given that DNA methylation is widespread throughout the mammalian genome and found on up to 80% of CpG dinucleotides (Bird, 2002) and that methylation of CGIs near gene promoters is often associated with gene repression (Meissner et al., 2008), we hypothesized that PRC2 might preferentially bind methylated CpG sequences. We first prepared unmethylated and methylated $(CpG)_{24}$ DNA substrates for binding analysis. As

shown in Figure 4.12b and c, we found that the methylated DNA, $(mCpG)_{24}$, bound with K_d^{app} = 11.3 ± 6.2 nM, which is about 50-fold stronger than the unmethylated $(CpG)_{24}$ substrate. A similar strong preference for PRC2 binding to methylated DNA was found for the *TERT* promoter, which contained 8 methylated CpG dinucleotides found *in vivo* as determined by bisulfite sequencing analysis (J. Stern, R. Paucek and T. Cech, manuscript in preparation).

To determine which subunit could be responsible for recognizing methylated CpG, we focused on AEBP2, which contains three well-conserved C2H2 zinc finger motifs. These motifs have been suggested to specifically recognize methylated CpG dinucleotides in other proteins (Buck-Koehntop et al., 2012). We purified PRC2(ΔAEBP2) 4-mer complex and repeated the binding experiments with the methylated and unmethylated DNA substrates. As shown in Figure 4.12d, PRC2(ΔAEBP2) did not have any binding preference for methylated DNA. To further corroborate this result, we introduced point mutations into each of AEBP2's three conserved C2H2 zinc finger motifs (schematic shown in Figure 4.13a), and found that this mutant PRC2-AEBP2 5-mer complex lost its specificity for methylated DNA (Figure 4.13b). Together, these data suggest that AEBP2 is a regulator that helps the PRC2 complex read DNA epigenetic marks and that DNA methylation and histone lysine methylation might be mutually reinforcing.

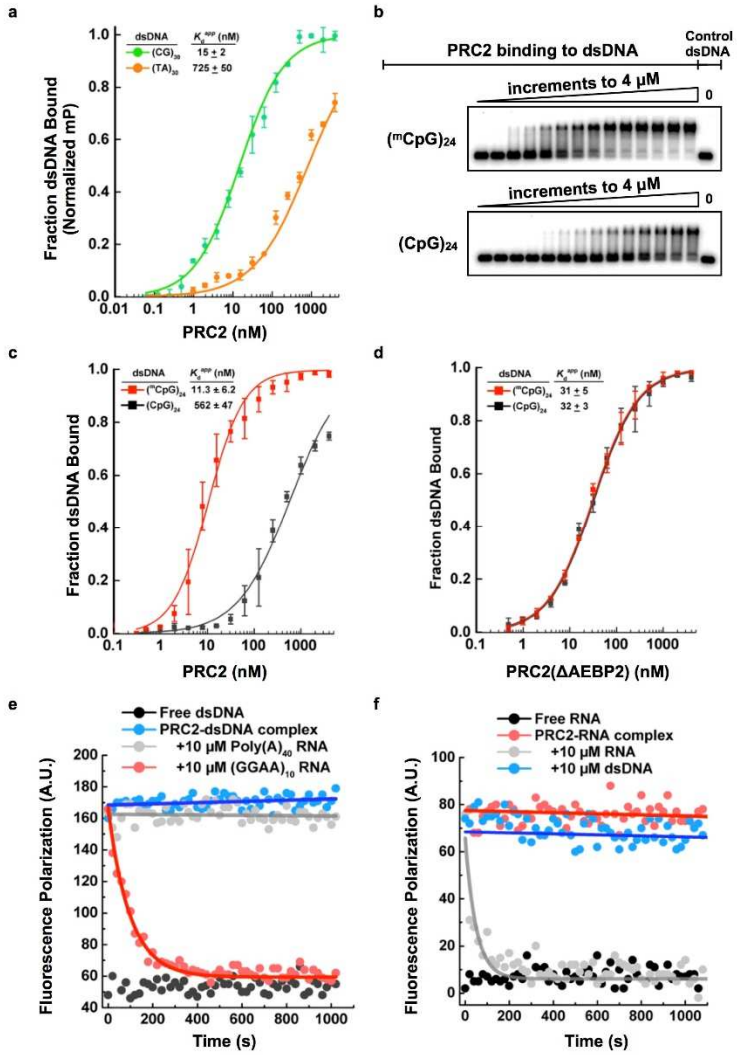


Figure 4.12. PRC2 preferentially binds CG-rich and CpG-methylated DNA. A. PRC2 binding to $(CG)_{30}$ dsDNA is substantially tighter than to dsDNA $(TA)_{30}$. Fluorescence polarization, mP = milli-polarization, 25 mM KCl. Error bars give SD, n = 3. **B**. EMSA gels show that PRC2 recognizes methylated CpG dinucleotides in dsDNA. Top: methylated $(mCpG)_{24}$ dsDNA substrate; bottom: unmethylated $(CpG)_{24}$ dsDNA substrate. Data points represent 2-fold titrations ranging from 0–4 μM PRC2. **C.** EMSA data in (b) fit with equilibrium binding curves shows that PRC2 has a substantial preference in binding to methylated DNA. 100 mM KCl in binding buffer. Error bars give SD, n = 3. **D**. Binding curves show that PRC2(ΔAEBP2) 4mer complex does not discriminate between methylated and unmethylated DNA. Thus the AEBP2 subunit is likely responsible for the recognition of methylated CpG dinucleotides. Error bars give SD, n = 3. **E**. Fluorescence polarization assays show pre-formed PRC2-dsDNA complex dissociation in the presence of RNA competitors: $(GGAA)_{10}$. $\tau(GGAA)_{10}$ = 94.6 ± 4.4 s. Error bars give SD, n=3. **F**. Fluorescence polarization assays showing the lack of dissociation of pre-formed PRC2-RNA complex using dsDNA competitor.

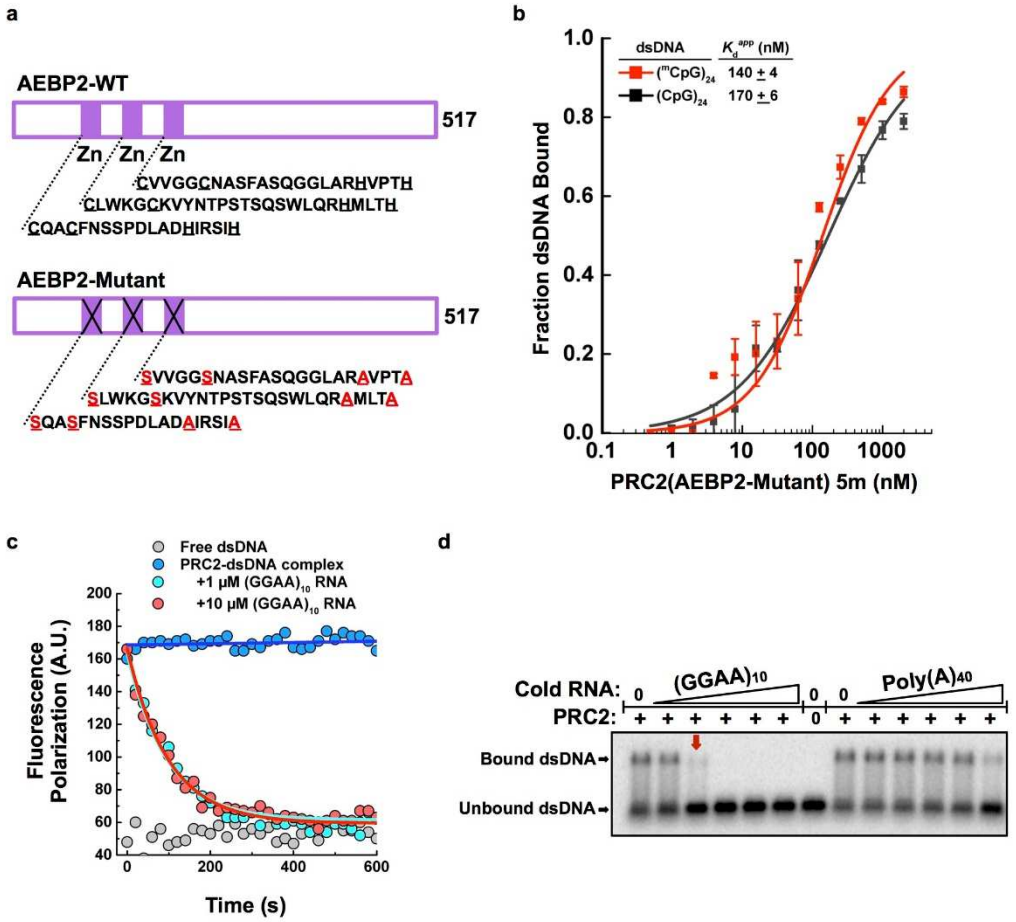


Figure 4.13. RNA and DNA bind competitively to PRC2. A. Schematic representation of the AEBP2 subunit. Mutations introduced to disrupt the three Zinc finger domains in AEBP2 are indicated in red text. **B.** Equilibrium binding curves appear to suggest that PRC2(AEBP2-Mutant) 5-mer complex lacks recognition of methylated CpG dinucleotides. Error bars give SD, n = 3. **C.** Fluorescence polarization assay to measure the dissociation rate of pre-formed PRC2-Alexa488-dsDNA complex in the presence of $(GGAA)_{10}$ RNA competitor. 5 nM of Alexa488-dsDNA probe was incubated with 200 nM of PRC2 for 30 min before adding 1 or 10 μM RNA competitor (200-fold or 2000-fold excess to fluorescent probe). **D.** Gel-based competition experiment showing that $(GGAA)_{10}$ RNA competes off PRC2 from pre-formed PRC2-dsDNA complexes, whereas negative control $Poly(A)_{40}$ RNA does not compete. RNAs were titrated from 0–10 μM in 3-fold increments.

RNA and DNA binding to PRC2 are mutually exclusive

Understanding the importance of linker DNA binding for PRC2-nucleosome interaction led us to hypothesize that RNA was not just inhibiting PRC2 binding to nucleosomes, but more precisely, was antagonizing PRC2 binding to protein-free linker DNA. Thus, we speculated that RNA and DNA binding are competitive. To test this hypothesis, we performed fluorescence polarization competition experiments where

RNAs were titrated into reactions of pre-formed PRC2-A488-dsDNA complexes. For competitor RNA substrates, we used $(GGAA)_{10}$ (which forms G-quadruplexes)(Wang et al., 2017), and a negative control $Poly(A)_{40}$. In the absence of PRC2, free dsDNA gave a fluorescence polarization signal of about 55 A.U. as shown in Figure 4.12e. When PRC2-dsDNA complexes were formed, the fluorescence polarization signal rose to ~170 A.U. PRC2 was effectively competed from DNA using excess $(GGAA)_{10}$ RNA with a half-life of 91 ± 5 s (Figure 4.12e). Decreasing the RNA competitor concentration 10-fold did not change the half-life substantially (Figure 4.13c), indicating that the competitor RNA traps PRC2 upon dissociation from DNA rather than actively invading the PRC2-DNA complex. Notably, the $Poly(A)_{40}$ RNA was inert and did not sequester PRC2. We also performed a gel-based competition where RNAs were titrated into reactions of pre-formed PRC2-dsDNA complexes (Figure 4.13d) and we observed similar results. Next, we reversed the competition experiment by titrating competitor DNA into reactions of pre-formed PRC2-A488-RNA complexes. For competitor DNA, we used $(CG)_{30}$. In the absence of PRC2, free RNA gave a fluorescence polarization signal of ~9 A.U. as shown in Figure 4.12f. When PRC2-RNA complexes were formed, the signal rose to ~80 A.U. As shown in Figure 4.12f, we found that 10 μM DNA competitor failed to compete PRC2 from RNA. Note, in the absence of tRNA competitor, the K_d^{app} values of RNA were even much tighter, in the picomolar range. These data collectively support the key conclusion that RNA and DNA have mutually exclusive binding to PRC2, with RNA binding having much higher affinity.

JARID2 and EZH1 alter PRC2 HMTase activity but not nucleosome binding or its inhibition by RNA

JARID2 and EZH1 are PRC2 components that have both been suggested to “load” PRC2 onto nucleosomes (Son et al., 2013). Therefore, we hypothesized that PRC2 affinity for nucleosomes would be enhanced upon the inclusion of JARID2 or upon substitution of EZH2 for EZH1. PRC2-JARID2 6-mer and PRC2-EZH1 5-mer complexes were expressed and purified (see Methods). In both cases, a multi-step chromatography approach was used to ensure the purity of the complex, and the final size-exclusion column chromatogram showed a mono-dispersed peak (PRC2-JARID2 6-mer chromatogram shown in Figure 4.14a). A typical SDS-PAGE gel showing the pure PRC2-JARID2 6-mer complex is presented in Figure 4.14b.

We first evaluated the enzymatic properties of each protein complex using nucleosome substrates. In our HMTase assay (Figure 4.14c), the PRC2-JARID2 6-mer complex auto-methylated JARID2, as has been reported (Sanulli et al., 2015). The JARID2 automethylation signal provides some additional evidence that our purification scheme preserves the integrity of JARID2. Furthermore, H3K27 methylation of nucleosomes by the reconstituted PRC2-JARID2 6-mer complex was markedly improved (around 3-fold) when compared to PRC2 5-mer complex (Figure 4.14c,d), consistent with published studies where JARID2 protein was titrated into PRC2-H3 reactions(Kaneko et al., 2014b),(Li et al., 2010). In the case of the PRC2-EZH1 5-mer complex, the enzymatic activity was substantially decreased (Figure 4.15a,b), in agreement with an *in vivo* study suggesting that EZH1 maintains basal H3K27 mono-methylation activity(Shen et al., 2008).

We thought that these differences in activity might foreshadow differences in nucleosome-binding activity. However, neither the PRC2-JARID2 6-mer complex nor the

PRC2-EZH1 5-mer complex had differences in nucleosome-binding activity when compared to the typical PRC2 5-mer complex (data not shown**).** In addition, both complexes showed trends similar to PRC2 5-mer in binding to RNA (Figure 4.15c) and to DNA (Figure 4.15d), and in RNA-mediated disruption of PRC2-nucleosome binding (Figure 4.14e).

