TITLE

Full Title: Establishing RNA-RNA interactions remodels lncRNA structure and promotes PRC2 activity

Short title: RNA-RNA interactions regulate PRC2 activity switch

Authors

Maggie M. Balas[1,2,3], Erik W. Hartwick[2,3,4], Chloe Barrington[1,2,3], Justin T. Roberts[1,2], Stephen K. Wu[1,2], Ryan Bettcher[2], April M. Griffin[2], Jeffrey S. Kieft[1,2,3], and Aaron M. Johnson[1,2,3]*

Affiliations

1 Molecular Biology Program.

2 Department of Biochemistry and Molecular Genetics.

3 RNA Bioscience Initiative University of Colorado Denver Anschutz Medical Campus 12801 East 17th Ave., Aurora, CO, United States.

4 Current address, Columbia University Department of Chemistry, New York, NY, United States.

*Corresponding Author

Aaron M. Johnson, PhD

e-mail: Aaron.m.johnson@CUAnschutz.edu

Tel: +1(303)724-3224

ABSTRACT

Human Polycomb Repressive Complex 2 (PRC2) catalysis of histone H3 lysine 27 methylation at certain loci depends on long noncoding RNAs (lncRNAs). Yet, in apparent contradiction, RNA is a potent catalytic inhibitor of PRC2. Here we show that intermolecular RNA-RNA interactions between the lncRNA HOTAIR and its targets can relieve RNA inhibition of PRC2. RNA bridging is promoted by heterogenous nuclear ribonucleoprotein B1, which uses multiple protein domains to bind HOTAIR regions via multi-valent protein-RNA interactions. Chemical probing demonstrates that establishing RNA-RNA interactions changes HOTAIR structure. Genome-wide HOTAIR/PRC2 activity occurs at genes whose transcripts can make favorable RNA-RNA interactions with HOTAIR. We demonstrate that RNA-RNA matches of HOTAIR with target gene RNAs can relieve the inhibitory effect of a single lncRNA for PRC2 activity after B1 dissociation. Our work highlights an intrinsic switch that allows PRC2 activity in specific RNA contexts, which could explain how many lncRNAs work with PRC2.

INTRODUCTION

Chromatin regulation can depend on long noncoding RNA (lncRNA) transcripts(1,2), though in many cases the exact mechanism of action for the RNA is unclear. The most well-established example of lncRNA regulation of chromatin is via the Xist RNA that is required for silencing of one copy of the X chromosome in female mammals(3). Another more recent example of a lncRNA associated with chromatin-based silencing is the HOTAIR transcript(4). Both of these lncRNAs are associated with the activity of the histone methyltransferase complex Polycomb Repressive Complex 2 (PRC2), which is involved in facultative heterochromatin formation(5). The chromatin of the inactive X chromosome and of HOTAIR-repressed genes is marked in a lncRNA-dependent manner by histone H3 lysine 27 trimethylation, the product of PRC2. There is active debate in the field as to how these lncRNAs regulate PRC2(6). Recent evidence favors a model where Xist "tunes" PRC2 activity only after PRC2 recruitment, while recruitment itself is mediated by other mechanisms(7). Transcription is also reduced by Xist and HOTAIR independent of PRC2(8,9), consistent with low levels of transcription promoting PRC2 activity(10). PRC2 binds to single-stranded RNA (ssRNA) with a preference of G-tracts and G-quadruplexes(11). The prevalence of G-tracts, especially at the 5' end of genes(12), explains why PRC2 interacts with many pre-mRNAs(13), perhaps sampling nascent transcripts on chromatin(14) due to relatively fast on and off rates for RNA binding(15). When RNA binds to PRC2, the methyltransferase activity for nucleosomes is inhibited(16,17), suggesting that some specific additional context is required for PRC2 activity when RNAs such as Xist or HOTAIR are associated at chromatin that is methylated by PRC2.

Specificity of PRC2 activity is multi-faceted and differs depending on the organism and whether initiating or maintaining heterochromatin. In *Drosophila*, PRC2 is largely dependent on a small set of specific DNA binding proteins to recruit the complex for initiation of heterochromatin (*de novo* methylation)(5). Maintenance of H3K27me3 is facilitated by the ability of the EED subunit of PRC2 to recognize H3K27me3, recruiting and stimulating PRC2 at previously-established heterochromatin regions(5). This mechanism is thought to play a role in spreading and maintenance of methylation, though DNA binding

proteins are still essential in this maintenance mechanism as well(5). In humans, PRC2 recruitment can occur by multiple, sometimes overlapping, mechanisms including DNA binding proteins with lower specificity that work in combination with each other or other co-factors such as chromatin-associated long noncoding RNA(5). The complex nature of PRC2 recruitment in mammals, often necessitating multiple molecular mechanisms for action, has made it difficult to establish basic rules for PRC2 catalytic activity, especially in *de novo* initiation of H3K27me3-triggered heterochromatin.

The finding that PRC2 binds with significant affinity to nearly every naturally-occurring lncRNA and mRNA tested(13,16,17), which then inhibits nucleosome methylation, has called into question the specificity of lncRNAs that were suggested to work with PRC2 to silence chromatin. However, these findings did help establish that the RNA being produced by an RNA polymerase on chromatin, nascent RNA, would be an intrinsic inhibitor of PRC2 activity(16), preventing initiation of heterochromatin at active genes. In fact, tethering a high-affinity RNA substrate for PRC2 directly to chromatin in the nucleus actively antagonizes PRC2 activity at normal target genes(12), due to the higher affinity of PRC2 for single RNAs with G-tracts than for chromatin(15,18). It has been proposed that the binding of PRC2 to nascent RNA may allow the complex to sample the landscape of the genome, searching for a context where PRC2 activity is promoted(13,14,19). This may occur through encounter of H3K27me3, guiding PRC2 to pre-established heterochromatin and allowing PRC2 activity(20). For *de novo* heterochromatin formation, it is less clear how the nascent RNA-inhibited PRC2 is activated. One model suggests that transcription shut down is required for PRC2 activity(10). This model is supported by the demonstrated antagonism of PRC2 by histone modifications of highly-active genes and nascent RNA(16,21). While both Xist and HOTAIR can turn off transcription upstream of PRC2 activity(8,9), the association of both of these lncRNAs with chromatin promotes H3K27me3 at lncRNA target loci. Therefore, even if the gene is repressed and little nascent RNA is present, the lncRNA is at the chromatin locus. Regardless of how PRC2 is recruited, Xist and HOTAIR RNAs inhibit PRC2. Therefore, lncRNA-induced transcriptional repression does not resolve the issue that an inhibitory chromatin-associated RNA is present at the site of PRC2 activity.

RNA and chromatin compete for PRC2 binding(15,18). Specifically, the linker DNA between nucleosomes competes with RNA(15). At equimolar nucleotide concentrations, ssRNA wins the competition for PRC2 binding over chromatin(15), which explains the inhibitory effect of RNA for PRC2 activity. The affinity of PRC2 for a ssRNA is much higher than for a hairpin version of the same nucleotide content(11), suggesting a more favorable context for chromatin to compete. Within the PRC2 mechanism of sampling chromatin via nascent RNAs and associating with lncRNAs on chromatin, we hypothesize there may be a context for double-stranded RNA that may tip the balance from RNA to chromatin for productive PRC2 interaction.

Double-stranded RNA (dsRNA) occurs in the nucleus, resulting from multiple mechanisms including intra- and inter-molecular RNA-RNA interactions. LncRNAs can make intermolecular RNA-RNA interactions with different types of transcripts. For example, Xist pairs with its antisense RNA, TsiX, during X inactivation(22). We have previously shown that HOTAIR can interact directly with target gene RNA, such as JAM2, through an imperfect RNA-RNA base-pairing match sequence(23). This specific matching region within the sequence of HOTAIR is predicted to have a propensity for stable RNA-RNA interactions with known target transcripts(24). In fact, the region is a “hotspot” for RNA-RNA interactions as identified by computational prediction querying HOTAIR against the entire mRNA transcriptome(25). Interestingly, this “hotspot” overlaps a region within HOTAIR that was found to be conserved across vertebrates (except teleosts) with potential for RNA-RNA interaction with a HOXD transcript(26). HOTAIR RNA-RNA interaction was discovered due to its association with the RNA binding protein (RBP), heterogenous nuclear ribonucleoprotein (hnRNP) B1. Importantly, hnRNP A2/B1 was found to regulate HOTAIR-dependent PRC2 activity in cells. Furthermore, the B1 isoform was found to bind preferentially to HOTAIR and its target transcripts over A2(23). This protein has two tandem RNA recognition motif (RRM) domains that can associate in a head-to-tail dimer, binding two RNAs in an anti-parallel nature(27), suggesting a mechanism to promote RNA-RNA base-pairing interactions. We have previously shown that B1 can promote RNA-RNA interactions between HOTAIR and an RNA from a target gene, JAM2, which suggested that RNA-RNA interactions have a role in chromatin regulation by PRC2(23) (Fig. 1A), though

the underlying mechanism behind this has remained unclear. LncRNAs such as Xist and HOTAIR can adopt favorable structured states(28,29), presenting an additional challenge for any proposed intermolecular RNA matching where intramolecular interactions occur.

In the current study, we gain mechanistic insight into how B1-mediated RNA-RNA interactions can modulate HOTAIR structure and function to promote PRC2 activity. We determine the necessary features of hnRNP B1 for HOTAIR interaction and identify the specific regions of HOTAIR that interact with B1. We use chemical probing of HOTAIR secondary structure to highlight the structural changes that occur when B1 and an RNA-RNA match engage with HOTAIR. We find that genome-wide HOTAIR-dependent PRC2 activity occurs at loci whose transcripts make more-favorable RNA-RNA interactions with HOTAIR. Finally, we demonstrate that specific intermolecular RNA-RNA interaction relieves the inhibitory nature of HOTAIR RNA for PRC2 methyltransferase activity on nucleosomes. By dissecting these molecular changes to RNA structure, we highlight a switch that can result in PRC2 recruitment and activation by a lncRNA on chromatin. This may be a general mechanism that applies to many contexts where RNA plays a role in potentiating PRC2 activity.

RESULTS

RNA-RNA interaction is directly promoted by hnRNP B1 but not the A2 isoform

We have previously identified hnRNP A2/B1 as an important component of HOTAIR-dependent PRC2 activity in breast cancer. We found B1 was enriched preferentially in RNA pulldown assays with HOTAIR using nuclear extracts and we subsequently demonstrated direct *in vivo* binding of hnRNP B1 to HOTAIR, over the highly abundant isoform hnRNP A2. Additionally, B1 also bound preferentially to HOTAIR target transcripts, over A2(23,24). B1 differs from A2 by the inclusion of exon 2, which encodes 12 unique amino acids on the N-terminus (Fig. 1B). To further profile the recognition mechanism of HOTAIR by hnRNP B1 we first tested whether the isoform preference, B1 over A2, that was observed in the nuclear extract pulldown was recapitulated with purified protein. Using recombinant A2 or B1 proteins, expressed in *E. coli*, we performed in vitro HOTAIR pulldown assays and found that B1 binds preferentially to HOTAIR

compared to A2 (Fig. 1B). Even a three-fold increase in A2 concentration did not recapitulate the same level of binding as B1 to HOTAIR. Little to no binding was observed for B1 to equal amounts of the control non-coding RNA, of similar size and GC content, that corresponds to the anti-sense sequence of the luciferase mRNA (Anti-luc), which we have used previously as a control(23). We conclude from this that the B1-specific N-terminal domain (NTD), directly confers preferential binding to HOTAIR (Fig. 1B). The presumed position of the NTD, based on the N-terminus position in the A2 isoform crystal structure(27) would place the NTD in proximity to bound RNA (Fig. 1C). This suggests the B1 NTD itself directly interacts with RNA as an extension of the RRM, to increase specificity, affinity, or both. Next we evaluated B1 vs. A2 in the in vitro RNA-RNA interaction assay we have previously used to characterize matching with the HOTAIR target mRNA JAM2. Briefly, the RAT-tagged JAM2 match RNA fragment was incubated with full-length HOTAIR or Anti-luc control RNA in the presence or absence of hnRNP A2 or B1, and then affinity purified via the RAT tag. The association of HOTAIR or Anti-luc with JAM2 was quantified by RT-qPCR (Fig. 1D). Consistent with previous results, B1 was able to stimulate significant HOTAIR interaction with the target gene JAM2 RNA and we found that this was not the case for A2. Moreover, B1 was able to promote a minimal amount of RNA interaction between JAM2 and the Anti-luc control RNA, suggesting that B1 can bridge two RNA molecules indirectly while they remain separated, not base-paired, as seen in the crystal structure for A2 (Fig. 1C).

