

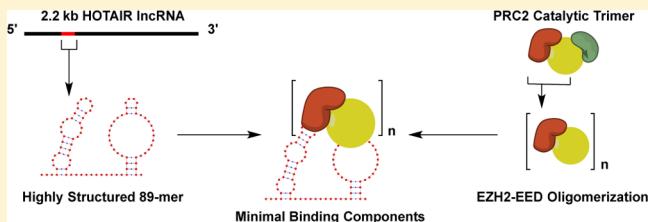
# Binding Interactions between Long Noncoding RNA HOTAIR and PRC2 Proteins

 Liang Wu,<sup>†</sup> Pierre Murat,<sup>†</sup> Dijana Matak-Vinkovic,<sup>†</sup> Adele Murrell,<sup>‡</sup> and Shankar Balasubramanian\*,<sup>†,‡</sup>
<sup>†</sup>Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, CB2 1EW, United Kingdom

<sup>‡</sup>Cancer Research U.K., Cambridge Institute, Li Ka Shing Centre, Robinson Way, Cambridge, CB2 0RE, United Kingdom

**S Supporting Information**

**ABSTRACT:** Long noncoding RNAs (lncRNAs) play a key role in the epigenetic regulation of cells. Many of these lncRNAs function by interacting with histone repressive proteins of the Polycomb group (PcG) family, recruiting them to gene loci to facilitate silencing. Although there are now many RNAs known to interact with the PRC2 complex, little is known about the details of the molecular interactions. Here, we show that the PcG protein heterodimer EZH2-EED is necessary and sufficient for binding to the lncRNA HOTAIR. We also show that protein recognition occurs within a folded 89-mer domain of HOTAIR. This 89-mer represents a minimal binding motif, as further deletion of nucleotides results in substantial loss of affinity for PRC2. These findings provide molecular insights into an important system involved in epigenetic regulation.



## INTRODUCTION

Long noncoding (lnc)RNAs are defined as RNA molecules over 200 nt in length which do not encode for a protein. Once considered to be transcriptional noise, lncRNAs have now been shown to regulate many key biological processes, including nuclear transport,<sup>1</sup> microRNA activity,<sup>2</sup> and epigenetic regulation.<sup>3,4</sup> One significant role of lncRNAs is the regulation of gene expression via a mechanism involving interaction with the epigenetic silencing complex Polycomb Repressive Complex 2 (PRC2), with ~20% of all lncRNA transcripts estimated to bind PRC2.<sup>5</sup>

The PRC2 complex contains the three core protein subunits EZH2, EED, and SUZ12. EZH2 contains a SET domain which catalyzes trimethylation of histone H3 at lysine 27 (H3K27 → H3K27Me3), a mark associated with transcriptional repression; SUZ12 and EED are also required for this EZH2 methyltransferase activity to occur.<sup>6</sup> The core PRC2 complex is unable to target and silence genomic regions by itself. Instead, many of the lncRNAs associated with PRC2 have been shown to act as cellular ‘address codes’, which guide PRC2 silencing to target specific regions of the genome where the lncRNA associates.<sup>7</sup> Given these lncRNAs direct epigenetic silencing via association with PRC2, their overexpression can lead to aberrant silencing of tumor suppressor genes, resulting in malignant cancerous phenotypes.<sup>8</sup>

*HOX Transcript Antisense RNA* (HOTAIR) is one of the most well-studied PRC2 interacting lncRNAs, first established as a regulator of HOX gene expression.<sup>9</sup> HOTAIR overexpression is associated with widespread gene expression changes and aggressive metastatic phenotypes in many cancers, including breast,<sup>10</sup> colorectal,<sup>11</sup> and hepatocellular carcinomas.<sup>12</sup> These aggressive phenotype changes are recapitulated in

vitro for cells with enforced HOTAIR overexpression,<sup>10,11</sup> suggesting that HOTAIR is responsible for driving these malignant characteristics.

Although the effects of HOTAIR overexpression have been well-characterized functionally,<sup>9,10</sup> the molecular details of its interaction with PRC2 are in need of further elucidation. EZH2 has been shown to interact with an in vitro transcribed HOTAIR RNA probe, suggesting it is involved in direct binding to HOTAIR.<sup>13</sup> However, the affinity of this interaction and possible contributions of other PRC2 subunits to HOTAIR binding were not established. For HOTAIR, a 300-mer domain at the 5' terminus has been found to be necessary and sufficient for interaction with PRC2.<sup>14</sup> However, the minimal HOTAIR domain required for PRC2 binding has yet to be defined.

Further knowledge of the molecular basis of HOTAIR-PRC2 binding will be important for understanding how HOTAIR overexpression can drive malignancy in cancers through PRC2. Moreover, a deeper understanding of the details of HOTAIR-PRC2 binding may provide a useful foundation for exploring this interaction as a target for small molecule intervention. Herein we describe a systematic investigation of the binding interaction between subunits of the PRC2 core catalytic heterotrimer (EZH2-EED-SUZ12, hereafter referred to as PRC2 3m) and the lncRNA HOTAIR. As part of this investigation, we report a minimal HOTAIR sequence responsible for PRC2 3m interaction, as well as a secondary structure for this domain derived from nuclease mapping. Collectively, these data give insight into the details of the

Received: August 9, 2013

Revised: October 29, 2013

Published: December 10, 2013

epigenetically important interaction between HOTAIR and PRC2.

## MATERIALS AND METHODS

**Preparation of Radiolabeled ncRNAs.** DNA templates for in vitro transcription were prepared by PCR amplification from a HOTAIR containing plasmid (a generous gift from Dr. Howard Chang - Stanford University) using primers listed in Table S1. RNAs were transcribed using the MEGAshortScript kit (Ambion) according to manufacturers protocol, followed by dephosphorylation with Antarctic Phosphatase (New England Biolabs). Dephosphorylated RNAs were 5' radiolabeled using T4 PNK (New England Biolabs) in the presence of [ $\gamma$ -<sup>32</sup>P]ATP (PerkinElmer). In all cases, radiolabeled RNAs were purified by gel extraction. RNA sequences used in this study are listed in Table S2.

