



**HAL**  
open science

# Contrôle de l'expression des gènes par les micro-ARN nucléaires : étude des mécanismes moléculaires

Raphaël Matégot

► **To cite this version:**

Raphaël Matégot. Contrôle de l'expression des gènes par les micro-ARN nucléaires : étude des mécanismes moléculaires. Biologie moléculaire. Université Côte d'Azur, 2018. Français. NNT : 2018AZUR4059 . tel-01981590

**HAL Id: tel-01981590**

**<https://theses.hal.science/tel-01981590>**

Submitted on 15 Jan 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



# THÈSE DE DOCTORAT

## Contrôle de l'expression des gènes par les micro-ARN nucléaires

Etude des mécanismes moléculaires

**Raphaël MATEGOT**

Centre Méditerranéen de Médecine Moléculaire Nice INSERM U1065

**Présentée en vue de l'obtention**

**du grade de Docteur en :**

Sciences de la Vie et de la Santé de  
l'Université Côte d'Azur

**Spécialité :**

Interaction Moléculaires et Cellulaires

**Dirigée par :** Dr. Michele TRABUCCHI

**Soutenue le :** 21 septembre 2018

**Devant le jury, composé de :**

**Dr. Valérie GRANDJEAN**

Présidente du jury CR1 Inserm  
C3M Nice France

**Dr. Eléonora LEUCCI**

Rapporteur Assistant Professeur  
Université de KU Leuven Belgique

**Dr. Stéphanie MILLEVOI**

Rapporteur CR1 Inserm  
CRCT Université de Toulouse France

**Dr. Michele TRABUCCHI**

DR2 Inserm, Directeur de Thèse  
C3M Nice France

# Contrôle de l'expression des gènes par les micro-ARN nucléaires

Etude des mécanismes moléculaires

**JURY :**

**Président du jury :**

Dr. Valérie GRANDJEAN, CR1 Inserm, C3M Nice France

**Rapporteurs :**

Dr. Eléonora LEUCCI, Assistant Professeur, Université de KU Leuven Belgique

Dr. Stéphanie MILLEVOI, CR1 Inserm, CRCT Université de Toulouse France

**Directeur de thèse :**

Dr. Michele TRABUCCHI, DR2 Inserm, C3M Nice France

## Contrôle de l'expression des gènes par les micro-ARN nucléaires: Etude des mécanismes moléculaires

La découverte de l'ARN interférence et des micro-ARN a permis de définir un principe majeur de régulation de l'expression des gènes, et a produit de nouveaux outils pour la médecine. Chez les mammifères, l'étude des fonctions des micro-ARN a été restreinte au cytoplasme, bien qu'ils soient aussi présents dans le noyau.

Cette thèse présente une série d'expériences visant à caractériser les facteurs moléculaires requis pour l'activité nucléaire des micro-ARN.

Nous avons débuté ce projet en explorant les partenaires ARN-dépendant de la protéine AGO2 par immunoprécipitation et spectrométrie de masse quantitative. Nous nous sommes concentrés sur trois de ces partenaires: SFPQ, PSPC1 et NONO, protéines nucléaires abondantes qui forment la famille de protéines DBHS (drosophila behavior and human splicing). Nous avons démontré que le complexe RISC nucléoplasmique est associé aux protéines SFPQ, PSPC1 et NONO dans plusieurs lignées cellulaires murines et humaines, d'une manière qui dépend de SFPQ.

Des expériences de type HITS-CLIP de la protéine AGO2 et de la protéine SFPQ dans des cellules souches de souris nous ont permis de montrer que SFPQ se lie préférentiellement aux 3'UTR longs en utilisant deux motifs spécifiques. Par l'intermédiaire de ces motifs, SFPQ contrôle significativement environ 20% de l'activité de ciblage de AGO2, ce qui est répercuté au niveau de la stabilité des ARN messagers cibles comme nous l'observons par analyse transcriptomique. Le mode d'action de la protéine SFPQ apparaît local et régule la liaison de AGO2 uniquement lorsque les sites de liaison de SFPQ sont proches (<500 nucléotides).

De plus, bien que SFPQ soit uniquement nucléaire, nous observons que SFPQ agit sur l'expression des ARNm cytoplasmiques. Cela suggère que la liaison et l'agrégation de la protéine SFPQ sur les ARNm programme le ciblage du 3'UTR par les miARN dans le noyau, d'une manière structurelle qui semble préservée dans le cytoplasme.

Enfin, nous avons montré que l'expression de SFPQ contrôle en particulier le programme de ciblage par let-7a, et module la transition des cellules souches vers l'état différencié.

Ces résultats contribuent à la diversité des mécanismes de régulation de l'activité des miARN.

Dans la deuxième partie du projet, nous avons exploré les partenaires ARN-indépendant de la protéine nucléaire AGO2. Nous avons découvert que la protéine AGO2 interagit avec le complexe CCR4-NOT1 et l'exosome nucléaire d'une manière indépendante de l'ARN.

Nous proposons une série d'expériences visant à confirmer ces résultats. Brièvement, l'hypothèse de travail qui semble la plus cohérente avec les données actuelles est la liaison directe de l'exosome au module CNOT2-CNOT3 du complexe CCR4-NOT1.

Ce modèle permettrait d'expliquer le mécanisme d'extinction des gènes par les miARN nucléaires qui reposerait donc sur leur interaction avec les complexes CCR4-NOT1 et exosome. Son mode opératoire comprendrait des protéines de liaison à l'ARN et des micro-ARN pour sélectionner les cibles.

Mots-clés : miARN, nucleus, SFPQ, post-transcription, RNA metabolism

Directeur de thèse: Dr. Michele Trabucchi, DR2 INSERM

## Control of gene expression by nuclear micro-RNA: Study of the molecular mechanisms

The discovery of RNA interference and micro-RNA has unravelled a fundamental principle of gene expression regulation, and has produced new tools for medicine. In mammals, the study of micro-RNA functions have been confined to the cytoplasm, although there is a growing body of evidence about their presence in the nucleus.

This thesis present a set of experiments directed towards understanding the molecular factors required for nuclear miRNA activity.

We started this project by exploring the RNA-dependent interactors of AGO2 by immunoprecipitation followed by quantitative mass spectrometry analysis. We focused on three partners: SFPQ, PSPC1 and NONO which are abundant nuclear proteins and belong to the DBHS family (drosophila behavior and human splicing). We demonstrated that the nucleoplasmic RISC complex associates with DBHS proteins in multiple murine and human cell lines in an SFPQ-dependent manner.

HITS-CLIP experiments of AGO2 and SFPQ proteins in mouse embryonic stem cells showed that SFPQ preferentially binds long 3'UTR using two specific motifs. Using these motifs, SFPQ significantly controls about 20% of local AGO2 binding activity and target mRNA stability as we observe by transcriptomic analysis. SFPQ mode of action appears to be local and SFPQ motifs are functional only when proximal to AGO2 binding sites in a window of 500 nucleotides.

Moreover, although SFPQ is only nuclear, we observe that SFPQ has an effect on cytoplasmic mRNAs, which suggests that SFPQ binding and aggregation to mRNA in the nucleus programs the 3'UTR for miRNA targeting, in a way that appears preserved in the cytoplasm.

In particular, we found that SFPQ controls let-7 targeting program and modulates embryonic stem cell differentiation into neuron cells.

These results contribute to expand the diversity of miRNA regulatory mechanisms.

In the second part of the project, we explored the RNA-independent interactors of AGO2. We discovered that nuclear AGO2 interacts with CCR4-NOT1 complex and the RNA exosome in an RNA-independent manner.

We propose a set of experiments to confirm these results, based on a new working model. Briefly, the most likely hypothesis to explain the current data is the direct binding of RNA exosome to the CNOT2-CNOT3 module of CCR4-NOT1 complex.

This model would explain the silencing mechanism of genes by nuclear miRNAs, whose activity would depend on the RISC complex interaction with CCR4-NOT1 complex and RNA exosome. Hence, the RNA exosome would use both RNA binding proteins and miRNAs to select targets for degradation, based on CCR4-NOT1 mode of action.

Keywords: miRNA, nucleus, SFPQ, post-transcriptional, RNA metabolism

Thesis Director: Dr. Michele Trabucchi, DR2 INSERM



Université Nice Côte d'Azur – Faculté des Sciences – Valrose  
Ecole Doctorale des Sciences de la Vie et de la Santé de Nice (ED85)  
Centre Méditerranéen de Médecine Moléculaire de Nice

## **Thèse de Doctorat**

Pour obtenir le titre de Docteur en Sciences de la Vie et de la Santé  
Spécialité : Interactions Moléculaires et Cellulaires

Présentée par  
**Monsieur Raphaël MATEGOT**

### **Contrôle de l'expression des gènes par les micro-ARN nucléaires: Etude des mécanismes moléculaires**

Thèse soutenue le  
**21 Septembre 2018**  
Au Centre Méditerranéen de Médecine Moléculaire de Nice

**Devant un jury de thèse composé de:**

**Dr. Valérie GRANDJEAN**  
**Dr. Stefania MILLEVOI**  
**Dr. Eleonora LEUCCI**  
**Dr. Michele TRABUCCHI**

Présidente du Jury  
Rapporteur  
Rapporteur  
Directeur de Thèse



Les travaux de doctorat présentés dans ce manuscrit ont été réalisés de Décembre 2014 à Décembre 2017 dans le Centre Méditerranéen de Médecine Moléculaire de Nice au sein de l'équipe  
« Control of Gene Expression »  
sous la direction du  
Dr. Michele TRABUCCHI

Je tiens à exprimer mon profond respect à tous les membres composant mon jury de thèse et à leur adresser mes remerciements de m'avoir honoré de faire partie de ce jury. Je souhaite remercier le Dr. Valérie GRANDJEAN d'avoir accepté de présider mon jury de thèse et pour ses précieux conseils.

Mes sincères remerciements s'adressent aussi au Dr. Stefania MILLEVOI et au Dr. Eleonora LEUCCI qui ont consacré de leur temps pour évaluer ce manuscrit.



## 1 Résumé

La découverte de l'ARN interférence et des micro-ARN a permis de définir un principe majeur de régulation de l'expression des gènes, et a produit de nouveaux outils pour la médecine. Chez les mammifères, l'étude des fonctions des micro-ARN a été restreinte au cytoplasme, bien qu'ils soient aussi présents dans le noyau.

Cette thèse présente une série d'expériences visant à caractériser les facteurs moléculaires requis pour l'activité nucléaire des micro-ARN.

Nous avons débuté ce projet en explorant les partenaires ARN-dépendant de la protéine AGO2 par immunoprécipitation et spectrométrie de masse quantitative. Nous nous sommes concentrés sur trois de ces partenaires: SFPQ, PSPC1 et NONO, protéines nucléaires abondantes qui forment la famille de protéines DBHS (drosophila behavior and human splicing). Nous avons démontré que le complexe RISC nucléoplasmique est associé aux protéines SFPQ, PSPC1 et NONO dans plusieurs lignées cellulaires murines et humaines, d'une manière qui dépend de SFPQ.

Des expériences de type HITS-CLIP de la protéine AGO2 et de la protéine SFPQ dans des cellules souches de souris nous ont permis de montrer que SFPQ se lie préférentiellement aux 3'UTR longs en utilisant deux motifs spécifiques. Par l'intermédiaire de ces motifs, SFPQ contrôle significativement environ 20% de l'activité de ciblage de AGO2, ce qui est répercuté au niveau de la stabilité des ARN messagers cibles comme nous l'observons par analyse transcriptomique. Le mode d'action de la protéine SFPQ apparait local et régule la liaison de AGO2 uniquement lorsque les sites de liaison de SFPQ sont proches (<500 nucléotides).

De plus, bien que SFPQ soit uniquement nucléaire, nous observons que SFPQ agit sur l'expression des ARNm cytoplasmiques. Cela suggère que la liaison et l'agrégation de la protéine SFPQ sur les ARNm programme le ciblage du 3'UTR par les miARN dans le noyau, d'une manière structurelle qui semble préservée dans le cytoplasme.

Enfin, nous avons montré que l'expression de SFPQ contrôle en particulier le programme de ciblage par let-7a, et module la transition des cellules souches vers l'état différencié.

Ces résultats contribuent à la diversité des mécanismes de régulation de l'activité des miARN.

Dans la deuxième partie du projet, nous avons exploré les partenaires ARN-indépendant de la protéine nucléaire AGO2. Nous avons découvert que la protéine AGO2 interagit avec le complexe CCR4-NOT1 et l'exosome nucléaire d'une manière indépendante de l'ARN.

Nous proposons une série d'expériences visant à confirmer ces résultats. Brièvement, l'hypothèse de travail qui semble la plus cohérente avec les données actuelles est la liaison directe de l'exosome au module CNOT2-CNOT3 du complexe CCR4-NOT1.

Ce modèle permettrait d'expliquer le mécanisme d'extinction des gènes par les miARN nucléaires qui reposerait donc sur leur interaction avec les complexes CCR4-NOT1 et exosome. Son mode opératoire comprendrait des protéines de liaison à l'ARN et des micro-ARN pour sélectionner les cibles.

Mots-clés : miARN, nucleus, SFPQ, post-transcription, RNA metabolism

Directeur de thèse: Dr. Michele Trabucchi, DR2 INSERM

The discovery of RNA interference and micro-RNA has unravelled a fundamental principle of gene expression regulation, and has produced new tools for medicine. In mammals, the study of micro-RNA functions have been confined to the cytoplasm, although there is a growing body of evidence about their presence in the nucleus. This thesis present a set of experiments directed towards understanding the molecular factors required for nuclear miRNA activity.

We started this project by exploring the RNA-dependent interactors of AGO2 by immunoprecipitation followed by quantitative mass spectrometry analysis. We focused on three partners: SFPQ, PSPC1 and NONO which are abundant nuclear proteins and belong to the DBHS family (drosophila behavior and human splicing). We demonstrated that the nucleoplasmic RISC complex associates with DBHS proteins in multiple murine and human cell lines in an SFPQ-dependent manner.

HITS-CLIP experiments of AGO2 and SFPQ proteins in mouse embryonic stem cells showed that SFPQ preferentially binds long 3'UTR using two specific motifs. Using these motifs, SFPQ significantly controls about 20% of local AGO2 binding activity and target mRNA stability as we observe by transcriptomic analysis. SFPQ mode of action appears to be local and SFPQ motifs are functional only when proximal to AGO2 binding sites in a window of 500 nucleotides.

Moreover, although SFPQ is only nuclear, we observe that SFPQ has an effect on cytoplasmic mRNAs, which suggests that SFPQ binding and aggregation to mRNA in the nucleus programs the 3'UTR for miRNA targeting, in a way that appears preserved in the cytoplasm.

In particular, we found that SFPQ controls let-7 targeting program and modulates embryonic stem cell differentiation into neuron cells.

These results contribute to expand the diversity of miRNA regulatory mechanisms.

In the second part of the project, we explored the RNA-independent interactors of AGO2. We discovered that nuclear AGO2 interacts with CCR4-NOT1 complex and the RNA exosome in an RNA-independent manner.

We propose a set of experiments to confirm these results, based on a new working model. Briefly, the most likely hypothesis to explain the current data is the direct binding of RNA exosome to the CNOT2-CNOT3 module of CCR4-NOT1 complex.

This model would explain the silencing mechanism of genes by nuclear miRNAs, whose activity would depend on the RISC complex interaction with CCR4-NOT1 complex and RNA exosome.

Hence, the RNA exosome would use both RNA binding proteins and miRNAs to select targets for degradation, based on CCR4-NOT1 mode of action.

Keywords: miRNA, nucleus, SFPQ, post-transcriptional, RNA metabolism

Thesis Director: Dr. Michele Trabucchi, DR2 INSERM

## 2 Table des matières

<b>1</b>	<b>Résumé</b> .....	<b>4</b>
<b>3</b>	<b>Remerciements</b> .....	<b>10</b>
<b>4</b>	<b>Organisation de la Thèse</b> .....	<b>11</b>
<b>5</b>	<b>Liste des abréviations</b> .....	<b>12</b>
<b>6</b>	<b>Liste des Figures &amp; Tableaux</b> .....	<b>17</b>
<b>7</b>	<b>Introduction et Sommaire</b> .....	<b>18</b>
<b>7.1</b>	<b>Introduction Générale</b> .....	<b>18</b>
<b>7.2</b>	<b>Les ARN non-codants</b> .....	<b>19</b>
7.2.1	L'hypothèse du monde ARN.....	19
7.2.2	Séquençage du génome et transcription généralisée du génome.....	19
7.2.3	Classification des ARN non-codants.....	20
<b>7.3</b>	<b>Les petits ARN non-codants</b> .....	<b>20</b>
7.3.1	Les micro-ARN.....	20
7.3.2	Les petits ARN interférents (siARN).....	21
7.3.2.1	ARN interférence chez les plantes.....	22
7.3.2.2	ARN interférence chez les champignons .....	22
7.3.2.3	ARN interférence chez le nématode .....	23
7.3.2.4	ARN interférence chez les vertébrés.....	24
7.3.3	Les piwi-ARN (piARN) .....	24
7.3.4	Fragments d'ARN de transfert (ARNt).....	25
7.3.5	Les fragments dérivés des ARN-Y .....	26
<b>7.4</b>	<b>Découverte des micro-ARN et phylogénie chez les métazoaires</b> .....	<b>27</b>
<b>7.5</b>	<b>Biogénèse des micro-ARN</b> .....	<b>29</b>
7.5.1	Spécificité de l'expression des micro-ARN.....	31
7.5.2	Régulation transcriptionnelle de l'expression des micro-ARN.....	32
7.5.3	Régulation post-transcriptionnelle de l'expression des micro-ARN .....	33
<b>7.6</b>	<b>Mécanismes d'action cytoplasmique des micro-ARN</b> .....	<b>34</b>
7.6.1	Inhibition de la traduction de l'ARNm .....	34
7.6.2	Déstabilisation et dégradation de l'ARNm .....	35
7.6.3	Biochimie du complexe effecteur miRISC.....	36
7.6.3.1	Structure et fonction des protéines Argonaute .....	36
7.6.3.2	Rôle du complexe CCR4-NOT1-CAF1 et des protéines GW182.....	38
7.6.3.2.1	Fonctions de déadénylation et de dégradation de l'ARNm.....	39

7.6.3.2.2	Fonction du complexe CCR4-NOT1 dans le décoiffage en 5' de l'ARNm.....	40
7.6.4	Séquestration des ARNm dans les « P-bodies » cytoplasmiques .....	41
7.6.5	Régulation de l'activité des micro-ARN par les protéines de liaison à l'ARN .....	42
<b>7.7</b>	<b>Ciblage des micro-ARN.....</b>	<b>44</b>
7.7.1	Mécanismes d'hybridation complémentaire des miARN à leur cibles.....	44
7.7.2	Prédiction des sites de liaison aux miARN .....	45
7.7.3	Cartographie expérimentale des sites de liaison des micro-ARN .....	48
7.7.3.1	Distribution de l'activité de liaison de la protéine Argonaute et des micro-ARN.....	49
7.7.3.2	Diversité des sites de liaisons : Sites non-canoniques et bulgés .....	49
7.7.4	Fréquence de l'hybridation centrale aux miARN .....	50
7.7.5	Influence de l'énergie libre de liaison du micro-ARN à la cible .....	51
7.7.6	Influence des modifications post-transcriptionnelles de la séquence des miARN....	51
<b>7.8</b>	<b>Micro-ARN et physiopathologie de l'organisme .....</b>	<b>52</b>
7.8.1	Rôle systémique des micro-ARN dans le développement des métazoaires .....	52
7.8.2	Rôle essentiel de certains micro-ARN dans le développement des mammifères .....	53
7.8.3	Dérégulations physiopathologiques de l'expression des micro-ARN .....	55
7.8.3.1	Dérégulations de l'expression des micro-ARN dans le cancer.....	55
7.8.4	Stratégies thérapeutiques visant à rétablir l'expression normale des micro-ARN ...	56
<b>7.9</b>	<b>Diversité de la localisation des micro-ARN .....</b>	<b>58</b>
7.9.1	Micro-ARN et réticulum endoplasmique .....	58
7.9.2	Micro-ARN dans les exosomes sécrétés .....	58
7.9.3	Micro-ARN dans les mitochondries.....	59
<b>7.10</b>	<b>Micro-ARN nucléaires.....</b>	<b>61</b>
7.10.1	Transport nucléaire des composants de l'ARN interférence et des micro-ARN.....	61
7.10.2	Fonctions nucléaires des micro-ARN chez les mammifères.....	62
7.10.2.1	Fonctions transcriptionnelles des siARN chez les mammifères .....	62
7.10.2.2	Fonctions transcriptionnelles des micro-ARN chez les mammifères .....	63
7.10.2.3	Fonctions dans la modulation de l'épissage alternatif .....	64
7.10.3	Activité de dégradation induite par les micro-ARN nucléaires.....	64
<b>7.11</b>	<b>Fonctions nucléaires d'autres petits ARN.....</b>	<b>65</b>
7.11.1	Élimination programmée de l'ADN chez les protozoaires ciliés.....	65
7.11.2	Méthylation de l'ADN dirigée par l'ARN chez les plantes.....	65
7.11.3	Formation de l'hétérochromatine chez la levure .....	69
7.11.4	Régulation des éléments transposables par les protéines Piwi.....	72
7.11.4.1	Découverte des rasi-ARN et protéines associées piwi dans les cellules germinales.....	72
7.11.4.2	Piwi-ARN et protéines associées dans les cellules germinales .....	73
7.11.4.3	Transmission maternelle du programme de ciblage piwi-ARN .....	74

7.11.4.4	Rôle fonctionnel des piARN et protéines PIwi .....	76
<b>8</b>	<b>Résultats .....</b>	<b>78</b>
<b>8.1</b>	<b>Etat de l'art au commencement de la thèse.....</b>	<b>78</b>
<b>8.2</b>	<b>Travail de thèse .....</b>	<b>78</b>
8.2.1	La protéine SFPQ module l'activité de régulation post-transcriptionnelle des gènes par les miARN dans le nucléoplasme .....	78
8.2.2	Mécanisme de l'extinction post-transcriptionnelle des gènes miARN-dépendante .....	81
8.2.3	Composants moléculaires nucléaires du complexe RISC.....	81
8.2.4	Des expériences complémentaires à réaliser .....	82
<b>8.3</b>	<b>Article 1. Post-transcriptional gene silencing mediated by microRNAs is controlled by nucleoplasmic SFPQ.....</b>	<b>84</b>
<b>8.4</b>	<b>Annexes .....</b>	<b>117</b>
<b>9</b>	<b>Discussion .....</b>	<b>122</b>
<b>9.1</b>	<b>Première partie du projet.....</b>	<b>122</b>
9.1.1	Confrontation des résultats avec les données de la littérature (1) .....	122
9.1.2	Fonctions connues des protéines DBHS et paraspeckles .....	122
9.1.3	Interaction entre les protéines DBHS et le complexe miRISC .....	123
9.1.4	Programmation du ciblage par les miARN dans le noyau .....	123
9.1.5	Fonction des protéines DBHS et paraspeckles dans la biogénèse des miARN .....	124
<b>9.2</b>	<b>Deuxième partie du projet.....</b>	<b>125</b>
9.2.1	Confrontation des résultats avec les données de la littérature (2) .....	125
9.2.2	Une interaction potentielle de l'exosome avec le complexe CCR4-NOT1 .....	125
9.2.2.1	Des indices pour le rôle du module CNOT2-CNOT3 .....	126
<b>10</b>	<b>Conclusions.....</b>	<b>133</b>
<b>11</b>	<b>Bibliographie .....</b>	<b>134</b>
<b>12</b>	<b>Note biographique.....</b>	<b>156</b>

Je dédie cette thèse à ma famille

### **3 Remerciements**

Je souhaite exprimer ma gratitude et mon respect à mon directeur de thèse, Dr. Michele Trabucchi, pour son temps, sa patience, ses conseils, son support, sa direction et son enthousiasme.

## 4 Organisation de la Thèse

La thèse est organisée en trois chapitres majeurs.

Le **Chapitre 1** (section 7) propose une introduction générale et décrit les mécanismes d'action canoniques des micro-ARN, d'après les données de la littérature.

Le **Chapitre 2** (section 8) présente les résultats des travaux conduits au cours de la présente thèse.

Le **Chapitre 3** (section 9) développe une discussion des résultats, en relief des travaux précédents.



## 5 Liste des abréviations

3'UTR : région 3' non-traduite

5'UTR : région 5' non-traduite

ADAR : adenosine deaminase acting on RNA

ADN : acide désoxyribonucléique

AGO1 : argonaute 1

AGO2 : argonaute 2

AGO3 : argonaute 3

AGO4 : argonaute 4

Air2p : arginine méthyltransférase-interacting RING finger protein 2

Amp1 : glutamate carboxypeptidase

Arb1 : argonaute binding protein 1

Arb2 : argonaute binding protein 2

ARC : argonaute siRNA chaperone complex

ARN : acide ribonucléique

ARN Pol II : ARN polymérase de type II

ARN Pol III : ARN polymérase de type III

ARN Pol IV : ARN polymérase de type IV

ARN Pol V : ARN polymérase de type V

ARNdb : ARN double-brin

ARNi : ARN interférence

ARNm : ARN messenger

ARNr : ARN ribosomal

ARNt : ARN de transfert

Arx : asterix

ATP : adénosine tri-phosphate

AtSN1 : arabidopsis thaliana SINE element family 1

DBHS : drosophila behavior and human splicing

MYC : myelocytomatosis virus homology proto-oncogene

Caf1 : CCR4-NOT1 associated factor 1

Cat-1 : cationic aminoacid transporter 1

Ccr4 : carbon catabolite repressive 4 (CNOT6)

CD44 : cell surface cluster of differentiation antigen 44

CDK1 : cyclin-dependent kinase 1 (cell division control protein 2, cdc2)

CDCA8 : cell division cycle associated a8

CDH1 : cadhérine-1

CDK8 : cyclin-dependent kinase 8

CenH : centromeric repeat homologous

Chp1 : chromodomain protein 1

Cid12 : caffeine-induced death protein 12

Cid14 : caffeine-induced death protein 14

Cim1 : CCR4-NOT1 interacting motif 1

CiRS-7 : circular RNA sponge for miR-7

Clr4 : cryptic loci regulator 4

CNOT1 : CCR4-NOT1 complex subunit 1

CNOT2 : CCR4-NOT1 complex subunit 2

CNOT3 : CCR4-NOT1 complex subunit 3

CNOT6 : CCR4-NOT1 complex subunit 6 (carbon catabolite repressive 4, Ccr4)

CNOT7 : CCR4-NOT1 complex subunit 7

CNOT9 : CCR4-NOT1 complex subunit 9

MT-CO1 : cytochrome C oxidase subunit 1, Cox1	EGFR : epithelial grow factor receptor
MT-CO2 : cytochrome C oxidase 2, Cox2	Egg : histone-lysine N-methyltransférase
CRP : c-reactive protein precursor	eggless
CSDC2 : cold shock domain-containing protein 2	EiF2C : eukaryotic initiation facteur 2 C
CUTs : cryptic unstable transcripts	EIF4E-T : eucaryotic initiation factor 4E transporter
Dcl-1 : dicer-like 1	EIF4G : eucaryotic initiation factor 4 G
Dcl-2 : dicer-like 2	ERBB2 : erythroblastic oncogene B
Dcp1 : mRNA decapping protein 1	Eri-1 : 3' -> 5' exoribonucléase 1
Dcp2 : mRNA decapping protein 2	ETS1 : external transcribed spacer 1
Dcr1 : dicer-1	EXOSC1 : exosome component 1
DDX17 : dead-box motif helicase 17	EXOSC10 : exosome component 10
DDX5 : dead-box motif helicase 5	EXOSC3 : exosome component 3
DDX6 : dead-box motif helicase 6	EXOSC4 : exosome component 4
DGCR8 : diGeorge critical region 8	EXOSC5 : exosome component 5
DIS3 : chromosome disjunction protein 3	EXOSC9 : exosome component 9
Dlc-3 : dicer-like 3	GERp95 : golgi and endoplasmic réticulum associated 95kDa protein
DND1 : dead-end protein homolog 1	GNPDA1 : glucosamine-6-phosphate isomerase 1
PUM1 : Pumilio 1	GW182 : gawky 182 kDa Gw-repeat rich protein
DNMT1 : ADN méthyltransférase de type 1	GWAS : genome wide association studies
DNMT3B : ADN méthyltransférase de type 3b	HBL-1 : basic/helix-loop transcription factor 1
DNMT3L : ADN méthyltransférase type 3-like	Hen1 : hua enhancer 1
Drm1 : domains rearranged methylase 1	HIF1 : hypoxia induced factor 1
Drm2 : domains rearranged methylase 2	HITS-CLIP : high-throughput sequencing cross-linking immunoprécipitation
E2F3 : adenovirus early promoter 2 binding protein factor 3	HNRNP-Q2 : heterogeneous nuclear ribonucleoprotein Q2
eARN : enhancer-ARN	HNRNPA1 : heterogeneous nuclear ribonucleoprotein A1
EDC3 : enhancer of decapping factor 3	

HOXB3 : homeobox protein 3	ME31B : maternal expression at 31B
HP1 : heterochromatin protein 1	RNA-binding protein
Hrr1 : helicase required for RNA interference 1	MET-1 : méthyltransférase 1
HSC70 : heat-shock cognate protein 70 kDa	MiARN : micro-ARN
HSP70 : heat-shock protein 70 kDa	MID : domaine middle
ELAVL1 : RNA binding protein Hu-antigen R	MILI : MIWI-like
IAP : intracisternal particle type A	MiRISC : complexe d'extinction des gènes induit par les miARN
ICOS : inducible T-cell costimulator precursor	MITF : microphthalmia-associated transcription factor 1
KH : K-homology (hnRNPK) domain	MIWI : mouse piwi-like
KLF4 : krüppel- like factor 4	MIWI2 : mouse piwi-like 2
KSRP : KH-homology splicing regulatory protein	Mmi1 : meiotic mRNA interception factor 1
KTF1 : kow-domain containing transcription factor 1	MOV10 : moloney leukemia virus 10
LATS2 : large tumor suppressor kinase 2	Mtr4 : mRNA transport regulator 4
Let-7 : letal 7	Mut-2 : mutator 2
LIN-14 : lineage 14	Mut-7 : mutator 7
Lin-4 : lineage 4	MYB : myeloblastosis transcription factor
LIN-28 : Lineage 28	Nab3 : nuclear polyadenylated RNA-binding protein 3
LINE : long interspersed nuclear element	NCL : nucleolin
LncRNA : long ARN non-codant	ND1 : NADH deshydrogenase subunit 1
LSD1 : lysine-specific histone demethylase 1	NEAT1 : nuclear enriched abundant transcript 1
LTR : long terminal repeat sequence	NELF : negative elongation factor
LYPLA1 : acyl-protein thioesterase 1	NEXT : nucleoplasmic exosome targeting complex
Ma : million d'années	NF45 : nuclear factor 45 kDa
MALAT1 : metastasis associated lung adenocarcinoma transcript 1	NF90 : nuclear factor 90 kDa
	NNS : nrd1-nab3-sen1 complex
	NONO : non-POU domain-containing octamer binding protein

NOP53 : nucleolar protein 53 kDa	PRE : Pumilio recognition element
NOS : nitric oxid synthase	PROMPTs : promoter upstream transcripts
Not1 : negative on TATA-less	PSPC1 : paraspeckle component 1
Nrd1 : negative regulator of differentiation 1	PTB : polypyrimidine tract binding protein
Nrpd1b : nuclear RNA polymérase D subunit 1a	PTGS : extinction post-transcriptionnelle des gènes
Nrpe : nuclear RNA polymérase D subunit 1b	Qde-1 : quelling defective 1
OCT4 : octamer binding transcription factor 4	Qde-2 : quelling defective 2Ras
paARN : promoter-associated ARN	RACE-PCR : rapid amplification of cDNA ends polymérase chain reaction
PABPC1 : poly(A) binding protein 1	Ras : rat sarcoma transformation factor
Pam : pabc recognition motif	rasiARN : repeat-associated small interfering RNAs
PAN2 : poly(A) nuclease 2	RBM7 : RNA-binding motif protein 7
PAN3 : poly(A) nuclease 3	RdDM : RNA-directed DNA methylation
Panx : panoramix	Rde-1 : RNAi defective 1
Papi : partner of piwi protein	Rde-2 : RNAi defective 2
PARN1 : poly(A) ribonuclease 1	Rde-3 : RNAi defective 3
Pat1 : DNA topoisomerase 2 associated protein 1	Rdr2 : RNA-dependent RNA polymérase 2
PAZ : domaine piwi argonaute zwillie	RDRC : RNA-directed RNA-polymérase complex
PCR : polymérase chain reaction	Rdrp1 : RNA-dependent RNA-polymérase 1
Pdd1 : programmed DNA degradation 1	REST : neural-restrictive silencing factor, RE1
PH : pleckstrin-homology domain	REr : réticulum endoplasmique rugueux
PiARN : piwi-ARN	RHOB : ras homology family member B
Pimet : piARN méthyltransférase	RISC : complexe d'extinction des gènes induit par l'ARN
PITX3 : pituitary homeobox 3	
Piwi : p-element induced wimpy testis	
PKR : protéine kinase R	
PNLDC1 : parn-like domain-containing protein 1	
Polr3d : polymérase type III subunit 3d	

RITS : RNA-induced transcriptional silencing complex  
 RNase : ribonucléase  
 TROVE2 : TROVE domain containing protein 2  
 Rrp6 : ribosomal RNA-processing protein 6  
 RT-qPCR : real-time quantitative polymérase chain reaction  
 RUNX1 : runt-related transcription factor 1  
 SCP1 : synaptotomal complex protein 1  
 SD : silencing domain  
 Sde-3 : silencing-defective 3  
 Sen1 : helicase tRNA-splicing endonuclease positive effector 1  
 SFPQ : splicing factor proline and glutamine-rich  
 SH3 : src-homology domain type 3  
 siARN : petit ARN interférent  
 SINE : short interspersed nuclear element  
 SKIV2L2 : superkiller viralicidic activity 2-like RNA helicase 2  
 SnoRNA : small nucleolar RNA  
 SNP : single nucleotide polymorphisms  
 SnRNA : small nuclear RNA  
 SOX2 : sex-determining region Y-box 2 transcription factor  
 Su(ste) : suppressor of stellate

Sup : transcriptional regulator  
 superman  
 Swi6 : mating-type switching gene 6  
 SLC7A1 : Solute carrier family, Cat-1  
 Tas3 : targeting complex subunit 3  
 TDRD7 : tudor domain containing protein 7  
 TGA : transcriptional gene activation  
 TGS : transcriptional gene silencing  
 TNRC6 : trinucleotide-repeat containing protein 6  
 Top2a : topoisomerase 2- $\alpha$   
 TRAMP : Trf4/Air2/Mtr4p polyadenylation complex  
 TRBP : TAR RNA binding protein  
 Trf4 : topoisomerase 1 related factor 4  
 Trf5 : topoisomerase 1 related factor 5  
 Twi1 : tetrahymena piwi homolog 1  
 Utp18 : U3 snoRNA-associated protein 18  
 VHC : virus de l'hépatite C  
 Wde : windei homolog of mbd1 chromatin-associated factor 1  
 XRN2 : 5' -> 3' exoribucléase 2  
 YY1 : ying yang protein 1  
 ZCCHC8 : zinc finger CCHC domain-containing protein 8  
 $\beta$ -Trcp1 : beta-transducin repeat-containing protein 1

## 6 Liste des Figures & Tableaux

<i>Figure 1. Evolution du nombre des miARN chez les espèces représentées.....</i>	<i>28</i>
<i>Figure 2. Etapes de la biogénèse des miARN .....</i>	<i>29</i>
<i>Figure 3. Structure du précurseur pri-miARN .....</i>	<i>31</i>
<i>Figure 4. Distribution des miARN dans différents tissus et une lignée cellulaire .....</i>	<i>31</i>
<i>Figure 5. Co-facteurs de la biogénèse des miARN au niveau post-transcriptionnel.....</i>	<i>33</i>
<i>Figure 6. Structure cristallographique du complexe AGO2/Micro-ARN/Cible.....</i>	<i>38</i>
<i>Figure 7. Régulation de l'activité des miARN par le micro-environnement du 3'UTR .....</i>	<i>43</i>
<i>Figure 8. Schéma représentant le mécanisme de ciblage des miARN.....</i>	<i>52</i>
<i>Figure 9. Représentation de la localisation cellulaire et extra-cellulaire des miARN .....</i>	<i>60</i>
<i>Figure 10. Schéma explicatif de la voie de méthylation de l'ADN dirigée par l'ARN.....</i>	<i>68</i>
<i>Figure 11. Schéma explicatif du complexe RITS chez la levure à fission .....</i>	<i>71</i>
<i>Figure 12. Schéma du mécanisme ping-pong dans la lignée germinale.....</i>	<i>75</i>
<i>Figure 13. Mécanisme d'inhibition transcriptionnelle des transposons par la protéine nucléaire Piwi.....</i>	<i>77</i>
<i>Figure 14. Une hypothèse pour expliquer le recrutement de l'exosome à AGO2. ....</i>	<i>129</i>
<i>Figure 15. Structures cristallographiques d'un exosome ou d'un complexe CNOT1-CNOT2-CNOT3. ....</i>	<i>131</i>
<i>Figure 16. Potentielle modulation dynamique de la transcription par les miARN nucléaires. ....</i>	<i>132</i>
<i>Tableau 1. Facteurs de transcription régulant la transcription de miARN.....</i>	<i>32</i>
<i>Tableau 2. Facteurs de régulation post-transcriptionnels des miARN .....</i>	<i>34</i>
<i>Tableau 3. Protéines de liaison à l'ARN et régulation du ciblage des miARN.....</i>	<i>44</i>
<i>Annexe 1. Localisation nucléaire d'AGO2 et micro-ARN.....</i>	<i>117</i>
<i>Annexe 2. Localisation nucléaire dynamique d'AGO2 et micro-ARN .....</i>	<i>118</i>
<i>Annexe 3. Interaction d'AGO2 avec l'exosome nucléaire.....</i>	<i>119</i>
<i>Annexe 4. Interaction d'AGO2 avec l'exosome nucléaire et CNOT1.....</i>	<i>120</i>
<i>Annexe 5. Interaction de l'exosome nucléaire avec AGO2.....</i>	<i>121</i>
<i>Boîte 1. L'exosome: Structure, co-facteurs et cibles.....</i>	<i>130</i>

## 7 Introduction et Sommaire

### 7.1 Introduction Générale

L'information génétique est stockée sous la forme de la double hélice de l'ADN dans chaque cellule<sup>1</sup>, et code les instructions pour la production de protéines, les molécules qui accomplissent les fonctions cellulaires. L'expression des gènes procède d'abord par la production d'une molécule d'ARN messager (ARNm) qui est copiée à partir de l'ADN, et sert de messenger pour transférer l'information génétique au niveau du ribosome pour la traduction en protéines. Il n'y a pas de production de protéines sans ARNm.

Chez les bactéries, l'ARNm est instable et dégradé rapidement après quelques cycles de traduction, par conséquent la plupart de la régulation de l'expression génique s'opère au niveau de la transcription<sup>2</sup>. Chez les eucaryotes, où le cycle cellulaire est plus lent, l'ARNm est plus stable et dès lors s'accumule. Cette accumulation de matériel permet de sculpter différentes quantités d'expression génique, par exemple en modifiant le taux de dégradation ou le stockage de la molécule d'ARNm.

Deux copies d'ADN et de chaque gène sont présentes dans chaque cellule du corps, cependant tous les gènes ne sont pas activés simultanément. Certains gènes remplissent des fonctions métaboliques ou de structure, qui sont essentielles et ubiquitaires ou exprimés dans tous les types cellulaires, alors que d'autres gènes remplissent des fonctions spécifiques et sont exprimés dans des types cellulaires spécialisés, parfois transitoirement en réponse à des stimuli environnementaux, ou pendant le développement. L'expression des gènes est modulée par une combinaison de mécanismes de régulation transcriptionnelle, post-transcriptionnelle et post-traductionnelle.

Les protéines de la famille Argonaute sont apparues chez les bactéries et participent à un système de défense contre les éléments mobiles et les acides nucléiques étrangers<sup>3,4</sup>. Chez les eucaryotes, des régions spécialisées du génome produisent des micro-ARN et, avec les protéines Argonaute, sont impliquées dans des mécanismes de régulation de l'information génétique endogène que nous abordons dans cette thèse.