B1 bridges RNAs in the absence of strong base-pairing potential

The crystal structure of the A2 tandem RRMs bound to RNA revealed a head-to-tail dimer in complex with two RNAs(27). Each RNA crosses the RRM of one protomer to the next and the two RNAs are aligned in an antiparallel nature (Fig. 1C). Based off of this RNA arrangement, where the dimer binds two molecules through single-strand engagement by the RRMs, we asked whether the ability of B1 to promote interaction of HOTAIR with its targets is dependent on base-pairing between the RNAs. We used the HOTAIR RNA-RNA interaction assay described above to test this. Matching of HOTAIR with its target does not require B1 in vitro, as we have previously shown(23). In fact, addition of B1 in our previously published RNA matching assays only modestly promoted the interaction between the RNAs. To test whether B1 can

promote interaction with RNAs that do not have strong complementarity, we reduced the concentration of the RNAs to promote more stringency and B1 dependence. Under these conditions, B1 stimulates HOTAIR interaction with the JAM2 target RNA roughly three-fold (Fig. 1F), compared to no significant background association (Fig. S1A). When the fairly extensive imperfect base-pairing interaction between JAM2 and HOTAIR is disrupted by deleting or mutating (changing to the complement base) the 64 nt interaction region on HOTAIR, B1 is able to recover nearly the same level of RNA-RNA interaction as with wildtype HOTAIR (Fig. 1F). This result suggests that B1 can bridge two RNAs due to the strength of binding to those individual RNAs and the ability of the dimer to interact with each RNA at the same time. Complementarity of two RNAs at, or proximal to, the B1 bridge would subsequently promote intermolecular base-pairing. The HOTAIR-JAM2 match is stable in the absence of B1, suggesting that, once formed, it would persist after B1 dissociation.

Single-nucleotide mapping of B1 UV crosslinks to HOTAIR identifies major direct interactions

We used the eCLIP (enhanced Cross-Linking and Immunoprecipitation) method with recombinant B1 and in vitro transcribed HOTAIR (in vitro eCLIP) to profile the interactions between the two(24). Briefly, UV crosslinking was performed on pre-incubated B1 and HOTAIR, the sample was digested with a low amount of RNase to generate RNA fragments that could be reverse transcribed and made into a sequencing library. The eCLIP protocol involves size selection and downstream ligation of one sequencing adapter to the 3' end of the cDNA as part of sequencing library preparation (Fig. 2A, S1C). Because a base that remains cross-linked to an amino acid "scar" can prematurely terminate reverse transcription, we mapped the termination sites of the in vitro eCLIP sequencing to better refine the specific site of direct B1 interaction on HOTAIR. The in vitro eCLIP results of B1 binding to HOTAIR revealed multiple regions with high reverse transcription (RT) stops. We observed six main peaks of RT stops from five locations on HOTAIR in the size of ~25-100 nucleotides (Fig. 2B). Four of these peaks fell within domain 1 of HOTAIR (nucleotides 1-530), suggesting HOTAIR domain 1 is important for B1-mediated function (Fig. 2B). Controls with non-crosslinked protein and a crosslinked non-binding protein yielded background levels of recovered RNA with poor mapping capability (Fig. S1D) and RT of HOTAIR alone demonstrated

that the B1 RT stops are specific (Fig. S1E). We conclude from this that B1 binds to multiple distinct locations within domain 1 of HOTAIR. Based on other lower-frequency RT stops that still produce a peak, we suspect additional secondary interactions are made, potentially facilitated by proximity of RNAs in tertiary structure near the primary interaction sites.

B1 C-terminal glycine-rich domain participates in HOTAIR engagement

The crystal structure of the A2 RRMs demonstrates a minimal complex for bridging of RNAs (Fig. 1C). We have also demonstrated that the B1 N-terminal domain is required for efficient HOTAIR engagement and promotion of RNA-RNA interactions (Fig. 1B). Though these interactions are important features in establishment of RNA-RNA interactions, we wished to also test other features of B1 that may participate in this activity. In addition to the NTD and tandem RRMs, B1 has a run of “RGG” motifs proximal to the second RRM domain, as well as a low-complexity glycine-rich C-terminal domain. We generated minimal RRM constructs for B1 and A2, an alanine substitution of all five RGGs in the full-length construct, and a construct with the C-terminal glycine-rich domain deleted. Equal amounts of protein loading and HOTAIR recovery is demonstrated by Coomassie gel and qPCR quantification. RNA pulldown experiments demonstrated a clear requirement for the C-terminal portion of B1, however the RGG motifs were unnecessary, suggesting the unstructured glycine-rich domain is required for tight binding to HOTAIR (Fig. 2C). The additional amino acids within glycine-rich domains have been proposed to influence the interplay of this region with other RBPs or RNA (30,31).

Chemical probing highlights B1 interaction with HOTAIR

To further investigate mechanistically how RNA-RNA interactions are facilitated, we performed chemical probing experiments on HOTAIR domain 1 with the JAM2 RNA match (62 nt) and/or B1 (Fig. 3A). RNA was chemically modified using 1-methyl-7-nitroisatoic anhydride (1M7), which reacts with the 2'-hydroxyl of the RNA backbone to form adducts on accessible or flexible nucleotides, primarily in single-stranded regions of the RNA (Fig. 3A). Adduct formation was quantified using primer extension via reverse transcription, where RT stops equate to reactivity to the modifier. Due to the strong RT stop introduced

when HOTAIR and JAM2 interact, we had to include JAM2 removal steps to the protocol to generate complete reactivity data (Supplementary Fig. S2). To analyze the capillary electrophoresis data, HiTRACE RiboKit (32-36) was used and subsequently normalized using published methods(37). The normalized reactivity values were evaluated for each experimental condition: HOTAIR only, HOTAIR + JAM2, HOTAIR + B1, HOTAIR + B1 + JAM2; and for control conditions HOTAIR + Poly A and HOTAIR + BSA (Fig. 3A,B, Supplementary Fig. S4). We observed reproducible chemical reactivity patterns across all conditions (see error trace plotted in Supplementary Fig. S3 and Supporting File SF1). We were able to detect subtle and larger changes in regions of HOTAIR upon the addition of JAM2 RNA and/or B1, consistent with our previous data and predictions for establishment of RNA-RNA interactions (Fig. 3B,C). Interestingly, we found the addition of B1 reduced chemical probing reactivity in many regions of HOTAIR (Fig. 3B,C). This decreased reactivity was highlighted at the B1-interaction sites identified from our in vitro eCLIP analysis (highlighted in blue, 141-172 nt, 304-314 nt, 460-523 nt) (Fig. 3D,E). These results support a model of multiple prominent B1 interactions with HOTAIR and indicate that the eCLIP and chemical probing data are in agreement. The extent of reduced reactivity we observed may be explained by either 1) a broader direct influence of the protein on reactivity, perhaps through proximity in three-dimensional space, or 2) a significant change to RNA structure induced by B1, perhaps increased secondary structure, leading to decreased 1M7 reactivity in regions of HOTAIR that B1 does not bind directly.

Establishment of RNA-RNA interactions alters HOTAIR structure

We next generated a difference map of chemical probing conditions, to quantify reactivity changes compared to HOTAIR RNA alone (Fig. 4A). To highlight regions of change, we employed a sliding-window analysis, trained using control conditions (see Supporting File SF4 and Materials and Methods), to identify regions of change above a control threshold value. We mapped these regions of change, according to each experimental condition, onto the previously determined secondary structure model of HOTAIR domain 1(28) (Fig. 4B-E). Analysis of HOTAIR + JAM2 demonstrated that JAM2 pairing with HOTAIR changes the 1M7 reactivity of HOTAIR in the regions of the secondary structure that surround the JAM2 base-pairing site (nucleotides 245-306) (Fig. 4B). This includes reduced reactivity across the base-pairing

site, consistent with the strong complementarity of the intermolecular RNA base-pairing match. As noted above, addition of B1 caused significant changes in reactivity to a majority of HOTAIR nucleotides, as evidenced by the many highlighted regions on the secondary structure (Fig. 4C). Of note, although the general trend of nucleotide reactivity was downward, B1 caused higher reactivity at specific regions of HOTAIR, suggesting these regions are more exposed and single-stranded.

The combination of both JAM2 and B1 also generated regions of altered reactivity for HOTAIR RNA, especially surrounding the JAM2 interaction site (Fig. 4D,E). Since B1 alone changes reactivity for so many nucleotides of HOTAIR, we subtracted those changes from the JAM2/B1 changes to mask these (Supplementary Fig. S5). By doing this, we can see that the JAM2/B1 condition is significantly different from B1 alone, despite the large extent of changes with B1 alone. The profile of this B1-subtracted data do not resemble the JAM2 condition. This suggests that JAM2 is able to either synergize with, or counteract, the effects of B1 alone. This is further emphasized by subtracting both JAM2 and B1 alone conditions from the JAM2/B1 reactivity (Fig. S5, bottom). This double-subtraction resulted in significant changes that persist, which cannot be accounted for by additive effects of the individual conditions. Altogether, the chemical probing data clearly show that the steps in establishing RNA-RNA interactions do remodel HOTAIR structure and suggest that the B1 and the intermolecular RNA base-pairing have individual and potentially synergistic effects on changing HOTAIR structure.

Conversion from ssRNA to dsRNA promotes PRC2 activity

It is not clear how PRC2 goes from RNA-mediated inhibition to an enzymatically active state in situations where a lncRNA like Xist or HOTAIR are known to be present and no pre-existing H3K27 methylation has occurred. PRC2 has been shown to have a strong affinity for ssRNA, but has a weaker affinity for duplex RNA, like that found in perfectly base-paired stem-loop structures(11). We analyzed RNA-RNA interaction predictions between HOTAIR and the entire transcriptome(24,25) and compared them to previously identified HOTAIR-dependent PRC2 targets in a model of HOTAIR-overexpressing triple negative breast

cancer cells(23,38). Genes that acquired new PRC2 activity when HOTAIR was overexpressed were biased towards those with transcripts that have more-favorable RNA-RNA interaction propensity with HOTAIR (Fig. 5A). This led us to hypothesize that duplex RNA might provide the correct context for PRC2 enzymatic activity by allowing PRC2 to transfer to chromatin via lower affinity binding of dsRNA. To determine this, we performed histone methyltransferase (HMTase) assays with single-stranded or double-stranded HOTAIR RNA and evaluated levels of H3K27me3. We first assessed optimal H3K27me3 activity in these assays using di-nucleosome templates composed of two 601 sequences surrounding 40-bp of linker DNA (Fig. 5B), recombinant PRC2 complex (Fig. 5C) and the co-factor JARID2 (Fig. 5D). We next annealed HOTAIR RNA with a titration of reverse complement RNA to form perfect duplex RNA and introduced these into the HMTase assays (Fig. 5E). We observed a distinct reduction in H3K27me3 in the presence of ssRNA vs no RNA. As we increased the amount of reverse complement RNA, H3K27me3 levels increased in a manner concurrent with formation of dsRNA (Fig. 5E). These results suggest that while a ssRNA inhibits PRC2 activity, formation of a duplex between the inhibitory RNA and its match is able to relieve this inhibition and promote PRC2-mediated H3K27me3.