**Generation of Recombinant Proteins.** Baculovirus for 6xHIS-EED and SUZ12 were a generous gift from Prof. Kristian Helin, University of Copenhagen. Transfer plasmid pFastBac-FLAG-EZH2 was a generous gift from Prof. Robert Kingston, Harvard Medical School. Proteins were expressed in High Five insect cells using baculoviral expression vectors. For dimers and trimers, the complexes were expressed by coinfection and purified together as a single unit, rather than mixing prepurified monomers. High Five cells were harvested 48 h after infection and purified according to tag. For FLAG tagged proteins (FLAG-EZH2 and FLAG-EZH2-SUZ12), pellets were resuspended in BC300 buffer (20 mM HEPES [pH 7.9], 300 mM KCl, 0.2 mM EDTA, 10% glycerol, 1 mM DTT, 0.2 mM PMSF, and complete protease inhibitors (Roche)), lysed using sonication, and extracts cleared by centrifugation. Cleared extracts were incubated with prewashed M2 beads (Sigma) for 4 h at 4 °C. Beads were then pelleted, briefly washed 3X with BC2000 (as above, but 2 M KCl), and once with BC300. Beads were eluted with 0.2 mg/mL 3xFLAG peptide (Sigma), and desalting using Zeba spin desalting columns (Pierce). Where necessary, proteins were further purified by gel filtration using a HiLoad 16/600 Superdex 200 column (GE healthcare), followed by concentration using Amicon ultra 10 kDa MWCO spin concentrators (Millipore). For 6xHis-EED, pellets were resuspended in Histrap buffer (as for BC300, but including 50 mM imidazole), lysed by sonication, and extracts cleared by centrifugation. Cleared extracts were loaded onto a Histrap FF column (GE Healthcare) and eluted with a 50–500 mM imidazole gradient. Combined peak fractions were further purified by gel filtration followed by spin concentration. For dual tagged complexes (FLAG-EZH2–6xHIS-EED and PRC2 3m), purification initially followed the procedure for 6xHis tag purification. Combined Histrap fractions were then incubated with prewashed M2 beads, and further purified following the procedure for FLAG purification. Protein purity was assessed by SDS-PAGE and Western blotting (SI Figure 1; antibodies listed in Table S3). Recombinant LSD1 (ab80379) was purchased from AbCam.

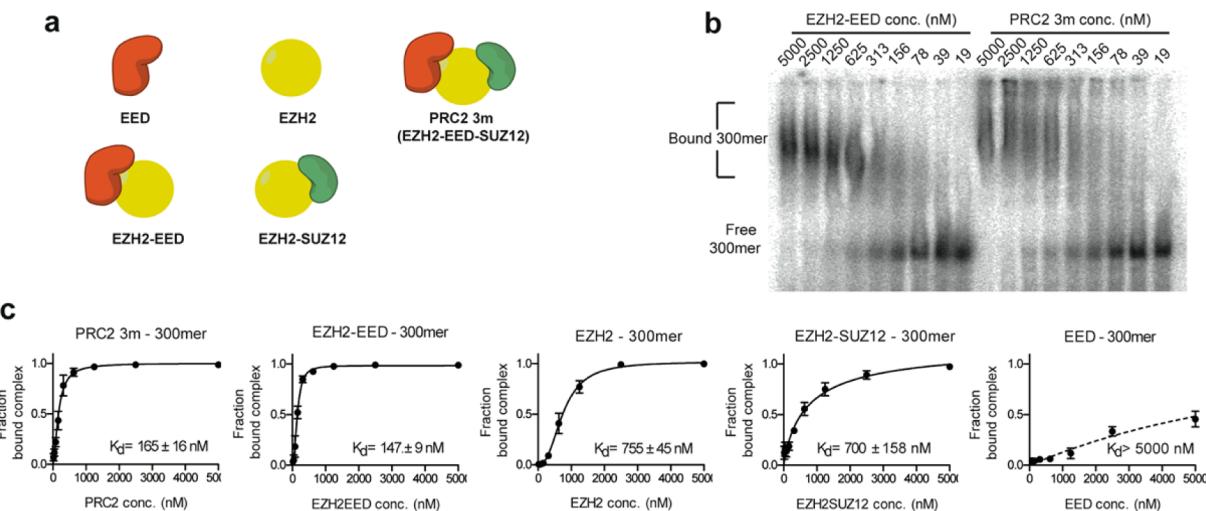
**Electrophoretic Mobility Shift Assays (EMSA).** Protein-RNA binding was carried out with indicated amounts of protein and 5' <sup>32</sup>P labeled RNA in binding buffer (50 mM Tris-HCl [pH 7.4], 100 mM KCl, 5 mM MgCl<sub>2</sub>, 10 mM  $\beta$ -ME, 0.1% IGEPAL CA-630), with yeast tRNA (Roche) at a concentration of 1.25  $\mu$ g/ $\mu$ L. Radiolabeled RNAs were annealed by heating to 95 °C, snap cooling to 4 °C, addition of binding buffer, followed by equilibration to rt over at least 20 min. Comparison

with a slow-cooling annealing procedure produced no difference in protein affinity. Binding reactions were incubated at rt for 45 min, then loaded onto a nondenaturing 6% (200-mers or smaller) or 3% (300-mers) 75:1 acrylamide:bisacrylamide gel. Gels were visualized using storage phosphor, and quantified using ImageQuant (GE Healthcare). Where multiple upper bands were present, all were quantified together as the 'bound' fraction unless specified otherwise. Dissociation constants ( $K_d$ ) were derived from data point fitting with Prism (Graphpad) according to the function for specific binding with hill slope:  $B = B_{\max}^*[X]^h/(K_d^h + [X]^h)$  – where [X] is the concentration of protein X, h is the Hill coefficient, B is the fraction of shifted complex, and  $B_{\max}$  is the maximal amount of complex formed.

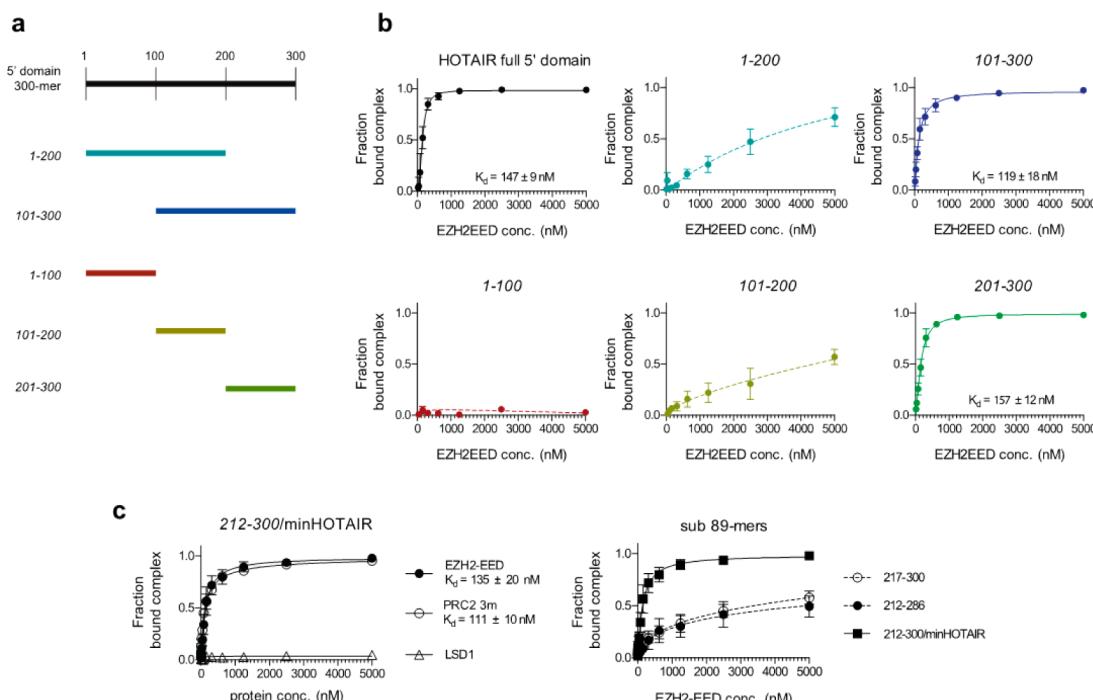
**Nondenaturing Nanoelectrospray Ionization Mass Spectrometry (Native Mass Spectrometry).** Intact EED-EZH2 complexes were prepared for native mass spectrometry experiment by buffer exchange to 500 mM ammonium acetate [pH 7.0] using Micro Bio-Spin 6 Chromatography Columns (Bio-Rad Laboratories, Inc.). Prior to the buffer exchange, protein complex solution was 161  $\mu$ M. The optimal concentration was experimentally determined using the mass spectrometer, giving a final concentration of ~15–25  $\mu$ M. Mass spectra were recorded on a Synapt HDMS instrument (Waters, Manchester, UK), and calibrated using cesium iodide (100 mg/mL). In all cases 2.5  $\mu$ L of sample solution was loaded into capillary with tips that were cut to give an inner diameter of 1–5  $\mu$ m, as previously described.<sup>15</sup> Typical values for MS parameters were: capillary voltage 1.9 kV, cone voltage 100 V, cone gas 40 L/h, extractor 1.9 V, ion transfer stage pressure 3.66 mbar, trap collision energy 15–35 V, transfer collision energy 14–19 V, trap  $4.23 \times 10^{-2}$  mbar, IMS pressure (5.01–5.02)  $\times 10^{-1}$  mbar, TOF analyzer pressure  $9.71 \times 10^{-7}$  mbar. Micromass MassLynx 4.1 was used for data acquisition and processing.