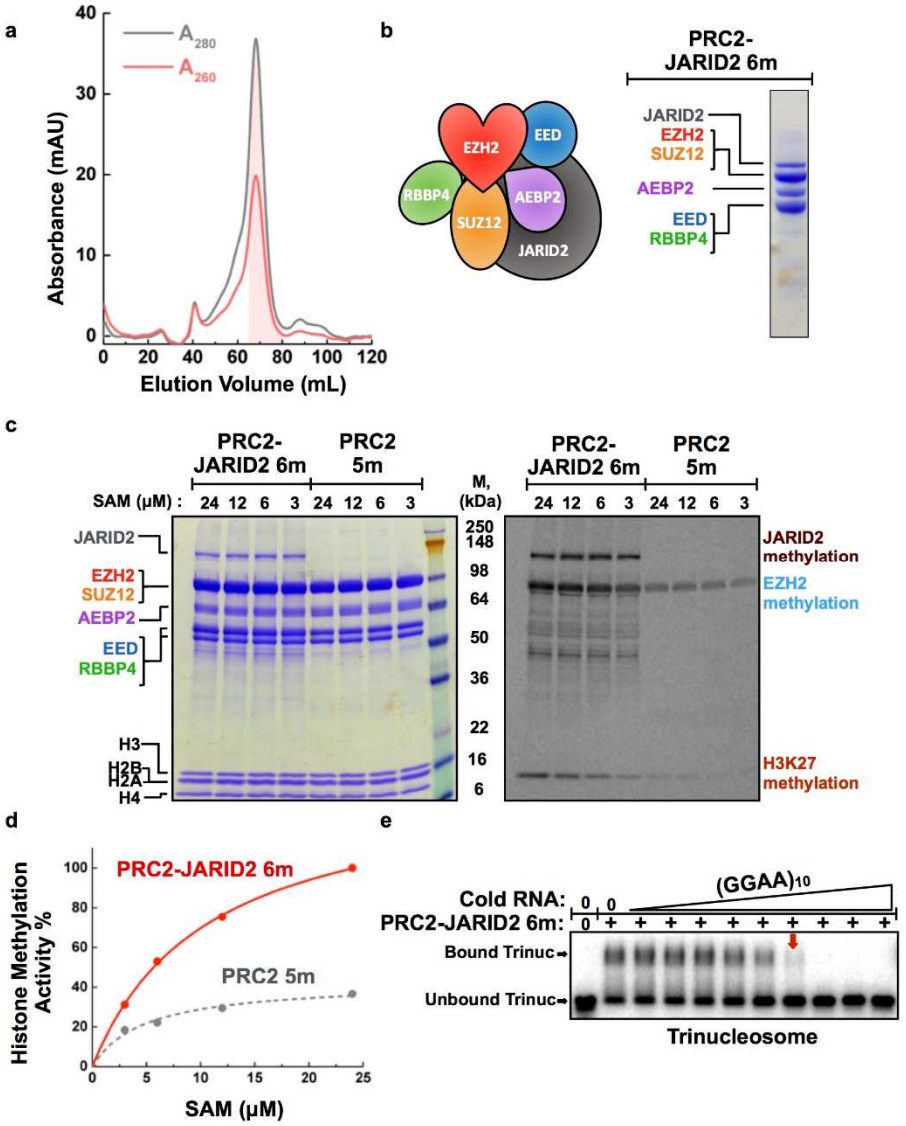


Figure 4.14. JARID2 enhances PRC2 histone methyltransferase activity but does not prevent eviction by RNA. A. Fractionation of PRC2-JARID2 6-mer over Sephacryl S-400 sizing column shows that the complex is monodisperse. A260/A280 ratio <0.7 indicates the lack of nucleic acid contamination. **B**. Typical SDS-PAGE gel showing the purity of our recombinant PRC2-JARID2 6-mer complex. **C**. A comparison of histone methyltransferase activity between PRC2-JARID2 6-mer and PRC2 5-mer using mononucleosome substrates shows that JARID2

substantially augments PRC2 activity. PRC2 complexes and histone proteins were visualized by Coomassie-staining (left-hand gel); methylation signal detected by ^{14}C-autoradigraphy (right-hand gel). **D**. Quantification of H3 methyltransferase activity. **E**. $(GGAA)_{10}$ RNA disrupts pre-formed complexes of PRC2 6m-trinucleosome. Complexes were fully disrupted (red box arrow) at a PRC2:RNA molar ratio of ~2:1. Data points represent 3-fold titrations ranging from 0–10 μM RNA.

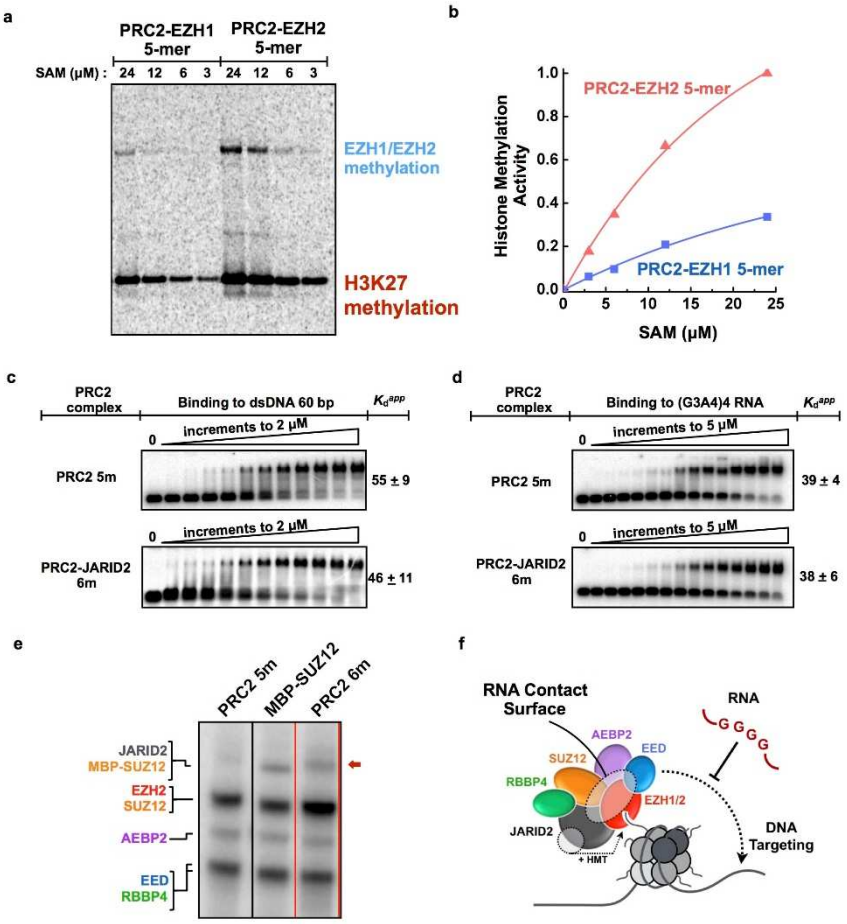


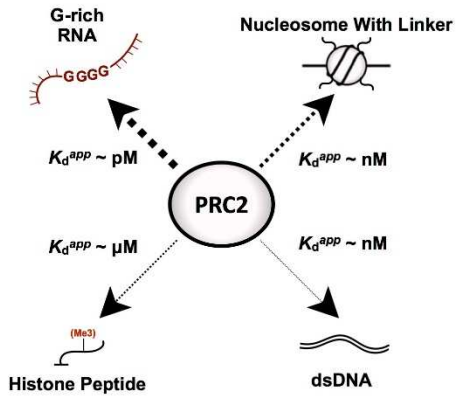
Figure 4.15. Substitution of EZH1 for EZH2 or inclusion of the JARID2 subunit in PRC2 does not alter trends in DNA and RNA-binding. A. A comparison of histone methyltransferase activity between PRC2-EZH1 5-mer and PRC2 5-mer using mononucleosome substrates shows that EZH1 has much reduced activity compared to EZH2. Notably, EZH1 appears to be auto-methylated similarly to EZH2. Methylation signal detected by ^{14}C-autoradigraphy (right-hand gel). **B.** Quantification of PRC2-EZH1 5-mer methyltransferase activity. **C.** EMSA gels comparing PRC2 5-mer and PRC2-JARID2 6-mer binding to 60 bp dsDNA. 5 nM of Alexa488-dsNA probe was used. Data points represent 2-fold increments ranging from 0–2 μM. Error bars give SD, n = 3. **D.** EMSA gels comparing PRC2 5-mer and PRC2-JARID2 6-mer binding to G-quadruplex forming (G3A4)4 RNA. For these RNA binding experiments, binding buffer contained fragmented yeast tRNA competitor, as previously described(Wang et al., 2017). Data points represent 3-fold increments ranging from 0–5 μM. Error bars give SD, n = 3. **E.** Cross-linking analysis of 4-thio-U-substituted 24 nt RNA to PRC2-JARID 6-mer shows that JARID2 is cross-linked to RNA (red arrow). **F.** Model showing the RNA contact region in the PRC2-JARID2 6-mer complex as it spans across five subunits(Wang et al., 2017) , and the role of JARID2 in allosterically regulating PRC2 activity.

a

PRC2 Substrate	Identity	K_d^{app} (nM)
RNA	$(GGAA)_{10}$ 40-mer	≤ 0.2
	$(GA)_{20}$ 40-mer	≤ 1.0
dsDNA	$(CpG)_{24}$	562 ± 47
	$(^{m}CpG)_{24}$	11.3 ± 6.2
Nucleosome	Mononucleosome with Linker DNA	280 ± 19
	Nucleosome Core particle	41,200 ± 15,600
Histone Peptide*	H3K4me3	>400,000[14]
	H3K27me3	108,000[14]

* Dissociation constants determined by ITC

b



c

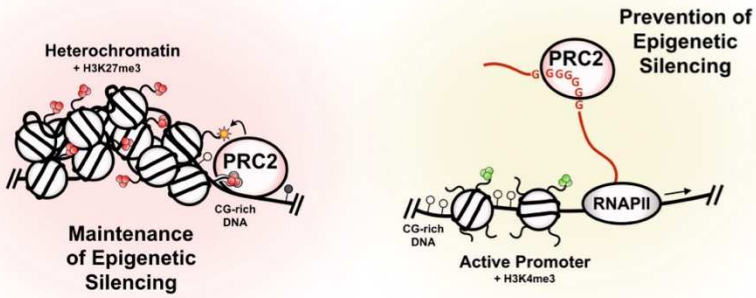


Figure 4.16. A model for PRC2-chromatin-RNA interactions and regulation of epigenetic silencing. A. A summary of relative PRC2 binding to RNA, DNA, nucleosomes, and histone peptides. All K_d^{app} values were determined by EMSA under the same conditions, in the absence of tRNA competitor, with the exception of the literature K_d values for histone peptide binding[14] which were determined by ITC. **B**. A comparison of binding affinity of PRC2 to various substrates reveals that PRC2 binds RNA >> nucleosomes with linker DNA, DNA >> histone tails. **C**. Model includes (**1**) maintenance of epigenetic silencing where high-affinity DNA-binding modules of PRC2 allow the complex to preferentially bind nucleosome-free regions and methylate an adjoining nucleosome, and (**2**) prevention of epigenetic silencing, where RNA binding suppresses H3K27 methylation by interfering with the DNA binding required by PRC2 to bind chromatin. Heterochromatin is characterized by H3K27me3 marks of repressed chromatin and dense nucleosome packing. Active promoters feature H3K4me3 marks. CpG islands featuring unmethylated (open circles) and methylated (filled circles) CpG dinucleotides in gene promoter regions may serve as a conduit for PRC2 targeting. Even though the model may be an oversimplification, it provides an interpretation of how epigenetic silencing could be switched on-and-off under certain chromatin states, as deduced from our quantitative study of PRC2 substrate binding.

DISCUSSION

Numerous studies have established the importance of PRC2 for epigenetic gene silencing in mammals and mapped the genomic sites of H3K27me3 deposition in multiple cell types. Yet fundamental mechanistic questions remained. How does PRC2 bind chromatin? How is PRC2 recruited to particular sites in the genome? Does RNA binding to PRC2 regulate these events, and how? Here we have answered these questions using quantitative biochemistry with purified PRC2 and reconstituted nucleosomes. There will no doubt be numerous exceptions *in vivo* to these "rules" established *in vitro*, and these exceptions will stimulate researchers to find the additional factors (e.g., non-histone chromosomal proteins or post-translational modifications) that allow the biological situation to diverge from the biochemical paradigm.

How does PRC2 bind chromatin?

This question had seemingly already been answered: PRC2 binds the N-terminal tails of histone H3 in nucleosomes. Yet, do such interactions provide sufficient binding energy, given measured affinities of histone peptide-PRC2 in the 40–400 μM range? Here, we find that the histone-free linker DNA has a central and direct role in the recruitment of PRC2 to an adjoining nucleosome core particle *in vitro*, with an observed binding affinity in the nanomolar range. In the absence of an available DNA linker, such as with a nucleosome core particle, PRC2 targeting to nucleosomes is substantially weakened *in vitro*. This may explain why ChIP-seq studies find PRC2 peaking near transcriptional start sites, which are typically nucleosome-free regions (Davidovich et al., 2013b)·(Kaneko et al., 2013b); however, the resolution of current ChIP-seq datasets does not allow for a definitive comparison of PRC2 binding sites and nucleosome positioning.

Nevertheless, one pioneering study has provided evidence of mutual exclusion between SUZ12 binding and nucleosome density on targeted CpG islands (Riising et al., 2014).

Binding of N-terminal tails of H3 peptides with PRC2 has been directly observed by X-ray crystallography (Jiao and Liu, 2015b). Furthermore, a body of literature finds genuine differences in binding of modified H3 peptides to PRC2, but all reported values are in the micromolar K_d range. It seemed possible that these affinities would be much stronger if the histone tails were presented in the context of nucleosomes, so it is perhaps not surprising that numerous PRC2 recruitment models have drawn upon the idea of pre-existing histone marks guiding and reinforcing PRC2 localization at target genes. In stark contrast to this expectation, we found that different histone marks make minor contributions to PRC2 binding of nucleosomes *in vitro*. Instead, our findings indicate that the mechanism of PRC2 targeting to chromatin is dependent on linker DNA. Thus, our new perspective is that nucleosome-free regions of chromatin and nucleosome linkers recruit PRC2 histone methyltransferase, after which it will bind histone tails in its vicinity. PRC2 will then be stimulated by existing H3K27me3 or repressed by binding H3K4me3, as in the canonical model. Inherent to this new model is that PRC2 enrichment at a target site can be explained by chromatin binding in hierachical fashion with nanomolar affinity DNA-binding components cooperating with micromolar affinity histone tail-binding modules.