## **7.2 Les ARN non-codants**

### **7.2.1 L'hypothèse du monde ARN**

Après le déchiffrement du code génétique, organisé en codons décryptés par les ARN de transfert en tant que molécules adaptatrices (chimère ARN-acide aminé), il a été proposé que les activités fondamentales des processus de la vie (support de l'hérédité, réplication et fonctions catalytiques métaboliques, puis traduction et synthèse des protéines) aient reposé exclusivement sur des molécules d'ARN dans les cellules ancestrales<sup>5</sup>. Par exemple, la première des enzymes était probablement une molécule d'ARN (ribozyme) avec des propriétés de répliqueuse, dont certains prototypes ont été (re)-produits en laboratoire par évolution in-vitro, ou optimisation aléatoire et sélection des séquences d'ARN vers l'activité recherchée<sup>6</sup>.

Dans le génome des eucaryotes, on pense que la naissance de nouveaux gènes codants peut s'effectuer à partir de la transition de certains ARN non-codants comme les enhancers ARN ou des longs ARN non-codants (lncRNAs) vers des ARN codants, ou ARNm<sup>7</sup>. Dans la section suivante nous abordons les propriétés de certains de ces ARN non-codants.

### **7.2.2 Séquençage du génome et transcription généralisée du génome**

Le séquençage du génome humain et la découverte récente de la transcription d'une grande partie de celui-ci a eu un impact majeur sur notre compréhension de la régulation des gènes<sup>8,9</sup>. Moins de trois pourcents du génome code des ARNm, mais plus des deux tiers du génome est transcrit et produit principalement des ARN non-codants. La plupart des ARN non-codants (ARNnc) sont exprimés dans des types cellulaires spécifiques ou à des moments précis du développement, mais ne code pas pour des protéines et agit pour activer ou réprimer l'expression des gènes.

Les ARNnc sont transcrits à partir de régions génomiques qui ont pendant longtemps été considérées comme de l'ADN "poubelle"<sup>10</sup>. Au cours des dix dernières années, de nombreuses études ont souligné le rôle émergent et fondamental de nouvelles catégories d'ARNnc impliqués dans la modulation de l'expression des gènes.



### 7.2.3 Classification des ARN non-codants

Il est possible de classer les ARNnc en deux catégories majeures: les ARNnc structurels et les ARNnc régulateurs.

Les ARNnc structurels sont ubiquitaires, ou exprimés de manière constitutive dans tous les types cellulaires. Ils comprennent entre autres les ARN de transfert (ARNt), les ARN ribosomiaux (ARNr), les petits ARN nucléaires (snRNA) et les petits ARN nucléolaires (snoRNAs).

De l'autre côté, l'expression des ARNnc régulateurs est modulable (comme celle des ARN messagers) et activée à des moments précis du développement où dans des types cellulaires spécialisés.

Les ARNnc régulateurs accomplissent des fonctions très diverses, et sont impliqués dans plusieurs mécanismes de régulation de l'expression des gènes, comme les modifications épigénétiques de la chromatine, l'activité transcriptionnelle, l'épissage alternatif, la dégradation de l'ARNm ou la régulation de la traduction des ARNm. Les ARNnc régulateurs sont arbitrairement groupés par leur taille: long ARN non-codants (>200 nt), ARNnc intermédiaire (30 à 200 nt) et petits ARNnc (20 à 30 nt)<sup>11-13</sup>.

## 7.3 Les petits ARN non-codants

### 7.3.1 Les micro-ARN

Parmi les petits ARNnc, les mieux caractérisés font partie de la classe des micro-ARN (22 nucléotides), qui est développée dans la présente thèse.

Brièvement, les micro-ARN (miARN) sont des opérateurs post-transcriptionnels qui régulent l'expression génique. Les miARN sont maturés à partir de long précurseurs d'ARN<sup>i</sup>, et s'associent à une protéine Argonaute<sup>ii</sup> pour cibler d'autres transcrits par hybridation complémentaire, en formant le complexe extincteur des gènes induit par les miARN (miRISC). Le complexe miRISC induit la diminution de l'expression des gènes par inhibition de la traduction et dégradation de l'ARNm.

Les miARN s'hybrident principalement aux régions 3'UTR des ARNm, par complémentarité parfaite de la séquence "seed" qui s'étend sur 6-8 nucléotides à

---

<sup>i</sup> Voir 7.5 Biogénèse des micro-ARN

<sup>ii</sup> Voir 7.6.3.1 Structure et fonction des protéines Argonaute

l'extrémité 5' du miARN (à partir du 2<sup>ème</sup> nucléotide en 5'). Ce type d'hybridation est le mécanisme canonique de ciblage par les miARN, mais il a été rapporté que la fréquence de l'hybridation non-canonique<sup>i</sup> des miARN peut dépasser 60% de l'activité de ciblage des miARN<sup>14-16</sup>.

Un miARN interagit avec plusieurs centaines de sites répartis dans le transcriptome codant d'une cellule, et un ARNm est potentiellement ciblé par plusieurs miARN différents, soulignant la complexité du réseau de régulation par les miARN.

Les estimations du nombre de miARN présents dans le génome sont variables et sont comprises entre 500 (estimation robuste) à 3000 chez l'humain<sup>17,18</sup>. Les miARN ainsi que d'autres petits ARNnc sont impliqués dans virtuellement tous les processus biologiques et peuvent être utilisés en tant que biomarqueurs<sup>19-27</sup>.

Les fonctions des miARN sont développées plus en détail dans les sections suivantes de la thèse.

### 7.3.2 Les petits ARN interférents (siARN)

Les petits ARN interférents (siARN) sont incorporés dans les protéines Argonautes, qui sont les protéines essentielles de l'ARN interférence.

Les petits ARN interférents peuvent être générés à partir de petits ARN introduits artificiellement dans la cellule<sup>28</sup> (méthode d'ARN interférence), mais peuvent également être produits de manière endogène, comme à partir d'ARN messagers transcrits de manière convergente (endo-siARN)<sup>29-32</sup>.

L'ARN interférence (ARNi) est à l'origine un système endogène d'inactivation intracellulaire des gènes lors de la pénétration d'acides nucléiques étrangers dans la cellule, et dont le mécanisme repose systématiquement sur la production d'un intermédiaire d'ARN double-brin (ARNdb) correspondant au gène ciblé.

Les gènes *Argonaute* ont été initialement identifiés à partir de cribles par mutagenèse visant à identifier des gènes régulateurs du développement chez les plantes, et ont été également impliqués dans le mécanisme d'inactivation des transgènes dans certains organismes génétiquement modifiés. Plus tard, les gènes *Argonaute* ont été reliés au

---

<sup>i</sup> Voir 7.7.3.2 Diversité des sites de liaisons : Sites non-canoniques et bulgés

mécanisme de l'ARN interférence à partir d'études menées notamment chez les plantes, les champignons et le nématode.

### 7.3.2.1 ARN interférence chez les plantes

Chez *Arabidopsis thaliana*, les mutations du gène *Argonaute-1* (AGO1) induisent un développement anormal des feuilles, une inflorescence et une atrophie des cotyledons qui se replient d'une manière évoquant les tentacules de céphalopodes comme ceux de la pieuvre *Argonauta Argo*, d'où le nom du gène<sup>33</sup>. Les mutations du gène AGO1 suppriment le mécanisme d'inactivation post-transcriptionnelle des transgènes, qui est parfois observé lorsque l'on tente de générer une plante transgénique<sup>34</sup>.

L'inactivation post-transcriptionnelle des transgènes est aussi appelé co-suppression, dont un exemple caractéristique a été observé chez la plante *Petunia hybrida*. En effet, dans le but de produire des pétales plus foncés, l'introduction d'une copie supplémentaire du gène codant la chalcone synthase<sup>i</sup> dans le génome de *Petunia hybrida* peut induire l'inactivation non seulement du transgène mais aussi des deux copies endogènes, ce qui génère alors des pétales partiellement colorés voir des pétales totalement blancs<sup>35,36</sup>. Lors du phénomène d'inactivation post-transcriptionnelle des gènes (PTGS) on observe la production de deux populations de petits ARN non-codants d'environ 25 nucléotides correspondant à la fois à la séquence des gènes inactivés, et à la séquence complémentaire<sup>37</sup>.

### 7.3.2.2 ARN interférence chez les champignons

Chez le champignon filamenteux *Neurospora Crassa*, on retrouve le même phénomène d'inactivation post-transcriptionnelle des transgènes qui est connu sous le nom de *quelling*, et est inactivé avec des mutations dans le gène AGO1 (*qde-2*), mais aussi une ARN-polymérase ARN-dépendante (*qde-1*) ou une protéine de type hélicase (*qde-3*)<sup>38</sup>.

---

<sup>i</sup> Un gène codant une enzyme impliquée dans la synthèse du pigment violet.

### 7.3.2.3 ARN interférence chez le nématode

Chez le nématode *Caenorhabditis elegans*, la voie ARN interférence peut être activée en administrant une molécule d'ARNdb<sup>i</sup> par ingestion de bactéries exprimant l'ARNdb, ou par micro-injection cellulaire de la molécule<sup>39,40</sup>.

Le phénomène d'ARN interférence est complètement inactivé suite aux mutations dans le gène *rde-1* (RNAi defective 1) dont la fonction ne semble pas influencer la captation intracellulaire de l'ARNdb, ni la stabilité intracellulaire ou la diffusion tissulaire de la molécule<sup>41</sup>. Le gène *rde-1* fait partie de la famille piwi/sting/argonaute/zwiller/eiF2C qui compte 22 membres chez les nématodes<sup>28</sup>. D'autres mutations dans les gènes *rde-2*, *rde-3*, *mut-2* ou *mut-7* induisent aussi une mobilisation des transposons, une augmentation de la perte de chromosomes et une fertilité réduite en plus de l'inactivation de la réponse ARNi.

Le mécanisme de l'ARN interférence repose essentiellement sur la molécule d'ARNdb, qui peut être produite en combinant des préparations individuelles du brin sens et du brin antisens. L'utilisation d'ARN simple-brin déclenche elle très peu ou pas d'ARN interférence<sup>42</sup>.

Chez *Caenorhabditis elegans*, l'inactivation des gènes qui se produit par l'ARN interférence diffuse de tissu en tissu, et est particulièrement efficace en raison de mécanismes d'amplification et de maintenance de la molécule par des ARN-polymérase ARN-dépendante<sup>43,44</sup>. De plus, l'inactivation est transmissible par la lignée germinale sur plusieurs générations mais uniquement en injectant les animaux hermaphrodites<sup>45</sup>.

L'ARN interférence peut être dirigée contre l'un ou l'autre des transcrits monocistroniques d'un opéron polaire<sup>46</sup> sans interférer avec la quantité de l'autre transcrit dans la cellule<sup>47</sup>. Ces premières études ont suggéré que l'ARN interférence déclenche principalement un mécanisme de destruction post-transcriptionnelle de l'ARNm ciblé, dont on observe la disparition à la fois dans le cytoplasme et le noyau<sup>47</sup>.

L'addition séparée du brin antisens dans un extrait cellulaire déclenche l'ARN interférence à peu près cent fois moins efficacement que l'ARN double-brin, qui est converti en une population de petits ARN interférents de 21-23 nucléotides et induisent le clivage de l'ARN ciblé à des intervalles réguliers de 21-23 nucléotides<sup>48</sup>. Le site de clivage de l'ARN cible correspond systématiquement aux position 10-11 des petits

---

<sup>i</sup> dont la taille est typiquement entre 500-1000 paires de bases.

ARN<sup>49</sup>, et la rupture de la liaison phosphodiester de l'ARN produit des extrémités 3'-hydroxyl et 5'-phosphate<sup>50</sup>.

#### **7.3.2.4 ARN interférence chez les vertébrés**

Chez les vertébrés, l'introduction d'ARNdb<sup>i</sup> dans la cellule déclenche une réponse antivirale de type interféron, basée sur la Protéine kinase R (PKR) et la stimulation de la RNase L par la 2'-5' oligoadénylate synthase.

Cette réponse antivirale constitue un mécanisme de dégradation généralisée (non-spécifique) des ARN dans la cellule y compris des ARN viraux double-brin, ce qui protège de la réplication et la diffusion du virus aux cellules voisines.

Ce mécanisme de défense entraîne une difficulté pour l'utilisation spécifique de la méthode d'ARN interférence chez les vertébrés.

L'activation spécifique de l'ARN interférence est cependant possible par l'utilisation d'ARNdb de taille optimale, dont la taille n'excède pas 21-23 nucléotides<sup>51,49</sup>.

#### **7.3.3 Les piwi-ARN (piARN)**

Les piwi-ARN (piARN) sont des petits ARN non-codants d'une taille de 23-30 nucléotides. Ils font partie d'un système spécialisé dans la régulation des transposons et éléments répétés, en particulier dans la lignée germinale. Les piARN s'associent aux protéines de la famille Piwi pour accomplir leurs fonctions d'extinction des transposons. Dans les tissus reproducteurs coexistent deux voies majeures de biogénèse des piARN qui approvisionnent le système de protéines piwi. La première voie génère des piARN à partir de matériel transposable accumulé dans le génome (les clusters de piARN), et est requise pour la maintenance de l'inactivation des éléments transposables résidents. Cette voie est aussi impliquée dans l'inactivation de certains transcrits pendant le stade pachytène de la méiose.

La seconde voie de biogénèse est activée spécifiquement dans la lignée germinale<sup>ii</sup>, et constitue une réponse adaptative contre les éléments transposables actifs<sup>iii</sup>, basée sur un mécanisme d'amplification.

---

<sup>i</sup> D'une taille supérieure à 30-80 paires de bases.

<sup>ii</sup> La voie d'amplification ping-pong, abordée dans la section 6.10.4

<sup>iii</sup> Qui représentent un potentiel colonisateur du génome.

Ce mécanisme (la voie d'amplification ping-pong) repose sur l'incorporation des fragments de transposons clivés dans le système Argonaute-3/Aubergine, en l'absence d'enzymes de type ARN-polymérase ARN-dépendante dans le génome des vertébrés<sup>52</sup>.

Le mode d'action des protéines de la famille Piwi est double et repose sur un mécanisme de clivage post-transcriptionnel des séquences de transposons mais également une inhibition transcriptionnelle des transposons au niveau nucléaire par le ciblage du transcrit naissant<sup>52</sup>.

#### 7.3.4 Fragments d'ARN de transfert (ARNt)

La deuxième classe de petits ARNnc les plus abondants dans la cellule après les micro-ARN semble dériver de fragments d'ARN de transfert (ARNt) de taille 28-34 nucléotides correspondant aux extrémités 5' (générés par la RNase P) ou 3' des ARNt matures, ou bien à l'extrémité 3' spécifique du précurseur ARNt (générés par la RNase Z)<sup>53,54</sup>. Alternativement, des fragments de taille 31-40 nucléotides sont générés par clivage au niveau de l'anticodon par l'enzyme RNY1 angiogénin chez la levure, et la biogénèse de certains fragments d'ARNt serait dépendante de Dicer<sup>55</sup>. De plus, la dégradation de certains ARNt est dépendante de la méthylation de cytosines par les enzymes de la famille DNMT2<sup>56</sup>, comme en aval de l'anticodon de l'ARNt<sup>Asp-GTC</sup> au niveau de la cytosine 38<sup>57</sup>. Par ailleurs, la transmission épigénétique de phénotypes par la lignée germinale serait dépendante de l'enzyme DNMT2<sup>58,59</sup>.

Les fragments d'ARNt sont particulièrement abondants dans le sperme, et semblent d'abord clivés dans l'épididyme puis délivrés aux spermatozoïdes par transfert de vésicules exosomales spécialisées (epididymosomes) dont le contenu est enrichi en fragments d'ARNt<sup>54</sup>.

La fonction des fragments d'ARNt est peu caractérisée, mais semble reposer sur la dégradation d'ARN contenant des éléments rétroviraux endogènes, et impliquée<sup>i</sup> dans la transmission héréditaire de phénotypes métaboliques par la lignée paternelle<sup>54</sup>.

Cette fonction de dégradation reposerait sur l'association de certains fragments d'ARNt avec le système de protéines Argonaute, avec une préférence pour AGO3 et AGO4<sup>62</sup>, mais aussi avec le système de protéines Piwi<sup>55</sup>.

---

<sup>i</sup> Avec les miARN, voir Gapp et al., 2016 & Grandjean et al., 2015, réfs <sup>60,61</sup>.

### 7.3.5 Les fragments dérivés des ARN-Y

Les ARN-Y sont des ARN non-codants de taille 80-120 nucléotides associés au déterminant Ro60 (TROVE2)<sup>63</sup>, présent de 1 à 4 copies fonctionnelles dans le génome des vertébrés<sup>64</sup>, mais aussi jusqu'aux bilatériens et certains procaryotes avec des ARN de structure similaire<sup>65</sup>. La protéine TROVE2 est impliquée dans un mécanisme de contrôle qualité de plusieurs classes d'ARN non-codants<sup>66</sup>, et la liaison de haute affinité des ARN-Y à TROVE2<sup>i</sup> semble bloquer ou moduler la fonction contrôle qualité de TROVE2<sup>68,69</sup>.

Les ARN-Y sont également essentiels pour la formation des fourches de réplication, lors de l'étape d'initiation de la réplication de l'ADN<sup>65,70,71</sup>.

Des fragments d'ARN-Y d'une taille de 22-36 nucléotides sont détectables dans plusieurs types cellulaires et cellules cancéreuses<sup>72,73</sup>. Cependant, la biogénèse des ARN-Y est inconnue, mais semble indépendante de Dicer et les fragments ne semblent pas s'associer aux protéines Argonaute<sup>64</sup>.

De plus, les fragments d'ARN-Y s'accumulent dans les cellules apoptotiques<sup>74</sup>, y compris dans les macrophages spumeux<sup>27</sup>, et peuvent être utilisés comme biomarqueurs prédictifs dans certaines pathologies comme l'athérosclérose<sup>25</sup>. Dans les macrophages, les ARN-Y stimulent les récepteurs Toll-like de type 7, et une apoptose caspase-3-dépendante<sup>27</sup>.

---

<sup>i</sup> La protéine TROVE2 est également associée aux rétrotransposons Alu et la modulation de la réponse inflammatoire à l'interféron, voir Hung et al., 2015, réf <sup>67</sup>.

#### 7.4 Découverte des micro-ARN et phylogénie chez les métazoaires

Les micro-ARN ont été découverts chez le nématode *Caenorhabditis elegans* par des études de mutations (*lin-4*) induisant la répétition de divisions cellulaires spécifiques de l'état larvaire, aussi appelées mutations de gènes hétérochroniques<sup>75,76</sup>. Les micro-ARN sont localisés dans des régions génomiques transcrites en ARN formant une structure en tige-boucle<sup>77</sup>, et après maturation régulent l'expression de gènes cibles par appariement complémentaire antisens<sup>75</sup>.

Chez l'humain, la moitié (54%) des miARN est localisée dans des transcrits intergéniques non-codants (lncRNAs)<sup>i</sup>, tandis que l'autre moitié (46%) est localisée dans les régions introniques de gènes codants<sup>79</sup>. Certains miARN comme ceux de la famille<sup>ii</sup> *let-7*<sup>80,81</sup>, ou ceux du locus *miR-17~92* sont situés dans des précurseurs polycistroniques, et parfois plusieurs régions dans le génome comme c'est le cas des membres de la famille *let-7*<sup>82</sup>.

Le génome humain code approximativement 519 miARN regroupés en 283 familles<sup>83</sup> (**Figure 1**), ce qui représente environ un tiers des estimations contenues dans la base de données miRBase<sup>84</sup>. Le répertoire de séquences de miARN chez l'humain est le produit de 585 acquisitions de miARN (318 familles) et 64 pertes de miARN (34 familles), depuis la divergence avec les éponges il y a ~800 million d'années (Ma). Des vagues majeures d'acquisition de miARN ont précédé l'émergence des bilatériens il y a 675 Ma (+39 miARN, 32 familles), des vertébrés il y a 530 Ma (+176 miARN, 45 familles), des mammifères placentaires ou euarchontoglires il y a 75 Ma (+144 miARN, 91 familles), et une part importante des miARN est apparue après la divergence avec les rongeurs (+179 miARN, 111 familles).

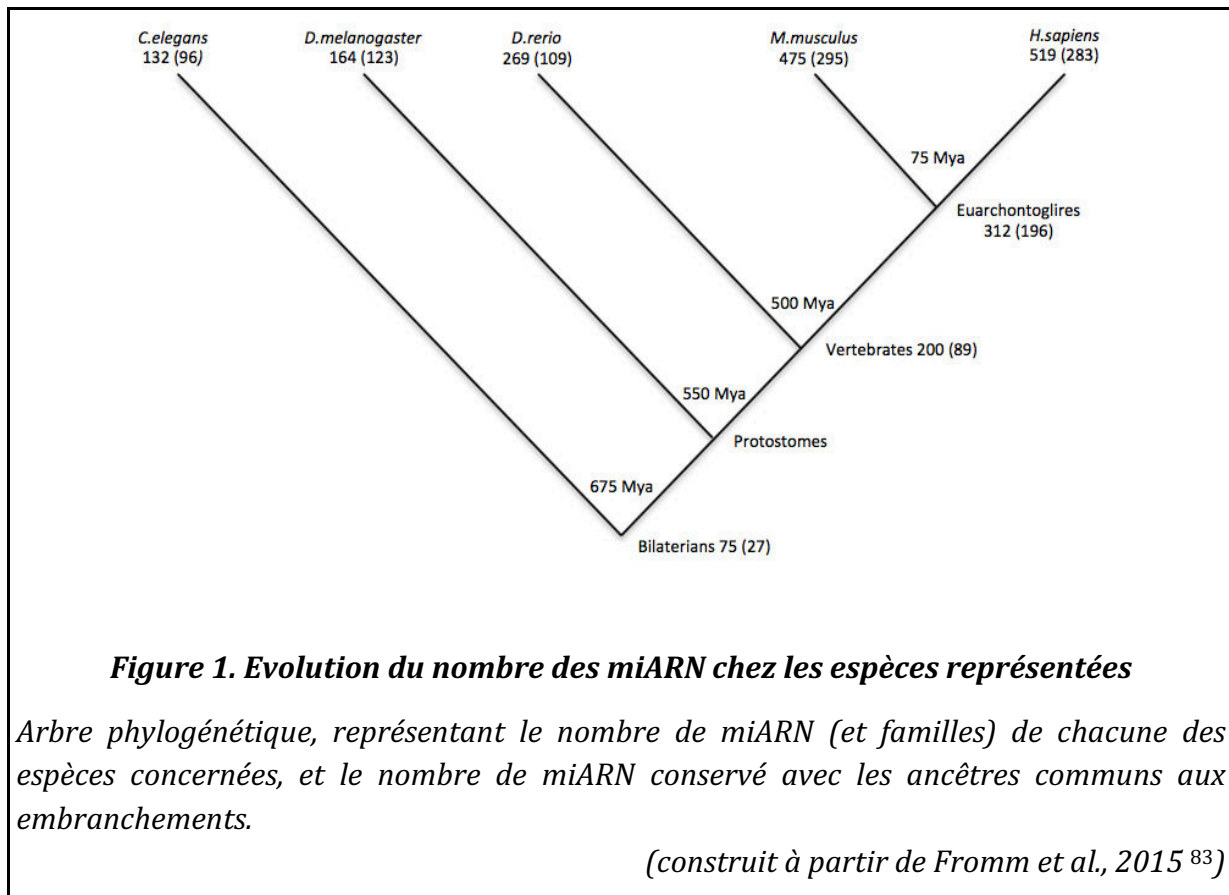
Dans l'ensemble, nous partageons 75 miARN (27 familles) avec les bilatériens, 200 miARN (89 familles) avec les vertébrés et 312 miARN (196 familles) avec les mammifères placentaires<sup>83,85</sup>.

---

<sup>i</sup> Quelques miARN comme le *miR-198* sont localisés dans les régions exoniques de gènes codants, voir Cullen 2004, réf <sup>78</sup>.

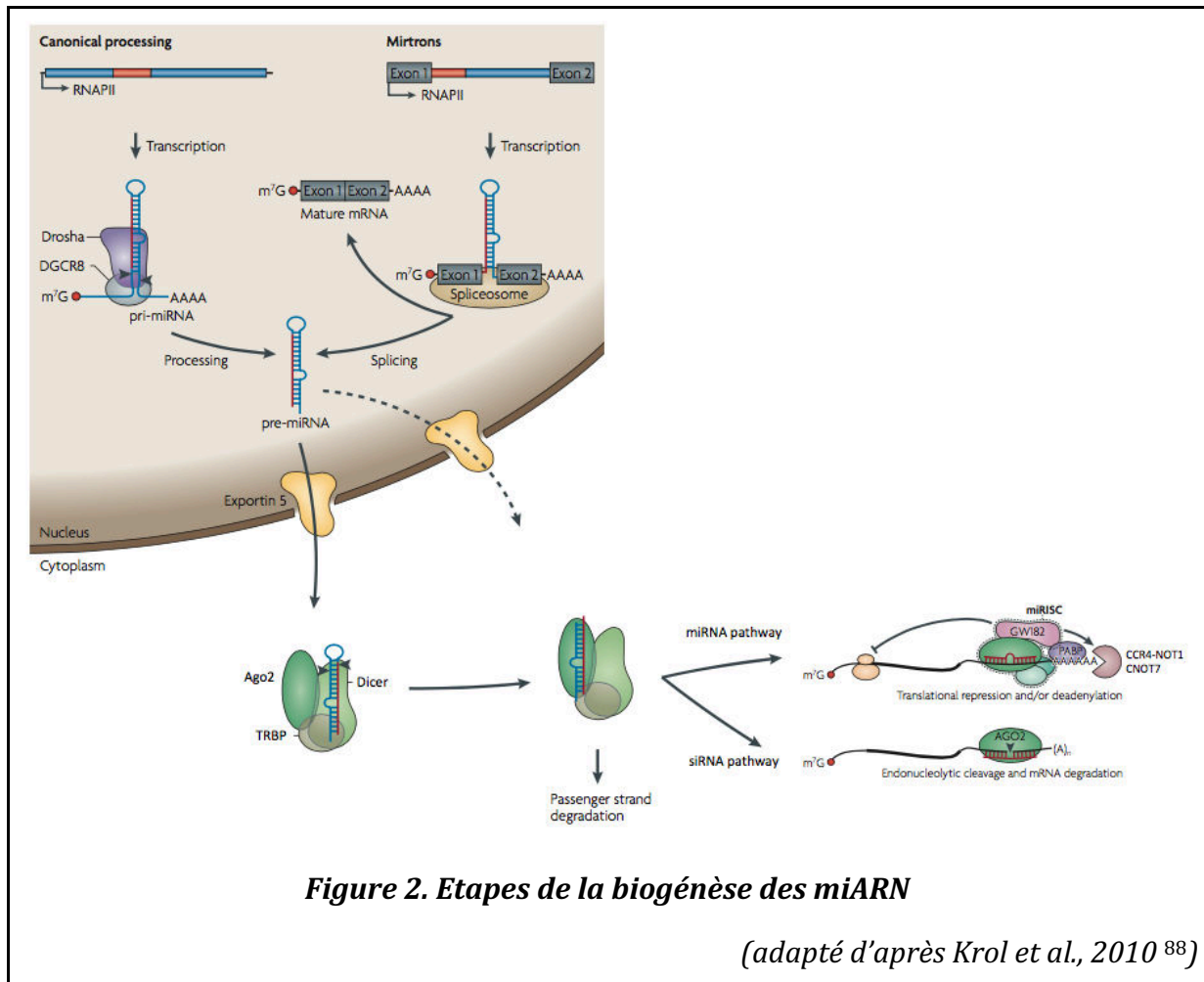
<sup>ii</sup> Qui partagent la même région seed, voir sections 7.7.2 et 7.8.2.





## 7.5 Biogénèse des micro-ARN

Les séquences de miARN sont localisées dans des long précurseurs d'ARN simple-brin contenant des régions en tige-boucle, puis maturés à une taille mature de ~22 nucléotides<sup>37,48,86,87</sup>.



Dans le noyau, le précurseur primaire des miARN<sup>89</sup> (pri-miARN, **Figure 2**) est d'abord transcrit<sup>i</sup> par l'ARN Pol II<sup>91</sup> et maturé en pré-miARN par le complexe microprocesseur contenant les protéines DGCR8 et DROSHA<sup>92,93</sup>. Le microprocesseur se fixe aux segments double-brin (35±1 bp) de(s) région(s) tige-boucle du pri-miARN (**Figure 3**) à l'aide d'un ou plusieurs motifs localisés dans les séquences simple-brin flanquantes, la boucle terminale ou bien dans le segment double-brin<sup>94</sup>, et positionne les domaines catalytiques de la nucléase DROSHA à une distance d'environ 13 nucléotides de la base du duplex, pour un clivage endonucléolytique de part et d'autre de la double hélice

<sup>i</sup> Le plus souvent par l'ARN Pol II, quelques miARN sont proches d'éléments Alu et transcrits par l'ARN Pol III, voir Borchert et al., 2006, réf<sup>90</sup>.

d'ARN avec deux bases de décalage<sup>95</sup>. Cette étape de maturation des pri-miARN prendrait place dans un corps nucléaire spécialisé appelé paraspeckle<sup>i</sup> 96,97. Le clivage par DROSHA définit l'extrémité 5' du miARN, et libère un fragment de ~60 nucléotides qui correspond au pré-miARN, qui est ensuite exporté du noyau vers le cytoplasme par une karyophérine, l'exportin-5<sup>98</sup>.

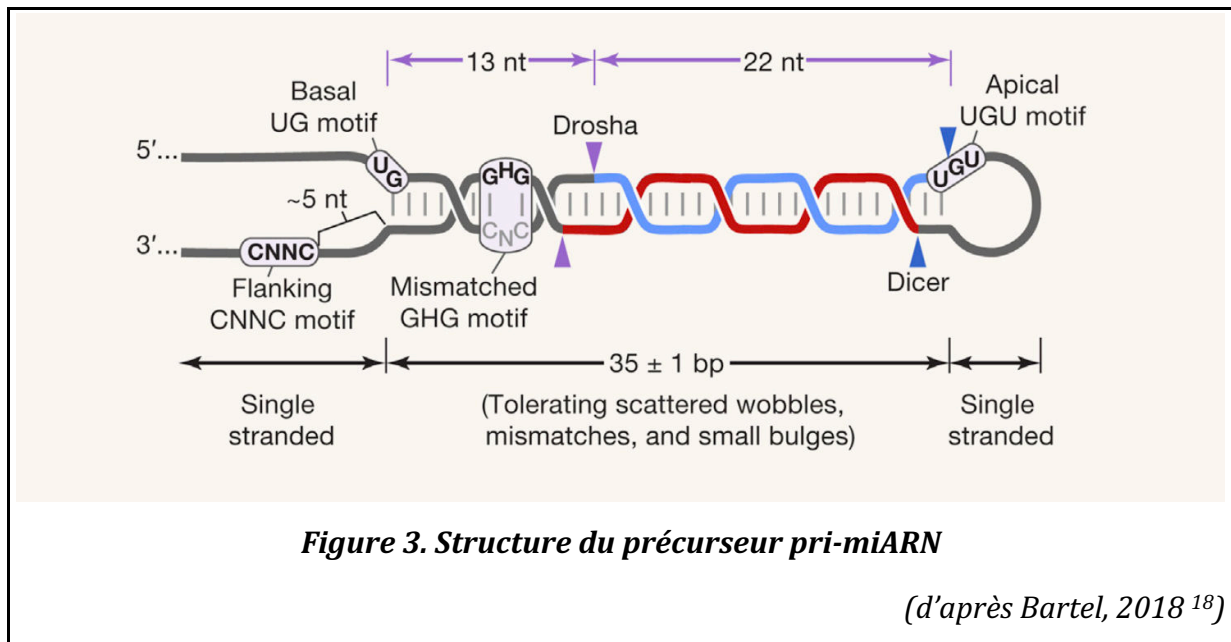
De plus, la maturation de certains miARN appelés mirtrons est indépendante du microprocesseur, et repose uniquement sur l'intégrité du spliceosome<sup>99-101</sup>. Ces miARN sautent l'étape de clivage par le complexe microprocesseur, et la plupart sont maturés directement par l'enzyme DICER à partir d'introns épissés<sup>ii</sup>.

Dans le cytoplasme, la boucle terminale est clivée par une deuxième nucléase appelée Dicer, dont le clivage décalé de part et d'autre de la double-hélice produit un ARN duplex de 19-22 nucléotides avec deux nucléotides protubérants à chaque extrémité 3' <sup>102</sup>. Chez les métazoaires, la taille médiane du pré-miARN est de 60 nucléotides composé d'une boucle terminale de 16 nucléotides ( $8 \leq N \leq 99$ ), et chaque brin mesure 22 nucléotides ( $20 \leq N \leq 26$ ) avec 20 nucléotides d'appariement parfait ( $16 \leq N \leq 25$ )<sup>83</sup>. Le miARN mature, aussi appelé séquence guide, est presque invariablement le brin dont l'extrémité 5' est la moins stable thermodynamiquement. Le brin guide est sélectivement incorporé dans une nucléase de la famille Piwi/Argonaute, dont la spécificité de séquence dépend du micro-ARN associé<sup>103</sup>. Chez les mammifères, cette protéine de la famille Piwi/Argonaute est principalement Argonaute-2 (AGO2) et constitue avec le miARN le complexe d'extinction des gènes induit par l'ARN (RISC ou miRISC)<sup>103</sup>. Le miARN est transféré de la protéine Dicer vers AGO2 avec l'aide de la protéine TRBP<sup>104</sup> (R2d2 chez la drosophile<sup>105</sup>) pendant que la protéine AGO2 est maintenue dans une conformation ouverte par l'action des protéines chaperonnes HSC70/HSP90, avec consommation d'énergie ATP<sup>106</sup>.

---

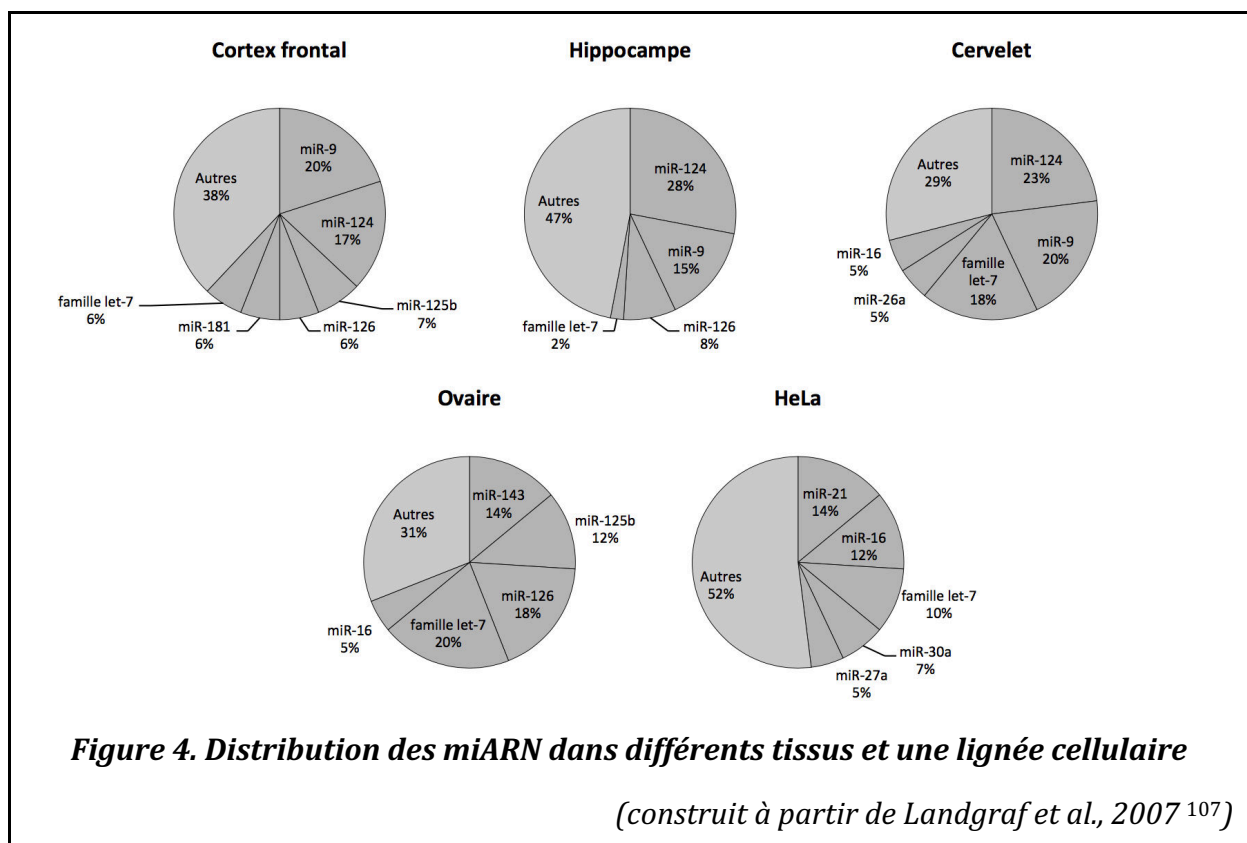
<sup>i</sup> Adjacents aux corps d'épissages appelés speckles.

<sup>ii</sup> Exceptionnellement, certains miARN introniques comme miR-106b sont maturés par le complexe microprocesseur.



### 7.5.1 Spécificité de l'expression des micro-ARN

L'expression des miARN est très spécifique (**Figure 4**), en général seulement 5 miARN différents représentent 50% de la quantité totale des miARN exprimés dans un type cellulaire<sup>107</sup>. Le nombre total de molécules de miARN présents dans une cellule est typiquement compris entre 50 000 à 100 000 molécules<sup>108</sup>.



### 7.5.2 Régulation transcriptionnelle de l'expression des micro-ARN

Les mécanismes qui contribuent à l'expression sélective des miARN reposent sur la régulation transcriptionnelle des loci de miARN par la liaison de facteurs de transcription spécifiques au promoteur<sup>109</sup> (**Tableau 1**), comme la stimulation du cluster oncogénique 17-92 par le facteur de transcription MYC<sup>110</sup>.

Les promoteurs de pri-miARN peuvent également être régulés à travers des mécanismes épigénétiques de méthylation par les enzymes DNMT1/DNMT3B comme pour les miR-124a, miR-34b/c, let-7, miR-9<sup>111</sup>.