**Native Gel Electrophoresis.** Native gels were run using the NativeBlue system (Invitrogen) following the manufacturers protocols. Gels were visualized by fixing in 40% MeOH/8% AcOH solution, followed by destaining in 10% AcOH for 5 h.

**Enzymatic Mapping and Footprinting.** 5' <sup>32</sup>P labeled RNA was digested with RNase I or V1 in the presence or absence of EZH2-EED. For mapping experiments, labeled RNA was mixed with 0.1 U RNase I, or 0.01 U RNase V1, and 1  $\mu$ g yeast tRNA in binding buffer (see EMSA protocol) and incubated at RT for 15 min. Cleavage reactions were stopped by ethanol precipitation. Footprinting experiments were carried out in the same way, except mixtures also contained 200 nM or 2  $\mu$ M EZH2-EED and 0.001 U RNase V1 was used. For alkaline hydrolysis ladder, 5' <sup>32</sup>P labeled RNA and 1  $\mu$ g yeast tRNA were hydrolyzed in 20  $\mu$ L alkaline solution (50 mM Na<sub>3</sub>PO<sub>4</sub>) by heating at 65 °C for 2 min. Hydrolysis was stopped by quenching in 0.5 M Tris-HCl [pH 7.4], and RNA was purified by ethanol precipitation. For RNase T1 ladder, 5' <sup>32</sup>P labeled RNA was heated at 50 °C for 5 min in denaturing buffer [20 mM sodium citrate (pH 5), 1 mM EDTA, and 7 M urea]. 0.1 U RNase T1 was added and the mixture incubated for 15 min at rt. Digested products were resuspended in TBE-Urea sample buffer (Invitrogen) and resolved by 20% denaturing (8 M urea) PAGE. Gels were visualized by autoradiography. Cleavage bands were quantified using ImageQuant and normalized against total lane radioactivity.



**Figure 1.** PRC2 subunit permutations and their affinities for HOTAIR. (a) Schematic of subunit permutations tested in this study. In all cases EZH2 is N-FLAG tagged and EED is N-6×His tagged. (b) Representative gel showing EMSAs for EZH2-EED and PRC2 3m. (c) Binding curves as determined by quantitative EMSA. EZH2-EED exhibits a very similar affinity for HOTAIR as PRC2 3m. Error bars represent the standard error of the mean (S.E.M.) calculated from three replicates.

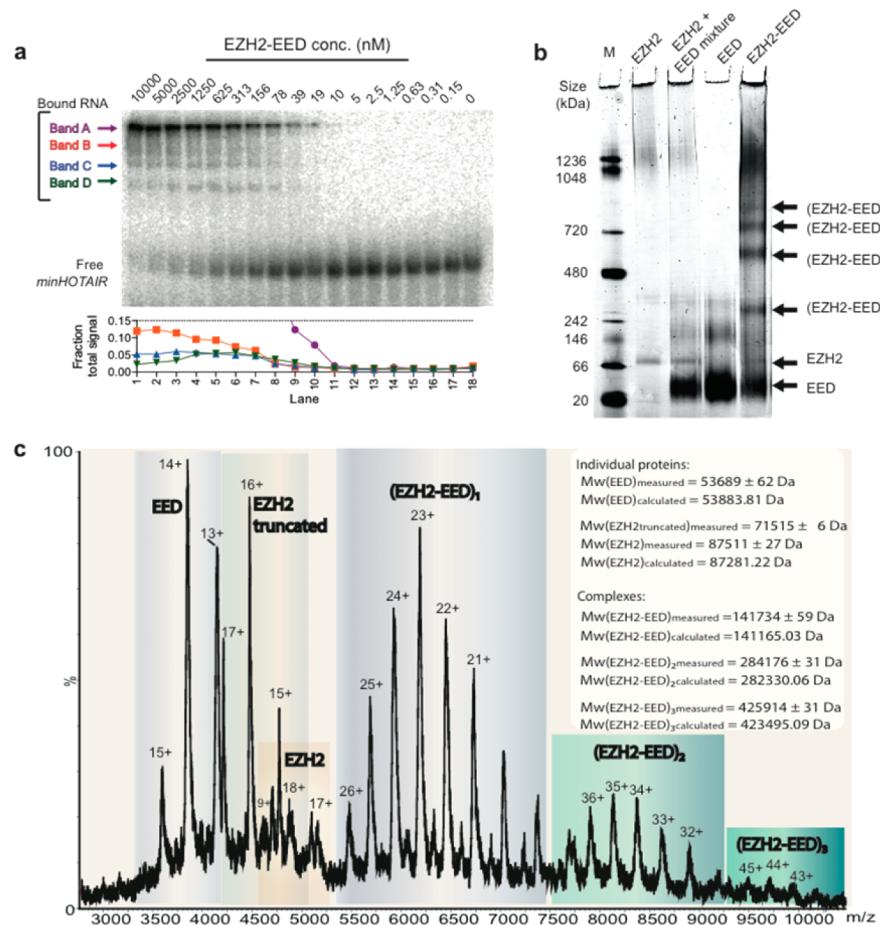


**Figure 2.** HOTAIR minimization and its effects on EZH2-EED binding. (a) Schematic of HOTAIR sections used in the initial part of deletional study. (b) Binding curves as determined by quantitative EMSA. 201–300 exhibits the same affinity for EZH2-EED as the 5' 300-mer, suggesting this RNA contains a minimal PRC2 interacting domain. (c) Further deletion gives 212–300 as the minimal PRC2 interacting domain of HOTAIR. 212–300/minHOTAIR binds with near identical affinity to EZH2-EED and PRC2 3m, while showing no affinity for LSD1. Further deletion of sections at the 5' or 3' end of minHOTAIR results in loss of affinity for EZH2-EED. Error bars represent the SEM calculated from three replicates.