What makes a genomic region a target for PRC2-mediated epigenetic silencing?

In search of a recruiting DNA element, our *in vitro* analysis showed that PRC2 prefers binding to CG-rich sequences. This finding is consistent with multiple studies showing that PRC2 target genes contain CG-rich sequences *in vivo* (Wang et al.,

2017), (Mendenhall et al., 2010b), (Ku et al., 2008), (Wu et al., 2011). Past studies have also proposed that CpG islands could be novel recruiters for PRC2 in mammals. Interestingly, CpG island and CG content have been shown to be key factors that promote nucleosome instability and depletion (Fenouil et al., 2012), (Deaton and Bird, 2011); this may further contribute to PRC2 binding CG-rich regions *in vivo*.

PRC2 has been suggested to be directly associated with DNA methylation and that loss of DNA methylation leads to loss of H3K27me3 and PRC2, which is in agreement with our *in vitro* results showing that methylated DNA binds optimally (Jin et al., 2009), (Viré et al., 2006). Even though some studies have argued that PRC2 only occupies unmethylated CpG islands, the discrepancy between these studies and our findings could come from a different composition of PRC2 subcomplexes. We have found that the AEBP2 subunit of PRC2 is responsible for reading methylated CpG dinucleotides. In fact, one recent study reports that AEBP2 plays a role in defining the mutually exclusive composition of PRC2 subcomplexes (Grijzenhout et al., 2016). AEBP2 may not be the only epigenetic regulator that associates with the PRC2 core complex. Indeed, other accessory subunits in PRC2 subcomplexes may promote exclusion from methylated CpG islands. Future studies are needed to test such hypotheses.

Drosophila has PREs, which recruit proteins such as PHO and other DNA-binding factors (Sengupta et al., 2004), (Klymenko et al., 2006). Such factors then recruit *Drosophila* PRC2 (dPRC2) in hierarchical fashion. Even though the sequence specificity of dPRC2 for PREs is not fully understood, this DNA-based recruitment mechanism is widely accepted. In mammals, no homologs of PREs have been identified, and neither have orthologs of the *Drosophila* PRC2-recruiting proteins (Schuettengruber et al., 2007).

However, our study has demonstrated that human PRC2 has in fact robust DNA-binding activity (with nanomolar K_d^{app}), and that select *Drosophila* PRE DNA substrates bind very well to PRC2 (data not shown). In addition, nucleosome-kinetic studies have suggested PREs are areas of histone replacement and reduced nucleosome occupancy(Mito et al., 2007). And our findings indicate that the intrinsic DNA-binding activity of human PRC2 is important for binding to chromatin and promoting activity. Therefore, this suggests that PRC2 recruitment in mammals and *Drosophila* may be more similar than previously thought.

JARID2

JARID2 has been identified and recognized as an important regulator of PRC2. Initial studies concluded that JARID2 negatively regulates PRC2 activity, but subsequent studies have reported that JARID2 is an activator (Shen et al., 2009), (Peng et al., 2009). This contradiction could simply arise from variations in substrates used in activity assays (unfolded histone H3 vs. histone octamer vs. nucleosome), or could be due to difficulties in protein purification (Kaneko et al., 2014b). In order to be more biologically relevant, we reconstituted PRC2-JARID2 6-mer complex by co-expressing JARID2 together with the other five subunits, and we then used well-assembled nucleosomes as our substrate in activity assays. Our results have shown that the incorporation of JARID2 substantially stimulates EZH2 catalytic activity, as indicated by an increase in H3K27 methylation and automethylation. This finding supports the conclusion that JARID2 may allosterically regulate PRC2, in agreement with previous findings from mixing experiments using PRC2 and JARID2 (Kaneko et al., 2014b),(Li et al., 2010). However, for our PRC2 complexes incorporating JARID2, we did not observe increased nucleosome-binding affinity as a

previous study has suggested (Son et al., 2013). This discrepancy could be due to the previous use of truncated JARID2 proteins to test nucleosome-binding affinity, rather than full-length JARID2 in complex with PRC2 as used here. Furthermore, JARID2 has been proposed to have an N-terminal noncoding RNA-binding region (RBR) (Kaneko et al., 2014c). We found that the inclusion of JARID2 did not alter PRC2 RNA-binding activity (Figure 4.15d), however, our RNA crosslinking analysis showed that JARID2 may contribute a surface that contacts RNA when it is in complex with PRC2 (Figure 4.15e and 4.15f).

Does RNA binding to PRC2 regulate these events, and how?

Why would the PRC2 gene-silencing complex be recruited to actively transcribed genes? Why does PRC2 not suppress transcriptional activity at those genes where it is bound? These perplexing observations can now be explained by the intrinsic properties of PRC2-RNA-nucleosome interactions. In highly expressed genes, PRC2 may be bound by RNA emerging from active genes but it is not able to promote silencing, because the RNA shuttles PRC2 away from chromatin. RNA inhibition of PRC2 binding was previously observed in non-equilibrium pull-down assays by Jenner (Beltran et al., 2016b). The current work now provides an unexpectedly simple mechanism for this eviction: (1) RNA and DNA bind PRC2 competitively, with the RNA having higher affinity. (2) DNA binding is necessary for PRC2 to bind nucleosomes. (3) Thus, PRC2 cannot bind nucleosomes in the presence of a saturating amount of RNA. Furthermore, these results provide mechanistic understanding for how RNA can hold PRC2 “poised but in check”(Kaneko et al., 2014b). The RNA is not an active-site inhibitor of PRC2 methyltransferase activity, as

evidenced by the unperturbed automethylation activity, but instead it simply prevents PRC2 from binding its histone substrate.

In conclusion

PRC2 recruitment to target sites is perhaps one of the most enigmatic aspects of its function. As a gene silencer with an intrinsic ability to propagate its own heritable enzymatic product, PRC2 must possess a repertoire of biochemical properties suitable to its biological function. Together, our findings reveal new insights into how PRC2 associates with nucleosomes and on the interplay of PRC2 between chromatin and RNA. As shown in the table of relative K_d values in Figure 4.16a and by the schematic in Figure 4.16b, PRC2 binds RNA >> nucleosomes with linker DNA, or DNA >> histone tails. Based on these relative affinities, as well as the passive role of RNA in dissociating PRC2-nucleosomes complexes, we propose our model in Figure 4.16c: the PRC2 gene-silencing complex undergoes recruitment to target genes by binding to DNA, with contributions in specificity conferred by CG-rich sequences or CpG islands. This DNA binding by PRC2 could promote scanning of nearby chromatin and the recognition of histone marks. Most of the affinity of PRC2 binding comes from its binding to DNA, not histones; its binding to histone tails is of course necessary for H3K27-trimethylation and for regulation, but it makes a minor contribution to affinity. In the case of active transcription units, on the other hand, the nascent pre-mRNA or nearby lncRNA transcripts bind PRC2 and prevent its deposition to chromatin; thus, the active state is maintained. These models will no doubt prove to be oversimplified in particular biological situations, where the presence of other proteins or RNP factors (either PRC2-bound or chromatin-bound) may overcome or compete with the intrinsic properties of the PRC2

complex. Yet understanding the fundamental properties of the PRC2-nucleosome interaction provides a framework for interpretation of specific instances where epigenetic gene silencing is maintained or is switched on or off.

MATERIALS AND METHODS

Protein expression and purification

Human PRC2-JARID2 6-mer complexes were expressed in insect cells similarly to PRC2 5-mer as previously described (Davidovich et al., 2013b), (Wang et al., 2017). In brief, sequences encoding human EZH2, SUZ12, EED, RBBP4, AEBP2, and JARID2 (UniProtDB entry isoform sequences Q15910-2, Q15022, O75530-1, Q09028-1, Q6ZN18-1, and Q92833-1, respectively) were cloned into the pfast-bac1 expression vector (Invitrogen) with PreScission-cleavable N-terminal hexahistidine-MBP tags. JARID2 had an additional 3XFLAG-tag before the PreScission cutting site. Standard Bac-to-Bac baculovirus expression system (Invitrogen) was used to generate baculovirus stocks according to manufacturer's protocol. The titer of each baculovirus stock was measured by gp64 detection (Expression Systems). Equal amounts of baculovirus of each subunit were used to infect sf9 cells (Invitrogen) at a density of 2.0 x 10^6 cells/ml. Following infection, the cells were incubated for 72 h (27°C, 130 rpm) before they were harvested. The harvested cells were snap-frozen with liquid nitrogen for later purification. PRC2 5-mer complex was purified as previously described (Wang et al., 2017) via a three-column purification scheme. For PRC2 6-mer complex, an additional FLAG-tag purification was incorporated, as follows. Cell extract was incubated with the amylose resin and washed thoroughly, followed by elution with 10 mM maltose. The elutate was concentrated to ~15 mg/ml (Amicon Ultra-15 Centrifugal Filter Unit, 30 kDa MWCO,

Millipore cat # UFC903024). Then, the concentrated eluate was incubated with equilibrated Anti-FLAG G1 affinity resin (GenScript L00432-25) in the cold room with slow agitation for 2-3 h. Beads were washed with 20 C.V. of FLAG-buffer (10 mM Tris, pH 7.5 at RT and 150 mM NaCl) and eluted with FLAG-buffer supplemented with 3XFLAG peptide to 0.2 mg/ml ApexBio A6001). Each of three elutions proceeded with slow agitation for 40-50 min in the cold room. The TCEP concentration was adjusted to 1 mM and the eluate was concentrated with Amicon Ultra-0.5 Centrifugal Filter Unit with 30-kDa cutoff (Millipore UFC503096), followed by digestion with PreScission protease at a mass ratio of 1:50 protease:protein. After completion of cleavage, protein complex was chromatographed on a 5 mL Hi-Trap Heparin column (GE, 17-0407-03), followed by fractionation over a HiPrep 16/60 Sephacryl S-400 HR sizing column. The 6-mer peak fractions were identified using SDS-PAGE, pooled, and concentrated as above. Final protein concentration was measured by absorbance at 280 nm, and the ratio of absorbance at 260 nm/280 nm <0.7 provided an indication of no nucleic acid contamination.

Nucleosome reconstitution

Histone octamer was assembled using equal molar amount of each histone in 2 M NaCl. Octamer was purified via a sizing column (GE, HiLoad 16/60 Superdex 200 HR). Puc19 plasmid containing 601-Widom positioning sequence was purified using GigaPrep (Qiagen 12191) and cut with EcoRV (NEB R3195M). Efficiency of cutting was determined using a 1% agarose gel. Then, the DNA was adjusted to ~1 mg/mL and purified via Mono Q column (GE 17-5167-01). Fractions containing the nucleosome template were

identified by agarose gel, pooled, and concentrated by ethanol precipitation. DNA was then dissolved in TE buffer.

Nucleosome reconstitutions were performed by standard protocol using salt dialysis(Dyer et al., 2004). In brief, nucleosome particles and arrays were reconstituted at a small scale from 1 to 6 μM in a volume of 25–50 μl. Widom-601 DNA (1-, 3-, or 12-repeats) were mixed with histone octamers and assembled by dialysis from RB-High buffer (2 M KCl, 10 mM Tris-HCl pH 7.5, 0.1 mM EDTA, 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 h, buffer was exchanged into TCS buffer (20 mM Tris-HCl pH 7.5, 1 mM EDTA, 1 mM DTT) and allowed to dialyze for 1-6 h. Nucleosomes were then transferred to LoBind tubes and stored at 4°C. After reconstitution, a portion of each sample was cut by EcoRI-HF (NEB R3101T) and analyzed by 5% native-PAGE on a 1% agarose gel. Based on ethidium bromide staining, the histone-saturation level of reconstituted nucleosomes could be assessed. In addition, each sample was subjected to AFM to further examine whether each reconstituted nucleosome batch was of good quality.

Atomic-force microscopy

For surface deposition, nucleosome samples were diluted with TCS buffer to a final concentration of 2 nM. 50 μL of diluted samples were pipetted onto APTES-modified mica (Grade V, SPI). Samples were then incubated at room temperature for 5 min. The mica discs were then rinsed with purified 18.2-MΩ deionized water. For samples intended for AFM imaging in air, the washed mica disks were then dried using a gentle N_2 gas flow,

perpendicular to the mica surface. For samples intended for imaging in liquid, the rinsed mica disks were quickly exchanged into imaging buffer (10 mM Tris-HCl pH 7.5, 3 mM $NiCl_2$). All samples were imaged by AFM immediately following preparation. Data acquisition was performed using a MFP-3D AFM or 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. 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. Image processing was carried out using publicly available software (Gwyddion).

Electrophoretic mobility shift assay (EMSA)

Reconstituted nucleosomes were radiolabeled in the cold room for 1 h using T4 PNK (NEB M0201L) by standard protocol, with the modification of reducing the concentration of $MgCl_2$ in the T4 PNK reaction buffer to 0.5 mM. After labeling, excess [γ-32P]-ATP in the reaction was removed by running the samples over a G50 Sephadex column (Roche 11 273 949 001). The column was pre-equilibrated with TCS buffer. The counts of nucleosomes were determined by liquid scintillation counting. Radiolabeled nucleosomes were diluted with 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, 0.05% v/v NP-40, 0.1 mg/ml bovine serum albumin, 5% v/v glycerol). Next, stock PRC2 was diluted with binding buffer and added to radiolabeled nucleosomes. Binding was carried out at 30°C for 30 min, followed by loading samples onto non-denaturing 0.7% agarose gel (Fisher BP160-100) buffered with 1XTBE at 4°C. Gel electrophoresis was carried out for 90 min at 66 V in an

ice box within a 4°C cold room. A Hybond N+membrane (Amersham, Fisher Scientific 45-000-927) and two sheets of Whatman 3 mm chromatography paper were put underneath the gel, which then was vacuum dried for 60 min at 80°C. Dried gels were exposed to phosphorimaging plates, which were scanned using a Typhoon Trio phosphorimager (GE Healthcare) for signal acquisition. Gel analysis was carried out with ImageQuant software (GE Healthcare) and data fitted to a sigmoidal binding curve using custom written code in MATLAB (MathWorks).

