# 博士論文

**Interactions between non-coding RNAs and proteins in mammals**

**(ほ乳類における非コードRNAとタンパク質の相互作用)**

ベタンクール メディナ フアン ギジェルモ

**Betancur Medina Juan Guillermo**

東京大学大学院新領域創成科学研究科

メディカルゲノム専攻

RNA機能研究分野

2014

# Contents

# 1. The mammalian non-coding transcriptome

In recent years it has become evident that in eukaryotes most of the genome is transcribed, and that a big portion of the transcriptome corresponds to non-coding RNAs (ncRNAs) that are found both in the nucleus and the cytoplasm. However, during many years only a few subclasses of ncRNAs were assigned functions, such as ribosomal RNAs (rRNAs) and transfer RNAs (tRNAs) that are involved in mRNA translation; small nucleolar RNAs (snoRNAs) that modify rRNAs; and small nuclear RNAs (snRNAs) that are involved in splicing (Birney et al., 2007; Carninci et al., 2005; Guttman et al., 2009; Mattick, 2004). Later, founding members of other novel classes of ncRNAs were also described, such as the microRNA (miRNA) *lin-4* in *C. elegans* that regulates the expression of the LIN-14 protein during development (Lee et al., 1993) and that eventually led to the discovery in mammals of miRNAs as well as other classes of short ncRNAs (sRNAs) including short interfering RNAs (siRNAs) and PIWI-interacting RNAs (piRNAs), among others. Also in the 1990s the H19 gene, an abundant fetal-specific hepatic transcript was described as the first mammalian gene whose product is a long non coding RNA (lncRNA) (Brannan et al., 1990).

Although there are a few ribozymes in mammals (Salehi-Ashtiani et al., 2006; Teixeira et al., 2004), most ncRNAs are catalytically inactive and require partner proteins to exert their functions, and many form intricate ribonucleoprotein (RNP) complexes, such as ribosomes and the telomerase complex.

Eukaryotic ribosomes, for example, contain around 80 proteins and are formed by 2 subunits, the small 40S and the large 60S subunits. In addition to proteins, the large subunit contains 3 rRNAs, 25S rRNA, 5S rRNA and 5.8S rRNA while the small subunit contains the 18S rRNA. Ribosomes must ensure that appropriate tRNAs are recruited according to the triplets in the mRNA and must also catalyze peptide bond formation between aminoacids during translation (Wilson and Doudna Cate, 2012). Another well characterized example of a RNP complex is the telomerase complex that catalyzes the addition of DNA repeats at the ends of linear chromosomes. The core components of the telomerase RNP complex are telomerase reverse transcriptase (TERT) and telomerase RNA (TER), an RNA template for the addition of the DNA repeats (Mason et al., 2011).

Among the more recently discovered ncRNA classes, miRNAs are ~22nt long

RNAs present mainly in the cytoplasm that form an RNP complex known as the RNA induced silencing complex (RISC) that has an Argonaute (Ago) protein at its core. In contrast, lncRNAs are a group of ncRNAs longer than 200nt that have complex secondary and tertiary structures. lncRNAs play many different functions, such as regulation of translation (Carrieri et al., 2012; Yoon et al., 2012), acting as decoys for proteins or other ncRNAs (Bernard et al., 2010; Poliseno et al., 2010; Tripathi et al., 2010), or as architectural elements of nuclear structures (Clemson et al., 2009; Chen and Carmichael, 2009; Sasaki et al., 2009; Sunwoo et al., 2009), among others. Therefore, they do not share a common intracellular location, and many lncRNAs interact with distinct arrays of proteins (and other RNAs) according to their individual functions. A subgroup of lncRNAs is involved in chromatin remodeling and some of them directly interact with various DNA or histone modifying complexes (Bertani et al., 2011; Nagano et al., 2008; Pandey et al., 2008; Sado et al., 2005; Sun et al., 2006; Wang et al., 2011; Yang et al., 2011) (Figure 1).

Since protein-RNA interactions are fundamental for the function of many ncRNAs, understanding of ncRNA interaction networks is fundamental to comprehend the role of ncRNA within more complex regulatory networks.

This work is divided in two parts. First I focus on the interaction between miRNAs and a protein called Dicer during RISC assembly, which ultimately leads to the posttranscriptional regulation of thousands of genes in mammalian cells. Second I analyze the relationship between two well known lncRNAs, Xist and HOTAIR, and the Polycomb-group (PcG) repressive complex 2 (PRC2), that together participate in the regulation of the inactivation of the X chromosome in female mammals, and of Hox gene expression during embryonic development and differentiation, respectively.

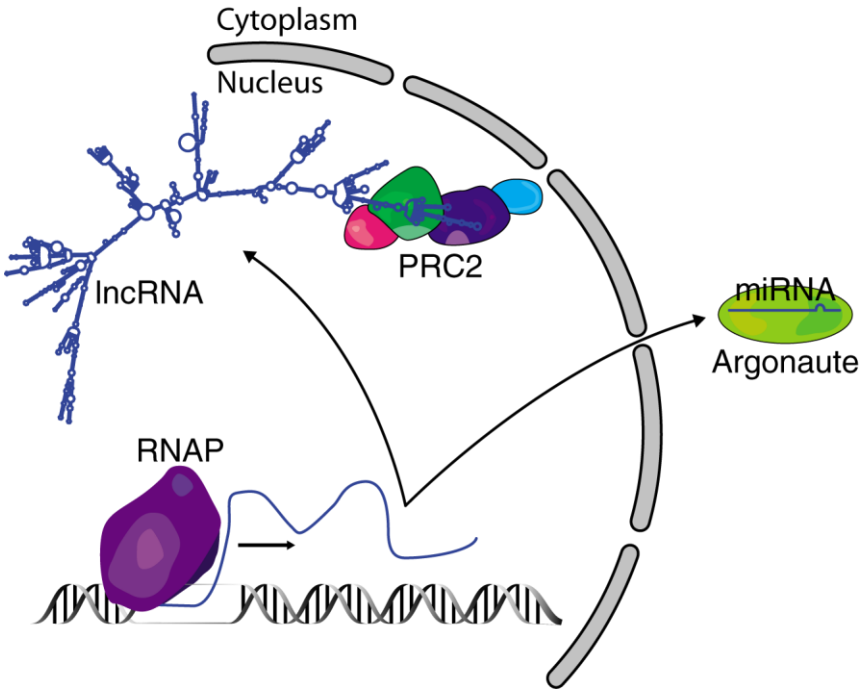



**Figure 1.** Multiple classes of ncRNAs are encoded in the mammalian genome. This work focuses on the protein-RNA interactions that lead, in the cytoplasm, to the assembly of RISC that has an Ago protein at its core; and in the nucleus, to the formation of RNP complexes between some lncRNAs and PRC2.

# 2. Part 1: Dicer is dispensable for asymmetric RISC loading in mammals

### 2.1 Small RNA biogenesis and RISC assembly

MicroRNAs are a class of small ncRNAs that fine tune the expression of genes at post-transcriptional level. It is estimated that around half of human genes are regulated by miRNAs and that they are involved in most biological processes. MicroRNAs are encoded in the genome of most eukaryotic organisms and are transcribed mainly by RNAPII to produce primary transcripts with stem loop structures up to several kilobases long. In the canonical processing pathway primary transcripts are first processed in the nucleus by the Microprocessor complex that contains Drosha, a type III RNase, and its cofactor DiGeorge syndrome critical region 8 (DGCR8). The distal end of the pre-miRNA stem loops is cleaved by Drosha, which results in the production of a ~65nt long pre-miRNA that is then exported to the cytoplasm by Exportin 5 (Exp5) (Ghildiyal and Zamore, 2009; Kim et al., 2009b). Once in the cytoplasm pre-miRNAs are recognized by a protein called Dicer that processes the end closer to the loop of the RNA to produce a miRNA/miRNA* duplex (Figure 2). Dicer possesses several functional domains: an N-terminal helicase domain, a domain with unknown function (DUF283), a PAZ domain, two RNase III domains, and a double stranded RNA (dsRNA)-binding domain (Figure 6A). In flies, there are two Dicer paralogs (Lee et al., 2004). Dicer-1 (Dcr-1) processes pre-miRNAs into miRNA/miRNA* duplexes (Lee et al., 2004; Tsutsumi et al., 2011), while Dicer-2 (Dcr-2) processes long dsRNAs to produce siRNA duplexes (Cenik et al., 2011; Lee et al., 2004; Welker et al., 2010), which can also be artificially introduced into cells. In contrast, mammals have a single Dicer that processes both pre-miRNAs and long dsRNAs (Provost et al., 2002; Zhang et al., 2002). In all cases fully processed RNA duplexes have 3’ overhangs and are phosphorylated at their 5’ends.

Dicer proteins often have dsRNA-binding proteins as their partners: Loquacious (Loqs) for fly Dcr-1 (Forstemann et al., 2005; Jiang et al., 2005; Saito et al., 2005), R2D2 for fly Dcr-2 (Liu et al., 2003), and TAR-binding protein (TRBP) or PKR activator (PACT) for mammalian Dicer (Chendrimada et al., 2005; Gregory et al., 2005; Haase et al., 2005; Lee et al., 2006; Maniataki and Mourelatos, 2005).

After processing by Dicer, small RNA duplexes are assembled into an effector complex known as RNA-induced silencing complex (RISC), that has an Argonaute

(Ago) protein at its core (Hammond et al., 2001) (Figure 2). In flies, miRNA/miRNA* and siRNA duplexes are preferentially sorted into two Ago paralogs, Ago1 and Ago2, respectively (Forstemann et al., 2007; Okamura et al., 2004; Tomari et al., 2007) (Figure 9A, B). In contrast, mammals lack such sorting mechanism and all four Ago proteins (Ago1–Ago4) can incorporate both miRNA/miRNA* and siRNA duplexes (Liu et al., 2004; Meister et al., 2004; Yoda et al., 2010) (Figure 9C).

RISC assembly can be divided into at least two steps (Kawamata and Tomari, 2010) (Figure 2). First, a small RNA duplex is loaded into the Ago protein to form pre-RISC. It is thought that small RNA (sRNA) duplexes are too bulky to fit into Ago, and that therefore a conformational change in the protein mediated by the Hsc70/Hsp90 chaperone machinery and the consumption of ATP is required for RISC loading (Iki et al., 2010; Iwasaki et al., 2010; Miyoshi et al., 2010) (Figure 3). Second, the duplex is unwound and one of the two strands (guide strand) is retained in Ago while the other strand (passenger strand) is discarded for degradation, resulting in the formation of mature RISC. Therefore, only one of the strands remains incorporated in RISC and guides the complex to target mRNAs that are recognized mainly through base pairing with the seed region of the guide strand, usually nucleotides 2-8 from its 5’ end (Doench and Sharp, 2004; Lewis et al., 2003). RISC regulates the expression of the target mRNAs by target cleavage, translational repression and/or deadenylation in a sequence specific manner (Figure 2).

The selection of which strand serves as the guide is not random but often asymmetric (Khvorova et al., 2003; Schwarz et al., 2003). In flies and mammals, it is widely accepted that the thermodynamic asymmetry of the ends of a small RNA duplex is a key factor that determines which of the two strands of the duplex serves as the guide. In general, the strand with the less stable 5’ end tends to serve as the guide strand, while the other strand with the more stable 5’ end is more likely discarded from the Ago protein during unwinding (Khvorova et al., 2003; Schwarz et al., 2003) (Figure 4). Importantly, the strand serving as the guide is already determined by the polarity of small RNA duplexes upon loading, before unwinding actually occurs (Kawamata and Tomari, 2010). In flies, the Dcr-2/R2D2 heterodimer, which is essential for Ago2-RISC assembly (Liu et al., 2003; Pham et al., 2004; Tomari et al., 2004a), senses the thermodynamic asymmetry of the siRNA duplex

(Tomari et al., 2004b). R2D2 orients siRNA duplexes by binding to the more stable end, which positions Dcr-2 at the opposite, less stable, end. The heterodimer then loads the siRNA duplex into fly Ago2 with the prearranged orientation (Tomari et al., 2004b) (Figure 5). Thus, Dcr-2/R2D2 has two distinct functions: dicing of long dsRNAs and asymmetric loading of siRNA duplexes into Ago2.

By analogy to the functions of fly Dcr-2/R2D2 in thermodynamic asymmetry sensing and Ago2-RISC loading, it is tempting to postulate that Dcr-1 and Loqs in flies and Dicer and TRBP (or PACT) in mammals might play similar roles in RISC assembly, in addition to their essential role in dicing (Chendrimada et al., 2005; Gregory et al., 2005; MacRae et al., 2008; Maniataki and Mourelatos, 2005; Miyoshi et al., 2009; Sakurai et al., 2011). Indeed, it was recently shown that, much as fly Dcr-2/R2D2, recombinant human Dicer/TRBP can bind to siRNA duplexes according to their thermodynamic asymmetry (Gredell et al., 2010; Noland et al., 2011). However, from a functional point of view, immunodepletion of human Dicer from HeLa lysate to an undetectable level did not hinder target cleavage activity of siRNAs (Martinez et al., 2002), and the efficiency of target gene silencing by siRNAs was apparently uncompromised in two independent lines of Dicer-knockout mouse embryonic stem (ES) cells (Kanellopoulou et al., 2005; Murchison et al., 2005). Moreover, siRNA-initiated canonical assembly of human Ago2-RISC was recently reconstituted in vitro without Dicer and TRBP (Ye et al., 2011). These accumulating observations raise the question of whether or not Dicer/TRBP binding plays a role in asymmetric RISC loading in mammals. This work shows that, in live cells and in vitro, Dicer is dispensable for asymmetric RISC assembly in mammals.

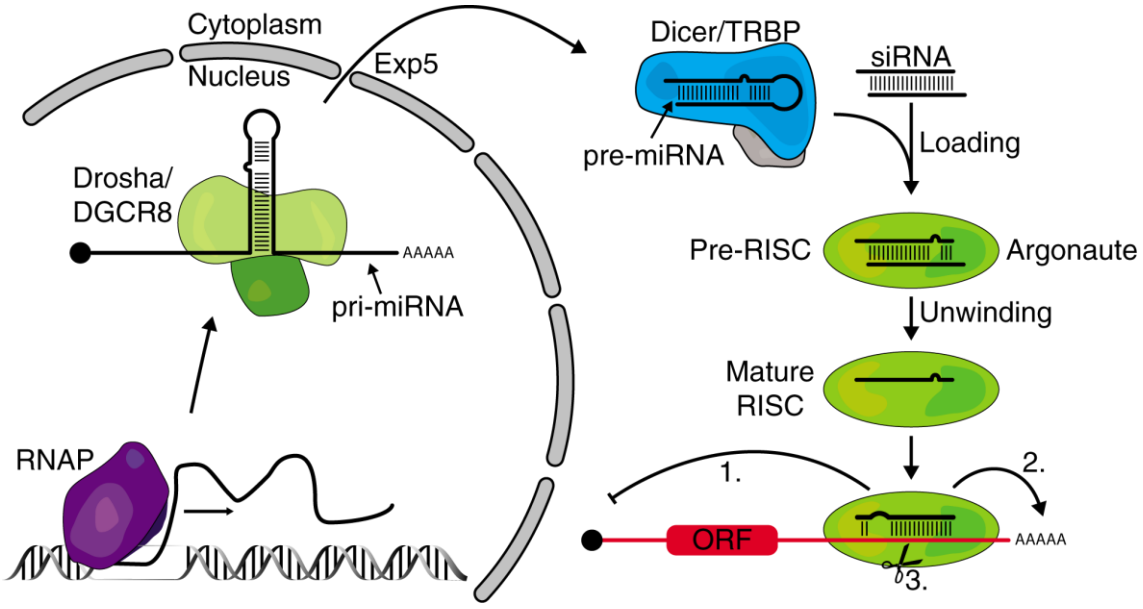



**Figure 2.** Scheme of miRNA biogenesis and RISC assembly in mammals. Mature RISC targets mRNAs and regulates their expression by (1) translational repression, (2) deadenylation, or (3) target cleavage. See text for further details.

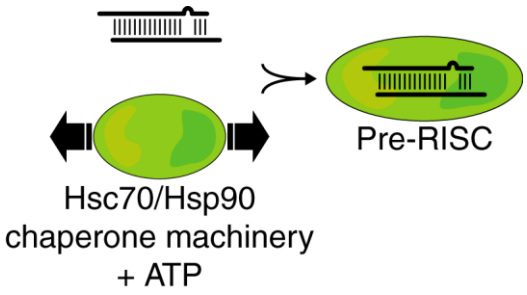



**Figure 3.** RISC loading requires the Hsc70/Hsp90 chaperone machinery. In flies and mammals the Hsc70/Hsp90 chaperone machinery is believed to induce a conformational change of Ago and to aid loading of the sRNA duplex to form pre-RISC.

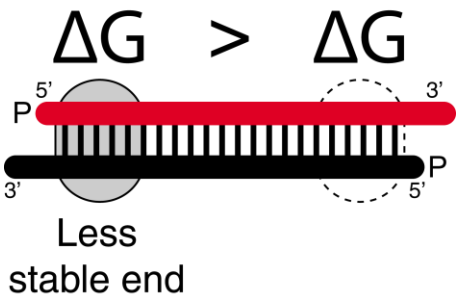



**Figure 4.** Small RNA duplexes are functionally asymmetric due mainly to the thermodynamic asymmetry of the ends of the duplex. The strand with its 5' end towards the less stable end of the duplex (ΔG>) has a higher probability to remain incorporated in RISC and function as guide strand (red). The opposite strand (black) is discarded and is known as passenger strand. Both strands of small RNA duplexes are phosphorylated at their 5' ends.

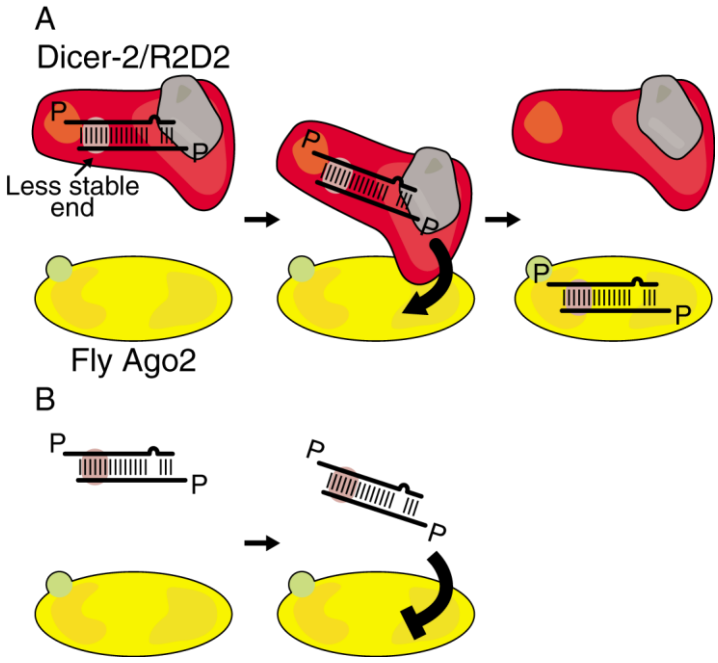



**Figure 5.** In flies the Dcr-2/R2D2 heterodimer is required for Ago2-RISC loading. A. Dcr-2/R2D2 binds siRNA duplexes asymmetrically and hands them over to Ago2 with the prearranged orientation. B. In the absence of Dcr-2/R2D2 no RISC loading occurs.

## 2.2 Materials and methods

### 2.2.1 General reagents

1× lysis buffer: 30 mM HEPES-KOH pH 7.4, 100 mM potassium acetate, 2 mM magnesium acetate.

40× reaction mix (does not correspond to 40× concentration, but to the volume necessary for 40 standard reactions): 133 mM potassium acetate, 9.33 mM magnesium acetate, 1.7 mM DTT, 3.33 mM ATP, 0.33 U/µl, 83.33 mM creatine monophosphate, 0.1 U/µl creatine phosphokinase in ultrapure water, in a final volume of 120 µl.

### 2.2.2 Plasmids constructions

For the construction of the mouse DICER1 expression vector (pCAGEN-DICER1), the *Dicer1* sequence was amplified from pBluescript II KS(–)-Dicer (Doi et al., 2003) and inserted in the *Eco*RI site of the pCAGEN vector using the In-Fusion Advantage PCR cloning kit (Clontech). The following primers were used.

| | | |
|---|---|---|
| Dicer: | Dcr-F | TTTTGGCAAAGAATTCGATGATTGAAAAGCCCTGCT |
| | Dcr-R | TATCCTCGAGGAATTCGTGGAGCTGTGGTTCTGGTC |

For the SBP-tagged DICER1 construct (pCAGEN-SBP-DICER1), the *Dicer1* and SBP sequences were amplified from pCAGEN-DICER1 or pASW (Iwasaki et al., 2010), respectively, and these two fragments were fused using the In-Fusion Advantage PCR cloning kit (Clontech). The following primers were used.

| | | |
|---|---|---|
| Dicer: | Dcr-F1 | GCAGGCTCCGCGGCCATGGCAGGCCTGCAGCTCAT |
| | Dcr-R1 | TATCCTCGAGGAATTCTCAGCTGTTAGGAACCTGAGGC |
| SBP: | SBP F | TTTTGGCAAAGAATTCCCATGGACGAGAAGACCACCGGC |
| | SBP R | GGCCGCGGAGCCTGCTTTTT |

To prepare target constructs for luciferase assays double stranded oligos containing perfectly complementary sequences to each strand of duplex A and

duplex B were synthesized (Hokkaido System Science) and cloned in the *Xho*I and *Not*I sites of the psiCHECK-2 vector (Promega).

### 2.2.3 Luciferase reporter assay

$4 \times 10^4$ *Dicer1*$^{-/-}$ mouse embryonic fibroblasts (MEF) cells (Yang et al., 2010; Yi et al., 2006) per well were seeded in 24-well plates and 24 h later were transfected with 0.8 µg of pCAGEN or pCAGEN-DICER1 using Lipofectamine 2000 according to the manufacturer's instructions. Twenty four hours after the initial transfection, cells were co-transfected with 0.25 µg of the target-bearing psiCHECK-2 constructs and with increasing concentrations of 5' pre-phosphorylated duplex A or duplex B, using DharmaFect Duo (Dharmacon). 48 h after the initial transfections, *Renilla* and Firefly luciferase expressions were measured using Dual Luciferase Reporter Assay System (Promega), and normalized to mock RNA transfections. Graphs were generated and $IC_{50}$ values were calculated using Igor Pro software (Wavemetrics).

### 2.2.4 Antibodies

The following antibodies were used for western blot and super shift assays: Anti-DICER1 "C-terminal" (sc-30226, Santa Cruz) anti-DICER1 "N-terminal" (Kanellopoulou et al., 2005), anti-AGO2 (2897, Cell Signaling), and normal rabbit IgG (2729, Cell Signaling).

### 2.2.5 Preparation of recombinant DICER1 and cell lysate

2 × 106 HEK293T cells per 15 cm dish were transfected with 33 µg pCAGEN-SBP-DICER1 and 132 µl FugeneHD transfection reagent (1:4 ratio). Cells were harvested 24 hpt and the packed cell volume (PCV) was estimated. Cells were resuspended in 1× PCV of 1× lysis buffer containing 1 mM DTT and 1× Protease Inhibitor Complex (Roche). Cells were dounced, the lysate was cleared at 17000 g for 20 min and incubated with 0.2× PCV of Streptavidin Sepharose High Performance (GE) beads for 3 h at 4 °C. Beads were washed 3 times with 1× lysis buffer containing 300 mM NaCl and 1 mM DTT; and rinsed in the same buffer without NaCl. rDICER1 was eluted from the beads for 1 h at 4 °C using 1× lysis buffer containing 50 % glycerol, 0.01 % BSA, 5 mM biotin and 1 mM DTT (Tsutsumi et al., 2011). For lysate preparation $1.3 \times 10^6$ *Dicer1*$^{-/-}$ MEF cells per 15

cm dish were grown over night and lysate was prepared in the same way as for DICER1 purification. Lysate was flash frozen after clearing at 17000 g for 20 min.

### 2.2.6 RNA oligos

The sequences of duplex A, duplex B and *let-7* siRNA duplex are included in figures 7 and 8 (Hokkaido System Science). The sequence of pre-*let-7* for dicing assays is (Hokkaido System Science):

| | |
|---|---|
| pre-*let-7* | UGAGGUAGUAGGUUGUAUAGUAGUAAUUACACAUCA<br>UACUAUACAAUGUGCUAGCUUUCU |

Target RNAs for in vitro RISC assembly reactions were RNA oligos perfectly complementary to strand 4 or strand 5 of *let-7* siRNA duplex (Figure 8) (Hokkaido System Science).

| | |
|---|---|
| Target for strand 4 | mUmCmUmUmCmAmCmUmAmUmAmCmAmAmCmCmUm<br>AmCmUmAmCmCmUmCmAmAmCmCmUmU |
| Target for strand 5 | mUmCmUmUmCmAmGmAmGmAmGmGmUmAmGmUmUm<br>GmGmUmUmGmUmAmUmAmAmCmCmUmU |

(m): 2'-*O*-methyl.

Target RNAs for cleavage assays were 182nt long RNA fragments containing single perfectly complementary target sites for strand 4 or 5 of *let-7* siRNA duplex at the 3' end. In vitro transcription templates were generated using KOD PlusNeo (Toyobo) and pGL3 basic (Promega) as template and the following primers.

| | | |
|---|---|---|
| Target for strand 4 for target cleavage: | 182nt targetF | GCGTAATACGACTCACTATAGTCACATCTC<br>ATCTACCTCC |
| | Strand4tgtR | CCCATTTAGGTGACACTATAGATTTACATC<br>GCGTGGATCTACTGGTCTGCCTAAAGAAG<br>GTTGAGGTAGTTGGTTGTATAGTGAAGAG<br>AGGAGTTCATG |
| Target for strand 5 for target cleavage: | 182nt targetF | GCGTAATACGACTCACTATAGTCACATCTC<br>ATCTACCTCC |
| | Strand5tgtR | CCCATTTAGGTGACACTATAGATTTACATC<br>GCGTGGATCTACTGGTCTGCCTAAAGAAG<br>GTTATACAACCAACTACCTCTCTGAAGAGA<br>GGAGTTCATG |

PCR products were purified with NucleoSpin PCR and Gel Clean-up kit (Macherey Nagel) and transcribed with T7-scribe standard RNA IVT kit (Cellscript).

### 2.2.7 RNA labeling and duplex preparation

Five picomol RNA were radiolabeled using T4 PNK (Takara) and 32P-γATP in a final volume of 10 µl. After a 1 h incubation at 37°C the sample was run on a G-25 column (GE), run on Urea-PAGE, excised from the gel, eluted precipitated, and resuspended in 1× lysis buffer. To prepare each duplex appropriate single stranded RNAs were mixed, heated at 95 °C for 2 minutes and annealed at room temperature for 30 min in 1× lysis buffer. Cold RNA duplexes were prepared similarly using cold ATP. RNA targets for cleavage assays were radiolabeled using ScriptCap m7G Capping System (Cellscript).

### 2.2.8 Native gel analysis, UV crosslinking, dicing and target cleavage assays

Methods were previously described in detail and were used with minor modifications (Haley et al., 2003; Kawamata and Tomari, 2011; Pellino et al., 2005; Sakurai et al., 2011; Yoda et al., 2010). In vitro RISC assembly reactions contained the following reagents and were incubated at 25 °C for 30 min.

| |
|---|
| 3 µl 40× reaction mix |
| 10 nM radiolabeled RNA duplex |
| 4 µl cell lysate |
| 1 µl rDICER1 |
| 10nM target RNA |
| Final volume: 10 µl (in 1× lysis buffer) |

Samples were run on a 1.4% agarose native gel without the addition of loading buffer (Figure 8D). For the experiments to detect complex D (Figure 8H, I), target RNA was omitted from the reaction and samples were run in a 4.5% polyacrylamide (29:1) native gel after addition of 10 µl Sucrose loading buffer (400 mg/ml sucrose, 2.5 mg/ml xylencyanol, 2.5 mg/ml bromophenol blue). For supershift assays, samples were mixed with 0.5 µl of anti-DICER1 “C-terminal” or normal rabbit IgG and incubated at 4 °C for 1 h before electrophoresis.