## RESULTS AND DISCUSSION

**EZH2-EED is the Minimal HOTAIR Interacting Unit of PRC2.** We began by measuring binding affinities between HOTAIR and defined subunit permutations of PRC2 3m to determine which subunits were required for binding. A quantitative Electrophoretic Mobility Shift Assay (EMSA) was used to determine the extent of protein-RNA complex formed for each subunit permutation over a range of protein concentrations.

Five permutations of the PRC2 3m complex were produced using baculovirus infection of insect cells (Figure 1a). The various combinations of PRC2 3m subunits were each affinity purified using either an N-Terminal FLAG tag present on EZH2 or an N-terminal 6×His tag on EED. Correct expression and purification of PRC2 3m permutations was checked by SDS-PAGE and Western blotting (SI Figure 1). Dimer and trimer complexes were produced by coexpression and purified together, rather than mixing of constituent monomers after purification. Phosphorylation at T345/T350 of EZH2 (in



**Figure 3.** EZH2-EED forms discrete oligomers. (a) Representative gel of minHOTAIR and EZH2-EED. Multiple EMSA bands are indicative of multiple stoichiometry of binding, including higher order complexes which are too large to enter the gel matrix. The fraction of highest mobility shifted band D (green line) increases initially, but then decreases at high protein concentrations, indicating higher order complex formation is dependent on protein concentration. (b) Native gel of EED, EZH2, EZH2 + EED mixture, and EZH2-EED heterodimer. Multiple bands are observed for EZH2-EED, which are not present in the mixture of monomers. Deviation from expected mass for EZH2-EED compared to ladder is likely due to deviation from globularity for the protein shape, or nonideal binding of the G-250 charge shift reagent used for native gel electrophoresis. (c) Mass spectra for intact EZH2-EED showing presence of discrete oligomeric EZH2-EED complexes. The main species is heterodimeric (EZH2-EED)<sub>1</sub>, which appears at 5300–6700 m/z region with charges 22<sup>+</sup>–26<sup>+</sup>. Higher oligomers (EZH2-EED)<sub>2</sub> and (EZH2-EED)<sub>3</sub> appear above 7500 m/z. Some dissociation of EZH2-EED into constituent EZH2 and EED monomers is also observed, as well as some fragmentation of the EZH2 subunit (also observed in SI Figure 1). Under the native mass spectrometry conditions, (EZH2-EED)<sub>1</sub> appears to be the main heterodimer component of the mixture.

mouse/human respectively) has been reported to enhance affinity of this subunit for lncRNA binding.<sup>13</sup> Western blotting indicated this modification was not present in the PRC2 3m permutations used in our studies (SI Figure 2). For the RNA binding partner we used the 300-mer 5' domain of HOTAIR previously established to be necessary and sufficient to bind PRC2 3m,<sup>14</sup> as the size of the full HOTAIR lncRNA (2.2 kb) rendered it intractable to study by EMSA.

Quantitative EMSA analysis showed robust binding between EZH2 alone and the HOTAIR 5' domain with  $K_d = 755 \pm 43$  nM, in agreement with previous reports suggesting EZH2 to be a subunit responsible for lncRNA interaction (although these studies did not report a binding affinity).<sup>13,16</sup> EZH2-SUZ12 was found to bind the HOTAIR 300-mer with comparable affinity to EZH2 alone ( $K_d = 755 \pm 43$  nM for EZH2 vs  $700 \pm 158$  nM for EZH2-SUZ12), indicating that the SUZ12 subunit does not significantly contribute to PRC2 3m interaction with the HOTAIR 5' 300-mer.

We observed a ~5-fold increase in HOTAIR affinity for the PRC2 3m complex ( $K_d = 165 \pm 16$  nM) compared to EZH2-

alone or EZH2-SUZ12, suggesting that the EED subunit is required to stabilize the interaction between lncRNA and protein. Of all subunit permutations tested, only EZH2-EED showed a comparable strength of interaction for HOTAIR as PRC2 3m ( $K_d = 165 \pm 16$  nM for PRC2 vs  $147 \pm 9$  nM for EZH2-EED). These data show that the EZH2-EED heterodimer represents the minimal component of PRC2 necessary for HOTAIR binding (Figure 1b,c). Because EZH2 alone binds to HOTAIR with a much greater affinity than EED alone ( $K_d = 755 \pm 43$  nM for EZH2 vs >5  $\mu$ M for EED), these data also suggest that EZH2 is likely to be the main HOTAIR binding component within the EZH2-EED heterodimer. EED may also play a direct role in HOTAIR binding, or alternatively, act by modulating EZH2 to increase its affinity for the RNA.

**An 89-mer Section of HOTAIR is the Minimal PRC2 Interacting Domain.** We next investigated the contribution of the RNA partner in the HOTAIR-PRC2 interaction, and minimization of the previously reported 300-mer domain. Using a deletional approach, we tested progressively shorter sections of the HOTAIR RNA for their ability to bind EZH2-

EED using EMSA. We reasoned that a minimal RNA domain that interacted with PRC2 3m in the same fashion as the previously reported 300-mer should display the same affinity for EZH2-EED or PRC2 as the longer RNA.

In vitro transcription was used to generate oligonucleotides corresponding to different sections of HOTAIR (Figure 2a and SI Figure 3), which were then assessed for their ability to bind the EZH2-EED heterodimer via EMSA. We found that the 200-mer 101–300 show a similar affinity for EZH2-EED as the full 300-mer 5' domain ( $K_d = 119 \pm 18$  nM for 101–300 vs  $147 \pm 9$  nM for the 300-mer). The 200-mer 1–200 showed only weak affinity for EZH2-EED, with binding not saturating even up to  $5 \mu\text{M}$  protein, implying that the EZH2-EED binding domain resided partially or completely within 201–300 (Figure 2b).

A second round of minimization using 100-mer RNA sections found 101–200 to possess only very weak EZH2-EED binding capability (no saturation up to  $5 \mu\text{M}$ ), whereas 201–300 retained full affinity for EZH2-EED ( $K_d = 157 \pm 12$  nM), strongly suggesting the protein binding domain resided within this final 100 nt of the HOTAIR 5' 300-mer domain. As expected from the 200-mer EMSAs, 1–100 showed no affinity for EZH2-EED.