Semisynthesis of H3 containing trimethyl-lysine (H2K4Kme3 and H3K27me3)

Semisynthetic histones were generated essentially as previously described (Nguyen et al., 2014) with some minor changes. Briefly, synthetic peptides were synthesized as C-terminal hydrazides by Fmoc-SPPS and converted into the corresponding thioesters by oxidation using $NaNO_2$ in the presence of thiols. Native chemical ligation was then performed with the requisite truncated H3 protein in ligation buffer (200 mM phosphate buffer, 6 M guanidine hydrochloride, pH 7.5) with 3% 2,2,2-trifluoroethanethiol (TFET). The cysteine-containing ligation product was subjected to radical-desulfurization to give the native H3 sequence bearing the Kme3 PTM at the desired position. Following purification by HPLC, the proteins were characterized by HPLC and ESI-MS.

***In vitro* histone methyltransferase assay**

Unless indicated otherwise, each 10 µl reaction contained 600 nM PRC2, 600 nM mononucleosome, and 12 µM *S*-[methyl-^{14}Cadenosylmethionine (PerkinElmer NEC363050UC) in Covfefe buffer (50 mM Tris-HCl pH 8.0 at 30°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). Reactions were incubated for 1 h at 30°C and stopped by adding 4X loading dye. Each reaction was then heated at 95°C for 5 min and loaded onto either 4-12% Bis-Tris gel (ThermoFisher NP0322BOX) or 10-20% Tris-Glycine gel (ThermoFisher XP10202BOX). Gels were first stained by Coomassie and scanned, then vacuum dried for 60 min at 80°C. Signal was acquired with a Typhoon Trio phosphorimager (GE Healthcare). Densitometry and analysis were carried out with ImageQuant software (GE Healthcare).

In testing RNA inhibition of histone methyltranferase activity, the reaction was set up as described above except 2.4 μM mononucleosome was used and RNA was titrated into the reaction from 60 μM (2-fold dilutions).

RNA and nucleosome competition

The reaction was set up as above except 1000-2000 c.p.m radiolabeled nucleosome was mixed first with unlabeled nucleosome, and then the mixture was added into each reaction. After the initial 30 min incubation, unlabeled competitor RNA was titrated into each reaction and an additional 30 min incubation was carried out at 30°C before the reaction was loaded onto 0.7% agarose gel.

RNA and DNA competition assay

The same reaction was set up as described in the RNA/nucleosome competition except DNA was radiolabeled.

Fluorescence polarization for binding reaction and competition assay

Each dsDNA containing 5' Alexa-488 fluorescent dye was synthesized, annealed, and purified by IDT. Each 40 μl reaction contained 5 nM fluorescent DNA and PRC2 in binding buffer (50 mM Tris-HCl pH 7.5, 25 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). Reaction was incubated for 30 min at 30°C, followed by measuring fluorescence polarization in a 384-well plate with Synergy 2 multi-mode plate reader (BioTek). For competitor assays, initial signals were measured (defined as t = 0) before adding competitors. RNA or dsDNA cold competitors were added into the reactions and the measurements were taken every 20 sec for 4 h.

MNase digestion analysis

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/μL and 0.8 Units/μL 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-dodecanucleosome binding reaction was prepared in a final volume of 45 μL and incubated at 30ºC for 30 min. Following incubation, 2.5 μl $CaCl_2$ was added to a final concentration of 5 mM, and 2.5 μl of diluted MNase was added to the mixture. The MNase reaction was allowed to proceed for approximately 10 min before quenching with 12.5 μl of 500 mM EDTA. Then, 2.5 μl of a 10 mg/ml stock solution of proteinase K was added to digest PRC2 proteins and histones by incubating for 30-60 min at 50ºC. 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 h. DNA bands were visualized by ethidium bromide staining. A 123-bp DNA ladder (1 μg per lane) was used as a size marker.

UV Crosslinking of RNA-PRC2 Complexes

RNA was synthesized (GE-Dharmacon) with 4-thio-U incorporated into the oligo (UUUGGGU[4-S-U]UGGGUUGGGUUGGGUU). This 24 nt oligo was end- labeled with gamma-^{32}P-ATP as described above. Trace amounts of hot RNA were refolded (as in (Wang et al., 2017), except in some instances BSA and tRNA were omitted from the refold buffer) or used without refolding and incubated with 200 nM–1 μM PRC2 protein. PRC2 5m or 6m or PRC2 that contained an uncleavable MBP tag on SUZ12 was incubated with RNA for 30 min at 30 °C. Samples were moved to siliconized glass coverslips on ice, placed in a Stratalinker with 365 nm bulbs, and exposed to light for 10–30 min. Sample were diluted with SDS loading dye and loaded onto a Nupage 4-12% Bis-Tris gel (Life Technologies). Gel electrophoresis was carried out for 60 min at 180 volts. Gels were vacuum dried at 80 °C for 30 min on Whatman 3mm chromatography paper. Dried gels were exposed to phosphorimaging plates and signal acquisition was performed using a Typhoon Trio phosphorimager (GE Healthcare).

Reference

Agris, P.F., Vendeix, F.A.P., and Graham, W.D. (2007). tRNA's wobble decoding of the genome: 40 years of modification. J. Mol. Biol. *366*, 1–13.

Aitken, C.E., and Lorsch, J.R. (2012). A mechanistic overview of translation initiation in eukaryotes. Nat. Struct. Mol. Biol. *19*, 568–576.

Akhtar, A., and Becker, P.B. (2000). Activation of transcription through histone H4 acetylation by MOF, an acetyltransferase essential for dosage compensation in Drosophila. Mol. Cell *5*, 367–375.

Alfano, C., Sanfelice, D., Babon, J., Kelly, G., Jacks, A., Curry, S., and Conte, M.R. (2004). Structural analysis of cooperative RNA binding by the La motif and central RRM domain of human La protein. Nat. Struct. Mol. Biol. *11*, 323–329.

Augui, S., Nora, E.P., and Heard, E. (2011). Regulation of X-chromosome inactivation by the X-inactivation centre. Nat. Rev. Genet. *12*, 429–442.

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.

Bannister, A.J., and Kouzarides, T. (1996). The CBP co-activator is a histone acetyltransferase. Nature *384*, 641–643.

Battiste, J.L., Pestova, T.V., Hellen, C.U.T., and Wagner, G. (2000). The eIF1A Solution Structure Reveals a Large RNA-Binding Surface Important for Scanning Function. Mol. Cell *5*, 109–119.

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. (2016a). The interaction of PRC2 with RNA or chromatin is mutually antagonistic. Genome Res. gr.197632.115.

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. (2016b). The interaction of PRC2 with RNA or chromatin is mutually antagonistic. Genome Res. gr.197632.115.

Bernstein, E., Duncan, E.M., Masui, O., Gil, J., Heard, E., and Allis, C.D. (2006). Mouse Polycomb Proteins Bind Differentially to Methylated Histone H3 and RNA and Are Enriched in Facultative Heterochromatin. Mol. Cell. Biol. *26*, 2560–2569.

Bertolotti, A., Lutz, Y., Heard, D.J., Chambon, P., and Tora, L. (1996). hTAF(II)68, a novel RNA/ssDNA-binding protein with homology to the pro-oncoproteins TLS/FUS and EWS is associated with both TFIID and RNA polymerase II. EMBO J. *15*, 5022–5031.

Bierhoff, H., Dammert, M.A., Brocks, D., Dambacher, S., Schotta, G., and Grummt, I. (2014). Quiescence-induced LncRNAs trigger H4K20 trimethylation and transcriptional silencing. Mol. Cell *54*, 675–682.

Bird, A. (2002). DNA methylation patterns and epigenetic memory. Genes Dev. *16*, 6–21.

Bose, D.A., Donahue, G., Reinberg, D., Shiekhattar, R., Bonasio, R., and Berger, S.L. (2017). RNA Binding to CBP Stimulates Histone Acetylation and Transcription. Cell *168*, 135–149.e22.

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

Brockdorff, N. (2013). Noncoding RNA and Polycomb recruitment. RNA *19*, 429–442.

Brooun, A., Gajiwala, K.S., Deng, Y.-L., Liu, W., Bolaños, 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.

Buck-Koehntop, B.A., Stanfield, R.L., Ekiert, D.C., Martinez-Yamout, M.A., Dyson, H.J., Wilson, I.A., and Wright, P.E. (2012). Molecular basis for recognition of methylated and specific DNA sequences by the zinc finger protein Kaiso. Proc. Natl. Acad. Sci. U. S. A. *109*, 15229–15234.

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

Calvio, C., Neubauer, G., Mann, M., and Lamond, A.I. (1995). Identification of hnRNP P2 as TLS/FUS using electrospray mass spectrometry. RNA N. Y. N *1*, 724–733.

Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R.S., and Zhang, Y. (2002). Role of histone H3 lysine 27 methylation in Polycomb-group silencing. Science *298*, 1039–1043.

Chapman, R.D., Heidemann, M., Hintermair, C., and Eick, D. (2008). Molecular evolution of the RNA polymerase II CTD. Trends Genet. TIG *24*, 289–296.

Chu, C., Zhang, Q.C., da Rocha, S.T., Flynn, R.A., Bharadwaj, M., Calabrese, J.M., Magnuson, T., Heard, E., and Chang, H.Y. (2015). Systematic Discovery of Xist RNA Binding Proteins. Cell *161*, 404–416.

Ciferri, C., Lander, G.C., Maiolica, A., Herzog, F., Aebersold, R., and Nogales, E. (2012). Molecular architecture of human polycomb repressive complex 2. eLife *1*.

Cifuentes-Rojas, C., Hernandez, A.J., Sarma, K., and Lee, J.T. (2014a). Regulatory Interactions between RNA and Polycomb Repressive Complex 2. Mol. Cell *55*, 171–185.

Cifuentes-Rojas, C., Hernandez, A.J., Sarma, K., and Lee, J.T. (2014b). Regulatory Interactions between RNA and Polycomb Repressive Complex 2. Mol. Cell *55*, 171–185.

Clark, M.D., Marcum, R., Graveline, R., Chan, C.W., Xie, T., Chen, Z., Ding, Y., Zhang, Y., Mondragón, A., David, G., et al. (2015). Structural insights into the assembly of the histone deacetylase-associated Sin3L/Rpd3L corepressor complex. Proc. Natl. Acad. Sci. *112*, E3669–E3678.

Cléry, A., Blatter, M., and Allain, F.H.-T. (2008). RNA recognition motifs: boring? Not quite. Curr. Opin. Struct. Biol. *18*, 290–298.

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.

Corden, J.L., Cadena, D.L., Ahearn, J.M., and Dahmus, M.E. (1985). A unique structure at the carboxyl terminus of the largest subunit of eukaryotic RNA polymerase II. Proc. Natl. Acad. Sci. U. S. A. *82*, 7934–7938.

Cowling, V.H. (2010). Regulation of mRNA cap methylation. Biochem. J. *425*, 295–302.

Da Cruz, S., and Cleveland, D.W. (2011). Understanding the role of TDP-43 and FUS/TLS in ALS and beyond. Curr. Opin. Neurobiol. *21*, 904–919.

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

Davidovich, C., Zheng, L., Goodrich, K.J., and Cech, T.R. (2013a). Promiscuous RNA binding by Polycomb repressive complex 2. Nat. Struct. Mol. Biol. *20*, 1250–1257.

Davidovich, C., Zheng, L., Goodrich, K.J., and Cech, T.R. (2013b). Promiscuous RNA binding by Polycomb repressive complex 2. Nat. Struct. Mol. Biol. *20*, 1250–1257.

Davidovich, C., Goodrich, K.J., Gooding, A.R., and Cech, T.R. (2014). A dimeric state for PRC2. Nucleic Acids Res. *42*, 9236–9248.

Davidovich, C., Wang, X., Cifuentes-Rojas, C., Goodrich, K.J., Gooding, A.R., Lee, J.T., and Cech, T.R. (2015a). Toward a Consensus on the Binding Specificity and Promiscuity of PRC2 for RNA. Mol. Cell *57*, 552–558.

Davidovich, C., Wang, X., Cifuentes-Rojas, C., Goodrich, K.J., Gooding, A.R., Lee, J.T., and Cech, T.R. (2015b). Toward a Consensus on the Binding Specificity and Promiscuity of PRC2 for RNA. Mol. Cell *57*, 552–558.

Deal, R.B., Henikoff, J.G., and Henikoff, S. (2010). Genome-Wide Kinetics of Nucleosome Turnover Determined by Metabolic Labeling of Histones. Science *328*, 1161–1164.

Deaton, A.M., and Bird, A. (2011). CpG islands and the regulation of transcription. Genes Dev. *25*, 1010–1022.

Delattre, M., Spierer, A., Tonka, C.H., and Spierer, P. (2000). The genomic silencing of position-effect variegation in Drosophila melanogaster: interaction between the heterochromatin-associated proteins Su(var)3-7 and HP1. J. Cell Sci. *113*, 4253–4261.

Deng, C., Li, Y., Zhou, L., Cho, J., Patel, B., Terada, N., Li, Y., Bungert, J., Qiu, Y., and Huang, S. (2016). HoxBlinc RNA Recruits Set1/MLL Complexes to Activate Hox Gene Expression Patterns and Mesoderm Lineage Development. Cell Rep. *14*, 103–114.

Deng, H.-X., Zhai, H., Bigio, E.H., Yan, J., Fecto, F., Ajroud, K., Mishra, M., Ajroud-Driss, S., Heller, S., Sufit, R., et al. (2010). FUS-immunoreactive inclusions are a common feature in sporadic and non-SOD1 familial amyotrophic lateral sclerosis. Ann. Neurol. *67*, 739–748.

Ding, J., Hayashi, M.K., Zhang, Y., Manche, L., Krainer, A.R., and Xu, R.M. (1999). Crystal structure of the two-RRM domain of hnRNP A1 (UP1) complexed with single-stranded telomeric DNA. Genes Dev. *13*, 1102–1115.

Dormann, D., Madl, T., Valori, C.F., Bentmann, E., Tahirovic, S., Abou-Ajram, C., Kremmer, E., Ansorge, O., Mackenzie, I.R.A., Neumann, M., et al. (2012). Arginine methylation next to the PY-NLS modulates Transportin binding and nuclear import of FUS. EMBO J. *31*, 4258–4275.