For UV cross-linking, after RISC assembly, samples were transferred to terasaki dishes placed on ice in 10 µl aliquots, and were irradiated at a wavelength of 254 nm for 5 minutes. After addition of SDS-PAGE loading buffer samples were heated at 95°C for 2 min and run on an SDS-PAGE.

For target cleavage assays typical RISC assembly reactions were modified to contain 500nM cold RNA duplexes and 10 nM cap-labeled target RNAs. Samples were treated with proteinase K, precipitated with ethanol and run on 8% Urea-PAGE.

Dicing reactions typically contained 3 µl 40× reaction mix, 5 nM radiolabeled pre-let-7, 4 µl cell lysate and 1 µl rDICER1 in 10 µl reaction mixture. Samples were incubated at 30 °C for 30 min, digested with proteinase K, precipitated with ethanol and run on 10% Urea-PAGE.

### 2.3 Reporter assays in live cells for asymmetric RISC assembly in mammals

To assess whether Dicer is necessary for asymmetric RISC assembly of small RNA duplexes in mammalian cells a *Dicer1*$^{-/-}$ mouse embryonic fibroblast (MEF) cell line was used. The cell line harbors a deletion on exon 22 of both Dicer alleles which introduces premature termination codons that cause non-sense mediated RNA decay (Yang et al., 2010; Yi et al., 2006) (Figure 6A). It was reported that an independently established *Dicer1*$^{-/-}$ mouse ES cell line (Murchison et al., 2005) expresses a C-terminally truncated DICER1 (Noland et al., 2011), which may potentially retain some functions of the full-length protein. Western blot analysis using two different antibodies against the N-terminal or the C-terminal regions of DICER1 confirmed the lack of expression of full-length or any specific truncated forms of DICER1 in the *Dicer1*$^{-/-}$ MEF cell line (Figure 6B). Additionally, incubation of *Dicer1*$^{-/-}$ MEF lysate with radiolabeled pre-*let-7* miRNA precursor validated that the cell lysate is deficient in dicing activity (Figure 6C). Dicer expression and function were effectively rescued by transfection with a DICER1-expressing construct (Figure 6D).

First the silencing efficiency of small RNA duplexes in the presence or absence of Dicer was examined. To this end, to minimize the complexity of mimicking natural miRNAs or targeting endogenous genes an arbitrary, functionally asymmetric siRNA duplex (duplex A) was prepared (Figure 7A). As a matter of convenience, the two strands of the duplex are referred to as strand 1 and strand 2. Strand 1 has a less stable 5' end than strand 2; therefore strand 1 is more likely to function as the guide strand (Figure 4). Based on duplex A, an artificial miRNA/miRNA*-like duplex (duplex B) was created (Figure 7B), by introducing internal mismatches and wobble base pairs, without altering the sequence of strand 1 (Yoda et al., 2010). Duplex B is composed of strand 1 and strand 3; according to the thermodynamic asymmetry, again, strand 1 is more likely to serve as the guide strand. Then reporters for each strand of these two duplexes were constructed by introducing a single target site perfectly complementary to each strand downstream of a *Renilla* luciferase (Rluc) reporter in the dual luciferase assay system. These constructs were co-transfected with the corresponding small RNA duplexes into *Dicer1*$^{-/-}$ MEF cells with or without Dicer expression rescued. The incorporation of each strand into mature RISC,

as a reflection of their potency to silence the respective target Rluc reporter, was assessed by dual luciferase assays.

As expected, RISC was assembled asymmetrically in presence of DICER1 with both duplex A and duplex B (Figure 7, +DICER1); in both cases strand 1 was more potent for target silencing. In the case of miRNA/miRNA*-like duplex B, strand 1 was two orders of magnitude more effective in silencing the reporter than strand 3 (Figure 7B and Table 1), while in the case of the siRNA duplex A, strand 1 was around 10-fold more potent than strand 2 (Figure 7A and Table 1). Importantly, for both duplexes, the silencing effect of each strand was unaffected regardless of whether DICER1 expression was rescued or not (compare +DICER1 and –DICER1 in Figure 7); and all the $IC_{50}$ values were essentially identical (Table 1). These findings suggest that mammalian Dicer is dispensable for duplex asymmetry sensing and RISC assembly in live cells.

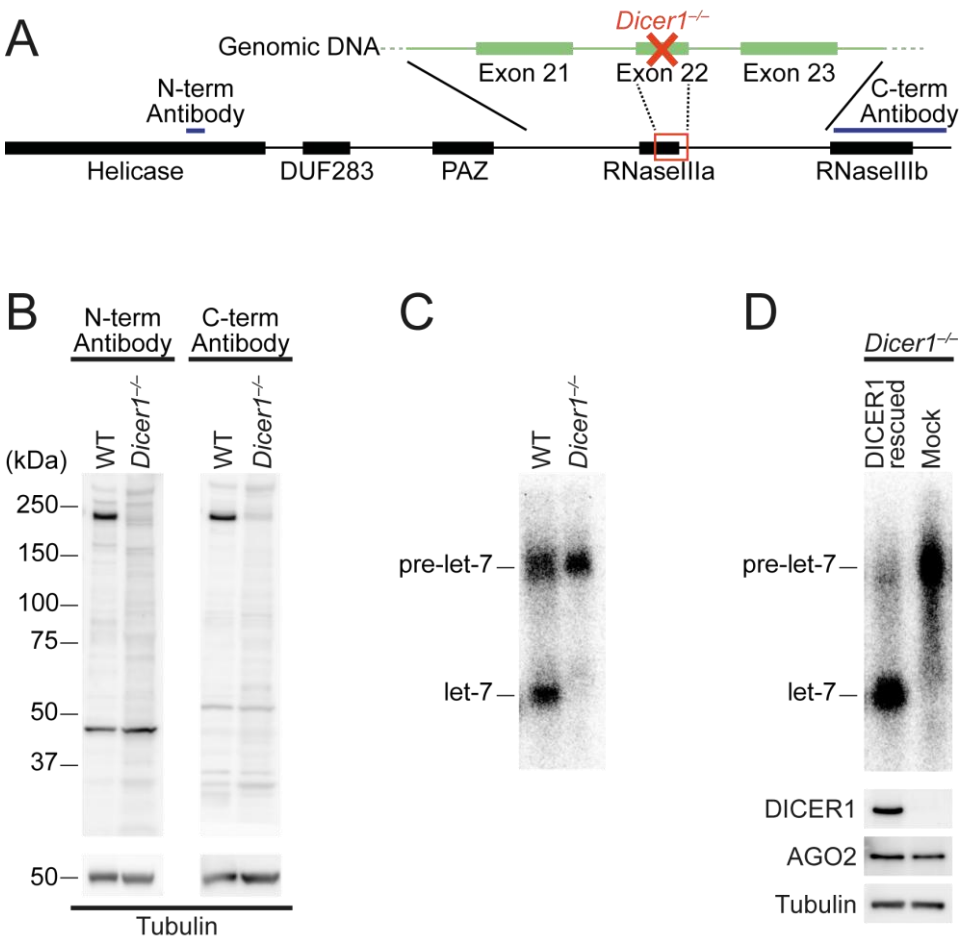



**Figure 6.** Characterization of the *Dicer1*$^{-/-}$ MEF cell line used in this study. A. Schematic representation of the *Dicer1* locus and DICER1 protein. Green blocks in the genomic DNA represent exons 21 to 23 of *Dicer1* and black blocks represent the domains of DICER1. Exon 22 encodes part of the first RNaseIII domain (red rectangle). In the *Dicer1*$^{-/-}$ cell line, exon 22 is deleted which results in premature termination codons that are predicted to cause non-sense mediated RNA decay. Regions recognized by antibodies used in this study are indicated by blue bars. B. Western blot using the antibodies indicated in A. Lack of full length and truncated forms of DICER1 in the *Dicer1*$^{-/-}$ cell line was confirmed. C. Lysate from the *Dicer1*$^{-/-}$ cell line lacks dicing activity. D. Dicing activity and DICER1 expression are rescued by transfection of a DICER1-expressing construct.

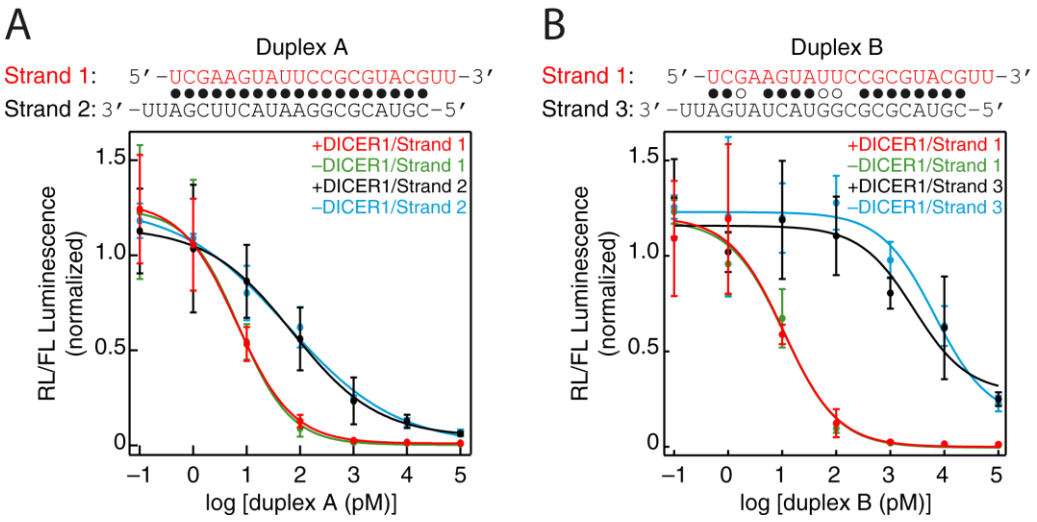



**Figure 7.** Mammalian Dicer is dispensable for asymmetric RISC assembly in live cells. A, B. Silencing of Rluc reporters by increasing concentrations of duplex A (A) or duplex B (B) in *Dicer1*$^{-/-}$ MEF cells with or without DICER1 expression rescued. Mean ± s.d. (n=3).

**Table 1.** log[IC50(pM)] values (mean ± s.d.) calculated from the data in Figure 7A and B. Essentially identical $IC_{50}$ values were obtained with or without Dicer expression rescued.

| | | **+DICER1** | **–DICER1** |
|---|---|---|---|
| **Duplex A** | **Strand 1** | 0.8 ± 0.2 | 0.9 ± 0.2 |
| | **Strand 2** | 1.9 ± 0.3 | 1.8 ± 0.2 |
| **Duplex B** | **Strand 1** | 1.0 ± 0.2 | 1.1 ± 0.1 |
| | **Strand 3** | 3.3 ± 0.3 | 3.8 ± 0.2 |

### 2.4 In vitro assays for asymmetric RISC assembly in mammals

Next, to explore any potential function of Dicer in RISC assembly the formation of pre-RISC and its conversion to mature RISC by an agarose native gel system was compared (Kawamata and Tomari, 2011) in the presence or absence of Dicer. Previously, in vitro RISC assembly activity has been assessed in HeLa or HEK293T lysate (Yoda et al., 2010), but in MEF lysate it is rather weak in comparison, and tends to fluctuate from batch to batch. Therefore, instead of comparing between *Dicer1*$^{-/-}$ MEF lysates with and without Dicer expression rescued by transfection, a strategy of re-supplementing *Dicer1*$^{-/-}$ MEF cell lysate with purified recombinant DICER1 (rDICER1; Figure 8A) was used. A control experiment showed that addition of rDICER1 in *Dicer1*$^{-/-}$ MEF cell lysate efficiently rescued dicing activity (Figure 8B). A functionally asymmetric siRNA duplex based on the *let-7* sequence (*let-7* siRNA duplex; Figure 8C) was used, which has been intensively studied with the agarose native gel system (Kawamata et al., 2009; Yoda et al., 2010). Either of the two strands (strand 4 and strand 5) was 5' $^{32}$P radiolabeled, and the duplexes were incubated in *Dicer1*$^{-/-}$ MEF lysate with or without rDICER1 re-supplementation. When samples were run on a native agarose electrophoresis two complexes could be detected (Figure 8D), as previously described. The complex at the top contains double stranded RNA, while the complex at the bottom contains single stranded RNA, and therefore correspond to pre-RISC and mature RISC, respectively (Kawamata et al., 2009; Yoda et al., 2010). Similar level of pre-RISC was detected when either of the two strands was radiolabeled, whereas markedly higher level of mature RISC was detected when strand 4 was radiolabeled, precisely reflecting the asymmetry of the duplex. Importantly, the presence or absence of Dicer did not affect the kinetics of asymmetric loading of the duplex or subsequent unwinding (Figure 8D).

To assess the direct loading of each strand into Ago, in vitro RISC assembly reactions were crosslinked with short-wave UV light (Figure 8E); and to determine the functional asymmetry of RISC in vitro the target cleavage activity of each strand was evaluated (Figure 8F, G). Note that no target cleavage was detected unless RISC was programmed with exogenous *let-7* siRNA, which confirms the specificity of the assay (Figure 8F). Crosslinking to Ago and target cleavage agreed well with the amount of mature RISC formed, and were unaffected by the presence or absence of

Dicer. Therefore, mammalian Dicer is dispensable for asymmetric RISC assembly and target cleavage of small RNA duplexes in vitro.

It was previously reported that, in mammalian cell lysate, siRNA duplexes form a complex containing Dicer and TRBP (referred to as complex D or pre-RLC), which can be separated by polyacrylamide native electrophoresis (Pellino et al., 2005; Sakurai et al., 2011). Indeed, such complex was readily detected with radiolabeled *let-7* siRNA duplex when *Dicer1*$^{-/-}$ MEF lysate was supplemented with rDICER1 (Figure 8H). Note that the complex indicated by an asterisk on figure 8H does not contain DICER1 and is therefore non-specific (Figure 8I). Unlike mature RISC, formation of complex D and crosslinking to Dicer showed no correlation with the amount of mature RISC formed and the target cleavage activity (compare Figures 8D, E, G, H), confirming that Dicer plays a minimum role in RISC assembly in mammals.

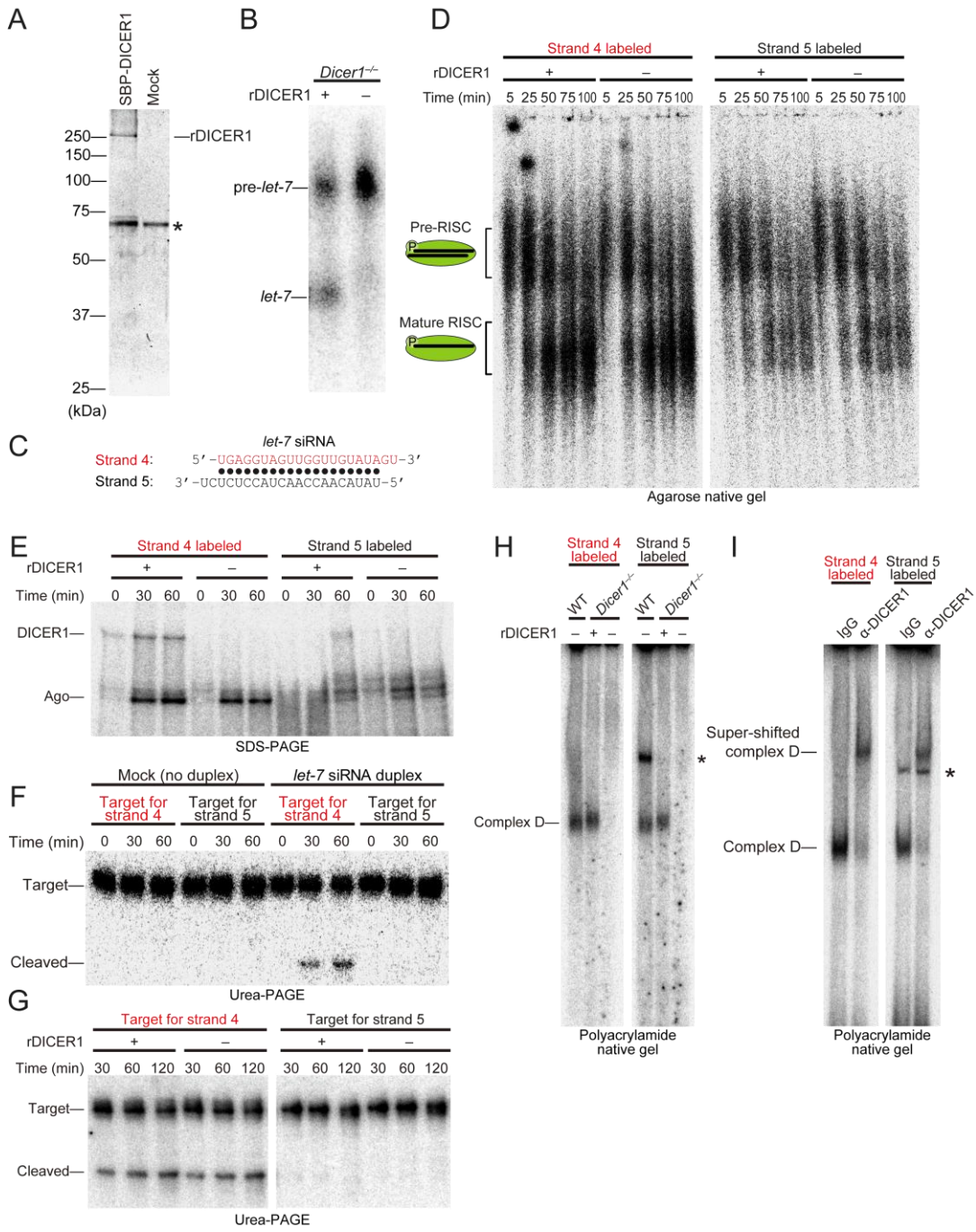



**Figure 8.** Mammalian Dicer is dispensable for asymmetric RISC assembly in vitro. A. Purification of rDICER1 from HEK293T cells expressing SBP-DICER1. Mock purification was performed from naïve HEK293T. Asterisk denotes a non-specific protein that binds to sptreptavidin sepharose. B. Dicer activity in *Dicer1*$^{-/-}$ lysate with or without rDICER1 supplementation. C. RNA duplex used in D–I. D. Pre-RISC and mature RISC formation with strand 4 or strand 5 in the presence or absence of DICER1. E. Short-wave UV crosslinking of strand 4 or strand 5 in the presence or absence of DICER1. F. A control experiment to confirm the specificity of the target cleavage assay in (G). No cleavage product was detected unless RISC

was programmed by exogenous *let-7* siRNA duplex. G. Target cleavage assay for strand 4 or strand 5 in the presence or absence of DICER1. H. Formation of Dicer containing complex (complex D) in the presence or absence of DICER1. I. Super shift assay of Complex D with anti DICER1 antibody. The complex marked with an asterisk in (H) does not contain Dicer.

### 2.5 Dicer is not required for asymmetric RISC assembly in mammals

In the fly Ago2-RISC assembly pathway, Dcr-2/R2D2 is absolutely required for loading of siRNA duplexes; in the absence of Dcr-2 or R2D2, no Ago2-RISC is formed from siRNA duplexes (Liu et al., 2003; Pham et al., 2004; Tomari et al., 2004a). For this reason, a complex containing Dcr-2/R2D2 is called RISC-loading complex (RLC) (Figure 5). In contrast, whether or not Dicer is required for RISC assembly in mammals has been a subject of debate. Previous multiple lines of evidence (Kanellopoulou et al., 2005; Martinez et al., 2002; Murchison et al., 2005) together with the findings of this work indicate that Dicer is not required for RISC assembly in mammals (Figure 9C).

Dcr-2/R2D2 binds to siRNA duplexes in an orientation according to their thermodynamic asymmetry (Tomari et al., 2004b) (Figure 5). Recombinant human Dicer/TRBP is also able to bind to siRNAs in an asymmetric manner, just like fly Dcr-2/R2D2 (Noland et al., 2011). However, the data shown here suggest that such Dicer/TRBP binding plays no essential role in asymmetric RISC assembly in mammals. How can then thermodynamic asymmetry of small RNA duplexes be sensed in mammals? Although it cannot be excluded that there are unknown factors for this function, the hypothesis that Argonaute proteins themselves can generally sense the thermodynamic asymmetry of small RNA duplexes is attractive. Structural studies of an archaeal Piwi protein and a bacterial Ago protein (Parker et al., 2005; Wang et al., 2008), and more recently of eukaryotic Ago proteins (Elkayam et al., 2012; Nakanishi et al., 2012; Schirle and MacRae, 2012), have shown that the 5' phosphate of the guide strand is anchored in a phosphate binding pocket at the interface between the MID and PIWI domains. In that position, paring between the 5' first base of the guide strand and the complementary base near the 3' end of the passenger strand is blocked, with the ends of the two strands directed in opposite directions (Figure 10). This structural conformation should favor the orientation of the small RNA duplex with the end that is more easily unwound (i.e., the less stable end) towards the phosphate-binding pocket. The high homology of the MID and PIWI domains, in particular the architecture of the C-terminal carboxylate that forms the phosphate binding pocket, suggests that the binding mode for a small RNA molecule is commonly shared by Ago proteins (Boland et al., 2011; Elkayam et al., 2012; Nakanishi et al., 2012; Parker et al., 2005; Schirle and MacRae, 2012; Wang et

al., 2008). In the case of fly Ago2-RISC assembly, Dcr-2/R2D2 binding might double-check or amplify the sensing of thermodynamic asymmetry in addition to Ago2 itself; an siRNA duplex needs to bind Dcr-2/R2D2 before entering Ago2 and yet the duplex might transiently shuttle between Dcr-2/R2D2 and Ago2, with each protein checking the asymmetry, before it is properly docked into Ago2. Supporting this view, crosslinking experiments show that the orientation of Dcr-2/R2D2 on the siRNA duplex is less asymmetric for the recombinant heterodimer than for embryo lysate (Tomari et al., 2004b). In contrast, such double-checking system may be lacking in mammals where Dicer/TRBP binding is uncoupled from RISC assembly. Interestingly, in dicer-mutant zebrafish embryos, injected miRNA duplexes were fully active for target silencing, indicating that RISC assembly does not require Dicer in fish as in mammals (Giraldez et al., 2006). Even in flies, simultaneous >95% depletion of Dcr-1 and Dcr-2, which abolished siRNA duplex-initiated Ago2-RISC assembly, did not affect assembly of miRNA/miRNA* duplexes into Ago1-RISC, suggesting that Dicer is dispensable for fly Ago1-RISC assembly (Kawamata et al., 2009). Therefore, although fly Ago2-RISC assembly is the best-characterized pathway to date, it is likely to represent a very unique pathway entirely dependent on Dcr-2/R2D2. Indeed, the four mammalian Ago proteins and fly Ago1 are more closely related to each other than to fly Ago2 (Kawamata et al., 2009; Yoda et al., 2010) (Figure 9). Moreover, fly Ago1 and human Ago1–4 show strong preference of uracil (U) at the 5’ end of the guide strand, which could in theory also help sensing the thermodynamic asymmetry, whereas fly Ago2 tends to favor 5’ cytosine (C) (Czech et al., 2009; Ghildiyal et al., 2010; Okamura et al., 2009). In summary, although the fly Ago-2 pathway has been extensively studied, generalization of its properties should be done with caution, and further study is required to clarify what makes this pathway so special.

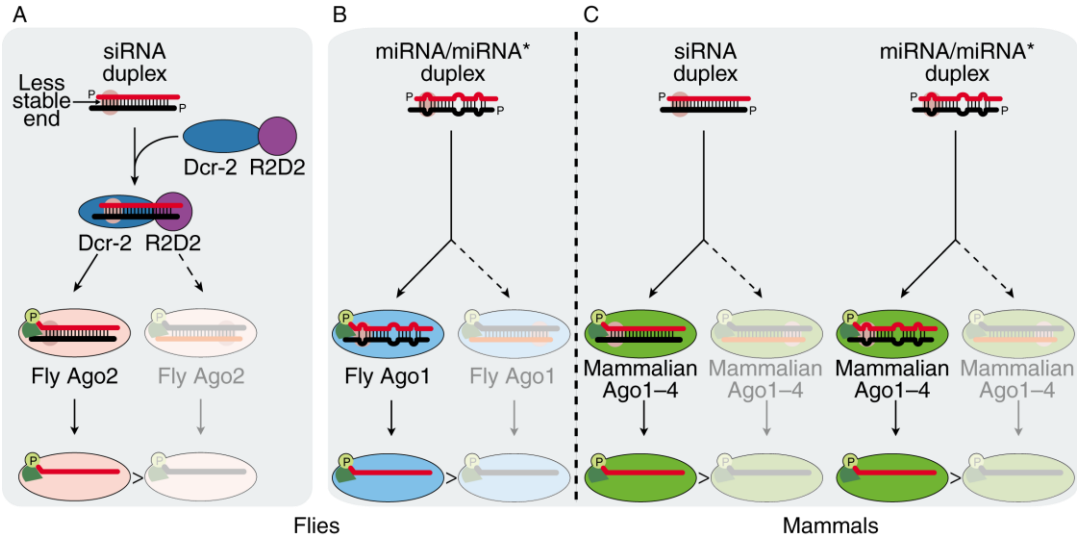



**Figure 9.** Scheme of the asymmetric assembly of RISC. Small RNA duplexes are preferentially loaded into Ago proteins with their less stable end towards the phosphate-binding pocket of Ago (P), which results in the selection of the red strand of the RNA duplex as the guide strand. If the duplex is occasionally loaded in the opposite orientation, the black strand functions as the guide. Red strand-loaded Ago is more abundant (>). In flies siRNAs and miRNA/miRNA* duplexes are actively sorted into Ago2- and Ago1-RISC, respectively. A. The Dcr-2/R2D2 heterodimer senses the asymmetry of siRNA duplexes and such binding is a prerequisite for Ago2-RISC assembly. B, C. In contrast to the fly Ago2 pathway, in the (B) fly Ago1 and (C) mammalian pathways Dicer is dispensable for asymmetric RISC loading. Modified from (Betancur et al., 2012).

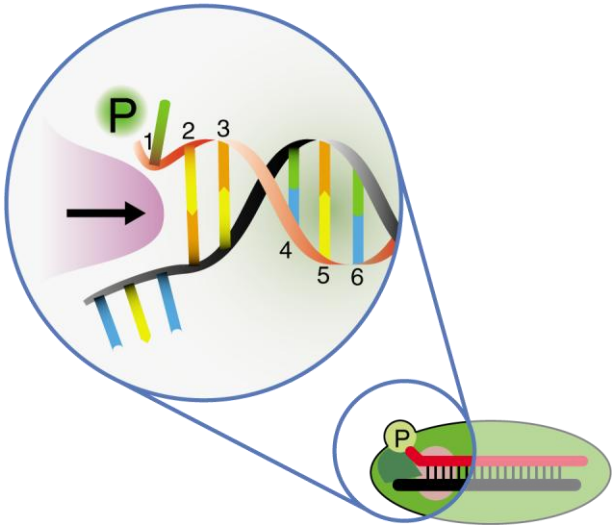



**Figure 10.** Scheme of the structural conformation of double stranded RNA-loaded Ago. The figure shows a close up of the phosphate binding pocket of Ago. The 5' phosphate (P) of the guide strand (red) is docked at the phosphate binding pocket of Ago (green shadow) between the PIWI and MID domains. In this conformation, an α-helix (arrow) blocks base pairing between the first base of the guide strand (1) and the corresponding base in the passenger strand (black). This conformation itself might favor the orientation of small RNA duplexes with their less stable end (which is more easily unwound) towards the phosphate binding pocket, over the opposite orientation.