Further deletion of 201–300 gave the 89-mer 212–300 as the shortest section of HOTAIR which could bind EZH2-EED with the same affinity as the full 300-mer (Figure 3c –  $K_d = 135 \pm 20$  nM for 212–300 vs  $147 \pm 9$  nM for the 300-mer). EMSA using 212–300 and the PRC2 3m complex showed a similar affinity to that found for EZH2-EED binding ( $K_d = 111 \pm 10$  nM for PRC2 vs  $135 \pm 20$  nM for EZH2-EED), indicating that the binding interaction was not an artifact arising from the more minimal EZH2-EED heterodimer. The very similar EZH2-EED/PRC2 3m affinity observed between 212–300 and longer oligomers suggests that this domain is readily available in the context of longer RNA sequences. As a negative binding control, the histone demethylase LSD1, known to interact with the full HOTAIR lncRNA at a 600-nt 3' domain,<sup>14</sup> showed no interaction with 212–300, suggesting that this 89-mer RNA is specific for PRC2 proteins. Deletion of further nucleotides either at the 5' or 3' ends of 212–300 significantly reduced affinity for EZH2-EED, thus showing that this 89-mer (hereafter referred to as minHOTAIR) represents the minimal PRC2 interacting domain of HOTAIR.

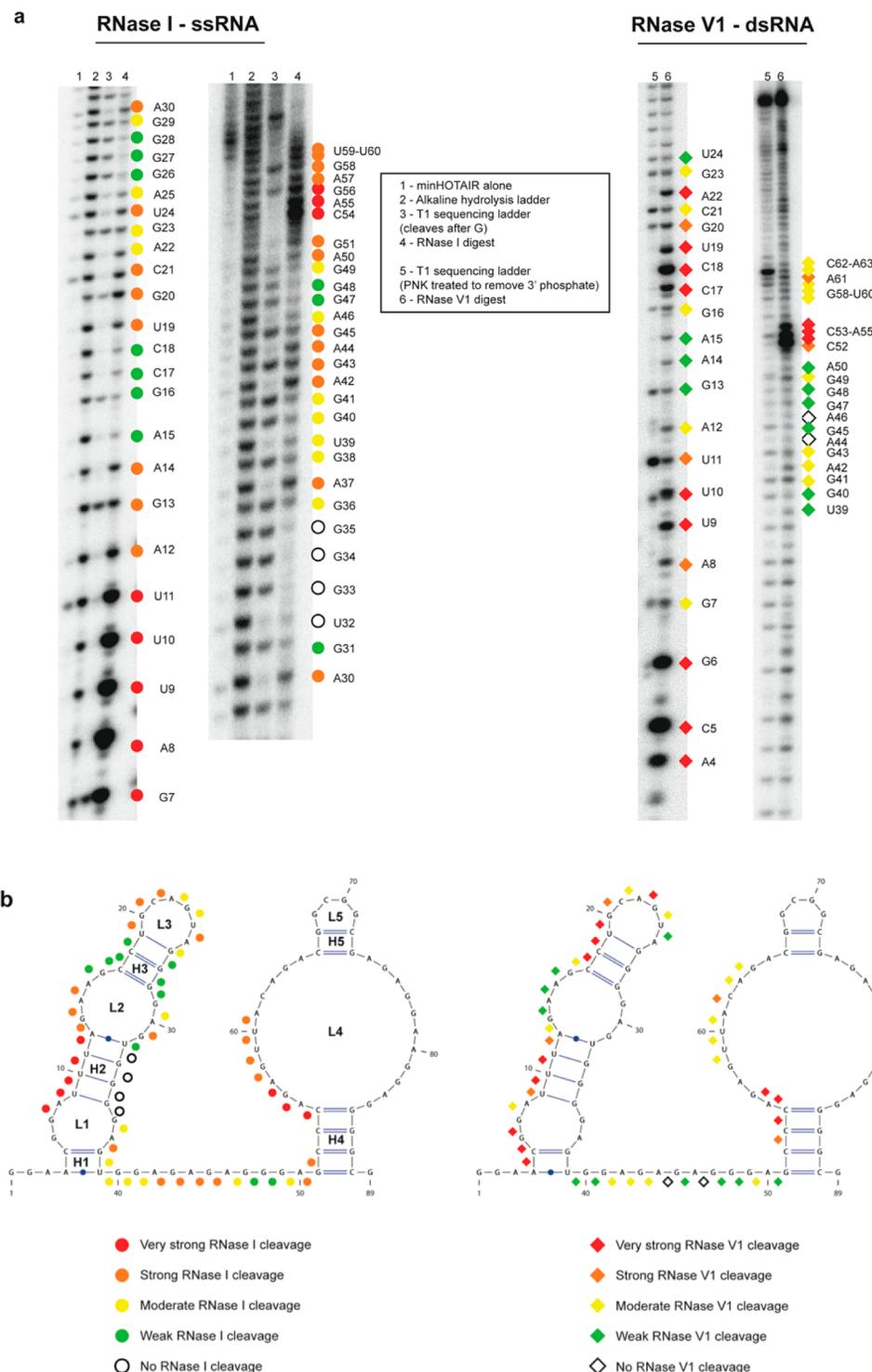
**EZH2-EED Forms Discrete Oligomers.** Reducing the length of lncRNA allowed us to use higher percentage gels for EMSA, giving increased resolution and allowing mobility differences between different protein-RNA stoichiometric species to be seen. This was not possible using the large 300 nt HOTAIR 5' domain RNA. We observed multiple upper bands in all EMSAs with either 201–300 or minHOTAIR along with EZH2-EED or PRC2 3m (Figure 3a and SI Figure 4). The design of our EMSA experiments, with the RNA component held at a very low concentration (~1 nM) compared to protein, suggested that these higher order species most likely corresponded to multiple proteins to RNA, rather than multiple RNA molecules associated with a single protein. In all EMSAs, the lowest mobility protein bound band appears to predominate among shifted complexes, suggesting a strong preference for formation of higher oligomers. In contrast, the fraction of the highest mobility shifted band was found to decrease at higher concentrations of protein (Figure 3a, Band D), consistent with a concentration dependent effect on the extent of oligomerization.

Because higher order complexes were visible for both EZH2-EED and PRC2 3m EMSAs, these results indicate that this oligomerization property arises from the EZH2-EED heterodimer, or one of its constituent subunits. To test for the presence of higher order complexes, we performed native protein gel electrophoresis on EZH2-EED, as well as EZH2 alone, EED alone, and a 1:1 stoichiometric mixture of the two monomers (Figure 3b). Native gels were carried out in the absence of minHOTAIR, in order to assess whether oligomerization was dependent on the presence of an RNA binding partner.

Using native gel electrophoresis, we observed the presence of higher order EZH2-EED complexes, with oligomers up to (EZH2-EED)<sub>4</sub> visible on the gel. Oligomers formed in the absence of minHOTAIR, suggesting this property is mediated by protein–protein interactions. Both EZH2 alone and EED alone appear on the gel as single main bands, with weaker minor bands at higher mass likely corresponding to impurities. Notably, the mixture of EZH2 and EED monomers did not show the same pattern of bands as observed for EZH2-EED, indicating that oligo-EZH2-EED formation cannot be recapitulated by simple mixing ex vivo, as might be expected for random aggregation, but is dependent on correct coexpression of the two proteins. Addition of minHOTAIR to EZH2-EED caused a small but reproducible upshift in all EZH2-EED bands on the native gel, but did not alter the proportion of bands, suggesting the RNA is able to bind all oligomeric states of the complex but does not affect the extent of oligomerization (SI Figure 5).