La transcription des miARN peut être augmentée par l'association avec les éléments de types enhancers et une forte concentration en facteurs de transcriptions, comme par exemple le cluster 290-295 dans les cellules souches embryonnaires<sup>109,112</sup>.

miARN / promoteur	Facteur de transcription	Rôle	Référence
miR-17-92	MYC	Active	110
miR-9	MYC	Active	113
miR-34/miR-107	P53	Active	114
miR-124	REST (RE1)	Inhibe	115
miR-133b	PITX3	Active	116
miR-210	Hif-1	Active	117
miR-145	P53	Active	118
miR-27a	RUNX1	Active	119
miR-15a	MYB	Inhibe	120
let-7	HBL-1	Inhibe	121
miR-7	YAN	Inhibe	122
Cluster 290-295	OCT4, SOX2, NANOG	Active	109

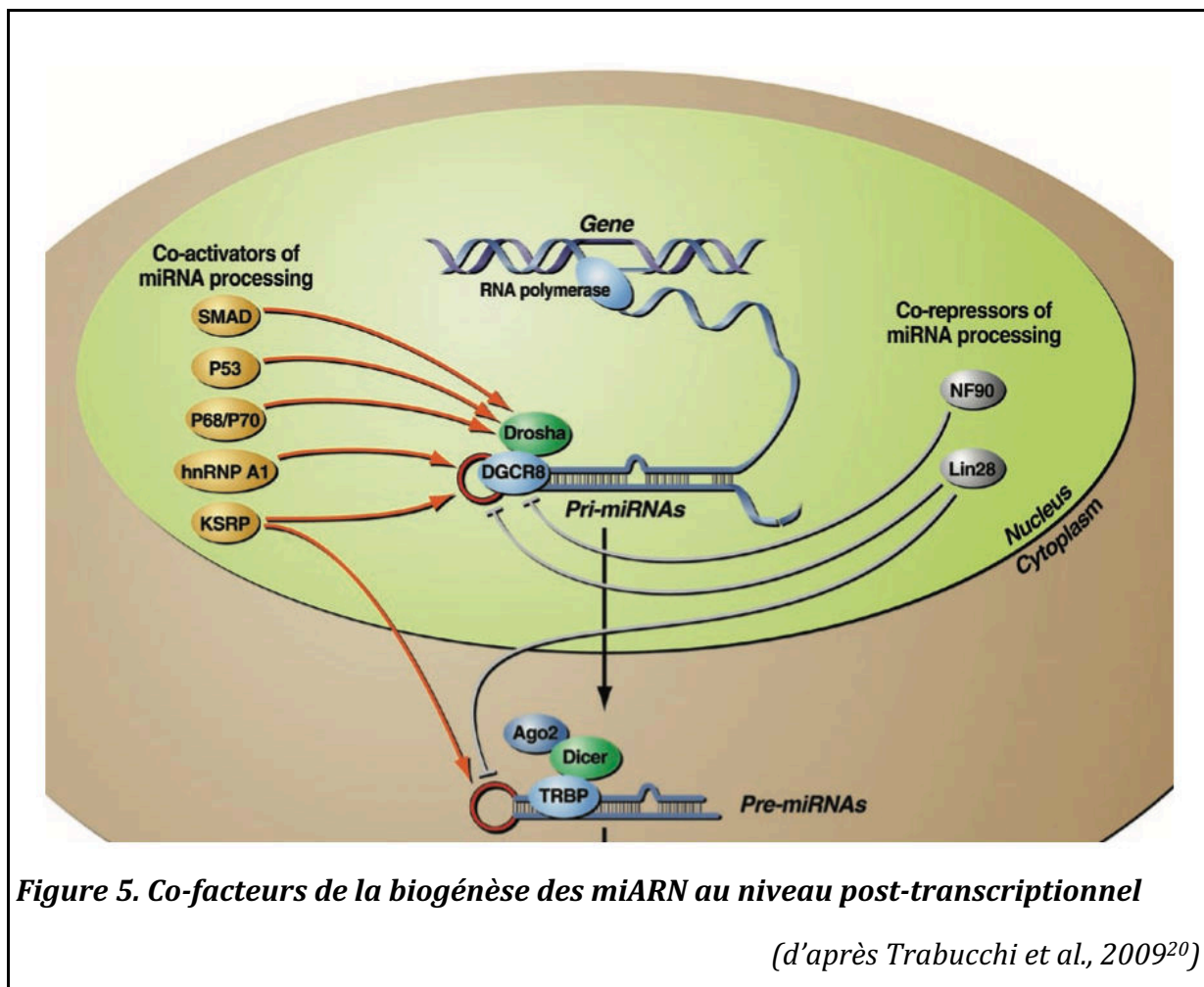
**Tableau 1. Facteurs de transcription régulant la transcription de miARN**

### 7.5.3 Régulation post-transcriptionnelle de l'expression des micro-ARN

Certaines protéines se fixent à la boucle terminale ou la base du duplex pri-miARN et bloquent ou stimulent la maturation du pré-miARN (**Figure 5 & Tableau 2**)<sup>123,124</sup>.

Par exemple, la protéine LIN28 reconnaît la boucle terminale du pre-let-7, et recrute une terminal-uridylyl transférase induisant la poly-uridylation de l'extrémité 3', ce qui stimule la dégradation du pre-let-7 par l'exonucléase cytoplasmique DIS3L2<sup>125</sup>.

De manière complémentaire, certaines protéines comme KSRP se fixent à la boucle terminale des pri-miARN et favorisent la maturation de certains miARN par le complexe DROSHA/DGCR8 et DICER<sup>124</sup>.



Protéine	Mode d'action	miARN(s)	Rôle	Référence
DDX5/ DDX17	Interaction avec Drosha	35% des miARN	Active	126,127
LIN28	Interaction avec la tige-boucle	Famille let-7	Inhibe	128
KSRP	Interaction avec la tige-boucle	Famille let-7 miR-155	Active Active	124 19
SMAD	Interaction avec DDX5	miR-21	Active	129
P53	Interaction avec Drosha	miR-16/143/206	Active	127
HNRNPA1	Interaction avec la tige-boucle	miR-18a	Active	130
NF90/ NF45	Interaction avec la tige-boucle	miR-30a/miR-21 pre-let-7a	Inhibe	131

**Tableau 2. Facteurs de régulation post-transcriptionnels des miARN**

## 7.6 Mécanismes d'action cytoplasmique des micro-ARN

Plusieurs mécanismes ont été successivement proposés pour rendre compte de l'inhibition de l'activité des gènes par les miARN. Dans cette section nous décrivons l'essentiel des fonctions des miARN et du complexe miRISC, qui ont été étudiées dans le cytoplasme.

### 7.6.1 Inhibition de la traduction de l'ARNm

Dans un premier temps, l'inhibition de la traduction en protéine de l'ARNm a été proposée comme principal mécanisme d'action des miARN. Cette hypothèse a été initialement formulée à partir d'études chez le nématode *C.elegans*, où il a été montré par des méthodes de quantification du transcrite par protection à la RNase une absence de changements dans la quantité d'ARNm ou du profil de distribution polysomal du gène cible LIN-14 suite à l'expression du micro-ARN lin-4<sup>75,76,132</sup>. De plus, les premières expériences menées dans des cellules de mammifères ont décrit l'inhibition d'un gène rapporteur (chloramphénicol acétyltransférase) sans modification détectable de la stabilité de l'ARNm<sup>133</sup>. Accessoirement, les premières expériences d'ancrage de la protéine AGO2 sur un gène rapporteur luciférase qui utilisent le système phage  $\lambda$  N-peptide et boucle BoxB n'ont pas révélé de changement notable de la quantité d'ARNm suite à l'inhibition de l'activité luciférase<sup>134</sup>.

En fait, il a finalement été montré par la suite que la distribution polysomale d'un gène rapporteur luciférase contenant trois sites de liaison à let-7 est bien modifiée suite à l'inhibition de let-7 via un oligonucléotide antisens, et passe de la fraction non-traduite à la fraction activement traduite (dense) dans des gradients de sucrose<sup>135</sup>.

Plusieurs études à haut-débit du profil de distribution polysomale des ARNm ont confirmé cette hypothèse<sup>136,137</sup>.

### 7.6.2 Déstabilisation et dégradation de l'ARNm

Bien que l'inhibition de la traduction puisse rendre compte d'une partie du mécanisme d'action des miARN, c'est la déstabilisation et la dégradation de l'ARNm qui semble être la cause principale de la diminution de l'expression protéique.

Les premières études mécanistiques ont rapporté une diminution de la quantité d'ARNm du gène *lin-14* en réponse à l'expression de *lin-4*, qui est effectivement observable en se basant sur une méthode de détection par northern blot<sup>138</sup>. De plus, la déstabilisation de l'ARNm d'un gène rapporteur luciférase a été observée suite à l'expression du miR-30 dans des cellules de mammifères<sup>139</sup>.

Au niveau mécanistique, l'insertion de six sites de liaison au miR-125 dans un gène rapporteur luciférase induit une diminution de la quantité d'ARNm en présence de miR-125<sup>140</sup>, et la déstabilisation de l'ARNm commence par la dégradation de la queue polyadénylée<sup>141</sup>. Dans un extrait cytoplasmique issu de cellules Krebs-2, la traduction *in-vitro* d'un gène rapporteur luciférase contenant six sites de liaison à let-7 est inhibée d'une manière let-7-dépendante et coiffe m<sup>7</sup>G-dépendante<sup>142</sup>.

En fait, la déadénylation de l'ARNm est la deuxième étape après l'inhibition traductionnelle, et est détectable après une incubation prolongée de l'ARN dans un extrait cytoplasmique<sup>143</sup>. Au niveau transcriptomique, la transfection de petits ARN interférents ou de miARN induit une diminution de la quantité de nombreux ARNm cibles<sup>144,145</sup>. Inversement, l'inhibition de l'activité des miARN *in-vivo* via des oligonucléotides antisens induit une stabilisation de nombreux ARNm<sup>146</sup>.



### 7.6.3 Biochimie du complexe effecteur miRISC

Les mécanismes d'inhibition traductionnelle et de dégradation de l'ARNm induits par les miARN reposent sur l'interaction des protéines Argonaute avec les protéines GW182<sup>147</sup> (famille TNRC6 chez les vertébrés) et le complexe CCR4-NOT1.

Dans ce chapitre nous détaillons les interactions fonctionnelles de ces composants.

#### 7.6.3.1 Structure et fonction des protéines Argonaute

Le composant principal du complexe miRISC, une protéine de la famille Argonaute, a été identifié à partir des études sur le mécanisme de l'ARN interférence.

En effet, les composants moléculaires du mécanisme d'ARN interférence ont été initialement fractionnés par centrifugation à haute vitesse. Dans des extraits de pancréas de rat, la protéine de la famille *Piwi/Argonaute*, GERp95 (AGO2), co-purifie avec les membranes intracellulaires et est enrichie dans la fraction microsomale<sup>148</sup>.

L'ARN interférence et l'activité de clivage est conservée chez pratiquement tous les eucaryotes<sup>i</sup>, y compris *Drosophila melanogaster*<sup>151</sup>. Les expériences sur la découverte du complexe RISC ont été menées dans les cellules de Schneider de la drosophile (cellules de type macrophage). L'activité de clivage associée aux petits ARN interférents sédimente dans le culot ribosomal et est partiellement éliminée du surnageant cytoplasmique S100<sup>152</sup>, tandis que l'activité qui génère les petits ARN à partir de l'ARN double-brin reste dans le surnageant cytoplasmique, et peut être reconstituée *in-vitro* uniquement à partir de l'immunoprécipitation de la RNase de type III CG4792 (Dicer) mais pas de la RNase III Drosha ou de l'hélicase *homeless*<sup>102</sup>.

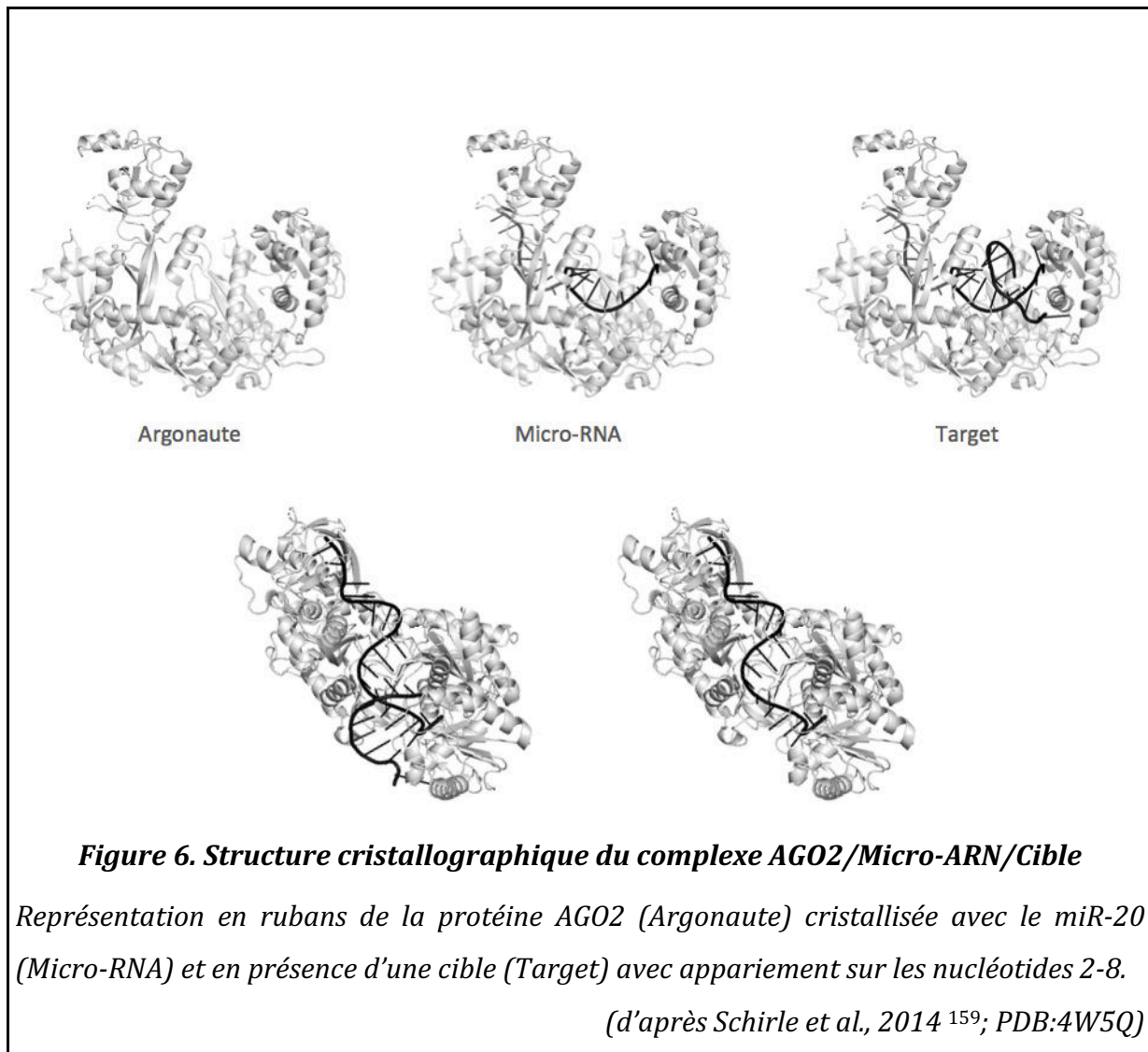
L'activité de clivage peut être extraite à partir de la solubilisation partielle du culot ribosomal après un traitement à haute force ionique. Le fractionnement par chromatographies successives de cette activité mène à l'isolation et l'identification d'un complexe de 500 kDa (le complexe RISC) contenant la protéine AGO2<sup>103</sup>, qui co-purifie également avec des miARN<sup>153</sup>.

Les petits ARN interférents sont incorporés dans la protéine AGO2 grâce à des interactions avec le squelette sucre-phosphate et des domaines de liaison spécifiques aux acides nucléiques (**Figure 6**). Les études menées sur la protéine AGO2 de *Drosophila*

---

<sup>i</sup> *Saccharomyces Cerevisiae* a perdu les gènes AGO1 et Dicer; et l'activité catalytique de la protéine AGO2 de *Danio Rerio* est peu efficace<sup>149,150</sup>.

*melanogaster* ont révélé que AGO2 contient un domaine de liaison aux acides nucléiques (le domaine PAZ) qui interagit seulement avec l'extrémité 3' du petit ARN mais pas de l'ARN double-brin. Les études menées sur le domaine Piwi (P-element induced wimpy testis<sup>154</sup>) de la protéine Argonaute de *Pyrococcus furiosus* ont révélé que ce domaine confère quant à lui les propriétés catalytiques à la protéine avec une structure proche de la RNase H<sup>155</sup>. La protéine AGO2 est essentielle pour le développement de la souris, et est chez les mammifères l'unique membre de la famille Argonaute (AGO1 à AGO4) qui conserve une activité catalytique<sup>156,157</sup>. Cette activité catalytique de type RNase H du domaine PIWI dépend de résidus carboxylate (Asp, Asp, Glu, ou DDE) qui sont essentiels pour la coordination d'un ion métallique ( $Mg^{2+}$ ) pour activation d'une molécule d'eau et attaque nucléophile de la liaison phosphodiester, d'après les expériences menées avec des protéines AGO2 recombinantes purifiées<sup>157</sup>. Enfin, les études menées sur la protéine Piwi de *Archaeoglobus fulgidus* et la protéine AGO2 humaine ont révélé que les protéines Argonaute/Piwi contiennent un domaine qui se lie uniquement aux extrémités 5'- $PO_4^{2-}$  des petits ARN (le domaine MID) grâce à des liaisons hydrogènes avec les atomes d'oxygène du groupement phosphate, des interactions d'empilement aromatique avec la base nucléotidique et la coordination octaédrique d'un ion magnésium<sup>158</sup>. La première nucléobase du miARN est dans une conformation inaccessible pour des liaisons hydrogènes avec la cible<sup>159</sup>. Ce premier nucléotide est presque invariablement une uridine chez les métazoaires (occasionnellement une adénosine)<sup>77</sup>, qui est également la nucléobase préférée de la poche hydrophobique avec une affinité plus de trente fois supérieure aux cytosines ou guanosines<sup>160</sup>. L'extrémité 5'- $PO_4^{2-}$  est nécessaire pour l'incorporation du brin guide et un mécanisme ARN interférence efficace<sup>161,162</sup>. Les extrémités 5'-OH d'un ARN double-brin sont converties en extrémités 5'- $PO_4^{2-}$  dans les extraits cellulaires et peuvent déclencher l'ARN interférence, cependant les protéines AGO2 eucaryotes recombinantes purifiées ne peuvent pas être reconstituées avec des petits ARN avec une extrémité 5'-OH<sup>155</sup>.



### 7.6.3.2 Rôle du complexe CCR4-NOT1-CAF1 et des protéines GW182

Chez les plantes, levures ou champignons, les protéines Argonautes sont découplées du complexe d'inhibition traductionnelle et de déadénylation CCR4-NOT1-CAF1.

Le génome des métazoaires contient un gène de la famille GW182 (plusieurs gènes chez les vertébrés, formant la famille TNRC6), et produit une protéine qui connecte le système Argonaute au complexe CCR4-NOT1-CAF1. Cette évolution élargit l'éventail des fonctions du complexe RISC en ajoutant la fonction d'inhibition traductionnelle et de déadénylation aux mécanismes d'action du RISC chez les métazoaires, en particulier en l'absence de clivage par la protéine Argonaute lors d'appariement imparfaits<sup>163</sup>. Le complexe CCR4-NOT1-CAF1 est composé d'une dizaine de protéines, parmi lesquelles le

complexe minimal composé de CNOT1, CNOT2, CNOT3, CNOT7 (CAF1) et CNOT9 figure parmi les plus conservés dans l'évolution.

Dans les cellules de Schneider, l'ancrage artificiel de la protéine GW182 à un transcrit luciférase induit la dégradation de l'ARNm d'une manière Dcp1/2 et Not1-dépendante<sup>164</sup>. De plus, l'inhibition du gène GW182 ou AGO1 produit des perturbations du transcriptome très similaires, et le fragment N-terminal de la protéine GW182 contient des motifs riches en glycine/tryptophane, et co-immunoprécipite avec AGO1<sup>164</sup>. Une combinaison de deux motifs GW ou WG (le motif-crochet Ago, ou Ago-hook motif) est essentielle pour l'interaction avec le domaine Piwi des protéines Argonaute<sup>i</sup> <sup>165</sup>. Il a par la suite été montré que les protéines GW182 sont directement responsables du recrutement des déadénylases (PAN2/PAN3), mais également du complexe CCR4-NOT1-CAF1 de déadénylation et d'inhibition de la traduction.

Une protéine importante pour le recrutement du complexe CCR4-NOT1-CAF1 par GW182 est la protéine CNOT9, qui est l'un des gènes les plus conservés dans l'évolution<sup>ii</sup>. Les interactions entre GW182 et CNOT1 sont stabilisées en présence de la protéine CNOT9. Les domaines répétés armadillo de CNOT9 présentent une structure en super hélice-alpha présentant deux poches hydrophobiques, qui peuvent accepter deux résidus tryptophane en tandem tels que ceux que l'on peut trouver dans la protéine GW182<sup>166,167</sup>. Chez la drosophile, la protéine CNOT9 interagit également avec des protéines de liaison à l'ARN telle que Bag-of-marbles, indépendamment de la liaison à GW182<sup>168</sup>.

#### **7.6.3.2.1 Fonctions de déadénylation et de dégradation de l'ARNm**

Dans les extraits d'ascites de Krebs-2, on observe la co-immunoprécipitation de la protéine AGO2 et la déadénylase CNOT7 (CAF1) en présence de nucléase micrococcalle. De plus, la déadénylation d'un gène rapporteur luciférase contenant six sites de liaison à let-7 est inhibée par l'addition d'un peptide compétiteur correspondant au Ago-hook motif (TNRC6B fragment 946-966), mais pas en présence du peptide compétiteur

---

<sup>i</sup> A l'exception de la protéine AGO2 chez la drosophile ou des protéines Piwi chez l'humain.

<sup>ii</sup> Aussi chez plantes, levures et champignons.

mutant (TNRC6B fragment 946-966, W952L et W962L). La région N-terminale de TNRC6C contient un motif d'interaction avec la protéine PABPC1 (Pam2 motif) qui est requis pour un niveau de déadénylation maximal du gène rapporteur. Le modèle proposé est la déstabilisation de l'interaction EIF4G-PABPC1, qui protège l'ARNm par circularisation<sup>143</sup>.

La région N-terminale de GW182 contient douze répétitions des motifs GW/WG où l'un des motifs GWG contribue davantage à la liaison avec AGO1 que les autres. Cependant l'ensemble des motifs doit être muté pour empêcher toute interaction avec AGO1<sup>169</sup>. De plus, le domaine C-terminal de GW182 (SD, fragment 1260-1690) interfère avec la traduction, accélère la dégradation de l'ARNm<sup>169,170</sup>, et contient un second motif d'interaction avec PABPC1 (Pam2 motif) position 1382<sup>171-173</sup>. Dans des expériences d'ancrage à un gène rapporteur, la mutation du motif Pam2 dans le domaine SD de TNRC6C réduit la vitesse de déadénylation du gène rapporteur, mais c'est l'extrémité C-terminale la plus distale de TNRC6C (1595-1690) qui contribue à une activité de déadénylation processive<sup>174</sup>.

En effet, la région C-terminale de TNRC6C recrute le complexe CCR4-NOT1 à l'aide de deux motifs LWG conservés (CIM1 position 1294-1300 et CIM2 position 1647-1660) d'une manière CNOT1-dépendante mais indépendamment de l'interaction avec PABPC1<sup>174</sup>. La région C-terminale recrute aussi les poly(A) nucléases PAN2 et PAN3 indépendamment du complexe CCR4-NOT1<sup>175,176</sup>. De plus, dans les expériences d'ancrage du complexe CCR4-NOT1 à un gène rapporteur, on peut observer que la dégradation de l'ARNm continue après que la queue polyadénylée a été ôtée, et se termine aux boucles BoxB<sup>174</sup>.

#### **7.6.3.2.2 Fonction du complexe CCR4-NOT1 dans le décoiffage en 5' de l'ARNm**

En plus du mécanisme de déadénylation, le complexe RISC interagit avec des facteurs impliqués dans le décoiffage en 5' de l'ARN. A l'intérieur des « processing-bodies » (P-bodies) cytoplasmiques, AGO2 interagit avec des complexes contenant DDX6 (RCK/P54)<sup>177</sup>. DDX6 est l'homologue humain du gène Dhh1p, une ARN hélicase à motif DEAD-box impliquée dans le décoiffage en 5' et dans l'inhibition de la traduction<sup>178</sup>.

Chez la levure, la surexpression du gène Dhh1p complémente le phénotype des souches délétées pour les déadénylases Ccr4 ou Pop2<sup>179</sup>. Inversement, on observe une accumulation d'ARN coiffés en 5' dans les souches  $\Delta$ Dhh1p ainsi qu'une augmentation de la demi-vie des ARNm<sup>178</sup>. Cet effet s'explique par l'interaction de Dhh1p avec le facteur de décoiffage Dcp1p dans des expériences de double-hybride<sup>180</sup>, ce qui entraîne une répression de la traduction *in-vitro*<sup>181</sup>. Chez la drosophile, on retrouve ce mécanisme avec l'homologue du gène DDX6 (ME31B), qui est requis pour l'activité des miARN en combinaison avec DCP1 et d'autres facteurs de décoiffage<sup>182</sup>.

C'est le domaine MIF4G de CNOT1 qui recrute directement la protéine DDX6, et peut aussi recruter simultanément la déadénylase CNOT7<sup>167</sup>. L'interaction de CCR4-NOT1-CAF1 avec DDX6 est requise pour une activité de dégradation maximale dans des expériences d'ancrage à un gène rapporteur<sup>167</sup>, et recrute les facteurs de décoiffage EDC3, Pat1<sup>183</sup> ou EIF4E-T<sup>184</sup>.

#### **7.6.4 Séquestration des ARNm dans les « P-bodies » cytoplasmiques**

En général, le complexe CCR4-NOT1-CAF1 présente un profil de marquage diffus dans le cytoplasme<sup>185</sup>, cependant il est possible d'observer une accumulation de CCR4 dans les P-bodies<sup>186</sup> et CNOT1 co-localise avec DCP1A<sup>187</sup>. Bien que l'essentiel des facteurs du catabolisme de l'ARN soient concentrés dans les P-bodies, l'assemblage des granules n'est pas essentiel pour observer l'activité de dégradation de l'ARNm par les miARN, mais par contre l'inhibition de l'activité des miARN ou du complexe RISC entraîne la disparition des P-bodies<sup>188</sup>. D'après des expériences de transcriptomique et protéomique, on observe que les gènes dont les ARN messagers sont enrichis dans les P-bodies sont traduits moins efficacement, la distribution de la queue poly(A) est hétérogène et montre une préférence pour des longueurs minimales ou maximales, mais les ARN messagers ne semblent pas être dégradés rapidement. Ces observations suggèrent davantage un rôle de stockage des ARNm ciblés par les miARN dans les P-bodies<sup>189</sup>.

### 7.6.5 Régulation de l'activité des micro-ARN par les protéines de liaison à l'ARN

Le ciblage des ARNm par les miARN peut être levé par le recrutement local de protéines de liaison à l'ARN (**Figure 7 & Tableau 3**). Certains ARN dé-réprimés quittent ainsi les P-bodies et l'état non-traduit pour réintégrer les ribosomes et la traduction.

Un des premiers exemples concerne la protéine ELAVL1 et ses sites de liaison dans la région 3'UTR du gène SLC7A1 (CAT1)<sup>190</sup>. Dans les cellules humaines d'hépatocarcinome, le ciblage de SLC7A1 par miR-122 est empêché par la liaison de ELAVL1 (HuR) à l'ARNm SLC7A1 suite à un stress de privation d'acides aminés<sup>190</sup>. Dans un autre exemple, la protéine DND1 empêche le ciblage de l'ARNm P27 par miR-221, ou le ciblage de l'ARNm NANOS1 par le miR-430 dans les cellules germinales primordiales<sup>191</sup>.

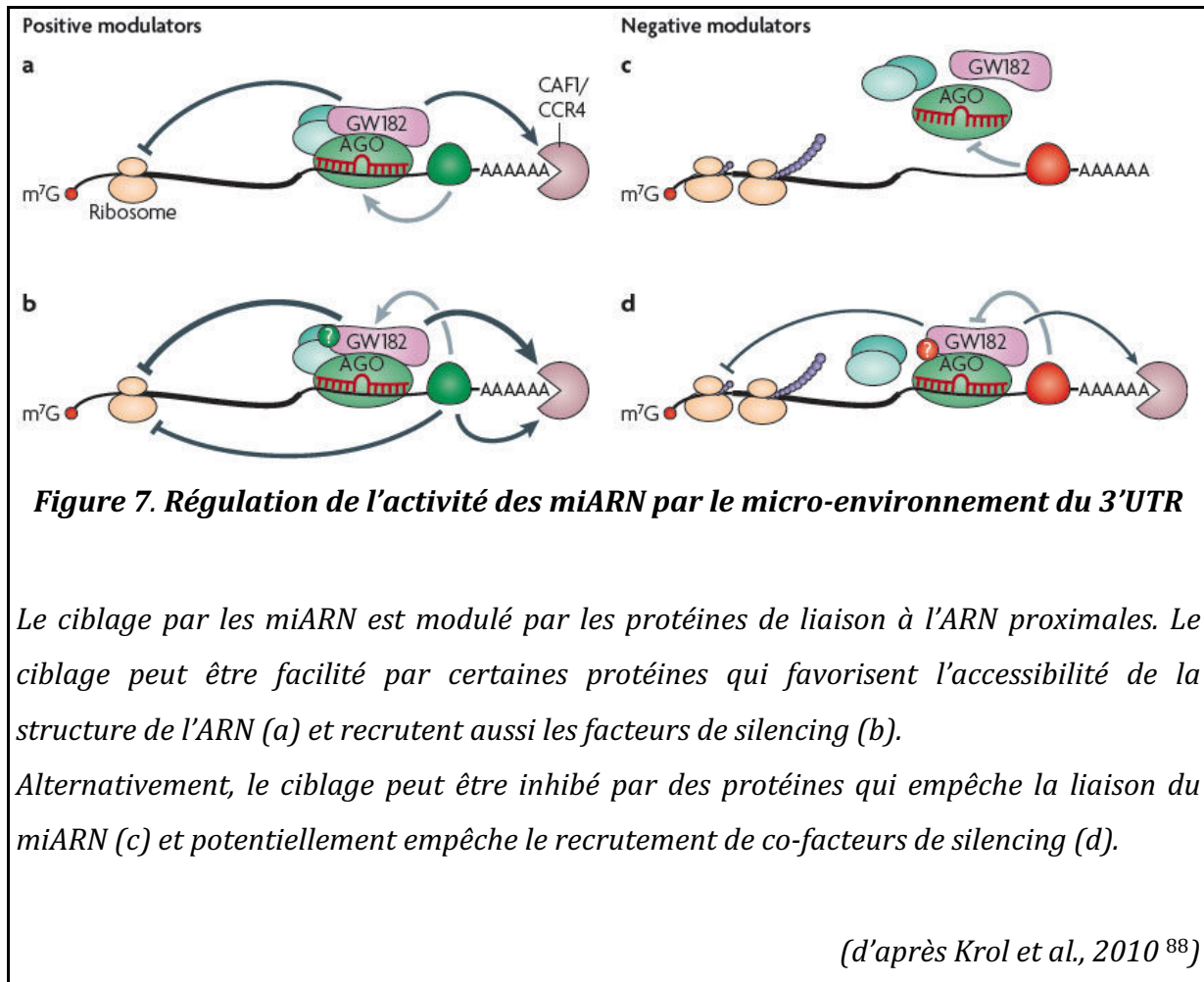
A l'inverse, certaines protéines de liaison à l'ARN facilitent l'activité de ciblage par les miARN<sup>192</sup>, et le ciblage de l'ARNm MYC est facilité par la protéine ELAVL1<sup>193</sup>.

Les fibroblastes engagés en division cellulaire expriment la protéine Pumilio-1 (PUM1) qui active le ciblage de l'ARNm P27 par miR-221/222, en se liant à des éléments de reconnaissance par Pumilio (PRE) localisés dans l'ARNm P27. Dans les fibroblastes quiescents, la protéine PUM1 n'est pas exprimée, et les éléments de reconnaissance PRE s'hybrident aux sites de liaison de miR-221/222 dans une conformation double-brin stables en l'absence de la protéine PUM1<sup>194</sup>. En plus de maintenir les sites de miR-221/222 dans une conformation accessible, les protéines Pumilio pourraient également faciliter l'assemblage du RISC en interagissant avec le complexe CCR4-NOT1-CAF1<sup>195</sup>. Enfin, certaines protéines sont impliquées à la fois dans l'inhibition et l'activation du ciblage par les miARN. Par exemple, la liaison de la protéine PTB à ses sites dans le 3'UTR de l'ARNm SCP1 inhibe le ciblage par miR-96 et miR-124, tandis que la liaison de PTB au 3'UTR de l'ARNm GNPDA1 facilite le ciblage par let-7<sup>196</sup>.

D'autres facteurs régulent l'activité des miARN, comme les ARN hélicases ou des protéines de liaison à l'ARN. Chez la drosophile, on observe des mutations dans le gène Armitage, une ARN hélicase ATP-dépendante à boîte DEA(H/D) qui correspond au gène Sde-3 (silencing defective-3) chez les plantes<sup>197</sup>. Ces mutations empêchent l'incorporation des petits ARN interférents dans le complexe RISC après l'étape de maturation par Dicer-2/R2D2<sup>198,199</sup>. Chez l'humain, le gène MOV10 (Armitage) est nécessaire pour le mécanisme d'ARN interférence<sup>147</sup>, et co-immunoprécipite avec les protéines Argonaute dans les P-bodies<sup>147</sup>. Dans les neurones, la protéine MOV10

augmente l'activité de miR-138 sur un sous-ensemble de ses cibles telles que LYPLA1 à la synapse<sup>200</sup>.

Un autre exemple de co-facteur concerne la protéine HNRNP-Q2 qui fixe la queue poly(A) des ARNm, déplace la protéine PABP et lève l'inhibition induite par les miARN de manière générale<sup>201</sup>.





Protéine	Gène cible	miARN régulé	Rôle	Référence
ELAVL1	CAT1	miR-122	Inhibe	190
	TOP2A	miR-548c	Inhibe	202
	NCL	miR-494	Inhibe	203
	MT-CO2	miR-16	Inhibe	204
	ERBB2	miR-331	Inhibe	205
	CRP	miR637	Inhibe	206
	MYC	let-7	Active	193
	RHOB	miR-19	Active	207
DND1	P27	miR-221	Inhibe	208
	LATS2	miR-372	Inhibe	208
	NANOS	miR-430	Inhibe	208
	TDRD7	miR-430	Inhibe	208
PUM1	P27	miR-221/222	Inhibe	194
	E2F3	miR-502/125b	Active	209
PTB	SCP1	miR-124	Inhibe	210
	SCP1	miR-96	Inhibe	210
	GNPDA1	miR-181b/196a/let-7b	Active	196
Igf2bp1	$\beta$ -Trcp1	miR-183	Inhibe	211
	MITF	miR-340	Inhibe	212
Rbm38	c-MYB	miR-150	Inhibe	213
Roquin	ICOS	miR-146	Active	214

**Tableau 3. Protéines de liaison à l'ARN et régulation du ciblage des miARN**

## 7.7 Ciblage des micro-ARN

### 7.7.1 Mécanismes d'hybridation complémentaire des miARN à leur cibles

L'hybridation du miARN à sa cible repose principalement sur l'appariement des nucléotides de la région 5' du miARN, ou la région de nucléation du duplex (région seed) aux ARN cibles. La région seed comprend typiquement les nucléotides 2 à 7 du miARN

qui sont pré-arrangés en forme hélicoïdale de type A (**Figure 6**) en l'absence de l'ARN cible<sup>159</sup>. Il s'agit de la région minimale requise pour l'hybridation du miARN à sa cible<sup>215</sup>, et l'on observe une complémentarité de séquence entre la région seed et des motifs impliqués dans une forme de régulation post-transcriptionnelle localisés dans le 3'UTR de nombreux ARNm<sup>216</sup>.

La constante de dissociation d'un complexe AGO2-miARN sur une cible comprenant entre 3 et 6 nucléotides complémentaires à la région seed ( $3 \leq N \leq 6$ ) est d'environ 1 seconde, tandis que sur une cible présentant 7 nucléotides complémentaires cette constante est supérieure à 300 secondes, d'après des expériences de transfert d'énergie de fluorescence par résonance en chambre microfluidique<sup>217</sup>. La substitution de 2 nucléotides dans la région seed entraîne l'inactivation du ciblage d'un gène rapporteur luciférase<sup>215</sup> ou d'un gène rapporteur GFP dans des embryons de zebrafish<sup>218</sup>.

### 7.7.2 Prédiction des sites de liaison aux miARN

L'hybridation avec la région seed (**Figure 8**) est le critère principal à la base des logiciels pour prédire les cibles des miARN <sup>219</sup>, mais d'autres paramètres complémentaires interviennent<sup>220</sup>.

Etant donné que la fréquence d'une séquence aléatoire de 6 nucléotides est d'environ 1 occurrence toutes les 4 kilobases de séquence génomique, chaque miARN possède virtuellement plusieurs centaines de milliers de sites de liaison dispersés dans le génome. Cependant, tous les sites ne sont pas fonctionnels, comme on peut l'observer par mutagenèse dirigée dans des expériences où le 3'UTR comprenant le site de liaison est cloné en aval d'un gène rapporteur luciférase<sup>219</sup>, ou en évaluant la régulation effective des gènes contenant les sites prédits par analyse transcriptomique sur puce microarray<sup>145</sup>. L'efficacité des sites de liaison dépend également de l'accessibilité et de l'occupation physique du site par le complexe RISC<sup>221</sup>.

Avec le séquençage du génome humain et d'autres espèces, les premières estimations informatiques du nombre de sites des miARN sur la base de la complémentarité de séquence avec la région seed (2-8) chez la souris et le rat ont fait état de 451 sites (400 gènes) ou 5.7 sites par miARN en utilisant un jeu de données de 79 miARN conservés chez les mammifères et un panel de 14 539 gènes orthologues. Dans ces analyses, un jeu

de données contrôle consistant en la région seed des miARN dont la séquence est randomisée (avec le même contenu di-nucléotidique) retourne 140 sites de liaison ou 1.8 sites par miARN contrôle/virtuel<sup>219</sup>. En utilisant des régions alternatives d'hybridation au miARN autres que la région seed, aucun site ne peut être prédit au-dessus de ce bruit de fond.

L'utilisation de critères d'hybridation moins stringents de la région seed (2-7) et l'ajout de génomes supplémentaires comme celui du poulet et du chien à l'analyse de la conservation des sites génère une estimation de 8 484 sites (2 767 gènes) pour un jeu de données de 148 miARN communs aux mammifères et 10 896 gènes orthologues, soit 57 sites par miARN<sup>222</sup>.

Si l'on autorise des mésappariements ou des bulges dans l'hybridation de la région seed (avec compensation de l'énergie libre de liaison), les estimations passent à 10 000 sites au-dessus du bruit de fond pour un jeu de données de 58 miARN communs aux amniotes et 18 567 gènes orthologues (en ignorant les séquences répétées), soit 172 sites par miARN<sup>223</sup>.

Etant donné la multiplicité des sites de liaisons, la fréquence de co-occurrence de sites de miARN différents sur le même ARNm a été évaluée afin d'identifier les cibles avec haute probabilité de régulation pour une combinaison de miARN. En effet, sur un jeu de donnée de 58 miARN communs aux amniotes, il existe au moins 1 ARNm commun à environ deux-tiers de toutes les combinaisons de triplets possibles (30 856 combinaisons), mais cette probabilité passe à 0.04 pour  $\geq 10$  ARNm et à 0.001 pour  $\geq 25$  ARNm<sup>223</sup>. De plus, les cibles à « effet de levier » (dont l'expression est fortement co-régulée par plusieurs sites et plusieurs miARN) sont rares<sup>224</sup>, les plus connues étant le ciblage de *lin-14* par *lin-4* (7 sites) et *let-7* (3 sites)<sup>75,76,225</sup>. En effet, entre *Drosophila melanogaster* et *Drosophila pseudoobscura*, il existe peu (<5%) de cibles qui possèdent plus d'un site conservé pour un miARN donné. Le cas de figure le plus répandu (presque 50% des cibles) est une combinaison d'un seul et unique site pour plusieurs miARN, et un grand nombre de gènes évitent les sites de liaison aux miARN tels que les gènes ubiquitaires ou l'ARN ribosomal. On observe aussi que les gènes à expression tissu-spécifique sont généralement dépourvus de sites de liaison pour les miARN co-exprimés dans le type cellulaire concerné<sup>226</sup>.

L'incorporation d'une méthode de distance phylogénétique pour détecter les régions locales de sélection positive dans les alignements de 3'UTR a permis une des analyses

les plus exhaustives de l'estimation du nombre de sites. En utilisant un jeu de données comprenant 87 familles de miARN et des alignements de séquences chez 22 vertébrés, il existe environ  $46\,441 \pm 2\,175$  sites ( $10\,739 \pm 564$  gènes) ou  $534 \pm 25$  sites par miARN au-dessus du bruit de fond pour les gènes codants (3'UTR) chez l'humain<sup>227</sup>. Il y a peu de sites avec des mésappariements ( $625 \pm 239$  sites) ou une hybridation complémentaire avec la partie 3' du miARN ( $714 \pm 315$  sites) qui sont sous sélection positive. A l'inverse, il y a un grand nombre de sites de liaison canoniques de type 6-mers (2-7), soit  $10\,970 \pm 1\,909$  sites (2-7), et  $8\,803 \pm 1\,276$  sites de type 6-mers décalés (3-8) qui semblent sélectivement maintenus, bien que leur efficacité soit plus faible par rapport aux 7-mers.

Les nucléotides adjacents aux sites de liaison des miARN sont davantage conservés. En particulier, la moitié (40%) des bases préférentiellement conservées en position 1 du site sont des adénosines, peu importe l'identité de la séquence du miARN sélectionné<sup>222</sup>. Cette particularité s'explique par l'existence d'une poche de surface de liaison aux nucléobases sur la protéine AGO2, qui s'accommode spécifiquement aux adénosines avec trois fois plus d'affinité que les cytosines, uridines ou guanosines<sup>159</sup>. Par conséquent, parmi toutes les possibilités d'hybridation d'un miARN à sa cible, la combinaison de type 8-mer ou plus (l'appariement parfait sur 7 nucléotides consécutifs avec une adénosine flanquante position 1) possède le potentiel de stabilité, et donc de régulation le plus important. Après la transfection d'un miARN, la régulation médiane des ARNm possédant un site 8-mer est d'environ 0.80-fold (20%), et est multipliée à 0.65-fold avec la présence de deux 8-mers. Il y a synergie des sites proches, de telle sorte que la régulation d'un site à 0.9-fold (10%) est démultipliée à 0.55-fold (45%) en présence d'un deuxième site compris entre 8 à 34 nucléotides du premier<sup>i</sup>, au lieu du produit des sites égal à 0.81-fold (19%) lorsqu'ils sont espacés d'une plus grande distance<sup>228</sup>.