# 3. Part 2: The RNA binding activity of PRC2

### 3.1 A subset of lncRNAs interacts with the PRC2 complex

Chromatin is a dynamic structure that needs to adjust to allow or repress gene expression according to diverse regulatory signals. During development and differentiation gene expression reprogramming is particularly important and it involves a concert of DNA and histone modifying complexes that regulate the coordinate transition of cells from the pluripotent to the differentiated state that leads to the formation of full organisms (Ringrose and Paro, 2004). In contrast, abnormal epigenetic states that result in gene expression patterns that partially resemble the transcriptional landscape of undifferentiated cells are hallmarks of some types of cancer (Egger et al., 2004).

Epigenetic marks are roughly classified as either activating or repressive and include DNA and histone modifications. Activating marks result in chromatin domains that are more accessible for transcription factors and RNA polymerases and are therefore actively transcribed; while repressive marks result in gene expression silencing (Figure 11). In mammals, DNA methylation is a stable mark of transcriptionally repressed chromatin that is usually transmitted to daughter cells (Smith and Meissner, 2013). In contrast histones tails undergo a variety of post-translational modifications, including methylation, acetylation, ubiquitination, phosphorylation among others, and depending on their nature and combination with other epigenetic marks have either activating or repressive effects. Histone modifications can be reversed after their initial deposition (Jenuwein and Allis, 2001).

Although the functions of many histone modifications are unknown, some of them have been consistently associated with specific roles in gene expression. In mammalian cells di/tri methylation of lysine 27 histone 3 (H3K27me2/3), a repressive mark, is usually associated with promoters of silent genes, while H4K4me3 is associated with actively transcribed promoters (Ringrose and Paro, 2004; Schuettengruber et al., 2007). However, H3K27me2/3 also forms clusters with smaller regions of H3K4me3, known as bivalent domains, a hallmark of regions rich in regulatory factors important during development in mouse embryonic stem (mES) cells that are resolved (activated or repressed) upon lineage commitment (Azuara et al., 2006; Bernstein et al., 2006; Mikkelsen et al., 2007). Other histone marks with known functions are H3K36me3 that is enriched in the bodies of actively transcribed

genes, while H3K9me3 and H3K20me3 associate with repetitive elements; and overlapping H3K9me3, H3K4me3, and DNA methylation are a signature of promoters of imprinted genes (Delaval et al., 2007; Feil and Berger, 2007; Mikkelsen et al., 2007).

H3K27me2/3 is deposited by PRC2, a member of the Polycomb group (PcG) protein family, on genes that are already silenced, and maintains transcription repressed throughout cell divisions (Ringrose and Paro, 2004). PRC2 is a multiprotein complex composed of 4 core components: EED, EZH2, RbAp46/48 and SUZ12 (Cao et al., 2002; Kuzmichev et al., 2002) (Figure 12), which are orthologs from the *Drosophila* Extra sexcombs (Esc), Enhancer of zeste E(z), a nuclear remodeling factor (Nurf55) and Suppressor of zeste 12 (Su(z)12), respectively (Schuettengruber and Cavalli, 2009). EZH2, through its SET domain, is the catalytic subunit of the complex (Cao et al., 2002), while the others subunits are required for maximum methyltransferase activity (Cao and Zhang, 2004). In addition EED has been reported to recognize pre-trimethylated H3K27 (Margueron et al., 2009), and RbAp48 is a histone chaperone that binds to histone 3 and 4 (Nowak et al., 2011), both through their WD40 domains (Figure 13). Moreover, there are some other accessory proteins that usually co-precipitate at sub-equimolar ratios, including Polycomblike proteins (PCLs) (Casanova et al., 2011), AEBP2 (Kim et al., 2009a) and JARID2 (Pasini et al., 2010). AEBP2 (Kim et al., 2009a) and JARID2, through its ARID domain (Kim et al., 2004; Kim et al., 2003), can bind DNA directly (Figure 12).

PRC2, however, does not bind DNA in a sequence specific manner. Instead, in flies PRC2 (and other PcG complexes) binds to DNA elements known as polycomb response elements (PRE) with the aid of several different co-factors. However, many of the genomic binding sites for such factors do not overlap with PcG-bound regions, but instead with active epigenetic marks, which indicates that those factors are not sufficient for PRC2 recruitment on their own (Muller and Kassis, 2006; Schuettengruber and Cavalli, 2009; Schuettengruber et al., 2007). In addition, in mammals, no PRE-like elements have been described, although many PRC2 binding sites are C+G rich, or correspond to CpG islands, (Ku et al., 2008; Margueron and Reinberg, 2011; Schuettengruber and Cavalli, 2009). Instead of requiring single recruiting cofactors, specific recruitment of PRC2 to its target loci is more likely the

result of a combinatorial effect of several independent events. Those events might include interactions engaged by EED and RbAp48 with histones (Margueron et al., 2009; Nowak et al., 2011), binding to DNA by JARID and AEBP2 (Kim et al., 2009a; Kim et al., 2004; Kim et al., 2003), and as recent reports suggest, the contribution of lncRNAs (Tsai et al., 2010; Zhao et al., 2008).

In the last decade it became apparent that most of the mammalian genome is transcribed (Carninci et al., 2005), and multiple efforts have focused on cloning and analyzing the vast amount of mammalian transcripts. One of the first systematic descriptions of lncRNAs came from the analysis of mouse cDNA libraries that led to the identification of a large pool of RNAs longer than 200nt with low coding potential (Carninci et al., 2005). Although initially many lncRNAs were described as transcriptional noise (Ponjavic et al., 2007; Struhl, 2007); it soon became clear that many have strict expression patterns and subcellular localizations (Amaral and Mattick, 2008; Dinger et al., 2008; Mercer et al., 2008), which suggests a high level of regulation. In addition biochemical and functional studies have demonstrated that at least a portion of them is important in processes as diverse as chromatin remodeling, regulation of transcription, splicing and translation (Mercer et al., 2009).

A subset of lncRNAs has been described to interact with DNA or histone modifying complexes and regulate epigenetic states, although in most cases details of the molecular mechanisms of the protein-RNA interaction are still not well understood (Lee, 2012). In mice Airn, and Kcnq1ot1 RNAs interact with G9a, a H3K9 methyltransferase, to regulate the expression of imprinted genes located the antisense orientation of their own loci (Nagano et al., 2008; Pandey et al., 2008). There are also reports that PRC1, another PcG repressor complex that mono ubiquitinates lysine 119 of histone H2A (H2AK119u1), is recruited to the tumor suppressor locus *INK4b/ARG/INK4a* by ANRIL lncRNA; and that MALAT and TUG lncRNAs interact with the PRC1 subunit PC2 to regulate the intranuclear location of transcriptional units (Yang et al., 2011). lncRNAs also associate with epigenetic activator complexes, such in the case of Hottip and Mistral RNAs that bind to MLL, a H3K4 methyltransferase, to activate HoxA gene transcription (Bertani et al., 2011; Wang et al., 2011). The DNA methyltransferase Dnmt3a is also recruited by the lncRNA Tsix to silence Xist, located in the antisense orientation (Sado et al., 2005; Sun et al., 2006).

In addition, RNA immunoprecipitation (RIP) of PRC2 demonstrated that the complex directly associates with many RNAs in mouse (Zhao et al., 2010) and human (Khalil et al., 2009) cells and that a portion of them corresponds to lncRNAs. In mice, around 1/5 of all PRC2-associated RNAs overlap H3K27me2/3-H3K9me3 bivalent domains or known SUZ12 binding sites, an indication that lncRNAs might be related to PRC2 recruitment or retention in a subset of PRC2 targets (Zhao et al., 2010). Further evidence of the functional significance of the lncRNA-PRC2 interaction came from knock down experiments that showed that depletion of some lncRNAs in human cells caused de-repression of genes normally silenced by PRC2 (Khalil et al., 2009).

Two of the lncRNAs that have consistently been shown to interact with PRC2 are the X-inactive-specific transcript (Xist) and Hox antisense intergenic RNA (HOTAIR). In humans Xist is a 19kb transcript transcribed from the X inactivation center (Xic) in the X chromosome (Brown et al., 1991). It is required for X chromosome inactivation (XCI), a process in which one of the two X chromosomes is randomly selected to be silenced during early development to equalize gene expression from the X chromosome between males and female mammals (Lyon, 1961; Russell, 1961). XCI is initiated by accumulation and spreading of Xist on the inactive X (Xi) (Avner and Heard, 2001; Brockdorff et al., 1992; Brown et al., 1992; Clemson et al., 1996; Plath et al., 2002), a process mediated by DNA binding proteins directly bound by Xist (Hasegawa et al., 2010; Jeon and Lee, 2011). Interestingly, introduction of ectopic Xic sequences in autosomes results in long range silencing, although inactivation is less efficient that for X-linked genes (Cattanach, 1974; Popova et al., 2006; Russell, 1963).

XCI requires the H3K27 methytransferase activity of PRC2 (Wang et al., 2001), and the H2AK119 mono ubiquitinase activity of PRC1 that together effectively silence gene expression from the target chromosome.

One model for the recruitment of PRC2 indicates that PRC2 directly binds Xist, which is tethered to the Xi (Hasegawa et al., 2010; Jeon and Lee, 2011), providing a platform for spreading H3K27me3 in cis (Zhao et al., 2008) (Figure 14). The region of Xist bound by PRC2 has been mapped to the 5’ end of the transcript, where there is a stretch of 8.5 repeats of a ~28nt sequence, known as A-repeat domain (Wutz et

al., 2002; Zhao et al., 2008). Although the most widely accepted conformation of the A-repeat domain is that each of its 8.5 repeated sequences adopt a double stem loop structure (Figure 15A and Supplementary figure 2), there is also some evidence that only hairpin 1 forms, and hairpin 2 is instead involved in the formation of inter-repeat duplex interactions (Duszczyk et al., 2011). Xist with a deletion of A-repeat domain still spreads on the Xi, but fails to induce inactivation (Wutz et al., 2002). It has also been postulated that a single repeat of is enough for direct binding of PRC2 to Xist (Kanhere et al., 2010; Zhao et al., 2008).

In contrast to Xist, HOTAIR, acts in trans to regulate Hox genes, a group of transcription factors that regulate body morphogenesis and the positional identity of cells. In humans HOTAIR is a 2.3kb long transcript transcribed from within the HoxC locus, on chromosome 12, and its target genes are in the HoxD locus, on chromosome 2; and in agreement with its role as a regulator of Hox genes, HOTAIR expression is restricted mainly to posterior and distal anatomical sites. Similar to Xist, HOTAIR is proposed to directly recruit PRC2 through a 300nt domain located at its 5' most end (Figure 15B), which results in the tri-methylation of H3K27 and silencing of the HoxD locus (Rinn et al., 2007; Tsai et al., 2010). However, it soon became clear that besides its role in the regulation of Hox genes, HOTAIR is also directly recruited to and regulates the expression of multiple genes (Chu et al., 2011; Gupta et al., 2010). HOTAIR is thought to recognize its target DNA locus with the help of the simultaneous recruitment of protein complexes that recognizes specific DNA motifs, such as the LSD1/CoREST/REST complex (Rinn et al., 2007; Tsai et al., 2010) (Figure 14).

HOTAIR plays a key role in reprogramming of the chromatin state in human cancer cells (Gupta et al., 2010); and HOTAIR knock down human fibroblasts exhibit transcriptional activation in trans in the HoxD locus (Rinn et al., 2007). However, although the sequence of HOTAIR is modestly conserved and its anatomical pattern of expression is similar between humans and mice (Rinn et al., 2007), HOTAIR knockout mice and MEFs display little phenotypic effects (Schorderet and Duboule, 2011), which is suggestive of recent acquisition of HOTAIR function in humans (Schorderet and Duboule, 2011).

Initial evidence of the direct interaction between Xist and HOTAIR with PRC2 came from electrophoretic mobility shift assays (EMSA) and RNA pull-down experiments with Xist and HOTAIR fragments. While some experiments concluded that SUZ12, but not EZH2 is responsible for the RNA binding activity of the complex (Kanhere et al., 2010), the opposite has also been shown by different groups (Kaneko et al., 2010; Tsai et al., 2010; Zhao et al., 2010; Zhao et al., 2008), and therefore the existing reports regarding the RNA binding activity of PRC2 are contradictory and in some cases mutually exclusive and this remains a matter of debate that this work aims to revisit (Figure 16).

Also, there is no known common RNA motif that PRC2 recognizes. Although the double stem loop structure of A-repeat has been postulated to be the PRC2-bound element of Xist, it is unknown whether PRC2 specifically inspects certain features of the sequence or structure or whether binding specificity is loose. In addition, other lncRNAs (e.g. HOTAIR) lack such repeats, and therefore whether PRC2 makes use of different mechanism to bind to different RNAs or not is still an open question.

To address these questions I performed a series of in vitro RNA binding experiments with recombinant PRC2 components and A-repeat RNA fragments with intact or disrupted double stem loop structures and HOTAIR fragments. I demonstrate that in contrast to previous reports 2 subunits of PRC2, EZH2 and SUZ12, have RNA binding activity. In addition, I show that both proteins also bind to mutated A-repeat fragments as well as to HOTAIR, that does not contain sequence or strucrure repeats; and also to unrelated RNAs, which is an indication that the double stem loop structures of A-repeat are not specifically recognized by PRC2 and that instead the RNA binding specificity of the complex is loose.

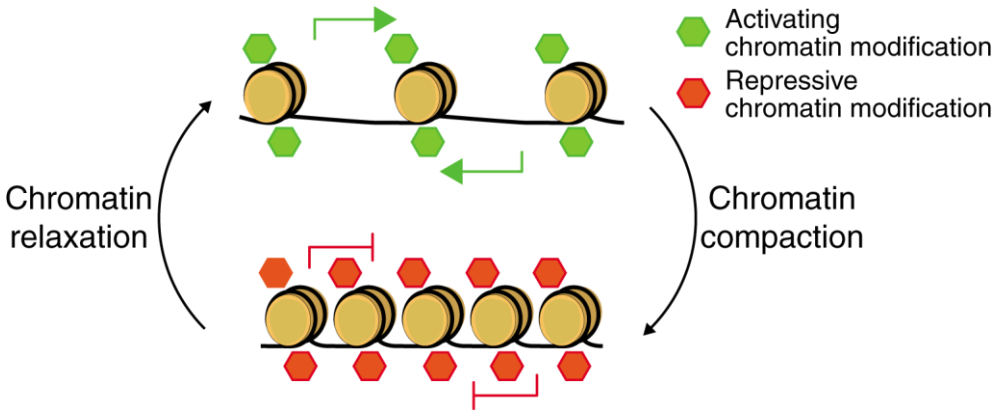



**Figure 11.** The two basic states of chromatin. Chromatin is a dynamic structure that particularly during development and differentiation, needs to adjust to allow or repress gene expression according to regulatory signals. There are two basic chromatin states: compacted or silent and relaxed or transcriptionally active. The shift between the two states is mediated by various DNA and histone modifying factors that induce a large array of chromatin modifications that can be broadly classified as activating or repressive.

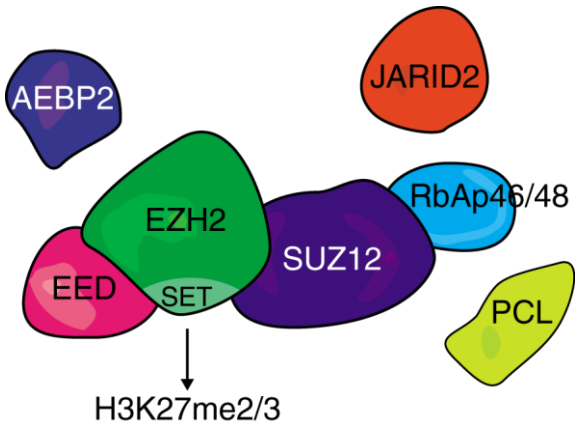



**Figure 12.** Composition of PRC2. PRC2 is composed of 4 core components: EED, EZH2, SUZ12 and RbAp46/48. EZH2 catalyzes the di/tri methylation of H3K27. AEBP2, JARID2 and PCL are accessory proteins.

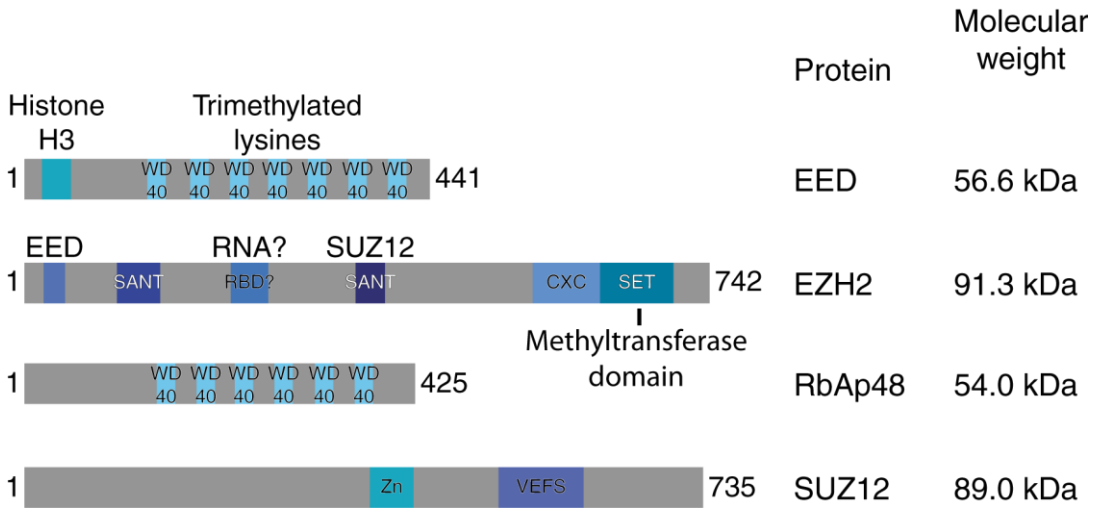



**Figure 13.** Domain structure of the 4 core components of PRC2. Labels above each cartoon indicate substrates that the corresponding domains (or regions) bind to. SANT: SW13, ADA2, N-CoR and TFIIIB DNA-binding domain; CXC: cysteine-rich domain; WD40: short 40aa motifs; SET: tritorax conserved domain; VEFS: domain conserved among VRN2-EMF2-FIS2-SUZ(Z)12; RBD?: putative RNA binding domain Zn: zinc finger domain.

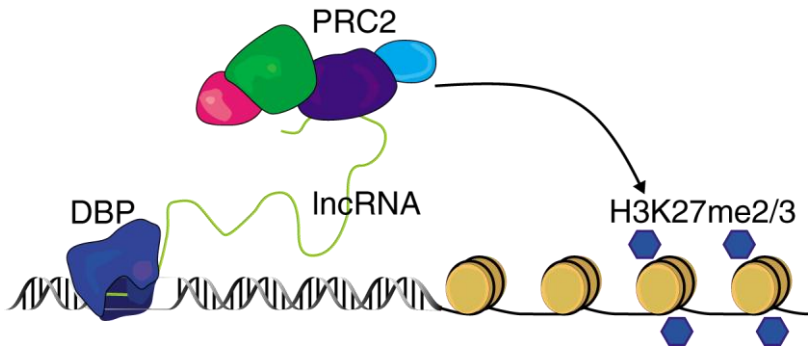



**Figure 14.** Model of the recruitment of PRC2 by Xist and HOTAIR. The lncRNA acts as a scaffold for the binding of PRC2 and additional DNA binding complexes that tether PRC2 to its target loci. In addition to the lncRNAs other factor are also necessary for specific recruitment of PRC2 (see text for details). DBP: DNA binding protein, H3K27me2/3: di/tri methylated lysine 27 of histone 3.

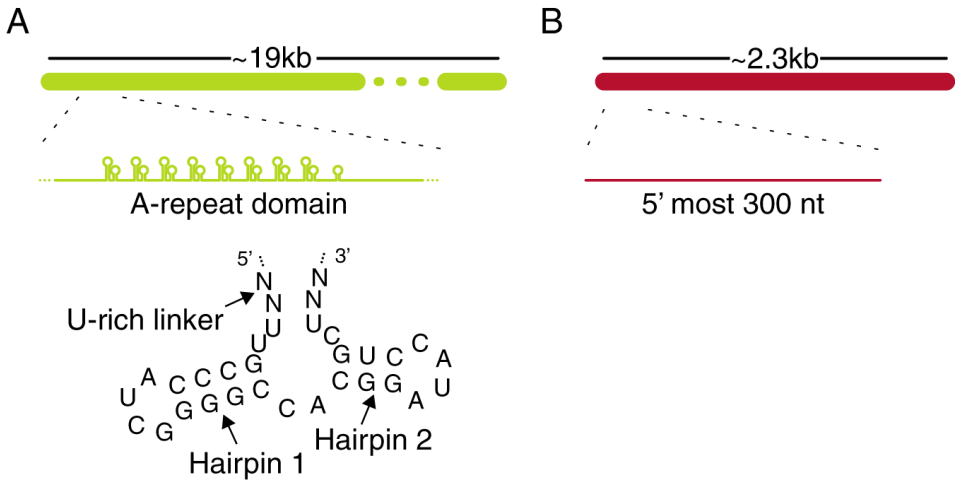



**Figure 15.** Human Xist and HOTAIR lncRNAs. A. Xist is a ~19 kb long transcript that contains a domain known as A-repeat towards its 5' end, which is the putative PRC2 binding region. A-repeat consists of 8.5 repeats of a double stem loop (below) connected by ~20 nt U rich linkers. B. HOTAIR is ~2.3 kb long and the putative region bound by PRC2 has been mapped to its 5' most 300 nt.

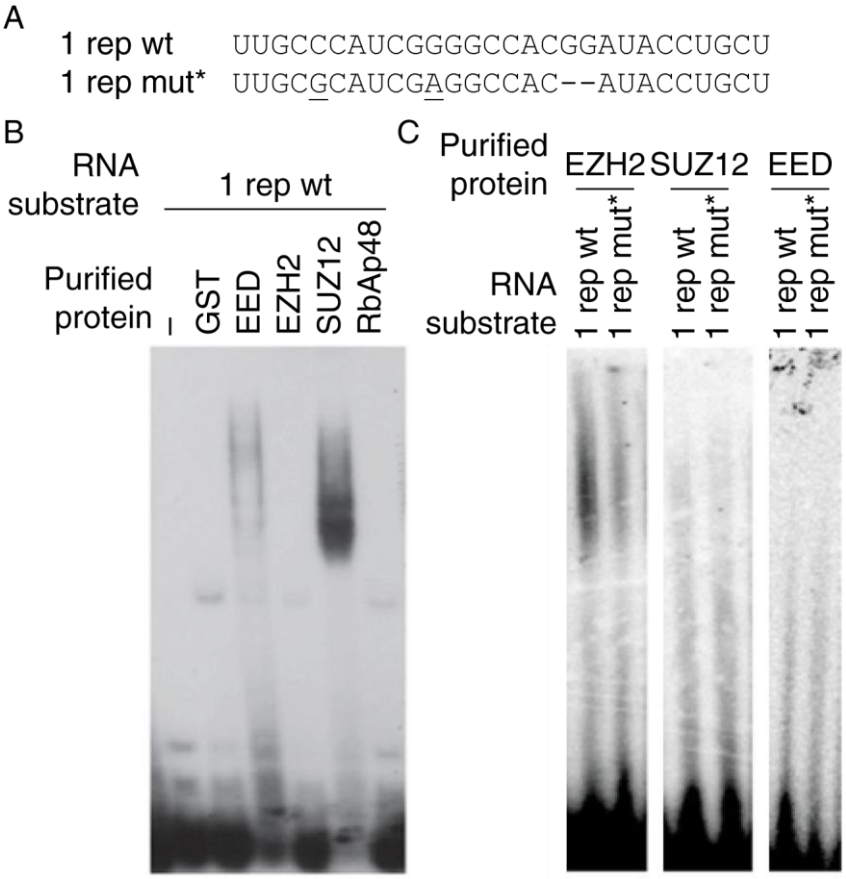



**Figure 16.** Previous attempts to identify the RNA binding subunit of PRC2 by EMSA. A. RNA fragments used as substrates. 1 rep wt folds into a double stem loop like the one shown in Figure 15A. B. One group (Kanhere et al., 2010) showed that SUZ12, but not EZH2 is the RNA binding subunit of PRC2. C. A different group (Zhao et al., 2010) found the opposite result using the same RNA substrate. In addition, a mutant, non structured RNA (1 rep mut*) was bound more weakly by EZH2. Figures modified from the references indicated.

## 3.2 Materials and methods

### 3.2.1 General reagents

1× lysis buffer and 40× reaction mix are described in section 2.2.1.

### 3.2.2 Plasmid constructions

Total RNA was extracted from HeLa S3 and HEK293T cells using Trizol. One microgram of RNA was reversed transcribed using PrimeScript RT reagent kit with gDNA eraser (Takara) following the instructions from the manufacturer.