HOTAIR target RNA-RNA matches promote PRC2 activity

Fully-duplex dsRNA was used above, which may be relevant for some lncRNA contexts, but HOTAIR makes imperfect matches with RNA targets. To begin to address if HOTAIR imperfect RNA-RNA matches with target mRNA could also promote PRC2 catalytic activity, we evaluated the HOTAIR matches with endogenous targets JAM2 and HOXD10 (Fig. 6A). Consistent with previous results, HOTAIR alone was able to inhibit PRC2 activity. When increasing ratios of HOTAIR-JAM2 duplex were present, the match relieved the inhibitory effect of the single-stranded HOTAIR fragment alone, thereby stimulating H3K27me3 (Fig. 6B). Equal molar ratios of HOTAIR to JAM2 resulted in the highest levels of H3K27me3 promotion (Fig. 6C). Similar results were observed with the HOTAIR-HOXD10 duplex RNA match (Fig. 6D), with equal molar ratios of match RNA relieving PRC2 inhibition the most (Fig. 6E). Titration of JAM2 or HOXD RNA alone had no effect on PRC2 (Supplementary Fig. S6A-C), highlighting that the duplex

formed with HOTAIR RNA is the effector. In contrast, a control PolyA RNA of similar size did not significantly relieve inhibition (Fig. 6F, Supplementary Fig. S6D).

The experiments above used RNA fragments surrounding the matching regions. These fragments were similar in size to the model RNA substrates used in previous biophysical studies of RNA-PRC2 interaction(11). Next we tested a much larger portion of HOTAIR RNA, using the HOTAIR domain 1 sequence (nt 1-530), to evaluate if the JAM2 match was sufficient to relieve inhibition of PRC2 activity. We found that HOTAIR domain 1 was a more-potent inhibitor of PRC2 by molar ratio, consistent with longer RNAs binding with higher affinity to PRC2. Strikingly, JAM2 was able to relieve PRC2 inhibition by this longer RNA (Fig. 6G,H). However, when we substitute HOTAIR domain 1 with a version in which the JAM2 interaction site is deleted, we observe no relief of inhibition (Fig. 6I, S6E). Together, these results demonstrate that the formation of dsRNA via HOTAIR matches with known genomic RNA targets like JAM2 and HOXD10, are sufficient to promote H3K27me3 in vitro. This suggests that endogenous RNA-RNA interactions between lncRNA and target nascent transcripts could facilitate PRC2 catalytic activity for gene silencing at specific genomic locations (Fig. 5A).

Finally, we addressed the contribution of B1 to PRC2 activity that is modulated by RNA-RNA interactions. Under identical conditions to HOTAIR Domain 1 and a JAM2 titration (Fig. 6D), we found that addition of B1 did not allow JAM2 to relieve PRC2 inhibition (Fig. 6J,K). This may be due to how B1 normally functions in a step-wise RNA-RNA interaction mechanism, where the protein facilitates RNA bridges, but must dissociate before PRC2 inhibition can be relieved. We note that B1 does increase HOTAIR reactivity in multiple regions, which persist when JAM2 matches to HOTAIR (Fig. 3,4). These exposed regions of HOTAIR that are induced by B1 may contribute to PRC2 inhibition until B1 dissociates, leaving the matched RNA state that facilitates relief of PRC2 inhibition by the lncRNA.

DISCUSSION

RNA binding to PRC2 inhibits its catalytic activity(16,17). It has remained unclear how this inhibition is relieved in contexts where RNA is present in a region of chromatin that has no previously-deposited H3K27 methylation, including lncRNA-associated loci and genes bearing nascent transcripts. Based on our previous observation that the HOTAIR lncRNA makes preferential interactions with hnRNP B1(23), a multi-valent RNA binding protein that promotes RNA-RNA interactions, we have further profiled the molecular basis of this mechanism and how it relates to the observation that HOTAIR can somehow promote PRC2 activity when overexpressed in cancers(38). We find that hnRNP B1 uses multiple domains to engage HOTAIR in a manner that can bridge it to a target gene RNA (Fig. 1-2). When HOTAIR matches with a target RNA, the lncRNA structure is remodeled by B1 and the matching RNA (Fig. 3-4). We find that HOTAIR-mediated PRC2 targets make more favorable RNA-RNA interaction with HOTAIR (Fig. 5A). In turn, the formation of duplex RNA between HOTAIR and its targets limits the ability of the lncRNA to inhibit PRC2 activity (Fig. 6). When B1 is still present, RNA matching is not sufficient to relieve PRC2 inhibition, suggesting B1 must dissociate from the RNA for PRC2 activity to be promoted. PRC2 binds to many individual RNAs in the nucleus(5), including nascent transcripts, presumably in an inactive state. Our results suggest a model where an RNA can be a positive effector of *de novo* PRC2 activity in a context where RNA-RNA interactions relieve PRC2 inhibition on chromatin (Fig. 7).

HOTAIR RNA engagement and RNA-RNA bridging by hnRNP B1

The crystal structure of the tandem RRMs of hnRNP A2/B1(27) demonstrates a potential for two RNAs to be engaged by an A2 or B1 dimer in an anti-parallel orientation. While the RRMs likely only bind single-stranded RNA, the adjacent RNA sequences are in a favorable orientation for base-pairing. This engagement is inter-molecular in the crystal structure and likely explains the ability of B1 to bring two RNAs together, even when the base-pairing potential between them is limited (Fig. 1F). There is potential for the same mode of engagement to work intramolecularly, as well, and this may underlie the multiple sites of direct interaction we observe for B1 on HOTAIR (Fig. 2B) and the ability of the protein to reduce chemical reactivity of multiple regions of HOTAIR (Fig. 3). A possible explanation for why the B1 N-terminal domain promotes HOTAIR binding is because it promotes B1 dimerization. Further work is

required to determine if this is the case. The C-terminal domain of B1 is necessary for HOTAIR binding and thus promotion of RNA-RNA interactions (Fig. 2C). The C-terminus of the related hnRNP A1 can bind RNA(30,31) and includes an intrinsically-disordered domain that has been shown to self-associate and phase separate at high concentrations(39). Whether all of these properties are important for the mechanism we have characterized remains to be determined; however, the ability to self-associate, at least into a dimeric state, would potentially promote the RRMs of two monomers forming the head-to-tail conformation that can promote RNA-RNA interactions (Fig. 2E).

We find that B1 cannot be present with the matched RNAs for relief of PRC2 inhibition (Fig. 6 J,K). We interpret this result by placing the function of B1 upstream in an RNA-RNA interaction mechanism model that ultimately leads to catalysis of H3K27 tri-methylation and heterochromatin formation (Fig. 7). In this model B1 must dissociate before PRC2 activity is promoted. The requirement for a matchmaker protein to "get out of the way" has been proposed in other molecular matchmaker models(40) and is consistent with the match itself being the ultimate effector, rather than the protein.

RNA-RNA interactions promote PRC2 activity

The ability of an RNA with base-pairing potential to relieve the inhibition of PRC2 that is imposed by a single RNA binding event may apply beyond the HOTAIR mechanism. There are multiple examples of lncRNAs with intermolecular RNA-RNA interaction capability involved in PRC2 activity(2). For example, some imprinted loci such as *Kcnq1* have antisense transcripts that induce PRC2 activity and repression occurring coincident with sense transcripts, present in an RNA "cloud" at the locus that is methylated by PRC2(41). Xist also has a perfect complement antisense transcript, TsiX(41). In mice, Tsix transcription promotes PRC2 activity at the Xist promoter, coincident with the formation of Xist:TsiX double-stranded RNA, to repress Xist expression on the active X chromosome(22,42). In fact, this Xist:TsiX dsRNA does not inhibit PRC2(17). These RNA matching capabilities may underlie how PRC2 can methylate chromatin in a cloud of a lncRNA that would otherwise repress methyltransferase activity.

LncRNA matching may occur with the nascent transcript at the locus where PRC2 activity is promoted. Although nascent transcription from highly active genes is likely to "win out" by inhibiting PRC2(13,16), there are multiple pieces of evidence suggesting that PRC2 is active in the presence of nascent transcripts at lowly-expressed genes. Xist and Kcnq1ot1 lncRNAs are both present with low levels of their antisense transcript (TsiX and the protein-coding gene Kcnq1, respectively) while PRC2 methylates the chromatin at these loci(22,43). More-generally, H3K27me3 deposition has been shown to occur in genes with moderate transcription activity(44-46). Additionally, building up paused RNA Polymerase II with the drug DRB leads to more accumulation of promoter H3K27me3 than does complete inhibition of transcription initiation(10). Interestingly, a recent study found that endonucleolytic cleavage of nascent transcripts by a Polycomb-associated enzyme complex is important for maintaining low expression of Polycomb-repressed genes, suggesting nascent transcription persists even after PRC2 activity occurs(47). These results suggest that PRC2 can deal with the inhibitory effects of a nascent transcript to achieve H3K27 methylation. Intermolecular RNA-RNA interactions are one potential mechanism to achieve this. In addition, recent work has highlighted that once H3K27me3 has been established, this modification can help further relieve inhibition of methyltransferase activity by RNA(20). Where the mechanisms mentioned above do not act, other mechanisms must exist to prevent nascent RNAs from inhibiting PRC2, such as additional RNA binding protein interactions with nascent RNA to mask it. Our findings fit into a model where the inhibitory effects of RNA on PRC2 catalytic activity can be overcome by specific intermolecular RNA-RNA interactions to promote de novo H3K27 methylation. This mechanism may operate with multiple lncRNA pathways in normal contexts and when aberrantly high lncRNA expression occurs in disease, such as with HOTAIR, that may drive improper PRC2 activity.

ACKNOWLEDGMENTS

We thank Kaushik Ragunathan, Srinivas Ramachandran, and members of the Johnson Lab for helpful comments on the manuscript. We thank Chen Davidovich and Tom Cech for PRC2 baculovirus constructs.

Funding:

This work was supported by NIH grants R35GM119575 (A.M.J.), R35GM118070 (J.S.K.), and T32GM008730 and F31CA247343 (J.T.R.); RNA Bioscience Initiative Scholar Awards (M.M.B. and E.W.H); and an RNA Bioscience Academic Enrichment Award (C.B.) from the University of Colorado School of Medicine RNA Bioscience Initiative. We thank the University of Colorado Cancer Center Genomics Core and the Protein Production, Monoclonal Antibody, and Tissue Culture Shared Resource, both supported by NIH grant P30-CA46934, for technical support.

Author Contributions:

M.M.B and A.M.J conceived the study. M.M.B, E.W.H, C.B, S.K.W performed experiments. M.M.B, E.H.W, and J.T.R performed bioinformatic analysis. R.B. and A.M.G. designed and purified protein constructs and provided experimental advice. M.M.B, E.W.H, J.S.K, and A.M.J wrote the manuscript. All authors reviewed and approved of the manuscript.

Competing Interests: The authors declare that they have no competing interests.

Data and Materials availability: All data related to the manuscript are included in main and supplemental materials. Specific materials generated during this study are available upon request.

MATERIALS AND METHODS

In vitro transcription of RNAs

In vitro transcription (IVT) of RNA was performed using the MEGAScript T7 kit (Thermo or Ambion), reactions were incubated for 4 hours at 37°C. RNA was treated with Turbo DNase for 15 minutes, and RNA was purified with the RNeasy kit (QIAGEN) 6% Urea polyacrylamide gel or and visualized by agarose bleach gel.

