論文の内容の要旨

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

There is increasing evidence that most of the mammalian genome is transcribed, and that a big portion of the transcripts have low protein-coding potential, and are therefore collectively called non-coding RNAs (ncRNAs). Although some of them posses catalytic activities, most require partner proteins to exert their functions. Among short ncRNAs (<200nt) are micro RNAs (miRNAs) and small interfering RNAs (siRNAs) that together with an Argonaute protein (Ago) form the RNA Induced Silencing Complex (RISC) and regulate gene expression at the post-transcriptional level. On the other hand, ncRNAs >200nt are known as long ncRNAs (lncRNAs), some of which interact with diverse histone or DNA modifying complexes and are involved in gene regulation at the transcriptional level. Two of the best known lncRNAs that fall in this category are X-inactive specific transcript (Xist) and HOX antisense intergenic RNA (HOTAIR). My research focuses on the protein-RNA interactions that occur between miRNAs, siRNAs and the protein Dicer during RISC assembly, and on the binding between the Polycomb Repressive Complex 2 (PRC2) and lncRNAs that leads to transcriptional silencing of PRC2 target genes.

Part I. Dicer is dispensable for asymmetric RISC assembly

MicroRNA primary transcripts are successively processed by two type III RNAses, Drosha in the nucleus and Dicer in the cytoplasm, which results in the production of mature miRNA duplexes. They, as well as exogenous siRNA duplexes are then loaded into an Argonaute (Ago) protein, of which there are 4 in mammals, and form pre-RISC. The duplex is then unwound and one strand of the duplex is discarded (passenger strand), resulting in the formation of mature RISC. Therefore, only one of the strands remains incorporated in the complex and serves as guide for RISC to regulate the expression of target mRNAs by target cleavage, translational repression and/or deadenylation in a sequence specific manner. Insights into the mechanism of guide strand selection have come mainly from studies of the Ago2 pathway in flies, where asymmetric assembly of Ago2-RISC requires the Dicer2-R2D2 heterodimer, which senses the thermodynamic asymmetry of the ends of the small RNA duplex, such that the strand with its 5’ end towards the least stable end of the duplex serves as guide (Figure 1B). However, the mechanism of asymmetry sensing in mammals remains unknown, and by analogy it has been postulated that in mammals Dicer might play a similar function as in the fly Ago2 pathway.

To clarify the role of Dicer in RISC assembly in mammals, the incorporation into mature RISC of both the guide and passenger strands of a miRNA-like duplex was assessed by cotransfection of *Renilla* luciferase reporter constructs bearing a target site for the guide or the passenger strand in *Dicer*$^{-/-}$ MEF cells with or without Dicer expression rescued. My results show that from a functional point of view RISC was assembled asymmetrically in presence or absence of Dicer as evidenced by the higher potency of strand 1 to silence the luciferase reporter (Figure 1A). The same experiment using an siRNA duplex produced a similar result. Additionally, further

analyses in vitro using a native gel system that allows the identification of pre- and mature RISC were performed. It was observed that the absence of Dicer neither affects RISC loading or its transition to mature RISC, nor the speed of assembly. These results indicate that Dicer is dispensable for asymmetry sensing and RISC assembly in mammals, and that the mechanism for sensing the thermodynamic stability of the ends of the small RNA duplex and that dictates its orientation in RISC is different from the Ago2-RISC pathway in flies (Figure 1B, C).