PRC2 cDNA: EED, EZH2, RbAp48 and Suz12 were amplified from HeLa S3 cDNA using PrimeStar Max (Takara) and the following primers pairs.

| | | |
|---|---|---|
| EED: | EED extF1 | TGGGCGCGATTTGCGACAGT |
| | EED extR1 | TGCTCTACGTGCCCTTACTAGCA |
| EZH2: | EZH2 extF1 | TCCGACACCCGGTGGGACTC |
| | EZH2 extR1 | GCAGCTGTTTCAGAGGAGGGGG |
| RbAp48: | RbAp48 extF1 | TCGACCCCAGGATTCCCCCG |
| | RbAp48 extR1 | GAAAACACCCACGGTTTGGGCT |
| SUZ12: | SUZ12 extF1 | GGGCGAGCGGTTGGTATTGCA |
| | SUZ12 extR1 | ACTCAACCACAGTGCTCGGAGT |

pCAGEN-SBP-EZH2, -Suz12 and pCAGEN-SBP-ps-EED, -RbAp48: The PRC2 component sequences were reamplified using a nested PCR strategy and the SBP tag was amplified from pASW (Iwasaki 2010). The fragments were cloned simultaneously into the *Eco*RI site of pCAGEN (Matsuda and Cepko, 2004) using Infusion (Clontech) to produce pCAGEN-SBP-EZH2, and -SUZ12. The following primers were used:

| | | |
|---|---|---|
| SBP: | SBP F2 | TTTTGGCAAAGAATTCCCATGGACGAGAA GACCACCGGC |
| | SBP R2 | GGCCGCGGAGCCTGCTTTTT |
| SUZ12: | SUZ12 F1 | GCAGGCTCCGCGGCCATGGCGCCTCAGAA GCAC |
| | SUZ12 R1 | TATCCTCGAGGAATTTGGGGTTAGAGCTTT TCAGAGT |
| EZH2: | EZH2 F1 | GCAGGCTCCGCGGCCATGGGCCAGACTGG GAAG |
| | EZH2 R1 | TATCCTCGAGGAATTGGGGAGGAGGTAGC AGATGT |

EED and RbAp48 were cloned with a cleavage site for HRV3C (PreScission [ps]) protease following the SBP tag. The following oligos containing the recognition site of the protease were annealed and cloned simultaneously with the SBP tag into the *Eco*RI site of pCAGEN using Infusion (Clontech) to produce pCAGEN-SBP-ps. The following primers were used.

| | | |
|---|---|---|
| SBP: | SBP F3 | TTTTGGCAAAAATTCCCATGGACGAGAAG ACCACCGGC |
| | SBP R2 | GGCCGCGGAGCCTGCTTTTT |
| ps: | ps F1 | GCAGGCTCCGCGGCCCTGGAAGTTCTGTTC CAGGGGCCCGAATTCCTCGAGGATA |
| | ps R1 | TATCCTCGAGGAATTCGGGCCCCTGGAAC AGAACTTCCAGGGCCGCGGAGCCTGC |

EED and RbAp48 were reamplified with a nested PCR and cloned into the *Eco*RI site of pCAGEN-SBP-ps. The following primers were used.

| | | |
|---|---|---|
| EED: | EED F2 | CTGTTCCAGGGGCCCATGTCCGAGAGGGA AGTGTC |
| | EED R1 | TATCCTCGAGGAATTAGGCAAAAGTATTTT ATCGAAGTC |
| RbAp48: | RbAp48 F1 | CTGTTCCAGGGGCCCATGGCCGACAAGGA AGCAGCC |
| | RbAp48 R1 | TATCCTCGAGGAATTCTAGGACCCTTGTCC TTCTGG |

pEFh-SBP-GFP: GFP was amplified from pMXs-IG (Cell Biolabs) and subcloned into the *Eco*RI site of pEFh-SBP (kind gift from Akio Yamashita, Yokohama City University) using Infusion (Clontech). The following primers were used.

| | | |
|---|---|---|
| GFP: | pEF-G-GFPF | TTTCAGGGCGAATTCATGGTGAGCAAGGG CGAGGA |
| | pEF-GFPR | GATTGTCGATGAATTTTACTTGTACAGCTC GTCCAT |

pColdI, pET28b, pGEX-6P-EED, -EZH2, -RbAp48, -Suz12 and -GFP: PRC2 components and GFP were amplified from pCAGEN or pEFh constructs and subcloned into the *Eco*RI site of pColdI using Infusion (Clontech). The following primers were used.

| | | |
|---|---|---|
| EED: | pColdEEDF | CGAGGGATCCGAATTCATGTCCGAGAGGG AAGTGTC |

| | | |
|---|---|---|
| | pColdEEDR | CGACAAGCTTGAATTAGGCAAAAGTATTTTATCGAAGTC |
| EZH2: | pColdEZF | CGAGGGATCCGAATTCATGGGCCAGACTGGGAAG |
| | pColdEZR | CGACAAGCTTGAATTGGGGAGGAGGTAGCAGATGT |
| RbAp48 | pColdRbF | CGAGGGATCCGAATTCATGGCCGACAAGGAAGCAGCC |
| | pColdRbR | CGACAAGCTTGAATTCTAGGACCCTTGTCCTTCTGG |
| Suz12: | pColdSUZF | CGAGGGATCCGAATTCATGGCGCCTCAGAAGCAC |
| | pColdSUZR | CGACAAGCTTGAATTTGGGGTTAGAGCTTTTCAGAGT |
| GFP: | pColdGFPF | CGAGGGATCCGAATTCATGGTGAGCAAGGGCGAGGA |
| | pColdGFPR | CGACAAGCTTGAATTTTACTTGTACAGCTCGTCCAT |

Cloning into pET28b and pGEX-6P with Infusion (Clontech) was done with the same primers, but changing the first 15nt of each primer for sequences homologous to the sequence surrounding the *Eco*RI site of each plasmid following the instructions from the manufacturer.

pDEST10-EED, -EZH2, -RbAp48, -SUZ12 and -GFP

PRC2 components and GFP were amplified from pCAGEN or pEFh constructs and subcloned into pENTR using pENTR/D-TOPO cloning kit (Invitrogen). The following primers were used.

| | | |
|---|---|---|
| EED: | EEDTOPOF1 | CACCATGTCCGAGAGGGAAGTGTC |
| | EEDTOPOR1 | AGGCAAAAGTATTTTATCGAAGTC |
| SUZ12: | SUZTOPOF1 | CACCATGGCGCCTCAGAAGCAC |
| | SUZTOPOR1 | TGGGGTTAGAGCTTTTCAGAGT |
| EZH2: | EZHTOPOF1 | CACCATGGGCCAGACTGGGAAG |
| | EZHTOPOR1 | GGGGAGGAGGTAGCAGATGT |
| RbAp48: | RbTOPOF1 | CACCATGGCCGACAAGGAAGCAGCC |
| | RbTOPOR1 | CTAGGACCCTTGTCCTTCTGG |
| GFP: | GFPTOPOF1 | CACCATGGTGAGCAAGGGCGAGGA |
| | GFPTOPOR1 | TTACTTGTACAGCTCGTCCAT |

The genes were subcloned into pDEST10 using LR recombinase (Invitrogen).

pENTR-hRepA: the A-repeat domain of human Xist was amplified from HEK293T cDNA using KOD Plus-NEO (Toyobo) and the following primers.

| | | |
|---|---|---|
| A-repeat | XistTOPOF | CACCAGTGTCTTCTTGACACGTCCTCCA |
| | XistR | AGAGTGCAACAACCCACAAAACCA |

The PCR product was cloned into pENTR using pENTR/D-TOPO cloning (Invitrogen).

pMK-RQ-8rep_mut: a mutant 8.5 rep plasmid was custom synthesized by Invitrogen's gene synthesis service.

pCRII-2_rep_mut: the following oligos were annealed and cloned using Zero Blunt TOPO PCR cloning kit (Invitrogen).

| | | |
|---|---|---|
| 2 rep | 2rep mut F | TCTTCCACTCTCTTTTCTATATTTTAGCAATCGGGGCTGCAGATACATAGTTTTATTATTTTTTCTTTAGCAAACGGGGCCGTAGATACATACC |
| | 2rep mut R | GGTATGTATCTACGGCCCCGTTTGCTAAAGAAAAAATAATAAAACTATGTATCTGCAGCCCCGATTGCTAAAATATAGAAAAGAGAGTGGAAGA |

### 3.2.3 In vitro transcription

Templates for in vitro transcription were synthesized using KOD Plus-Neo (Toyobo) and the following primers and plasmid templates.

2 rep wt: Template: pENTR-hRepA. Primers:

| | |
|---|---|
| hRepA_T7_346-372F | CTTAATACGACTCACTATAGTCTTCCACTCTCTTTTCTATATTTTGC |
| hRepA_410-439 | GGCAGGTATCCACGGCCCCGTTGGGCAAAG |

2 rep mut: Template: pCRII-2rep_mut. Primers:

| | |
|---|---|
| hRepA_T7_346-372mutF | CTTAATACGACTCACTATAGTCTTCCACTCTCTTTTCTATATTTTAG |
| hRepA_410-439mutR | GGTATGTATCTACGGCCCCGTTTGCTAAAG |

4 rep wt: Template: pENTR-hRepA. Primers:

| | |
|---|---|
| hRepA_T7_346-372F | CTTAATACGACTCACTATAGTCTTCCACTCTCTTTTCTATATTTTGC |

| | |
|---|---|
| hRepA_499-529R | AGCAGGTATCCGATACCCCGATGGGCTAAGG |

4 rep mut: Template: pMK-RQ-8rep_mut. Primers:

| | |
|---|---|
| hRepA_T7_346-372mutF | CTTAATACGACTCACTATAGTCTTCCACTCTCTTTTCTATATTTTAG |
| hRepA_499-529mutR | AGTATGTATCTGATACCCCGATTGCTTAAGG |

8.5 rep wt: Template: pENTR-hRepA. Primers:

| | |
|---|---|
| hRepA_T7_346-372F | CTTAATACGACTCACTATAGTCTTCCACTCTCTTTTCTATATTTTGC |
| hRepA_749-769 | TCCATAAAAAGCACCGATGGG |

8.5 rep mut: Template: pMK-RQ-8rep_mut. Primers:

| | |
|---|---|
| hRepA_T7_346-372mutF | CTTAATACGACTCACTATAGTCTTCCACTCTCTTTTCTATATTTTAG |
| hRepA_749-769mutR | TCCATAAAAAGCATCGATGTG |

HOTAIR 1-308: Template: pLZRS-HOTAIR (Gupta et al., 2010) (Addgene plasmid 26110). Primers:

| | |
|---|---|
| Hotair_T7_1-20F | GCGTAATACGACTCACTATAGGACTCGCCTGTGCTCTGGAG |
| Hotair_289-308R | AATAAAGACGCCCCTCCTTC |

HOTAIR 1-102: Template: pLZRS-HOTAIR (Gupta et al., 2010) (Addgene plasmid 26110). Primers:

| | |
|---|---|
| Hotair_T7_1-20 | GCGTAATACGACTCACTATAGGACTCGCCTGTGCTCTGGAG |
| Hotair_83-102R | TCTGTGGAACTCCCAGGCCT |

HOTAIR 103-204: Template: pLZRS-HOTAIR (Gupta et al., 2010) (Addgene plasmid 26110). Primers:

| | |
|---|---|
| Hotair_T7_103-122F | GCGTAATACGACTCACTATAGCCAACACCCCTGCTCCTGGC |
| Hotair_185-204 | TGCGAAGCGGAGCAGGACCT |

HOTAIR 205-308: Template: pLZRS-HOTAIR (Gupta et al., 2010) (Addgene plasmid 26110). Primers:

| | |
|---|---|
| Hotair_T7_205-224 | GCGTAATACGACTCACTATAGGTGGAATGGAACGGATTTAG |
| Hotair_289-308R | AATAAAGACGCCCCTCCTTC |

GFP 102nt: Template: pMXs-IG (Cell Biolabs). Primers:

| | |
|---|---|
| GFP_T7_2650-2670F | CTTAATACGACTCACTATAGGGCCACAAGTTCAGCGTGTCC |
| GFP_2732-2751R | CCAGGGCACGGGCAGCTTGC |

GFP 318nt: Template: pMXs-IG (Cell Biolabs). Primers:

| | |
|---|---|
| GFP_T7_2650-2670F | CTTAATACGACTCACTATAGGGCCACAAGTTCAGCGTGTCC |
| GFP_2950-2967R | GTCGATGCCCTTCAGCTC |

PCR products were gel purified with PCR and Gel Clean-up kit (Macherey Nagel) and in vitro transcribed using T7-scribe standard RNA IVT kit (Cellscript). The full sequences of all RNAs are shown in Supplementary Table 1.

### 3.2.4 Other RNAs

Other RNAs were chemically synthesized by Operon. Sequences are indicated in Supplementary Table 1.

### 3.2.5 Baculovirus system protein expression

The DH10Bac strain of *E. coli* was transformed with each of the pDEST10 constructs, and cells were shaken at 37 °C for 4 h before plating in LB agar containing 50 μg/ml kanamycin, 10 μg/ml tetracycline, 7 μg/ml gentamycin, 100 μg/ml X-gal and 40 μg/ml IPTG. Verified white colonies were grown in 15 ml

selective medium over night. Cells were pelleted and resuspended in 0.3 ml solution 1 (15 mM Tris HCl pH 7.4, 10 mM EDTA, 100 µg/ml RNase A). Then 0.3 ml of solution 2 (0.2 N NaOH, 1% SDS) were added and the mixture was incubated at room temperature for 5 min, then neutralized with 0.3 ml 3 M potassium acetate pH 5.5, and placed on ice for 5 min. After centrifugation at 14000 g for 10 min the supernatant was transferred to a new tube and precipitated with isopropanol, washed with 70 % ethanol and finally the pellet was resuspended in 40 µl ultrapure water.

To prepare P1 viral stock, one million Sf9 cells (Invitrogen) were plated per well of 6 well plates in 2 ml SF-900 II. Separately 6 µl Cellfectin (Invitrogen) and 5 µl bacmid were diluted in 100 µl SF-900 II medium. After lightly mixing and incubating for 15 min at room temperature, 800 µl SF-900 II medium were added and mix was added to cells. After incubating at 28 °C for 5 h the medium was removed and 2 ml fresh SF-900 II were added. Cells were incubated at 28 °C and the supernatant (P1) collected 72 h post transfection.

Sf9 cells were seeded in 20 ml SF-900 II at a density of 1.5 x $10^6$ cells/ml in 250 ml roller bottles and infected with 1.2 ml of the P1 stock. Cells were incubated in a rolling incubator at 28 °C for 72 h, when the supernatant was recovered (P2). P2 was further amplified by infecting 100 ml of Sf9 cells at a density of 1.5 x $10^6$ cells/ml with 7.5 ml of P2. P3 was recovered 48 h later.

For protein expression 500 ml Sf9 cells at a density of 1.5 x $10^6$ cells/ml were infected with 30 ml P3 and grown in the presence of 4% FBS. Cells were recovered 48h later, pelleted and flash frozen.

### 3.2.6 Protein expression in *E. coli*

Bacteria were transformed with pCOLD1, pET28b or pGEX-6P based constructs using standard methods and grown over night in 15 ml of selective medium. In every case, 1 L of medium containing the appropriate antibiotics was inoculated with 5 ml of preculture and cells were grown at 37 °C until they reached an O.D. between 0.4 and 0.5, and protein expression was induced by addition of IPTG (0.1-1 mM) and grown at 15 or 37 °C. After 24h cells were precipitated, washed with PBS and flash frozen.

### 3.2.7 Protein expression in HEK293T cells

Typically 8 x $10^6$ HEK293T cells were seeded per 15 cm dish in 15 ml D-MEM supplemented with 10 % FBS. Around 18 h later cells were transfected. Two milliliters of unsupplemented D-MEM was mixed with 40 µg plasmid and 240 µl PEI (7.5 mM polyethylenimine "MAX" [MW 40000] – Polysciences pH 8.0) (1:6 ratio). The transfection complexes were incubated at room temperature for 20 min and then added to cells. Cells were harvested 24 h post transfection and lysed immediately for protein purification (see below).

### 3.2.8 Protein purification

Baculovirus system and *E. coli*: Frozen cell pellets were resuspended in 10 ml buffer A (30 mM HEPES-KOH pH 7.4 at 4 °C, 200 mM potassium acetate, 2 mM magnesium acetate, 5 % glycerol, 1× Complete protease inhibitor cocktail, 1 mM DTT [or 0.2 mM TCEP for His tagged proteins] and 20 mM imidazole [only for His tagged proteins]) per 1 L *E. coli* or 500 ml Sf9 cells and sonicated for 3 min. The lysate was cleared at 10000 g for 10 min 3 times, and finally filtered with a 0.45 µm filter. Proteins were separated using HisTrap or GSTrap columns (GE Healthcare) and a gradient of buffer A to 100 % buffer B (buffer A containing 400 mM imidazole [for His tagged proteins] or 10 mM reduced glutathione [for GST tagged proteins]) in and Äkta instrument. Each fraction was run on SDS-PAGE and after CBB staining fractions containing the proteins were pooled and dialyzed against buffer C (30 mM HEPES-KOH pH 7.4 at 4°C, 100 mM potassium acetate, 2 mM magnesium acetate, 20 mM KCl, 5 % glycerol, 1 mM DTT) for ~2 h. The proteins were further purified in a MonoS column (MonoQ in the case of EED and RbAP48) (GE Healthcare) with a gradient of buffer C to 60 % buffer D (buffer C containing 1 M KCl). Fractions containing the proteins were pooled and dialyzed against 1× lysis buffer, 10 % glycerol, 1 mM DTT for 2 h-overnight, aliquoted and flash frozen. EED and RbAp48 pools were buffer exchanged using PD-10 columns (GE) instead of dialysis.

HEK293T cells: The pelleted cell weight (PCW) was estimated (typically ~6 g from sixty 15cm dishes) and cells were resuspended in 1× PCW volume 0.5 M high salt buffer (20 mM HEPES-KOH pH 8.0 at 4 °C, 1.5 mM $MgCl_2$, 0.5 M KCl, 0.2 mM EDTA, 25 % glycerol, 1 mM DTT, 1× complete protease inhibitor coctail). Cells were lysed with 15 strokes in a 7 ml douncer with a tight pestle on ice and further incubated at 4 °C with rotation for 30 min. The lysate was then centrifuged at

100000g for 1 h. The supernatant was collected and diluted with 1× PCW volume of low salt buffer (high salt buffer containing 20 mM KCl) (Abmayr et al., 2006). The lysate was mixed with 0.1× PCW volume of Streptavidin Sepharose High Performance beads (GE) and rotated at 4 °C for 10 min. The beads were precipitated and washed 3 times with 1 ml 0.5 M NaCl wash buffer (1× lysis buffer, 0.1 % tween-20, 0.5 M NaCl, 1 mM DTT, 1 mM PMSF) and rinsed twice with wash buffer (without NaCl). To elute the proteins, 500 µl elution buffer (1× lysis buffer, 2.5 mM biotin, 10 % glycerol, 1 mM DTT, 1 mM PMSF) were added, the samples were vortexed for a 2-3 seconds, spun down and the supernatant was recovered. This was repeated for a total of 5 elution fractions. Fractions were pooled, concentrated with VivaSpin 2 ml columns (Sartorius), aliquoted and flash frozen.

Quantification of the proteins was done by SDS-PAGE followed by CBB staining using BSA standard curves.

### 3.2.9 RNA labeling and quantification

RNAs were quantified using Nanodrop (Thermo). Five picomoles of chemically synthesized RNAs were radiolabeled at their 5' ends using T4 PNK (Takara) and $^{32}$P-γATP in a final volume of 10 µl for 1 h at 37 °C. Two picomoles of in vitro transcribed RNAs were radiolabeled at their 3' ends using yeast poly A polymerase (Afymetrix) and $^{32}$P-αATP in a final volume of 10 µl for 20 min at 37 °C. One microliter aliquots were taken from each reaction for the generation of standard curves. The remaining sample was run on a G-25 column (GE), run on Urea-PAGE, excised from the gel, eluted and precipitated. The radiolabeled RNAs were resuspended in 100 µl 1× lysis buffer, heated at 95 °C for 2 min and left at room temperature for 30 min to allow spontaneous refolding. The concentration of the purified RNA was estimated by running the sample in Urea-PAGE along with a standard curve. RNAs were set to a final concentration of 1-10 nM.

### 3.2.10 Gel shift and crosslinking assays

All experiments were done using low-protein-binding tubes. Typical RNA binding reactions contained 5 µl recombinant protein (~50 nM), 3 µl 40× mix, 1µl radiolabeled RNA (10nM) and 1 µl 1× lysis buffer or cold competitor RNA (final volume = 10ul). The samples were incubated at 25°C for 15 minutes, and then transferred to ice. For gel shift assays, 10 µl of sucrose loading buffer (400 mg/ml

sucrose, 2.5 mg/ml xylencyanol and 2.5 mg/ml bromophenol blue) were added and 10 µl were run in a 1× TBE 4.5% polyacrilamide gel (acrylamide/bisacrylamide 29:1) containing 2 mM $MgCl_2$, at 4 °C. Complexes formed with short RNAs (28 nt) were run for 1 h, with ~100 nt for 1 h 20 min and >200 nt for 2 h. Care was taken to run the electrophoresis as soon as possible after complex formation.

For crosslinking, after incubation at 25 °C the samples were transferred to terasaki dishes placed on ice in 10 µl aliquots, and were irradiated for 7 min with 254 nm UV light. After addition of SDS-PAGE loading buffer samples were heated at 95 °C for 2 min and run in an 8% wide range SDS-PAGE.

### 3.2.11 Filter binding assay

A Protran BA85 nitrocellulose membrane (GE Healthcare), a Hybond N+ nylon membrane (GE Healthcare) and a filter paper were pre-rinsed in 1× lysis buffer and assembled in that order (nitrocellulose membrane on top) in a dot-blotter equipped with vacuum. Each well was pre rinsed with 10 µl of 1× lysis buffer and vacuum was applied for ~15 sec. RNA binding reactions were assembled as above in a final volume of 6 µl using RNA stocks set to 1nM. After incubation for 15 min at 25 °C samples were placed on iced and 5 µl were loaded per well under vacuum. Each well was washed twice with 10 µl of 1× lysis buffer followed by ~15 sec and a final ~10 min drying step under vacuum. Membranes were air dried and exposed to phosphorimaging plates. The intensity per sample was measured using Multi Gauge (Fujifilm) and the fraction of bound RNA was calculated [fraction bound=bound RNA/(bound RNA+unbound RNA)]. Graphs were generated using IgorPro software (Wavemetrics) and data was fitted to Hill curves.

### 3.3 Recombinant protein expression and purification

To assess the RNA binding activity of the different subunits of PRC2, recombinant proteins for each of the 4 core components were expressed and purified using several different methodologies. Initially, the baculovirus expression system in Sf9 cells was tested because many proteins have been successfully produced in large quantities with the advantage that the host is eukaryotic. EED and RbAp48 were expressed at high levels, and large amounts of pure protein could be obtained after chromatographic purification (see materials and methods). In contrast, EZH2 and SUZ12 were expressed at very low levels (detectable only by western blot) and purification was not possible (Figure 17A).

Therefore, expression of EZH2 and SUZ12 was also tested using various combinations of the *E.coli* strains, vectors, and growth temperatures indicated in Table 2.

Only small amounts of SUZ12, but not of EZH2 could be purified using Rosetta2 DE3 cells, a pColdI-SUZ12 vector and induction temperature of 15 °C, but not other combinations of parameters (Figure 17B).

Finally, to try to purify EZH2 and try to obtain a larger amount of SUZ12 HEK293T cells were transfected with plasmids encoding the streptaviding binding protein (SBP) tag that binds with high affinity to streptavidin. Because EZH2 and SUZ12 are expressed at low levels in HEK293T cells around ~15 g of pelleted cells were used per purification as starting material (around one hundred and fifty 15cm dishes). In comparison, the SBP-GFP control shown in figure 17C was purified from ~0.5 g of pelleted cells (five 15cm dishes). Single step purification, followed by elution with biotin produced relatively pure proteins (Figure 17C).

Although SUZ12 expressed in *E. coli* cells was used for initial experiments, the low concentration of the protein that could be obtained (~40nM after concentrating) prevented the use of this recombinant protein for some of the experiments shown here. Therefore unless otherwise indicated PRC2 components purified from HEK293T cells were used for all experiments (EED and RbAp48 were easily purified using any of the expression systems).

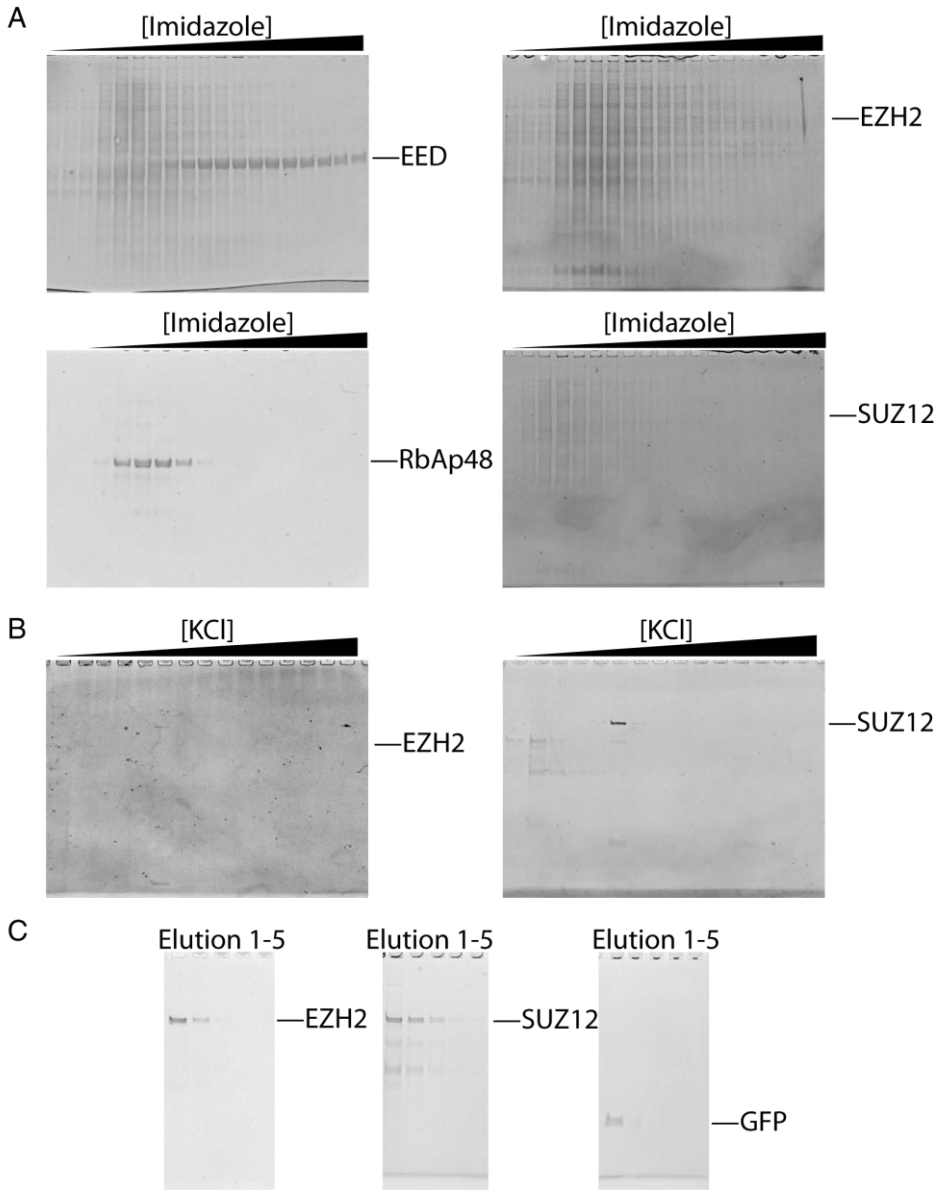



**Figure 17.** Examples of CBB staining of partially purified recombinant PRC2 components. A. His-tagged PRC2 components expressed in Sf9 cells using the baculovirus system. Proteins were initially purified on HisTrap columns and eluted with increasing concentration of imidazole. EED and RbAp48 were recovered in good amounts but no bands corresponding to EZH2 and SUZ12 were detected. EED and RbAp48 were further purified on a MonoQ column. B. His-tagged EZH2 and SUZ12 expressed in *E. coli* using pCold1 vectors. Proteins were initially purified on HisTrap columns, further purified with a MonoS column and eluted with increasing concentration of KCl. A small amount of SUZ12 could be recovered, but not of EZH2. C. SBP-tagged EZH2, SUZ12 and GFP control expressed in HEK293T cells. The three proteins were successfully purified.

**Table 2.** Different factors that were assessed for expression of EZH2 and SUZ12 in *E. coli* cells.

| ***E. coli* strain** | **Vector (tag)** | **Growth temperature** |
|---|---|---|
| Rosetta2 DE3 | pColdI (His) | 37°C |
| BL-21 | pGEX-6p (GST) | 15°C |
| JM-109 | pET28-b (His) | |

### 3.4 EZH1 and SUZ12 are RNA binding proteins

Because which is the RNA binding subunit of the PRC2 complex still remains controversial, I first set to revisit this question. To evaluate this, I first performed EMSAs, that allow the identification of protein-RNA complexes according to their migration on a native PAGE. For initial experiments recombinant EED and RbAp48 expressed in Sf9 cells, Suz12 expressed in *E. coli* and EZH2 and GFP control expressed in HEK293T cells were used (Figure 18A). Equimolar amounts of each protein were incubated with various radiolabeled substrates: an RNA fragment that contains the 8.5 repeats of A-repeat domain of Xist (8.5 rep wt) that was previously shown to be required for the silencing activity of Xist (Wutz et al., 2002); and 1 rep wt, that corresponds to the first double stem loop repeat of A-repeat and that was shown to be bound by PRC2 in vitro (Kanhere et al., 2010; Zhao et al., 2010; Zhao et al., 2008) (Figure 16 and 18B). Similar to previous observations (Kanhere et al., 2010) SUZ12, but not any of the other subunits of PRC2 or a GFP control shifted the Xist fragments, but not a control RNA of the same length as 8.5 rep mut or 1 rep with mutations in its sequence that disrupt the secondary structure (Figures 18B). Note that 1 rep mut used in this study is the same length as 1 rep wt, unlike 1 rep mut* used previously (compare 1 rep mut in Figure 18B with 1 rep mut* in Figure 16A). The 8.5 rep wt fragment is only slightly shifted due to its length (422nt).