Au-delà du rôle de la région seed, on observe la corrélation de certains paramètres avec l'efficacité des sites. En se basant sur leur performance par analyse transcriptomique, les sites les plus efficaces sont localisés dans des régions légèrement enrichies en nucléotides A et U ( $\leq 5\%$ )<sup>228</sup>, ce qui s'explique par une énergie libre moins élevée de la structure de l'ARN, et un accès facilité au complexe AGO2-miARN<sup>229</sup>. De plus, une

---

<sup>i</sup> Cette particularité est inexplicquée, la stabilisation du ciblage par un complexe RISC avec deux ou trois protéines Argonaute/miARN pourrait être à l'origine de cet effet.

hybridation complémentaire avec la partie 3' du miARN, en particulier aux nucléotides 13-16 est associée à une meilleure efficacité<sup>228</sup>. Cependant, bien que les analyses informatiques montrent que la complémentarité avec les nucléotides 13-16 est le type d'hybridation de 4-mers le mieux conservé après la région seed<sup>230</sup>, et semble compenser pour une hybridation imparfaite avec la région seed comme dans le cas du ciblage de lin-41 par let-7<sup>80</sup>, ce mécanisme d'hybridation est peu répandu<sup>231</sup> et représente moins de 2% de la totalité des sites préférentiellement conservés des miARN chez l'humain<sup>227</sup>. Enfin, les sites localisés dans le premier quart ou le dernier quart du 3'UTR sont généralement plus efficaces sauf lorsqu'ils sont proches du codon stop (<15 nucléotides), tandis que les sites situés dans la région codante ou le 5'UTR sont sur le chemin du ribosome et sont peu ou pas efficaces en raison de l'interférence avec la machinerie de traduction<sup>232</sup>.

Au niveau protéomique et suite à la transfection d'un miARN, environ la moitié (55/90) des protéines réprimées d'au moins 50% dérivent d'ARNm comprenant au moins un site de type 7-mer, et environ un tiers des protéines (152/420) réprimées d'au moins 25% dérivent d'ARNm comprenant un site de type 7-mer. Les sites de type 6-mer montrent quant à eux une régulation relativement faible<sup>233</sup>. Par exemple, dans un modèle de neutrophiles murins délétés pour le miR-223, on observe une stabilisation d'un sous-ensemble de protéines dérivant d'ARNm contenant des sites pour miR-223 sans pour autant accompagnée d'une stabilisation de l'ARNm, cependant les protéines dérégulées d'au moins 50% dérivent systématiquement d'ARNm dont la stabilisation est détectable<sup>233</sup>.

### 7.7.3 Cartographie expérimentale des sites de liaison des micro-ARN

Un grand nombre de cibles peut être prédit sur la base de l'hybridation avec la région seed, mais seulement 43% des ARNm contenant un site de type 8-mer sont régulés de manière détectable ( $\geq 15\%$ ) sur puce microarray, ce qui veut dire que les méthodes de prédiction in-silico surestiment le nombre de gènes régulés en pratique<sup>228</sup>. Les sites qui sont physiquement fixés par la protéine AGO2 peuvent être identifiés en irradiant les cellules aux ultraviolets pour induire la fixation covalente de la protéine AGO2 à ses cibles, ce qui permet après immunoprécipitation de purifier la protéine et les ARN associés de manière particulièrement stringente à travers une étape de gel dénaturant,

en combinaison avec un traitement à la RNase pour isoler et séquencer des fragments de taille médiane 50 nucléotides afin de préciser la région occupée (HITS-CLIP<sup>234</sup>).

### **7.7.3.1 Distribution de l'activité de liaison de la protéine Argonaute et des micro-ARN**

Les premières études menées avec la méthode de HITS-CLIP sur la distribution de la protéine AGO2 au niveau transcriptomique ont été menées dans le néocortex de la souris. Ces études ont révélé que près de 40% de l'activité de ciblage de AGO2 est concentrée dans la région 3'UTR<sup>221</sup>. Le reste de l'activité de ciblage de AGO2 est réparti entre les régions codantes et introniques ainsi que des transcrits intergéniques, et seulement 1% de l'activité de liaison est observée dans la région 5'UTR. Parmi les fragments d'ARN associés à AGO2, la probabilité d'observer des sites de liaison aux miARN exprimés dans le néocortex tels que miR-9 ou miR-30 suit une loi de distribution leptokurtique, ou centrée sur les fragments clustérisés<sup>i</sup>, en contraste avec les miARN non-exprimés dans le néocortex où la probabilité de distribution est aplatie ou platykurtique. La sélectivité des miARN et des sites est très élevée, pour chaque transcrite ciblé il y a en moyenne seulement deux sites majoritairement occupés par AGO2. Il est possible d'identifier ~655 sites de liaison par miARN dans le transcriptome de ces cellules, ce qui corrobore les estimations bio-informatiques (voir section 7.7.2).

### **7.7.3.2 Diversité des sites de liaisons : Sites non-canoniques et bulgés**

Dans le néocortex de la souris, il existe cependant 27% de l'activité de liaison de AGO2 qui est orpheline c'est-à-dire qu'aucun site canonique pour la plupart (88%) des miARN exprimés dans le néocortex n'est détectable<sup>221</sup>. Parmi ces clusters orphelins, le motif dont l'enrichissement est le plus significatif est un motif non-canonique qui correspond à la région seed de miR-124 avec un nucléotide « bulgé » ou renflé entre la position 5-6, et représente environ 6% du total de l'activité de liaison de AGO2. Le nucléotide qui forme un bulge avec la région seed de miR-124 est préférentiellement une guanosine, selon un modèle où le nucléotide engage d'abord une liaison avec la position pivot 6 de miR-124 et transitionne en bulge après que l'hybridation a progressé sur les nucléotides 6-8. Ce type de site non-canonique représenterait entre ≤15-22% de l'activité de liaison

---

<sup>i</sup> L'identification des régions de liaison dépend des programmes bio-informatiques utilisés pour reconstruire les clusters<sup>16</sup>. Dans ces études, le programme CIMS a été principalement utilisé.

de AGO2 dans le néocortex<sup>14</sup>. On observe aussi une fixation de AGO2 sur des ARN circulaires tels que cirS-7 (CDR1-AS) à de multiple sites correspondant à miR-7 avec hybridation complémentaire dans la région 3' <sup>235</sup>.

Dans une autre étude menée sur des lymphocytes T-helper murins, on observe également que 34% de l'activité de liaison de AGO2 dans les régions 3'UTR est orpheline et ne contient aucun site canonique pour les 40 miARN les plus exprimés tels que miR-155. Dans les même lymphocytes délétés pour miR-155, on observe une perte d'activité de liaison pour 191 sites répartie sur 175 gènes. Il y a 108 sites canoniques avec hybridation parfaite sur la seed-région, distribués entre les 7-mers avec adénosine (35 sites), les 8-mers (34 sites) suivis des 7-mers (25 sites) et 6-mers (14 sites). Parmi les 83 clusters restants ou orphelins, la moitié est constituée d'une combinaison de sites non-canoniques avec hybridation de la seed-région sur 6 à 8 nucléotides et un mésappariement position 5 (11 sites), un appariement de type wobble G:U position 5 (10 sites), un mésappariement position 7 (6 sites), un bulge position 6 (3 sites) ou deux bulge position 6 et 7 (3 sites). Chacun des sites non-canoniques est observé au maximum 5 fois et on observe une tendance de ces sites à s'hybrider de manière imparfaite avec la région 3' et les nucléotides 12-18 de miR-155. Les sites non-canoniques sont moins efficaces que les sites canoniques<sup>229</sup>.

#### **7.7.4 Fréquence de l'hybridation centrale aux miARN**

Le mode le moins fréquent d'hybridation du miARN est le ciblage central, c'est à dire avec les nucléotides 4-14 ou 5-15 du miARN. Ce mode de ciblage déclenche potentiellement le clivage par AGO2. Le clivage d'un ARN est possible avec une hybridation minimum de onze nucléotides (5-15) dans un système in-vitro d'extrait cellulaires S100, mais le clivage significatif de l'ARN requiert une hybridation consécutive sur au moins quinze nucléotides (2-16)<sup>230</sup>. Le clivage d'une cible quasi-parfaite voir parfaitement complémentaire par un miARN est observé chez les mammifères comme le ciblage du gène HOXB8 (avec un appariement G:U wobble position 5) par miR-196<sup>236</sup>, cependant les analyses informatiques montrent que le mode d'hybridation central est très rare. En effet, pour les miARN communs aux mammifères il existe en moyenne 13 sites de ce type dans la région 3'UTR sur un total de 19 109 gènes, et aucun des sites n'est conservé<sup>230</sup>.

Ce mode d'hybridation est beaucoup plus fréquent chez les plantes avec un rapport signal-sur-bruit de fond de 10:1, les sites sont conservés chez le riz et généralement enrichis dans les gènes avec une fonction de facteur de transcription<sup>237,238</sup>.

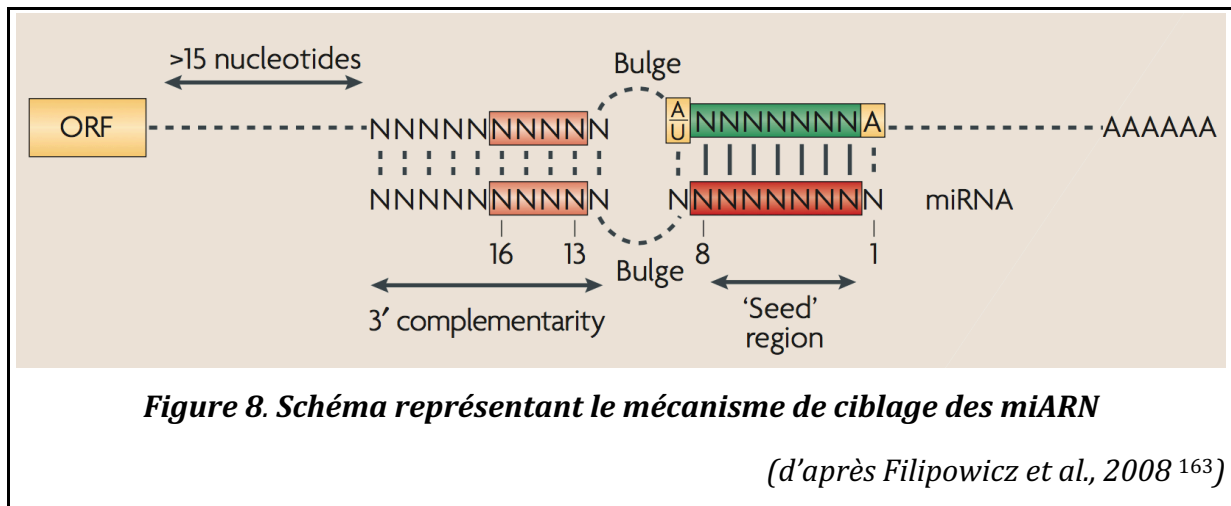
### 7.7.5 Influence de l'énergie libre de liaison du micro-ARN à la cible

L'énergie de liaison de l'hybride miARN-cible influence l'efficacité du site. Les appariements de type wobble (G:U) dans la région seed diminuent l'efficacité du site, en raison de la contribution d'une liaison hydrogène de moins là où une paire de base Watson-Crick canonique (G:C) est possible. Une des premières études a rapporté qu'un appariement de type wobble G:U position 3 abolit le ciblage d'un gène rapporteur luciférase<sup>215</sup>. Chez les nématodes, un appariement de type G:U sur la position 3, ou deux appariements de type G:U sur position 2 et 6 réduit l'efficacité de ciblage d'un gène rapporteur GFP par le miARN *lsy-6* d'environ 20%<sup>239</sup>. De plus, l'énergie libre des miARN dont la région seed est riche en nucléotides AU comme le miARN *lsy-6* ( $\Delta G^\circ = -3.65$  kcal.mol<sup>-1</sup>) ou miR-26 ( $\Delta G^\circ = -5.85$  kcal.mol<sup>-1</sup>) est moins élevée que l'énergie libre médiane des régions seed des miARN communs aux vertébrés ( $\Delta G^\circ = -7.87$  kcal.mol<sup>-1</sup>)<sup>240</sup>. Une raison supplémentaire proposée pour expliquer l'efficacité plus faible qui est observée par rapport à d'autres miARN est un effet de titration dû à la fréquence plus élevée de ce type de sites dans les régions 3'UTR, dont la composition est généralement enrichie en dinucléotides AU<sup>241</sup>.

### 7.7.6 Influence des modifications post-transcriptionnelles de la séquence des miARN

La séquence des miARN peut aussi être modifiée post-transcriptionnellement par editing ou déamination de l'adénosine (A → I). L'editing des miARN est plus fréquent chez l'homme que chez la souris, ainsi que dans le cerveau. L'editing a été observé pour 20 à 30 % des miARN (143 miARN matures), en particulier à la position 6 du miARN où l'on observe en moyenne 23.66% de transitions A → I (= G)<sup>107</sup>. L'editing est opéré au niveau du pri-miRNA par les enzymes ADARs, avec une préférence pour les sites UAG et concerne potentiellement 16% des pri-miARN<sup>242</sup>.





## 7.8 Micro-ARN et physiopathologie de l'organisme

### 7.8.1 Rôle systémique des micro-ARN dans le développement des métazoaires

Chez le nématode *Caenorhabditis elegans*, la délétion du gène *Dicer-1* et donc l'invalidation systémique de la biogénèse des miARN (et de la voie ARN interférence), entraîne des anomalies hétérochroniques et produit des animaux adultes, mais stériles. Certains programmes développementaux ne s'opèrent pas, comme la fusion des cellules de couture (seam cells) pendant la transition du stade larvaire L4 au stade adulte, et l'alaie est absent<sup>243</sup>. De plus, le développement de la lignée germinale des animaux hermaphrodites est anormal, le noyau des oocytes est plus volumineux, il n'y a plus d'auto-fertilisation et la vulve des animaux est éclatée (phénotype burst vulva) ce qui rappelle les mutations dans *let-7*<sup>244</sup>.

La délétion homozygote du gène *Dicer* est létale chez la souris, et entraîne la déplétion des cellules souches embryonnaires, l'absence de marquage OCT4/Brachyury à 7 jours et demi de développement embryonnaire (E7.5) et un arrêt du développement pendant la gastrulation<sup>245</sup>. De plus, l'activité catalytique de AGO2 est requise pour la maturation de miR-451, un miARN indispensable pour l'érythropoïèse dont la structure tige-boucle est maturée d'une manière indépendante de DICER<sup>246</sup>.

Ci-dessous j'illustre deux exemples du rôle physiopathologique des miARN, à savoir durant le développement et dans le cancer. J'ai choisi ces deux exemples en raison de l'abondance de la littérature sur ces sujets.

### 7.8.2 Rôle essentiel de certains micro-ARN dans le développement des mammifères

La délétion individuelle de certains miARN entraîne aussi des anomalies développementales chez les métazoaires<sup>18</sup>. Les phénotypes les plus sévères sont souvent observés après l'invalidation de plusieurs miARN de la même famille, en raison de la redondance de la région seed entre les membres de la famille. Parmi les 90 familles de miARN conservés chez les vertébrés, 69 délétions ont été générées chez la souris dont 11 causent des phénotypes létaux avec pénétrance complète telle que  $\Delta$ miR-1-1/miR-1-2 (létalité post-natale à P10, défauts cardiaques<sup>247,248</sup>),  $\Delta$ miR-9-2/miR-9-3 (létalité post-natale à P7, petits hémisphères cérébraux et bulbes olfactifs, neurogénèse corticale compromise dans le télencéphale<sup>249</sup>),  $\Delta$ miR-17-92 cluster (létalité périnatale, hypoplasie pulmonaire et défauts dans le septum inter-ventriculaire cardiaque<sup>250</sup>; la délétion partielle du cluster soit  $\Delta$ 17,  $\Delta$ 18a et  $\Delta$ 92-1 entraîne uniquement une hypoplasie pulmonaire et cause une létalité périnatale avec pénétrance complète<sup>251</sup>),  $\Delta$ miR-219-1/miR-219-2 (létalité néonatale avec hypomyélinisation du système nerveux central<sup>252</sup>),  $\Delta$ miR-137 (létalité embryonnaire entre E4.5–E11.5<sup>253</sup>), famille  $\Delta$ miR-181 (létalité embryonnaire<sup>254</sup>),  $\Delta$ miR-199a2 (létalité post-natale à P42, atrophie de l'hippocampe et du cortex, similarités avec le syndrome de Rett<sup>255</sup>),  $\Delta$ miR-216a/miR216b/miR-217 cluster (létalité embryonnaire à E9.5<sup>256</sup>),  $\Delta$ miR-218-1/miR-218-2 (létalité néonatale, akinésie, défaut de respiration et absence de jonctions neuromusculaires<sup>257</sup>),  $\Delta$ miR-128-2 (hyperactivité et crises d'épilepsies fatales entre 1 et 2 mois<sup>258</sup>) et  $\Delta$ miR-302a-d cluster (létalité embryonnaire entre E17.5–E18.5, différenciation neuroépithéliale prématurée et échec de la fermeture du tube neural<sup>259</sup>).

Il y a 6 délétions qui causent des phénotypes létaux avec pénétrance incomplète, tels que la famille  $\Delta$ miR-29 (létalité embryonnaire quasi-complète avec une atrophie post-natale et absence de cellules souches hématopoïétiques<sup>260</sup>), la famille  $\Delta$ miR-34/449 (>50% de létalité post-natale à P10, retards de croissance, ciliopathie du cil primaire avec fertilité compromise et défauts dans la clairance muco-ciliaire<sup>261</sup>),  $\Delta$ miR-124-1 (66% de létalité post-natale à P20, maladie débilitante avec apoptose neuronale et atrophie du cerveau<sup>262</sup>),  $\Delta$ miR-133-a1/miR-133-a2 (>50% de létalité post-natale à P10, développement cardiaque anormal<sup>263</sup>),  $\Delta$ miR-126 (50% de létalité embryonnaire, angiogénèse et intégrité vasculaire compromise<sup>264</sup>),  $\Delta$ miR-205 (létalité post-natale à P10, défauts de croissance de l'épiderme et des cellules souches du follicule pileux<sup>265</sup>).

La délétion individuelle de la plupart des miARN restants (52/69) cause des phénotypes viables chez la souris mais avec des perturbations le plus souvent dans la réponse immunitaire, ou des défauts de croissance et différenciation cellulaire dans des tissus spécifiques<sup>18</sup>.

Enfin, il y a pour l'instant peu de délétions qui ont été générées parmi les 96 miARN (88 familles) conservés uniquement chez les mammifères placentaires en raison du nombre moins élevé de cibles prédites et caractérisées, qui est dû à un bruit de fond plus important dans les analyses de conservation. Un des exemples concerne la délétion du cluster miR-290-295 exprimé préférentiellement dans les cellules souches, qui cause 75% de létalité embryonnaire à E8.5–E18.5 avec un échec de la migration des cellules germinales primordiales dans les ovaires, entraînant une stérilité chez les femelles<sup>266</sup>.

Le dosage allélique inadéquat des miARN peut aussi entraîner des pathologies à transmission mendélienne. Chez l'humain, la substitution d'un nucléotide (G>A) en position 4 de la région seed du miR-96 cause un défaut de maturation du pré-miARN et entraîne un syndrome autosomal dominant de surdité progressive<sup>267</sup>. Certains cas du syndrome de Feingold sont causés par la microdélétion germinale hémizygote du cluster miR-17~92, et entraînent une pathologie autosomale dominante caractérisée par une microcéphalie, petite taille et des anomalies digitales. Chez la souris, l'haploinsuffisance du cluster miR-17~92 phénocopie les défauts de croissance du squelette observés chez les patients atteints du syndrome de Feingold<sup>268</sup>.

L'intégrité de la plupart des miARN apparaît donc critique pour le développement des métazoaires. En effet, chez l'humain peu de miARN (10%) contiennent des polymorphismes mononucléotidiques, le plus souvent dans le brin passager ou en dehors de la séquence guide, et il n'y a quasiment aucune variation génétique dans les régions seeds ou la boucle terminale<sup>269</sup>.

### 7.8.3 Dérégulations physiopathologiques de l'expression des micro-ARN

#### 7.8.3.1 Dérégulations de l'expression des micro-ARN dans le cancer

L'inhibition de la biogénèse des miARN ou la dérégulation de l'expression des miARN est une caractéristique des tumeurs<sup>270</sup>.

Certains miARN ont un effet suppresseur de tumeur et sont délétés ou inhibés dans les cellules cancéreuses. Par exemple, l'inhibition de l'expression ou l'haploinsuffisance du cluster miR-15/16 (13q14 locus) est fréquente dans les cas de leucémie lymphoïde chronique<sup>271</sup>. Les locus contenant les miARN sont souvent localisés à des sites chromosomiques fragile et des régions délétées, tels que les loci contenant let-7a-2, let-7g, let-7c ou miR-34a-1 et miR-34a-2 dans le cancer du poumon<sup>272</sup>. Let-7 est l'un des miARN les plus fréquemment inhibés dans les cancers, y compris celui du poumon, en raison du ciblage de l'oncogène RAS<sup>273</sup>. L'inhibition de l'expression des miARN observée dans le cancer peut également résulter de mutations dans un facteur de transcription comme P53<sup>274,275</sup>. La perte de fonction de P53 compromet l'activation transcriptionnelle de ses cibles, comprenant entre autres les miARN de la famille miR-34 qui répriment la prolifération cellulaire<sup>114</sup>.

A l'inverse, certains miARN sont oncogéniques. L'amplification ou l'augmentation de l'expression du cluster miR-17~92 (13q31 locus) est fréquente dans les leucémies lymphoïdes chroniques, et le développement des tumeurs est accéléré avec l'expression concomitante de miR-17 et miR-19b dans un modèle murin de cette leucémie<sup>276</sup>. Des analyses transcriptomiques comparatives entre des biopsies cancéreuses et des tissus sains ont montré des dérégulations récurrentes pour certains cancers. Par exemple, l'augmentation de l'expression de miR-192, miR-194 et miR-215 est fréquente dans le cancer du colon<sup>277</sup>.

La réduction globale de la biogénèse des miARN par des mutations ou l'inhibition du complexe microprocesseur DGCR8/DROSHA, et/ou de DICER est observée dans certains cancers. Cette diminution globale de l'expression des miARN favorise la différenciation des cellules et augmente le potentiel tumorigène<sup>278</sup>.

En effet, on retrouve ce type de mécanisme lors de l'inhibition de let-7 par stratégie antisense pendant la reprogrammation des cellules souches induites pluripotentes avec les facteurs de Yamanaka OSK (OCT4, SOX2, KLF4), qui est équivalente à l'efficacité de

reprogrammation du cocktail complet OSKM (OCT4, SOX2, KLF4, MYC)<sup>279</sup>. Dans le cancer des ovaires, de plus faibles niveaux de la protéine DROSHA ou DICER sont associés à un mauvais pronostic<sup>280</sup> et on observe des hotspots mutationnels dans le domaine RNase IIIb de la protéine Dicer (au niveau des résidus impliqués dans la coordination d'ions métalliques) qui affectent l'activité de clivage et de maturation des pré-miARN<sup>281</sup>. Dans les tumeurs de Wilms, il est fréquent d'observer des points chauds mutationnels alléliques dans la protéine Drosha au niveau des domaines RNase III (59/534 cas, 11%), ou des délétions hémizygotiques de la protéine DGCR8 (20/534 cas, 4%) avec des points chauds mutationnels dans le domaine de liaison au pré-miARN de l'allèle restant<sup>282</sup>.

#### **7.8.4 Stratégies thérapeutiques visant à rétablir l'expression normale des micro-ARN**

L'administration systémique de miARN ou d'inhibiteurs de miARN visant à restaurer l'état physiologique représente une stratégie thérapeutique contre certains cancers<sup>283</sup>. Les inhibiteurs antisens de miR-122 (par acide nucléique bloqué 2'-4'; ou antisens Galactose-N-Acétyl conjugué) ont été les premières stratégies thérapeutiques à entrer en essais cliniques de Phase I et Phase II pour le traitement de l'infection par le virus de l'hépatite C (VHC), qui peut entraîner un cancer du foie<sup>284,285</sup>. Il s'agit toutefois d'un cas particulier, en effet le VHC est stabilisé par la liaison de miR-122 et AGO2 à l'extrémité 5' de l'ARN viral, qui n'est pas coiffé et susceptible à la dégradation 5' -> 3'<sup>286</sup>.

Un autre essai clinique récent pour le traitement de plusieurs tumeurs solides comme le carcinome hépatocellulaire, le cancer du pancréas ou le cholangiocarcinome consiste en l'injection systémique dans la circulation sanguine du miR-34a administré par stratégie liposomale. Bien que certains patients aient montré une réponse prolongée au traitement, l'essai clinique a été interrompu suite à plusieurs anaphylaxies et réponses immunitaires sévères adverses au traitement, ce qui soulève des interrogations quant au type de liposome utilisé<sup>287</sup>. D'autres stratégies d'administration reposent sur des méthodes de ciblage des liposomes encapsulant l'acide nucléique, comme les vésicules lipidiques sécrétées d'origine bactérienne, recouvertes d'anticorps ciblant le récepteur EGFR et utilisées pour délivrer le miR-15/16 dans un essai clinique pour le traitement du cancer du foie (mesomiR-1 ou TargomiRs)<sup>288,289</sup>. Deux autres traitements basés sur des inhibiteurs de miARN sont en essai clinique de Phase I, comme l'antimiR-155 pour le

traitement du lymphome T cutané et l'antimiR-103/107 pour le traitement de la stéatohépatite non-alcoolique (NASH) aussi appelée maladie du soda. Il existe également une stratégie basée sur l'administration du miR-29 dans la sclérodermie systémique associée à certains cancers<sup>283</sup>.

## 7.9 Diversité de la localisation des micro-ARN

### 7.9.1 Micro-ARN et réticulum endoplasmique

Les études biochimiques initiales du complexe RISC ont montré une association préférentielle aux membranes intracellulaires<sup>i</sup> y compris le réticulum endoplasmique rugueux (REr, fraction microsomale)<sup>102,148,152</sup>. Ces premières observations ont contribué à l'étude des mécanismes d'action du complexe RISC principalement dans le cytoplasme<sup>ii</sup>.

En effet, dans le cytoplasme la protéine AGO2 et les miARN semblent co-sédimer quasiment exclusivement avec la membrane externe du REr (marqueur calnexine) dans des gradients de sucrose<sup>290</sup>. Chez *Arabidopsis thaliana*, l'ancrage de AGO1 au REr serait induit par la protéine Amp1<sup>291</sup>.

Au niveau fonctionnel, une étude chez les mammifères a rapporté que le ciblage par let-7 sur un gène rapporteur contenant trois sites de liaison à let-7 entraîne une redistribution des niveaux de transcrits restants et une rétention préférentielle au niveau des polysomes du REr<sup>292</sup>.

L'ensemble de ces observations suggèrent qu'une partie de l'inhibition de la traduction induite par le complexe RISC intervient au niveau du réticulum endoplasmique, et qu'une fraction importante de AGO2 et des miARN est localisée au niveau du REr.

### 7.9.2 Micro-ARN dans les exosomes sécrétés

La présence des miARN dans les exosomes a été observée pour la première fois dans des exosomes sécrétés de mastocytes de foie chez la souris<sup>293</sup>. Des populations de miARN semblent sélectivement enrichies dans les exosomes par rapport à leur concentration intracellulaire, et il a été proposé que le mécanisme d'adressage repose sur une molécule cargo comme les ARN circulaires<sup>294</sup>.

Les petits ARN non-codants sont stables dans la circulation sanguine et peuvent être utilisés comme biomarqueurs de certaines pathologies. Par exemple, le taux de miR-141 dans le sérum est plus élevé chez les patients atteints de cancer de la prostate au stade métastatique que chez les individus sains<sup>295</sup>. Certains miARN exprimés par le placenta comme miR-141 sont détectés dans le plasma maternel durant le troisième trimestre de

---

<sup>i</sup> Voir section 7.6.3.1. Structure et fonction des protéines Argonaute.

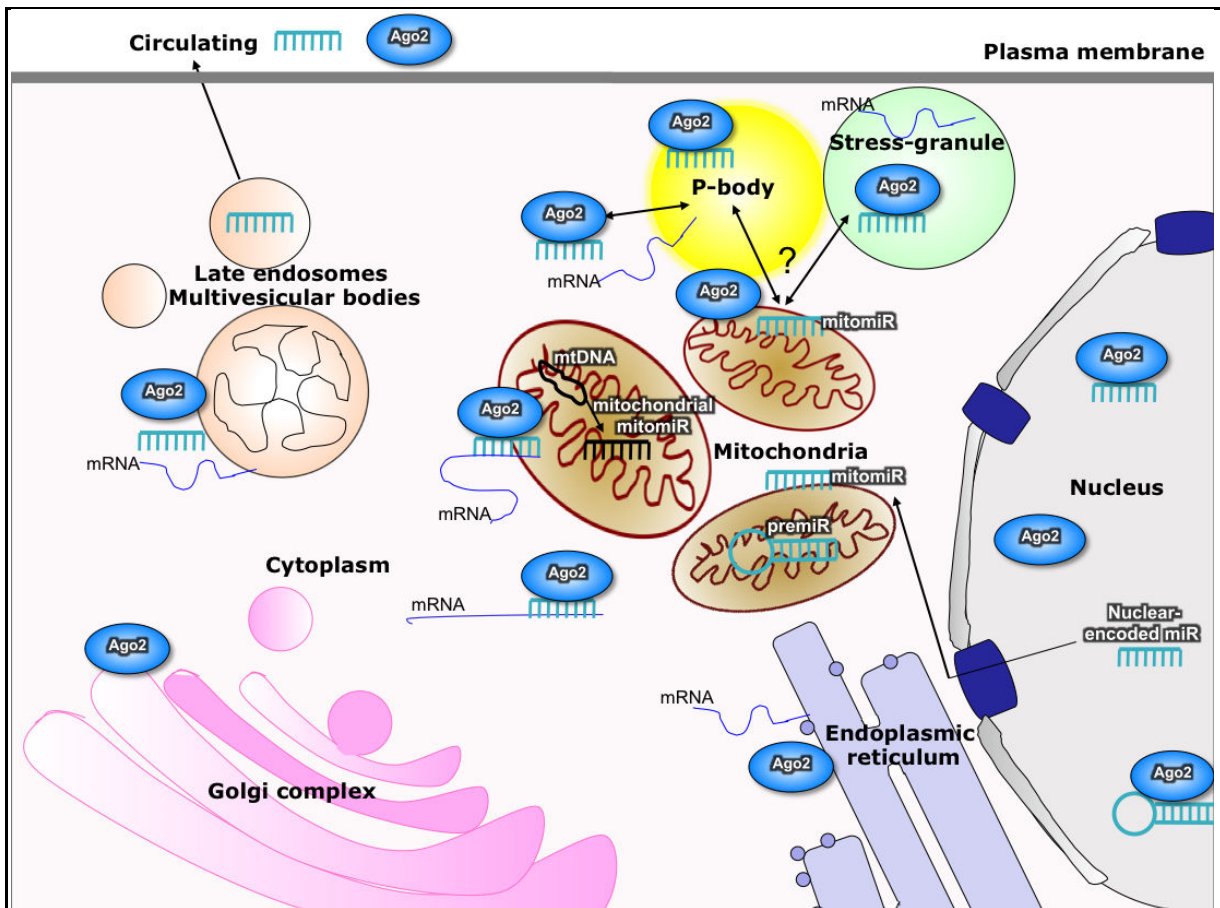
<sup>ii</sup> Et une vigilance quant à l'étude des fonctions nucléaires des miARN.

la grossesse et disparaissent après l'accouchement<sup>296</sup>. Au niveau physiologique, les exosomes qui sont sécrétés par une lignée cellulaire de cancer mammaire au stade métastatique sont capables d'induire une inhibition de la zonula occludens de cellules épithéliales réciendaires, ce qui suggère que le contenu en protéines et ARN des exosomes serait assez conséquent pour réguler un processus physiologique<sup>297</sup>.

### 7.9.3 Micro-ARN dans les mitochondries

Les miARN ont également été détectés dans les mitochondries (**Figure 9**) de foie adulte de rat<sup>298</sup>, de souris<sup>299</sup> et dans les mitochondries humaines<sup>300</sup>. Certains miARN semblent absents de la mitochondrie, indépendamment de leur abondance cellulaire. Par exemple le mir-122 n'est pas enrichi dans la mitochondrie du foie bien qu'il soit le miARN le plus exprimé, ce qui suggère un mécanisme d'import sélectif des miARN à la mitochondrie<sup>299</sup>. Dans les cardiomyocytes humains, la protéine AGO2 est présente dans les mitochondries alors que les protéines DICER ou TRBP sont absentes<sup>301</sup>. La localisation des miARN dans la mitochondrie est dynamique, les taux mitochondriaux de miR-1 augmentent pendant la différenciation des myoblastes humains en myocytes, et il semble que certains transcrits mitochondriaux tels que ND1 et MT-CO1 (COX1) soient stabilisés suite au ciblage par miR-1 et AGO2, ce qui s'expliquerait par l'absence des protéines effecteurs du complexe RISC à la mitochondrie comme TNRC6<sup>302</sup>.





**Figure 9. Représentation de la localisation cellulaire et extra-cellulaire des miARN**

*Dynamique de localisation intracellulaire de la protéine AGO2 et des miARN matures. Après l'étape de maturation du pré-miARN dans le cytoplasme, les miARN matures et leurs cibles localisent aux membranes intracellulaires telles que le complexe de golgi ou le réticulum endoplasmique. Les miARN localisent également aux granules cytoplasmiques appelées processing bodies (P-bodies) ou les granules induits par le stress (Stress-granules). Plus récemment les miARN ont été détectés à la mitochondrie ou dans les fluides corporels comme la circulation sanguine.*

*De plus, les miARN matures sont réimportés dans le noyau où ils régulent l'activité de transcrits nucléaires.*

*(d'après Bandiera et al., 2014<sup>303</sup>)*

## 7.10 Micro-ARN nucléaires

Les miARN ont été détectés dans le noyau des cellules animales<sup>156,304,305</sup>. Accessoirement, l'ARN interférence peut être déclenchée dans le noyau sur la base de l'activité catalytique de AGO2<sup>306</sup>.

La plupart des miARN cytoplasmiques sont aussi présents dans le noyau<sup>307</sup> et semblent être importés de manière non-sélective. Cependant certains miARN comme miR-29b apparaissent enrichis dans le noyau<sup>308</sup>, bien qu'un mécanisme d'import sélectif reste à déterminer. Il pourrait reposer sur la co-localisation préférentielle du miARN avec ses cibles et sites de liaison dans le noyau.

Certaines études ont décrit la présence voir l'enrichissement nucléolaire de certains miARN comme miR-206<sup>309,310</sup>. Il est aussi possible d'observer une séquestration nucléolaire de protéine fusion AGO2 dans certaines lignées cellulaires (HeLaS3)<sup>311</sup>.

L'identification des micro-ARN nucléaires a soulevé un intérêt pour les fonctions régulatrices qu'ils pourraient exercer sur les populations d'ARN non-codants régulateurs.

### 7.10.1 Transport nucléaire des composants de l'ARN interférence et des micro-ARN

L'import nucléaire de AGO2 repose principalement sur l'importine-8<sup>312</sup> tandis que le transport des protéines GW182 (TNRC6) repose sur l'importine- $\beta$ <sup>185</sup>. L'essentiel des interactions constituant le complexe RISC est préservé dans le noyau, telles que les protéines AGO2, les protéines TNRC6, le complexe CCR4-NOT1, TRBP, DICER et HSP90<sup>313,314</sup>.

## 7.10.2 Fonctions nucléaires des micro-ARN chez les mammifères

### 7.10.2.1 Fonctions transcriptionnelles des siARN chez les mammifères

La conservation de mécanismes épigénétiques comme ceux existant chez les plantes (RdDM) et la levure (RITS)<sup>i</sup> a été examinée chez les mammifères. Il a d'abord été observé des événements de méthylation du promoteur en réponse à des siARN ectopiques complémentaires au promoteur introduits dans des cellules humaines, en se basant sur des expériences de PCR du promoteur avec enzymes de restriction sensibles à la méthylation<sup>315</sup>. Dans un autre cas, il a été observé le dépôt de la marque épigénétique H3K9me2 sur un promoteur après introduction de siARN de séquence complémentaire, d'une manière qui dépend de la protéine AGO1<sup>316</sup>.

A l'opposé, dans d'autres études on observe une activation transcriptionnelle des promoteurs ciblés par des siARN, qui s'accompagne d'une déméthylation des histones (H3K9me)<sup>317</sup>, ou la réduction de l'acétylation de la lysine 9 de l'histone 3 (H3K9Ac) ou de la lysine 14 de l'histone 3 (H3K14Ac) avec augmentation de la méthylation de la lysine 4 (H3K4me2 et H3K4me3)<sup>318</sup>.

Il a été proposé que les mécanismes de l'inactivation transcriptionnelle (TGS) ou de l'activation transcriptionnelle (TGA) chez les mammifères reposent sur la présence de transcrits antisens des promoteurs ciblés<sup>319,320</sup>, et dépend de la nature du transcrit qui dans certains contextes agirait localement comme inhibiteur ou activateur de la transcription.

Dans une autre étude, un siARN ciblant le promoteur du gène MYC inhibe la transcription en entraînant la diminution du recrutement de l'ARN Pol II au gène cible, mais ne semble pas dépendre des marques épigénétiques H3K9me2, H3K27me3, l'acétylation de l'histone 3, ni de la méthylation de l'ADN<sup>321</sup>.

La dégradation d'un transcrit sens associé au promoteur (paARN) semble requise pour l'inhibition de la transcription du gène proximal, et repose principalement sur la protéine AGO2. Les niveaux du paARN diminuent suite au ciblage par le siARN, sans toutefois nécessiter le clivage par AGO2. En effet, l'utilisation d'un siARN avec trois

---

<sup>i</sup> Voir sections 7.11.2 & 7.11.3.

nucléotides centraux mésappariés entraîne également la diminution des niveaux du paRNA<sup>321</sup>. Le rôle majeur de la protéine AGO2<sup>i</sup> dans le mécanisme d'inhibition transcriptionnelle a également été observé dans d'autres contextes comme aux gènes codants les récepteurs à la progestérone<sup>323</sup>.

### 7.10.2.2 Fonctions transcriptionnelles des micro-ARN chez les mammifères

Les propriétés d'activation ou d'inhibition transcriptionnelle attribuées aux siARN ont aussi été observées pour certains miARN endogènes.

Chez l'humain, quelques miARN comme miR-320 sont complémentaires à des régions en amont (-200/-1) de certains promoteurs comme POLR3D, et semblent induire le recrutement de la protéine AGO1, du complexe répresseur polycomb (EZH2) et la triméthylation de la lysine 27 de l'histone 3 (H3K27me3)<sup>324</sup>.

Dans la lignée cellulaire d'adénocarcinome pulmonaire A549, des transcrits bidirectionnels chevauchant la région promoteur en amont du gène MT-CO2 (COX2) sont détectés par RACE-PCR, dont certains contiennent deux sites de liaison pour le miR-589 entre -10 et -60 bases. Le ciblage par miR-589 de certains de ces transcrits à faible nombre de copies induit l'activation du gène MT-CO2, d'une manière qui semble indépendante de l'activité de clivage de AGO2<sup>325</sup>.

Dans des fibroblastes sénescents humains, AGO2 semble s'accumuler dans le noyau et induire la répression transcriptionnelle de gènes cibles de E2F comme CDK1 (cdc2) et CDCA8 qui possèdent des sites de liaison à let-7f en amont (-400/-100) du promoteur. De plus, AGO2 co-immunoprécipite préférentiellement avec des marques épigénétiques répressives comme H3K9me2 et H3K27me3<sup>326</sup>.