In order to confirm observations from native gels, we also carried out native mass spectrometry on EZH2-EED. The mild electrospray ionization used in this technique allows for large protein complexes to be detected without fragmentation. We were able to detect several oligomeric species for EZH2-EED: the heterodimer (EZH2-EED)<sub>1</sub> was observed as the main component under native mass spectrometry conditions, with progressive decrease in signal intensity for (EZH2-EED)<sub>2</sub> and (EZH2-EED)<sub>3</sub> higher order complexes (Figure 3c). Native mass spectrometry of the EED subunit alone did not show any oligomerization, further indicating oligomerization is specific to the EZH2-EED heterodimer (SI Figure 6). Oligo-EZH2-EED complexes were always observed as multiples of a 1:1 heterodimer, further suggesting these oligomeric species are likely to be functional, rather than simply aggregates of EZH2 and EED which may form in any ratio.

Altogether, our data show that oligomerization is intrinsic to EZH2-EED, and is likely to be a functional property of the heterodimer, rather than the result of nonspecific aggregation. Interestingly, there are many well-documented examples of P<sub>c</sub>G proteins clustering as a mechanism for spreading chromatin silencing. One important example found in many cell types is the formation polycomb bodies,<sup>17</sup> large, dense regions rich in P<sub>c</sub>G proteins, which act as silencing domains catalyzing heterochromatin formation along large portions of the genome. X-chromosome inactivation, which begins with initial recruitment of PRC2 by lncRNA *Xist*, also involves clustering of PRC2 to create a P<sub>c</sub>G rich silent X-chromosome.<sup>18</sup> More recently, a study mapping lncRNA binding sites across the genome found HOTAIR to nucleate PRC2 binding on chromatin: HOTAIR peaks on chromatin appear as sharp foci, around which PRC2 spreads to form wider polycomb domains.<sup>19</sup> Our observation that EZH2-EED can



**Figure 4.** RNase I and V1 structural probing minHOTAIR. (a) Representative autoradiogram of RNase I and V1 mapping for minHOTAIR. Colored circles represent extent of nuclease cleavage for RNase I. Colored diamonds represent extent of nuclease cleavage for RNase V1. Because the RNA is 5' radiolabeled, only the 5' cleavage point for a RNase V1 cleaved duplex will be visualized on the sequencing gel. Note that PNK treatment of RNase T1 digest (lane 5) results in appearance of several spurious bands which do not correspond to a G in minHOTAIR. (b) RNAfold predicted structure of minHOTAIR with RNase I and V1 digestion pattern overlaid. Helices and Loops are annotated on the left diagram: H - Helix, L - Loop.

form higher order oligomers is in good agreement with all of these previous findings.

**Minimal PRC2 Binding Element of HOTAIR is Highly Structured.** We next carried out nuclease mapping experiments to determine the secondary structure of minHOTAIR. Nuclease digestion can be used to determine RNA 2D structure

by identifying single and double stranded regions of a folded RNA molecule. Nuclease mapping has previously been used as part of a strategy to map the *Xist* RepA lncRNA, suggesting a complex 2D structure distinct from the dual hairpin loops which had previously been suggested for *Xist* RepA.<sup>20</sup> We have also previously reported a similar approach used to probe the

structure of the miRNA precursor pre-Let-7g.<sup>21</sup> For mapping experiments on minHOTAIR we utilized two nucleases with contrasting specificities: RNase I, which cleaves ssRNA nucleotides regardless of base, and RNase V1, which is specific for duplex RNA.

Nuclease mapping of minHOTAIR revealed several regions protected from RNase I, and strongly cleaved by RNase V1, implying minHOTAIR to be a highly structured domain of the HOTAIR lncRNA (Figure 4a). Inputting these nuclease mapping patterns as constraints into the structure prediction program RNAfold<sup>22–24</sup> produced a 2D structure comprising two duplex containing regions (G1-U39 and G51-G89), connected by a 10 nt ssRNA linker (G40-A50) (Figure 4b). This structure was in good agreement with both RNase I and V1 digest patterns, with the exception of strong RNase I cleavage observed at the H2 helix, and V1 cleavage in some predicted ssRNA regions (L3 and part of L4). Notably, this proposed structure is distinct from the small dual-hairpin structures previously predicted by folding algorithms for some PRC2 interacting ncRNAs,<sup>16,25</sup> demonstrating the value of experimental mapping for RNA secondary structure determination.

RNase I cleavage at H2 may be due to conformational flexibility of the RNA, resulting in a dynamic equilibrium between the predicted duplex structure and one where the bases are single stranded. We note that RNase V1 also cleaves strongly at H2, consistent with the predicted duplex at this region, and supporting the hypothesis that bases in H2 can interconvert between ssRNA and duplex conformations. V1 cleavage observed in predicted single stranded regions may also be the result of a dynamic conformational equilibrium, or alternatively may indicate the presence of tertiary interactions between predicted ssRNA bases and other regions of minHOTAIR.

Alongside the structural mapping, we also carried out <sup>1</sup>H NMR spectroscopic analysis of minHOTAIR, to study conformational heterogeneity within the RNA structure. Interestingly, we observed a small degree of K<sup>+</sup> dependent Hoogsteen interaction within the <sup>1</sup>H NMR spectrum of minHOTAIR indicating a possible alternative G-quadruplex structure (SI Figure 7a,b). In the absence of K<sup>+</sup>, the <sup>1</sup>H NMR spectrum displayed well-resolved peaks indicative of a predominant single species in solution. G-quadruplexes are noncanonical nucleic acid structures which can form in G-rich nucleic acids. A quadruplex comprises two or more stacked G-tetrads: planar arrangements of four Hoogsteen hydrogen bonded guanine bases, which are further stabilized by coordination to a central monovalent cation.<sup>26</sup> The sequence of minHOTAIR contains a G-rich region which closely resembles a putative quadruplex forming sequence (G26-G49),<sup>27</sup> although one of the G tracts is interrupted, which would create a bulge in any resulting quadruplex.<sup>28</sup>

Because G-quadruplexes are stabilized by monovalent cations (usually K<sup>+</sup>), increased Hoogsteen interactions resulting from addition of K<sup>+</sup> is strongly indicative of formation of this noncanonical structure.<sup>29</sup> Comparison of peaks in the NMR titration spectra suggests that only a very small proportion of minHOTAIR bases are involved in K<sup>+</sup> dependent Hoogsteen interactions, implying the quadruplex is in equilibrium with a more favored duplex structure (SI Figure 7c). Because of this, we suggest that a quadruplex is not likely to be significant for interaction between minHOTAIR and EZH2-EED. However, transient formation of a quadruplex provides additional

explanation for the strong RNase I digestion at H2 due to loss of the corresponding duplex.

Lastly, we carried out nuclease footprinting experiments to determine if EZH2-EED interacted with specific sites on the predicted minHOTAIR secondary structure. Nuclease footprinting is carried out in a similar fashion to the mapping experiments described previously, but digests are carried out in the presence of the RNA binding partner. Changes in the nuclease digest pattern resulting from addition of the binding partner are indicative of sites involved in the binding interaction. A decrease in digestion at a region of the RNA indicates the protein binds to this site, or alternatively that the RNA conformation changes so that it is no longer recognized by the nuclease. Similarly, sites of increased digestion suggest a change in RNA conformation upon protein binding, rendering the bases more susceptible to nuclease. Footprinting was carried out using the same nucleases as for the mapping experiments (RNase I and RNase V1), in the presence of 200 nM ( $\sim K_d$  as determined by EMSA) or 2  $\mu$ M ( $\sim 10 \times K_d$ ) EZH2-EED, as well as a no protein control.