In addition, a ~300nt fragment from the 5' end of HOTAIR previously shown to be bound by PRC2 (Tsai et al., 2010) (Figure 15B, herein called HOTAIR-A), was dissected into 3 shorter fragments (HOTAIR-B–D) (Figure 18C). Similar to A-repeat fragments, HOTAIR-A, HOTAIR-D, but not HOTAIR B or C or control RNAs, were shifted exclusively by SUZ12 (Figure 19B). Since this indicates that the element of HOTAIR bound by SUZ12 is present within the HOTAIR-D fragment, HOTAIR-D was divided into 4 additional 26nt long oligos (Figure 18C). They were used for EMSA, and SUZ12 complexes were detected with all of them. Note that an unrelated 27nt RNA that forms a stem loop structure (see Supplementary Figure 4) was also bound by SUZ12, which might be related to the ability of SUZ12 to bind many short RNAs with stem loop structures (Kanhere et al., 2010).

An important observation is that when all 4 proteins are mixed a supershift of the RNA-containing complex is expected due to the increase in the total molecular

weight; however, in all cases the complexes detected with all proteins together migrated the same as when SUZ12 was used alone. This indicates that the recombinant proteins could not assemble into an intact PRC2 complex.

However, the EMSA results indicate that the 8.5 repeat stretch of Xist, and specifically its first repeat (Kanhere et al., 2010; Zhao et al., 2008), and the 5' most 300nt fragment of HOTAIR (Tsai et al., 2010) are bound by the SUZ12 subunit of PRC2. In addition the region of HOTAIR bound by PRC2 was narrowed down to fragment HOTAIR-C, which corresponds to nucleotides 205-308 of full length HOTAIR.

The EMSA results suggest that SUZ12 is the RNA binding subunit of the complex, in agreement with at least one previous report (Kanhere et al., 2010), but in contrast with the results of other groups that identified EZH2 as the RNA binding subunit using a similar methodology (Zhao et al., 2010; Zhao et al., 2008).

Binding of SUZ12 to 1 rep wt (and lack of binding of the other components thereof) was confirmed using recombinant proteins expressed in HEK293T cells (Figure 20A, B, C). Because it is possible that purified EZH2 and SUZ12 might be contaminated with endogenous PRC2 components, the purified proteins were analyzed by western blot. Although trace amounts of endogenous EZH2 were detected in the purified SUZ12 preparation and vice versa, the major component in both cases was the intended protein (Figure 20B).

Because crosslinking experiments allow the identification of RNA binding proteins according to their molecular weight, this type of assay was used to confirm the EMSA results. For this, after incubation of the radiolabeled RNAs with each protein the samples were crosslinked with shortwave UV light and then run on SDS-PAGE. 1 rep wt RNA fragment was used as a substrate, which adds only ~9kDa to the molecular weight of the protein when covalently crosslinked. Surprisingly, bands corresponding to complexes of the expected molecular weight of EZH2-1 rep wt and SUZ12-1 rep wt were detected with both proteins, but not with EED, RbAp48 or a GFP control (Figure 20D), in disagreement with the EMSA results, where no EZH2 complexes were observed.

Crosslinking experiments with longer A-repeat and HOTAIR fragments could not be performed because they cause a big shift in the migration of EZH2 and SUZ12 in SDS-PAGE due to their large molecular weight, which prevents the effective identification of the proteins according to their molecular weight. Crosslinking followed by RNAse I treatment to trim the unprotected ends of the RNAs resulted in broad smears in the gel which were difficult to analyze.

To further confirm the identity of the RNA binding proteins and to rule out that there might be some RNA binding contaminant of molecular weight similar to EZH2 and SUZ12 in the protein preparation, the samples were immunoprecipitated with antibodies specific for EZH2 or SUZ12 after crosslinking. A signal corresponding to the expected molecular weight of the protein-RNA complexes could be effectively detected with both proteins (Figure 20E), which confirms that both EZH2 and SUZ12 are RNA binding proteins.

Together these findings reconcile the previous contradictory reports that had argued that either EZH2 or SUZ12, but not both, possesses RNA binding activity.

Differences in the methodologies used for protein expression and purification might be responsible for the previously contradictory results. As evidenced here, the EZH2-1 rep wt interaction cannot be detected by EMSA and this might be a reason why previous reports failed to recognize this dual RNA binding activity of the complex.

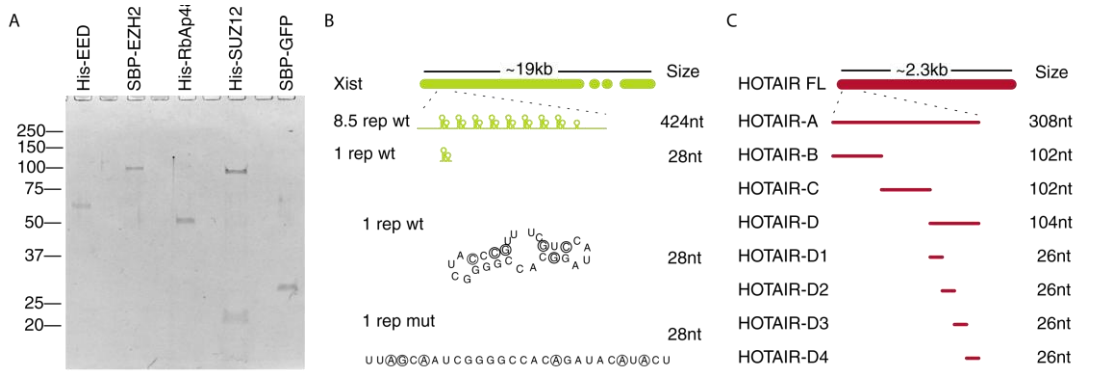



**Figure 18.** Purified PRC2 components and RNA fragments used for EMSA. A. EED and RbAp48 expressed in Sf9 cells, SUZ12 expressed in *E. coli* and EZH2 expressed in HEK293T cells were used for experiments shown in Figure 19. B. Xist RNA fragments used in Figure 19A. The sequence and structure of 1 rep wt and 1 rep mut are shown at the bottom. C. HOTAIR RNA fragments used in Figure 19B. Fragments of GFP mRNA of approximately the same length were used as controls (GFP-B: 423nt, GFP-C: 318nt, GFP-D: 102 nt) (see Supplementary figure 1 for details).

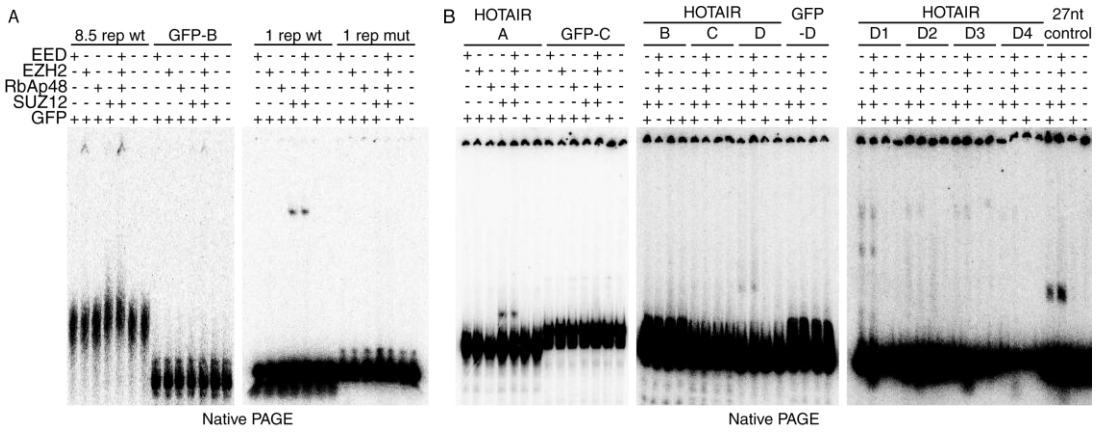



**Figure 19.** EMSA for the identification of the RNA binding subunit of PRC2. EMSA with (A) Xist and (B) HOTAIR RNA fragments.

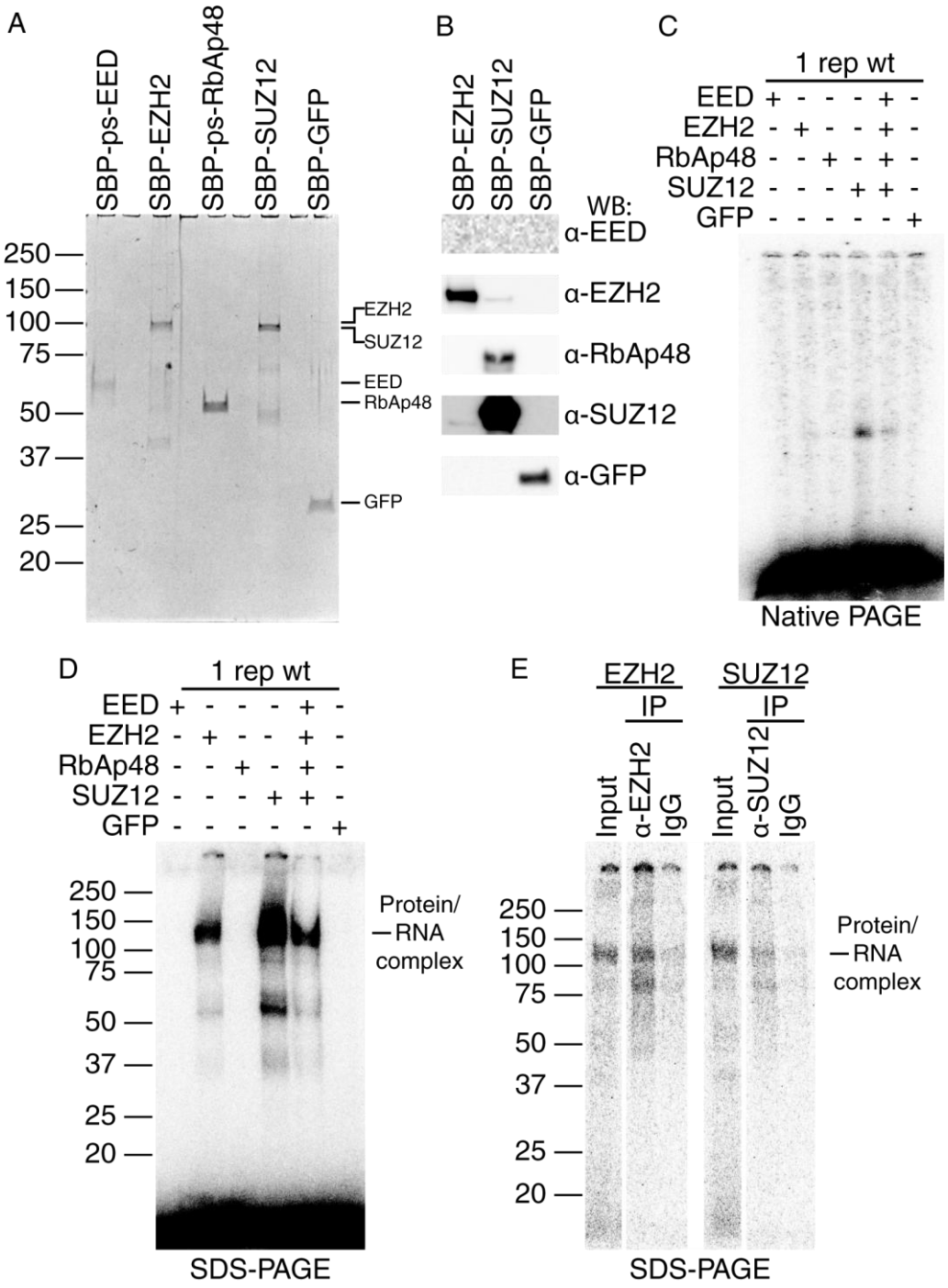



**Figure 2016.** EZH2 and SUZ12 are RNA binding proteins. A. Purified recombinant components of PRC2 purified from HEK293T cells. ps: HRV3C (PreScission) protease cleavage site. B. Western blot of purified PRC2 core proteins. Only small amounts of other endogenous PRC2 components co-purify with EZH2 and SUZ12. C, D. EMSA (C) and crosslinking (D) with 1 rep wt using proteins purified from HEK293T cells. E. IP of crosslinked EZH2- and SUZ12-1 rep wt complexes.

### 3.5 EZH2 and SUZ12 have only a slight preference for structured short RNAs

It has been suggested that PRC2 specifically recognizes the double stem loop structure of A-repeat and other short ncRNAs with similar structures (Kanhere et al., 2010; Zhao et al., 2010; Zhao et al., 2008). However, crosslinking experiments showed that 1 rep wt and 1 rep mut are bound by both EZH2 and SUZ12 (Figure 21A). To directly assess the affinity of both proteins for 1 rep wt and 1 rep mut I performed competition assays with increasing concentrations of non-radiolabeled wildtype and mutant RNAs (Figure 21B). Both RNAs compete with radiolabeled 1 rep wt and mut, for crosslinking to EZH2 and SUZ12, but the wild type RNA is a slightly stronger competitor. These data show that both proteins have only a small preference for the structured RNA which suggests that although such secondary structure might be preferred it is not an absolute determinant for binding to EZH2 or SUZ12.

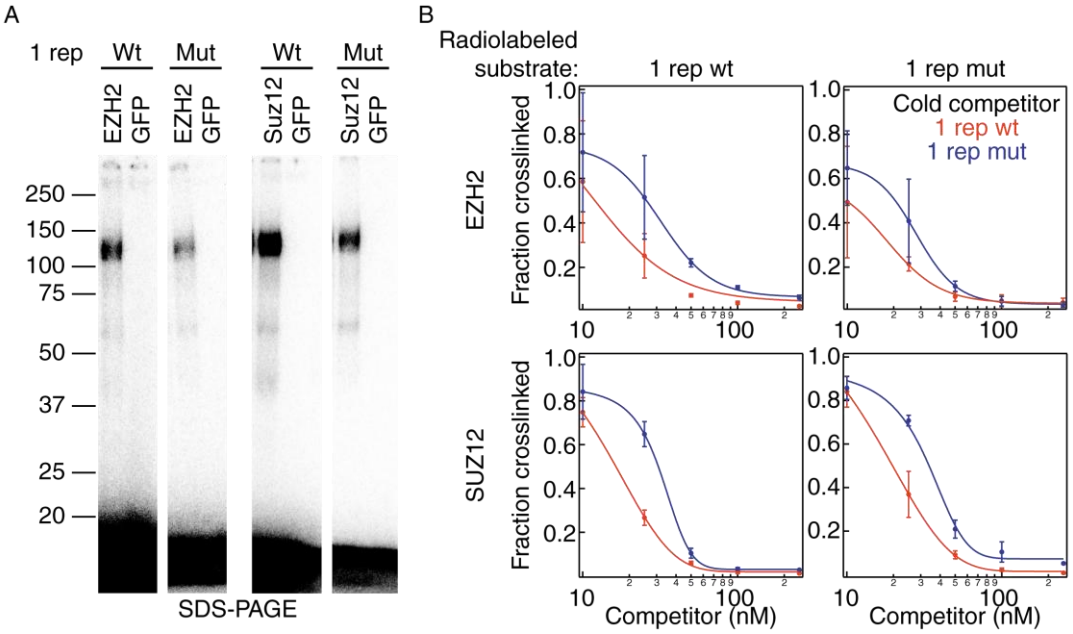



**Figure 21.** EZH2 and SUZ12 only slightly prefer 1 rep wt. A. EZH2 and SUZ12 crosslink with both 1 rep wt and mut. B. Competitions assays between radiolabeled 1 rep wt and 1 rep mut with cold RNA oligos. Mean ± s.d. (n=3).

### 3.6 EZH2 and SUZ12 have loose RNA binding specificity

EMSA results showed that the fraction of 1 rep wt bound by SUZ12 is small, and in comparison affinity for 8.5 rep wt was much higher (note that although 8.5 rep wt is only slightly shifted by SUZ12, basically no RNA remains unbound in comparison to the GFP control or Mock lanes) (Figure 19A). Each of the 8.5 repeats of the human A-repeat domain adopts a double stem loop structure similar to 1 rep wt (Wutz et al., 2002) (Figure 15A and Supplementary figure 2), and therefore the repeats can in theory act cooperatively to bind to EZH2 and SUZ12. For comparison, 1 rep, 2 rep, 4 rep and 8.5 rep wt and mut RNA fragments were prepared (Figure 22). Mutations in each of the double stem loop repeats were introduced at approximately the same positions as in 1 rep mut (Figure 18B and Supplementary table 1) and disruption of the secondary structure of each repeat was verified using RNAfold (Gruber et al., 2008) (Supplementary figure 2). Analysis of the overall secondary structure of each of the RNA fragments showed that despite the mutations, the RNA still folds into complex secondary structures different from the wildtype fragments. Also note that the RNAfold prediction of the secondary structure of any of the wt RNA fragments containing more than one repeat does not correspond to the expected sequence of double stem loop repeats (Wutz et al., 2002) (Figure 15A and Supplementary figure 3), indicating that the 8.5 double stem loop structure is not the most thermodynamically stable conformation (the one with the minimum free energy).

For a more quantitative measurement of the fraction of RNA bound by EZH2 and SUZ12, increasing amounts of recombinant proteins were incubated with each RNA and the binding strength was assessed by filter binding assays. In this methodology samples are applied using vacuum to two superposed membranes: a nitrocellulose membrane on top retains protein-bound RNA, while free RNA passes through and is captured by a positively charged nylon membrane below. The fraction of bound RNA is then calculated (Figure 22B and C).

Although 1 rep wt and mut are detectably bound by EZH2 and SUZ12, binding is very weak and ~95% of the RNA remains unbound, even at the highest protein concentration. Because the basal affinity of both proteins for a single repeat of the A-repeat domain is low, this might explain why mutations that affect the secondary

structure of 1 rep have a mild effect in the competition experiments (Figure 21B). In contrast, when 2 rep wt, 4 rep wt and 8.5 rep wt RNAs are used at equimolar concentration, the fraction of bound RNA greatly increases and reaches near saturation in the case of 8.5 rep wt (Figure 23A and B). In the 2 rep wt, 4 rep wt and 8.5 rep wt experiments the absolute concentration of the double stem loop unit is increased 2, 4 and 8.5 times respectively, but this sole increase in concentration is not responsible for the increase in the affinity to EZH2 and SUZ12 as evidenced by the weak binding of both proteins to 1 rep wt at 10 fold concentration (Figure 23A and B).

Mutations in 2 rep abolish its binding to EZH2 and SUZ12. However, in contrast, 4 rep mut and 8.5 rep mut bound as strongly as 4 rep wt and 8.5 rep wt even though the structure of each of its repeats is disrupted, and its overall secondary structure is drastically changed (see discussion below and Supplementary figures 2 and 3). Due to the length of 4 rep and 8.5 rep RNAs (184 and 424 nt respectively) it is expected that despite the introduction of mutations, new conformations are adopted, and that just like in the case of 1 rep mut, EZH2 and SUZ12 are able to bind in the absence of the double stem loop structures. I hypothesize that mutations in 4 and 8.5 rep resulted in unexpected general secondary structure elements that EZH2 and SUZ12 can bind, but that are not present in 2 rep mut.

Binding of EZH2 and SUZ12 to full length HOTAIR (FL), as well as HOTAIR-A (1-308 nt) and -D (205-308 nt) was also assessed by filter binding assays. Unlike Xist, HOTAIR does not contain any apparent sequence or structure repeats, and increasing the length of the RNA did not enhance binding over the basal binding of the HOTAIR-D fragment (Figure 23C and D). As observed in EMSA experiments, it is possible that a PRC2-binding element is restricted to this region; however, note that binding of HOTAIR fragments was weaker than binding of 8.5 rep fragments.

Finally, a set of RNA fragments of different lengths that correspond to fractions of the GFP mRNA, that are not expected to play any regulatory role of PRC2 function, were also tested. Although affinity for the shortest fragment (102 nt) was low, longer RNAs (318 nt and 645 nt long) were bound to levels similar to HOTAIR (Figure 23E and F). In comparison, control experiments using GFP protein purified from HEK293T cells following the same protocol as for purification of EZH2 and SUZ12

showed no detectable binding of Xist, HOTAIR or GFP RNA fragments (Figure 23G).

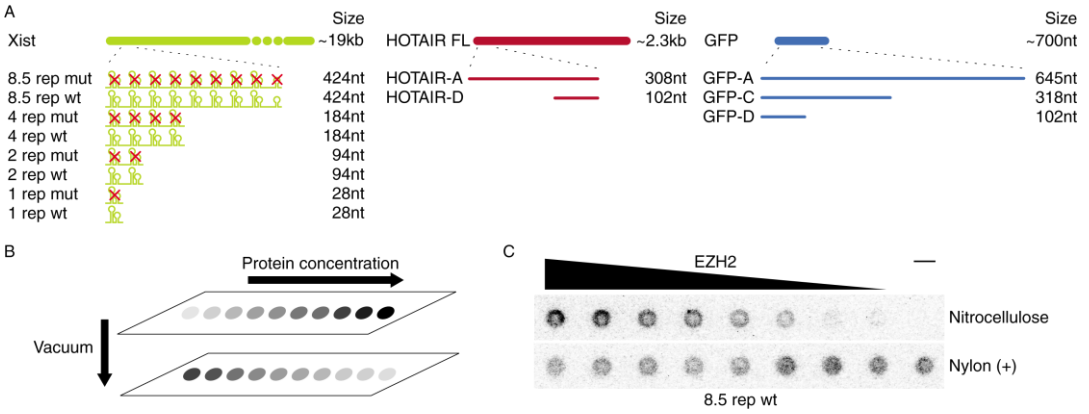



**Figure 22.** Filter binding assay method. A. Summary of RNA fragments used for filter binding assays in Figure 23. B. For filter binding assays samples are loaded onto two superposed membranes using vacuum. The membrane on top (nitrocellulose) retains protein-bound RNA and free RNA is captured by the membrane at the bottom (nylon, positively charged). C. An actual example of a filter binding assay using increasing concentration of purified EZH2 and 8.5 rep wt RNA as substrate.

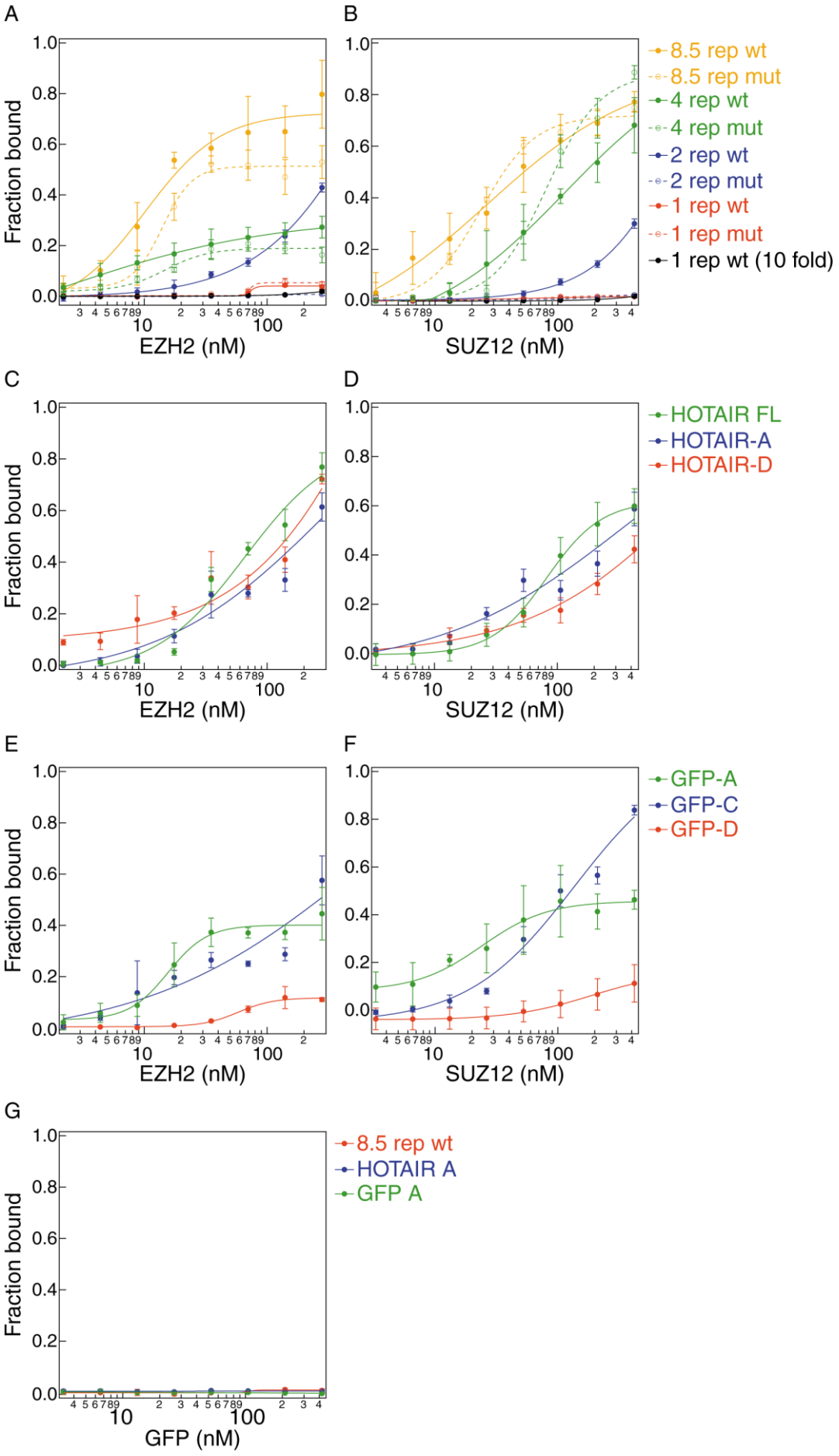

A
B
C
D
E
F
G
Fraction bound
EZH2 (nM)
SUZ12 (nM)
GFP (nM)
8.5 rep wt
8.5 rep mut
4 rep wt
4 rep mut
2 rep wt
2 rep mut
1 rep wt
1 rep mut
1 rep wt (10 fold)
HOTAIR FL
HOTAIR-A
HOTAIR-D
GFP-A
GFP-C
GFP-D
8.5 rep wt
HOTAIR A
GFP A

**Figure 23.** Filter binding assays with purified EZH2, SUZ12 and GFP. A, B. Binding of EZH2 (A) and SUZ12 (B) to Xist fragments. For 1 rep wt (10 fold) the concentration of 1 rep wt was raised from 0.1nM to 1nM for comparison. C, D. Binding of EZH2 (C) and SUZ12 (D) to HOTAIR fragments. E,F. Binding of EZH2 (E) and SUZ12 (F) to GFP mRNA fragments. G. Binding of purified GFP to selected Xist, HOTAIR and GFP mRNA fragments. Mean ± s.d. (n=>3).