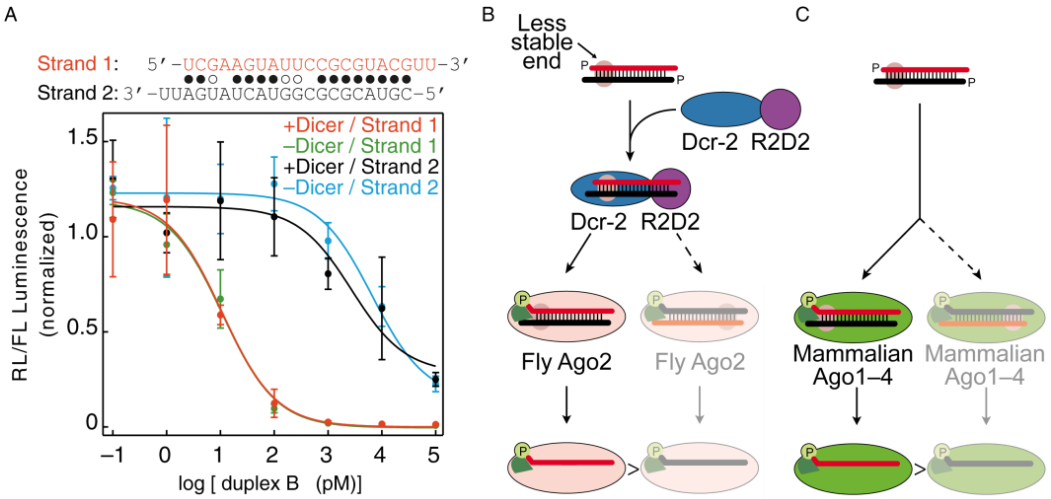

Figure 1. RISC assembles asymmetrically in the absence of Dicer. A. A miRNA-like duplex was functionally asymmetric to silence Renilla luciferase reporter constructs in *Dicer*$^{-/-}$ MEF cells after rescuing (+Dicer) or not (–Dicer) the expression of Dicer. B, C. Small RNA duplexes are preferentially 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 (>). B. In flies the Dicer-2/R2D2 heterodimer senses the asymmetry of siRNA duplexes and such binding is a prerequisite for Ago2-RISC assembly. C. In contrast to the fly Ago2 pathway, in mammals Dicer is dispensable for asymmetric RISC loading.

Part II. The RNA binding activity of the PRC2 complex

Xist is one of the best studied models of lncRNAs that are important factors in the regulation of epigenetic states. In humans Xist is required for X chromosome dosage compensation in female mammals. During early development one of the two X chromosomes is selected to be silenced and the selected chromosome is coated by Xist. Histone modifying complexes are recruited, including PRC2 that induces di/tri-methylation of lysine 27 of histone 3 (H3K27me2/3), which effectively silences gene expression from the target chromosome. Another lncRNA, HOTAIR, transcribed from within the HOX C locus on chromosome 12, recruits PRC2 to the HOX D locus, on chromosome 2. One model for the recruitment of PRC2 indicates that Xist and HOTAIR, tethered to their target loci via interaction with protein complexes with DNA binding activity, directly bind PRC2 and provide a platform for targeting H3K27me2/3 to adjacent regions. The domains of HOTAIR and Xist bound by PRC2 have been mapped to the 5’ ends of the transcripts. In the case of Xist it contains a stretch of 8.5 repeats of a double stem loop, known as A-repeat. It has also been postulated that a single repeat of Xist is enough for direct binding of PRC2.

The effector complex, PRC2, is composed of 4 core components: EED, EZH2, RbAp46/48 and SUZ12. EED recognizes pre-established histone modifications, EZH2 is the histone methyltransferase subunit, and RbAp48 binds to histone 3 and 4. Additionally, electrophoretic mobility shift assays (EMSA) have shown that SUZ12, but not EZH2 is responsible for the RNA binding activity of the complex. However, the opposite has also been shown by different groups, and therefore the existing reports regarding the RNA binding activity of PRC2 are contradictory.

To clarify those contradictory findings I used recombinant proteins for the four components of PRC2. The recombinant proteins were incubated with a radiolabeled 28nt RNA fragment that corresponds to the first repeat of Xist (1 rep wt). By EMSA, protein-RNA complexes were detected exclusively in the presence of SUZ12. However, UV crosslinking experiments showed that not only SUZ12, but also EZH2 was able to bind 1 rep wt, but not EED, RbAp48 or a GFP control. Since the EZH2-1 rep wt interaction cannot be detected by EMSA, this might be a reason why previous reports failed to recognize this dual RNA binding activity of the complex.

In addition, it has also been suggested that PRC2 specifically recognizes the double stem loop structure of the repeats of Xist. To asses this hypothesis, binding of EZH2 and SUZ12 to 1 rep wt and 1 rep mut, a mutant version predicted not to form any stable secondary structure, was assayed by crosslinking. Although binding was detected with both forms of the RNA, competition assays between wt and mut RNAs showed that both proteins have a slight preference for the structured RNA. Therefore the double stem loop structure might be a favorable secondary structure for binding to EZH2 and SUZ12, but it is not an absolute determinant.