Initial footprinting was carried out using the ssRNA specific nuclease RNase I. No changes were observed in the digestion pattern with RNase I, either at 200 nM or 2  $\mu$ M concentrations of EZH2-EED (SI Figure 8). This suggested that EZH2-EED may bind to regions of minHOTAIR which are already structured, and therefore resistant to RNase I even without protein present.

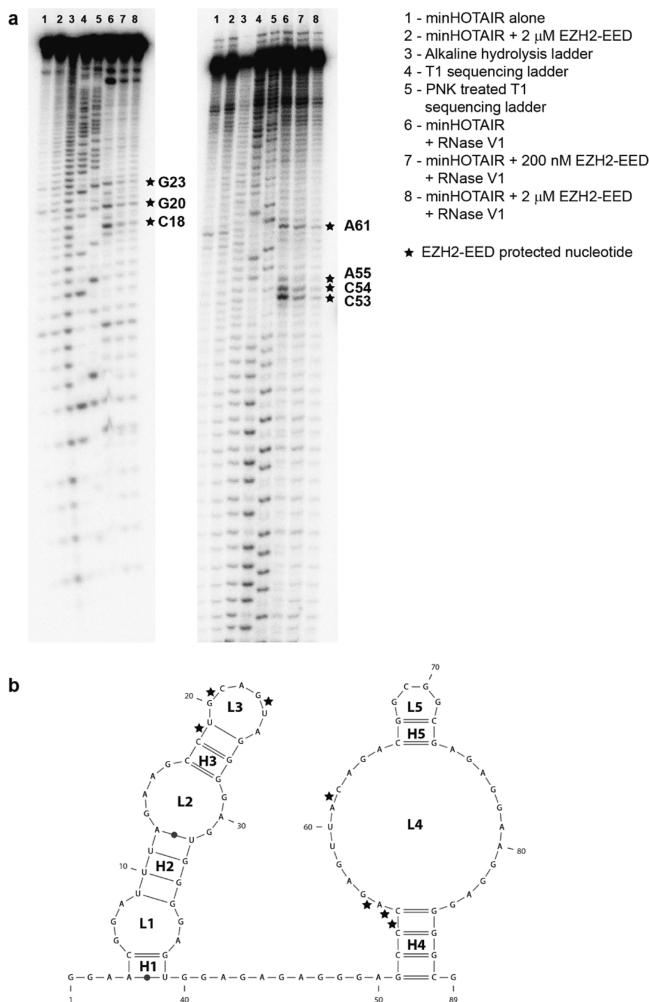
Using RNase V1, we detected protection from nuclease in two duplex regions of minHOTAIR—most strongly at the H4-L4 junction, but also around the hairpin loop L3 (Figure 5a,b). Because these two sites do not share any sequence or predicted structural homology, it is unlikely they correspond to two equivalent binding sites for the protein. These two protected regions may correspond to binding sites for the EZH2 and EED subunits which interact at different sites on the RNA. Alternatively, further tertiary interactions in the minHOTAIR RNA may bring these domains together to form a unified binding site for the EZH2-EED protein. Although the exact protein binding sites on minHOTAIR remain to be elucidated, these well-defined footprinting patterns we have observed indicate that EZH2-EED binding to minHOTAIR is likely to be mediated by a specific RNA motif, rather than the result of binding by nonspecific interactions.

## CONCLUSIONS

Herein, we have described the first systematic study of the minimal elements required for the interaction between lncRNA HOTAIR and the PRC2 3m complex. These results provide insight into the details of a highly important pathway involved in regulation of epigenetic silencing, which has been implicated in driving aggression in a wide range of cancers.<sup>10–12</sup>

We have shown that the heterodimer EZH2-EED exhibits the same binding affinity for HOTAIR as the core PRC2 3m complex, and that the 89 nt minHOTAIR RNA shows the same binding affinity to PRC2 3m as a previously reported 300 nt PRC2 binding RNA domain. Nuclease mapping and footprinting experiments revealed that the minHOTAIR RNA is highly structured, and that protein binding causes protection from RNase V1 at two separate regions of a predicted minHOTAIR secondary structure.

Previous PRC2 RNA immunoprecipitation studies have suggested ~20% of the lncRNA transcriptome is able to bind PRC2.<sup>5</sup> However, the molecular basis of these interactions is



**Figure 5.** RNase V1 footprinting of minHOTAIR interaction sites with EZH2-EED. (a) Representative autoradiograms of EZH2-EED footprinting with minHOTAIR. Gels were run for either 4 h (left) or 7 h to resolve relevant area of sequence. Stars represent sites of V1 protection upon binding of EZH2-EED to minHOTAIR. (b) Predicted structure of minHOTAIR with footprinted nucleotides (stars) overlaid.

not well understood. Interestingly, the minHOTAIR motif we have presented here differs significantly from the tandem dual hairpin motif first established in Xist RepA, which has been suggested to be necessary for lncRNA interaction with the PRC2 complex.<sup>16,25</sup> This lack of consensus may reflect the existence of several lncRNA-PRC2 binding modes, which may each correspond to distinct functions of the PRC2 complex. Experimentally verified minimal PRC2 binding motifs such as minHOTAIR provide a useful starting point for addressing such issues of interaction diversity. By using the proposed minHOTAIR secondary structure as a reference point in homology searches with other PRC2 interacting lncRNAs, it may be possible to identify the extent of RNA species that bind PRC2 in a 'HOTAIR-like' fashion. In this way, a comprehensive profile can gradually be established for the RNA binding capabilities of a key epigenetic regulator.

In this study, we also observed the presence of multiple stoichiometric protein:RNA species in EMSAs for both PRC2-3m and EZH2-EED with smaller sections of HOTAIR. This was shown to be a consequence of the EZH2-EED heterodimer clustering to form discrete higher order oligomers. These

observations correspond well to numerous literature reports of clustering behaviors in PcG mediated silencing.<sup>17-19</sup> An interesting hypothesis may be that the clustering often seen in PcG silencing systems is (in part) facilitated by an intrinsic ability of the EZH2-EED heterodimer to oligomerize. An intrinsic oligomerization activity would allow PRC2 complexes to rapidly spread from their initial recruitment sites on chromatin, and may be a mechanism for the cell to rapidly establish de novo epigenetic silencing via a protein based positive-feedback clustering mechanism.

To summarize, we have elucidated new details of the molecular interaction between lncRNA HOTAIR and the PRC2 complex. Given that HOTAIR is known to drive metastasis in a range of cancers by redirecting PRC2 silencing, this vital interaction may in due course prove to be a worthy target for intervention, with a view to reversing the malignant effects of HOTAIR overexpression. Further knowledge of the details of HOTAIR-PRC2 binding may help facilitate such efforts to target this RNA–protein interaction. Insights into the structural basis of HOTAIR-PRC2 binding may also be applicable to many other PRC2 interacting lncRNAs, helping to shed light on the interactions and mechanisms of an important class of epigenetic regulators.