### 3.7 PRC2, a dual RNA binding complex

PRC2 is one of the key regulators of epigenetic changes that regulate development and differentiation. However, how the complex is specifically recruited to genomic target loci is not well understood. lncRNAs have emerged as important co-factors for PRC2 recruitment, and in fact the PRC2 transcriptome is large, and a big portion corresponds to lncRNAs (Khalil et al., 2009; Zhao et al., 2010). At least a subset of those lncRNAs are implicated in the regulation of PRC2 target genes, as demonstrated in knock down experiments (Khalil et al., 2009).

lncRNAs can act in cis and trans. Cis acting RNAs are intrinsically tethered to their target loci during transcription and can serve as a platform for the recruitment of chromatin modifying complexes that bind towards their 5' end (Figure 24A), which in theory could be the case of Xist during the initiation of XCI, but could not explain Xist and PRC2 spreading over the whole X chromosome. However, this mechanism of action is difficult to prove in most cases because it is complicated to differentiate it from effects caused by the act of transcription itself. For example, currently there is a debate about whether the expression of the imprinted Igf2r gene is controlled by the Airn gene lncRNA RNA product (through its interaction with G9a, discussed in section 3.1.) or by Airn transcriptional overlap over the Igfr2 promoter (Latos et al., 2012; Nagano et al., 2008). In addition, both cis and trans acting lncRNAs can be tethered to their target loci through triple DNA-RNA interactions via Hoogsteen base pairing, although to date examples of this type of regulation are scarce (Schmitz et al., 2010) (Figure 24B). A final mode of regulation is the scaffold model in which lncRNAs act as scaffolds for the simultaneous recruitment of PRC2 (or other complexes) and protein complexes with affinity for specific DNA motifs (Tsai et al., 2010) (Figure 24C). Such is the case of Xist and HOTAIR, although to date it cannot be ruled out that other mechanisms of action can also be important (Brockdorff, 2013; Lee, 2012; Wang and Chang, 2011).

PRC2 is a multiprotein RNP complex. It had been previously shown that EZH2 and SUZ12 specifically bind 1 rep wt, but not unstructured mutant RNAs (Kanhere et al., 2010; Zhao et al., 2010; Zhao et al., 2008); and binding of EZH2 to a ~300nt region at the 5' end of HOTAIR was demonstrated (Kaneko et al., 2010; Tsai et al., 2010). However, some of those reports are mutually exclusive and indicate that

either one or the other, but not both proteins can bind RNAs (Kanhere et al., 2010; Zhao et al., 2010; Zhao et al., 2008). The present work establishes for the first time that in contrast to the previous assumption PRC2 has not only one, but two subunits with RNA binding activity.

Methodological differences might be responsible for this discrepancy. First, previous reports used PRC2 subunits expressed in baculovirus (Tsai et al., 2010; Zhao et al., 2008) and *E. coli* (Kanhere et al., 2010) for EMSA. Differences in posttranslational modifications depending on the source of the recombinant proteins might affect their RNA binding activity. In my work proteins expressed in HEK293T cells were used, a system that has the advantage that proteins can undergo all native posttranslational modifications; at the cost of lower purity, as small amounts of endogenous PRC2 components are co-purified (Figure 20B). To partially overcome this problem, an SBP tag was used for purification of each protein using streptavidin beads. The affinity of SBP for streptavidin is very high and allows the quick and clean purification of proteins in a single purification step (Keefe et al., 2001). In addition, the identity of the RNA binding proteins was confirmed by immunoprecipitation of the protein-RNA complexes with antibodies specific for EZH2 and SUZ12 (Figure 20E).

Moreover, as evidenced in Figures 19A and 20C, the interaction between EZH2 and 1 rep wt cannot be detected by EMSA. This assay is a very powerful technique to study the DNA or RNA binding affinity of proteins, but several factors, such as the stability of the complex during native electrophoresis, and the fact that the degree of migration is affected by multiple factors other than the molecular weight of the complex, can prevent the detection of some complexes (Hellman and Fried, 2007). Therefore EMSA results should be analyzed with caution and preferably should be confirmed with other methods. On the other hand, crosslinking with short UV light covalently links RNAs with aminoacid residues located in close proximity. Unlike in EMSA, the covalent bond fixes the protein-RNA complex and it is stable even under the denaturing conditions of SDS-PAGE, which facilitates the detection of unstable interactions, which could be the case of the binding between EZH2 and RNAs. It is possible that such instability is biologically significant, i.e. the EZH2-RNA interaction might be a transient step during the stabilization of the interaction of the RNA with PRC2, but additional studies are required to assess this idea.

In the case of EZH2 a region between aminoacids 342 and 370 has been postulated as the RNA binding domain of the protein and its activity is controlled by the cell cycle-regulated phosphorylation of threonine 345 (Kaneko et al., 2010) (Figure 13). In contrast, very little is known about the RNA binding activity of SUZ12, and in fact, about its domain structure (Figure 13), and therefore the RNA binding domain of SUZ12 remains to be identified.

Previous studies of the RNA binding activity of PRC2 focused mainly on the interaction of PRC2 subunits with a 28 nt single repeat of Xist's A-repeat. This led to the general assumption that the sequence and structural elements required for specific binding to PRC2 can be contained within such a small RNA. In fact, short RNAs (50-200nt long) that resemble the double stem loop structure of 1 rep are transcribed from many PRC2 target genes (Kanhere et al., 2010), are bound by SUZ12 and cause gene repression in cis. Although in the competition experiments (Figure 21B) the wild type RNA is a slightly stronger competitor, the double stem loop structure appears not to be an absolute requirement for binding of 1 rep to PRC2 as evidenced by the ability of 1 rep mut to compete with the wild type, structured 1 rep RNA for binding to EZH2 and SUZ12. However, the low basal affinity of 1 rep for EZH2 and SUZ12 is very low (Figure 23A and B) and instead both proteins prefer longer A-repeat fragments (Figure 23A and B). An independent study also found that fully assembled PRC2 affinity for various RNA substrates is enhanced as the length of the RNA increases (Davidovich et al., 2013), although such effect was not observed in this study with HOTAIR fragments, and was not very evident with GFP-derived RNAs (Figure 23C, D, E and F).

Interestingly, mutation of the 2 repeats of 2 rep wt completely abrogates binding (Figure 23A and B); but unexpectedly, affinity for 4 rep mut and 8.5 rep mut was similar to the affinity for the wt versions of the RNAs with both proteins. Mutations in 2 rep, 4 rep and 8.5 rep were introduced with the aim of disrupting the structure of each individual repeat, but controlling the overall structure of RNAs over 100 nt long with the introduction of mutations is challenging as invariably new secondary structures are created (Supplementary figure 3). I hypothesize that the mutations introduced into 4 rep and 8.5 rep resulted in the generation of unexpected RNA elements (possible secondary structures) that can be bound by PRC2 with low stringency, but that did not arise in the 2 rep mut fragment.

Therefore, although the contribution of the double stem loop structures for binding to PRC2 is difficult to define at this moment, the current evidence suggests that although PRC2 might bind to the double stem loop structures of A-repeat, specificity is loose and the RNA binding subunits of PRC2 can also bind promiscuously to random RNA elements present in the mutant versions of the A-repeat fragments, and even in unrelated RNAs, such as GFP mRNA (Figure 23E and F) or bacterial RNAs (Davidovich et al., 2013) that do not contain sequence or structure repeats. However, it remains intriguing that the affinity of EZH2 and SUZ12 for A-repeat fragments was maximized when the full length domain was used.

Supporting the point of view that the RNA binding specificity of PRC2 is loose, two recent independent studies have reported that in vivo promiscuous PRC2 binding to RNA (probed by the EZH2 subunit) is widespread. EZH2-bound target RNAs (not necessarily ncRNAs) arise from both active and repressed genes (Kaneko et al., 2013), and the binding sites are preferentially located in the 5' region of the RNAs and equally among introns and exons (Davidovich et al., 2013), which is consistent with a model in which EZH2 is recruited to nascent RNAs and scans transcripts throughout the genome. One interesting observation is that a fraction of the EZH2-bound RNAs arise from genes devoid of H3K27me2/3 despite having low, but detectable PRC2 occupancy (Davidovich et al., 2013; Kaneko et al., 2013). One hypothesis is that nascent RNAs act as repressors of the methyltransferase activity of EZH2, and therefore, H3K27me2/3 will only be deposited after the pre-establishment of a silent state by transcription factors and other epigenetic modifying complexes that stop transcription of the "repressive" RNAs. This model is consistent with the role of PRC2 in maintenance, rather than de novo establishment of transcriptional repression (Kaneko et al., 2013). This model, however, is not compatible with the role of some lncRNAs, such as HOTAIR, that are believed to recruit PRC2 in trans, or Xist that spreads to locations distant from its transcription site; processes that are independent of the act of transcription of the lncRNA.

Importantly, neither of the two studies excludes that some particular RNA-PRC2 interactions might be specific, and qualitatively different from the RNA scanning function of PRC2. However, whether PRC2 has two RNA-related roles is an open question. It is possible that only a subpopulation of PRC2 complexes scans nascent transcripts throughout the genome and senses the local epigenetic environment. In

fact, it is not clear whether the RNA scanning activity of EZH2 corresponds to fully assembled PRC2, or whether it might be the activity of a different EZH2-related complex or of a PRC2-independent role of EZH2 (Goff and Rinn, 2013; Kaneko et al., 2013). It is known, however, that PRC2 activity is affected by the surrounding epigenetic landscape; for example, PRC2 recognizes pre-established H3K27me2/3 (Margueron et al., 2009) and in certain conditions the methyltransferase activity of EZH2 is inhibited by the activating mark H3K4me3 (Schmitges et al., 2011; Voigt et al., 2012). In this context, deposition of new H3K27me2/3 would require the concert of several simultaneous signals that mark genes as transcriptionally silent: absence of "repressive" nascent transcripts and of the activating mark H3K4me3; plus the presence of pre-deposited H3K27me2/3. In addition, recognition of DNA sequences by the accessory proteins AEBP2 and JARID2, interaction with histone 3 and 4 by RbAp48 and possibly signals mediated by lncRNAs might also be required (Goff and Rinn, 2013; Kaneko et al., 2013) (Figure 25). Another subpopulation of PRC2 might have more specific and restricted RNA binding activity, and might use lncRNAs as co-factors to locate its target loci, but a fully assembled PRC2, additional factors and a specific intra nuclear micro-environment (still unknown) might be required in such cases. This might be the case of some lncRNA, that in contrast to nascent RNAs do not inhibit the catalytic activity of PRC2, but whose knock down instead causes de-repression of genes normally silenced by PRC2 (such as Xist and HOTAIR) (Khalil et al., 2009).

Finally, although the significance of the RNA binding by two different components of PRC2 awaits further investigation, it might be an indication of the complexity of signals that are required by the timely and precise deposition of H3K27me2/3 according to detailed regulatory signals. The RNA elements might be transferred between EZH2 and SUZ12; or tentatively, both subunits might bind cooperatively to target RNAs, or even to different classes of RNAs simultaneously, which would add an additional layer of complexity to the regulation of gene expression mediated by PRC2.

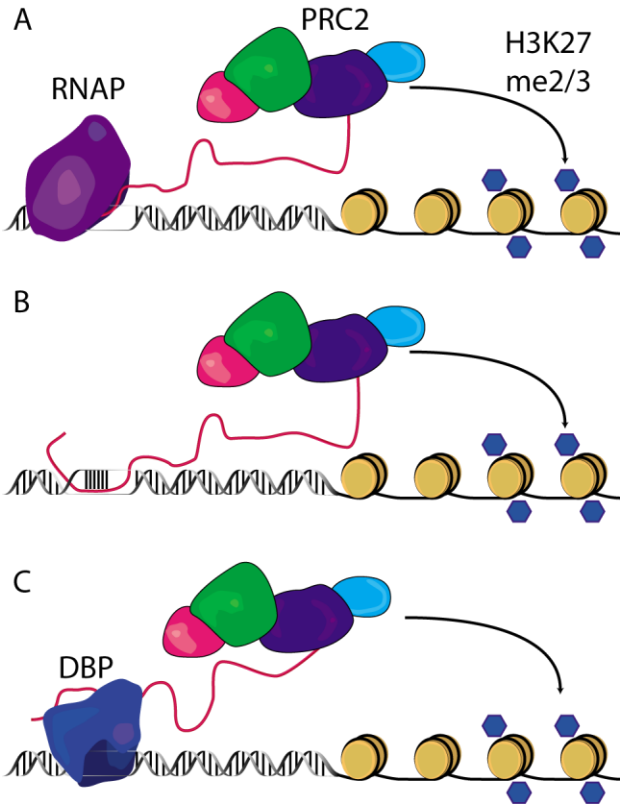



**Figure 24.** Models for the recruitment of RNA-mediated recruitment of epigenetic modifying factors. A. Co-transcriptional recruitment. B. DNA-RNA hybrid model. C. Scaffold model. RNAP: RNA polymerase. DBP: DNA binding protein.

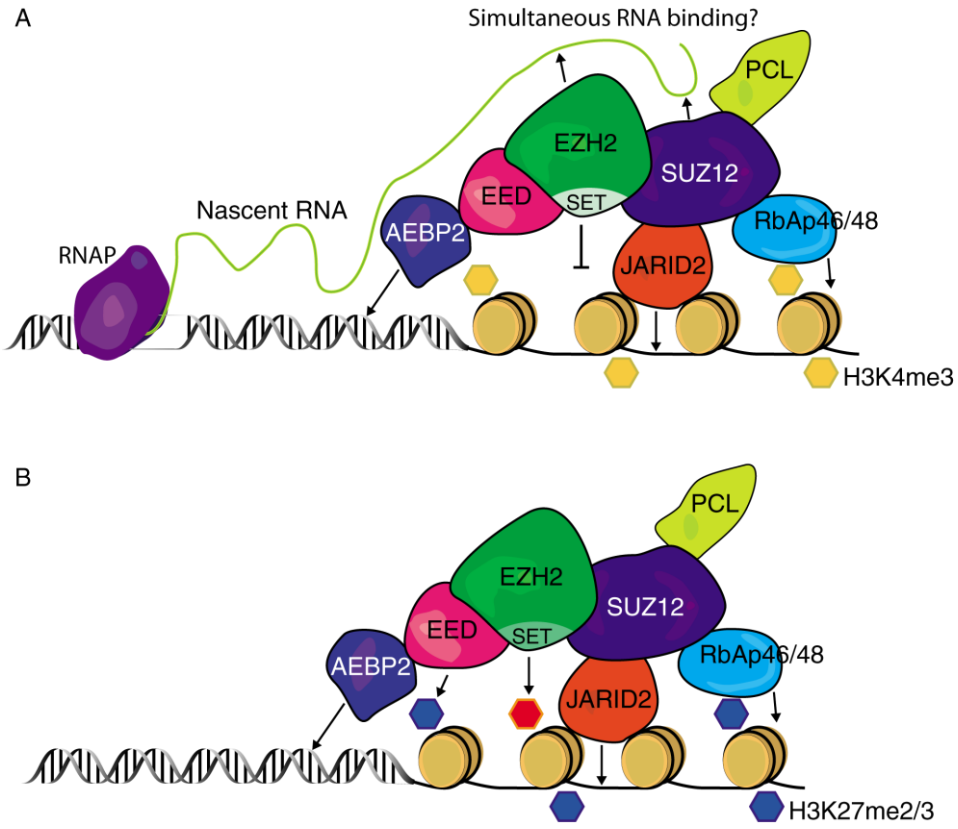



**Figure 25.** Model for the modulation of the methyltransferase activity of scanning-PRC2 by the epigenetic landscape of the target locus. A. PRC2 binds to a nascent transcript through its subunits EZH2 and SUZ12 (whether both proteins bind simultaneously is unknown) in a locus with deposited H3K4me3 (yellow), which together are marks of an actively transcribed locus. Under these conditions no H3K27me2/3 is deposited. B. In the absence of nascent transcripts and in the presence of pre-established H3K27me2/3 (blue), PRC2 deposits new H3K27me2/3 (red). Recognition of DNA by JARID2 and AEBP2; and binding to histone 3 and 4 by RbAp48 might be additional signals that help to sense the surrounding epigenetic environment. Compare with Figure 14

# 4. Concluding remarks

In 1958 Francis Crick postulated his famous theories of the sequence hypothesis and the central dogma that stated that in living cells information flows in a sequential way from DNA to RNA and finally intro proteins, and additionally that no information transfer occurs from proteins (Crick, 1970; Crick, 1958). Early works had already recognized the importance of proteins as enzymes in most cellular reactions as well as architectural scaffolds together with other macromolecules such as lipids and polysaccharides. However, the notion that DNA and RNA were mere media for the storage and transfer of information for the production of proteins remained unchallenged for a long time.

However, now more than ever it is evident that complex interconnections between all the macromolecules that form the cell are required for the existence of living beings, and that RNAs are key players in many biochemical reactions in addition to their fundamental role as mRNAs.

ncRNAs are involved in numerous cellular processes, from regulation of transcription, splicing, formation of cellular structures in the nucleus; to translation, and post-transcriptional regulation in the cytoplasm. In fact, ncRNA expression is as tightly regulated as the expression of protein coding genes (Guttman et al., 2009), and most likely ncRNA expression is also regulated by other ncRNAs.

Although some ncRNAs function as ribozymes, most of them require the concert of protein partners to exert their functions, and therefore the study of the composition and assembly of RNPs is has become a popular and exciting research topic. Each RNA class, just like protein classes, has assigned roles and it must find its molecular partners and targets within the vast and crowded micro-environment that is a live cell. And just like proteins, RNAs have several levels of organization, and therefore different layers of information that can be exploited for the formation of RNP complexes with its binding partners. The first layer is their primary sequence information; second, their secondary structure that gets more complex as the length of the RNA increases (see Supplementary figures 3, 4 and 5 for examples); and lastly, their three dimensional conformation alone or together with other RNA or protein partners.

miRNA precursors for example have to go through a series of processing steps until miRNAs are assembled into effector complexes to regulate gene expression.

Each step requires the precise spatiotemporal coordination of multiple factors and the recognition of various RNA elements that ensure that the correct RNAs are finally assembled into the appropriate RNPs. Cellular checkpoints of RNA fidelity include sensing RNA features such as RNA structure, the length of stems, and loops in hairpins, or the presence of specific chemical modifications, such as methyl or phosphoryl groups (Figure 26). Fly Dcr-1, for example, checks the loop size of pre-miRNAs and the distance between the 3' overhang and the tip of the loop to produce miRNAs of the correct size (Tsutsumi et al., 2011), and the phosphate binding pocket of Ago requires a phosphate group at the 5' end of the guide strand of small RNA duplexes during RISC loading (Kawamata et al., 2011).

Although less understood, lncRNAs are no different. Xist, HOTAIR and other nuclear RNAs undergo initial processing steps that are no different from those of protein coding mRNAs, such as capping, splicing and polyadenylation; and whether most lncRNA undergo specific processing steps remains largely unknown. However, unlike mRNAs, nuclear RNAs need to be retained in the nucleus, and some of them locate at very specific intra nuclear sites following strict cues from the surrounding environment. Due to the variety of protein-lncRNA interactions, it has been difficult to establish which are the specific sequence or structural elements that the partner proteins recognize. It is also very likely the current knowledge we have will be redefined in the next few years as more is discovered about the roles of lncRNAs. PRC2, for example, was thought to specifically recognize the double stem loop structure present in the A-repeat domain of Xist, but recent studies, including this one, have shown that the complex has the ability to bind many different RNAs that lack such motifs with apparent relaxed specificity, although under some circumstances PRC2 seems to prefer longer RNAs over short fragments (Davidovich et al., 2013) (Figure 23A and B). However, just like with miRNAs, it is expected that in some cases post transcriptional modifications, or secondary structure features are recognized by lncRNA binding partners. Additionally, due to their length, lncRNAs can bind multiple protein complexes simultaneously and can serve as platforms for the multimerization of RNP complexes (Figure 26).

Although we learned a lot in the past few years about the complexities of the world of ncRNAs, still many exciting questions remain to be answered and multiple

research avenues remain open. In addition, it is likely that still new classes of ncRNAs remain to be uncovered.

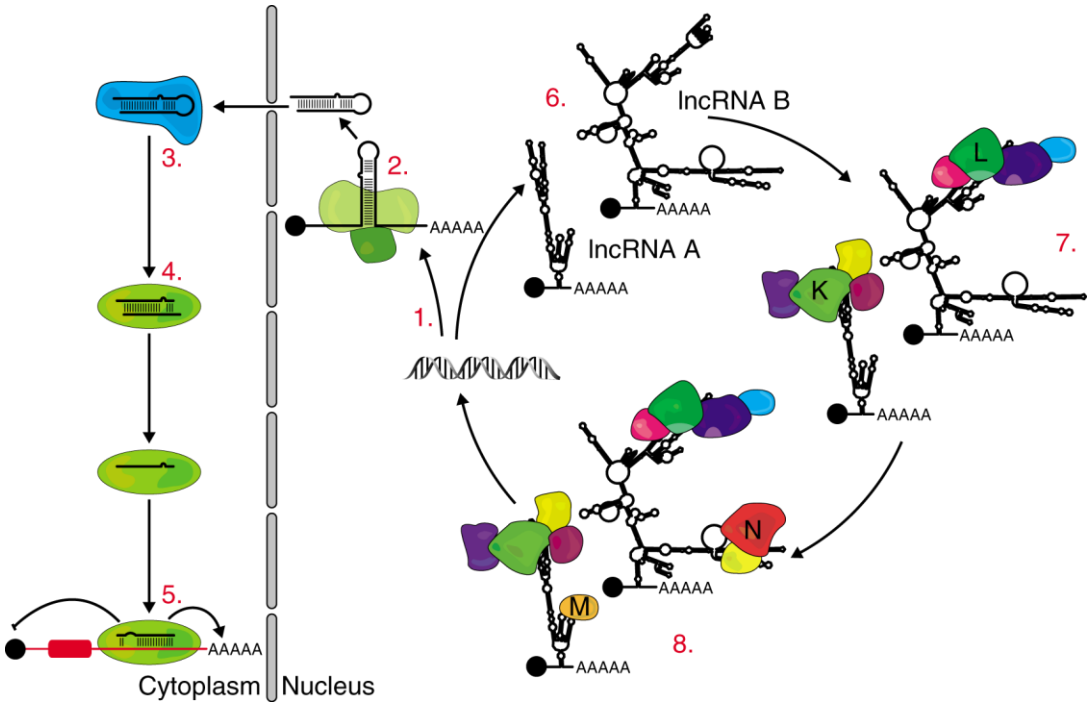
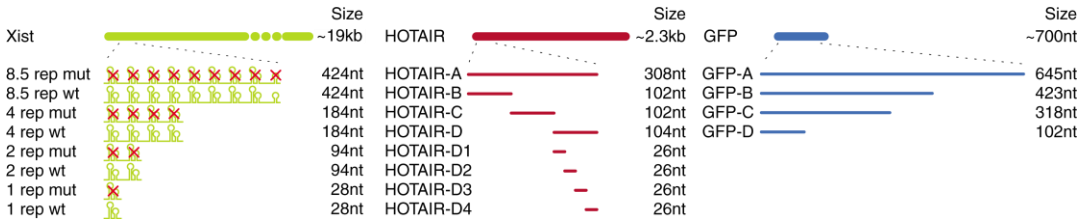


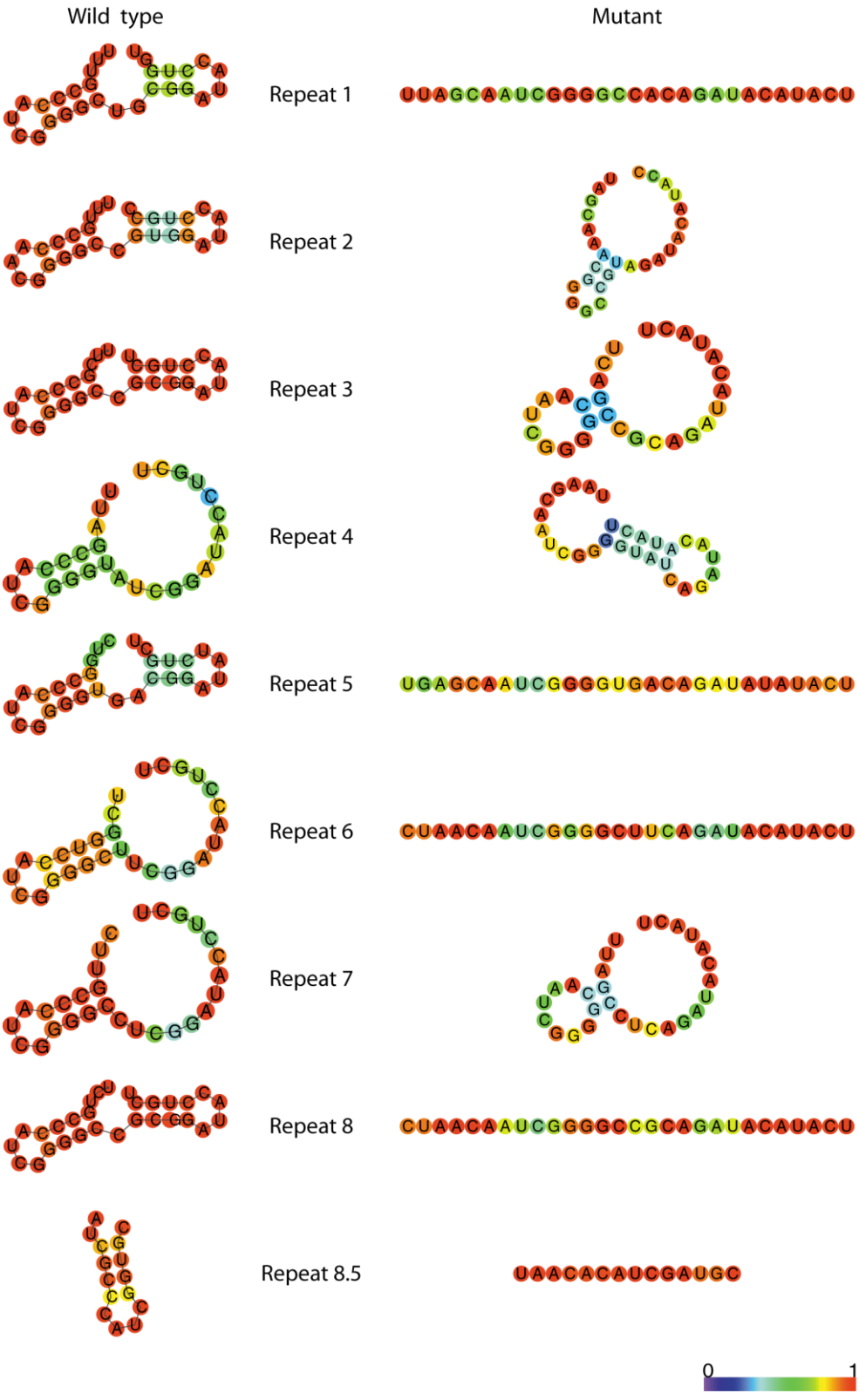

**Figure 26.** RNA elements of miRNAs and lncRNAs. Both miRNA and lncRNA precursors are transcribed in the nucleus and most are polyadenylated (AAAAA) and capped (black circle). MicroRNAs are processed by a specific subset of proteins that recognize RNA elements from the characteristic hairpin structure present in pri-miRNAs, such as the length and shape of the stem and the loop (2. and 3.). Next the miRNA duplex is loaded into Ago, which prefers double stranded RNAs and requires a 5' phosphate on the guide strand (4.). Then the miRNA duplex is unwound and the guide strand hybridizes with a target mRNA in particular by base pairing of the seed sequence (5.). On the other hand different lncRNAs (lncRNA A and B) adopt diverse secondary and tertiary structures and some of them are retained in the nucleus (6.). It is unknown whether there are specific processing factors for lncRNAs. Different lncRNAs may interact with distinct protein complexes (lncRNA A interacts with complex K, while lncRNA B binds to complex L) and can serve as scaffolds for the multimerization of RNP complexes (simultaneously, lncRNA A may bind protein M and lncRNA B may interact with complex N). Although lncRNAs have many dissimilar functions, some of them regulate epigenetic modifications

.