In vitro RNA pulldown experiments

15 nM of IVT 10X MS2 tagged RNA (Full-length HOTAIR or Anti-Luc) was rotated (end-over-end) at room temperature for 15 minutes with 80 nM of recombinant hnRNP B1 or A2 in EMB 300 Buffer (10 mM Hepes pH 7.9, 300 mM NaCl, 3 mM MgCl2, 0.5% NP-40, 10% Glycerol, 0.1 mM PMSF, 0.5 DTT), RNase inhibitor (NEB) and 20 ug of competitor yeast tRNA (Roche) in a total volume of 300 µL per sample. At the same

time, 300 nM MS2-MBP was prebound to 20 µL of amylose resin (NEB) in EMB 300 Buffer, RNase inhibitor (NEB) and 1% BSA, and rotated at room temperature for 15 minutes. The MS2-MBP amylose resin was then added to each IVT-hnRNP sample and incubated an additional 15 minutes rotating at room temperature. Resin was washed 4X in 800 µL EMB 300 Buffer and then protein association was analyzed by Western blot using antibody for hnRNP A2/B1 (Abcam #ab6102). Additionally, 10% of each sample was used for RNA analysis, where RNA was isolated by phenol/chloroform extraction, purified by ethanol precipitation, and quantified by RT-qPCR.

Purification of PRC2 complex using Baculovirus expression system

Human PRC2 was purified essentially as described(13), using individual pFastBac1 constructs of EZH2, SUZ12, EED, RBBP4, and AEBP2 with HMBP-PrS (6XHis, MBP, and Prescission Protease sequences) N-terminal tags (courtesy of C. Davidovich and T. Cech). Briefly, equal MOI of each viral construct was used to infect Sf9 cells for 72 hours at 27ºC. Cells were harvested in PBS and snap frozen. Cells were thawed and resuspended in Lysis Buffer (10 mM Tris-HCl pH 7.5, 250 mM NaCl, 0.5% NP-40, 1 mM TCEP, 1X Complete Protease Inhibitor (Roche)), 20 mL per 500 mL culture, and slowly rocked for 30 minutes at 4ºC. All further steps done at 4ºC, but never on ice. Lysate was clarified at 29,000 RCF for 30 minutes. Supernatant was incubated with 0.75 mL amylose resin (NEB) for 2 hours. Sample was poured into a column and resin was washed with 8 mL Lysis Buffer, 12 mL Wash Buffer 1 (10 mM Tris-HCl pH 7.5, 500 mM NaCl, 1 mM TCEP), 12 mL Wash Buffer 2 (10 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM TCEP). Protein was eluted with Wash Buffer 2 + 10 mM maltose. 0.5 mL fractions were collected until minimal protein was detected. Protein was concentrated ~10-fold. Sample was incubated with PreScission Protease (GE Healthcare) overnight. Full PRC2 complex was separated by size-exclusion chromatography on a 24 mL Superose 6 column (GE Healthcare) in 10 mM Tris-HCl pH 7.5, 250 mM NaCl, 1 mM TCEP. Five-subunit complex containing fractions were concentrated, glycerol was added to 10%, and sample was aliquoted and flash-frozen.

Recombinant protein purification in E. coli

Human hnRNP B1 (~37.4 kDa), A2 (~36 kDa), and B1 truncations were performed as previously described(23). Human GST-JARID2 (119-574) was expressed in BL21 (DE3) CodonPlus E. coli cells overnight at 18ºC. Cells were lysed on ice using 2 mg/mL lysozyme and sonication in PBS (500 mM NaCl total), 1 mM EDTA, 1 mM EGTA, 1 mM PMSF, 0.5% Triton-X 100, 15 mM DTT. Sample was spun at 29,000 RCF at 4ºC and supernatant was incubated with rocking with Pierce Glutathione Agarose Resin (Thermo Fisher) for 2-3 hours. Resin was washed with at least 20 times volume of PBS (350 mM NaCl total), 1 mM EDTA, 1 mM EGTA, 0.1 mM PMSF, 0.5% Triton-X 100, 0.1 mM DTT. Protein was eluted in 50 mM Tris-HCl pH 8.0, 350 mM NaCl, 1 mM DTT, 10 mM glutathione. Sample was dialyzed in 50 mM Tris-HCl pH 8.0, 350 mM NaCl, 1 mM DTT, 5% glycerol, aliquoted and flash frozen.

In vitro RNA-RNA interaction assays

Performed as previously described(23) using 125 nM RAT-tagged JAM2 RNA fragment (IVT from gBlock, IDT) incubated with 5 nM IVT versions of full-length HOTAIR, HOTAIR with the JAM2 site deleted, HOTAIR with the JAM2 site mutated (to its complement base), or control RNA Anti-luciferase (anti-sense to the luciferase mRNA) at equal amounts by nanogram (UV quantification). Protein additions included 50 nM recombinant hnRNP B1, hnRNP A2 and 500 nM PP7-protein A fusion protein purified from E. coli. The PP7 fusion protein is required to pulldown RAT-tagged RNA with IgG conjugated Dynabeads, with minimal RNA recovery of both tagged and untagged RNA in the absence of the PP7 fusion protein (Supplementary Fig. S1A) with B1 binding contingent on JAM2 (Supplementary Fig. S1B).

RNA quantification by qRT-PCR

Reverse transcription was performed using the cDNA High Capacity Kit (Life Technologies). Standard curves using IVT JAM2, HOTAIR or Anti-luciferase were used to calculate amounts of RNA recovered. RNA inputs for experiments and standard curve samples were tested for integrity by agarose or acrylamide gel. RT-qPCR was performed using Sybr Green master mix (Takyon), with two qPCR replicates performed for each sample. Technical replicates were averaged prior to analysis of biological replicates.

In vitro *eCLIP-seq*

In vitro eCLIP-seq was performed as previously described(24). 1.78 pmol HOTAIR RNA and 8.9 or 17.8 pmol recombinant hnRNP B1 (see (23) for sequence details) were incubated in in 100 µL RNA refolding buffer (20 mM HEPES-KOH pH 7.9, 100 mM KCl, 3 mM $MgCl_2$, 0.2 EDTA pH 8.0. 20% Glycerol, 0.5 mM PMSF, 0.5 DTT) for 20 minutes at room temperature. The mixture was diluted to 250 µL in refolding buffer and UV-crosslinked twice in one well of a 24-well plate at 250 mJ and 254 nm wavelength, with mixing by pipette in between. B1-RNA crosslinked samples were treated with 0.1 ng RNase A for 4 minutes at 37ºC and 1200 rpm mixing, then stopped with 200 U Murine RNase Inhibitor (NEB). Following this, the in vitro samples were subjected to end repair, adaptor ligation, SDS-PAGE and transfer to nitrocellulose, and the remainder of the eCLIP-seq protocol, then sequenced multiplexed with other eCLIP-seq libraries as previously described(24). PCR products from different cycle number were analyzed by gel to avoid over-amplification (Supplementary Figure S1C). Control samples HOTAIR-B1 non-crosslinked, B1-RRMs crosslinked, and HOTAIR only (without gel and transfer steps) were included (Supplementary Figure S1D,E).

Chemical probing and analysis

Chemical probing procedure was similar to(28,36). Specifically, 20 pmol in vitro transcribed and purified HOTAIR RNA was incubated in 500 µL reactions with equimolar amounts of JAM2 RNA, hnRNP B1 or both in RNA refolding buffer (50 mM HEPES pH 7.4, 200 mM KCl, 5 mM $MgCl_2$, 0.1 mM EDTA) at room temperature for 20 minutes, then divided into two tubes, each containing 245 µL. The reaction was started by addition of 544 nmol 1M7 (+) (1-methyl-7-nitroisatoic anhydride), or of an equal amount of pure DMSO as a control (–). Samples were incubated for 5 min at 37°C and reactions were quenched with 5 µL 0.5M MES-NaOH. Samples with the JAM2 RNA underwent a JAM2 removal step where they were incubated with a JAM2 DNA complement oligo, heated at 94°C for 3 minutes to denature RNA and slow cool at room

temperature for 10 minutes. All chemically modified RNA was purified using Trizol extraction + isopropanol precipitation and reverse transcribed using 0.2 μM RNA with SuperScript III reverse transcriptase (Thermo) at 48°C for 1 hour using a fluorescently labeled primer (IDT): 5'-/5-6FAM/. Labeled DNA products were eluted in HiDi formamide spiked Gene Scan ROX 1000 size standard (Thermo). Samples were run on an Applied Biosystems 3500 XL capillary electrophoresis system and the data was analyzed using HiTRACE RiboKit(32-36) with MatLab (Math Works).

The HiTRACE normalized data, error, and replicate experiments ($n \geq 3$) are contained in Supporting File SF1. The HiTRACE normalized data for each condition were subsequently normalized using published methods to better compare each condition on the same scale to prevent extremely reactive positions from dominating the data analysis(37). Specifically, outlier positions within the data were temporarily excluded from the data and the remaining top 10% of the data were used to calculate a mean reactivity value used to normalize the data. The highly reactive data points were included in the data but represent extremely reactive positions. The inter-quartile range for each experiment was then multiplied by a constant of 6 (empirically chosen) where 6 times the IQR is considered an extremely reactive position. The error generated from the original HiTRACE analysis was carried through this normalization by converting the error from the HiTRACE analysis to percent error. The percent error for each nucleotide position was then multiplied by the normalized value reproducing the error for all positions in each experiment. The normalized values and corresponding error used for further analysis can be found in Supporting File SF2. Difference mapping of the data was completed by subtracting the HOTAIR only mean normalized data from the experimental mean normalized data. The mean-normalized and difference data are represented in heatmap form using Morpheus (https://software.broadinstitute.org/morpheus/) and the raw values are reported in Supporting File SF3.

To determine a region of change, the absolute value of the difference data was smoothed by using a sliding-window-mean for every 3 nucleotide positions along the region of interest. A threshold value was

551 determined by calculating the mean of the smoothed data from normalized PolyA control data minus the
552 HOTAIR only data. The absolute-mean value and standard deviation were calculated to be 0.067 and
553 0.049, respectively. A threshold value of 0.17 would represent the mean reactivity value plus two times
554 the standard deviation and was subsequently applied to the HOTAIR + JAM2 experiments. A threshold
555 value of 0.32 was generated using the same procedure for the BSA control experiments (mean = 0.12,
556 std = 0.1) and applied to HOTAIR + B1 and HOTAIR + B1 +JAM2 experiments. The regions of change
557 were defined as 5 consecutive nucleotides (JAM2 alone) or 4 consecutive nucleotides (B1 and JAM2/B1)
558 above the threshold value where two consecutive flanking positions had values below the threshold value.
559 The regions of change are located in Supporting File SF4. The identified regions of change were then
560 displayed back onto the secondary structural model for HOTAIR (Fig. 4B-E). All mathematical treatments
561 for the normalized data, data subtractions for difference data, and defining the regions of change were
562 performed using python in Jupyter Notebook. Graphs for the correlation plots of HOTAIR with the control
563 data (BSA and PolyA) and experimental samples (JAM2, B1, and JAM2 + B1) in Supplementary Fig. S4
564 and graphs for the normalized reactivity with error shading (Fig. S3) were generated using python in
565 Jupyter Notebook.
566
567 Replicate information for each experiment: HOTAIR only (4 replicates); HOTAIR with PolyA control (4
568 replicates); HOTAIR with BSA control (3 replicates); HOTAIR with JAM2 experiment (5 replicates);
569 HOTAIR with hnRNP B1 experiment (3 replicates); HOTAIR with JAM2 and B1 experiment (4
570 replicates).
571
572
573 ***RNA-RNA interaction analysis of genome-wide PRC2 targets***
574 Using a previously published database of computationally predicted interactions between human lncRNAs
575 and the entire transcriptome(25), a list of 40,740 annotated mature transcripts and computational
576 predictions for RNA-RNA interaction with HOTAIR were analyzed. This list was sorted by the predicted
577 minimum free energy found among the interactions contained within each pair of RNA sequences. We

compared the HOTAIR RNA-RNA interaction list with previously published HOTAIR-dependent PRC2 targets in MDA-MB-231 breast cancer cells identified by ChIP-seq(23,38), a combined list of 885 genes from multiple groups. Histograms displayed as a fraction of the total identified for each list were plotted together relative to predicted minimum free energy and a t test was performed comparing the two distributions.