Dans d'autres cas, on retrouve une activation transcriptionnelle des promoteurs par des miARN, comme dans le cas du promoteur du gène codant CDH1 qui possède un site de liaison (-645/-621) de type 8-mer au miR-373, ou bien le gène codant CSDC2 qui possède un site de liaison quasi-parfait au miR-373<sup>327</sup>.

Un autre exemple concerne le promoteur du gène codant l'interleukine-24 qui contiendrait un site de liaison (-127/-107) de type (3-9)-mer au miR-205, et s'accompagne de l'acétylation des histones H3 et H4<sup>328</sup>.

---

<sup>i</sup> La protéine de la famille Argonaute la plus exprimée, voir Dongmei et al., 2012, réf <sup>322</sup>.

### 7.10.2.3 Fonctions dans la modulation de l'épissage alternatif

L'épissage alternatif des gènes est modulé par les marques épigénétiques locales des histones<sup>329</sup>. Dans les cellules humaines, le ciblage de séquences introniques par des siARN est accompagné par le dépôt de la méthylation de la lysine 27 de l'histone 3 (H3K27me3) et la méthylation de la lysine 9 de l'histone 3 (H3K9me2). Ces marques épigénétiques ralentissent la progression locale de l'ARN polymérase, et favorisent l'inclusion d'exons alternatifs (en particulier des sites d'épissage faibles) d'une manière AGO1-dépendante<sup>330</sup>. La protéine AGO2 est retrouvée sur la chromatine et co-immunoprécipite avec une population de petits ARN de 19-30 nucléotides, ainsi que des facteurs d'épissage et de remodelage de la chromatine. Les petits ARN associés à AGO2 correspondent principalement à des séquences introniques. Ils sont d'orientation colinéaire au sens de la transcription et localisés environ -20 à -40 bases en amont du site intronique accepteur d'épissage 3'. La déplétion de AGO1 ou AGO2 empêche la déposition locale de la marque H3K9me3 et compromet l'inclusion de l'exon variant du gène CD44<sup>311</sup>.

### 7.10.3 Activité de dégradation induite par les micro-ARN nucléaires

L'activité de clivage de AGO2 peut être déclenchée par certains miARN comme le miR-671, qui est préférentiellement nucléaire et clive l'ARN circulaire CDR1-AS qui contient un site complémentaire parfait<sup>331</sup>.

De plus, quelques études ont rapporté une activité de dégradation de l'ARNm induite par les miARN dans le noyau indépendamment du clivage par AGO2.

Dans une lignée cellulaire de lymphome de Hodgkin, le miR-9 semble cibler et induire la dégradation de l'ARN non-codant MALAT1 qui contient deux sites de liaison à miR-9, dont un site de type 6-mer décalé imparfait<sup>332</sup>.

Chez la souris, le miR-709 est localisé préférentiellement dans le noyau et induit la diminution de l'expression des miR-15/16 à travers un site de liaison dans le pri-miR-15/16<sup>333</sup>.

Cependant, le mécanisme de dégradation induit par les miARN nucléaires est inconnu.

## 7.11 Fonctions nucléaires d'autres petits ARN

Les protéines de la famille Piwi/Argonaute s'associent à d'autres petits ARN non-codants de biogénèse distincte des miARN. Le complexe ainsi formé est impliqué dans d'autres processus de régulations de l'expression des gènes, en particulier au niveau nucléaire. Dans ce chapitre, je décris la fonction et le mécanisme d'action de ces autres petits ARN, en contraste avec le mécanisme d'action des miARN et du complexe RISC dans le noyau chez les mammifères.

### 7.11.1 Elimination programmée de l'ADN chez les protozoaires ciliés

Chez le protozoaire cilié *Tetrahymena thermophila*, le gène *Tw1* de la famille Piwi est requis pour l'élimination programmée des séquences d'ADN éliminées internes (IES) du nouveau macronoyau formé pendant la conjugaison<sup>334</sup>. Pendant la conjugaison, le génome du micronoyau est transcrit dans les deux directions, et produit des ARNdb. Ces ARNdb sont maturés par l'enzyme Dicer-1 en petits ARN non-codants de 28 nucléotides appelés scanRNAs. Les scanRNAs sont ensuite incorporés dans la protéine *Tw1*<sup>335</sup>. Selon le modèle proposé, les scanRNAs sont d'abord filtrés sur l'ancien macronoyau avec une dégradation des scanRNAs complémentaires, produisant un tamis restant de scanRNAs qui programme les réarrangements génomiques du nouveau macronoyau<sup>336</sup>. Pendant ce processus, on observe la déposition transitoire d'une marque épigénétique, la di-méthylation de l'histone 3 à la lysine 9 (H3K9me2) aux séquences d'ADN éliminées internes environ trois heures avant le réarrangement génomique programmé. La marque H3K9me2 co-localise avec la protéine *Pdd1* qui possède un chromodomaine de liaison à la marque H3K9me2<sup>337</sup>. Dans les cellules  $\Delta Pdd1$ , on observe un délai de plusieurs heures dans l'accumulation des scanRNAs, une absence de déposition de la marque H3K9me2 et un échec de l'élimination programmée de l'ADN.

### 7.11.2 Méthylation de l'ADN dirigée par l'ARN chez les plantes

Chez les plantes, les protéines Argonautes guident la déposition de certaines marques épigénétiques comme le groupement 5'-methyl sur les cytosines (**Figure 10**), selon un mécanisme connu sous le nom de Méthylation de l'ADN dirigée par l'ARN (RdDM)<sup>338</sup>.

Dans les plantes à tabac, certaines copies artificiellement intégrées de l'ADN complémentaire du viroïde du tubercule de pomme de terre peuvent être méthylées,

d'après l'analyse southern blot de l'ADN avec des enzymes de restriction sensibles à la méthylation. Cependant, cette méthylation est observée seulement après que la réplication autonome du viroïde a commencé, par la production d'ARN viral<sup>i</sup>.

Dans les plants transgéniques où aucune méthylation des copies intégrées au génome n'est observée, il est possible de déclencher leur méthylation *de novo*, en infectant la plante avec l'ARN du viroïde<sup>339</sup>.

Dans un autre exemple, le phénomène de co-suppression (par méthylation) des promoteurs génomiques de type NOS est observable dans certaines plantes transformées, mais seulement dans le cas où l'ADN recombinant produisant des transcrits de la séquence promoteur NOS s'est inséré en double exemplaire dans une configuration tandem inversé<sup>340</sup>.

Cette conformation particulière des transgènes résulte en la transcription d'ARN double-brin, et la génération de petits ARN non-codants<sup>341</sup>. Ce type d'inactivation transcriptionnelle est transmissible à la descendance, et apparaît stable même après l'excision du transgène<sup>340</sup>.

Le mécanisme d'inactivation transcriptionnelle repose sur les composants de l'ARN interférence tels que l'ARN polymérase ARN-dépendante de type 2 (Rdr2<sup>342</sup>), un gène de type Dicer-3 (Dcl-3<sup>342</sup>), un gène appelé Sde-3, le gène Argonaute-4<sup>343</sup>, les méthyltransférases Drm2/Drm1 (famille DNA méthyltransférase 3, DNMT3) pour la déposition *de novo* de la méthylation<sup>344</sup> et la méthyltransférase Met-1 pour la conservation et la transmission de la méthylation aux cellules filles<sup>345</sup>.

Chez la plante *Arabidopsis Thaliana*, les mutations dans le gène *Argonaute-4* suppriment la déposition de la marque épigénétique H3K9me2 sur certains gènes comme *Sup* et certains éléments transposables comme *AtSN1*<sup>346</sup>. La marque H3K9me2 déclenche normalement le recrutement de l'ARN polymérase de type IV (spécifique aux plantes) par la fixation de la protéine *Shh1* à la marque H3K9me2<sup>347</sup>. L'ARN polymérase de type IV est recrutée sur des régions génomiques correspondant à des transposons, des régions répétées ou des transgènes, et produit des transcrits primaires<sup>348</sup>.

Les transcrits primaires sont convertis en transcrits double-brin par l'ARN polymérase ARN-dépendante *Rdr2*, puis maturés en siARN par *Dcl-3* mais pas *Dcl-2* (spécialisée dans la maturation des siARN dérivés de virus) ni *Dcl-1* (spécialisée dans la maturation

---

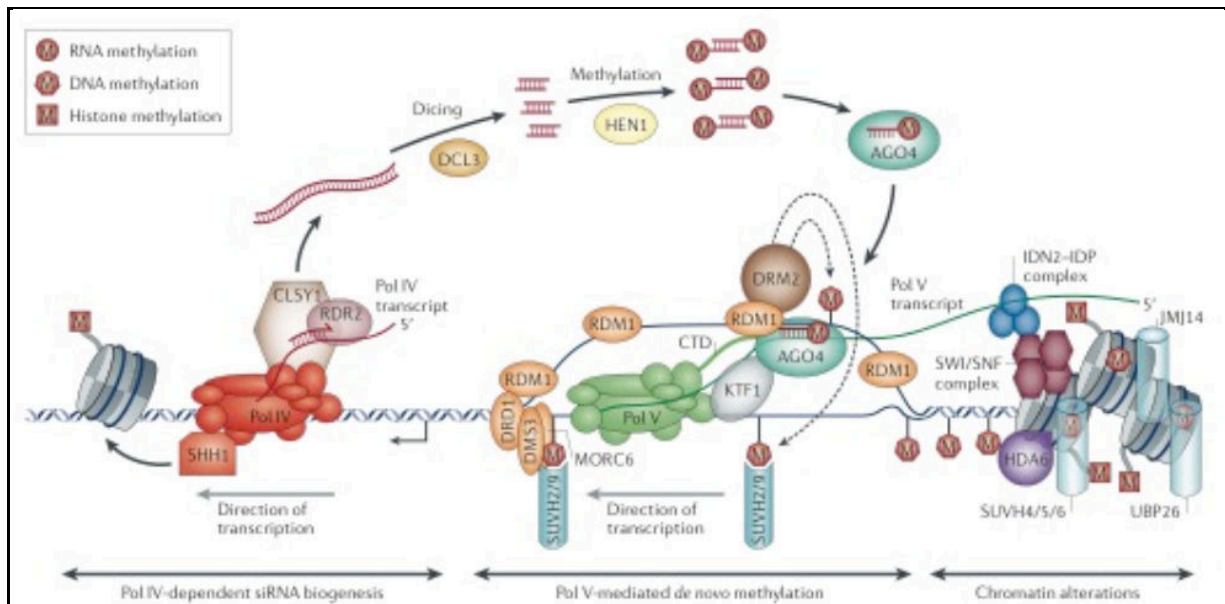
<sup>i</sup> La réplication de l'ARN de ce viroïde passe par une amplification circulaire de l'ARN.

des miARN)<sup>342</sup>, et sont 2'-O-méthylés à l'extrémité 3' par l'enzyme Hen1<sup>349</sup>. De plus, l'ARN polymérase de type V (ARN Pol V, spécifique aux plantes) est nécessaire pour l'accumulation des siARN, ainsi que le mécanisme de méthylation<sup>350</sup> qui opère en aval de l'ARN polymérase de type IV<sup>351</sup>.

En effet, la protéine AGO4 co-localise avec l'ARN Pol V mais pas l'ARN polymérase IV, et semble se lier au domaine C-terminal de la sous-unité Nrp1b (Nrpe) de l'ARN Pol V<sup>352</sup>, qui contient des motifs GW/WG de type Ago-hook<sup>353</sup>. De plus, AGO4 co-immunoprécipite avec KTF1, qui pourrait également fixer directement AGO4 par l'intermédiaire de motifs GW/WG de type Ago-hook<sup>351</sup>.

Un des mécanismes de méthylation proposé est le recrutement de la méthyltransférase Drm2 par les complexes AGO4-siARN sur les transcrits naissants synthétisés par l'ARN Pol V<sup>354</sup>.





**Figure 10. Schéma explicatif de la voie de méthylation de l'ADN dirigée par l'ARN**

Représentation du mécanisme de méthylation de l'ADN dirigée par l'ARN. De gauche à droite : certains loci portant la marque épigénétique H3K9me sont transcrits en précurseurs simple-brin par l'ARN Pol IV qui est recrutée par la protéine Shh1 interagissant avec l'histone méthylé. Il y a synthèse du brin complémentaire par l'ARN polymérase ARN-dépendante Rdr2. L'ARN double-brin est mûré en petits ARN interférents par l'enzyme Dicer-like 3 (Dcl-3), stabilisés par 2'-O-Méthylation à l'extrémité 3' par l'enzyme Hen1 et incorporés dans une protéine Argonaute-4 (AGO4). Les complexes AGO4-siARN interagissent avec (i) des transcrits naissants synthétisés par l'ARN Pol V, (ii) le domaine carboxy-terminal de l'ARN Pol V via un motif Ago-hook ou (iii) la protéine de liaison KTF1 par l'intermédiaire d'un motif Ago-hook. AGO4 recrute l'ADN méthyltransférase Drm2 par l'intermédiaire de la protéine Rdm1 et induit la déposition de groupements 5'-méthyl sur les cytosines. Le recrutement de l'ARN polymérase V est facilité par la présence locale initiale de marques épigénétiques répressives sur le site, et flanquant le site (chromatin alterations).

(d'après Matzke et al., 2014<sup>354</sup>)

### 7.11.3 Formation de l'hétérochromatine chez la levure

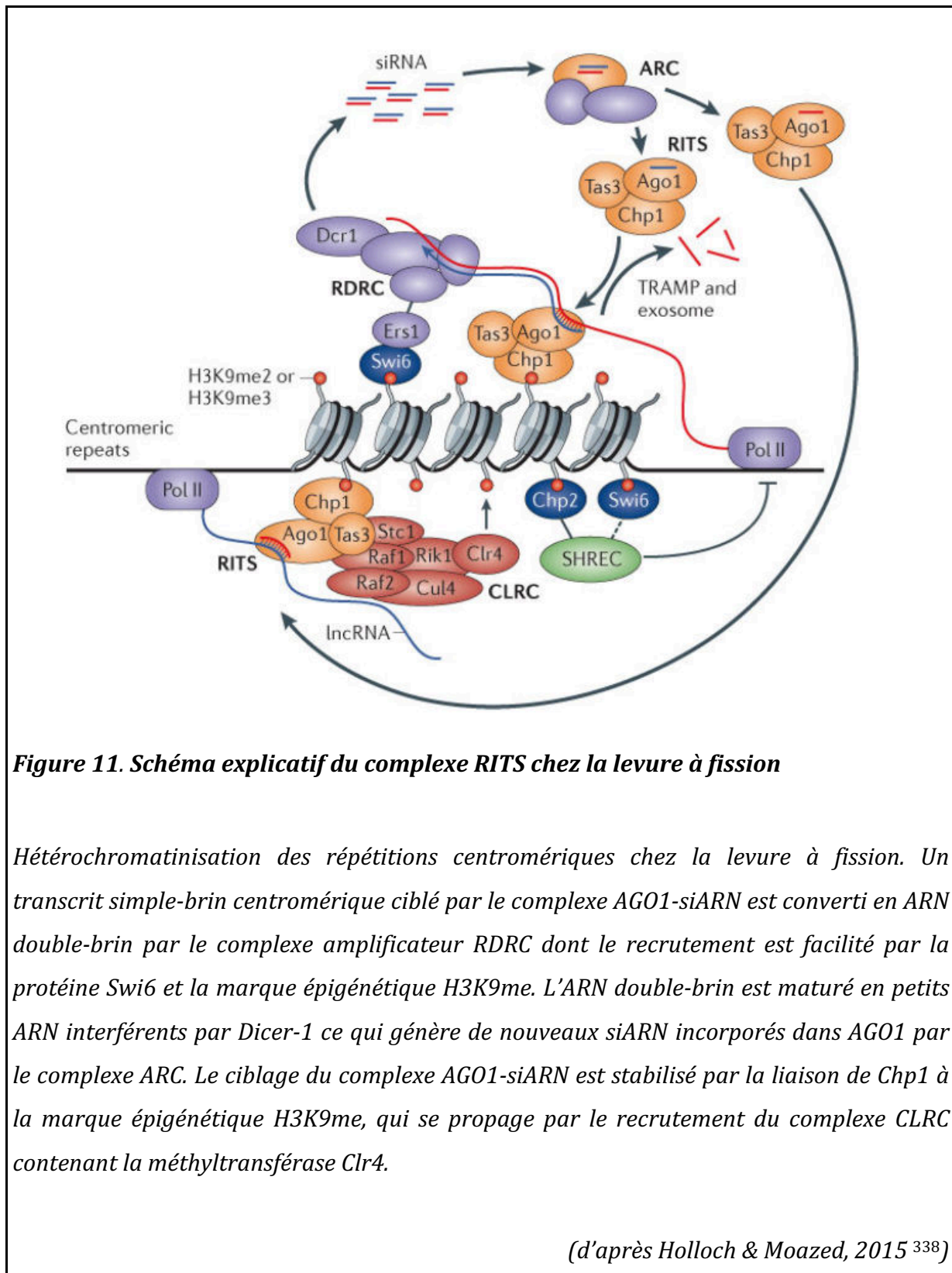
Chez la levure à fission *Saccharomyces Pombe*, les composants de l'ARN interférence AGO1, Dicer-1 (Dcr1) et une ARN polymérase ARN-dépendante (Rdrp1) sont impliqués dans la formation de l'hétérochromatine aux répétitions d'homologie aux centromères (CenH, répétitions externes) sur le locus de conjugaison définissant le type (mating-type locus)<sup>355</sup>. En l'absence des gènes AGO1, Dcr1 ou Rdrp1, on observe l'accumulation de transcrits de taille 1.4-2.4 kb correspondant aux répétitions centromériques, cependant un transgène inséré dans la partie centrale des répétitions reste silencieux<sup>356</sup>. La protéine chromodomaine pombe 1 (Chp1<sup>357</sup>), la protéine AGO1 et la sous-unité du complexe de ciblage 3 (Tas3) forment un complexe protéique appelé RNA-induced transcriptional silencing complex (RITS) qui co-immunoprécipite avec des siARN correspondant aux répétitions centromériques<sup>358</sup>. Le domaine piwi de AGO1 fixe directement la protéine Tas3 par l'intermédiaire du motif Ago-hook<sup>165</sup>, et la fonction catalytique de AGO1 est requise pour le clivage du brin passager et l'activité normale du complexe RITS<sup>359</sup>. La localisation centromérique et l'accumulation des siARN dans le complexe RITS reposent sur la maturation des transcrits centromériques primaires par les gènes Rdrp1, Hélicase requise pour l'ARNi (Hrr1) et une poly(A) polymérase homologue à la famille Trf4/5 (Cid12), qui forment un complexe amplificateur de l'ARN dirigé par l'ARN (RDRC)<sup>360</sup>. Les siARN sont transférés dans le complexe RITS par l'intermédiaire du complexe ARC composé des protéines Arb1 et Arb2<sup>361,362</sup>.

Le processus d'hétérochromatinisation peut être récapitulé par l'ancrage de la protéine Tas3 aux transcrits naissants via le système  $\lambda$ N-peptide and BoxB, ce qui entraîne une production de siARN et la méthylation de la lysine 9 de l'histone 3 (H3K9me) d'une manière RITS-dépendante, RDRC-dépendante, Swi6-dépendante et Clr4-dépendante<sup>363</sup>. Le modèle proposé est le ciblage co-transcriptionnel du complexe RITS sur le transcrit centromérique naissant (**Figure 11**), d'une manière qui semble facilitée par la présence de la marque épigénétique H3K9me qui stabiliserait l'ancrage du complexe à la chromatine<sup>364</sup>. L'hétérochromatinisation apparaît partiellement restreinte en cis<sup>365</sup>, en raison d'une limitation de la diffusion des siARN par une exoribonucléase appelé Eri-1<sup>363</sup>.

Bien que le processus d'hétérochromatinisation induite par le complexe RITS réduise fortement le taux de transcription, il existe un niveau résiduel de transcription naissante (et de production de siARN), qui est dégradée principalement par l'exosome nucléaire

d'une manière Rrp6-dépendante, Dis3-dépendante et Cid-14-dépendante (une polyadénylate polymérase homologue à la famille Trf4)<sup>366</sup>. Le gène Cid14 présente une homologie avec le gène rde-3 chez le nématode, et fait partie de la superfamille des nucléotidyltransférases polymérase  $\beta$ <sup>367</sup>.

La délétion du gène Cid14 entraîne la quasi-disparition des siARN dans le complexe RITS, néanmoins un faible dosage des siARN semble suffisamment efficace pour déposer la marque épigénétique H3K9me2 et induire l'hétérochromatinisation<sup>366</sup>. En fait, dans les souches  $\Delta$ Cid14, le système de surveillance de l'ARN ribosomal (ARNr) par le complexe TRAMP semble compromis, et l'ARNr défectueux entre à la place dans la voie de l'ARN interférence d'une manière RDRC-dépendante, ce qui déplace l'équilibre du complexe RITS vers la saturation par des siARN antisens de l'ARNr<sup>368</sup>.



#### 7.11.4 Régulation des éléments transposables par les protéines Piwi

##### 7.11.4.1 Découverte des rasi-ARN et protéines associées piwi dans les cellules germinales

Chez la drosophile, les protéines *Piwi* sont requises pour l'intégrité de la lignée germinale<sup>369</sup>.

En effet, il a été initialement identifié que la délétion de certains composants de l'ARN interférence comme *sting* (aubergine, une protéine de la famille piwi) ou *homeless* (*spindle-E*, une ARN hélicase à boîte DE-H) entraîne la stérilité des males. Ce phénotype résulte de la formation de cristaux de protéine *Stellate* dans les spermatocytes<sup>370,371</sup>, qui s'accumulent suite à la dérégulation du locus *Stellate*. De plus, les mutations dans le gène *homeless* entraînent la dérégulation des éléments transposables et des séquences répétées en tandem dans la lignée germinale<sup>372</sup>.

En temps normal, le locus *Stellate* est réprimé par le locus suppresseur de *Stellate* *Su(ste)*<sup>i</sup>. Le locus *Su(ste)* est transcrit dans les deux directions en raison de l'insertion antisens d'un transposon *Hoppel*, ce qui entraîne la production de siARN de 24-29 nucléotides. Ces siARN sont impliqués dans la suppression de l'expression du locus *Stellate* d'une manière *sting*-dépendante et *homeless*-dépendante<sup>372</sup>.

Cette nouvelle classe de siARN de 24-29 nucléotides dérivés de régions répétées a été initialement appelée rasiARN. La biogénèse des rasiARN est indépendante de la voie ARN interférence canonique (*Dcr-1*, *Dcr-2* ou *AGO2*). Les rasiARN correspondent essentiellement aux petits ARN que nous appelons aujourd'hui piwi-ARN.

De nombreuses études ont confirmé le rôle des rasiARN et du système piwi dans le ciblage et la répression des éléments transposables et répétés.

En effet, les rasiARN sont principalement d'orientation antisens aux régions répétées et s'associent avec les protéines *Piwi* et *Aubergine*<sup>373</sup>. Les mutations dans les gènes *Piwi*, *sting* et *homeless* suppriment la variegation des éléments-P<sup>ii</sup> et induisent une perte de la marque épigénétique répressive H3K9me2 aux éléments-P<sup>374</sup>. Les mutations dans

---

<sup>i</sup> Le locus *Stellate* est lié à l'Y tandis que le locus *Su(Ste)* est lié à l'X.

<sup>ii</sup> Transposons de type II chez la drosophile

homeless induisent aussi une délocalisation de la protéine hétérochromatique 1 (HP1) depuis les répétitions centromériques sur l'ensemble des chromosomes polytènes<sup>375</sup>.

#### 7.11.4.2 Piwi-ARN et protéines associées dans les cellules germinales

Après la découverte des rasiARN, une nouvelle dénomination en piwi-ARN (piARN, l'équivalent des rasiARN) a été reprise pour définir cette nouvelle classe de petits ARN non-codants et le système piwi a été caractérisé plus en détails.

Le système piwi est composé de trois protéines piwi (Piwi, et Aubergine, Ago3) et deux voies d'approvisionnement des piARN. L'une des voies alimente le système Piwi et l'autre alimente le système Aubergine, Ago3. De plus, le système Piwi, Aubergine, Ago3 communique physiquement.

La voie de biogénèse primaire et l'accumulation des piARN dans les protéines Piwi est Armitage-dépendante et Zucchini-dépendante<sup>i</sup>. De plus Zucchini co-immunoprécipite avec Aubergine<sup>376</sup>.

De l'autre côté, on trouve la voie d'amplification ping-pong (**Figure 12**) qui est Aubergine-dépendante, krimper-dépendante, spindle-E-dépendante, vasa-dépendante<sup>377</sup> et Ago3-dépendante<sup>378</sup>.

Les protéines Piwi et Aubergine sont principalement associées à des séquences de piARN antisens aux transposons dont l'extrémité 5' est souvent une Uridine<sup>ii</sup>, et la protéine Ago3 est associée à des séquences de piARN colinéaires (sens) aux transposons qui comportent souvent une adénosine position 10<sup>369</sup>. De plus, on observe une relation particulière entre les piARN de Ago3 et Aubergine, qui sont complémentaires antisens sur les premiers 10 nucléotides<sup>380</sup>.

En l'absence de la protéine Ago3, seuls les piARN associés à Aubergine collapsent, ce qui compromet l'intégrité des cellules souches de la lignée germinale, mais les piARN associés à la protéine Piwi sont préservés, en particulier ceux qui dérivent du cluster piARN appelé *flamenco*<sup>378</sup>.

---

<sup>i</sup> Enzyme de maturation du précurseur piwi-ARN.

<sup>ii</sup> L'extrémité 3' des piARN est également 2'-O-méthylés par l'enzyme Pimet (piARN methyltransférase<sup>379</sup>).

La plupart des piARN sont dérivés de la transcription de régions clusterisées contenant des insertions imbriquées de transposons et qui sont localisées près des régions péricentromériques et télomériques. Ces clusters (6 à 7 clusters par chromosome) sont pauvres en gènes et en séquences répétées, d'une taille de 10-83 kb, produisant une variété de 10 à 4500 piARN avec une forte préférence alternée pour un seul des deux brins<sup>381</sup>.

Les loci de piARN semblent attirer et accumuler un cimetière d'éléments transposables comme mémoire immunitaire de l'exposition à ces éléments<sup>52</sup>. Dans la plupart des loci de piARN, on trouve des éléments transposables qui peuvent être insérés dans les deux directions.

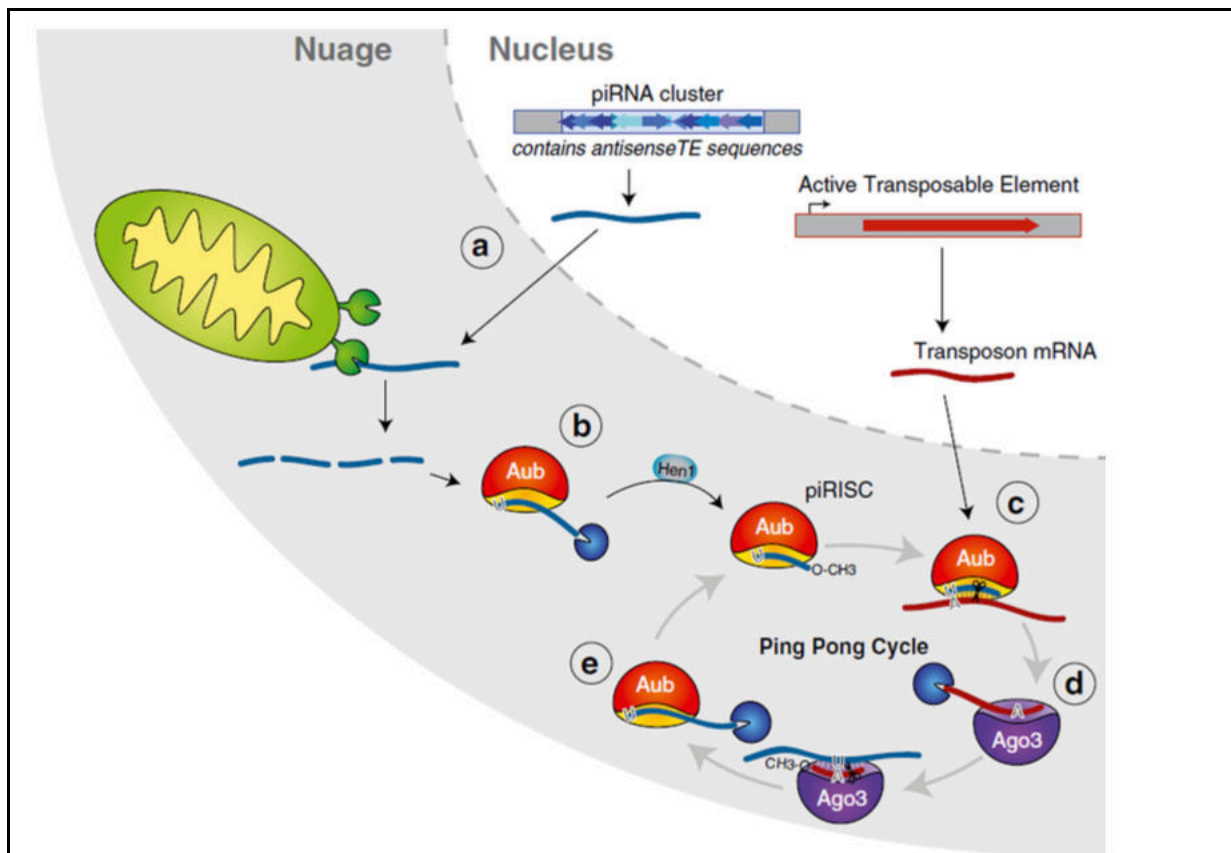
Cependant, on observe certaines exceptions comme le cluster *flamenco*, qui contient essentiellement des transposons insérés dans une seule direction. Ce cluster particulier est actif uniquement dans les cellules somatiques du follicule qui entourent et protègent l'oocyte. Les piARN dérivant du cluster *flamenco* ciblent entre autre les transposons rétroviraux *gypsy* en trans d'une manière Piwi-dépendante.

Dans les cellules souches germinales, la protéine Piwi est principalement nucléaire tandis que les protéines Aubergine et Argonaute-3 sont majoritairement cytoplasmiques. La protéine Ago3 est enrichie dans des granules perinucléaires denses aux électrons qui sont appelées nuage. De plus, une petite quantité de protéines Piwi et Aubergine est déposée et transmise dans la région postérieure du cytoplasme de l'oocyte, là où est programmée la lignée germinale primordiale de l'embryon<sup>369</sup>.

#### **7.11.4.3 Transmission maternelle du programme de ciblage piwi-ARN**

La plupart du programme de ciblage piARN est transmis par la lignée maternelle, avec la déposition dans l'oocyte de l'essentiel des protéines Piwi/Aubergine. Les copies génomiques de transposons divergentes de l'ADN maternel (par exemple des long éléments inter-espacés nucléaires de type I, ou des éléments P actifs) qui sont transmises par l'ADN paternel déclenchent une réponse d'amplification ping-pong dans le zygote. Bien que le zygote soit viable, l'amplification ping-pong dans la lignée germinale féminine échoue à cibler efficacement les nouveaux transposons qui se mobilisent et entraînent des phénotypes stériles connus sous le nom de dysgénésie des

hybrides. L'insertion d'un ou deux éléments P actifs dans un locus de cluster piARN maternel naïf suffit à générer une amplification ping-pong robuste et un zygote fertile<sup>382</sup>.



**Figure 12. Schéma du mécanisme ping-pong dans la lignée germinale**

**(a)** Dans la lignée germinale et les cellules somatiques du follicule épithélial, les piARN primaires sont générés à partir de précurseurs simple-brins transcrits à partir des loci de piARN. Le transcript primaire est exporté dans le cytoplasme et clivé à des intervalles réguliers par l'enzyme Zucchini<sup>383</sup>. Ce clivage détermine l'extrémité 5' du piARN, qui est ensuite transféré dans une protéine Piwi ou Aubergine **(b)**, tandis que l'extrémité 3' est taillée par des enzymes différentes selon les espèces, telles que l'enzyme PARN1 chez *C.elegans*<sup>384</sup>, Nibbler (homologue de mut-7) chez *D.Melanogaster*<sup>385</sup> ou PNLDC1 chez *Bombyx Mori*<sup>386</sup> et *Mus musculus*<sup>387</sup>. La seconde méthode de génération des piARN repose sur la voie d'amplification ping-pong, qui constitue une réponse adaptative spécifique aux transposons actifs. Les piARN générés par ce mécanisme sont appelés piARN secondaires. Le mécanisme ping-pong est activé uniquement dans les cellules germinales et prend place dans les granules cytoplasmiques périnucléaires qui sont appelées nuage. Dans cette voie, c'est la protéine Ago3 (contenant des piARN générés à partir de transposons sens) qui clive



*le précurseur primaire<sup>377</sup>. L'amplification réside dans le fait au lieu d'être dégradé, une partie du transcrit clivé par Ago3 est recyclée et maturée en piARN par l'incorporation dans une protéine Aubergine, qui peut à son tour cliver un transcrit complémentaire. Le mécanisme se propage par une succession de clivages réciproques (ping-pong) générant des piARN phasés de 10 nucléotides, et est rendu possible par la mémoire et la transcription antisens d'un élément transposable similaire contenu dans le loci piARN. De plus, plusieurs protéines impliquées dans la biogénèse des piARN comme Zucchini, Papi et PARN1/PNLDC1 sont localisées ou ancrées à la surface de la mitochondrie<sup>384,386,388</sup>.*

*(d'après Toth et al., 2016<sup>389</sup>)*

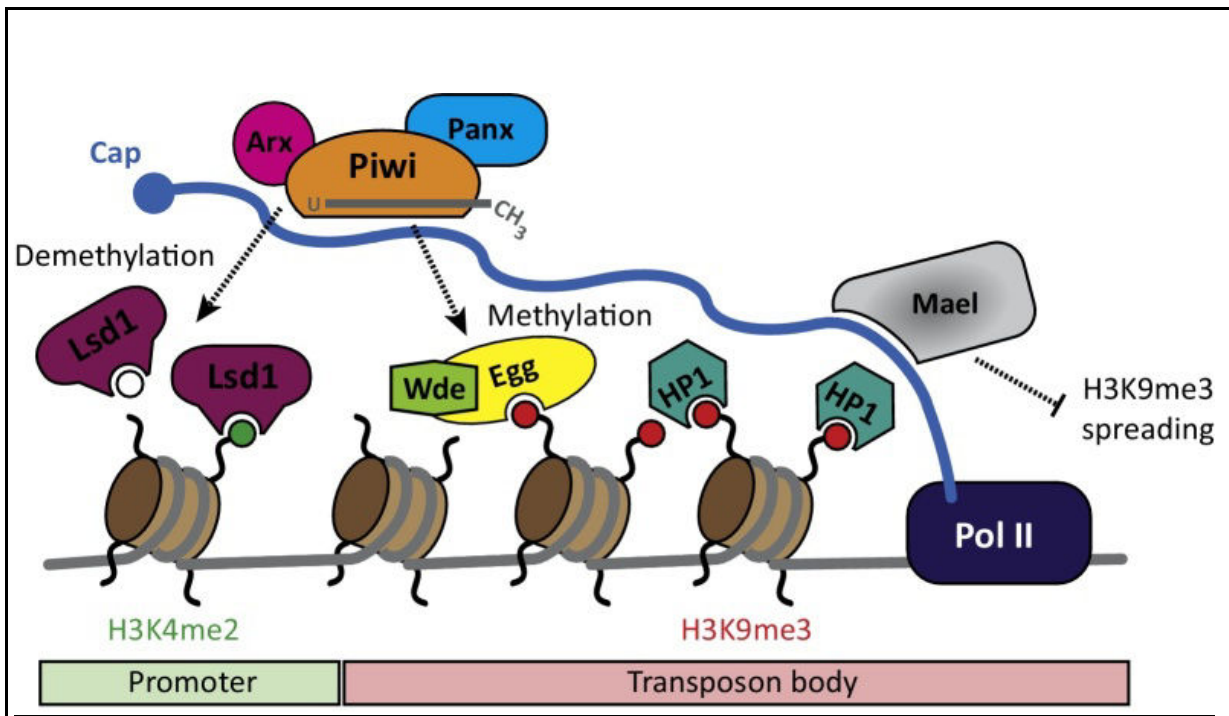
#### **7.11.4.4 Rôle fonctionnel des piARN et protéines Piwi**

Une des fonctions des piARN est la protection du génome contre la prolifération des éléments transposables, qui peuvent être mobilisés pendant la déméthylation généralisée qui est programmée au cours du développement précoce de l'embryon<sup>52</sup>. Le rôle fonctionnel de la plupart des piARN, qui sont exprimés préférentiellement pendant le stade pachytène de la méiose, reste cependant inconnu.

La fonction d'inactivation des transposons est assurée par le clivage des éléments transposables au niveau post-transcriptionnel mais également par une inactivation au niveau transcriptionnel (**Figure 13**).

En effet, chez la drosophile, la protéine nucléaire Piwi est requise pour la déposition de marques épigénétiques de type H3K9me3 sur les éléments transposables, indépendamment de son activité catalytique<sup>390</sup>. La déposition de la marque H3K9me3 requiert les protéines Asterix (Arx)<sup>391</sup> et Panoramix (Panx)<sup>392</sup>.

En conclusion, les piARN représentent une classe de petits ARN non-codants de biogénèse distincte des miARN, dont le rôle primordial est le maintien de l'intégrité du génome, principalement dans la lignée germinale et pendant le développement.



**Figure 13. Mécanisme d'inhibition transcriptionnelle des transposons par la protéine nucléaire Piwi**

Dans les ovaires de la drosophile, les transcrits naissants de transposons sont ciblés par le complexe protéique formé de Piwi, Astérix (Arx) et Panoramix (Panx). Le complexe induit la déposition de marques épigénétiques répressives comme H3K9me3, par le recrutement des protéines Eggless (Egg), Windei (Wde) et la protéine hétérochromatique 1 (HP1). Les marques épigénétiques activatrices sont effacées par la protéine déméthylase spécifique des lysines (LSD1) et la diffusion en cis de la marque épigénétique H3K9me3 est limitée par la protéine Maelstrom.

(d'après Czech & Hannon 2016<sup>52</sup>)

## 8 Résultats

### 8.1 Etat de l'art au commencement de la thèse

J'ai intégré l'équipe du Dr. Trabucchi (C3M) en Janvier 2015, en tant que doctorant du programme Labex Signallife, après avoir enchaîné trois périodes de travail de 1 mois de Octobre 2014 à Décembre 2014 dans les équipes respectives du Dr. Bardoni (IPMC), Dr. Barbry (IPMC) puis Dr. Trabucchi (C3M) où j'ai eu finalement eu l'opportunité d'effectuer ma thèse.

A mon arrivée dans l'équipe, un des projets majeurs en cours concernait l'étude de la protéine de liaison à l'ARN SFPQ et son rôle dans la régulation du ciblage des ARNm par les miARN. Cette protéine a été découverte et caractérisée avant mon arrivée dans le laboratoire sur la base d'une spectrométrie de masse des interacteurs de la protéine AGO2 destinée à identifier de nouveaux partenaires du complexe RISC.

De manière intéressante, cette expérience a révélé plusieurs protéines nucléaires comme celles des paraspeckles, dont l'observation appelait à une étude approfondie des fonctions nucléaires du complexe RISC conjuguée à la protéine SFPQ.

### 8.2 Travail de thèse

#### 8.2.1 La protéine SFPQ module l'activité de régulation post-transcriptionnelle des gènes par les miARN dans le nucléoplasme

Au cours de ma première année de thèse, une partie du travail s'est donc articulée autour du projet initial et a fait l'objet d'un article publié dans Nature Communications (Bottini et al., 2017), dans lequel j'ai contribué en tant que second auteur.

Nous avons débuté ce projet en explorant les partenaires ARN-dépendant de la protéine AGO2 par immunoprécipitation et spectrométrie de masse quantitative. Parmi les interactants ARN-dépendants, nous nous sommes concentrés sur trois protéines nucléaires abondantes : SFPQ, PSPC1 et NONO qui forment la famille drosophila behavior and human splicing (DBHS). Nous avons démontré que le complexe RISC nucléoplasmique est associé aux protéines SFPQ, PSPC1 et NONO dans plusieurs lignées cellulaires murines et humaines, d'une manière qui dépend de SFPQ.

Des expériences de type HITS-CLIP de la protéine AGO2 et/ou de la protéine SFPQ dans des cellules souches nous ont permis de montrer que SFPQ se lie préférentiellement aux

3'UTR longs en utilisant deux motifs spécifiques. En effet, SFPQ contrôle significativement environ 20% de l'activité de liaison de AGO2, ce qui est répercuté au niveau transcriptomique. Cependant, cette activité concerne uniquement les sites de liaison de SFPQ proches (<500 nucléotides) de AGO2.