Du, Z., Lee, J.K., Tjhen, R., Li, S., Pan, H., Stroud, R.M., and James, T.L. (2005). Crystal Structure of the First KH Domain of Human Poly(C)-binding Protein-2 in Complex with a C-rich Strand of Human Telomeric DNA at 1.7 Å. J. Biol. Chem. *280*, 38823–38830.

Dyer, P.N., Edayathumangalam, R.S., White, C.L., Bao, Y., Chakravarthy, S., Muthurajan, U.M., and Luger, K. (2004). Reconstitution of nucleosome core particles from recombinant histones and DNA. Methods Enzymol. *375*, 23–44.

Fan, J., Xing, Y., Wen, X., Jia, R., Ni, H., He, J., Ding, X., Pan, H., Qian, G., Ge, S., et al. (2015). Long non-coding RNA ROR decoys gene-specific histone methylation to promote tumorigenesis. Genome Biol. *16*, 139.

Faustino, N.A., and Cooper, T.A. (2003). Pre-mRNA splicing and human disease. Genes Dev. *17*, 419–437.

Fenouil, R., Cauchy, P., Koch, F., Descostes, N., Cabeza, J.Z., Innocenti, C., Ferrier, P., Spicuglia, S., Gut, M., Gut, I., et al. (2012). CpG islands and GC content dictate

nucleosome depletion in a transcription-independent manner at mammalian promoters. Genome Res. *22*, 2399–2408.

Fiesel, F.C., and Kahle, P.J. (2011). TDP-43 and FUS/TLS: cellular functions and implications for neurodegeneration. FEBS J. *278*, 3550–3568.

Foster, C.T., Dovey, O.M., Lezina, L., Luo, J.L., Gant, T.W., Barlev, N., Bradley, A., and Cowley, S.M. (2010). Lysine-Specific Demethylase 1 Regulates the Embryonic Transcriptome and CoREST Stability. Mol. Cell. Biol. *30*, 4851–4863.

Frank, D.N., and Pace, N.R. (1998). Ribonuclease P: Unity and Diversity in a tRNA Processing Ribozyme. Annu. Rev. Biochem. *67*, 153–180.

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

Ghildiyal, M., and Zamore, P.D. (2009). Small silencing RNAs: an expanding universe. Nat. Rev. Genet. *10*, 94–108.

Gil, J., and O'Loghlen, A. (2014). PRC1 complex diversity: where is it taking us? Trends Cell Biol. *24*, 632–641.

Goodrich, J.A., and Kugel, J.F. (2007). Binding and Kinetics for Molecular Biologists (CSHL Press).

Grijzenhout, A., Godwin, J., Koseki, H., Gdula, M.R., Szumska, D., McGouran, J.F., Bhattacharya, S., Kessler, B.M., Brockdorff, N., and Cooper, S. (2016). Functional analysis of AEBP2, a PRC2 Polycomb protein, reveals a Trithorax phenotype in embryonic development and in ESCs. Dev. Camb. Engl. *143*, 2716–2723.

Grishin, N.V. (2001). KH domain: one motif, two folds. Nucleic Acids Res. *29*, 638–643.

Grote, P., Wittler, L., Hendrix, D., Koch, F., Währisch, S., Beisaw, A., Macura, K., Bläss, G., Kellis, M., Werber, M., et al. (2013). The tissue-specific lncRNA Fendrr is an essential regulator of heart and body wall development in the mouse. Dev. Cell *24*, 206–214.

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

Hansen, K.H., Bracken, A.P., Pasini, D., Dietrich, N., Gehani, S.S., Monrad, A., Rappsilber, J., Lerdrup, M., and Helin, K. (2008). A model for transmission of the H3K27me3 epigenetic mark. Nat. Cell Biol. *10*, 1291–1300.

He, L., and Hannon, G.J. (2004). MicroRNAs: small RNAs with a big role in gene regulation. Nat. Rev. Genet. *5*, 522–531.

Hewitt, C., Kirby, J., Highley, J.R., Hartley, J.A., Hibberd, R., Hollinger, H.C., Williams, T.L., Ince, P.G., McDermott, C.J., and Shaw, P.J. (2010). Novel FUS/TLS mutations and pathology in familial and sporadic amyotrophic lateral sclerosis. Arch. Neurol. *67*, 455–461.

Hino, S., Kohrogi, K., and Nakao, M. (2016). Histone demethylase LSD1 controls the phenotypic plasticity of cancer cells. Cancer Sci. *107*, 1187–1192.

Hoell, J.I., Larsson, E., Runge, S., Nusbaum, J.D., Duggimpudi, S., Farazi, T.A., Hafner, M., Borkhardt, A., Sander, C., and Tuschl, T. (2011). RNA targets of wild-type and mutant FET family proteins. Nat. Struct. Mol. Biol. *18*, 1428–1431.

Iko, Y., Kodama, T.S., Kasai, N., Oyama, T., Morita, E.H., Muto, T., Okumura, M., Fujii, R., Takumi, T., Tate, S., et al. (2004). Domain architectures and characterization of an RNA-binding protein, TLS. J. Biol. Chem. *279*, 44834–44840.

Ishigaki, S., Masuda, A., Fujioka, Y., Iguchi, Y., Katsuno, M., Shibata, A., Urano, F., Sobue, G., and Ohno, K. (2012). Position-dependent FUS-RNA interactions regulate alternative splicing events and transcriptions. Sci. Rep. *2*, 529.

Ishizu, H., Siomi, H., and Siomi, M.C. (2012). Biology of PIWI-interacting RNAs: new insights into biogenesis and function inside and outside of germlines. Genes Dev. *26*, 2361–2373.

Iwasaki, Y.W., Siomi, M.C., and Siomi, H. (2015). PIWI-Interacting RNA: Its Biogenesis and Functions. Annu. Rev. Biochem. *84*, 405–433.

Jackson, R.J., Hellen, C.U.T., and Pestova, T.V. (2010). The mechanism of eukaryotic translation initiation and principles of its regulation. Nat. Rev. Mol. Cell Biol. *11*, 113–127.

James, T.C., and Elgin, S.C. (1986). Identification of a nonhistone chromosomal protein associated with heterochromatin in Drosophila melanogaster and its gene. Mol. Cell. Biol. *6*, 3862–3872.

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

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

Jin, B., Yao, B., Li, J.-L., Fields, C.R., Delmas, A.L., Liu, C., and Robertson, K.D. (2009). DNMT1 and DNMT3B modulate distinct polycomb-mediated histone modifications in colon cancer. Cancer Res. *69*, 7412–7421.

Johansson, H.E., Dertinger, D., LeCuyer, K.A., Behlen, L.S., Greef, C.H., and Uhlenbeck, O.C. (1998). A thermodynamic analysis of the sequence-specific binding of RNA by bacteriophage MS2 coat protein. Proc. Natl. Acad. Sci. U. S. A. *95*, 9244–9249.

Kadrmas, J.L., and Beckerle, M.C. (2004). The LIM domain: from the cytoskeleton to the nucleus. Nat. Rev. Mol. Cell Biol. *5*, 920–931.

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.

Kaneko, S., Son, J., Shen, S.S., Reinberg, D., and Bonasio, R. (2013a). PRC2 binds active promoters and contacts nascent RNAs in embryonic stem cells. Nat. Struct. Mol. Biol. *20*, 1258–1264.

Kaneko, S., Son, J., Shen, S.S., Reinberg, D., and Bonasio, R. (2013b). PRC2 binds active promoters and contacts nascent RNAs in embryonic stem cells. Nat. Struct. Mol. Biol. *20*, 1258–1264.

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

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

Kaneko, S., Bonasio, R., Saldaña-Meyer, R., Yoshida, T., Son, J., Nishino, K., Umezawa, A., and Reinberg, D. (2014c). Interactions between JARID2 and noncoding RNAs regulate PRC2 recruitment to chromatin. Mol. Cell *53*, 290–300.

Kanhere, A., Viiri, K., Araújo, C.C., Rasaiyaah, J., Bouwman, R.D., Whyte, W.A., Pereira, C.F., Brookes, E., Walker, K., Bell, G.W., et al. (2010). Short RNAs are transcribed from repressed polycomb target genes and interact with polycomb repressive complex-2. Mol. Cell *38*, 675–688.

Kato, M., Han, T.W., Xie, S., Shi, K., Du, X., Wu, L.C., Mirzaei, H., Goldsmith, E.J., Longgood, J., Pei, J., et al. (2012). Cell-free Formation of RNA Granules: Low Complexity Sequence Domains Form Dynamic Fibers within Hydrogels. Cell *149*, 753–767.

Keller, C., Kulasegaran-Shylini, R., Shimada, Y., Hotz, H.-R., and Bühler, M. (2013). Noncoding RNAs prevent spreading of a repressive histone mark. Nat. Struct. Mol. Biol. *20*, 994–1000.

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.

Kiernan, M.C., Vucic, S., Cheah, B.C., Turner, M.R., Eisen, A., Hardiman, O., Burrell, J.R., and Zoing, M.C. (2011). Amyotrophic lateral sclerosis. Lancet Lond. Engl. *377*, 942–955.

Kim, K.H., and Roberts, C.W.M. (2016). Targeting EZH2 in cancer. Nat. Med. *22*, 128–134.

Klymenko, T., Papp, B., Fischle, W., Köcher, T., Schelder, M., Fritsch, C., Wild, B., Wilm, M., and Müller, J. (2006). A Polycomb group protein complex with sequence-specific DNA-binding and selective methyl-lysine-binding activities. Genes Dev. *20*, 1110–1122.

Kodama, Y., Shumway, M., Leinonen, R., and International Nucleotide Sequence Database Collaboration (2012). The Sequence Read Archive: explosive growth of sequencing data. Nucleic Acids Res. *40*, D54-56.

Ku, M., Koche, R.P., Rheinbay, E., Mendenhall, E.M., Endoh, M., Mikkelsen, T.S., Presser, A., Nusbaum, C., Xie, X., Chi, A.S., et al. (2008). Genomewide Analysis of PRC1 and PRC2 Occupancy Identifies Two Classes of Bivalent Domains. PLOS Genet. *4*, e1000242.

Kwon, I., Kato, M., Xiang, S., Wu, L., Theodoropoulos, P., Mirzaei, H., Han, T., Xie, S., Corden, J.L., and McKnight, S.L. (2013). Phosphorylation-regulated binding of RNA polymerase II to fibrous polymers of low-complexity domains. Cell *155*, 1049–1060.

Lagier-Tourenne, C., Polymenidou, M., and Cleveland, D.W. (2010). TDP-43 and FUS/TLS: emerging roles in RNA processing and neurodegeneration. Hum. Mol. Genet. *19*, R46-64.

Lagier-Tourenne, C., Polymenidou, M., Hutt, K.R., Vu, A.Q., Baughn, M., Huelga, S.C., Clutario, K.M., Ling, S.-C., Liang, T.Y., Mazur, C., et al. (2012). Divergent roles of ALS-linked proteins FUS/TLS and TDP-43 intersect in processing long pre-mRNAs. Nat. Neurosci. *15*, 1488–1497.

Laity, J.H., Lee, B.M., and Wright, P.E. (2001). Zinc finger proteins: new insights into structural and functional diversity. Curr. Opin. Struct. Biol. *11*, 39–46.

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

LeCuyer, K.A., Behlen, L.S., and Uhlenbeck, O.C. (1995). Mutants of the Bacteriophage MS2 Coat Protein That Alter Its Cooperative Binding to RNA. Biochemistry (Mosc.) *34*, 10600–10606.

Lerga, A., Hallier, M., Delva, L., Orvain, C., Gallais, I., Marie, J., and Moreau-Gachelin, F. (2001). Identification of an RNA binding specificity for the potential splicing factor TLS. J. Biol. Chem. *276*, 6807–6816.

Lewis, P.W., Müller, M.M., Koletsky, M.S., Cordero, F., Lin, S., Banaszynski, L.A., Garcia, B.A., Muir, T.W., Becher, O.J., and Allis, C.D. (2013). Inhibition of PRC2 Activity by a Gain-of-Function H3 Mutation Found in Pediatric Glioblastoma. Science 1232245.

Li, G., Margueron, R., Ku, M., Chambon, P., Bernstein, B.E., and Reinberg, D. (2010). Jarid2 and PRC2, partners in regulating gene expression. Genes Dev. *24*, 368–380.

Lim, D.A., Huang, Y.-C., Swigut, T., Mirick, A.L., Garcia-Verdugo, J.M., Wysocka, J., Ernst, P., and Alvarez-Buylla, A. (2009). Chromatin remodelling factor Mll1 is essential for neurogenesis from postnatal neural stem cells. Nature *458*, 529–533.

Liu, X., Niu, C., Ren, J., Zhang, J., Xie, X., Zhu, H., Feng, W., and Gong, W. (2013). The RRM domain of human fused in sarcoma protein reveals a non-canonical nucleic acid binding site. Biochim. Biophys. Acta *1832*, 375–385.

Lodish, H., Berk, A., Zipursky, S.L., Matsudaira, P., Baltimore, D., and Darnell, J. (2000). The Three Roles of RNA in Protein Synthesis.

Lunde, B.M., Moore, C., and Varani, G. (2007). RNA-binding proteins: modular design for efficient function. Nat. Rev. Mol. Cell Biol. *8*, 479–490.

Maison, C., Bailly, D., Peters, A.H.F.M., Quivy, J.-P., Roche, D., Taddei, A., Lachner, M., Jenuwein, T., and Almouzni, G. (2002). Higher-order structure in pericentric heterochromatin involves a distinct pattern of histone modification and an RNA component. Nat. Genet. *30*, 329–334.

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

Margueron, R., Justin, N., Ohno, K., Sharpe, M.L., Son, J., Drury Iii, W.J., Voigt, P., Martin, S.R., Taylor, W.R., De Marco, V., et al. (2009). Role of the polycomb protein EED in the propagation of repressive histone marks. Nature *461*, 762–767.

Maris, C., Dominguez, C., and Allain, F.H.-T. (2005). The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. FEBS J. *272*, 2118–2131.

Mastrocola, A.S., Kim, S.H., Trinh, A.T., Rodenkirch, L.A., and Tibbetts, R.S. (2013). The RNA-binding protein fused in sarcoma (FUS) functions downstream of poly(ADP-ribose) polymerase (PARP) in response to DNA damage. J. Biol. Chem. *288*, 24731–24741.