Nucleosome reconstitution

Di-nucleosomes were assembled using salt dialysis, as previously described(48). To generate a DNA template for chromatin reconstitution, a 343-bp PCR product consisting of 2x 601 positioning sequences separated by 40 bp of linker sequence was cloned into pUC57 vector backbone. PCR product was purified using Nucleospin DNA purification kit (Macherey-Nagel). Chromatin was reconstituted by salt dialysis: DNA template and human core histones were dialyzed 18-24 hours at 4°C from Hi salt buffer (10 mM Tris-HCl pH 7.6, 2M NaCl, 1 mM EDTA, 0.05% NP-40, 5 mM BME) to Lo salt buffer (10 mM Tris-HCl pH 7.6, 50 mM NaCl, 1 mM EDTA, 0.05% NP-40, 5 mM BME). Chromatin was dialyzed for an additional 1 hour in Lo salt buffer and concentrated using Ultra-4 10K Centrifugal Filter Device (Amicon) and stored at 4°C for no longer than one month.

RNA annealing

RNA pre-annealing was performed by heating RNA at 94°C for 4 minutes in annealing buffer (6 mM HEPES pH 7.5, 60 mM KCl, 1 mM $MgCl_2$), then slow cooling on bench for 40 minutes followed by placing samples on ice.

Histone methyltransferase assays

HMTase assays were performed in a total volume of 15 µl containing HMTase buffer (10 mM HEPES, pH 7.5, 2.5 mM $MgCl_2$, 0.25 mM EDTA, 4% Glycerol and 0.1 mM DTT) with 75 µM S-Adenosylmethionine (SAM, NEB), varying amounts ssRNA and duplex RNA (see above), 600 nM JARID2, 360 nM of dinucleosomes, and 600-660 nM recombinant human PRC2 complexes under the following conditions.

The reaction mixture was incubated for 30 minutes at 30°C and stopped by adding 12 µl of Laemmli sample buffer (Biorad). After HMT reactions, samples were incubated for 5 minutes at 95°C and separated on SDS-PAGE gels. Gels were then subjected to wet transfer (30% MeOH transfer buffer) of histones to 0.22 µm PVDF membranes (Millipore) and protein was detected by Western blot analysis using primary αRb H3K27me3 antibody (Millipore #07-449), secondary antibody (Biorad #170-6515), H3-HRP (Abcam #ab21054). Similar experiments were performed, except that the total ribonucleotide concentrations of all RNAs used were kept constant (Supplementary Fig. S2).

REFERENCES

1. Bonasio, R., and Shiekhattar, R. (2014) Regulation of transcription by long noncoding RNAs. *Annu Rev Genet* **48**, 433-455

2. Wu, S. K., Roberts, J. T., Balas, M. M., and Johnson, A. M. (2020) RNA matchmaking in chromatin regulation. *Biochem Soc Trans. DOI* 10.1042/BST20191225. *In press.*

3. Loda, A., and Heard, E. (2019) Xist RNA in action: Past, present, and future. *PLoS Genet* **15**, e1008333

4. Rinn, J. L., Kertesz, M., Wang, J. K., Squazzo, S. L., Xu, X., Brugmann, S. A., Goodnough, L. H., Helms, J. A., Farnham, P. J., Segal, E., and Chang, H. Y. (2007) Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* **129**, 1311-1323

5. Yu, J. R., Lee, C. H., Oksuz, O., Stafford, J. M., and Reinberg, D. (2019) PRC2 is high maintenance. *Genes Dev* **33**, 903-935

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

7. Almeida, M., Bowness, J. S., and Brockdorff, N. (2020) The many faces of Polycomb regulation by RNA. *Curr Opin Genet Dev* **61**, 53-61

8. Zylicz, J. J., Bousard, A., Zumer, K., Dossin, F., Mohammad, E., da Rocha, S. T., Schwalb, B., Syx, L., Dingli, F., Loew, D., Cramer, P., and Heard, E. (2019) The Implication of Early Chromatin Changes in X Chromosome Inactivation. *Cell* **176**, 182-197 e123

9. Portoso, M., Ragazzini, R., Brencic, Z., Moiani, A., Michaud, A., Vassilev, I., Wassef, M., Servant, N., Sargueil, B., and Margueron, R. (2017) PRC2 is dispensable for HOTAIR-mediated transcriptional repression. *EMBO J*

10. 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

11. 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 e1055

12. Beltran, M., Tavares, M., Justin, N., Khandelwal, G., Ambrose, J., Foster, B. M., Worlock, K. B., Tvardovskiy, A., Kunzelmann, S., Herrero, J., Bartke, T., Gamblin, S. J., Wilson, J. R., and Jenner, R. G. (2019) G-tract RNA removes Polycomb repressive complex 2 from genes. *Nat Struct Mol Biol* **26**, 899-909

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

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

15. Wang, X., Paucek, R. D., Gooding, A. R., Brown, Z. Z., Ge, E. J., Muir, T. W., and Cech, T. R. (2017) Molecular analysis of PRC2 recruitment to DNA in chromatin and its inhibition by RNA. *Nat Struct Mol Biol* **24**, 1028-1038

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

17. Cifuentes-Rojas, C., Hernandez, A. J., Sarma, K., and Lee, J. T. (2014) Regulatory interactions between RNA and polycomb repressive complex 2. *Mol Cell* **55**, 171-185

18. 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., Ule, J., and Jenner, R. G. (2016) The interaction of PRC2 with RNA or chromatin is mutually antagonistic. *Genome Res* **26**, 896-907

19. Davidovich, C., Wang, X., Cifuentes-Rojas, C., Goodrich, K. J., Gooding, A. R., Lee, J. T., and Cech, T. R. (2015) Toward a consensus on the binding specificity and promiscuity of PRC2 for RNA. *Mol Cell* **57**, 552-558

20. Zhang, Q., McKenzie, N. J., Warneford-Thomson, R., Gail, E. H., Flanigan, S. F., Owen, B. M., Lauman, R., Levina, V., Garcia, B. A., Schittenhelm, R. B., Bonasio, R., and Davidovich, C. (2019) RNA exploits an exposed regulatory site to inhibit the enzymatic activity of PRC2. *Nat Struct Mol Biol* **26**, 237-247

21. Voigt, P., Leroy, G., Drury, W. J., 3rd, Zee, B. M., Son, J., Beck, D. B., Young, N. L., Garcia, B. A., and Reinberg, D. (2012) Asymmetrically modified nucleosomes. *Cell* **151**, 181-193

22. Ogawa, Y., Sun, B. K., and Lee, J. T. (2008) Intersection of the RNA interference and X-inactivation pathways. *Science* **320**, 1336-1341

23. Meredith, E. K., Balas, M. M., Sindy, K., Haislop, K., and Johnson, A. M. (2016) An RNA matchmaker protein regulates the activity of the long noncoding RNA HOTAIR. *RNA* **22**, 995-1010

24. Nguyen, E. D., Balas, M. M., Griffin, A. M., Roberts, J. T., and Johnson, A. M. (2018) Global profiling of hnRNP A2/B1-RNA binding on chromatin highlights LncRNA interactions. *RNA Biol*, 1-13

25. Terai, G., Iwakiri, J., Kameda, T., Hamada, M., and Asai, K. (2016) Comprehensive prediction of lncRNA-RNA interactions in human transcriptome. *BMC Genomics* **17 Suppl 1**, 12

677 26. Nepal, C., Taranta, A., Hadzhiev, Y., Pundhir, S., Mydel, P., Lenhard, B., Muller, F., and Andersen,
678 J. B. (2020) Ancestrally Duplicated Conserved Noncoding Element Suggests Dual Regulatory
679 Roles of HOTAIR in cis and trans. *iScience* **23**, 101008

680 27. Wu, B., Su, S., Patil, D. P., Liu, H., Gan, J., Jaffrey, S. R., and Ma, J. (2018) Molecular basis for
681 the specific and multivariant recognitions of RNA substrates by human hnRNP A2/B1. *Nat*
682 *Commun* **9**, 420

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

685 29. Smola, M. J., Christy, T. W., Inoue, K., Nicholson, C. O., Friedersdorf, M., Keene, J. D., Lee, D.
686 M., Calabrese, J. M., and Weeks, K. M. (2016) SHAPE reveals transcript-wide interactions,
687 complex structural domains, and protein interactions across the Xist lncRNA in living cells. *Proc*
688 *Natl Acad Sci U S A* **113**, 10322-10327

689 30. Kumar, A., and Wilson, S. H. (1990) Studies of the strand-annealing activity of mammalian hnRNP
690 complex protein A1. *Biochemistry* **29**, 10717-10722

691 31. Mayeda, A., Munroe, S. H., Caceres, J. F., and Krainer, A. R. (1994) Function of conserved
692 domains of hnRNP A1 and other hnRNP A/B proteins. *EMBO J* **13**, 5483-5495

693 32. Lee, S., Kim, H., Tian, S., Lee, T., Yoon, S., and Das, R. (2015) Automated band annotation for
694 RNA structure probing experiments with numerous capillary electrophoresis profiles.
695 *Bioinformatics* **31**, 2808-2815

696 33. Kladwang, W., Mann, T. H., Becka, A., Tian, S., Kim, H., Yoon, S., and Das, R. (2014)
697 Standardization of RNA chemical mapping experiments. *Biochemistry* **53**, 3063-3065

698 34. Kim, H., Cordero, P., Das, R., and Yoon, S. (2013) HiTRACE-Web: an online tool for robust
699 analysis of high-throughput capillary electrophoresis. *Nucleic Acids Res* **41**, W492-498

700 35. Yoon, S., Kim, J., Hum, J., Kim, H., Park, S., Kladwang, W., and Das, R. (2011) HiTRACE: high-
701 throughput robust analysis for capillary electrophoresis. *Bioinformatics* **27**, 1798-1805

702 36. Hartwick, E. W., Costantino, D. A., MacFadden, A., Nix, J. C., Tian, S., Das, R., and Kieft, J. S.
703 (2018) Ribosome-induced RNA conformational changes in a viral 3'-UTR sense and regulate
704 translation levels. *Nat Commun* **9**, 5074

705 37. Hajdin, C. E., Bellaousov, S., Huggins, W., Leonard, C. W., Mathews, D. H., and Weeks, K. M.
706 (2013) Accurate SHAPE-directed RNA secondary structure modeling, including pseudoknots.
707 *Proc Natl Acad Sci U S A* **110**, 5498-5503

708 38. Gupta, R. A., Shah, N., Wang, K. C., Kim, J., Horlings, H. M., Wong, D. J., Tsai, M. C., Hung, T.,
709 Argani, P., Rinn, J. L., Wang, Y., Brzoska, P., Kong, B., Li, R., West, R. B., van de Vijver, M. J.,
710 Sukumar, S., and Chang, H. Y. (2010) Long non-coding RNA HOTAIR reprograms chromatin state
711 to promote cancer metastasis. *Nature* **464**, 1071-1076