Finally, I examined how the number of Xist repeats affects the protein-RNA binding affinity. For a more quantitative measurement, an increasing amount of recombinant protein was incubated with 1 rep wt or RNA fragments containing 2 or 8 repeats (2 rep wt and 8 rep wt), the binding strength was assessed by filter binding assays, and the fraction of bound RNA was calculated (Figure 2A). Although 1 rep wt is detectably bound by EZH2 and SUZ12, binding is very weak. In contrast, when 2 rep wt or 8 rep wt RNAs were used at equimolar concentration, the fraction of bound RNA significantly increased, which cannot be explained by the sole increase in the concentration of the double stem loop unit, as evidenced by the low binding of 1 rep wt at 10 fold concentration. Instead, this result suggests that there might be a cooperative RNA binding effect when more than one repeat is present, or that higher order RNA elements might be required to achieve specific binding to SUZ12 and EZH2. Disruption of the stem loop structures (mut RNAs) had mixed effects. While 2 rep mut was barely bound, 8 rep mut was bound with similar strength to the wt version. This might be due to unexpected secondary structures also bound by EZH2 and SUZ12, created by the introduction of mutations, and which may not form in the shorter 2 rep mut RNA.

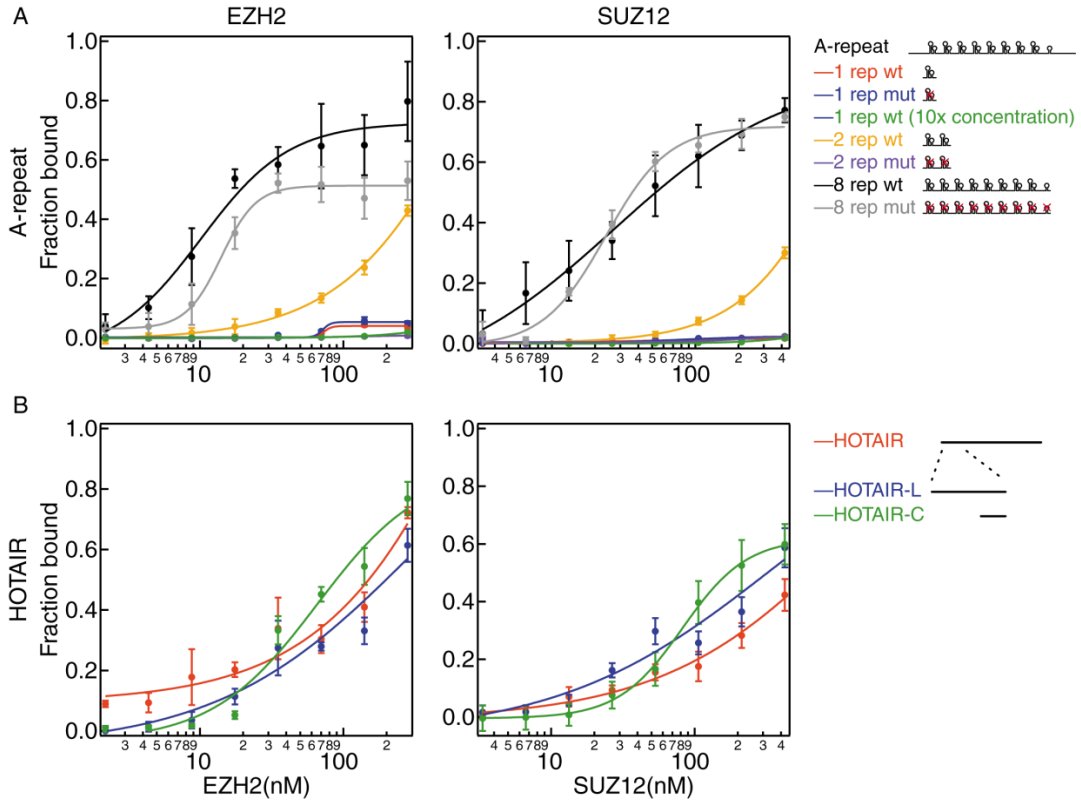

Figure 2. Filter binding assay with a series of (A) wild type and mutant Xist fragments with increasing number of repeats and (B) full length (HOTAIR) and fragments 1-308 (L) and 205-305 (C) of HOTAIR.

In contrast, HOTAIR does not possess any obvious secondary structure repeats. Binding to EZH2 and SUZ12 showed a different trend from Xist as the interaction is not enhanced by increases in the length of the RNA, which might be an indication that in this case there is a single PRC2 binding element contained in the shortest RNA used (HOTAIR-C), and that the PRC2 binding mechanism might be different from binding to Xist.

In summary, my results show that two PRC2 components, EZH2 and SUZ12, have RNA binding activity, and that both proteins have a slight preference for short structured RNAs. However, affinity for a single repeat of Xist is low and instead, binding becomes stronger as the number of repeats increases, although further studies are required to determine the actual contribution of the double stem loop structure of the repeats to PRC2 binding. The significance of the RNA binding by two different components of the complex awaits further investigation, but it is possible that EZH2 and SUZ12 might act together to stabilize the interaction with longer RNAs.