## ■ ASSOCIATED CONTENT

## S Supporting Information

Supporting Information and methods, Supplementary Figures S1–S6, and Supplementary Tables S1, S2, and S3. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## AUTHOR INFORMATION

## Corresponding Author

\*Tel: +44 (0)1223 336347. E-mail: sb10031@cam.ac.uk.

## Funding

The Balasubramanian lab is supported by program funding from Cancer Research UK. LW is supported by a Cancer Research UK studentship.

## Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

The authors would like to thank Dr. Lovorka Stojic for helpful advice and comments throughout the course of this project.

## ■ REFERENCES

- (1) Willingham, A. T., Orth, A. P., Batalov, S., Peters, E. C., Wen, B. G., Aza-Blanc, P., Hogenesch, J. B., and Schultz, P. G. (2005) A strategy for probing the function of noncoding RNAs finds a repressor of NFAT. *Science* 309, 1570–1573.

(2) Poliseno, L., Salmena, L., Zhang, J., Carver, B., Haveman, W. J., and Pandolfi, P. P. (2010) A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature* 465, 1033–1038.

(3) Mercer, T. R., and Mattick, J. S. (2013) Structure and function of long noncoding RNAs in epigenetic regulation. *Nat. Struct. Mol. Biol.* 20, 300–307.

(4) Wang, K. C., and Chang, H. Y. (2011) Molecular mechanisms of long noncoding RNAs. *Mol. Cell* 43, 904–914.

(5) Khalil, A. M., Guttmann, M., Huarte, M., Garber, M., Raj, A., Rivea Morales, D., Thomas, K., Presser, A., Bernstein, B. E., van Oudenaarden, A., Regev, A., Lander, E. S., and Rinn, J. L. (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.

- (6) Cao, R., and Zhang, Y. (2004) SUZ12 is required for both the histone methyltransferase activity and the silencing function of the EED-EZH2 complex. *Mol. Cell* 15, 57–67.
- (7) Batista, P. J., and Chang, H. Y. (2013) Long noncoding RNAs: cellular address codes in development and disease. *Cell* 152, 1298–1307.
- (8) Tsai, M. C., Spitale, R. C., and Chang, H. Y. (2011) Long intergenic noncoding RNAs: new links in cancer progression. *Cancer Res.* 71, 3–7.
- (9) 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.
- (10) Gupta, R. A., Shah, N., Wang, K. C., Kim, J., Horlings, H. M., Wong, D. J., Tsai, M. C., Hung, T., Argani, P., Rinn, J. L., Wang, Y., Brzoska, P., Kong, B., Li, R., West, R. B., van de Vijver, M. J., Sukumar, S., and Chang, H. Y. (2010) Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* 464, 1071–1076.
- (11) Kogo, R., Shimamura, T., Mimori, K., Kawahara, K., Imoto, S., Sudo, T., Tanaka, F., Shibata, K., Suzuki, A., Komune, S., Miyano, S., and Mori, M. (2011) Long noncoding RNA HOTAIR regulates polycomb-dependent chromatin modification and is associated with poor prognosis in colorectal cancers. *Cancer Res.* 71, 6320–6326.
- (12) Geng, Y. J., Xie, S. L., Li, Q., Ma, J., and Wang, G. Y. (2011) Large intervening non-coding RNA HOTAIR is associated with hepatocellular carcinoma progression. *J. Int. Med. Res.* 39, 2119–2128.
- (13) Kaneko, S., Li, G., Son, J., Xu, C. F., Margueron, R., Neubert, T. A., and Reinberg, D. (2010) Phosphorylation of the PRC2 component Ezh2 is cell cycle-regulated and up-regulates its binding to ncRNA. *Genes Dev.* 24, 2615–2620.
- (14) 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.
- (15) Hernandez, H., and Robinson, C. V. (2007) Determining the stoichiometry and interactions of macromolecular assemblies from mass spectrometry. *Nat. Protoc.* 2, 715–726.
- (16) Zhao, J., Sun, B. K., Erwin, J. A., Song, J. J., and Lee, J. T. (2008) Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* 322, 750–756.
- (17) Pirrotta, V., and Li, H. B. (2010) A view of nuclear Polycomb bodies. *Curr. Opin. Genet. Dev.* 22, 101–109.
- (18) Wutz, A. (2011) Gene silencing in X-chromosome inactivation: advances in understanding facultative heterochromatin formation. *Nat. Rev. Genet.* 12, 542–553.
- (19) Chu, C., Qu, K., Zhong, F. L., Artandi, S. E., and Chang, H. Y. (2011) Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. *Mol. Cell* 44, 667.
- (20) Maenner, S., Blaud, M., Fouillen, L., Savoye, A., Marchand, V., Dubois, A., Sanglier-Cianferani, S., Van Dorsselaer, A., Clerc, P., Avner, P., Visvikis, A., and Branlant, C. (2010) 2-D Structure of the A Region of Xist RNA and Its Implication for PRC2 Association. *PLoS Biol.* 8, e1000276.
- (21) Lightfoot, H. L., Bugaut, A., Armisen, J., Lehrbach, N. J., Miska, E. A., and Balasubramanian, S. (2011) A LIN28-dependent structural change in pre-let-7g directly inhibits dicer processing. *Biochemistry* 50, 7514–7521.
- (22) Hofacker, I. L., Fontana, W., Stadler, P. F., Bonhoeffer, L. S., Tacker, M., and Schuster, P. (1994) *Monatsh. Chem.* 125, 167–188.
- (23) McCaskill, J. S. (1990) The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers* 29, 1105–1119.
- (24) Zuker, M., and Stiegler, P. (1981) Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucl. Acids Res.* 9, 133–148.
- (25) Kanhere, A., Viiri, K., Araujo, C. C., Rasaiyah, J., Bouwman, R. D., Whyte, W. A., Pereira, C. F., Brookes, E., Walker, K., Bell, G. W., Pombo, A., Fisher, A. G., Young, R. A., and Jenner, R. G. (2010) Short RNAs are transcribed from repressed polycomb target genes and interact with polycomb repressive complex-2. *Mol. Cell* 38, 675–688.
- (26) Neidle, S., and Balasubramanian, S., Eds. (2006) *Quadruplex Nucleic Acids*, RSC Biomolecular Sciences, Cambridge, U.K.
- (27) Huppert, J. L., and Balasubramanian, S. (2005) Prevalence of quadruplexes in the human genome. *Nucl. Acids Res.* 33, 2908–2916.
- (28) Mukundan, V. T., and Phan, A. T. (2013) Bulges in G-quadruplexes: broadening the definition of G-quadruplex-forming sequences. *J. Am. Chem. Soc.* 135, 5017–5028.
- (29) Bugaut, A., Murat, P., and Balasubramanian, S. (2012) An RNA hairpin to G-quadruplex conformational transition. *J. Am. Chem. Soc.* 134, 19953–19956.