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

McHugh, C.A., Chen, C.-K., Chow, A., Surka, C.F., Tran, C., McDonel, P., Pandya-Jones, A., Blanco, M., Burghard, C., Moradian, A., et al. (2015). The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. Nature *521*, 232–236.

McMahon, K.A., Hiew, S.Y.-L., Hadjur, S., Veiga-Fernandes, H., Menzel, U., Price, A.J., Kioussis, D., Williams, O., and Brady, H.J.M. (2007). Mll Has a Critical Role in Fetal and Adult Hematopoietic Stem Cell Self-Renewal. Cell Stem Cell *1*, 338–345.

Meissner, A., Mikkelsen, T.S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B.E., Nusbaum, C., Jaffe, D.B., et al. (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature *454*, 766–770.

Mendenhall, E.M., Koche, R.P., Truong, T., Zhou, V.W., Issac, B., Chi, A.S., Ku, M., and Bernstein, B.E. (2010a). GC-Rich Sequence Elements Recruit PRC2 in Mammalian ES Cells. PLOS Genet. *6*, e1001244.

Mendenhall, E.M., Koche, R.P., Truong, T., Zhou, V.W., Issac, B., Chi, A.S., Ku, M., and Bernstein, B.E. (2010b). GC-Rich Sequence Elements Recruit PRC2 in Mammalian ES Cells. PLOS Genet. *6*, e1001244.

Mihaly, J., Mishra, R.K., and Karch, F. (1998). A Conserved Sequence Motif in Polycomb-Response Elements. Mol. Cell *1*, 1065–1066.

Millard, C.J., Varma, N., Saleh, A., Morris, K., Watson, P.J., Bottrill, A.R., Fairall, L., Smith, C.J., and Schwabe, J.W. (2016). The structure of the core NuRD repression complex provides insights into its interaction with chromatin. eLife *5*, e13941.

Minajigi, A., Froberg, J.E., Wei, C., Sunwoo, H., Kesner, B., Colognori, D., Lessing, D., Payer, B., Boukhali, M., Haas, W., et al. (2015). A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation. Science *349*, aab2276.

Mitchell, S.F., Walker, S.E., Algire, M.A., Park, E.-H., Hinnebusch, A.G., and Lorsch, J.R. (2010). The 5′-7-methylguanosine cap on eukaryotic mRNAs serves both to stimulate canonical translation initiation and block an alternative pathway. Mol. Cell *39*, 950–962.

Mito, Y., Henikoff, J.G., and Henikoff, S. (2007). Histone Replacement Marks the Boundaries of cis-Regulatory Domains. Science *315*, 1408–1411.

Moindrot, B., Cerase, A., Coker, H., Masui, O., Grijzenhout, A., Pintacuda, G., Schermelleh, L., Nesterova, T.B., and Brockdorff, N. (2015). A Pooled shRNA Screen Identifies Rbm15, Spen, and Wtap as Factors Required for Xist RNA-Mediated Silencing. Cell Rep. *12*, 562–572.

Monfort, A., Di Minin, G., Postlmayr, A., Freimann, R., Arieti, F., Thore, S., and Wutz, A. (2015). Identification of Spen as a Crucial Factor for Xist Function through Forward Genetic Screening in Haploid Embryonic Stem Cells. Cell Rep. *12*, 554–561.

Moretti, R., Donato, L.J., Brezinski, M.L., Stafford, R.L., Hoff, H., Thorson, J.S., Dervan, P.B., and Ansari, A.Z. (2008). Targeted chemical wedges reveal the role of allosteric DNA modulation in protein-DNA assembly. ACS Chem. Biol. *3*, 220–229.

Morey, L., Santanach, A., and Croce, L.D. (2015). Pluripotency and Epigenetic Factors in Mouse Embryonic Stem Cell Fate Regulation. Mol. Cell. Biol. *35*, 2716–2728.

Muchardt, C., Guillemé, M., Seeler, J.-S., Trouche, D., Dejean, A., and Yaniv, M. (2002). Coordinated methyl and RNA binding is required for heterochromatin localization of mammalian HP1α. EMBO Rep. *3*, 975–981.

Müller, J., and Bienz, M. (1991). Long range repression conferring boundaries of Ultrabithorax expression in the Drosophila embryo. EMBO J. *10*, 3147–3155.

Müller, J., Hart, C.M., Francis, N.J., Vargas, M.L., Sengupta, A., Wild, B., Miller, E.L., O'Connor, M.B., Kingston, R.E., and Simon, J.A. (2002a). Histone Methyltransferase Activity of a Drosophila Polycomb Group Repressor Complex. Cell *111*, 197–208.

Müller, J., Hart, C.M., Francis, N.J., Vargas, M.L., Sengupta, A., Wild, B., Miller, E.L., O'Connor, M.B., Kingston, R.E., and Simon, J.A. (2002b). Histone Methyltransferase Activity of a Drosophila Polycomb Group Repressor Complex. Cell *111*, 197–208.

Muñoz, M.J., de la Mata, M., and Kornblihtt, A.R. (2010). The carboxy terminal domain of RNA polymerase II and alternative splicing. Trends Biochem. Sci. *35*, 497–504.

Musco, G., Stier, G., Joseph, C., Morelli, M.A.C., Nilges, M., Gibson, T.J., and Pastore, A. (1996). Three-Dimensional Structure and Stability of the KH Domain: Molecular Insights into the Fragile X Syndrome. Cell *85*, 237–245.

Myasnikov, A.G., Simonetti, A., Marzi, S., and Klaholz, B.P. (2009). Structure–function insights into prokaryotic and eukaryotic translation initiation. Curr. Opin. Struct. Biol. *19*, 300–309.

Nekrasov, M., Wild, B., and Müller, J. (2005). Nucleosome binding and histone methyltransferase activity of *Drosophila* PRC2. EMBO Rep. *6*, 348–353.

Nguyen, C.D., Mansfield, R.E., Leung, W., Vaz, P.M., Loughlin, F.E., Grant, R.P., and Mackay, J.P. (2011). Characterization of a family of RanBP2-type zinc fingers that can recognize single-stranded RNA. J. Mol. Biol. *407*, 273–283.

Nguyen, U.T.T., Bittova, L., Müller, M.M., Fierz, B., David, Y., Houck-Loomis, B., Feng, V., Dann, G.P., and Muir, T.W. (2014). Accelerated chromatin biochemistry using DNA-barcoded nucleosome libraries. Nat. Methods *11*, 834–840.

O'Carroll, D., Erhardt, S., Pagani, M., Barton, S.C., Surani, M.A., and Jenuwein, T. (2001). The polycomb-group gene Ezh2 is required for early mouse development. Mol. Cell. Biol. *21*, 4330–4336.

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

Pang, Y.L.J., Poruri, K., and Martinis, S.A. (2014). tRNA synthetase: tRNA Aminoacylation and beyond. Wiley Interdiscip. Rev. RNA *5*, 461–480.

Parker, R., and Song, H. (2004). The enzymes and control of eukaryotic mRNA turnover. Nat. Struct. Mol. Biol. *11*, 121–127.

Peng, J.C., Valouev, A., Swigut, T., Zhang, J., Zhao, Y., Sidow, A., and Wysocka, J. (2009). Jarid2/Jumonji coordinates control of PRC2 enzymatic activity and target gene occupancy in pluripotent cells. Cell *139*, 1290–1302.

Postepska-Igielska, A., Giwojna, A., Gasri-Plotnitsky, L., Schmitt, N., Dold, A., Ginsberg, D., and Grummt, I. (2015). LncRNA Khps1 Regulates Expression of the Proto-oncogene SPHK1 via Triplex-Mediated Changes in Chromatin Structure. Mol. Cell *60*, 626–636.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinforma. Oxf. Engl. *26*, 841–842.

Radman-Livaja, M., and Rando, O.J. (2010). Nucleosome positioning: how is it established, and why does it matter? Dev. Biol. *339*, 258–266.

Ray, D., Kazan, H., Cook, K.B., Weirauch, M.T., Najafabadi, H.S., Li, X., Gueroussov, S., Albu, M., Zheng, H., Yang, A., et al. (2013). A compendium of RNA-binding motifs for decoding gene regulation. Nature *499*, 172–177.

Riising, E.M., Comet, I., Leblanc, B., Wu, X., Johansen, J.V., and Helin, K. (2014). Gene Silencing Triggers Polycomb Repressive Complex 2 Recruitment to CpG Islands Genome Wide. Mol. Cell *55*, 347–360.

Ringrose, L., and Paro, R. (2007). Polycomb/Trithorax response elements and epigenetic memory of cell identity. Dev. Camb. Engl. *134*, 223–232.

Rinn, J.L., and Chang, H.Y. (2012). Genome Regulation by Long Noncoding RNAs. Annu. Rev. Biochem. *81*, 145–166.

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.

Rogelj, B., Easton, L.E., Bogu, G.K., Stanton, L.W., Rot, G., Curk, T., Zupan, B., Sugimoto, Y., Modic, M., Haberman, N., et al. (2012). Widespread binding of FUS along nascent RNA regulates alternative splicing in the brain. Sci. Rep. *2*, 603.

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.

Sanulli, S., Justin, N., Teissandier, A., Ancelin, K., Portoso, M., Caron, M., Michaud, A., Lombard, B., da Rocha, S.T., Offer, J., et al. (2015). Jarid2 Methylation via the PRC2 Complex Regulates H3K27me3 Deposition during Cell Differentiation. Mol. Cell *57*, 769–783.

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

Schmitges, F.W., Prusty, A.B., Faty, M., Stützer, A., Lingaraju, G.M., Aiwazian, J., Sack, R., Hess, D., Li, L., Zhou, S., et al. (2011a). Histone Methylation by PRC2 Is Inhibited by Active Chromatin Marks. Mol. Cell *42*, 330–341.

Schmitges, F.W., Prusty, A.B., Faty, M., Stützer, A., Lingaraju, G.M., Aiwazian, J., Sack, R., Hess, D., Li, L., Zhou, S., et al. (2011b). Histone Methylation by PRC2 Is Inhibited by Active Chromatin Marks. Mol. Cell *42*, 330–341.

Schones, D.E., Cui, K., Cuddapah, S., Roh, T.-Y., Barski, A., Wang, Z., Wei, G., and Zhao, K. (2008). Dynamic regulation of nucleosome positioning in the human genome. Cell *132*, 887–898.

Schuettengruber, B., Chourrout, D., Vervoort, M., Leblanc, B., and Cavalli, G. (2007). Genome Regulation by Polycomb and Trithorax Proteins. Cell *128*, 735–745.

Schuettengruber, B., Martinez, A.-M., Iovino, N., and Cavalli, G. (2011). Trithorax group proteins: switching genes on and keeping them active. Nat. Rev. Mol. Cell Biol. *12*, 799–814.

Schwartz, J.C., Ebmeier, C.C., Podell, E.R., Heimiller, J., Taatjes, D.J., and Cech, T.R. (2012). FUS binds the CTD of RNA polymerase II and regulates its phosphorylation at Ser2. Genes Dev. *26*, 2690–2695.

Schwartz, J.C., Wang, X., Podell, E.R., and Cech, T.R. (2013). RNA seeds higher-order assembly of FUS protein. Cell Rep. *5*, 918–925.

Sengupta, A.K., Kuhrs, A., and Müller, J. (2004). General transcriptional silencing by a Polycomb response element in Drosophila. Development *131*, 1959–1965.

Shen, X., Liu, Y., Hsu, Y.-J., Fujiwara, Y., Kim, J., Mao, X., Yuan, G.-C., and Orkin, S.H. (2008). EZH1 Mediates Methylation on Histone H3 Lysine 27 and Complements EZH2 in Maintaining Stem Cell Identity and Executing Pluripotency. Mol. Cell *32*, 491–502.

Shen, X., Kim, W., Fujiwara, Y., Simon, M.D., Liu, Y., Mysliwiec, M.R., Yuan, G.-C., Lee, Y., and Orkin, S.H. (2009). Jumonji modulates polycomb activity and self-renewal versus differentiation of stem cells. Cell *139*, 1303–1314.

Shi, H., and Moore, P.B. (2000). The crystal structure of yeast phenylalanine tRNA at 1.93 A resolution: a classic structure revisited. RNA N. Y. N *6*, 1091–1105.

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

Son, J., Shen, S.S., Margueron, R., and Reinberg, D. (2013). Nucleosome-binding activities within JARID2 and EZH1 regulate the function of PRC2 on chromatin. Genes Dev. *27*, 2663–2677.

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.

Tachibana, M., Sugimoto, K., Nozaki, M., Ueda, J., Ohta, T., Ohki, M., Fukuda, M., Takeda, N., Niida, H., Kato, H., et al. (2002). G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis. Genes Dev. *16*, 1779–1791.

Takahama, K., Takada, A., Tada, S., Shimizu, M., Sayama, K., Kurokawa, R., and Oyoshi, T. (2013). Regulation of telomere length by G-quadruplex telomere DNA- and TERRA-binding protein TLS/FUS. Chem. Biol. *20*, 341–350.

Tan, A.Y., and Manley, J.L. (2010). TLS inhibits RNA polymerase III transcription. Mol. Cell. Biol. *30*, 186–196.

Tan, A.Y., Riley, T.R., Coady, T., Bussemaker, H.J., and Manley, J.L. (2012). TLS/FUS (translocated in liposarcoma/fused in sarcoma) regulates target gene transcription via single-stranded DNA response elements. Proc. Natl. Acad. Sci. U. S. A. *109*, 6030–6035.

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.

Tripsianes, K., Friberg, A., Barrandon, C., Brooks, M., Tilbeurgh, H. van, Seraphin, B., and Sattler, M. (2014). A Novel Protein-Protein Interaction in the RES (REtention and Splicing) Complex. J. Biol. Chem. *289*, 28640–28650.

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.

Viré, E., Brenner, C., Deplus, R., Blanchon, L., Fraga, M., Didelot, C., Morey, L., Van Eynde, A., Bernard, D., Vanderwinden, J.-M., et al. (2006). The Polycomb group protein EZH2 directly controls DNA methylation. Nature *439*, 871–874.

Vo, N., and Goodman, R.H. (2001). CREB-binding Protein and p300 in Transcriptional Regulation. J. Biol. Chem. *276*, 13505–13508.

Vorlíčková, M., Kejnovská, I., Sagi, J., Renčiuk, D., Bednářová, K., Motlová, J., and Kypr, J. (2012). Circular dichroism and guanine quadruplexes. Methods San Diego Calif *57*, 64–75.