712 39. Molliex, A., Temirov, J., Lee, J., Coughlin, M., Kanagaraj, A. P., Kim, H. J., Mittag, T., and Taylor,
713 J. P. (2015) Phase separation by low complexity domains promotes stress granule assembly and
714 drives pathological fibrillization. *Cell* **163**, 123-133

715 40. Sancar, A., and Hearst, J. E. (1993) Molecular matchmakers. *Science* **259**, 1415-1420

716 41. Lee, J. T., and Bartolomei, M. S. (2013) X-inactivation, imprinting, and long noncoding RNAs in
717 health and disease. *Cell* **152**, 1308-1323

718 42. Ohhata, T., Hoki, Y., Sasaki, H., and Sado, T. (2008) Crucial role of antisense transcription across
719 the Xist promoter in Tsix-mediated Xist chromatin modification. *Development* **135**, 227-235

720 43. Lewis, A., Mitsuya, K., Umlauf, D., Smith, P., Dean, W., Walter, J., Higgins, M., Feil, R., and Reik,
721 W. (2004) Imprinting on distal chromosome 7 in the placenta involves repressive histone
722 methylation independent of DNA methylation. *Nat Genet* **36**, 1291-1295

723 44. Young, M. D., Willson, T. A., Wakefield, M. J., Trounson, E., Hilton, D. J., Blewitt, M. E., Oshlack,
724 A., and Majewski, I. J. (2011) ChIP-seq analysis reveals distinct H3K27me3 profiles that correlate
725 with transcriptional activity. *Nucleic Acids Res* **39**, 7415-7427

726 45. Brookes, E., de Santiago, I., Hebenstreit, D., Morris, K. J., Carroll, T., Xie, S. Q., Stock, J. K.,
727 Heidemann, M., Eick, D., Nozaki, N., Kimura, H., Ragoussis, J., Teichmann, S. A., and Pombo, A.
728 (2012) Polycomb associates genome-wide with a specific RNA polymerase II variant, and
729 regulates metabolic genes in ESCs. *Cell Stem Cell* **10**, 157-170

730 46. Wei, C., Xiao, R., Chen, L., Cui, H., Zhou, Y., Xue, Y., Hu, J., Zhou, B., Tsutsui, T., Qiu, J., Li, H.,
731 Tang, L., and Fu, X. D. (2016) RBFox2 Binds Nascent RNA to Globally Regulate Polycomb
732 Complex 2 Targeting in Mammalian Genomes. *Mol Cell* **62**, 982

733 47. Zhou, H., Shipkovenska, G., Kalocsay, M., Zhang, J., Luo, Z., Gygi, S. P., and Moazed, D. (2019)
734 An RNA degradation complex required for silencing of Polycomb target genes. *bioRxiv*,
735 2019.2012.2023.887547

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

739 49. Wright, P. R., Georg, J., Mann, M., Sorescu, D. A., Richter, A. S., Lott, S., Kleinkauf, R., Hess, W.
740 R., and Backofen, R. (2014) CopraRNA and IntaRNA: predicting small RNA targets, networks and
741 interaction domains. *Nucleic Acids Res* **42**, W119-123
742

743

744 **FIGURE LEGENDS**

745 **FIGURE 1. HOTAIR intermolecular interaction is promoted by hnRNP B1. (A)** Model of B1-mediated
746 HOTAIR RNA-RNA interactions with nascent target RNA leading to PRC2 activity and gene silencing. (**B)**
747 In vitro RNA pulldown with MS2-tagged HOTAIR or Anti-luc with recombinant B1 or A2. “A2+” is 3x the
748 concentration of A2. Minus MS2-MBP fusion protein pulldown included to account for background bead
749 binding. Western blot analysis using A2/B1 antibody. RNA recovery quantified by qPCR (n=3). **(C)** Crystal
750 structure of RRM domains of A2/B1 in complex with 10 mer RNA (yellow)(27). Two molecules of the
751 tandem RRMs are shown in purple/green. Blue circles highlight N-terminus. Adapted from (27)

http://creativecommons.org/licenses/by/4.0/. **(D)** Schematic of RNA-RNA interaction assay: RAT-tagged JAM2 and HOTAIR IVTs incubated +/- recombinant hnRNP B1. JAM2 tethered by PP7 coat protein on magnetic beads. Recovery of HOTAIR by JAM2 pulldown quantified by RT-qPCR and protein by Western. **(E)** Assays from 1D with HOTAIR or Anti-luc RNA +/- hnRNP B1 or A2. (n=6). **(F)** As in 1E with full-length HOTAIR, HOTAIR with the JAM2 interaction site deleted or mutated +/- hnRNP B1 (n = 3). Error bars in 1E,F represent standard deviations. P-values determined using two-way ANOVA and two-tailed student t tests.

FIGURE 2. Examining hnRNP B1 specific interactions with the lncRNA HOTAIR using in vitro eCLIP mapping. (A) Schematic of in vitro eCLIP experiments: recombinant B1 incubated with IVT HOTAIR. HOTAIR-B1 complexes were UV-crosslinked. RNA was fragmented with limited RNase A treatment, followed by in vitro eCLIP protocol (see Methods and (24)). **(B)** (*Top)* Mapping of reverse transcription termination events from HOTAIR-B1 in vitro eCLIP as a measure of direct protein crosslinking (HOTAIR domains alternately shaded violet). Termination sites normalized to read count with significant peaks determined by values greater than 5000. (*Bottom*) Zoom in on domain 1 (1-530) for titration experiments highlights multiple B1 interaction sites (shaded in grey). **(C)** (*Top*) Diagram of constructs including construct with all RGGs mutated ("5XRGG MUT") and B1 glycine-rich domain deletion ("ΔGR"). (*Bottom*) Assays as in 1B with recombinant truncated versions of the constructs depicted above. Western for A2/B1. Intensities should be compared to input, since the antibody recognizes the constructs differentially. Equal protein loading for samples demonstrated by Coomassie gel with equal amounts of protein loaded as in each pulldown. Bar graph of HOTAIR recovery as percent input from each pulldown was quantified by qPCR (n=2).

FIGURE 3. Chemical probing of HOTAIR highlights B1 interactions. (A) Diagram of the IVT HOTAIR domain 1 construct for chemical probing experiments. Schematic representation of 1M7 chemical probing of HOTAIR domain 1 with B1 and/or the JAM2 fragment (54 nt). **(B)** Heatmap of normalized reactivity for

778 HOTAIR only (HA only), HOTAIR + JAM2 (HAJ2), HOTAIR + B1 (HAB1) and HOTAIR + B1 + JAM2
779 (HAB1J2). White values represent non-reactive nucleotides and red values represent more reactive
780 nucleotides. **(C)** Line graphs of normalized reactivity for each condition. Light blue shaded boxes highlight
781 the in vitro B1 eCLIP sites identified. Light red shaded box highlights the RNA-RNA interaction site. **(D)**
782 Boxplots of normalized reactivity for all nucleotide positions, the JAM2 interaction region (245-306 nt), and
783 B1 eCLIP-derived binding sites (141-172 nt, 304-314 nt, 460-523 nt). Error bars represent standard
784 deviations. P-values determined using one-way ANOVA multiple comparison tests between HA only
785 compared to each variable condition (*P < 0.01, **P < 0.005, ***P < 0.0001). **(E)** Bar graphs for normalized
786 reactivity of HOTAIR only and HOTAIR + B1 at specific eCLIP B1 binding sites, as well as a minimally
787 changed control region, by nucleotide.

788

789 **FIGURE 4. Establishment of RNA-RNA interactions alters HOTAIR structure. (A)** Heatmap of
790 HOTAIR reactivity changes upon addition of JAM2 (ΔJ2), hnRNP B1 (ΔB1) or both (ΔB1J2), compared to
791 HOTAIR only. Red values become more reactive, white values don't change, and blue values become
792 less reactive when JAM2, B1 or both are present. **(B)** Prominent regions of change mapped to the HOTAIR
793 domain 1 secondary structure(28) (with permission) with either JAM2 RNA, **(C)** hnRNP B1 (highlighted in
794 grey) or **(D,E)** both JAM2+B1 using either control RNA threshold analysis (highlighted in light green) or
795 control protein threshold analysis (highlighted in light purple). Regions of change were determined as
796 described in Materials and Methods. In vitro eCLIP B1 binding sites are highlighted in blue and the JAM2
797 RNA-RNA interaction site is highlighted in green.

798

799 **FIGURE 5. Duplex RNA promotes PRC2 activity. (A)** Histograms comparing the predicted minimum
800 free energy for RNA-RNA interactions between HOTAIR and 40,740 RNAs across the transcriptome
801 (grey) or between HOTAIR and 885 transcripts from genes that gain PRC2 activity when HOTAIR is
802 overexpressed in breast cancer cells (red). Data from(23,25,38). **(B)** Native gel of di-nucleosomes
803 reconstituted via salt dialysis using a DNA template containing two 601 sequences surrounding 40-bp of

linker DNA. DNA and nucleosome samples were run on a 5% native polyacrylamide gel and stained with SYBR Gold. **(C)** Recombinant human PRC2 complex includes SUZ12, EZH2, EED, RBBP4 and AEBP2, analyzed by SDS-PAGE and stained with Coomassie blue. **(D)** Histone methyltransferase assay (HMTase assay) was performed with recombinant PRC2 complex, di-nucleosomes, S-Adenosylmethionine (SAM) with and without the co-factor JARID2 (amino acids 119-574). PRC2 activity was determined by SDS-PAGE followed by H3K27me3 and total H3 Western blot analysis. **(E)** Native 0.5X TBE gel of RNA annealing titration with HOTAIR forward and reverse fragments to show formation of dsRNA. HMTase assay with annealed HOTAIR dsRNA titration analyzed by Western blot.

FIGURE 6. HOTAIR matching with target RNA promotes PRC2 activity. (A) Native PAGE of RNA annealing titration with HOTAIR fragments and either JAM2 or HOXD10 matching RNAs. **(B)** RNA-RNA interaction between HOTAIR and JAM2 predicted by IntaRNA(49). HMTase assay performed with recombinant PRC2, di-nucleosomes, HOTAIR fragment (62 nt) and JAM2 match (54 nt) titration. PRC2 activity determined by H3K27me3 and total H3 Western. **(C)** Quantification of 6B (n=3). **(D)** IntaRNA result of HOTAIR and HOXD10. HMTase as in 6B with HOTAIR fragment (31 nt) HOXD10 match (37 nt) titration. **(E)** Quantification of 6D (n=5). **(F**) As in 6D, with polyA instead of HOXD10 (n=4). **(G)** HMTase as above with HOTAIR domain 1 (nucleotides 1-530) and JAM2 match (62 nt) titration. **(H)** Quantification of 6G (n=3). **(I)** As in 6G, except using HOTAIR with JAM2 match site deleted “HOTAIR D1_del”. **(J)** As in 6G, except with the addition of hnRNP B1. **(K)** Quantification of 6J (n=3). (**C,E,F,H,I,K**) Percent relief of inhibition is normalized to H3 signal and relative to no RNA and described as % relief from HOTAIR only reaction. Error bars represent standard deviations. P-values were determined using unpaired t tests with a 95% confidence interval.

Figure 7. Model for HOTAIR-mediated chromatin silencing via intermolecular RNA-RNA interactions. Binding of PRC2 to single-stranded regions of HOTAIR inhibits enzymatic activity. B1 promotes RNA-RNA matching of HOTAIR with target nascent RNA via bridging of the RNAs and

conformational changes in the lncRNA to promote intermolecular base-pairing. B1 dissociates from the RNAs, promoting a conformation that may reduce PRC2 affinity for those RNAs through formation of the dsRNA match, thereby increasing PRC2 interaction with chromatin, leading to H3K27me3 and transcriptional repression.