De plus, nous avons observé que cette régulation s'étend aux ARNm cytoplasmiques. Ce résultat suggère que la liaison et l'agrégation de la protéine SFPQ à l'ARN programme la structure du 3'UTR et donc les possibilités de ciblage par les miARN dans le noyau, et ceci d'une manière qui semble préservée dans le cytoplasme.

Enfin, nous avons montré en particulier que l'expression de SFPQ contrôle le programme de ciblage par let-7a, et module la transition des cellules souches vers l'état différencié.

Ces résultats contribuent à la diversité des mécanismes de régulation de l'activité des miARN.

#### Mon apport à ce travail:

Durant cette première année, j'ai développé et adapté un protocole stringent de fractionnement cytoplasme/nucléoplasme<sup>i</sup> et chromatine<sup>ii</sup> (section 8.3 **Méthodes: Cellular fractionation**), dans le but d'étudier les fonctions nucléaires des miARN et de AGO2, tout en évaluant le rôle des protéines de la famille DBHS (paraspeckles) dans ce contexte.

Brièvement, les cellules ont été collectées par traitement à la trypsine et lavées une fois avec une solution saline tamponnée au phosphate (PBS 1X). Les culots cellulaires ont été lysés pendant 10 minutes sur glace en utilisant 380µL de tampon HLB par 10.10<sup>6</sup> cellules (HLB, 10mM Tris-Cl pH 7.5, 0.3% NP-40, 10mM NaCl, 10% glycérol, 3mM MgCl<sub>2</sub>, Inhibiteurs de protéases (cOmplete, Roche), Inhibiteurs de RNase (RNaseOut, Life tech). Les culots de noyaux ont été lavés trois fois avec la solution HLB, et lysés pendant 10 min sur glace avec 380µL de solution Modifiée Wuarin-Schibler (MWS, 20mM Tris-Cl pH 7.5, 300mM NaCl, 1% NP-40, 4mM EDTA, Inhibiteurs de protéases (cOmplete, Roche), Inhibiteurs de RNase (RNaseOut, Life tech).

Les culots de chromatine précipitée obtenus à partir du protocole décrit ci-dessus ont été solubilisés d'une autre manière, comme décrit dans<sup>394</sup>.

---

<sup>i</sup> Voir Gagnon et al., Nature Protocols 2014, réf<sup>393</sup>.

<sup>ii</sup> Voir Minajigi et al., Science 2015, réf<sup>394</sup>.

Brièvement, les culots de chromatine ont été lavés trois fois avec la solution MWS, et resuspendus dans 250µL de solution tampon chromatine (Tampon chromatine, 50mM Tris-Cl pH 7.5, 0.5% NP-40, 0.1% sarkosyl), Inhibiteurs de protéases (cOmplete, Roche), Inhibiteurs de RNase (RNaseOut, Life tech).

La mixture de chromatine a été brièvement soniquée trois fois pendant 5" avec intermittences sur un Vibracell Bioblock Scientific (Puissance 40% Watts et amplitude 40%), puis supplémentée avec 25 unités de DNase I (RQ1 DNase 1u/µL, Promega) ainsi qu'une concentration finale 1X de tampon RQ1 DNase I, et incubée à 37°C pendant 45 minutes. La solution a été ajustée à 0.6% sarkosyl, 150mM de chlorure de lithium et incubée à 37°C pendant 15 minutes supplémentaires pour solubiliser la chromatine native dans un volume final de 380µL.

Des aliquots de préparations cytoplasmiques, nucléoplasmiques et chromatiques ont été cryogénisés dans l'azote liquide et stockés à -80°C, et les ARN ont été extraits comme décrit dans <sup>393</sup>.

Ces expériences ont permis d'identifier la présence de AGO2 et des miARN dans le nucléoplasme (section 8.3, **Figure 3a**) mais également dans la chromatine des cellules (données non-publiées présentées en section 8.4, **Annexe 1 & Annexe 2**), tandis que la protéine AGO2 et les miARN interagissent avec SFPQ, NONO et PSPC1 seulement dans le nucléoplasme (section 8.3, **Supplemental figure 4**).

Je me suis ensuite intéressé aux fonctions d'un miARN en particulier (let-7a), et j'ai exploré la régulation de cibles nucléaires comme la cible canonique LIN28A, en me basant sur un système de lignée cellulaire en culture transfectées transitoirement par let-7a (section 8.3, **Supplemental figure 5E**).

Des expériences de RT-qPCR ont permis de confirmer la diminution de l'ARNm nucléaire *Lin28a* en présence de let-7a (données non-publiées). Cet effet est restauré suite à la déplétion concomitante des protéines SFPQ, NONO ou PSPC1 (données non-publiées). Par ailleurs, de nombreuses expériences d'immunoprécipitation suivies d'analyse RT-qPCR ont permis d'observer une association de AGO2 et des paraspeckles à l'ARNm nucléaire *Lin28a* (données non-publiées).

J'ai ensuite évalué le mode de régulation de l'ARNm nucléaire *Lin28a* par let-7a, en considérant un mécanisme transcriptionnel et/ou post-transcriptionnel.

Dans un premier temps, j'ai évalué le potentiel de régulation transcriptionnelle à travers des expériences de nuclear run-on<sup>i</sup> (section 8.3, **Méthodes: nuclear run-on**) et j'ai optimisé en parallèle une approche originale par global run-on sequencing<sup>ii</sup> (GRO-Seq). Cependant, les expériences pilotes n'ayant pas permis d'identifier clairement une diminution de la transcription de LIN28A en présence de let-7a (section 8.3, **Supplemental figure 9B & Supplemental figure 9C**), nous n'avons pas poursuivi cette hypothèse plus en détails.

Les résultats ont été plus probants en examinant la stabilité de l'ARNm *Lin28a* par traitement à l'actinomycine D.

En présence de let-7a, j'ai observé la dégradation de l'ARNm nucléaire *Lin28a*. Par ailleurs, cet effet semble restauré suite à la déplétion concomitante de la protéine SFPQ, mais pas la protéine NONO ou PSPC1 (section 8.3, **Figure 5d**).

### **8.2.2 Mécanisme de l'extinction post-transcriptionnelle des gènes miARN-dépendante**

Dans la seconde partie du travail de thèse, je me suis intéressé au mécanisme de dégradation de l'ARNm par les miARN nucléaires. En effet, bien que certaines études aient rapporté une activité de dégradation de l'ARNm par les miARN dans le noyau, le mécanisme moléculaire est encore inconnu.

### **8.2.3 Composants moléculaires nucléaires du complexe RISC**

Nous nous sommes donc intéressés à la composition moléculaire du RISC nucléaire par des expériences de co-immunoprécipitation, avec une approche candidat ciblée sur certains facteurs exonucléolytiques tels que l'exosome nucléaire, ou le complexe CCR4-NOT1.

Pour étudier ces interactions, j'ai exprimé des protéines fusion Flag-AGO2 ou Flag-EXOSC10 dans les cellules 293T, et immunoprécipité les protéines nucléoplasmiques avec un anticorps anti-Flag.

Les protéines Flag-AGO2 co-immunoprécipitent avec les protéines endogènes EXOSC10 et CNOT1 d'une manière résistante à la RNase, mais pas avec SFPQ en présence de RNase (données non-publiées présentées en section 8.4, **Annexe 4A**). De manière

---

<sup>i</sup> Voir Morris et al., Nature Protocols 2015

<sup>ii</sup> Voir Lis et al., Science 2008

complémentaire, les protéines Flag-EXOSC10 co-immunoprécipitent avec les protéines endogènes AGO2 d'une manière résistante à la RNase (données non-publiées présentées en section 8.4, **Annexe 4B**).

Ces expériences nous permettent d'affirmer que le complexe nucléaire RISC contient l'exosome nucléaire d'une manière indépendante de l'ARN.

Ces observations m'ont conduit à interroger le mécanisme moléculaire de l'interaction de l'exosome nucléaire avec le complexe RISC, dont nous proposons un modèle de travail basé sur la littérature (**Figure 14**), ainsi que des expériences complémentaires pour confirmer les résultats obtenus.

#### **8.2.4 Des expériences complémentaires à réaliser**

Pour confirmer les résultats obtenus, il serait intéressant de s'appuyer sur les éléments publiés, ainsi que l'argumentation que nous développons dans cette thèse (voir section **9.2.2**). Brièvement, l'hypothèse de travail qui semble la plus cohérente au regard des données actuelles est la liaison directe de l'exosome au module CNOT2-CNOT3 du complexe CCR4-NOT1.

Dans ce cadre, il serait intéressant de tester la stabilité de l'interaction du complexe CCR4-NOT1 et de l'exosome dans certaines des souches Not générées par Reed et al., 1980 <sup>395</sup> et Collart et al., 1993 <sup>396</sup>, comme le mutant tronqué cdc39-2 (CNOT1 tronqué en C-terminal), ou le mutant déstabilisé Not1-10 (G1943D, ou G2223D chez l'humain, voir Lee et al., 1998 <sup>397</sup>), et plus précisément certains des mutants Not2 dans les régions Not-box conservées (K134E, R156G, voir Russell et al., 2002 <sup>398</sup>).

Ces expériences permettraient d'identifier dans un premier temps quelle partie de la protéine échafaudage Not1 est importante pour le recrutement de l'exosome, puis quel module est important pour le recrutement de l'exosome.

D'autres expériences reposeraient sur la confirmation des interactions (GST-pulldown), ou bien in-fine dans un système minimal in-vitro composé d'un complexe purifié CNOT1-CNOT2-CNOT3 (voir Boland et al., 2013 <sup>399</sup>) et un exosome reconstitué (voir Greimann et al., 2008 <sup>400</sup>). La stabilité des interactions, y compris avec les mutants, pourrait être évaluée par filtration sur gel suivie de détection UV ou coomassie, diffusion

de lumière multi-angle et indice de réfraction différentielle, ou bien par titration calorimétrique isotherme. Il serait ensuite intéressant de caractériser la structure du complexe CNOT1-CNOT2-CNOT3-Exosome par cristallographie aux rayons X, ou autre méthode d'analyse structurale.

Des expériences fonctionnelles basées sur l'ancrage artificiel du module CNOT2-CNOT3 (et plus particulièrement des régions Not-box) sur un gène rapporteur, avec la déplétion concomitante de l'exosome permettrait d'apporter des informations sur la relation fonctionnelle entre le complexe CCR4-NOT1 et l'exosome. Une validation complémentaire par l'addition de peptides compétiteurs<sup>i</sup> pourrait être utilisée, avec les informations recueillies sur la structure du complexe.

S'ils venaient à être concluants, les résultats de ces expériences permettraient d'apporter une interprétation nouvelle des fonctions du complexe CCR4-NOT1 et des miARN dans le noyau, en particulier quant à la sélection post-transcriptionnelle de l'expression de certains ARN régulateurs tels que les enhancer-ARN par l'exosome nucléaire.

Si toutefois les expériences préliminaires s'avéraient non-concluantes, nous devons postuler qu'il existe un ou plusieurs composant(s) adaptateur(s) différent(s) entre le module CNOT2-CNOT3 et l'exosome, dont l'identité est impossible à prédire à l'heure actuelle, bien que cette hypothèse semble improbable.

---

<sup>i</sup> Méthode utilisée pour démontrer la relation fonctionnelle entre la protéine Argonaute et GW182, voir section 7.6.3.2.1.








**8.3 Article 1. Post-transcriptional gene silencing mediated by microRNAs is controlled by nucleoplasmic SFPQ**

## ARTICLE

DOI: 10.1038/s41467-017-01126-x

OPEN

# Post-transcriptional gene silencing mediated by microRNAs is controlled by nucleoplasmic Sfpq

Silvia Bottini <sup>1,2</sup>, Nedra Hamouda-Tekaya<sup>1,2</sup>, Raphael Mategot<sup>1,2</sup>, Laure-Emmanuelle Zaragosi <sup>3</sup>,  
Stephane Audebert <sup>4</sup>, Sabrina Pisano<sup>5</sup>, Valerie Grandjean<sup>1,2</sup>, Claire Mauduit<sup>1,2,6,7</sup>, Mohamed Benahmed<sup>2,8</sup>,  
Pascal Barbry <sup>3</sup>, Emanuela Repetto<sup>1,2</sup> & Michele Trabucchi <sup>1,2</sup>

There is a growing body of evidence about the presence and the activity of the miRISC in the nucleus of mammalian cells. Here, we show by quantitative proteomic analysis that Ago2 interacts with the nucleoplasmic protein Sfpq in an RNA-dependent fashion. By a combination of HITS-CLIP and transcriptomic analyses, we demonstrate that Sfpq directly controls the miRNA targeting of a subset of binding sites by local binding. Sfpq modulates miRNA targeting in both nucleoplasm and cytoplasm, indicating a nucleoplasmic commitment of Sfpq-target mRNAs that globally influences miRNA modes of action. Mechanistically, Sfpq binds to a sizeable set of long 3'UTRs forming aggregates to optimize miRNA positioning/recruitment at selected binding sites, including let-7a binding to Lin28A 3'UTR. Our results extend the miRNA-mediated post-transcriptional gene silencing into the nucleoplasm and indicate that an Sfpq-dependent strategy for controlling miRNA activity takes place in cells, contributing to the complexity of miRNA-dependent gene expression control.

<sup>1</sup>INSERM U1065, C3M, Team Control of Gene Expression (10), 151 route de St-Antoine-de-Ginestière, B.P. 2 3194, Nice 06204, France. <sup>2</sup>Université Côte d'Azur, INSERM, C3M, 151 route de St-Antoine-de-Ginestière, B.P. 2 3194, Nice 06204, France. <sup>3</sup>Université Côte d'Azur, CNRS, IPMC, Valbonne, France. <sup>4</sup>CRCM, Marseille Protéomique, Institut Paoli-Calmettes, Aix Marseille University, INSERM, CNRS, 27 bd Leï Roure, BP 30059, Marseille 13273, France. <sup>5</sup>Université Côte d'Azur, CNRS, INSERM, IRCAN, Faculty of Medicine, 28 Av. Valombrose, Nice 06107, France. <sup>6</sup>Université Lyon 1, UFR Médecine Lyon Sud, Lyon F-69921, France. <sup>7</sup>Hospices Civils de Lyon, Hôpital Lyon Sud, Laboratoire d'Anatomie et de Cytologie Pathologiques, Pierre-Bénite F-69495, France. <sup>8</sup>Centre Hospitalier Universitaire de Nice, Département de Recherche Clinique et d'Innovation, Nice F-06001, France. Silvia Bottini and Nedra Hamouda-Tekaya contributed equally to this work. Correspondence and requests for materials should be addressed to M.T. (email: [mtrabucchi@unice.fr](mailto:mtrabucchi@unice.fr))

**M**icroRNAs (miRNAs) are functional small RNAs and fundamental components of gene expression programs that regulate many biological processes, including cell proliferation, differentiation, and death<sup>1</sup>. Like other small RNAs, miRNAs can be used as biomarkers for human disorders<sup>2</sup>. miRNAs associate with Argonaute (Ago) proteins, mainly Ago2, to form the miRNA-induced silencing complex (miRISC) and target mRNAs<sup>3</sup>. Although miRNA-dependent silencing has mainly been described in the cytoplasm, a growing body of evidence indicates that miRNAs are also functional in the nucleus<sup>4–6</sup>. Canonically, miRNAs use a sequence of 6–8 nucleotides (nt) at their 5' end, called the seed region, to block the translation or promote the degradation of target mRNAs<sup>1</sup>. Despite efforts to develop bioinformatics tools to predict miRNA-binding sites, it has been demonstrated that prediction approaches can be misleading, yielding approximately 70% false or negative targets<sup>7</sup>. Such a low prediction efficiency might be improved by considering the activity of RNA-binding proteins that bind to mRNAs to control miRNA targeting<sup>1</sup>. In support of this mechanism, it has been shown that some RNA-binding proteins associate with specific mRNAs and interfere with specific miRNA-binding sites to either inhibit or enhance miRNA targeting<sup>8</sup>. This result leads to the concept of a sequence micro-environment surrounding miRNA-binding sites that plays an important role in regulating miRNA activity<sup>8</sup>. Although some examples of such a mechanism have been described, including those for Hu-Antigen R (HuR)<sup>9</sup> and Deadend I (Dnd1)<sup>10</sup>, much remains to be learned about the molecular mechanism(s) underlying the roles of the sequences surrounding miRNA-binding sites, and whether this feature is general of miRNA targeting regulation or is confined to specific cases.

Herein, to explore the RNA dependency of miRNA activity we used a quantitative proteomic analysis to identify RNA-dependent Ago2 interactors. Among the identified RNA-dependent interactors, we focused our investigation on Splicing factor proline/glutamine-rich protein (Sfpq). We found that Sfpq interacts with nucleoplasmic miRISC in different human and mouse cell lines. We demonstrated that Sfpq directly promotes miRNA targeting through local binding, ultimately facilitating miRNA-dependent degradation. Although Sfpq only interacts with miRISC in the nucleoplasm, it appeared to modulate miRNA targeting in both the nucleoplasm and cytoplasm. This result indicated that a nucleoplasmic commitment for Sfpq-target mRNAs globally influences miRNA targeting in both cellular compartments. We found that Sfpq preferentially binds to long 3' UTRs through long sequences that harbor multiple copies of two distinct Sfpq-binding motifs that we had identified. Observations by atomic force microscope (AFM) further showed that Sfpq aggregates onto target 3'UTRs. This process ultimately leads to the position/recruitment optimization of miRNAs at specific binding sites, including let-7a targeting of the Lin28A 3'UTR. In stem cells, Sfpq regulated the let-7-dependent gene expression program toward a neuron-like phenotype differentiation. Our results unveil an unanticipated role for Sfpq in post-transcriptionally promoting miRNA-dependent gene silencing from the nucleoplasm to the cytoplasm of a sizeable subset of mRNAs that have long 3'UTRs. These results highlight the importance of nuclear miRNA targeting and the sequence features of mRNAs for post-transcriptional miRNA programs during gene expression regulation.

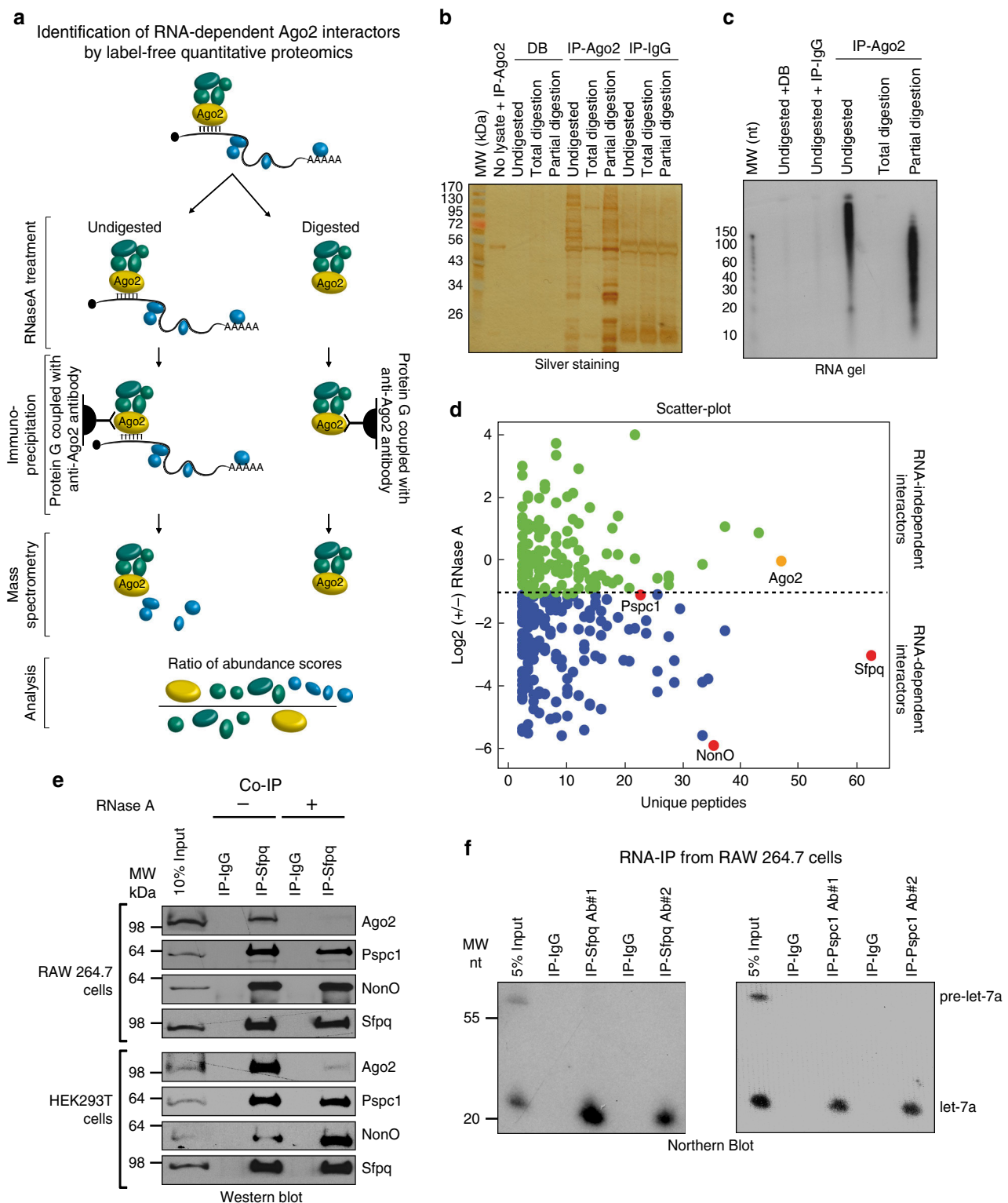
## Results

**Identification of RNA-dependent Ago2 interactors.** To identify RNA-dependent Ago2 interactors, we performed a label-free quantitative mass spectrometry (MS) analysis<sup>11</sup> of

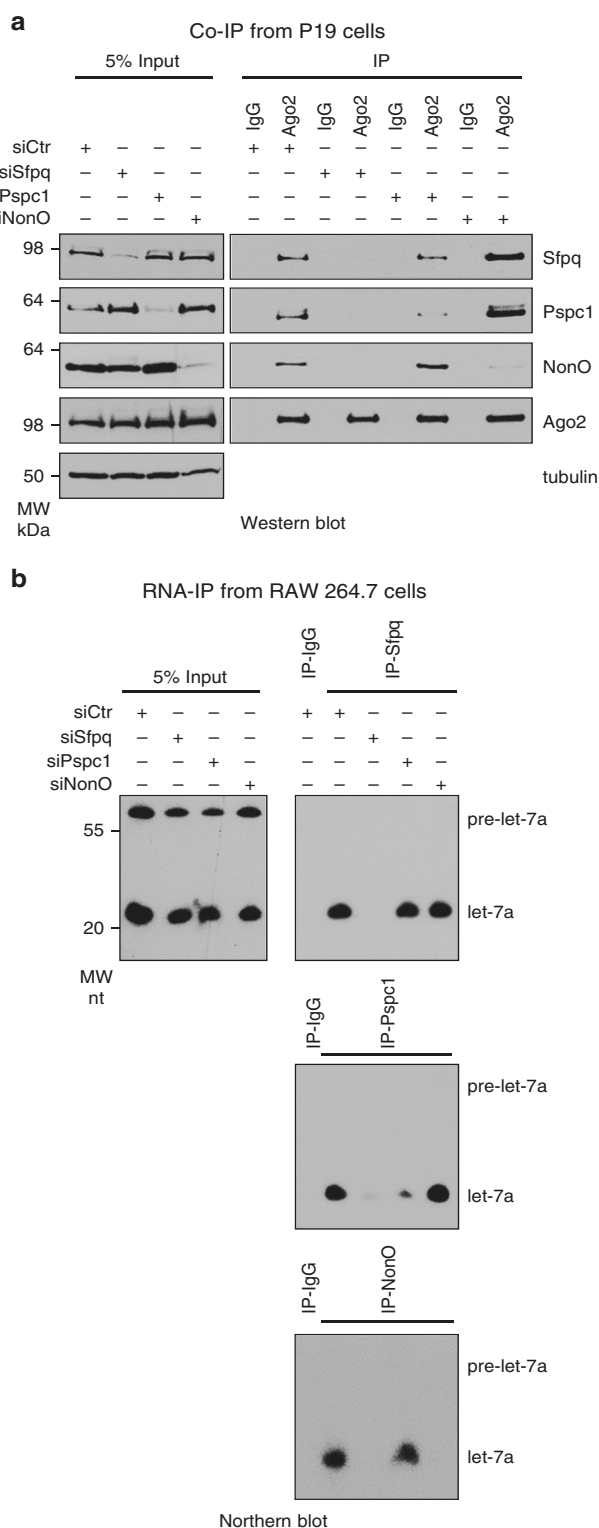
immunoprecipitated Ago2-containing complexes obtained from mouse RAW 264.7 cell extracts that were either undigested (–) RNase) or totally digested ((+)RNase) (Fig. 1a). Immunoprecipitation (IP) was revealed by SDS-PAGE-silver staining or an RNA gel (Fig. 1b, c). MS analysis identified 915 proteins that might interact with Ago2 (Supplementary Data 1). For quantitative analysis, we applied different filters based on the following criteria: (i) the reproducibility of three replicates using a PCA-based procedure; (ii) the application of a stringent cutoff point for the Mascot score to identify unique peptides; and (iii) the enrichment of proteins identified (abundance score) in the Ago2 immunoprecipitated samples with respect to the IgG IP (ANOVA test,  $p$ -value  $\leq 0.05$ ; Supplementary Data 2). Together, these filters enabled the removal of background noise and resulted in the identification of 299 different proteins in both conditions. After statistical analysis, we identified 133 Ago2 RNA-independent and 166 RNA-dependent interactors (Fig. 1d, Supplementary Fig. 1a, b, and Supplementary Data 3).

Many of the Ago2 interactors found in this study were previously independently identified as being associated with Ago2. Indeed, among the RNA-independent interactors, we identified many miRISC components, including members of the GW protein family (Trinucleotide Repeat Containing 6—TNRC6—in mammals), and CCR4-NOT Transcription Complex Subunit 1 (CNOT1), which, together with DEAD-Box Helicase 6 (DDX6, an RNA-dependent interactor), regulates miRNA-dependent deadenylation of target mRNAs<sup>12</sup>. As expected, we also found several known RNA-dependent interactors, such as Polyadenylate-Binding Protein 1 (PABP1)<sup>13</sup>, Elav1 (also called HuR)<sup>14</sup>, the putative DExD box helicase Moloney Leukemia Virus 10 (MOV10)<sup>13</sup>, and La Ribonucleoprotein Domain Family Member 1 (Larp1)<sup>15</sup>. Notably, although it has been shown that Ago2 also associates with proteins involved in miRNA biogenesis<sup>15</sup>, we did not find Dicer, Trans-Activation Responsive RNA-Binding Protein 2 (Trbp2), or Heat Shock Protein 90 (Hsp90) in our analysis, suggesting that the antibody we used specifically immunoprecipitated Ago2 associated with mature miRNAs and not precursors (Supplementary Fig. 1c, d). Taken together, despite some differences, our list of 299 Ago2 interactors is similar to that reported for a previous SILAC quantitative proteomic analysis<sup>15</sup>. Therefore, our results corroborate previous works, indicating the high specificity of our analysis, and extend them by identifying novel interactors.

To study the biological and mechanistic relevance of RNA as a mediator of the interactions between Ago2 and other proteins for miRNA function, we decided to focus on Sfpq, Paraspeckle Component 1 (Pspc1), and Non-POU Domain Containing, Octamer-Binding (NonO), which form a protein complex<sup>16</sup>. These three proteins were among the most abundant RNA-dependent Ago2 interactors (red points in Fig. 1d). All three proteins contain two RNA recognition motif (RRM) domains that confer binding specificity to RNA<sup>17</sup> and a coiled-coil domain that mediates protein aggregation<sup>18</sup>. Sfpq and NonO are multifunctional RNA-binding proteins that can regulate different steps of the mRNA life cycle, including splicing, nuclear localization, and degradation<sup>19, 20</sup>. Pspc1 function is poorly understood. Interestingly, these proteins are known to associate with the long non-coding RNA Nuclear Paraspeckle Assembly Transcript 1 (Neat1) to form the paraspeckle, a nucleoplasmic compartment of approximately 0.2–1  $\mu$ m in size with several physiopathological functions<sup>16</sup>, including nuclear retention of mRNAs<sup>21</sup>, transcriptional regulation<sup>22</sup>, cancer pathogenesis<sup>23</sup>, and viral infection<sup>24</sup>. Interactions with Ago2 were also observed for Sfpq in previous MS analyses in HEK293T cells and in hepatocellular carcinoma<sup>15, 25</sup>. To validate the specific associations between Ago2 and Sfpq, Pspc1, and NonO, we performed co-IP



**Fig. 1** Sfpq, Pspc1, and NonO are components of Ago2 complex and interact with let-7a. **a** Overview of the proteomic method used to identify RNA-dependent proteins interacting with Ago2. **b** Silver staining of an SDS-PAGE analysis from the IP with the anti-Ago2 antibody, the protein G conjugated with the Dynabeads (DB), or the anti-Ago2 antibody alone incubated without cell lysate. The samples were untreated, treated with  $10 \mu\text{g ml}^{-1}$  RNase A for 30 min at room temperature (for partial digestion), or treated with  $10 \text{ mg ml}^{-1}$  RNase A for 30 min at room temperature (for total digestion). **c** Radioactive images of a TBE-Urea gel showing signal from  $^{32}\text{P}$ -labeled RNA fragments of samples untreated, treated with  $10 \mu\text{g ml}^{-1}$  RNase A for 30 min at room temperature (for partial digestion), or treated with  $10 \text{ mg ml}^{-1}$  RNase A for 30 min at room temperature (for total digestion). **d** Scatter-plot of the log base 2 ( $\pm$ ) RNase A ratios (abundance scores) plotted with the unique peptides for each identified protein. Each spot is a different protein. **e** Co-IP of endogenous Sfpq and Ago2, Pspc1, or NonO in RAW 264.7 and HEK293T cells. When indicated, cell lysates were incubated at room temperature with RNase A ( $10 \text{ mg ml}^{-1}$ ) for 30 min. **f** Sfpq and Pspc1 interact with mature let-7a but not with the precursor. RAW 264.7 cell extracts were immunoprecipitated with two different antibodies for each indicated protein. The RNA was purified and analyzed by Northern blotting



**Fig. 2** Sfpq mediates the interaction between miRISC and Pspc1 or NonO. **a** Co-IP of endogenous Ago2 with Sfpq, Pspc1, or NonO from P19 cells. Cells were transfected with the indicated siRNAs and analyzed by Western blotting. Ago2 and tubulin served as controls. **b** RNA-IP of let-7a with the endogenous Sfpq, Pspc1, or NonO. RAW 264.7 cells were transfected with the indicated siRNAs. Cell extracts were immunoprecipitated with the indicated antibodies and the RNA was purified and analyzed by Northern blotting

experiments with the endogenous proteins in different human and mouse cell lines, including RAW 264.7, P19, and HEK293T cells (Fig. 1e and Supplementary Fig. 2a). In all three cellular models, we confirmed each protein as an integral component of the Ago2 complex. These results were also validated by co-immunoprecipitating Sfpq-tagged and Ago2-tagged proteins or by GST pull-down experiments (Supplementary Fig. 2b, c). RNase treatment confirmed the RNA dependency of the Ago2 interactions; however, it did not affect the interactions among Sfpq, Pspc1, and NonO (Fig. 1e and Supplementary Fig. 2a). To determine whether they also associate with miRNAs, we performed an ribonucleoprotein complexes immunoprecipitation (RNA-IP) followed by Northern blotting for selected miRNAs in different cell lines. Sfpq and Pspc1 interact with let-7a and miR-23b (Fig. 1f and Supplementary Fig. 2d).

Overall, these results validated our MS analysis and raised the possibility that the Sfpq-Pspc1-NonO complex could regulate miRNA functions in an RNA-dependent fashion.

### Nucleoplasmic Ago2 interacts with Sfpq and its complex.

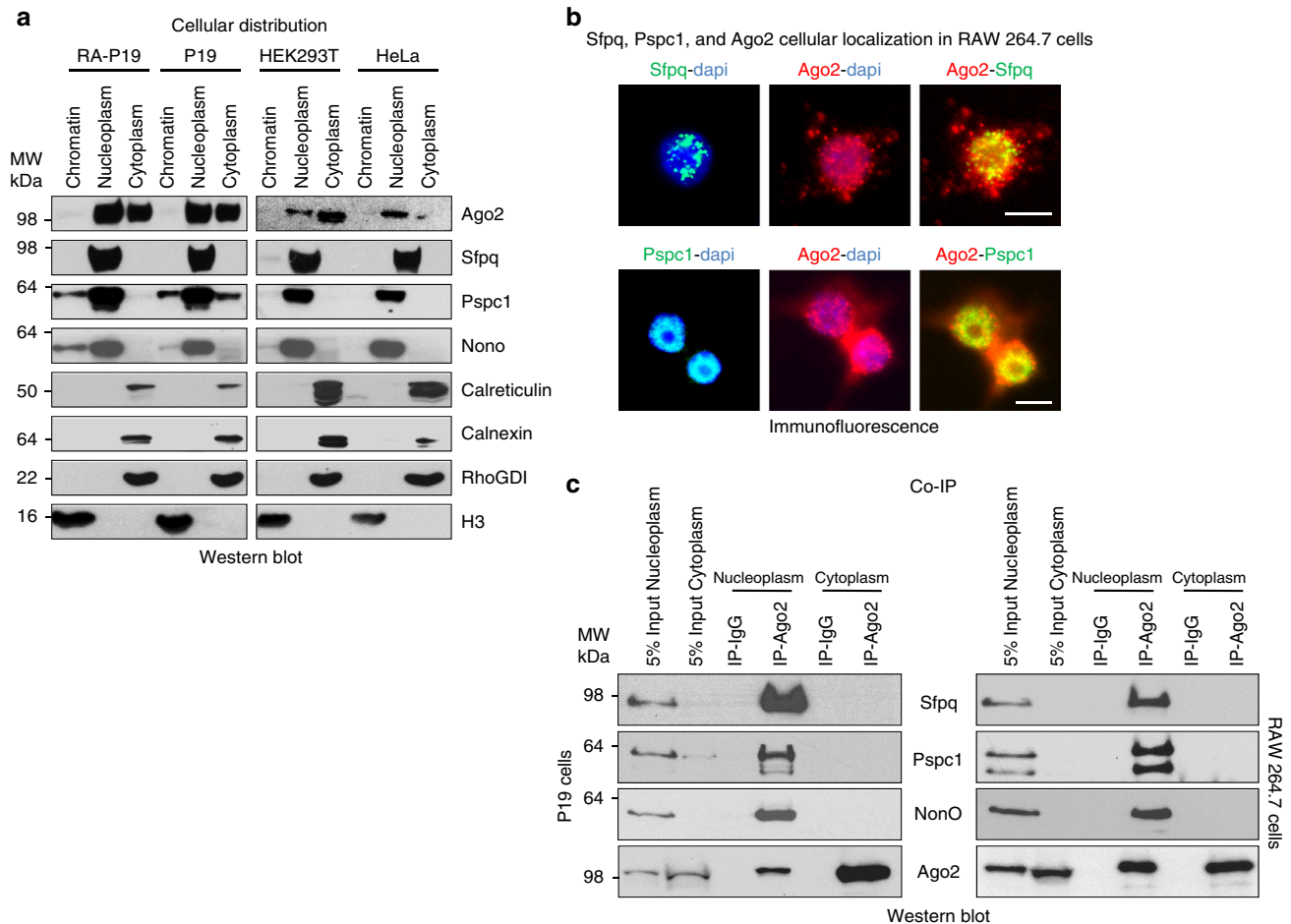
Because Sfpq, Pspc1, and NonO form a complex and share the ability to directly bind single-stranded RNAs, we determined whether one of them specifically mediated the interaction with miRISC. Thus, we singularly knocked down each protein in mouse and human cells and checked for interactions with Ago2 by co-IP and with miRNAs by RNA-IP. As shown in Fig. 2a and Supplementary Fig. 3a and b, Sfpq knockdown specifically inhibited the interaction between Ago2 and either Pspc1 or NonO, whereas Pspc1 or NonO knockdown did not affect the interaction between Sfpq and Ago2. Consistently, Sfpq knockdown also inhibited the interaction between let-7a and either Pspc1 or NonO, whereas Pspc1 or NonO knockdown did not affect the interaction between Sfpq and let-7a (Fig. 2b and Supplementary Fig. 3c).

miRNAs are bound to nuclear Ago2, indicating the existence of a nuclear miRNA pathway<sup>6</sup>. Interestingly, we found that nuclear miRISC co-localizes with Sfpq, Pspc1, and NonO in the nucleoplasm in both human and mouse cell lines by Western and Northern blot analyses (Fig. 3a and Supplementary Fig. 4a). Protein markers for the cytoplasm and the endoplasmic reticulum were absent from nuclear preparations (Fig. 3a). As a second method for testing the nuclear localization of Sfpq and Pspc1, we used co-immunostaining, which clearly indicated that Sfpq and Pspc1 localize in the nucleus, whereas Ago2 is localized in both compartments (Fig. 3b). Co-IP experiments confirmed that Ago2 interacts with Sfpq-Pspc1-NonO complex in the nucleoplasm, but not in the cytoplasm or on chromatin (Fig. 3c and Supplementary Fig. 4b, c). Together, our results indicate that Sfpq mediates the interaction of both Pspc1 and NonO with nucleoplasmic miRISC. Therefore, Sfpq appears to mediate the sequence specificity of this complex to modulate nucleoplasmic miRISC activity.

### Sfpq promotes miRNA binding on a subset of binding sites.

To investigate whether Sfpq associates to all or to a subset of miRNAs, we performed an Sfpq RNA-IP experiment in RAW 264.7 cells followed by high-throughput small RNA sequencing analysis and compared the results with the Ago2 RNA-IP-sequencing data (Supplementary Data 4). We found that 224 and 259 mature miRNAs associated with Sfpq and Ago2, respectively (Supplementary Data 4 and Supplementary Fig. 5a). All 224 miRNAs enriched in the Sfpq IP were also associated with Ago2 (Fig. 4a). Most of the 35 remaining miRNAs that associated solely with Ago2 were miRNA passenger strands (Supplementary Data 4).