# 5. Supplementary information

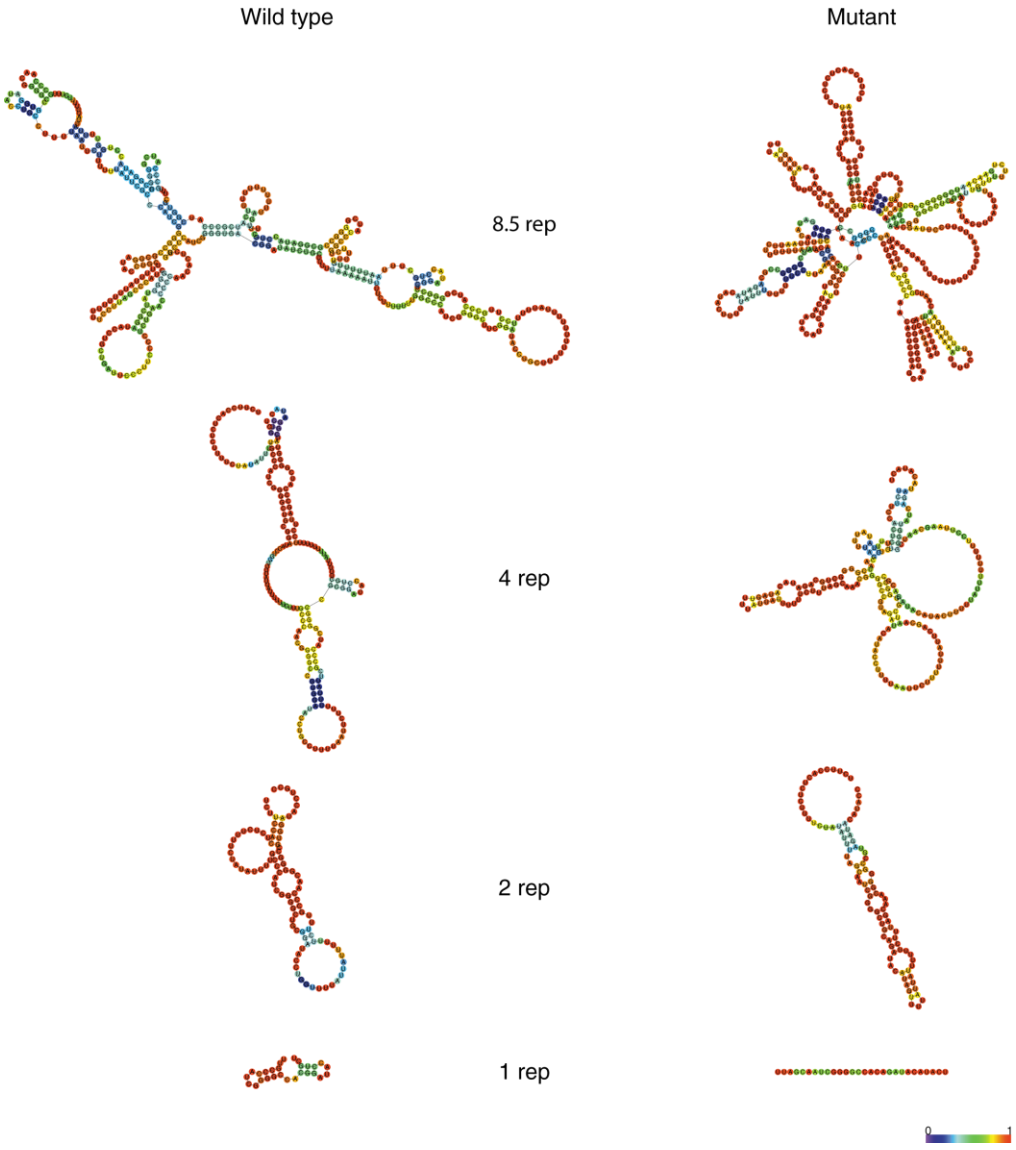



**Supplementary figure 1.** Summary of all the RNA fragments used in Part 2 of this study.

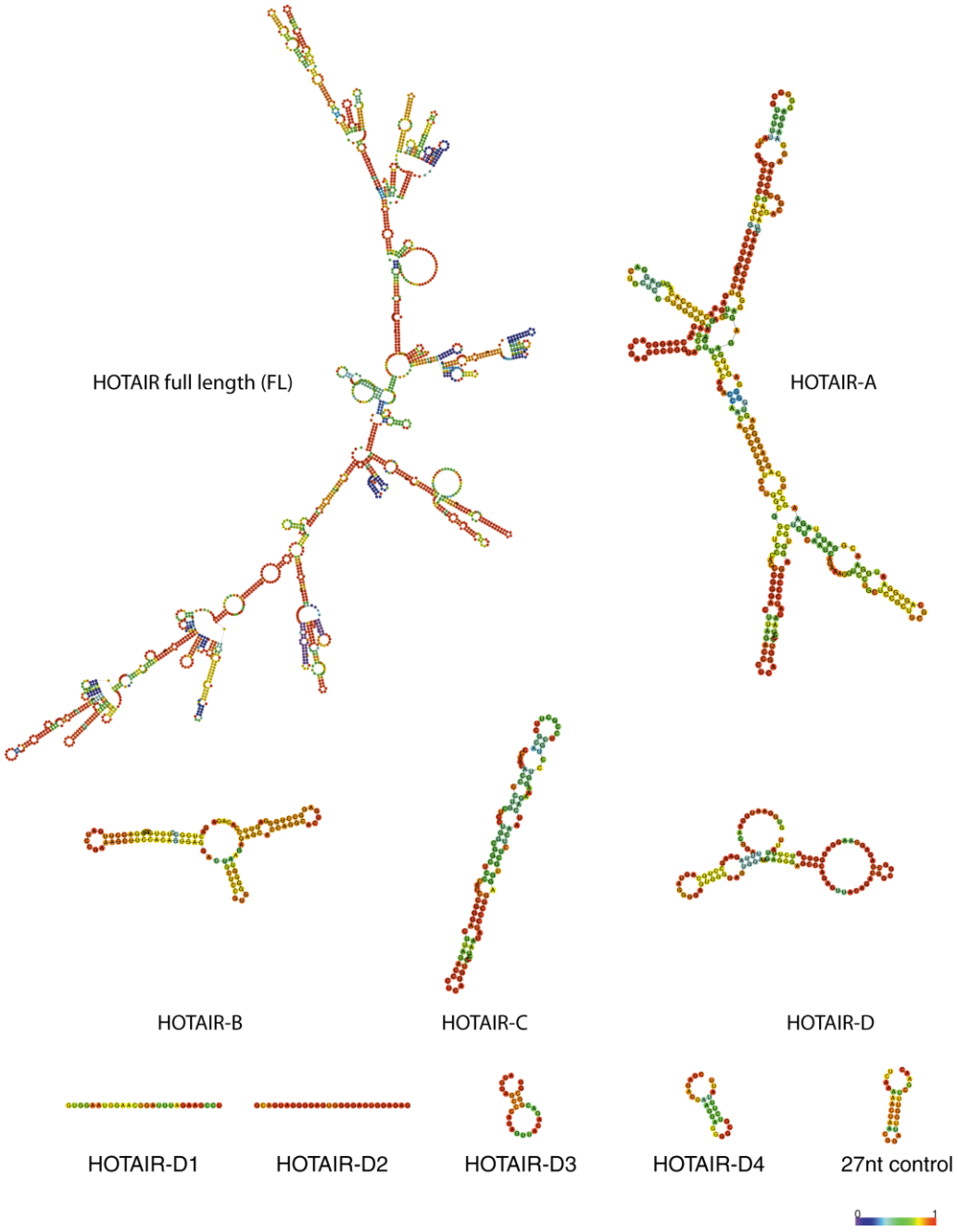



**Supplementary figure 2.** Prediction of the minimum free energy (MFE) secondary structures of each of the 8.5 stem loop structures of Xist (left). The figures on the right side correspond to the structures of mutant versions of the RNA fragments after the introduction of point mutations that disrupt the secondary structures. See

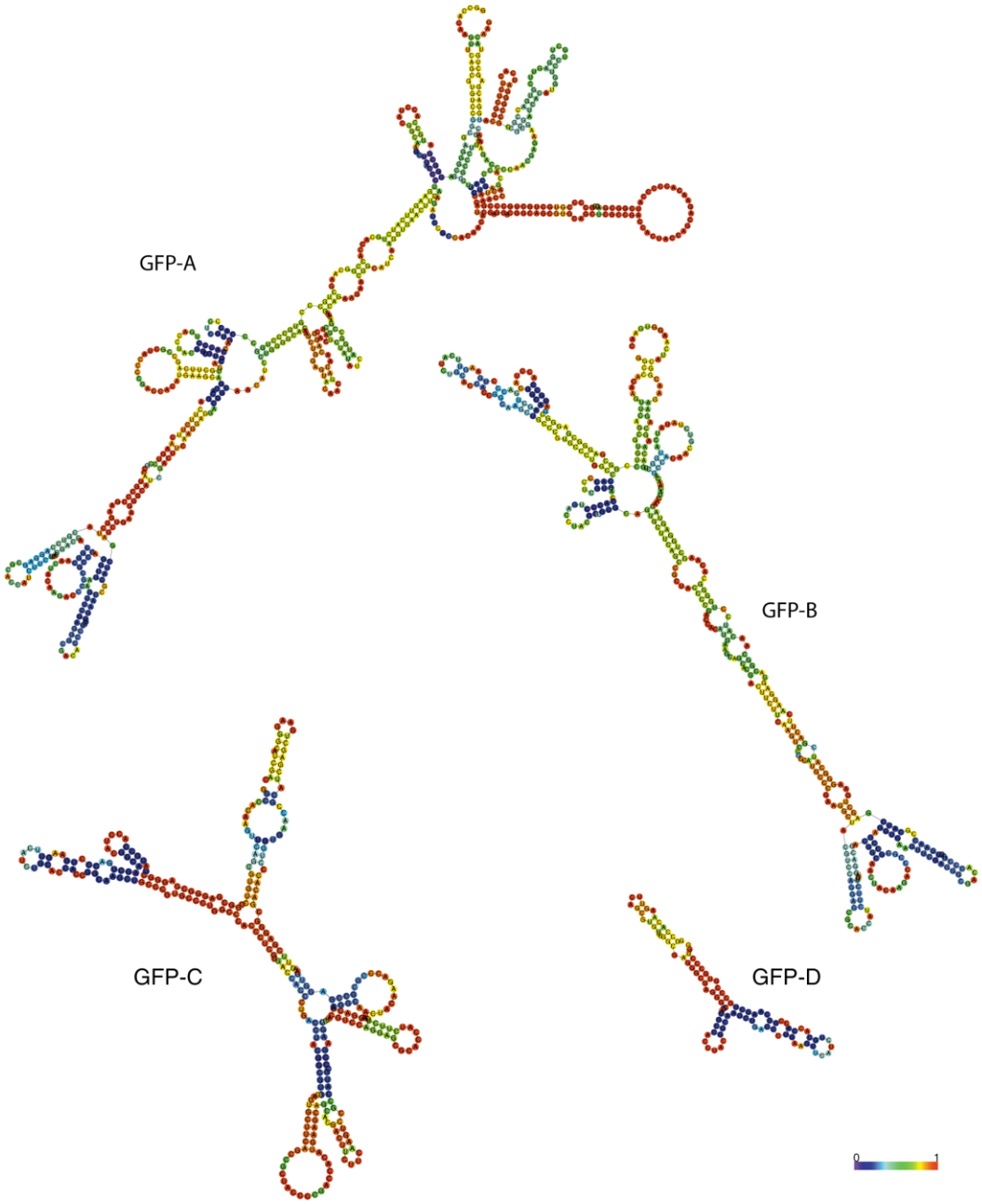
supplementary table 1 for detailed descriptions of the mutations. Predictions were perfomed using RNAfold with default parameters (Gruber et al., 2008). The color scale indicates base-pairing probabilities. In the case of unpaired regions the scale refers to the probability of being unpaired.




**Supplementary figure 3.** Prediction of the minimum free energy (MFE) secondary structures of Xist fragments used in this study. Predictions were perfomed using RNAfold with default parameters (Gruber et al., 2008). The color scale indicates base-pairing probabilities. In the case of unpaired regions the scale refers to the probability of being unpaired.




**Supplementary figure 4.** Prediction of the minimum free energy (MFE) secondary structures of HOTAIR fragments used in this study. Predictions were perfomed using RNAfold with default parameters (Gruber et al., 2008). The color scale indicates base-pairing probabilities. In the case of unpaired regions the scale refers to the probability of being unpaired.




**Supplementary figure 5.** Prediction of the minimum free energy (MFE) secondary structures of GFP fragments used in this study. Predictions were perfomed using RNAfold with default parameters (Gruber et al., 2008). The color scale indicates base-pairing probabilities. In the case of unpaired regions the scale refers to the probability of being unpaired.

**Supplementary table 1.** Sequences of the RNAs used in Part 2 of this study. Underlined sequences correspond to mutated nucleotides.

| Name | Length | Sequence |
|---|---|---|
| **8.5 rep wt** | 424 | TCTTCCACTCTCTTTTCTATATTTTGCCCATCGGGGCTGCGGATACCTGGTTTTATTATTTTTTCTTTGCCCAACGGGGCCGTGGA<br>TACCTGCCTTTTAATTCTTTTTTATTCGCCCATCGGGGCCGCGGATACCTGCTTTTTATTTTTTTTTCCTTAGCCCATCGGGGTAT<br>CGGATACCTGCTGATTCCCTTCCCCTCTGAACCCCCAACACTCTGGCCCATCGGGGTGACGGATATCTGCTTTTTAAAAATTTTCT<br>TTTTTTGGCCCATCGGGGCTTCGGATACCTGCTTTTTTTTTTTTTTATTTTTCCTTGCCCATCGGGGCCTCGGATACCTGCTTTAAT<br>TTTTGTTTTTCTGGCCCATCGGGGCCGCGGATACCTGCTTTGATTTTTTTTTTTCATCGCCCATCGGTGCTTTTTATGGA |
| **8.5 rep mut** | 424 | TCTTCCACTCTCTTTTCTATATTTTAGCAATCGGGGCTGCAGATACATAGTTTTATTATTTTTTCTTTAGCAAACGGGGCCGTAGA<br>TACATACCTTTTAATTCTTTTTTATTCAGCAATCGGGGCCGCAGATACATACTTTTTATTTTTTTTTCCTTAAGCAATCGGGGTAT<br>CAGATACATACTGATTCCCTTCCCCTCTGAACCCCCAACACTCTGAGCAATCGGGGTGACAGATATATACTTTTTAAAAATTTTCT<br>TTTTTTGAACAATCGGGGCTTCAGATACATACTTTTTTTTTTTTTTATTTTTCCTTAGCAATCGGGGCCTCAGATACATACTTTAAT<br>TTTTGTTTTTTCTGAACAATCGGGGCCGCAGATACATACTTTGATTTTTTTTTTTCATAACACATCGATGCTTTTTATGGA |
| **4 rep wt** | 184 | TCTTCCACTCTCTTTTCTATATTTTGCCCATCGGGGCTGCGGATACCTGGTTTTATTATTTTTTCTTTGCCCAACGGGGCCGTGGA<br>TACCTGCCTTTTAATTCTTTTTTATTCGCCCATCGGGGCCGCGGATACCTGCTTTTTATTTTTTTTTCCTTAGCCCATCGGGGTAT<br>CGGATACCTGCT |
| **4 rep mut** | 184 | TCTTCCACTCTCTTTTCTATATTTTAGCAATCGGGGCTGCAGATACATAGTTTTATTATTTTTTCTTTAGCAAACGGGGCCGTAGA<br>TACATACCTTTTAATTCTTTTTTATTCAGCAATCGGGGCCGCAGATACATACTTTTTATTTTTTTTTCCTTAAGCAATCGGGGTAT<br>CAGATACATACT |
| **2 rep wt** | 94 | TCTTCCACTCTCTTTTCTATATTTTGCCCATCGGGGCTGCGGATACCTGGTTTTATTATTTTTTCTTTGCCCAACGGGGCCGTGGA<br>TACCTGCC |
| **2 rep mut** | 94 | TCTTCCACTCTCTTTTCTATATTTTAGCAATCGGGGCTGCAGATACATAGTTTTATTATTTTTTCTTTAGCAAACGGGGCCGTAGA<br>TACATACC |
| **1 rep wt** | 28 | UUGCCCAUCGGGGCCACGGAUACCUGCU |
| **1 rep mut** | 28 | UUAGCAAUCGGGGCCACAGAUACAUACU |
| **HOTAIR full length (FL)** | 2145 | GACTCGCCTGTGCTCTGGAGCTTGATCCGAAAGCTTCCACAGTGAGGACTGCTCCGTGGGGGTAAGAGAGCACCAGGCACTGAGGC<br>CTGGGAGTTCCACAGACCAACACCCCTGCTCCTGGCGGCTCCCACCCGGGACTTAGACCCTCAGGTCCCTAATATCCCGGAGGTGC<br>TCTCAATCAGAAAGGTCCTGCTCCGCTTCGCAGTGGAATGGAACGGATTTAGAAGCCTGCAGTAGGGGAGTGGGGAGTGGAGAGAG<br>GGAGCCCAGAGTTACAGACGGCGGCGAGAGGAAGGAGGGGCGTCTTTATTTTTTTAAGGCCCCAAAGAGTCTGATGTTTACAAGAC<br>CAGAAATGCCACGGCCGCGTCCTGGCAGAGAAAAGGCTGAAATGGAGGACCGGCGCCTTCCTTATAAGTATGCACATTGGCGAGAG |

| Name | Length | Sequence |
|---|---|---|
| | | AAGTGCTGCAACCTAAACCAGCAATTACACCCAAGCTCGTTGGGGCCTAAGCCAGTACCGACCTGGTAGAAAAAGCAACCACGAAG<br>CTAGAGAGAGAGCCAGAGGAGGGAAGAGAGCGCCAGACGAAGGTGAAAGCGAACCACGCAGAGAAATGCAGGCAAGGGAGCAAGGC<br>GGCAGTTCCCGGAACAAACGTGGCAGAGGGCAAGACGGGCACTCACAGACAGAGGTTTATGTATTTTTATTTTTTAAAATCTGATT<br>TGGTGTTCCATGAGGAAAAGGGAAAATCTAGGGAACGGGAGTACAGAGAGAATAATCCGGGTCCTAGCTCGCCACATGAACGCCCA<br>GAGAACGCTGGAAAAACCTGAGCGGGTGCCGGGGCAGCACCCGGCTCGGGTCAGCCACTGCCCCACACCGGGCCCACCAAGCCCCG<br>CCCCTCGCGGCCACCGGGGCTTCCTTGCTCTTCTTATCATCTCCATCTTTATGATGAGGCTTGTTAACAAGACCAGAGAGCTGGCC<br>AAGCACCTCTATCTCAGCCGCGCCCGCTCAGCCGAGCAGCGGTCGGTGGGGGGACTGGGAGGCGCTAATTAATTGATTCCTTTGGA<br>CTGTAAAATATGGCGGCGTCTACACGGAACCCATGGACTCATAAACAATATATCTGTTGGGCGTGAGTGCACTGTCTCTCAAATAA<br>TTTTTCCATAGGCAAATGTCAGAGGGTTCTGGATTTTTAGTTGCTAAGGAAAGATCCAAATGGGACCAATTTTAGGAGGCCCAAAC<br>AGAGTCCGTTCAGTGTCAGAAAATGCTTCCCCAAAGGGGTTGGGAGTGTGTTTTGTTGGAAAAAAGCTTGGGTTATAGGAAAGCCT<br>TTCCCTGCTACTTGTGTAGACCCAGCCCAATTTAAGAATTACAAGGAAGCGAAGGGGTTGTGTAGGCCGGAAGCCTCTCTGTCCCG<br>GCTGGATGCAGGGGACTTGAGCTGCTCCGGAATTTGAGAGGAACATAGAAGCAAAGGTCCAGCCTTTGCTTCGTGCTGATTCCTAG<br>ACTTAAGATTCAAAAACAAATTTTTAAAAGTGAAACCAGCCCTAGCCTTTGGAAGCTCTTGAAGGTTCAGCACCCACCCAGGAATC<br>CACCTGCCTGTTACACGCCTCTCCAAGACACAGTGGCACCGCTTTTCTAACTGGCAGCACAGAGCAACTCTATAATATGCTTATAT<br>TAGGTCTAGAAGAATGCATCTTGAGACACATGGGTAACCTAATTATATAATGCTTGTTCCATACAGGAGTGATTATGCAGTGGGAC<br>CCTGCTGCAAACGGGACTTTGCACTCTAAATATAGACCCCAGCTTGGGACAAAAGTTGCAGTAGAAAAATAGACATAGGAGAACAC<br>TTAAATAAGTGATGCATGTAGACACAGAAGGGGTATTTAAAAGACAGAAATAATAGAAGTACAGAAGAACAGAAAAAAAATCAGCA<br>GATGGAGATTACCATTCCCAATGCCTGAACTTCCTCCTGCTATTAAGATTGCTAGAGAATTGTGTCTTAAACAGTTCATGAACCCA<br>GAAGAATGCAATTTCAATGTATTTAGTACACACACAGTATGTATATAAACACAACTCACAGAATATATTTTCCATACATTGGGTAG<br>GTATGCACTTTGTGTATATATAATAATGTATTTTCCATGCAGTTTTAAAATGTAGATATATTAATATCTGGATGCATTTTC |
| **HOTAIR-A** | 308 | GACTCGCCTGTGCTCTGGAGCTTGATCCGAAAGCTTCCACAGTGAGGACTGCTCCGTGGGGGTAAGAGAGCACCAGGCACTGAGGC<br>CTGGGAGTTCCACAGACCAACACCCCTGCTCCTGGCGGCTCCCACCCGGGACTTAGACCCTCAGGTCCCTAATATCCCGGAGGTGC<br>TCTCAATCAGAAAGGTCCTGCTCCGCTTCGCAGTGGAATGGAACGGATTTAGAAGCCTGCAGTAGGGGAGTGGGGAGTGGAGAGAG<br>GGAGCCCAGAGTTACAGACGGCGGCGAGAGGAAGGAGGGGCGTCTTTATT |
| **HOTAIR-B** | 102 | GACTCGCCTGTGCTCTGGAGCTTGATCCGAAAGCTTCCACAGTGAGGACTGCTCCGTGGGGGTAAGAGAGCACCAGGCACTGAGGC<br>CTGGGAGTTCCACAGA |
| **HOTAIR-C** | 102 | CCAACACCCCTGCTCCTGGCGGCTCCCACCCGGGACTTAGACCCTCAGGTCCCTAATATCCCGGAGGTGCTCTCAATCAGAAAGGT<br>CCTGCTCCGCTTCGCA |
| **HOTAIR-D** | 104 | GTGGAATGGAACGGATTTAGAAGCCTGCAGTAGGGGAGTGGGGAGTGGAGAGAGGGAGCCCAGAGTTACAGACGGCGGCGAGAGGA<br>AGGAGGGGCGTCTTTATT |
| **HOTAIR-D1** | 26 | GTGGAATGGAACGGATTTAGAAGCCT |

| Name | Length | Sequence |
|---|---|---|
| **HOTAIR -D2** | 26 | GCAGTAGGGGAGTGGGGAGTGGAGAG |
| **HOTAIR -D3** | 26 | AGGGAGCCCAGAGTTACAGACGGCGG |
| **HOTAIR -D4** | 26 | CGAGAGGAAGGAGGGGCGTCTTTATT |
| **27nt control** | 27 | UCAAAAACUAACGGAUUGGUUUCGAAC |
| **GFP-A** | 645 | GGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAA GCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGC ACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGC GCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGG GCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGA TCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTG CCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGT GACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAG |
| **GFP-B** | 423 | GGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAA GCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGC ACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGC GCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGG GCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAAC |
| **GFP-C** | 318 | GGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAA GCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGC ACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGC GCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGAC |
| **GFP-D** | 102 | GGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAA GCTGCCCGTGCCCTGG |

## 6. Acknowledgements (Agradecimientos)

Estoy muy agradecido con mi familia por todo su apoyo y por aceptar que esté tan lejos de casa por tantos años. Also to my very good friends in Colombia and the amazing new friends I have made in Japan for making every day happier.

I am also particularly thankful to Professor Yukihide Tomari for his insightful comments and ideas during the 4 years I have been member of his laboratory. Also to Mayuko Yoda for her invaluable help with experiments of the Dicer project. And all members of the laboratory, past and present, have all in one way or another contributed to the success of this research. Thank you Shintaro Iwasaki, Harumi Takeda, Tomoko Kawamata, Hiro-Oki Iwakawa, Hiroshi Sasaki, Natsuko Izumi, Eriko Matsuura, Yuichiro Mishima, Takashi Fukaya, Mayu Yoshikawa, Maki Kobayashi, Akihisa Tsutsumi, Pieter Bas Kwak, Yayoi Endo, Ken Naruse, Mariko Watanabe, Sho Niinuma, Kenji Okomori, Kyungmin Baek, Shinpei Kawaoka, Miyuki Mitomi, Yuri Kato, Kaori Kiyokawa and Masami Tashiro.

Finally I appreciate all the constructive comments from the reviewers of this thesis: Professors Shin-Ichi Nakagawa, Hitoshi Satoh, Nobuyoshi Akimitsu, Shuya Fukai and Takuya Ueda.

I am grateful to Alexander Tarakhovsky for providing the *Dicer1*$^{-/-}$ MEF cell line, Chrysi Kanellopoulou for the anti-DICER1 antisera, Kumiko Ui-Tei for pBluescript II KS(–)-Dicer, Connie Cepko for pCAGEN, Akio Yamashita for pEFh-SBP and Howard Chang for pLZRS-HOTAIR. I was supported by a scholarship from the Ministery of Education, Culture, Sports, Science and Technology (MEXT) of Japan.

# References


Abmayr, S.M., Yao, T., Parmely, T., and Workman, J.L. (2006). Preparation of nuclear and cytoplasmic extracts from mammalian cells. Curr Protoc Pharmacol *Chapter 12*, Unit12 13.

Amaral, P.P., and Mattick, J.S. (2008). Noncoding RNA in development. Mamm Genome *19*, 454-492.

Avner, P., and Heard, E. (2001). X-chromosome inactivation: counting, choice and initiation. Nat Rev Genet *2*, 59-67.

Azuara, V., Perry, P., Sauer, S., Spivakov, M., Jorgensen, H.F., John, R.M., Gouti, M., Casanova, M., Warnes, G., Merkenschlager, M., *et al.* (2006). Chromatin signatures of pluripotent cell lines. Nat Cell Biol *8*, 532-538.

Bernard, D., Prasanth, K.V., Tripathi, V., Colasse, S., Nakamura, T., Xuan, Z., Zhang, M.Q., Sedel, F., Jourdren, L., Coulpier, F., *et al.* (2010). A long nuclear-retained non-coding RNA regulates synaptogenesis by modulating gene expression. EMBO J *29*, 3082-3093.

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., *et al.* (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell *125*, 315-326.

Bertani, S., Sauer, S., Bolotin, E., and Sauer, F. (2011). The noncoding RNA Mistral activates Hoxa6 and Hoxa7 expression and stem cell differentiation by recruiting MLL1 to chromatin. Mol Cell *43*, 1040-1046.

Betancur, J.G., Yoda, M., and Tomari, Y. (2012). miRNA-like duplexes as RNAi triggers with improved specificity. Front Genet *3*, 127.

Birney, E., Stamatoyannopoulos, J.A., Dutta, A., Guigo, R., Gingeras, T.R., Margulies, E.H., Weng, Z., Snyder, M., Dermitzakis, E.T., Thurman, R.E., *et al.* (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature *447*, 799-816.

Boland, A., Huntzinger, E., Schmidt, S., Izaurralde, E., and Weichenrieder, O. (2011). Crystal structure of the MID-PIWI lobe of a eukaryotic Argonaute protein. Proc Natl Acad Sci U S A *108*, 10466-10471.

Brannan, C.I., Dees, E.C., Ingram, R.S., and Tilghman, S.M. (1990). The product of the H19 gene may function as an RNA. Mol Cell Biol *10*, 28-36.

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

Brockdorff, N., Ashworth, A., Kay, G.F., McCabe, V.M., Norris, D.P., Cooper, P.J., Swift, S., and Rastan, S. (1992). The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. Cell *71*, 515-526.