Figure 1

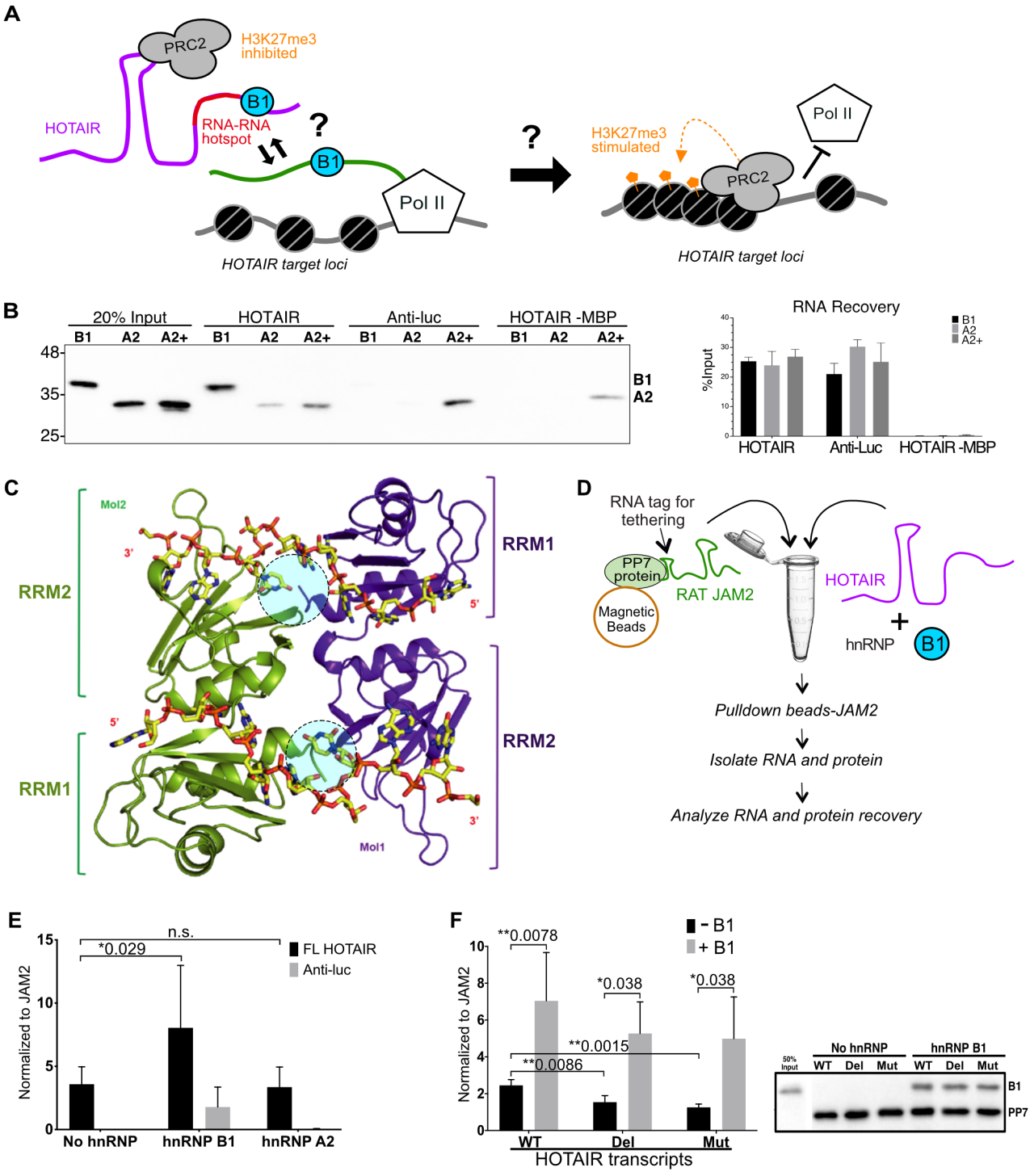


Figure 2

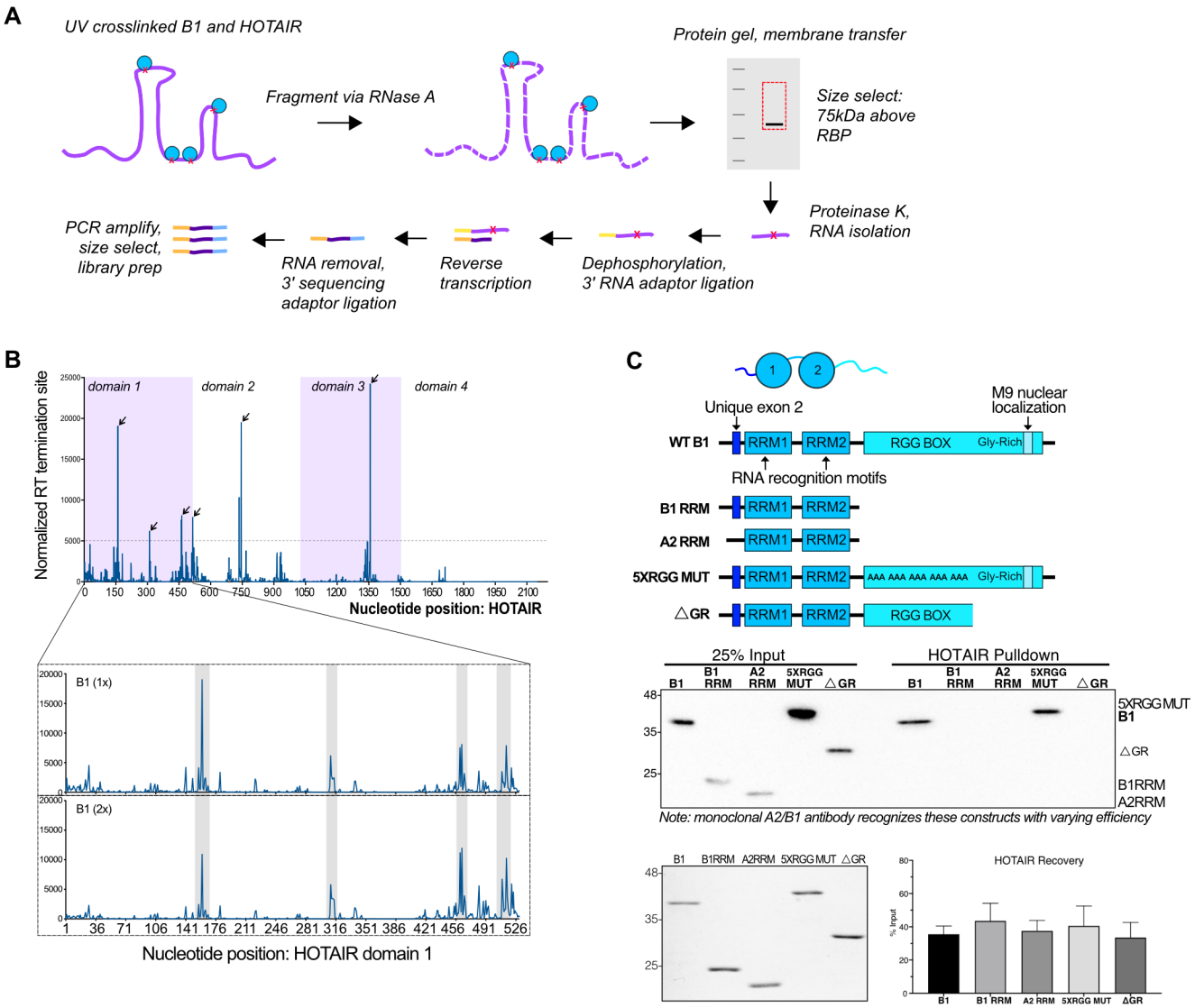


Figure 3

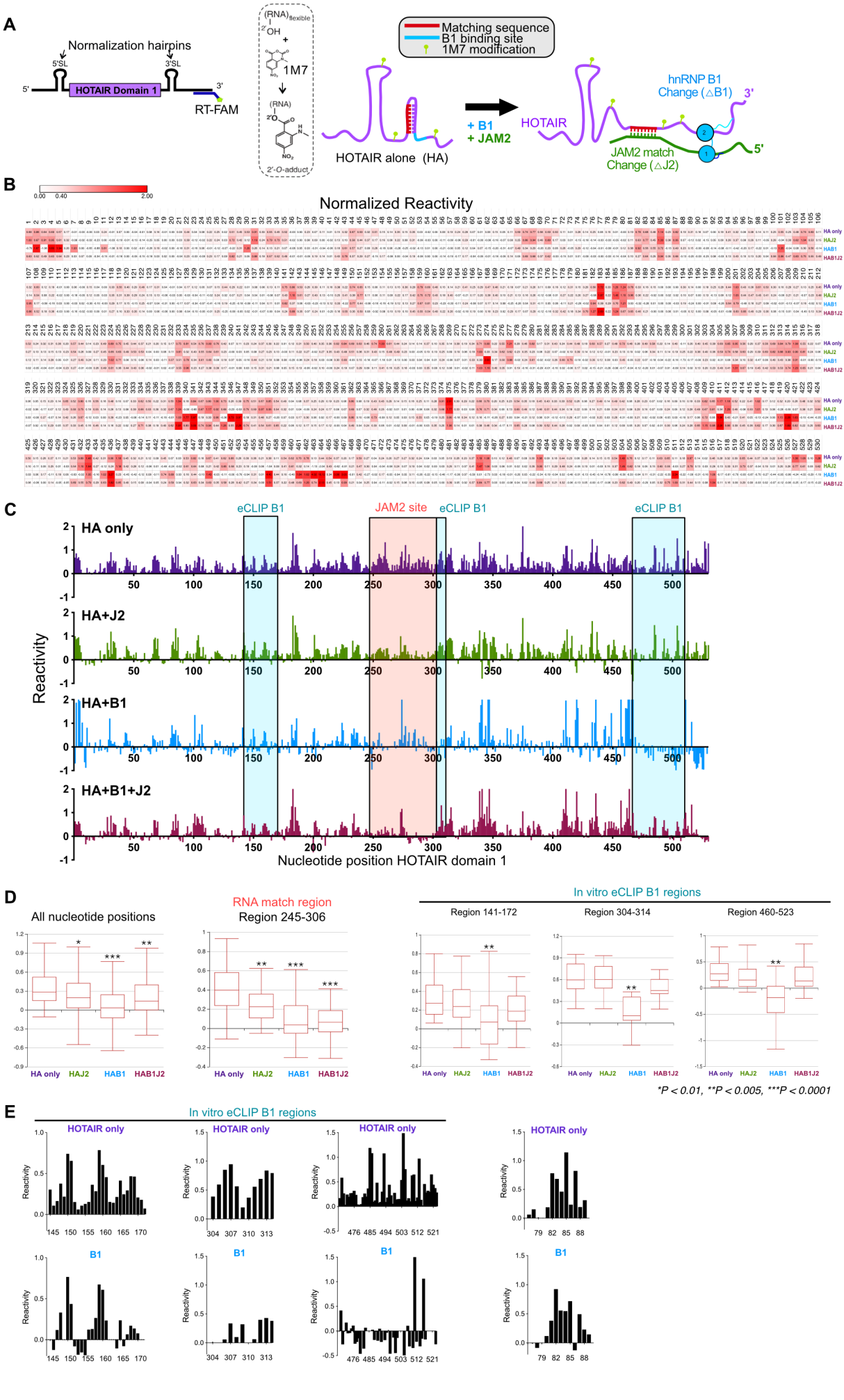
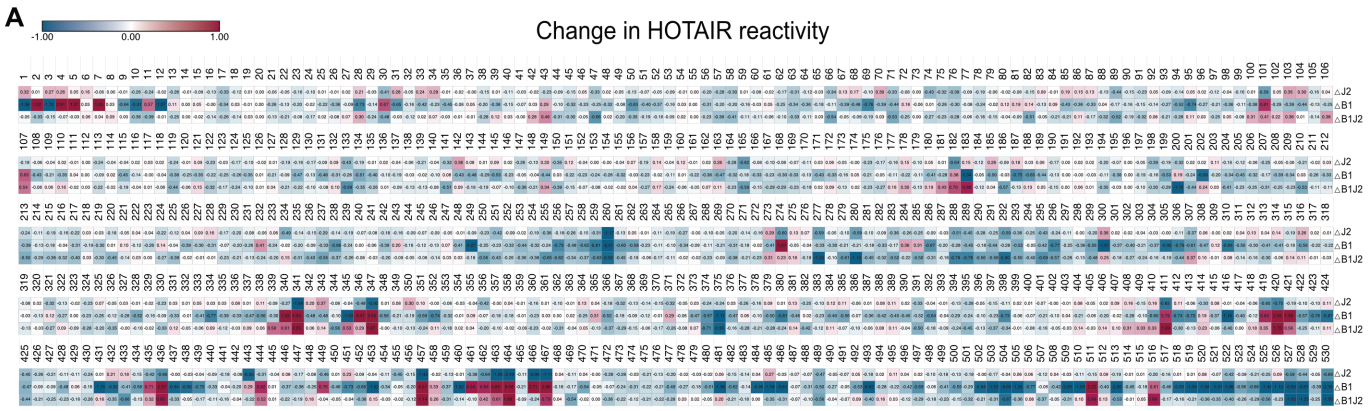
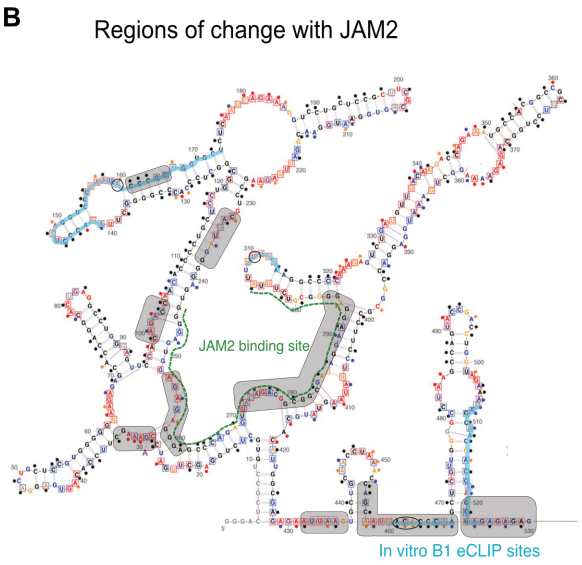