Wagschal, A., Sutherland, H.G., Woodfine, K., Henckel, A., Chebli, K., Schulz, R., Oakey, R.J., Bickmore, W.A., and Feil, R. (2008). G9a Histone Methyltransferase Contributes to Imprinting in the Mouse Placenta. Mol. Cell. Biol. *28*, 1104–1113.

Wang, X., and Tanaka Hall, T.M. (2001). Structural basis for recognition of AU-rich element RNA by the HuD protein. Nat. Struct. Mol. Biol. *8*, 141–145.

Wang, K.C., Yang, Y.W., Liu, B., Sanyal, A., Corces-Zimmerman, R., Chen, Y., Lajoie, B.R., Protacio, A., Flynn, R.A., Gupta, R.A., et al. (2011). A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. Nature *472*, 120–124.

Wang, W.-Y., Pan, L., Su, S.C., Quinn, E.J., Sasaki, M., Jimenez, J.C., Mackenzie, I.R.A., Huang, E.J., and Tsai, L.-H. (2013). Interaction of FUS and HDAC1 regulates DNA damage response and repair in neurons. Nat. Neurosci. *16*, 1383–1391.

Wang, X., Arai, S., Song, X., Reichart, D., Du, K., Pascual, G., Tempst, P., Rosenfeld, M.G., Glass, C.K., and Kurokawa, R. (2008). Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription. Nature *454*, 126–130.

Wang, X., Goodrich, K.J., Gooding, A.R., Naeem, H., Archer, S., Paucek, R.D., Youmans, D.T., Cech, T.R., and Davidovich, C. (2017). Targeting of Polycomb Repressive Complex 2 to RNA by Short Repeats of Consecutive Guanines. Mol. Cell *65*, 1056–1067.e5.

Widlak, P., and Garrard, W.T. (2006). Unique features of the apoptotic endonuclease DFF40/CAD relative to micrococcal nuclease as a structural probe for chromatin. Biochem. Cell Biol. Biochim. Biol. Cell. *84*, 405–410.

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.

Wilusz, J.E., Sunwoo, H., and Spector, D.L. (2009). Long noncoding RNAs: functional surprises from the RNA world. Genes Dev. *23*, 1494–1504.

Winter, J., Jung, S., Keller, S., Gregory, R.I., and Diederichs, S. (2009). Many roads to maturity: microRNA biogenesis pathways and their regulation. Nat. Cell Biol. *11*, 228–234.

Wu, H., D'Alessio, A.C., Ito, S., Wang, Z., Cui, K., Zhao, K., Sun, Y.E., and Zhang, Y. (2011). Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells. Genes Dev. *25*, 679–684.

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

Xu, C., Bian, C., Yang, W., Galka, M., Ouyang, H., Chen, C., Qiu, W., Liu, H., Jones, A.E., MacKenzie, F., et al. (2010). Binding of different histone marks differentially regulates the activity and specificity of polycomb repressive complex 2 (PRC2). Proc. Natl. Acad. Sci. *107*, 19266–19271.

Yang, L., Lin, C., Liu, W., Zhang, J., Ohgi, K.A., Grinstein, J.D., Dorrestein, P.C., and Rosenfeld, M.G. (2011). ncRNA- and Pc2 methylation-dependent gene relocation between nuclear structures mediates gene activation programs. Cell *147*, 773–788.

Yap, K.L., Li, S., Muñoz-Cabello, A.M., Raguz, S., Zeng, L., Mujtaba, S., Gil, J., Walsh, M.J., and Zhou, M.-M. (2010). 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.

Yuan, W., Wu, T., Fu, H., Dai, C., Wu, H., Liu, N., Li, X., Xu, M., Zhang, Z., Niu, T., et al. (2012a). Dense chromatin activates Polycomb repressive complex 2 to regulate H3 lysine 27 methylation. Science *337*, 971–975.

Yuan, W., Wu, T., Fu, H., Dai, C., Wu, H., Liu, N., Li, X., Xu, M., Zhang, Z., Niu, T., et al. (2012b). Dense chromatin activates Polycomb repressive complex 2 to regulate H3 lysine 27 methylation. Science *337*, 971–975.

Zappulla, D.C., and Cech, T.R. (2004). Yeast telomerase RNA: A flexible scaffold for protein subunits. Proc. Natl. Acad. Sci. U. S. A. *101*, 10024–10029.

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.

Zhao, J., Sun, B.K., Erwin, J.A., Song, J.-J., and Lee, J.T. (2008a). Polycomb Proteins Targeted by a Short Repeat RNA to the Mouse X Chromosome. Science *322*, 750–756.

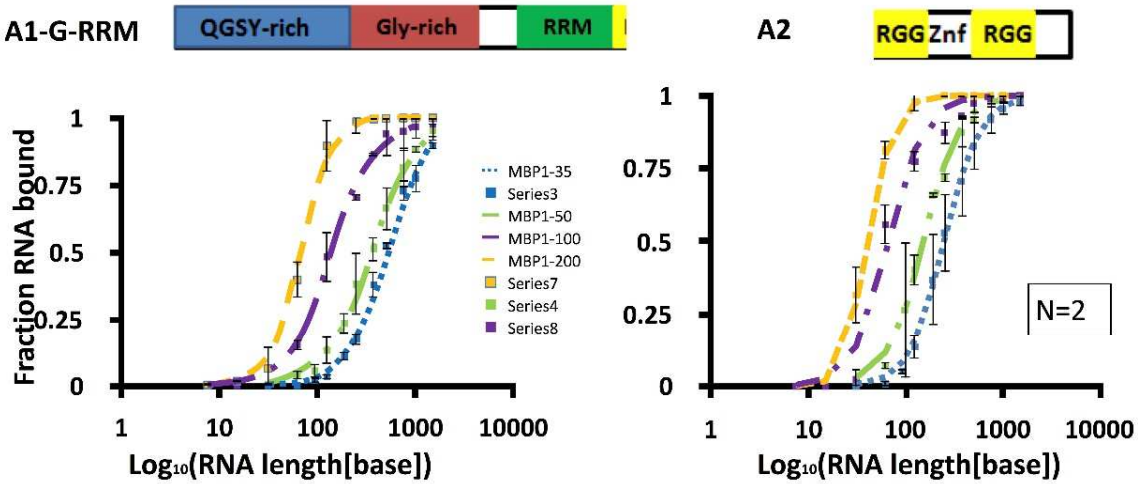
Zhao, J., Sun, B.K., Erwin, J.A., Song, J.-J., and Lee, J.T. (2008b). Polycomb Proteins Targeted by a Short Repeat RNA to the Mouse X Chromosome. Science *322*, 750–756.

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.

Zylicz, J.J., Dietmann, S., Günesdogan, U., Hackett, J.A., Cougot, D., Lee, C., and Surani, M.A. (2015). Chromatin dynamics and the role of G9a in gene regulation and enhancer silencing during early mouse development. eLife *4*, e09571.

Appendix A: Both RRM and Zinc Finger motifs of FUS bind RNA in a length-dependent manner

As shown in Chapter II, FUS binds RNA promiscuously. Both N-terminal and C-terminal truncations containing RRM and Zinc Finger, respectively, interact with RNA. We hypothesized that either RRM or Znf is in charge of the promiscuous recognition of RNA by FUS. To test this hypothesis, we repeated the MBP-RNA serial binding as we performed in Chapter II with both of the truncated proteins. We expected one truncation to preserve its ability to bind RNA in a length-dependent manner, as in the case of full-length FUS, with the other truncated protein lacking promiscuous RNA binding ability. Surprisingly, for both truncations we observed an incremental increase in affinity with increasing RNA length as we previously observed for full-length FUS. By quantifying binding affinities for both truncations, the binding curve of protein and each RNA began to shift from right to left as the RNA length increased (Figure A2), indicating a more tight interaction. Plotting log (K_d) versus log (RNA length) revealed a linear relationship between dissociation constant and RNA length with a slope value of −1 (Figure A2) for both truncations. Together, this data suggests that both truncations bind RNA promiscuously and are identical to WT FUS in terms of their RNA-binding properties.



Dissociation Constant (Kd) and Hill Coefficients(n):	
MBP1-35: Kd=494±38nM n=2.1±0.1	MBP1-35: Kd=280±58nM n=2.5±0.2
MBP1-50: Kd=349±56nM n=1.9±0.1	MBP1-50: Kd=167±18nM n=2.3±0.3
MBP1-100: Kd=139 ± 7nM n=1.9±0.2	MBP1-100: Kd=67±9nM n=2.4±0.2
MBP1-200: Kd=69±3nM n=3.8±0.1	MBP1-200: Kd=40±5nM n=3.4±0.1

Figure A1. Binding curves of FUS truncations binding to RNAs of different length. Left, binding curves for FUS N-terminal truncations containing RRM and RNAs comprising 35, 50, 100 and 200 bases of *E.coli* MBP mRNAs. Right, binding curves for FUS C-terminal truncations containing Zinc Finger and RNAs comprising 35, 50, 100 and 200 bases of *E. coli* MBP mRNAs. Error bars represent the range of two or three replicates.

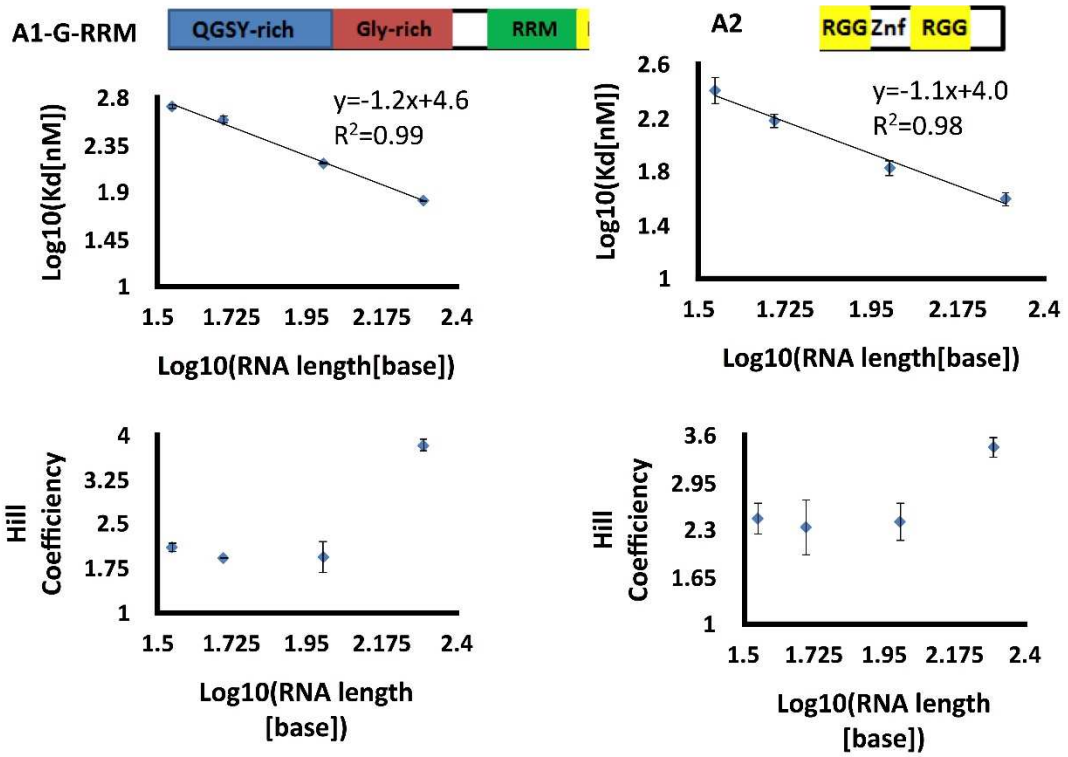


Figure A2. Additional analysis of binding curves of (A1) for both N-terminal (Left) and C-terminal (Right) truncations. Plotting log (K_d^{app}) versus log (RNA length) revealed a linear relationship with a slope of −1.

Appendix B: The affinity of FUS binding to RNA is not influenced by salt concentration

In chapter II, I demonstrated that FUS binds RNA promiscuously. One simple hypothesis to explain such promiscuous RNA binding is that FUS interacts with RNA electrostatically. To test this hypothesis, FUS-RNA binding reactions were performed at six different KCl concentrations (Figure B1). Unexpectedly, the binding affinity did not change at different concentrations of KCl (Figure B2). This suggests that the interaction between FUS and RNA is not primarily driven by electrostatics. In addition, we observed that the Hill co-efficient of FUS-RNA binding decreased as the concentration of KCl increased (Figure B2). In other words, the degree of cooperativity of FUS binding to RNA is altered by salt concentration. Perhaps higher concentration of KCl could hinder FUS-FUS interaction, lowering cooperativity.