**Fig. 3** Sfpq, Pspc1, and NonO interact with Ago2 in the nucleoplasm. **a** Immunoblot analysis of chromatin, nucleoplasm, or cytoplasm from HeLa, HEK293T, P19, or RA-treated P19 cells using antibodies directed to the indicated proteins. **b** Co-immunofluorescence analysis of RAW 264.7 cells stained with the indicated antibodies. Scale bar corresponds to 10  $\mu$ m. **c** Co-IP of endogenous Ago2 and the indicated proteins from nucleoplasmic or cytoplasmic extracts from RAW 264.7 or P19 cells

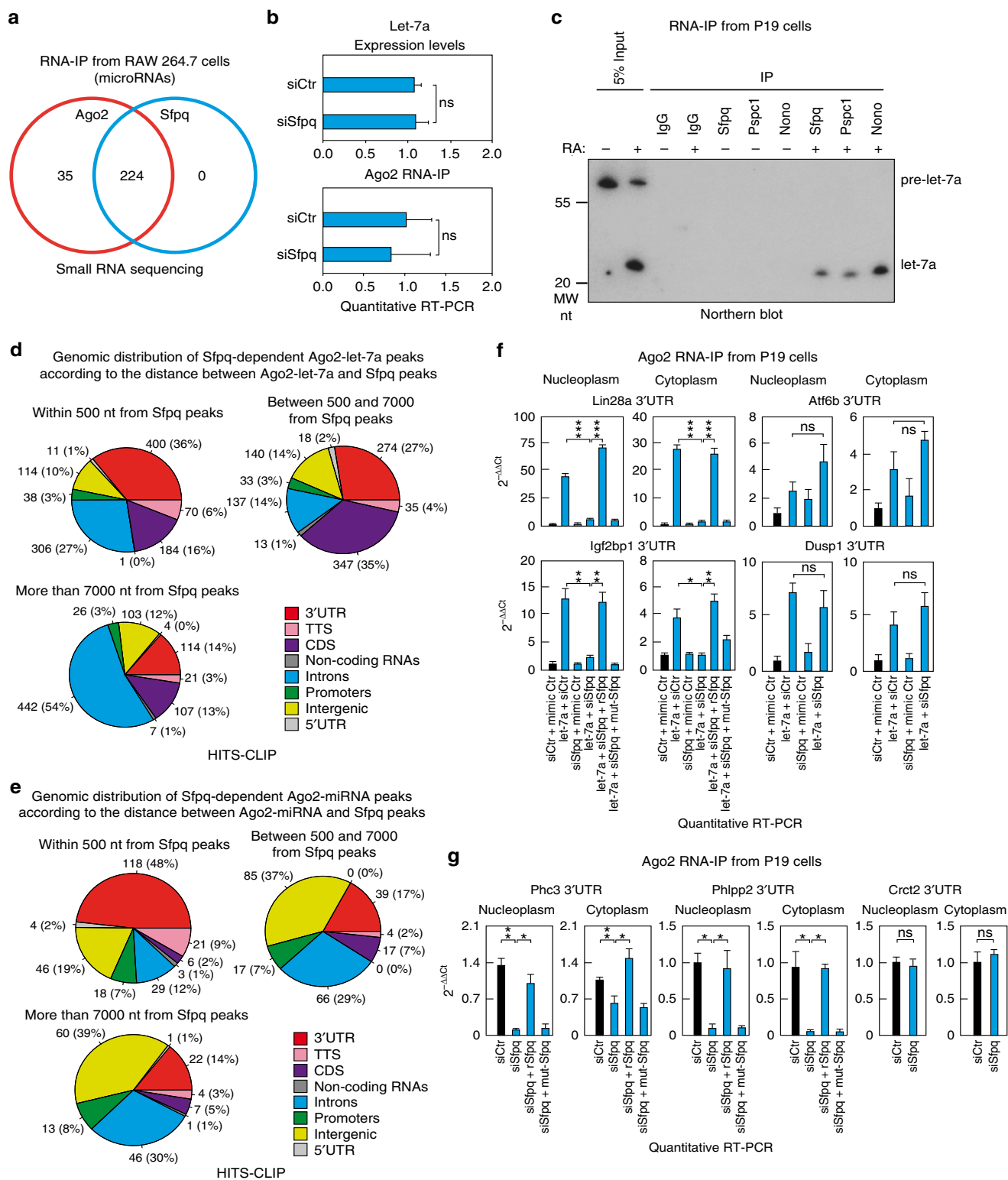
Therefore, we concluded that all the miRNAs associated with Sfpq are functionally active, given that they also associate with Ago2. These miRNAs include the let-7 family, miR-21, miR-31, miR-106b, and miR-23b. To validate this analysis, RNA-IP followed by either Northern blot or quantitative reverse-transcriptase PCR (RT-qPCR) analyses in both RAW 264.7 and HEK293T cells confirmed that let-7a, miR-23b, miR-125b, and miR-24 associate with Sfpq (Fig. 1f and Supplementary Figs. 2d, 5b). Notably, Sfpq only co-immunoprecipitated with mature miRNAs, not with precursors. These results indicate that Sfpq may be involved in the miRNA mode of action but not in miRNA biogenesis. Indeed, Sfpq knockdown did not affect miRNA expression or Ago2-loaded miRNA levels (Fig. 4b). Therefore, these results prompted us to investigate the possible role of Sfpq in regulating miRNA activity in gene silencing.

We hypothesized that similar to other RNA-binding proteins, Sfpq could modulate miRNA-binding activity. To test this hypothesis, we performed High-Throughput Sequencing of RNA isolated by CrossLinking IP (HITS-CLIP) on Ago2 in the presence or absence of Sfpq. We performed this analysis on ectopically expressed let-7a and endogenous miRNAs in stem cells. We selected let-7a as a model for the following reasons: (i) let-7 is the most associated miRNA with Sfpq (Supplementary Data 4); (ii) ingenuity pathway analysis from the Sfpq RNA-IP experiment followed by small RNA sequencing showed an

enrichment of cell cycle and cell development biological processes (Supplementary Fig. 5c), and let-7 controls both cell cycle and cell development<sup>26</sup>; (iii) both Sfpq and let-7 control nervous system development<sup>27, 28</sup>; and (iv) let-7 has been largely used as a model to study miRNA biology and biochemistry<sup>29</sup>. Although not expressed in stem cells, let-7 expression is induced upon retinoic acid (RA) stimulation to promote cell cycle arrest and differentiation<sup>30</sup>. As shown in Fig. 4c, Sfpq, Pspc1, and NonO co-immunoprecipitated with mature let-7a in RA-stimulated P19 cells, a cellular model for embryonic stem cells. Therefore, to investigate the impact of Sfpq on the modulation of let-7a-binding activity, we conducted HITS-CLIP mapping of Ago2 before and after Sfpq knockdown in let-7a-transfected P19 and control cells (Supplementary Data 5). Levels of ectopic Ago2-loaded let-7a in transfected P19 cells were comparable to that of endogenous miRNAs (Supplementary Fig. 5d), and its cellular localization was similar to that of RA-treated P19 cells (Supplementary Figs. 4a, 5e). Briefly, HITS-CLIP analysis identified 4202 Ago2-peaks specific for the let-7a condition (Ago2-let-7a peaks), which were distributed in different parts of the transcriptome, including the 3'UTR and protein coding sequence (CDS), as the main target substrates of let-7a (Supplementary Fig. 6a and Supplementary Data 5)<sup>1</sup>. To assess whether Sfpq directly regulates the identified Ago2-let-7a peaks, we first mapped the Sfpq-binding sites by comparing the Sfpq

HITS-CLIP in siSfpq-transfected P19 cells with sicontrol-transfected cells. Sfpq peaks identified by HITS-CLIP mainly co-localized with the Ago2 peaks in the 3'UTR (Supplementary Data 5 and Supplementary Fig. 6b). Hence, we calculated the distance between the Ago2-let-7a peaks and the closest Sfpq peaks. All the Ago2-let-7a peaks were then grouped into the

following three categories: (1) close to Sfpq peaks, within a distance of 500 nt; (2) far from Sfpq peaks, between a distance of 500–7000 nt; and (3) very far from Sfpq peaks, with a distance of 7000 nt or more (virtually not bound to Sfpq). We reasoned that the close distance would provide information about the local and direct roles of Sfpq in modulating Ago2-binding activity, whereas



far and very far distances would represent indirect effects of Sfpq knockdown, providing important internal controls for our analyses. Using this approach, we identified 1134 Ago2-let-7a peaks that were reduced in the absence of Sfpq, with a close distance to an Sfpq peak (Supplementary Fig. 6c). Among them, only those that mapped in the 3'UTR were significantly enriched compared with those with far or very far distances from Sfpq peaks (Fig. 4d; Fischer exact test  $p$ -value =  $2.2E-16$  for both comparisons). To determine whether 500 nt is the critical distance to define a direct regulation of Ago2-binding activity by Sfpq, we reasoned that within this distance, but not further, we should observe a significant reduction in Ago2-let-7a peaks upon Sfpq knockdown compared to a random distribution of the distance calculated computationally. Briefly, we divided the distance between Sfpq peaks and the reduced Ago2-let-7a peaks upon Sfpq knockdown from 0 to 1000 nt in bins of 250 nt each in the 3'UTRs. Then, we computationally shuffled 10,000 times the relative position of Sfpq and the reduced Ago2-let-7a peaks upon Sfpq knockdown in each bin and calculated the  $Z$ -score as a statistical test (Supplementary Fig. 6d). This analysis indicated that (i) at the closest distances between Sfpq and the Sfpq-dependent Ago2-let-7a peaks, the functional connections between Ago2 and Sfpq are greater and (ii) that this functional connection is significant up to a distance of 500 nt. Therefore, these data suggest that 500 nt is the maximal distance that can be considered to likely indicate a direct regulation of Ago2-binding activity by Sfpq. Importantly, among the peaks that are directly controlled by Sfpq, we found different let-7a-binding sites to those already validated in the mRNAs encoding for Lin28A<sup>31</sup>, Hic2<sup>32</sup>, Mier2, and Igf2bp1<sup>33</sup>. Similarly, we found that the direct reduction of the binding sites for the endogenously expressed miRNAs (endogenous Ago2-miRNA peaks; Supplementary Fig. 6e)<sup>34</sup> in P19 cells upon Sfpq knockdown statistically only occurred in the 3'UTR (Fig. 4e and Supplementary Fig. 6f; Fischer exact test,  $p$ -values =  $3.18E-13$  and  $9E-13$  for the close vs. far distance and close vs. very far distance comparisons, respectively). Importantly, we found that 24 and 8.2% of Sfpq-independent Ago2-let-7a peaks and endogenous Ago2-miRNA peaks, respectively, are located in the 3'UTR within a close distance to Sfpq peaks (<500 nt), suggesting that only a subset of miRNA-binding sites is directly controlled by Sfpq. To facilitate the accessibility and correct interpretation of all our HITS-CLIP results, we created an online database with a user-friendly interface containing the genomic coordinates for both the Ago2 and Sfpq peaks in each category we considered (<http://trabucchiab.unice.fr/SITO/index.php#>). Together, this analysis indicates that the 3'UTR is the substrate by which Sfpq can directly promote miRNA targeting at selected binding sites.

To validate these results, we performed Ago2 RNA-IP experiments followed by RT-qPCR. P19 cells were transfected with let-7a and/or with a different siSfpq from those used for the HITS-CLIP experiment (Supplementary Fig. 6g). We separated

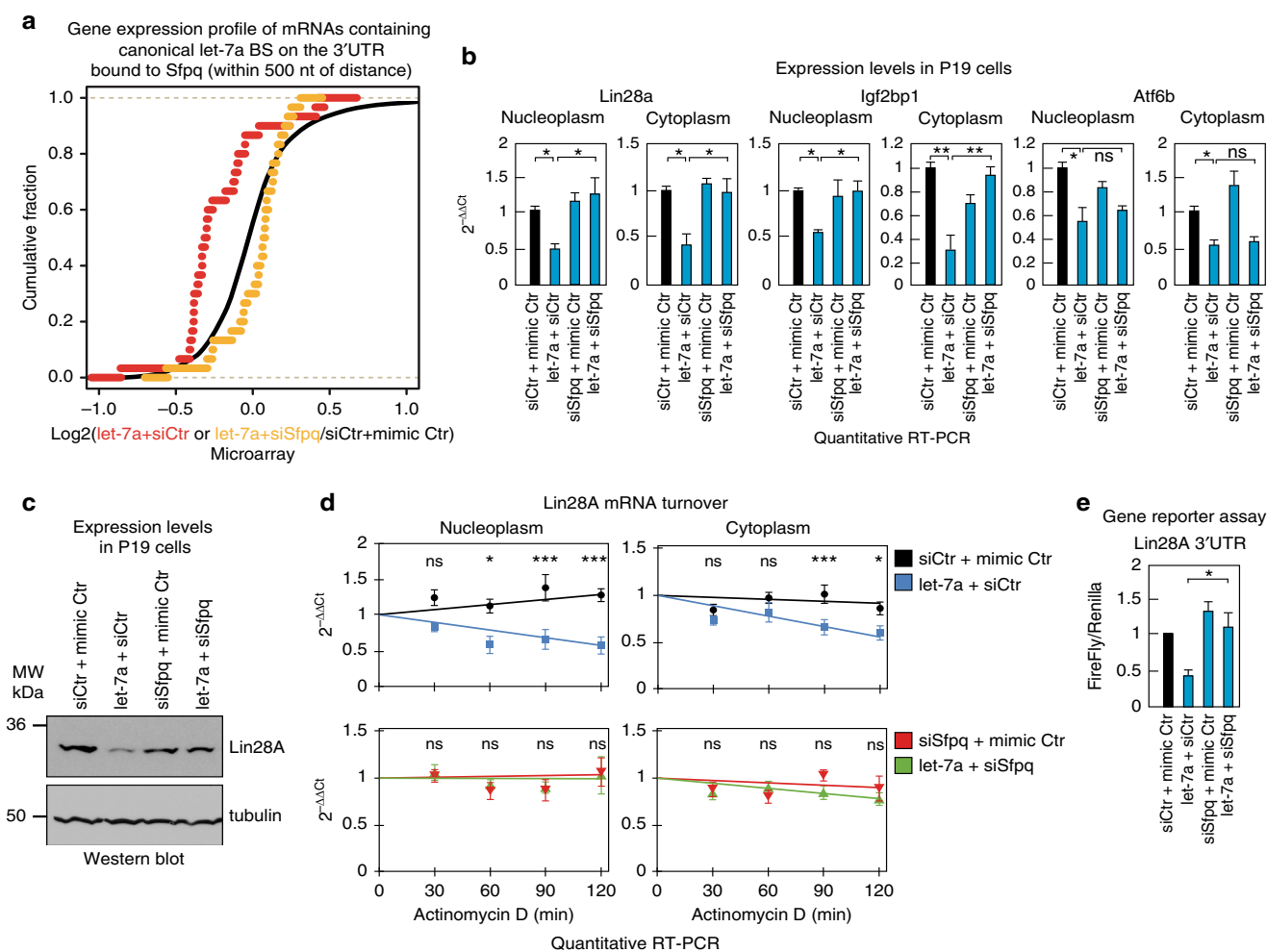
the nucleoplasm from the cytoplasm, partially digested the RNA and performed RNA-IP followed by RT-qPCR from each compartment. Interestingly, Sfpq knockdown impaired the activity of let-7a-binding sites on the Lin28A and Igf2bp1 3'UTRs in both the nucleoplasm and cytoplasm, but not on the Atf6b and Dusp1 3'UTRs, which were controls for Sfpq-independent let-7a-binding sites (Fig. 4f). By incubating the lysates from the siSfpq-let-7a-transfected cells with the recombinant wild-type Sfpq before RNA digestion, we were able to rescue this effect (Fig. 4f). The recombinant Sfpq mutant with L535, L539, L546, and M549 substituted to alanine was unable to rescue the miRNA targeting. The alanine mutations were designed to disrupt Sfpq ability to aggregate but not its binding activity, which was previously reported<sup>18</sup>. These data indicate that Sfpq binding and aggregation directly promote miRNA targeting at selected binding sites. Similar data were also obtained in human NTERA-2 stem cells and for the endogenous stem cell-specific miR-302b<sup>35, 36</sup> in both P19 and NTERA-2 cells (Fig. 4g and Supplementary Fig. 6h). Therefore, these data indicate that Sfpq directly controls a subset of miRNA-binding sites through local binding. Despite its nucleoplasmic localization, Sfpq enhances miRNA-binding activity in both the nucleoplasm and cytoplasm, suggesting that it promotes a nucleoplasmic commitment of mRNAs to globally control miRNA targeting.

#### Sfpq controls mRNA silencing by specific miRNA-binding sites.

To assess whether Sfpq-dependent miRNA targeting correlates with changes in target mRNA gene expression levels, we analyzed the mRNA expression profile in control and let-7a-transfected P19 cells upon Sfpq knockdown (Supplementary Data 6). Consistent with our HITS-CLIP data, Sfpq knockdown significantly rescued the magnitude of downregulation of those transcripts containing direct Sfpq-dependent Ago2-let-7a peaks in the 3'UTR (Supplementary Fig. 7a; Wilcoxon test,  $p$ -value =  $4E-7$ ), but not for other transcripts containing Sfpq-independent (far or very far distances between Ago2-let-7a and Sfpq peaks) or Sfpq-independent Ago2-let-7a peaks (Supplementary Fig. 7b). Because canonical miRNA-binding sites confer a more potent downregulation<sup>37</sup>, we analyzed the rescue of the silencing for the 12 Ago2-let-7a peaks that contain canonical let-7a-binding sites in the 3'UTR and whose binding is directly promoted by Sfpq, namely, Lin28A, Igf2bp1, Hic2, Mier2, Notch2, Map1b, Bbx, Skil, Lamp2, Tmem194, Ash1l, and Gns. As shown in Fig. 5a, in this case, the let-7a-mediated silencing and the rescue upon Sfpq knockdown was more striking than the full set of direct Sfpq-dependent target 3'UTRs (Wilcoxon test,  $p$ -value =  $4E-6$ ). This analysis was validated by performing RT-qPCR analyses on six of the selected let-7a target mRNAs, including four Sfpq-dependent and two Sfpq-independent targets, namely, Lin28A, Hic2, Mier2, Igf2bp1, Atf6b, and Dusp1 (Supplementary Fig. 7c). This

**Fig. 4** Sfpq promotes miRNA targeting at selected binding sites. **a** Venn diagram of Ago2 or Sfpq RNA-IP-enriched miRNAs found by small RNA sequencing analysis. **b** Let-7a expression levels (upper panel) and RNA-IP of Ago2 and let-7a (lower panel) in control or siSfpq-transfected RAW 264.7 cells. RNA extracts were analyzed by RT-qPCR. Data are presented as the mean  $\pm$  s.e.m. ( $n = 6$ ) and normalized to U2 snRNA or the input, respectively. **c** Sfpq, Pspc1, and NonO interact with mature let-7a in RA-treated P19 cells. Cell extracts were immunoprecipitated with the indicated antibodies and RNA was purified and analyzed by Northern blotting. Genomic distribution of either the **d** Ago2-let-7a or **e** endogenous Ago2-miRNA peaks decreased upon Sfpq knockdown, according to the distance to Sfpq peaks by HITS-CLIP analysis. These data show the prevalence of the Sfpq-dependent Ago2 peaks in the 3'UTR when Sfpq binds closely (<500 nt). RNA-IP of Ago2 and the indicated 3'UTRs for either **f** let-7a or **g** the endogenously expressed miR-302b-binding sites. P19 cells were transfected with the indicated molecules. Nucleoplasm and cytoplasm fractions were separated. The indicated cell lysates were incubated with either 100 nM full-length recombinant wild-type Sfpq or the Sfpq-214-598 quadruple mutant (L535A, L539A, L546A, and M549A) for 30 min at room temperature. Before IP, the lysate was partially digested with  $10 \mu\text{g ml}^{-1}$  RNase A for 30 min at room temperature. RNA was purified from the immunocomplexes and from 5% of the input and analyzed by RT-qPCR using oligonucleotide probes surrounding the miRNA-binding sites identified by HITS-CLIP. Data are presented as the mean  $\pm$  s.e.m. ( $n = 3$ ) and normalized to their own inputs. Student's  $t$ -test (for **b**) or one-way ANOVA followed by Tukey's post hoc test (for **f, g**) with  $*p < 0.05$  and  $**p < 0.01$ . ns not significant, TTS transcription termination site



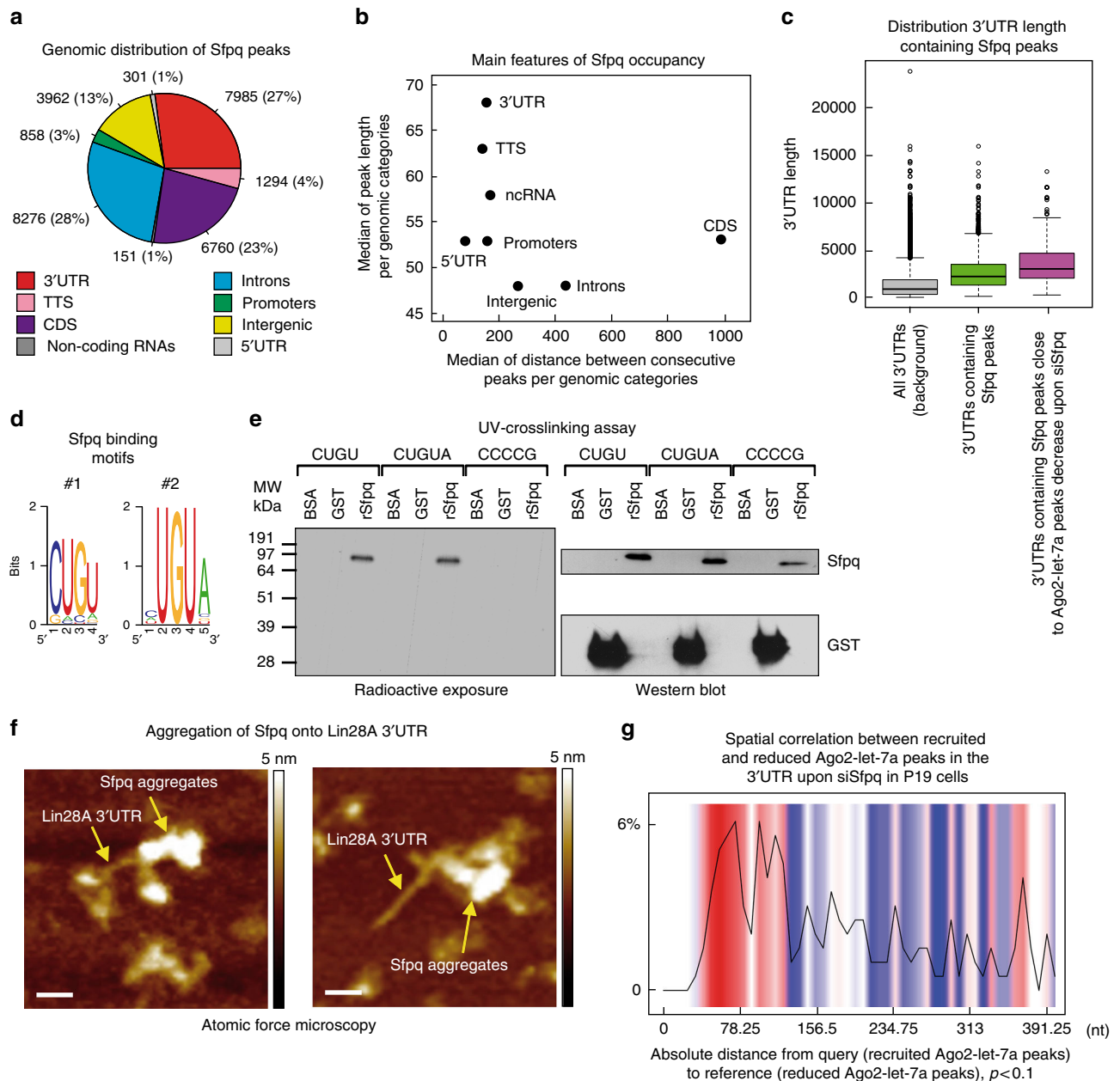


**Fig. 5** Sfpq promotes post-transcriptional silencing mediated by miRNAs in both nucleoplasm and cytoplasm. **a** Gene expression differences in let-7a-transfected P19 cells upon Sfpq knockdown and control. Differential expression is plotted for mRNAs containing canonical let-7a-binding sites (BS) localized in the 3'UTR within a close distance to Sfpq peaks (<500 nt) and reduced upon Sfpq knockdown. **b** P19 cells were transfected with the indicated molecules. Nucleoplasm and cytoplasm were separated to measure the expression levels of the indicated let-7a-target mRNAs. RNA was purified and analyzed by RT-qPCR. Data are normalized with U2 snRNA and presented as the mean  $\pm$  s.e.m. ( $n = 3$ ). **c** Immunoblot analysis of Lin28A and tubulin in P19 cells transfected with let-7a and/or siSfpq. **d** Quantitative RT-PCR analysis of the half-life of Lin28A transcript in let-7a-transfected P19 cells compared to control (upper panels) or siSfpq + let-7a-transfected cells compared to siSfpq control (lower panels). Total RNA from either nucleoplasm (left panels) or cytoplasm (right panels) was isolated at the indicated times after addition of actinomycin D. Data are normalized with  $\beta$ 2-microglobulin and presented as the mean  $\pm$  s.e.m. ( $n = 6$ ). **e** Relative luciferase activity of reporter constructs containing the mouse Lin28A 3'UTR sequence in HEK293T cells transfected with let-7a and siSfpq as indicated. The data were normalized using Renilla activity and presented as the mean  $\pm$  s.e.m. ( $n = 4$ ). One-way ANOVA followed by Tukey's post hoc test: \* $p < 0.05$ , \*\* $p < 0.01$ , ns not significant

validation was performed with a different siRNA from that used for the transcriptomic analysis. Similarly, Sfpq knockdown significantly upregulated the steady-state expression of those transcripts that contain direct Sfpq-dependent Ago2 peaks with canonical binding sites for the 20 most expressed endogenous miRNAs in the 3'UTR (Supplementary Fig. 7d; Wilcoxon test,  $p$ -value = 0.04769), but not that of the other transcripts containing indirect Sfpq-independent Ago2 peaks with canonical miRNA-binding sites (Wilcoxon test,  $p$ -value = 0.07729 for far and 0.9939 for very far distances, respectively; Supplementary Fig. 7d). However, no significant upregulation was observed upon Sfpq knockdown when we considered the whole set of mRNAs with endogenous Ago2 peaks containing canonical or non-canonical miRNA-binding sites (Supplementary Fig. 7e). The absence of any siSfpq effects on the whole data set of endogenous Ago2

peaks could be due to the very mild downregulation conferred by endogenous non-canonical miRNA-binding sites<sup>37</sup>.

As shown in Fig. 5b and Supplementary Fig. 8a, the downregulation of selected direct Sfpq-dependent mRNAs targeted by let-7a or by the endogenously expressed miR-302b was rescued in both the nucleoplasm and cytoplasm by Sfpq knockdown, but not the Sfpq-independent miRNA-target mRNA control Atf6b mRNA. Similar results were also obtained in human NTERA-2 cells (Supplementary Fig. 8b). This Sfpq-dependent mechanism of miRNA targeting controls let-7a-dependent differentiation programs in stem cells (Supplementary Fig. 8c). Therefore, these data support the hypothesis that Sfpq plays a role in regulating miRNA silencing on specific binding sites, which globally impact miRNA-dependent gene expression programs.



**Fig. 6** Sfpq aggregates onto long 3'UTRs to modulate the accessibility of miRNA-binding sites. **a** Genomic distribution of Sfpq peaks by HITS-CLIP analysis. **b** Scatter plot of length and frequency of Sfpq peaks in different genomic regions. **c** Distribution of the 3'UTR length of the indicated subpopulation of transcripts. **d** Result of the bioinformatic analysis for de novo search of Sfpq-binding motif from Sfpq HITS-CLIP data. **e** UV-crosslinking assay to analyze the interaction of the recombinant Sfpq (100 nM) with the  $^{32}\text{P}$ -labeled RNA oligonucleotides containing two copies of the CUGU or CUGUA, respectively, but not to the CCCCCG negative control sequence. **f** Top view of two topographic AFM images of typical Sfpq-Lin28A 3'UTR complexes. The color bar on the right represents the height scale with a maximum corresponding to 5 nm. Scale bar corresponds to 20 nm. **g** Spatial correlation of the recruited Ago2-let-7a peaks compared to the decreased ones. In the x-axis is reported the distance in nt between the reduced and the new recruited peaks on the 3'UTR, whereas in the y-axis is reported the percentage of recruited peaks. The black line is the data density, the red bars indicate that the number of recruited peaks is higher than that expected by chance, the blue bars indicate the opposite, whereas the white bars indicate no recruitment

To further inspect the direct role of Sfpq in controlling miRNA targeting in cells, we focused on a particular known let-7a target—the oncogene Lin28A, also known as *Lin-28 homolog A*, which regulates the self-renewal of stem cells and cancer stem cells<sup>38</sup>. Sfpq knockdown in either P19 or NTERA-2 cells abrogated the let-7a-mediated decrease in the steady-state levels of Lin28A protein in either P19 or NTERA-2 cells (Fig. 5c and Supplementary Fig. 9a). According to a previous work<sup>31</sup>, we observed a shorter half-life for endogenous nucleoplasmic and cytoplasmic

Lin28A mRNA in let-7a-transfected P19 cells compared with the control cells. This shortened half-life was rescued by Sfpq knockdown (Fig. 5d). Nuclear run-on (NRO) experiments ruled-out a transcriptional modulation of Lin28A by let-7a or Sfpq (Supplementary Fig. 9b, c). Overall, these data indicate that Sfpq globally modulates the Lin28A mRNA turnover rate through let-7a, resulting in a significant decrease in the steady-state expression levels of the mRNA protein.

To further confirm the direct involvement of Sfpq in controlling let-7a silencing activities on the Lin28A 3'UTR, we used a reporter plasmid in which the Lin28A 3'UTR was cloned downstream of a luciferase open reading frame. Overexpression of let-7a significantly reduced the luciferase activity in HEK293T cells transfected with the reporter plasmid, which was rescued upon Sfpq knockdown (Fig. 5e). However, siSfpq failed to produce any effects on a reporter construct containing only the six let-7a-binding sites (Supplementary Fig. 9d) and on the empty vector (Supplementary Fig. 9e). By contrast, Sfpq knockdown also failed to produce any effects on the miR-125b-mediated downregulation of the Lin28A 3'UTR<sup>39</sup> (Supplementary Fig. 9f), confirming the presence of Sfpq-independent binding sites even with close Sfpq peaks, which was also observed by the HITS-CLIP analysis. Moreover, Sfpq-binding activity on the Lin28A mRNA is independent of let-7a activity (Supplementary Fig. 9g). Overall, these data suggest that Sfpq binding to specific cis-elements on the Lin28A 3'UTR optimizes the positioning/recruitment of miRNAs to selected binding sites.

**Sfpq aggregates on target 3'UTRs to promote miRNA targeting.** To gain insights into the mechanism by which Sfpq regulates specific miRNA-binding sites in the 3'UTR, we investigated the Sfpq HITS-CLIP data in P19 cells to uncover the presence of any peculiar feature(s) of Sfpq-binding activity. Briefly, we found approximately 30,000 Sfpq peaks, with the majority of them mapping in introns and 3'UTRs (Fig. 6a). Interestingly, Sfpq peaks in the 3'UTRs are more elongated and closer to one another than those that map in other genomic regions (Fig. 6b). Transcription termination sites (TTS) and non-coding RNAs show similar trends, whereas Sfpq peaks in the introns show opposite features. Additionally, as shown in Fig. 6c, Sfpq occupancy is enriched in long 3'UTRs compared to the entire 3'UTR data set (Wilcoxon test,  $p$ -value = 2.2E-16). The 3'UTR length was even longer when we selected the 3'UTRs that contain at least one Sfpq peak and Sfpq-dependent let-7a-binding sites (Wilcoxon test,  $p$ -value = 2.2E-16 compared to the length of 3'UTRs-containing Sfpq peaks). These data suggest that Sfpq preferentially binds to long 3'UTRs, forming long portions of occupancy.

Next, we looked for specific Sfpq-binding motifs. Because RRM domains bind an average of 4 nt<sup>40</sup>, we looked for motifs of 4–6 nt in length in Sfpq peaks from the HITS-CLIP data set. Briefly, we searched for de novo motifs by dividing the Sfpq HITS-CLIP data set into the following three groups: (i) all Sfpq peaks; (ii) Sfpq peaks located in the 3'UTR; and (iii) Sfpq peaks within a close distance (<500 nt) to the Sfpq-dependent Ago2 peaks. In this analysis, we have found enriched motifs with a core composed of the UGU sequence (Supplementary Fig. 10a). Using motif clusterization<sup>41</sup>, we obtained two 4- and 5 nt-enriched consensus motifs (Fig. 6d). To validate these two motifs, we used recombinant Sfpq and synthetic RNA in UV-crosslinking and electrophoresis mobility shift (EMSA) assays (Fig. 6e and Supplementary Fig. 10b). BSA, GST, and Ago2, which were used as controls, did not show any binding activity to the newly discovered Sfpq-binding sequences; neither did a negative control sequence with recombinant Sfpq (Fig. 6e and Supplementary Fig. 10b, c). Importantly, these two binding motifs were significantly enriched in the Sfpq peaks located in the 3'UTR that map very close to Sfpq-dependent Ago2-let-7a peaks with respect to all the 3'UTR Sfpq peaks (Fisher exact test,  $p$ -value = 2.2E-16).

Because Sfpq aggregates to form polymers on nucleic acids<sup>17, 18</sup> to facilitate miRNA targeting (Fig. 4f, g), we hypothesized that the two Sfpq-binding motifs would serve as substrates to promote

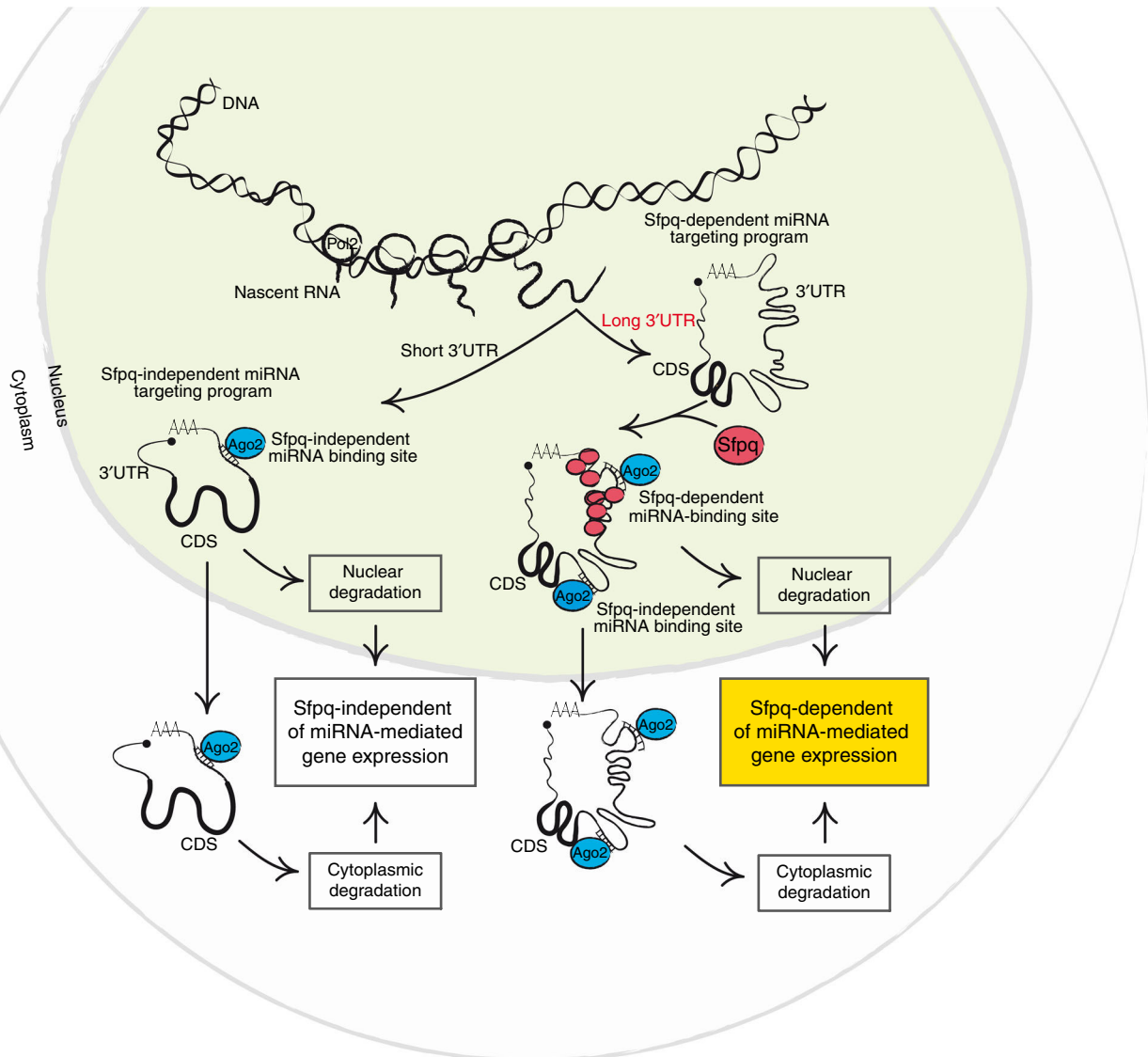
Sfpq aggregation on target 3'UTRs. To test this hypothesis, we performed Pearson and Spearman correlation tests between the number of occurrences of binding motifs within the peak sequences and the peak length. We found a higher positive correlation in Sfpq peaks that map within 500 nt from Sfpq-dependent Ago2-let-7a peaks and their length (Spearman correlation for motif#1: 0.56; Pearson correlation for motif#1: 0.85; Spearman correlation for motif#2: 0.54; and Pearson correlation for motif#2: 0.79), than in all the Sfpq peaks in the 3'UTR (Spearman correlation for motif#1: 0.55; Pearson correlation for motif#1: 0.72; Spearman correlation for motif#2: 0.48; and Pearson correlation for motif#2: 0.62), or in all the Sfpq peaks (Spearman correlation for motif#1: 0.48; Pearson correlation for motif#1: 0.65; Spearman correlation for motif#2: 0.42; and Pearson correlation for motif#2: 0.58). This analysis indicated that the highest frequency of Sfpq-binding sites occurred close to direct Sfpq-dependent Ago2 peaks, thus supporting the hypothesis that Sfpq has a tendency to form long aggregates that can ultimately promote the positioning/recruitment of miRNAs on selected and close binding sites. In fact, the Lin28A 3'UTR contains 34 Sfpq-binding sites, and recombinant wild-type Sfpq forms aggregates along its sequence, as we demonstrated using AFM (Fig. 6f and Supplementary Fig. 11a), but not the recombinant Sfpq-214–598 quadruple mutant (L535A, L539A, L546A, and M549A), which was unable to aggregate while binding to the RNA (Supplementary Figs. 10c, 11b)<sup>18</sup>. BSA did not bind to Lin28 3'UTR, which is removed from the untreated muscovite mica surface in the washing step (Supplementary Fig. 11c). Furthermore, in support of this conclusion, our HITS-CLIP analyses demonstrated that 70% of the 3'UTRs in which Ago2-let-7a peaks are directly controlled by Sfpq (Sfpq peaks within 500 nt of distance) undergo recruitment of new Ago2 peaks upon Sfpq knockdown, whereas for the endogenous Ago2-miRNA peaks, this value is 98% (Supplementary Data 7). Spatial correlation between the recruited and reduced Ago2 peaks upon Sfpq knockdown indicated that recruitment mainly occurs close to the Sfpq-dependent Ago2 peaks (Fig. 6g). These data suggest that the presence of Sfpq locally modulates miRNA accessibility/positioning to specific binding sites by forming aggregates that may modulate the secondary structure of the target 3'UTR.

Finally, RNA-sequencing analysis of control and siSfpq-transfected P19 cells ruled out any roles for Sfpq in regulating alternative splicing or alternative polyadenylation sites in the 3'UTRs in which Sfpq directly promotes miRNA targeting (Supplementary Fig. 12a, b, and Supplementary Data 8).

Overall, these data indicate that the presence of specific Sfpq-binding sites determines the fate of a cohort of mRNAs, where Sfpq forms long aggregates in the 3'UTR to modulate 3'UTR folding for the proper positioning/recruitment of miRNAs to selected binding sites, whereas avoiding random binding of miRNAs that would not be effective.

## Discussion

The interplay between miRNAs and RNA-binding proteins has been dubbed “the post-transcriptional regulatory code”; by interacting and competing to binding sites, these post-transcriptional modulators dictate metabolic impact on cognate RNAs to regulate stability, localization, and protein synthesis<sup>8</sup>. Our study added Sfpq to the list of other RNA-binding proteins, including HuR, Pumilio, polypyrimidine tract-binding protein, cytoplasmic polyadenylation element-binding proteins, Dnd1, and RNA-binding motif protein 38 (RBM38), that have been shown to be involved in regulating miRNA targeting in mammalian cells<sup>42</sup>. In comparison with previous studies where specific RNA-binding proteins appear to regulate some cytoplasmic



**Fig. 7** A model for the Sfpq-dependent control of miRNA targeting

miRNA targeting, we found that Sfpq targets a sizeable subset of long nuclear 3'UTRs to promote miRNA targeting.

Previous studies have reported the presence of endogenous miRNA pathway components in the nucleus<sup>6, 43, 44</sup>. According to these reports, miRNAs are loaded into miRISC in the cytoplasm and then imported into the nucleus by Importin 8. Indeed, although miRISC and its factors TNRC6 and the CCR4-NOT deadenylase complex are localized in the nucleus, miRISC loading factors, such as Hsp90, are absent<sup>6</sup>. Our data indicate that Sfpq specifically binds to selected long 3'UTRs and forms aggregates that likely modulate the secondary structure of target 3'UTRs to promote optimal positioning/recruitment functions of miRISC to specific binding sites, which leads to miRNA-dependent mRNA degradation. Because Sfpq interacts in an RNA-dependent fashion only with nucleoplasmic miRISC, but not on the chromatin, we concluded that Sfpq only associates with miRNA-target mRNAs post-transcriptionally (Fig. 7). This mechanism is conserved across species and in different cell types (Supplementary Fig. 13a), indicating that our results have uncovered a general strategy by which cells possess a specific pathway to control miRNA targeting in long 3'UTRs via an Sfpq-dependent mechanism. This finding contributes to the diversity of miRNA

modes of action and impacts on post-transcriptional gene expression regulation. Whether the partners of Sfpq, namely, Pspc1, NonO, and the long non-coding RNA Neat1, are part of this mechanism is still an open question.