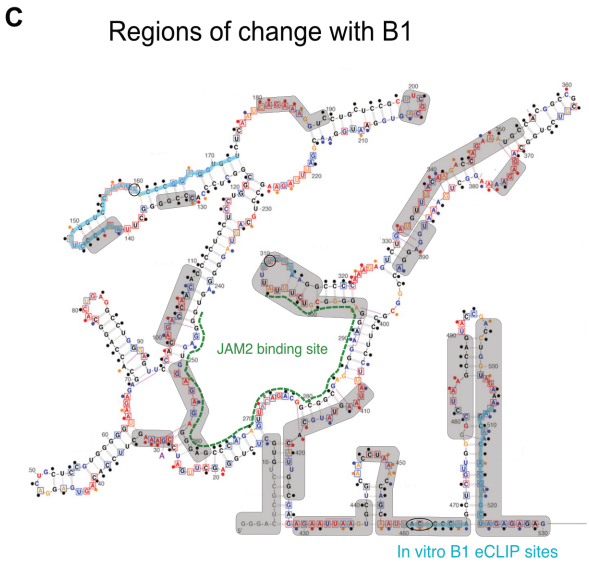
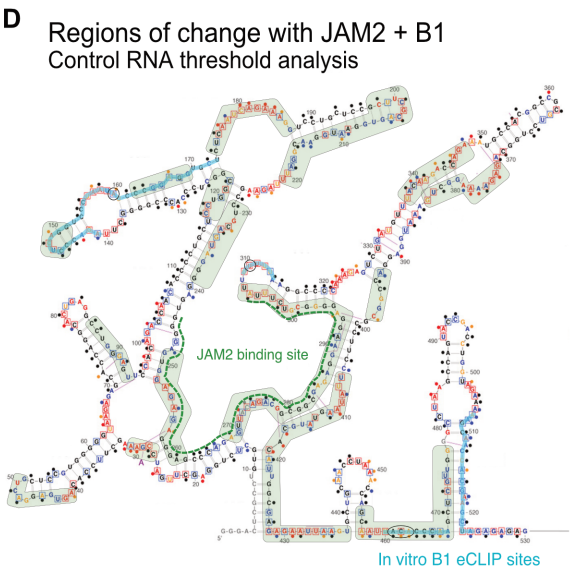


Figure 4









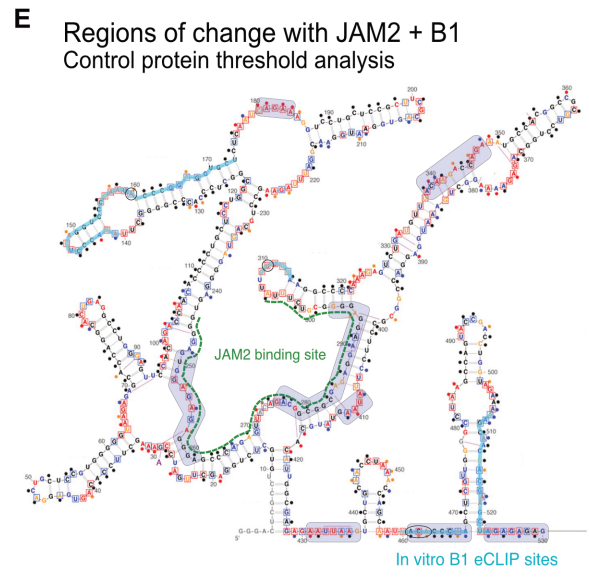


Figure 5

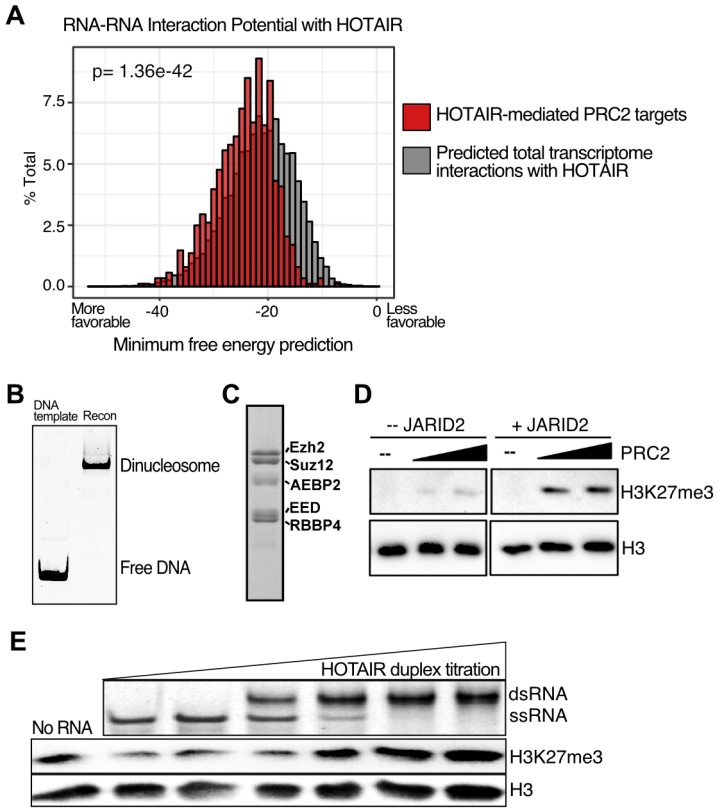


Figure 6

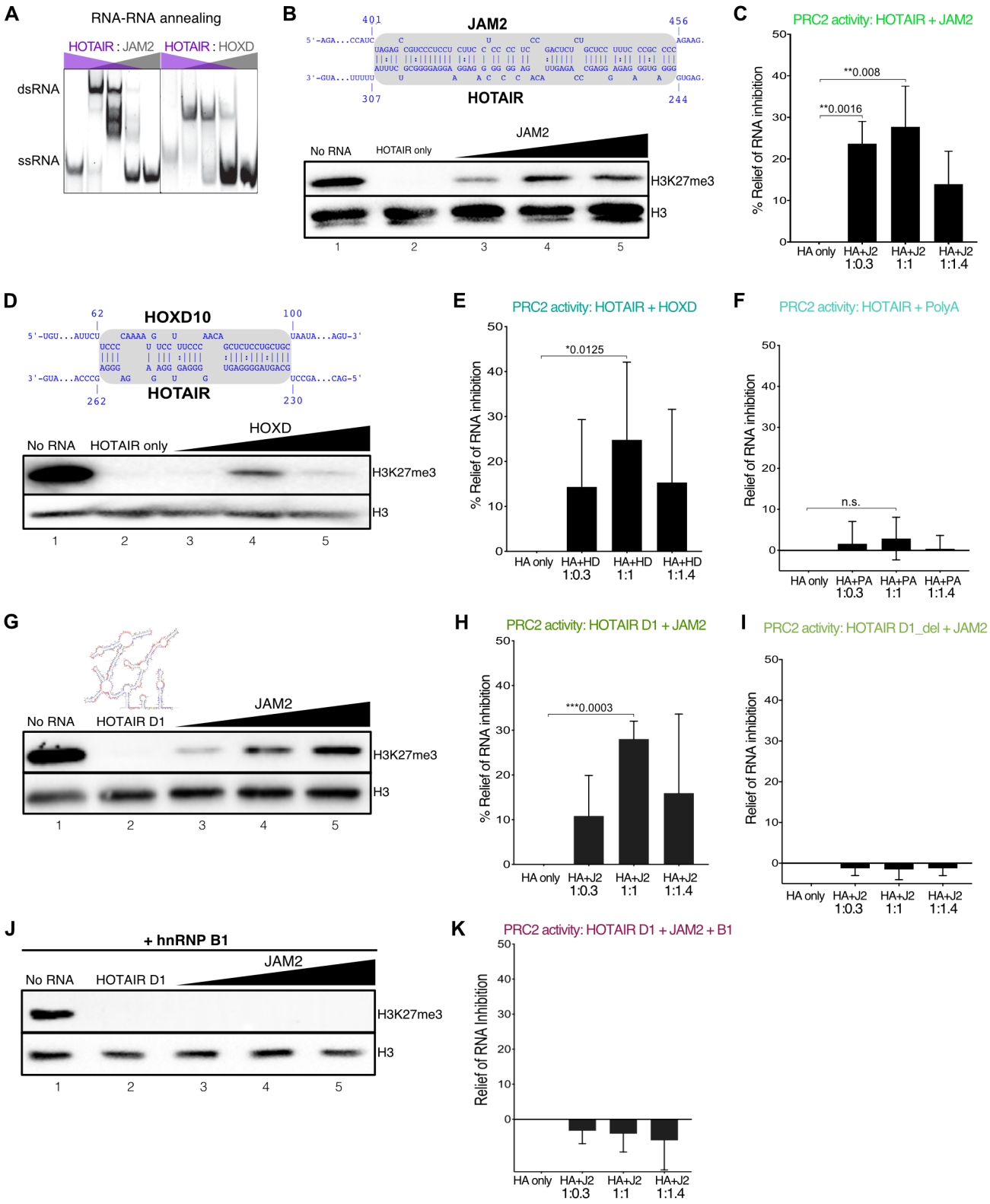


Figure 7