Brown, C.J., Ballabio, A., Rupert, J.L., Lafreniere, R.G., Grompe, M., Tonlorenzi, R., and Willard, H.F. (1991). A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. Nature *349*, 38-44.

Brown, C.J., Hendrich, B.D., Rupert, J.L., Lafreniere, R.G., Xing, Y., Lawrence, J., and Willard, H.F. (1992). The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. Cell *71*, 527-542.

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

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.

Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M.C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C., *et al.* (2005). The transcriptional landscape of the mammalian genome. Science *309*, 1559-1563.

Carrieri, C., Cimatti, L., Biagioli, M., Beugnet, A., Zucchelli, S., Fedele, S., Pesce, E., Ferrer, I., Collavin, L., Santoro, C., *et al.* (2012). Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat. Nature *491*, 454-457.

Casanova, M., Preissner, T., Cerase, A., Poot, R., Yamada, D., Li, X., Appanah, R., Bezstarosti, K., Demmers, J., Koseki, H., *et al.* (2011). Polycomblike 2 facilitates the recruitment of PRC2 Polycomb group complexes to the inactive X chromosome and to target loci in embryonic stem cells. Development *138*, 1471-1482.

Cattanach, B.M. (1974). Position effect variegation in the mouse. Genet Res *23*, 291-306.

Cenik, E.S., Fukunaga, R., Lu, G., Dutcher, R., Wang, Y., Tanaka Hall, T.M., and Zamore, P.D. (2011). Phosphate and R2D2 restrict the substrate specificity of Dicer-2, an ATP-driven ribonuclease. Mol Cell *42*, 172-184.

Clemson, C.M., Hutchinson, J.N., Sara, S.A., Ensminger, A.W., Fox, A.H., Chess, A., and Lawrence, J.B. (2009). An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles. Mol Cell *33*, 717-726.

Clemson, C.M., McNeil, J.A., Willard, H.F., and Lawrence, J.B. (1996). XIST RNA paints the inactive X chromosome at interphase: evidence for a novel RNA involved in nuclear/chromosome structure. J Cell Biol *132*, 259-275.

Crick, F. (1970). Central dogma of molecular biology. Nature *227*, 561-563.

Crick, F.H. (1958). On protein synthesis. Symp Soc Exp Biol *12*, 138-163.

Czech, B., Zhou, R., Erlich, Y., Brennecke, J., Binari, R., Villalta, C., Gordon, A., Perrimon, N., and Hannon, G.J. (2009). Hierarchical rules for Argonaute loading in Drosophila. Mol Cell *36*, 445-456.

Chen, L.L., and Carmichael, G.G. (2009). Altered nuclear retention of mRNAs containing inverted repeats in human embryonic stem cells: functional role of a nuclear noncoding RNA. Mol Cell *35*, 467-478.

Chendrimada, T.P., Gregory, R.I., Kumaraswamy, E., Norman, J., Cooch, N., Nishikura, K., and Shiekhattar, R. (2005). TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. Nature *436*, 740-744.

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*, 667-678.

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.

Delaval, K., Govin, J., Cerqueira, F., Rousseaux, S., Khochbin, S., and Feil, R. (2007). Differential histone modifications mark mouse imprinting control regions during spermatogenesis. EMBO J *26*, 720-729.

Dinger, M.E., Amaral, P.P., Mercer, T.R., Pang, K.C., Bruce, S.J., Gardiner, B.B., Askarian-Amiri, M.E., Ru, K., Solda, G., Simons, C., *et al.* (2008). Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. Genome Res *18*, 1433-1445.

Doench, J.G., and Sharp, P.A. (2004). Specificity of microRNA target selection in translational repression. Genes Dev *18*, 504-511.

Doi, N., Zenno, S., Ueda, R., Ohki-Hamazaki, H., Ui-Tei, K., and Saigo, K. (2003). Short-interfering-RNA-mediated gene silencing in mammalian cells requires Dicer and eIF2C translation initiation factors. Curr Biol *13*, 41-46.

Duszczyk, M.M., Wutz, A., Rybin, V., and Sattler, M. (2011). The Xist RNA A-repeat comprises a novel AUCG tetraloop fold and a platform for multimerization. RNA *17*, 1973-1982.

Egger, G., Liang, G., Aparicio, A., and Jones, P.A. (2004). Epigenetics in human disease and prospects for epigenetic therapy. Nature *429*, 457-463.

Elkayam, E., Kuhn, C.D., Tocilj, A., Haase, A.D., Greene, E.M., Hannon, G.J., and Joshua-Tor, L. (2012). The structure of human argonaute-2 in complex with miR-20a. Cell *150*, 100-110.

Feil, R., and Berger, F. (2007). Convergent evolution of genomic imprinting in plants and mammals. Trends Genet *23*, 192-199.

Forstemann, K., Horwich, M.D., Wee, L., Tomari, Y., and Zamore, P.D. (2007). Drosophila microRNAs are sorted into functionally distinct argonaute complexes after production by dicer-1. Cell *130*, 287-297.

Forstemann, K., Tomari, Y., Du, T., Vagin, V.V., Denli, A.M., Bratu, D.P., Klattenhoff, C., Theurkauf, W.E., and Zamore, P.D. (2005). Normal microRNA maturation and germ-line stem cell maintenance requires Loquacious, a double-stranded RNA-binding domain protein. PLoS Biol *3*, e236.

Ghildiyal, M., Xu, J., Seitz, H., Weng, Z., and Zamore, P.D. (2010). Sorting of Drosophila small silencing RNAs partitions microRNA* strands into the RNA interference pathway. RNA *16*, 43-56.

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

Giraldez, A.J., Mishima, Y., Rihel, J., Grocock, R.J., Van Dongen, S., Inoue, K., Enright, A.J., and Schier, A.F. (2006). Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. Science *312*, 75-79.

Goff, L.A., and Rinn, J.L. (2013). Poly-combing the genome for RNA. Nat Struct Mol Biol *20*, 1344-1346.

Gredell, J.A., Dittmer, M.J., Wu, M., Chan, C., and Walton, S.P. (2010). Recognition of siRNA asymmetry by TAR RNA binding protein. Biochemistry *49*, 3148-3155.

Gregory, R.I., Chendrimada, T.P., Cooch, N., and Shiekhattar, R. (2005). Human RISC couples microRNA biogenesis and posttranscriptional gene silencing. Cell *123*, 631-640.

Gruber, A.R., Lorenz, R., Bernhart, S.H., Neubock, R., and Hofacker, I.L. (2008). The Vienna RNA websuite. Nucleic Acids Res *36*, W70-74.

Gupta, R.A., Shah, N., Wang, K.C., Kim, J., Horlings, H.M., Wong, D.J., Tsai, M.C., Hung, T., Argani, P., Rinn, J.L., *et al.* (2010). Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature *464*, 1071-1076.

Guttman, M., Amit, I., Garber, M., French, C., Lin, M.F., Feldser, D., Huarte, M., Zuk, O., Carey, B.W., Cassady, J.P., *et al.* (2009). Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature *458*, 223-227.

Haase, A.D., Jaskiewicz, L., Zhang, H., Laine, S., Sack, R., Gatignol, A., and Filipowicz, W. (2005). TRBP, a regulator of cellular PKR and HIV-1 virus expression, interacts with Dicer and functions in RNA silencing. EMBO Rep *6*, 961-967.

Haley, B., Tang, G., and Zamore, P.D. (2003). In vitro analysis of RNA interference in Drosophila melanogaster. Methods *30*, 330-336.

Hammond, S.M., Boettcher, S., Caudy, A.A., Kobayashi, R., and Hannon, G.J. (2001). Argonaute2, a link between genetic and biochemical analyses of RNAi. Science *293*, 1146-1150.

Hasegawa, Y., Brockdorff, N., Kawano, S., Tsutui, K., and Nakagawa, S. (2010). The matrix protein hnRNP U is required for chromosomal localization of Xist RNA. Dev Cell *19*, 469-476.

Hellman, L.M., and Fried, M.G. (2007). Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. Nat Protoc *2*, 1849-1861.

Iki, T., Yoshikawa, M., Nishikiori, M., Jaudal, M.C., Matsumoto-Yokoyama, E., Mitsuhara, I., Meshi, T., and Ishikawa, M. (2010). In vitro assembly of plant RNA-induced silencing complexes facilitated by molecular chaperone HSP90. Mol Cell *39*, 282-291.

Iwasaki, S., Kobayashi, M., Yoda, M., Sakaguchi, Y., Katsuma, S., Suzuki, T., and Tomari, Y. (2010). Hsc70/Hsp90 chaperone machinery mediates ATP-dependent RISC loading of small RNA duplexes. Mol Cell *39*, 292-299.

Jenuwein, T., and Allis, C.D. (2001). Translating the histone code. Science *293*, 1074-1080.

Jeon, Y., and Lee, J.T. (2011). YY1 tethers Xist RNA to the inactive X nucleation center. Cell *146*, 119-133.

Jiang, F., Ye, X., Liu, X., Fincher, L., McKearin, D., and Liu, Q. (2005). Dicer-1 and R3D1-L catalyze microRNA maturation in Drosophila. Genes Dev *19*, 1674-1679.

Kaneko, S., Li, G., Son, J., Xu, C.F., Margueron, R., Neubert, T.A., and Reinberg, D. (2010). Phosphorylation of the PRC2 component Ezh2 is cell cycle-regulated and up-regulates its binding to ncRNA. Genes Dev *24*, 2615-2620.

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

Kanellopoulou, C., Muljo, S.A., Kung, A.L., Ganesan, S., Drapkin, R., Jenuwein, T., Livingston, D.M., and Rajewsky, K. (2005). Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing. Genes Dev *19*, 489-501.

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

Kawamata, T., Seitz, H., and Tomari, Y. (2009). Structural determinants of miRNAs for RISC loading and slicer-independent unwinding. Nat Struct Mol Biol *16*, 953-960.

Kawamata, T., and Tomari, Y. (2010). Making RISC. Trends Biochem Sci *35*, 368-376.

Kawamata, T., and Tomari, Y. (2011). Native gel analysis for RISC assembly. Methods Mol Biol *725*, 91-105.

Kawamata, T., Yoda, M., and Tomari, Y. (2011). Multilayer checkpoints for microRNA authenticity during RISC assembly. EMBO Rep *12*, 944-949.

Keefe, A.D., Wilson, D.S., Seelig, B., and Szostak, J.W. (2001). One-step purification of recombinant proteins using a nanomolar-affinity streptavidin-binding peptide, the SBP-Tag. Protein Expr Purif *23*, 440-446.

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

Khvorova, A., Reynolds, A., and Jayasena, S.D. (2003). Functional siRNAs and miRNAs exhibit strand bias. Cell *115*, 209-216.

Kim, H., Kang, K., and Kim, J. (2009a). AEBP2 as a potential targeting protein for Polycomb Repression Complex PRC2. Nucleic Acids Res *37*, 2940-2950.

Kim, T.G., Chen, J., Sadoshima, J., and Lee, Y. (2004). Jumonji represses atrial natriuretic factor gene expression by inhibiting transcriptional activities of cardiac transcription factors. Mol Cell Biol *24*, 10151-10160.

Kim, T.G., Kraus, J.C., Chen, J., and Lee, Y. (2003). JUMONJI, a critical factor for cardiac development, functions as a transcriptional repressor. J Biol Chem *278*, 42247-42255.

Kim, V.N., Han, J., and Siomi, M.C. (2009b). Biogenesis of small RNAs in animals. Nat Rev Mol Cell Biol *10*, 126-139.

Ku, M., Koche, R.P., Rheinbay, E., Mendenhall, E.M., Endoh, M., Mikkelsen, T.S., Presser, A., Nusbaum, C., Xie, X., Chi, A.S., *et al.* (2008). Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. PLoS Genet *4*, e1000242.

Kuzmichev, A., Nishioka, K., Erdjument-Bromage, H., Tempst, P., and Reinberg, D. (2002). Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. Genes Dev *16*, 2893-2905.

Latos, P.A., Pauler, F.M., Koerner, M.V., Senergin, H.B., Hudson, Q.J., Stocsits, R.R., Allhoff, W., Stricker, S.H., Klement, R.M., Warczok, K.E., *et al.* (2012). Airn transcriptional overlap, but not its lncRNA products, induces imprinted Igf2r silencing. Science *338*, 1469-1472.

Lee, J.T. (2012). Epigenetic regulation by long noncoding RNAs. Science *338*, 1435-1439.

Lee, R.C., Feinbaum, R.L., and Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell *75*, 843-854.

Lee, Y., Hur, I., Park, S.Y., Kim, Y.K., Suh, M.R., and Kim, V.N. (2006). The role of PACT in the RNA silencing pathway. EMBO J *25*, 522-532.

Lee, Y.S., Nakahara, K., Pham, J.W., Kim, K., He, Z., Sontheimer, E.J., and Carthew, R.W. (2004). Distinct roles for Drosophila Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. Cell *117*, 69-81.

Lewis, B.P., Shih, I.H., Jones-Rhoades, M.W., Bartel, D.P., and Burge, C.B. (2003). Prediction of mammalian microRNA targets. Cell *115*, 787-798.

Liu, J., Carmell, M.A., Rivas, F.V., Marsden, C.G., Thomson, J.M., Song, J.J., Hammond, S.M., Joshua-Tor, L., and Hannon, G.J. (2004). Argonaute2 is the catalytic engine of mammalian RNAi. Science *305*, 1437-1441.

Liu, Q., Rand, T.A., Kalidas, S., Du, F., Kim, H.E., Smith, D.P., and Wang, X. (2003). R2D2, a bridge between the initiation and effector steps of the Drosophila RNAi pathway. Science *301*, 1921-1925.

Lyon, M.F. (1961). Gene action in the X-chromosome of the mouse (Mus musculus L.). Nature *190*, 372-373.

MacRae, I.J., Ma, E., Zhou, M., Robinson, C.V., and Doudna, J.A. (2008). In vitro reconstitution of the human RISC-loading complex. Proc Natl Acad Sci U S A *105*, 512-517.

Maniataki, E., and Mourelatos, Z. (2005). A human, ATP-independent, RISC assembly machine fueled by pre-miRNA. Genes Dev *19*, 2979-2990.

Margueron, R., Justin, N., Ohno, K., Sharpe, M.L., Son, J., Drury, W.J., 3rd, Voigt, P., Martin, S.R., Taylor, W.R., De Marco, V., *et al.* (2009). Role of the polycomb protein EED in the propagation of repressive histone marks. Nature *461*, 762-767.

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

Martinez, J., Patkaniowska, A., Urlaub, H., Luhrmann, R., and Tuschl, T. (2002). Single-stranded antisense siRNAs guide target RNA cleavage in RNAi. Cell *110*, 563-574.

Mason, M., Schuller, A., and Skordalakes, E. (2011). Telomerase structure function. Curr Opin Struct Biol *21*, 92-100.

Matsuda, T., and Cepko, C.L. (2004). Electroporation and RNA interference in the rodent retina in vivo and in vitro. Proc Natl Acad Sci U S A *101*, 16-22.

Mattick, J.S. (2004). RNA regulation: a new genetics? Nat Rev Genet *5*, 316-323.

Meister, G., Landthaler, M., Patkaniowska, A., Dorsett, Y., Teng, G., and Tuschl, T. (2004). Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. Mol Cell *15*, 185-197.

Mercer, T.R., Dinger, M.E., and Mattick, J.S. (2009). Long non-coding RNAs: insights into functions. Nat Rev Genet *10*, 155-159.

Mercer, T.R., Dinger, M.E., Sunkin, S.M., Mehler, M.F., and Mattick, J.S. (2008). Specific expression of long noncoding RNAs in the mouse brain. Proc Natl Acad Sci U S A *105*, 716-721.

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.K., Koche, R.P., *et al.* (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature *448*, 553-560.

Miyoshi, K., Okada, T.N., Siomi, H., and Siomi, M.C. (2009). Characterization of the miRNA-RISC loading complex and miRNA-RISC formed in the Drosophila miRNA pathway. RNA *15*, 1282-1291.

Miyoshi, T., Takeuchi, A., Siomi, H., and Siomi, M.C. (2010). A direct role for Hsp90 in pre-RISC formation in Drosophila. Nat Struct Mol Biol *17*, 1024-1026.

Muller, J., and Kassis, J.A. (2006). Polycomb response elements and targeting of Polycomb group proteins in Drosophila. Curr Opin Genet Dev *16*, 476-484.

Murchison, E.P., Partridge, J.F., Tam, O.H., Cheloufi, S., and Hannon, G.J. (2005). Characterization of Dicer-deficient murine embryonic stem cells. Proc Natl Acad Sci U S A *102*, 12135-12140.

Nagano, T., Mitchell, J.A., Sanz, L.A., Pauler, F.M., Ferguson-Smith, A.C., Feil, R., and Fraser, P. (2008). The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. Science *322*, 1717-1720.

Nakanishi, K., Weinberg, D.E., Bartel, D.P., and Patel, D.J. (2012). Structure of yeast Argonaute with guide RNA. Nature *486*, 368-374.

Noland, C.L., Ma, E., and Doudna, J.A. (2011). siRNA repositioning for guide strand selection by human Dicer complexes. Mol Cell *43*, 110-121.

Nowak, A.J., Alfieri, C., Stirnimann, C.U., Rybin, V., Baudin, F., Ly-Hartig, N., Lindner, D., and Muller, C.W. (2011). Chromatin-modifying complex component Nurf55/p55 associates with histones H3 and H4 and polycomb repressive complex 2 subunit Su(z)12 through partially overlapping binding sites. J Biol Chem *286*, 23388-23396.

Okamura, K., Ishizuka, A., Siomi, H., and Siomi, M.C. (2004). Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways. Genes Dev *18*, 1655-1666.

Okamura, K., Liu, N., and Lai, E.C. (2009). Distinct Mechanisms for MicroRNA Strand Selection by Drosophila Argonautes. Molecular Cell *36*, 431-444.

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

Parker, J.S., Roe, S.M., and Barford, D. (2005). Structural insights into mRNA recognition from a PIWI domain-siRNA guide complex. Nature *434*, 663-666.

Pasini, D., Cloos, P.A., Walfridsson, J., Olsson, L., Bukowski, J.P., Johansen, J.V., Bak, M., Tommerup, N., Rappsilber, J., and Helin, K. (2010). JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells. Nature *464*, 306-310.

Pellino, J.L., Jaskiewicz, L., Filipowicz, W., and Sontheimer, E.J. (2005). ATP modulates siRNA interactions with an endogenous human Dicer complex. RNA *11*, 1719-1724.

Pham, J.W., Pellino, J.L., Lee, Y.S., Carthew, R.W., and Sontheimer, E.J. (2004). A Dicer-2-dependent 80s complex cleaves targeted mRNAs during RNAi in Drosophila. Cell *117*, 83-94.

Plath, K., Mlynarczyk-Evans, S., Nusinow, D.A., and Panning, B. (2002). Xist RNA and the mechanism of X chromosome inactivation. Annu Rev Genet *36*, 233-278.

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.

Ponjavic, J., Ponting, C.P., and Lunter, G. (2007). Functionality or transcriptional noise? Evidence for selection within long noncoding RNAs. Genome Res *17*, 556-565.

Popova, B.C., Tada, T., Takagi, N., Brockdorff, N., and Nesterova, T.B. (2006). Attenuated spread of X-inactivation in an X;autosome translocation. Proc Natl Acad Sci U S A *103*, 7706-7711.

Provost, P., Dishart, D., Doucet, J., Frendewey, D., Samuelsson, B., and Radmark, O. (2002). Ribonuclease activity and RNA binding of recombinant human Dicer. EMBO J *21*, 5864-5874.

Ringrose, L., and Paro, R. (2004). Epigenetic regulation of cellular memory by the Polycomb and Trithorax group proteins. Annu Rev Genet *38*, 413-443.

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

Russell, L.B. (1961). Genetics of mammalian sex chromosomes. Science *133*, 1795-1803.

Russell, L.B. (1963). Mammalian X-chromosome action: inactivation limited in spread and region of origin. Science *140*, 976-978.

Sado, T., Hoki, Y., and Sasaki, H. (2005). Tsix silences Xist through modification of chromatin structure. Dev Cell *9*, 159-165.

Saito, K., Ishizuka, A., Siomi, H., and Siomi, M.C. (2005). Processing of pre-microRNAs by the Dicer-1-Loquacious complex in Drosophila cells. PLoS Biol *3*, e235.

Sakurai, K., Amarzguioui, M., Kim, D.H., Alluin, J., Heale, B., Song, M.S., Gatignol, A., Behlke, M.A., and Rossi, J.J. (2011). A role for human Dicer in pre-RISC loading of siRNAs. Nucleic Acids Res *39*, 1510-1525.

Salehi-Ashtiani, K., Luptak, A., Litovchick, A., and Szostak, J.W. (2006). A genomewide search for ribozymes reveals an HDV-like sequence in the human CPEB3 gene. Science *313*, 1788-1792.

Sasaki, Y.T., Ideue, T., Sano, M., Mituyama, T., and Hirose, T. (2009). MENepsilon/beta noncoding RNAs are essential for structural integrity of nuclear paraspeckles. Proc Natl Acad Sci U S A *106*, 2525-2530.

Schirle, N.T., and MacRae, I.J. (2012). The crystal structure of human Argonaute2. Science *336*, 1037-1040.

Schmitges, F.W., Prusty, A.B., Faty, M., Stutzer, A., Lingaraju, G.M., Aiwazian, J., Sack, R., Hess, D., Li, L., Zhou, S., *et al.* (2011). Histone methylation by PRC2 is inhibited by active chromatin marks. Mol Cell *42*, 330-341.

Schmitz, K.M., Mayer, C., Postepska, A., and Grummt, I. (2010). Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes. Genes Dev *24*, 2264-2269.

Schorderet, P., and Duboule, D. (2011). Structural and functional differences in the long non-coding RNA hotair in mouse and human. PLoS Genet *7*, e1002071.

Schuettengruber, B., and Cavalli, G. (2009). Recruitment of polycomb group complexes and their role in the dynamic regulation of cell fate choice. Development *136*, 3531-3542.

Schuettengruber, B., Chourrout, D., Vervoort, M., Leblanc, B., and Cavalli, G. (2007). Genome regulation by polycomb and trithorax proteins. Cell *128*, 735-745.

Schwarz, D.S., Hutvagner, G., Du, T., Xu, Z., Aronin, N., and Zamore, P.D. (2003). Asymmetry in the assembly of the RNAi enzyme complex. Cell *115*, 199-208.

Smith, Z.D., and Meissner, A. (2013). DNA methylation: roles in mammalian development. Nat Rev Genet *14*, 204-220.

Struhl, K. (2007). Transcriptional noise and the fidelity of initiation by RNA polymerase II. Nat Struct Mol Biol *14*, 103-105.

Sun, B.K., Deaton, A.M., and Lee, J.T. (2006). A transient heterochromatic state in Xist preempts X inactivation choice without RNA stabilization. Mol Cell *21*, 617-628.

Sunwoo, H., Dinger, M.E., Wilusz, J.E., Amaral, P.P., Mattick, J.S., and Spector, D.L. (2009). MEN epsilon/beta nuclear-retained non-coding RNAs are up-regulated upon muscle differentiation and are essential components of paraspeckles. Genome Res *19*, 347-359.

Teixeira, A., Tahiri-Alaoui, A., West, S., Thomas, B., Ramadass, A., Martianov, I., Dye, M., James, W., Proudfoot, N.J., and Akoulitchev, A. (2004). Autocatalytic RNA cleavage in the human beta-globin pre-mRNA promotes transcription termination. Nature *432*, 526-530.

Tomari, Y., Du, T., Haley, B., Schwarz, D.S., Bennett, R., Cook, H.A., Koppetsch, B.S., Theurkauf, W.E., and Zamore, P.D. (2004a). RISC assembly defects in the Drosophila RNAi mutant armitage. Cell *116*, 831-841.

Tomari, Y., Du, T., and Zamore, P.D. (2007). Sorting of Drosophila small silencing RNAs. Cell *130*, 299-308.

Tomari, Y., Matranga, C., Haley, B., Martinez, N., and Zamore, P.D. (2004b). A protein sensor for siRNA asymmetry. Science *306*, 1377-1380.

Tripathi, V., Ellis, J.D., Shen, Z., Song, D.Y., Pan, Q., Watt, A.T., Freier, S.M., Bennett, C.F., Sharma, A., Bubulya, P.A., *et al.* (2010). The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. Mol Cell *39*, 925-938.

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.

Tsutsumi, A., Kawamata, T., Izumi, N., Seitz, H., and Tomari, Y. (2011). Recognition of the pre-miRNA structure by Drosophila Dicer-1. Nat Struct Mol Biol *18*, 1153-1158.

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.

Wang, J., Mager, J., Chen, Y., Schneider, E., Cross, J.C., Nagy, A., and Magnuson, T. (2001). Imprinted X inactivation maintained by a mouse Polycomb group gene. Nat Genet *28*, 371-375.

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

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

Wang, Y., Juranek, S., Li, H., Sheng, G., Tuschl, T., and Patel, D.J. (2008). Structure of an argonaute silencing complex with a seed-containing guide DNA and target RNA duplex. Nature *456*, 921-926.

Welker, N.C., Pavelec, D.M., Nix, D.A., Duchaine, T.F., Kennedy, S., and Bass, B.L. (2010). Dicer's helicase domain is required for accumulation of some, but not all, C. elegans endogenous siRNAs. RNA *16*, 893-903.

Wilson, D.N., and Doudna Cate, J.H. (2012). The structure and function of the eukaryotic ribosome. Cold Spring Harb Perspect Biol *4*.

Wutz, A., Rasmussen, T.P., and Jaenisch, R. (2002). Chromosomal silencing and localization are mediated by different domains of Xist RNA. Nat Genet *30*, 167-174.

Yang, J.S., Maurin, T., Robine, N., Rasmussen, K.D., Jeffrey, K.L., Chandwani, R., Papapetrou, E.P., Sadelain, M., O'Carroll, D., and Lai, E.C. (2010). Conserved vertebrate mir-451 provides a platform for Dicer-independent, Ago2-mediated microRNA biogenesis. Proc Natl Acad Sci U S A *107*, 15163-15168.

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

Ye, X., Huang, N., Liu, Y., Paroo, Z., Huerta, C., Li, P., Chen, S., Liu, Q., and Zhang, H. (2011). Structure of C3PO and mechanism of human RISC activation. Nat Struct Mol Biol *18*, 650-657.

Yi, R., O'Carroll, D., Pasolli, H.A., Zhang, Z., Dietrich, F.S., Tarakhovsky, A., and Fuchs, E. (2006). Morphogenesis in skin is governed by discrete sets of differentially expressed microRNAs. Nat Genet *38*, 356-362.

Yoda, M., Kawamata, T., Paroo, Z., Ye, X., Iwasaki, S., Liu, Q., and Tomari, Y. (2010). ATP-dependent human RISC assembly pathways. Nat Struct Mol Biol *17*, 17-23.

Yoon, J.H., Abdelmohsen, K., Srikantan, S., Yang, X., Martindale, J.L., De, S., Huarte, M., Zhan, M., Becker, K.G., and Gorospe, M. (2012). LincRNA-p21 suppresses target mRNA translation. Mol Cell *47*, 648-655.

Zhang, H., Kolb, F.A., Brondani, V., Billy, E., and Filipowicz, W. (2002). Human Dicer preferentially cleaves dsRNAs at their termini without a requirement for ATP. EMBO J *21*, 5875-5885.

Zhao, J., Ohsumi, T.K., Kung, J.T., Ogawa, Y., Grau, D.J., Sarma, K., Song, J.J., Kingston, R.E., Borowsky, M., and Lee, J.T. (2010). Genome-wide identification of polycomb-associated RNAs by RIP-seq. Mol Cell *40*, 939-953.

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.