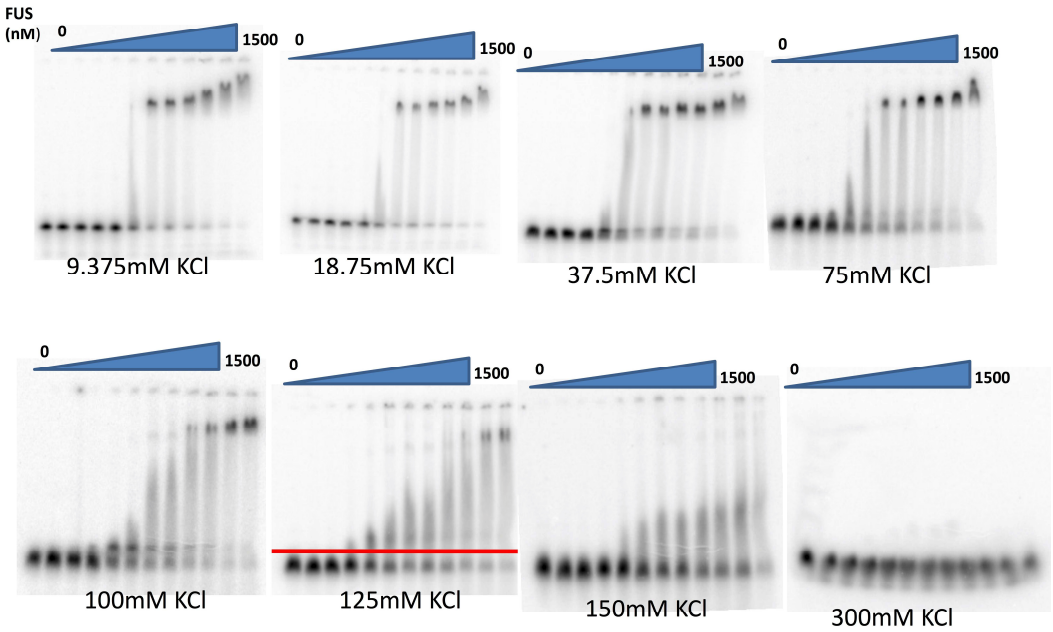
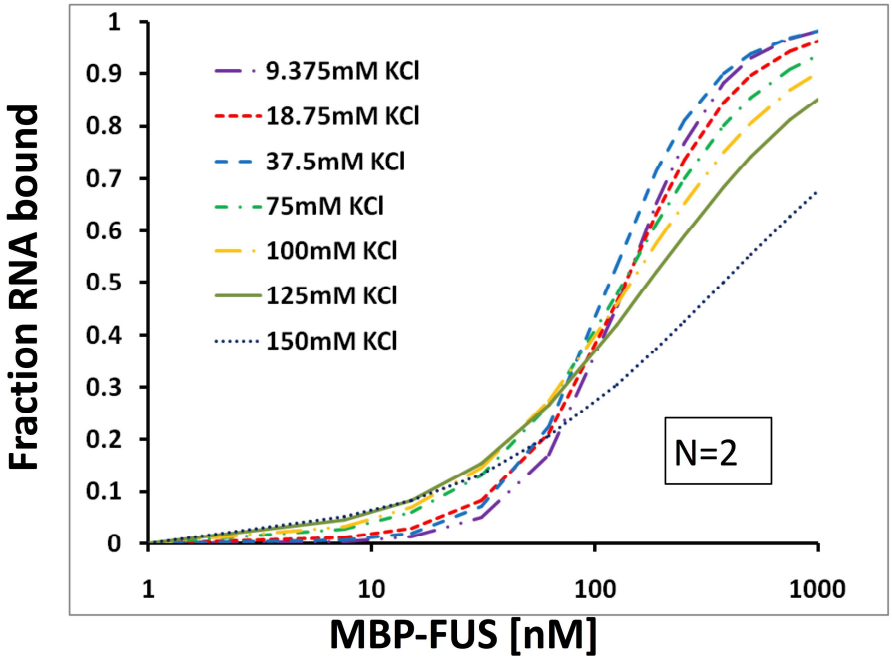


Figure B1. Representative EMSA gels of FUS binding to RNA at various salt concentrations. RNA binding experiments were performed at a salt concentration range of 9.375 mM to 300 mM. For all binding experiments, FUS concentration was titrated in 2-fold increments to 1500 nM.



Dissociation Constant (K_d^{app}) and Hill Coefficients(*n*):

9.375mM KCl:	K_d^{app}=137±9nM	*n*=2.0±0.1
18.75mM KCl:	K_d^{app}=135±10nM	*n*=1.7±0.2
37.5mM KCl:	K_d^{app}=111±2nM	n=1.9±0.1
75mM KCl:	K_d^{app}=132±18nM	n=1.4±0.4
100mM KCl:	K_d^{app}=144±4nM	n=1.2±0.1
125mM KCl:	K_d^{app}=173±27nM	n=1.0±0.1
150mM KCl:	K_d^{app}=375±24nM	n=0.8±0.1

Figure B2. FUS binding to RNA as a function of salt concentration. Equilibrium binding curves were generated by quantifying EMSA gels in Figure B1. Error bars represent an N = 2.

Appendix C: H3K27 methylation by PRC2 is regulated via a cis-acting disordered loop in EZH2

PRC2 is a histone methyltransferase, and is well known for its ability to methylate H3. As discussed in Chapter VI, RNA has been reported to inhibit PRC2 catalytic activity (Lee; Reingberg). Thus we began to test PRC2 catalytic activity using nucleosome substrates in the presence of RNA. PRC2 was incubated with *in vitro* reconstituted nucleosomes in the presence of ^{14}C-SAM as a substrate and the bands corresponding to methylation of H3K27 were detected by radiography (Figure C1.A). Surprisingly, another band other than H3K27me band was observed in the gel (Figure C1.A).

Literature has suggested that PRC2 is able to methylate JARID2 and perhaps other subunits. Therefore, we hypothesized that this signal came from PRC2 automethylating its subunits. To test this hypothesis and to determine which subunit is being methylated, if any, methylation experiments were repeated using PRC2 bearing an uncleavable MBP tag on a single subunit. Therefore, retarded electrophoretic mobility would be seen (Figure C1.B) and EZH2 has been suggested to be the major subunit methylated (Figure C1.B).

To further test whether EZH2 is methylated *in cis* or *in trans*, we purified a “catalytic dead” mutant (H694A on EZH2) and mixed it with MBP-tagged EZH2 (Data not shown). If EZH2 is methylated *in cis*, the amount of H694A methylation signal would not increase by mixing it with MBP-tagged EZH2. If methylation occurs *in trans*, an increased amount of H694A methylation signal would be expected. Even though some residual catalytic activity could still be observed in H694A, mixing H694A EZH2 with MBP-EZH2 in a 1:1 ratio did not increase the very weak catalytic activity of H694A EZH2. This suggested EZH2 is methylated *in cis*. To further corroborate the result, another PRC2 mutant was

made containing a Y726F mutation on EZH2. This mutation removes the tyrosine residue containing an essential hydroxyl group that initializes the S2N reaction that is the basis of the methyltransferase reaction. As anticipated, this mutant has fully abolished EZH2 catalytic activity (Figure C1.C). Mixing Y726F-EZH2 with MBP-EZH2 at 1:1 ratio did not show any signal in Y726F-EZH2, which is consistent with H694A-EZH2 (Figure C1.D). Collectively, these results suggest EZH2 automethylates itself *in cis*.

Mass-Spec analysis of PRC2 identified residues K510, KK514/515, K740 as those which are auto-methylated (Figure C2.A). By phylogenetic sequence alignment, the three lysine residues were found to be within the same CXC domain in EZH2 and are well conserved, even in Drosophila (Figure C2.B). To further confirm the results, a triple mutant PRC2 (K510A, KK514/K515AA) was generated. This mutant gave a 60% reduction in the highest SAM concentration tested compared to WT (Figure C2.C and D), indicating K510, KK514/515 were methylated by EZH2 and additional methylation site still existing, possibly K710.

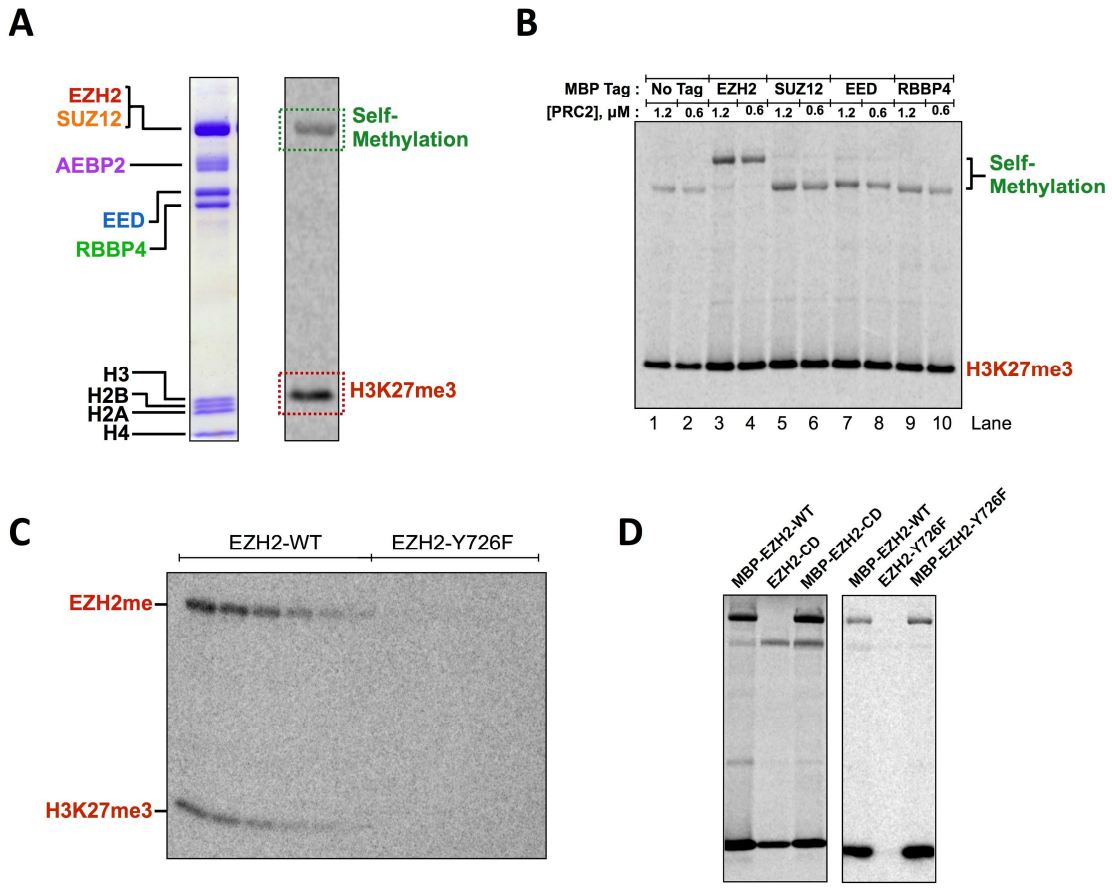


Figure C1. PRC2 methylates its EZH2 subunit. A. HMTase assays showing PRC2 enzymatic activity to mononucleosome and automethylation of PRC2. **B.** HMTase assays showing EZH2 is the subunit methylated by PRC2. HMTase assays were performed with one subunit containing uncleavable MBP tag and mononucleosmes. **C.** catalytic dead mutation (Y726F) abolished H3K27 methylation as well as EZH2 methylation. Reactions were carried out between PRC2 and mononucleosome, in the presence of cofactor ^{14}C-SAM. **D**. Mixing experiments between WT and catalytic dead mutation showing that automethylation occurs *in cis*. Experiments were performed by mixing WT PRC2 1:1 with catalytic dead mutant (Y726F).

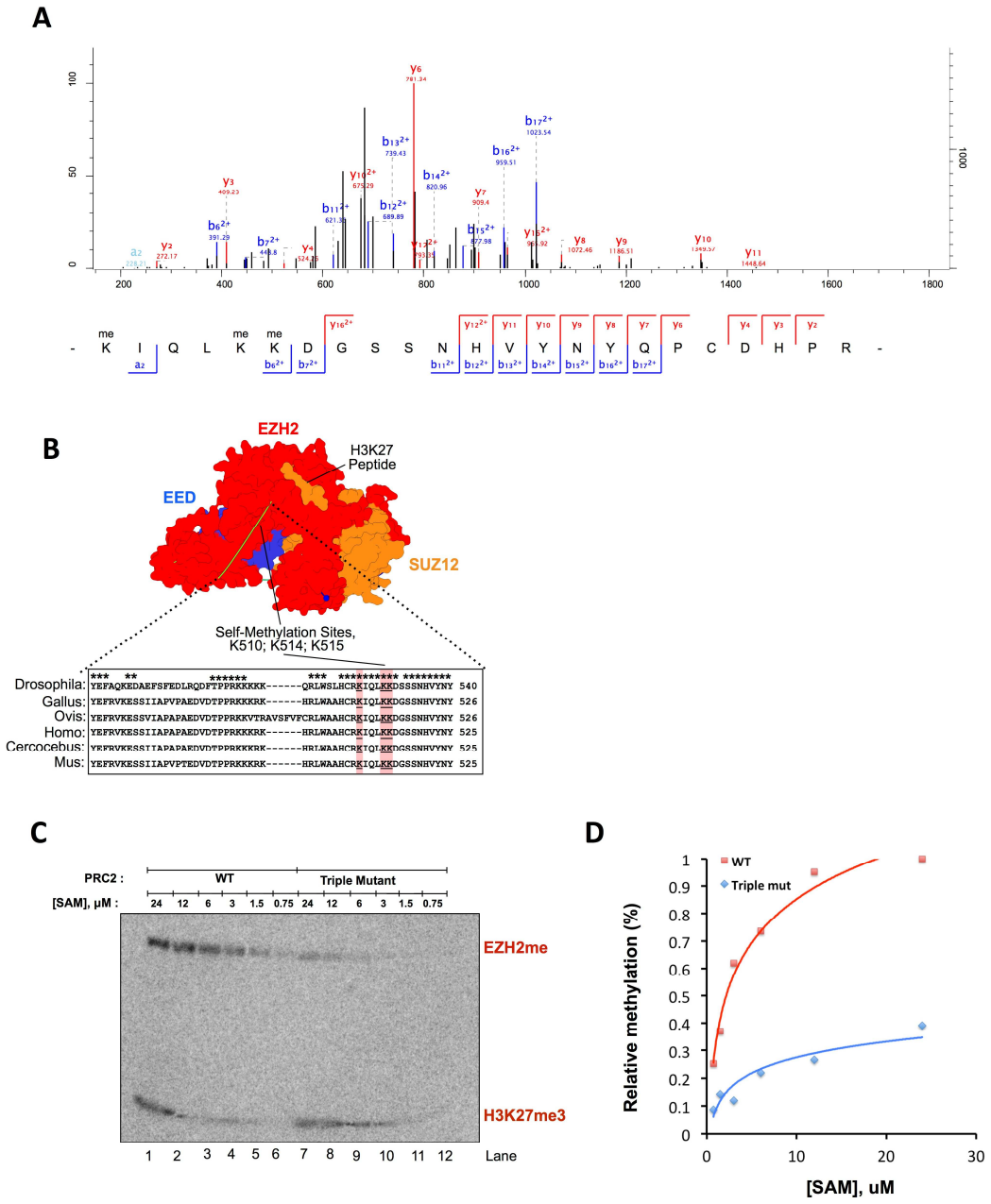


Figure C2. K510A, KK514/K515AA of EZH2 were the key residues methylated by PRC2 itself. A. mass spec analysis showing K510A, KK514/K515AA were residues methylated in HMTase assays. **B.** Three lysine (K510A, KK514/K515AA) were conserved shown in sequence alignments. **C.** HMTase assays showing that triple lysine mutant reduced EZH2 automethylation signal substantially. **D**. Quantifications of C.