A plethora of observations accumulated over many years strongly suggest that there is an intimate link between the various steps in the post-transcriptional gene expression pathway, including mRNA export, translation, stability, and localization<sup>45</sup>. Here, we demonstrated that nucleoplasmic Sfpq determines the fate of a sizeable set of mRNAs by acting at the 3'UTR and influencing both nucleoplasmic and cytoplasmic miRNA targeting, including with Lin28A mRNA in stem cells. These data mechanistically reveal for the first time an Sfpq-dependent link between nuclear and cytoplasmic miRNA-dependent silencing (Fig. 7). In conclusion, our findings widen the investigation of post-transcriptional silencing beyond the traditional cytoplasmic point of view to include a nucleoplasmic mechanism.

## Methods

**Proteomic analysis.** The antibody we used to immunoprecipitate Ago2 has been extensively used to immunoprecipitate endogenous Ago2 in several HITS-CLIP analyses, including<sup>46, 47</sup>, but it has not been used for MS analysis. IP by IgG and



protein G served as controls, as well as just the anti-Ago2 antibody not incubated with cell lysate. As a quality control for the specificity of the IP, we tested the immunoprecipitated pellet by Northern blotting against let-7a (Supplementary Fig. 1c) and by Western blotting against Ago2 (Supplementary Fig. 1d). The immunopellet from either undigested or totally digested RNA was further treated for MS analysis.

To determine the RNA dependency of the Ago2 interactors identified, we summed the abundance scores for each identified protein from the three replicates in each condition as one single value and used the base 2 logarithm of the ratio of the abundances of the (+)RNase and (-)RNase conditions. As a standard method for RNA-dependent association with the immunoprecipitated protein (Ago2), we normalized the RNA-dependency ratio to set the ratio of Ago2 itself to one based on the assumption that enrichment of the immunoprecipitated protein itself should not be RNA-dependent. According to Klass et al.<sup>48</sup>, we set up a cutoff for grouping the proteins as RNA-dependent or RNA-independent interactors by generating a null distribution by modeling RNA-independent associations (red line in Supplementary Fig. 13b). To this goal, we made the following two assumptions: (i) after normalization, any RNA-dependent values more than zero have a true value of zero and the observed variation from zero is due to noise; and (ii) the noise is symmetric about zero. Using these two assumptions, we took the RNA-dependence values higher than zero and built the null distribution symmetric about zero. Thus, we calculated the LOD (log of odds) for each interactor in the RNA-dependent distribution (black line in Supplementary Fig. 13b) vs. the null distribution (red line). Those interactors with a greater likelihood of being in the RNA-dependent distribution (LOD < 0) were considered Ago2 interactors in an RNA-dependent fashion (dashed line in Supplementary Fig. 13b and Supplementary Data 3).

**Mass spectrometry.** For MS, proteins from immunoprecipitated samples were loaded on Bis-tris acrylamide gels (Life Technologies) and migrated as soon as proteins stacked in a single band. Proteins were stained with Imperial Blue (Pierce, Rockford, IL), excised from the gel and digested with high sequencing grade trypsin (Promega, Madison, WI, USA). Then, we performed MS analysis according to Shevchenko et al.<sup>49</sup>. Briefly, gel excisions were washed and destained with 100 mM (NH<sub>4</sub>)HCO<sub>3</sub>, shrunk using 100 mM ammonium bicarbonate in 50% acetonitrile and let them dry at room temperature. After rehydration using 10 mM DTT in 25 mM ammonium bicarbonate pH 8.0 for 45 min at 56 °C the solution was replaced by 55 mM iodoacetamide in 25 mM ammonium bicarbonate pH 8.0 for 30 min at room temperature in the dark. Excised gels were then washed twice in 25 mM ammonium bicarbonate and shrunk for 5 min in 25 mM ammonium bicarbonate in 50% acetonitrile. They were then reswollen in 25 mM ammonium bicarbonate pH 8.0 supplemented with 12.5 ng µl<sup>-1</sup> trypsin (Promega) for 1 h at 4 °C and incubated overnight at 37 °C. Finally, digested peptides were harvested by two extractions: the first in 5% formic acid and the second one in 5% formic acid in 60% acetonitrile. Both extracts were pooled and dried.

The resulting samples were injected in duplicate in 0.1% TFA 4% acetonitrile and analyzed by liquid chromatography (LC)-tandem mass spectrometry (MS/MS) in an LTQ-Orbitrap-Velos (Thermo Electron, Bremen, Germany) online with a nanoLC Ultimate 3000 chromatography system (Dionex, Sunnyvale, CA). Peptides were separated on a Dionex Acclaim PepMap RSLC C18 column, according to the manufacturer's instructions. For peptide ionization in the nanospray source, spray voltage was set at 1.4 kV and the capillary temperature at 275 °C. Instrument method for the Orbitrap Velos was set up in data-dependent mode to switch consistently between MS and MS/MS. MS spectra were acquired with the orbitrap in the range of *m/z* 400–1700 at a FWHM resolution of 30,000 measured at 400 *m/z*. 445.120025 ions were used as lock mass for internal calibration. The 10 abundant precursor ions were selected and collision-induced dissociation fragmentations were performed in the ion trap on the 10 most intense precursor ions measured to have maximum sensitivity and maximum amount of MS/MS data. 500 counts was set as signal threshold for an MS/MS event. Charge state screening was enabled to exclude precursors with 0 and 1 charge states. Dynamic exclusion was enabled with a repeat count of 1, exclusion list size 500 and exclusion duration of 30 s.

**Bioinformatic analysis.** Raw files generated from MS analysis were processed with Proteome Discoverer 1.4 (Thermo Fisher Scientific). This software was used to search data via in-house Mascot server (version 2.4.1; Matrix Science Inc., London, UK) against the mouse subset (16,718 sequences) of the SwissProt database (2015\_06). Database search were done using the following setting: a maximum of two trypsin miscleavages allowed, methionine oxidation, N-terminal protein acetylation as variable modifications, and cysteine carbamido-methylation as fixed modification. A peptide mass tolerance of 6 ppm and a fragment mass tolerance of 0.8 Da were used for search analysis. Only peptides with high stringency Mascot score threshold (identity, false discovery rate (FDR) <1%) were selected and used for protein identification (unique peptides).

Relative intensity-based label-free quantifications (abundance scores) were processed using Progenesis LC-MS software (version 4.1; Nonlinear Dynamics, Newcastle, UK), according to the manufacturer's instruction (Nonlinear Dynamics, Newcastle, UK). First, raw LC Orbitrap MS data were imported and LC-MS heatmap of retention time and *m/z* were generated. Features from these LC-MS were automatically aligned and filtered to retain signals crossing the following

parameters; retention time 10–110 min, *m/z* 300–1700, charge state 2–6 and 3 or more isotopes. MS-MS spectra were exported into peak list as Mascot generic files (MGF) to allow protein identification using the following inclusion options: the three highest intensity precursors for each feature and precursor intensity over 25%. MGF files were used to search data via in-house Mascot server version 2.4.1 (Matrix Science Inc., London, UK) against the Mouse database subset of the SwissProt database (version 2015\_06). Only peptides (unique peptides) adjusted to 1% FDR (identity) and with ion score cut-off of 30 were exported from Mascot results and imported back to Progenesis LC-MS for protein grouping and quantification. Total ion intensity signal from each of the individual peptides generated protein quantification. Any conflicting peptide identifications were removed from the measurements of the quantified proteins. Univariate one-way analysis of variance (ANOVA) was performed within Progenesis LC-MS to calculate the protein *p*-value according to the sum of the normalized abundances across all runs.

We analyzed the enrichment of gene ontology (GO) terms, using the tool GoTermFinder<sup>50</sup>, and protein domains by protein families database, using the tool SMART<sup>51</sup> and python in-house scripts (Supplementary Data 3). The majority of Ago2 interactors identified were RNA-binding proteins. Although, our data clearly demonstrated an enrichment of RNA-related terms and domains for both RNA-dependent and RNA-independent interactors of Ago2, the proteins showing RNA-dependency were particularly enriched in post-transcriptional regulation terms and molecular functions, such as mRNA stabilization and mRNA binding (Supplementary Fig. 1a). In addition, RNA-dependent interactors were significantly enriched for known RNA protein-binding domains, including RRM and K Homology domain (Supplementary Fig. 1b), demonstrating the efficacy of the method used here.

**HITS-CLIP method and analysis.** We used a protocol according to Darnell's laboratory<sup>52</sup> and further modified by Wang et al.<sup>53</sup>. Briefly, 10 cm diameter dishes of P19 cells were transfected with combinations of mimic let-7a, siSfpq, siCtr, or mimic Ctr for 48 h to select the following conditions: (i) sicontrol (siCtr) and mimic Ctr; (ii) siCtr and mimic let-7a; (iii) siSfpq and mimic Ctr; (iv) siSfpq and mimic let-7a. All four conditions were used to perform Ago2 HITS-CLIP analysis, whereas only the first and the third conditions were used to perform Sfpq HITS-CLIP analysis. Afterward, cells were irradiated once at 400 mJ cm<sup>-2</sup> with 254 nm UVC light and lysate were then RNase digested using 10 µg ml<sup>-1</sup> of RNase I. Crosslinked Ago2 or Sfpq were recovered via overnight IP at 4 °C with 1 µg of either monoclonal anti-mouse Ago2 antibody (Wako Chemicals) or a mixed of two anti-Sfpq antibodies (Ab38148 from Abcam and B92 from Fischer Scientific) complexed with Protein-G Dynabeads. Immunoprecipitates were washed three times with High-Salt Buffer and PNK buffer, end labeled with 25 µCi <sup>32</sup>P gamma-ATP using PNK for 20 min at 37 °C then with 1 mM of cold ATP for 20 min at 37 °C. After three washes with PNK buffer the protein-RNA complexes were eluted from the beads using 1 × NuPage Loading buffer supplemented with 10% of 2-mercaptoethanol at 70 °C for 10 min. Protein-RNA complexes were then resolved on a 10% Bis-Tris Gel and transferred to nitrocellulose. Membranes were exposed onto a cassette at -80 °C for 1–2 h to obtain an autoradiograph film. Subsequently, protein-RNA complexes migrating in the 90–130 kD or 75–110 kD ranges for Ago2 and Sfpq, respectively, were excised (Supplementary Fig. 13c, d). RNA was recovered from the nitrocellulose using Proteinase K treatment followed by phenol-chloroform extraction and ethanol precipitation. Isolated RNA was subjected to small RNA library preparation with the CleanTag™ Ligation Kit for Small RNA Library Prep (Trilink Biotech) using a ¼ dilution of 3' and 5' adapters. cDNA products were subjected to PCR amplification for 20 or 22 cycles for Sfpq and Ago2 experiments, respectively, and were then purified on Purelink PCR micro kit columns (Invitrogen). PCR products represented cDNA inserts of 20–50 nt. Libraries were quantified using both bioanalyzer (Agilent) and Qubit (Invitrogen) and were subjected to sequencing by Ion Proton sequencer (Life Technologies).

**Bioinformatic analysis.** Reads were mapped onto the mm10 genomes using STAR<sup>54</sup>, then unmapped reads were mapped with bowtie2<sup>55</sup>. Total number of reads mapped are reported in Supplementary Data 5. The reproducibility of the replicates for each condition was assessed by performing principal component analysis using the R package "htSeqTools" (R package version 1.14.0). Afterward, replicates were merged together to assess bioinformatics analysis. Ago2 peaks specific of let-7a-transfected cells condition were found by comparing let-7a-transfected samples with the control using dCLIP program<sup>56</sup>. We filtered out peaks with score less than 7 and have identified 4202 Ago2-differential binding peaks (here called Ago2-let-7a peaks, Supplementary Data 5). To identify let-7a canonical-binding sites, we screened the identified Ago2-let-7a peaks sequences in the 3' UTR looking for the reverse complement of the seed sequence (UACCUC). Identical approach was used to identify the canonical-binding sites for the 21 most endogenously expressed miRNAs (Supplementary Data 5). To map Sfpq-binding sites we used dCLIP program by comparing Sfpq HITS-CLIP in siSfpq-transfected P19 cells with siCtr-transfected cells. We only selected peaks with a score greater than 7.

To investigate the potential role of Sfpq to promote let-7a targeting, we checked whether the identified Ago2-let-7a peaks were reduced in absence of Sfpq. We first selected the Ago2-let-7a peaks enriched in let-7a-transfected P19 cells condition

compared to let-7a-siSfpq-transfected P19 cells by using dCLIP program. Second, we selected Ago2-let-7a peaks that do not show any change when comparing siSfpq-transfected P19 cells with the siCtr condition by using dCLIP program. We reasoned that in the first filter we identify all Sfpq-dependent Ago2-let-7a peaks, whereas in the second filter we double check the real let-7a dependency of the identified peaks. To merge the information from the two filters, we measured the overlap between the full set of 4202 Ago2-let-7a peaks with the peaks coming from these two filters by calculating Jaccard index (minimal threshold index of 0.10 for overlapping peaks). At the end, we found that 3381 Ago2-let-7a peaks overlap with the peaks coming from these two filters, indicating their dependency on Sfpq.

To assess the Sfpq dependency of the endogenous miRNA-binding sites, we selected Ago2 peaks that overlap (Jaccard index > 0.10) between the full set of 2065 Ago2-miRNA peaks identified by HITS-CLIP analysis from P19 cells<sup>34</sup> and the Ago2-miRNA peaks enriched in siCtr-transfected P19 cells over the siSfpq-transfected P19 cells condition identified by dCLIP analysis. We found that 245 Ago2-miRNA peaks overlap with the peaks coming from such a filter, indicating their dependency on Sfpq.

Ago2-let-7a peaks recruited in absence of Sfpq were identified by comparing siSfpq-let-7a-transfected P19 cells to let-7a-transfected P19 cells with dCLIP analysis and selecting peaks with a score greater than 7. For bindings sites of endogenously expressed miRNAs, we analyzed Ago2-miRNA peaks in siSfpq-transfected P19 cells using pyoclip<sup>57</sup> with the same protocol as described in Bottini et al.<sup>34</sup>. In Supplementary Data 7 is shown the recruited Ago2-let-7a or Ago2-miRNA peaks in siSfpq-transfected P19 cells.

Spatial correlation of the Sfpq-dependent recruited Ago2 peaks (for both let-7a and endogenous miRNAs) compared to the decreased ones was performed with the R package GenometricCorr<sup>58</sup>. Briefly, we found that the two sets of Sfpq-dependent recruited and reduced Ago2 peaks are spatially correlated across the genome, calculated the significance as a deviation from a non-uniform distribution of one set of peaks over to the other one.

All analyses were performed with custom python scripts where not specified elsewhere.

**User-friendly website for Ago2 and Sfpq CLIP data sets.** We created a database in MySQL language with a web interface developed in php language to easily access to the entire mapping of Sfpq, Ago2-let-7a, or Ago2-miRNA peaks identified by HITS-CLIP analysis. The database can be interrogated by the following four different keys: (1) “gene identifier”, which provides all peaks on the requested gene; (2) “genomic category”, which provides all peaks that map in specific genomic category, including 3'UTR, 5'UTR, CDS, introns, TTS, promoters, intergenic, and ncRNA; (3) “miRNA name”, to access to all peaks bearing the signature of let-7a or the 21 most expressed miRNAs in P19 cells (Supplementary Data 5); and (4) “peak category”, which provides all peaks that belong to a specific category as assessed by this study, including Sfpq peaks, Ago2 peaks, Sfpq-dependent decreased Ago2 peaks localized within 500 nt from Sfpq peaks, Sfpq-dependent decreased Ago2 peaks localized between 500 and 7000 nt from Sfpq peaks, and Sfpq-dependent decreased Ago2 peaks localized more than 7000 nt from Sfpq peaks. One or more keys can be combined to explore the data from different points of view. The database is freely accessible at <http://trabuchilab.unice.fr/SITO/index.php#>.

**Search for Sfpq-binding motifs and clusterization.** To identify the putative binding motif(s) of Sfpq that would play a role in miRNA targeting in the 3'UTR, we divided the data set of Sfpq peaks from the HITS-CLIP analysis in the following three subsets: (i) all Sfpq peaks; (ii) Sfpq peaks located in the 3'UTR; and (iii) Sfpq peaks located in the 3'UTR within a distance of 500 nt from the Sfpq-dependent reduced Ago2-let-7a peaks. Then, we launched de novo motif discovery with a length between 4 and 6 nt. We used two motif finding programs: Homer<sup>59</sup> and DREME<sup>60</sup>. We observed that all these analyses reported significantly enriched motifs with a core composed by UGU. Thus, we performed motif clusterization with the program “Gimmemotifs”<sup>41</sup> (Supplementary Fig. 10a) and found two consensus motifs of 4 and 5-nt long, respectively (Fig. 6d). Each motif corresponds to many different words that belong to the motif with different probabilities. Hence, to set up a threshold on the number of words to look for, we calculated the occurrences of the two motifs in the sequences of the three Sfpq peak subsets previously selected. We used the tool FIMO<sup>61</sup> using different *p*-value cutoffs associated to the motif likelihood established by the program. For cutoffs higher than 0.01 for motif#1 and 0.005 for motif#2, we found many occurrences superimposed. Thus, to avoid superimposition, we established the aforementioned cutoffs, which correspond to two words for the motif#1, namely, CUGU and CUGA, and four words for motif#2, namely, CUGUA, AUGUA, UUGUA, and GUGUA.

**RNA-IP and RNA sequencing.** For RNA-IP, cells lysates were immunoprecipitated with Protein G-Dynabeads (Invitrogen)-coupled antibodies at 4 °C overnight<sup>2</sup>. When required, cell lysates were incubated at room temperature with RNase A (10 µg ml<sup>-1</sup>, Ambion) for 30 min for partial RNA digestion before the IP. When indicated, cell lysates were also incubated at room temperature with 100 nM recombinant human Sfpq wild-type or Sfpq-214–598 quadruple mutant (L535A, L539A, L546A, and M549A) for 30 min before RNA digestion and IP. Total RNA

was prepared using Trizol (Invitrogen) and RNA columns (Qiagen), and analyzed by either small RNA sequencing or quantitative RT-PCR.

For small RNA sequencing 45 ng of RNA from each IP were subjected to small RNA library preparation with the CleanTag™ Ligation Kit for Small RNA Library Prep, using a ¼ dilution of the 3' and 5' adapters and performing 15 cycles of PCR. PCR products were purified on Purelink PCR micro kit columns (Invitrogen). Final libraries were quantified with bioanalyzer (Agilent) and Qubit (Invitrogen), and subjected to sequencing by Solid sequencer (Life Technologies). Sequenced fragments were mapped with the LifeScope 2.5.1 pipeline (Life Technologies) and annotated according to the Ensembl non-coding RNA database.

**Bioinformatics analysis.** First, we checked the quality of the replicates using the PCA, then we used the webserver omiRas<sup>62</sup> to map the reads on the mouse genome, to normalize the data, and to compare each RNA-IP sample to the input. To avoid detecting differences at the level of transcriptional noise, we required a minimum of 20 reads for each small RNA identified. In order to specifically select small RNAs enriched in RNA-IP samples respect to the input, we employed Gaussian mixture modeling of the small RNA population distribution (Supplementary Fig. 5a) and calculated the LOD to filter out the background distribution (LOD > 0) from the input (Supplementary Data 4).

**RNA-sequencing experiments and analysis.** To check whether Sfpq may control miRNA targeting by alternative splicing or polyadenylation sites, we performed a RNA-sequencing analysis from siSfpq-transfected P19 cells and control using the Truseq Stranded mRNA kit (Illumina) from 1 µg total RNA and performing 15 PCR cycles. The final library size was centered on 290 bp. Paired-end sequencing was performed on a NextSeq 500 sequencer (Illumina) on a Mid 150 flowchip. The quality of replicates was assessed using the R package “htSeqTools”. Normalization of RNA-seq data was performed with the DESeq package available from Bioconductor (<http://www.bioconductor.org>). To identify differential splicing events between samples, we used the software SpliceTrap<sup>63</sup> (Supplementary Data 8). To assess whether Sfpq interferes with alternative polyadenylation sites and the length of the 3'UTRs we used the software DaPars<sup>64</sup>. Our analyses did not find any differential splicing or shortening-lengthening events regulated by Sfpq in those 3' UTRs on which Sfpq promotes miRNA targeting (Supplementary Fig. 12a).

**Additional bioinformatics and statistical analyses.** GO-term analysis was done with the webtool “GoTermFinder”<sup>50</sup> and the pictures were generated using the tool REVIGO<sup>65</sup> and custom R script. miRNA seed searching in peak sequences was performed with a custom python script that takes as input the peak sequences and the miRNA sequences, then calculates the reverse complement of the seed region using the library “Biopython”<sup>66</sup>. Venn diagrams were calculated with the webtool “Venny” (<http://bioinfogp.cnb.csic.es/tools/venny/index.html>). All statistical analyses were performed with the statistical software R, all other analyses were performed with custom python scripts where not specified otherwise.

**Microarray analysis.** For mRNA profiling analysis three replicates of P19 cells were transfected with siSfpq, mimic let-7a, siSfpq + mimic let-7a, or control (Supplementary Data 6). 48 h after transfection, total RNA was isolated by using RNeasy kit (Qiagen). RNA samples were labeled with Cy3 dye using the low RNA input QuickAmp kit (Agilent) as recommended by the supplier. 400 ng of labeled cRNA sample was hybridized on 8 × 60 K high-density SurePrint G3 gene mouse GE 8 × 60 K Agilent microarrays. Assessment of the quality of replicates was performed using the Bioconductor package “ArrayQualityMetrics”. Normalization of microarray data was performed with the Limma package available from Bioconductor (<http://www.bioconductor.org>) using the quantile methods.

**Cellular fractionation.** Cellular fractionation of cytoplasm, nucleoplasm, and chromatin was performed as previously described<sup>67</sup>, with a modification for the chromatin solubilization according to a previous protocol. Briefly, genomic DNA pellet was resuspended for treatment with Turbo DNase I in the DNase I digestion buffer (50 mM Tris pH 7.5, 0.5% Nonidet-P 40, 0.1% sodium lauroyl sarcosine, 1× Complete protease inhibitors) at 37 °C for 45 min with intermittent vortexing. The genomic DNA was further solubilized by adding 1% SDS, 0.3 M lithium chloride, 25 mM EDTA, and 25 mM EGTA and incubated at 37 °C for 15 min.

For IP experiments the salinity of 300 µg of protein lysate from cytoplasmic, nucleoplasmic, or genomic DNA solutions was adjusted to 150 mM of NaCl.

**Reagents and antibodies.** Glutathione-agarose, Actinomycin D, mouse IgG, rabbit IgG, anti-Flag, and anti-GST antibodies were purchased from Sigma. Knockdown experiments were performed with Smartpool siRNAs (Thermo Fisher Scientific), namely, mouse Sfpq (L-044760-01 and E-044760-00), human Sfpq (E-006455-00 and L-006455-00), mouse Pspc1 (L-49216-01), and mouse NonO (L-048587-00). Each catalog number of Smartpool siRNAs corresponds to a set of four different siRNAs. Two different Smartpool siRNA catalog numbers that target the same gene are composed by different siRNAs.

Protein-G Dynabeads were purchased from Thermo Fisher Scientific. PNK was purchased from New England Biolabs. Mouse monoclonal anti-Ago2 antibody (2D4) for HITS-CLIP, RNA-IP, and co-IP was purchased from Wako Chemicals. Rabbit anti-Sfpq (Ab38148), goat NonO (ab50411) antibodies were from Abcam.

Mouse monoclonal anti-c-Myc (9E10), mouse anti-BrdU, mouse anti-Tubulin, goat anti-GST, goat anti-actin (I-19), Rabbit anti-Pspc1 (sc-84577), and mouse anti-Pspc1 (sc-374387) antibodies were from Santa Cruz. Mouse anti-Sfpq antibody (B92) was purchased from Fischer Scientific. Mouse anti-Ago2 antibody for IP in human cells and immunoblotting was purchased from Sigma (WH0027161M1). Rabbit anti-Ago2 antibody for immunostaining and immunoblotting was purchased from Cell Signaling (C34C6). The immunofluorescence was revealed by using fluorescent secondary antibodies: Alexa Fluor 488-conjugated anti-mouse IgG (H + L) antibody (Cell Signaling), Alexa Fluor 594-conjugated anti-rabbit IgG (H + L) antibody (Molecular Probes). Slides were coverslipped in Vectashield Mounting Medium with DAPI (Vector Laboratories). Primary antibodies were diluted to  $1 \mu\text{g ml}^{-1}$  or  $0.1 \mu\text{g ml}^{-1}$  for immunoblotting or immunofluorescence, respectively. Secondary antibodies were diluted, according to the manufacturer's instructions.

**Cell transfection and immunoblotting.** RAW 264.7, P19, NTERA-2, HeLa, and HEK293T cells were purchased from ATCC. No mycoplasma contamination was assessed. Cells were transiently transfected for 48 h with Lipofectamine RNAiMax or 2000 (Invitrogen), according to the manufacturer's instructions. siRNAs and mimic miRNAs were transfected at the final concentration of 24 nM. Plasmids were transfected at the final concentration recommended by the manufacturer. When required, cell lysates were incubated at room temperature with RNase A ( $10 \text{ mg ml}^{-1}$ , Ambion) for 30 min for full RNA digestion. Three-hundred micrograms of proteins were immunoprecipitated with Protein G-Dynabeads (Invitrogen)-coupled antibodies for 16 h at  $4^\circ\text{C}$  with rotation. Immunoprecipitates were washed four times with the lysis buffer and resuspended in SDS protein loading buffer. Proteins were subjected to SDS-PAGE, electroblotted onto PVDF membranes, and probed with the indicated antibodies. Western blot analyses were performed with  $30 \mu\text{g}$  of protein lysate where not specified otherwise.

**Immunofluorescence.** For immunofluorescence, cells were fixed in 10% neutral formalin and incubated overnight with the primary antibodies<sup>68</sup>: anti-Sfpq, anti-Pspc1, or anti-Ago2 antibodies. Slides were coverslipped in Vectashield Mounting Medium with Dapi (Invitrogen). As control for immunofluorescence, slides were incubated with only the Alexa Fluor 488-conjugated anti-mouse or Alexa Fluor 594-conjugated anti-rabbit antibodies but not with primary antibodies. No signal was detected in these conditions. The results were analyzed on a Leica DM5500B microscope with a HAMAMATSU camera ORGA-ER.

**Luciferase reporter assay.** HEK293T cells (80% confluence in 96-well plates) were transfected with Lipofectamine 2000<sup>®</sup>, according to the manufacturer's instructions. Luciferase reporter assay was performed according to the manufacturer's instructions.

**RNA EMSA.** <sup>32</sup>P-labeled RNAs were synthesized and used as substrates for EMSA<sup>69</sup>. Briefly, <sup>32</sup>P-labeled RNAs were incubated with recombinant protein in the binding buffer (75 mM Hepes 7.9, 150 mM NaCl, 75 mM KCl, 0.1 mM EDTA, 5 mM DTT, 5 mM PMSF, and 15% glycerol) for 15 min at room temperature. The <sup>32</sup>P-labeled RNA-recombinant protein complex was resolved in a 10% acrylamide native gel. Finally, gel was dried and exposed at different time onto a film.

**Northern blot.** Total RNA was isolated from cells using Trizol (Invitrogen), resolved on 10% polyacrylamide-urea gels, and electroblotted onto HyBond N+ membranes. Membranes were hybridized overnight with radiolabeled DNA antisense oligonucleotides to miRNAs in ExpressHyb solution (Clontech). After hybridization, membranes were washed three times with  $2\times$  SSC and 0.05% SDS, twice with  $0.1\times$  SSC and 0.1% SDS, exposed overnight onto a film. The same blot was hybridized (upon stripping in boiling 0.1% SDS) with up to three distinct oligonucleotide probes (Supplementary Table 1).

**mRNA and miRNA quantitative expression analysis.** RNA expression by quantitative RT-PCR was performed by using standard procedures. Briefly, total RNA was isolated from cells using miRNeasy kit (Qiagen) and cDNA was synthesized with a random hexanucleotides using Superscript III or n-code kit for miRNAs (Life Technologies). Quantitative RT-PCRs using Sybr Green were performed on a StepONE system (Applied Biosystem). Expression was considered undetectable with Ct value  $\geq 40$ . The target expression value was normalized with a couple of reference genes: either U6-U2 snRNAs or actin- $\beta$ -microglobulin. The relative expression level was further normalized by the  $2^{-\Delta\Delta\text{Ct}}$  method when indicated. The Student's *t*-test or the one-way ANOVA followed by Tukey's post hoc test was performed to assess statistical significance. The primer sequences are detailed in Supplementary Table 1.

**Nuclear run-on.** We performed NRO followed by RT-qPCR<sup>70</sup>. Briefly, nuclei were prepared as described in "Cellular fractionation" method. To produce nascent RNA, nuclei were incubated with the run-on buffer ( $5\times$  buffer: 0.5 M Tris pH 8, 1 M  $\text{Mg}_2\text{Cl}_2$ , 1.5 M KCl, 200 mM DTT, 100 mM rNTPs, and 2 mM BrdUTP) for 30 min at  $30^\circ\text{C}$ . Trizol-extracted RNA was immunoprecipitated with anti-BrdU

antibody overnight and eluted. RT-qPCR was performed from input and immunoprecipitated samples. Five  $\mu\text{l}$  of  $1.5 \text{ pg } \mu\text{l}^{-1}$  in vitro transcribed 300 nt spike-in BrU-GFP (positive control) and fire-fly luciferase (FF—negative control) were added to  $100 \mu\text{l}$  of extracted NRO-RNAs (final concentration  $0.15 \text{ pg } \mu\text{l}^{-1}$ ). Nascent UTP-actin or BrU-actin B was used as internal control for NRO experiments (Supplementary Fig. 9c, right panel).

**Plasmids and recombinant proteins.** pMIR-REPORT plasmid containing the wild-type sequence of mouse Lin28a 3'UTR downstream the fire-fly luciferase CDS for gene reporter assay was kindly gifted by Dr. Baltimore. pet15b plasmid containing  $6\times$ His-tagged full-length human Sfpq or the Sfpq-214–598 quadruple mutant (L535A, L539A, L546A, and M549A) were kindly provided by Drs. Lynch and Bond, respectively. Sequence encoding only for the two RRM domains of human Sfpq was cloned in-house into pGEX-4T plasmid (GE Healthcare Life Sciences). Production of recombinant proteins were performed by the Protein Expression Facility at the University of Aix-Marseille. The purity of the recombinants was tested with protein gel stained by Imperial or Coomassie blue dye or by Sfpq immunoblotting. Recombinant Ago2 was purchased from Sino Biological Inc. Luciferase reporter gene plasmid containing 6xlet-7a-binding sites was kindly gifted by Dr. J.G. Belasco. HA-Sfpq was generously provided by Dr. S. Sacconi. MYC-Ago2, Flag-Ago2, and Flag-Ago1 were provided by Addgene. Oligonucleotides for cloning are listed in Supplementary Table 1.

**Atomic force microscope.** The in vitro transcribed Lin28A 3'UTR was heated at  $95^\circ\text{C}$  for 1 min and immediately cooled in ice. We then added the binding buffer (120 mM NaCl, 24 mM Tris-HCl pH 7.5, 4 mM  $\text{MgCl}_2$ ) containing the recombinant Sfpq proteins (wild-type or Sfpq-214–598 quadruple mutant (L535A, L539A, L546A, and M549A)) or BSA to a final protein-RNA concentration ratio corresponding to 121 nM/13 nM, and incubated at  $4^\circ\text{C}$  for 30 min. Ten  $\mu\text{l}$  of this solution was applied to the untreated muscovite mica surface and incubated for 10 min. After the incubation, we washed the surface with 1 ml of distilled water to wash away the unbound RNA and dried under a gentle flux of nitrogen gas. In our experimental condition, the mica surface is not enough positively charged to allow the absorption of unbound RNA, which is removed from the surface in the washing step. Thus, RNA is only visible on the mica surface upon interaction with protein.

Imaging was performed using the Multimode8 AFM connected to the Nanoscope V controller (Bruker Nano Surfaces, Santa Barbara, CA, USA) equipped with E-scanner, operating in PeakForce mode in air at room temperature using ScanAsyst-air probes (Bruker Nano Surfaces, Santa Barbara, CA, USA). Images of  $1 \mu\text{m}^2$  corresponding to  $512 \times 512$  pixels were captured with a scanning rate of 1 Hz using a PeakForce frequency of 2 kHz, a PeakForce setpoint of 0.03 V, and a Z limit of 700 nm. Raw AFM images were processed only for background slope removal (flattening).

**Data availability.** All sequencing and microarray for mRNA profiling were submitted to the GEO database under accession number series GSE89033.

Received: 22 December 2016 Accepted: 18 August 2017

Published online: 30 October 2017

## References

- Pasquinelli, A. E. MicroRNAs and their targets: recognition, regulation and an emerging reciprocal relationship. *Nat. Rev. Genet.* **13**, 271–282 (2012).
- Repetto, E. et al. RNY-derived small RNAs as a signature of coronary artery disease. *BMC Med.* **13**, 259 (2015).
- Ha, M. & Kim, V. N. Regulation of microRNA biogenesis. *Nat. Rev. Mol. Cell Biol.* **15**, 509–524 (2014).
- Ameyar-Zazoua, M. et al. Argonaute proteins couple chromatin silencing to alternative splicing. *Nat. Struct. Mol. Biol.* **19**, 998–1004 (2012).
- Benhamed, M., Herbig, U., Ye, T., Dejean, A. & Bischof, O. Senescence is an endogenous trigger for microRNA-directed transcriptional gene silencing in human cells. *Nat. Cell Biol.* **14**, 266–275 (2012).
- Gagnon, K. T., Li, L., Chu, Y., Janowski, B. A. & Corey, D. R. RNAi factors are present and active in human cell nuclei. *Cell Rep.* **6**, 211–221 (2014).
- Liu, C. et al. CLIP-based prediction of mammalian microRNA binding sites. *Nucleic Acids Res.* **41**, e138 (2013).
- Jens, M. & Rajewsky, N. Competition between target sites of regulators shapes post-transcriptional gene regulation. *Nat. Rev. Genet.* **16**, 113–126 (2015).
- Srikantan, S., Tominaga, K. & Gorospe, M. Functional interplay between RNA-binding protein HuR and microRNAs. *Curr. Protein Pept. Sci.* **13**, 372–379 (2012).
- Kedde, M. et al. RNA-binding protein Dnd1 inhibits microRNA access to target mRNA. *Cell* **131**, 1273–1286 (2007).
- Asara, J. M., Christofk, H. R., Freemark, L. M. & Cantley, L. C. A label-free quantification method by MS/MS TIC compared to SILAC and spectral counting in a proteomics screen. *Proteomics* **8**, 994–999 (2008).



12. Mathys, H. et al. Structural and biochemical insights to the role of the CCR4-NOT complex and DDX6 ATPase in microRNA repression. *Mol. Cell* **54**, 751–765 (2014).
13. Hock, J. et al. Proteomic and functional analysis of argonaute-containing mRNA-protein complexes in human cells. *EMBO Rep.* **8**, 1052–1060 (2007).
14. Bhattacharyya, S. N., Habermacher, R., Martine, U., Closs, E. I. & Filipowicz, W. Relief of microRNA-mediated translational repression in human cells subjected to stress. *Cell* **125**, 1111–1124 (2006).
15. Frohn, A. et al. Dicer-dependent and -independent Argonaute2 protein interaction networks in mammalian cells. *Mol. Cell Proteomics* **11**, 1442–1456 (2012).
16. Bond, C. S. & Fox, A. H. Paraspeckles: nuclear bodies built on long noncoding RNA. *J. Cell Biol.* **186**, 637–644 (2009).
17. Yarosh, C. A. et al. TRAP150 interacts with the RNA-binding domain of PSF and antagonizes splicing of numerous PSF-target genes in T cells. *Nucleic Acids Res.* **43**, 9006–9016 (2015).
18. Lee, M. et al. The structure of human SFPQ reveals a coiled-coil mediated polymer essential for functional aggregation in gene regulation. *Nucleic Acids Res.* **43**, 3826–3840 (2015).
19. Fox, A. H. & Lamond, A. I. Paraspeckles. *Cold Spring Harb. Perspect. Biol.* **2**, a000687 (2010).
20. Hirose, T. et al. NEAT1 long noncoding RNA regulates transcription via protein sequestration within subnuclear bodies. *Mol. Biol. Cell* **25**, 169–183 (2013).
21. Chen, L. L. & Carmichael, G. G. 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 (2009).
22. Imamura, K. et al. Long noncoding RNA NEAT1-dependent SFPQ relocation from promoter region to paraspeckle mediates IL8 expression upon immune stimuli. *Mol. Cell* **53**, 393–406 (2014).
23. Adriaens, C. et al. p53 induces formation of NEAT1 lncRNA-containing paraspeckles that modulate replication stress response and chemosensitivity. *Nat. Med.* **22**, 861–868 (2016).
24. Zhang, Q., Chen, C. Y., Yedavalli, V. S. & Jeang, K. T. NEAT1 long noncoding RNA and paraspeckle bodies modulate HIV-1 posttranscriptional expression. *MBio.* **4**, e00596–e00512 (2013).
25. Zhang, H. et al. PABPC1 interacts with AGO2 and is responsible for the microRNA mediated gene silencing in high grade hepatocellular carcinoma. *Cancer Lett.* **367**, 49–57 (2015).
26. Johnson, C. D. et al. The let-7 microRNA represses cell proliferation pathways in human cells. *Cancer Res.* **67**, 7713–7722 (2007).
27. Lowery, L. A., Rubin, J. & Sive, H. Whitesnake/sfpq is required for cell survival and neuronal development in the zebrafish. *Dev. Dyn.* **236**, 1347–1357 (2007).
28. Wulczyn, F. G. et al. Post-transcriptional regulation of the let-7 microRNA during neural cell specification. *FASEB J.* **21**, 415–426 (2007).
29. Roush, S. & Slack, F. J. The let-7 family of microRNAs. *Trends Cell Biol.* **18**, 505–516 (2008).
30. Parsons, X. H., Parsons, J. F. & Moore, D. A. Genome-scale mapping of microRNA signatures in human embryonic stem cell neurogenesis. *Mol. Med. Ther.* **1**, 105 (2012).
31. Wu, L. & Belasco, J. G. Micro-RNA regulation of the mammalian lin-28 gene during neuronal differentiation of embryonal carcinoma cells. *Mol. Cell Biol.* **25**, 9198–9208 (2005).
32. La Torre, A., Georgi, S. & Reh, T. A. Conserved microRNA pathway regulates developmental timing of retinal neurogenesis. *Proc. Natl. Acad. Sci. USA* **110**, E2362–E2370 (2013).
33. Gurtan, A. M. et al. Let-7 represses Nr6a1 and a mid-gestation developmental program in adult fibroblasts. *Genes Dev.* **27**, 941–954 (2013).
34. Bottini, S. et al. From benchmarking HITS-CLIP peak detection programs to a new method for identification of miRNA-binding sites from Ago2-CLIP data. *Nucleic Acids Res.* **45**, e71 (2017).
35. Greve, T. S., Judson, R. L. & Belloch, R. MicroRNA control of mouse and human pluripotent stem cell behavior. *Annu. Rev. Cell Dev. Biol.* **29**, 213–239 (2013).
36. Khalili, M. et al. Down-regulation of miR-302b, an ESC-specific microRNA, in Gastric Adenocarcinoma. *Cell J.* **13**, 251–258 (2012).
37. Agarwal, V., Bell, G. W., Nam, J. W. & Bartel, D. P. Predicting effective microRNA target sites in mammalian mRNAs. *Elife* **4**, e05005 (2015).
38. Zhou, J., Ng, S. B. & Chng, W. J. LIN28/LIN28B: an emerging oncogenic driver in cancer stem cells. *Int. J. Biochem. Cell Biol.* **45**, 973–978 (2013).
39. Chaudhuri, A. A. et al. Oncomir miR-125b regulates hematopoiesis by targeting the gene Lin28A. *Proc. Natl. Acad. Sci. USA* **109**, 4233–4238 (2012).
40. Mitchell, S. F. & Parker, R. Principles and properties of eukaryotic mRNPs. *Mol. Cell* **54**, 547–558 (2014).
41. van Heeringen, S. J. & Veenstra, G. J. GimmeMotifs: a de novo motif prediction pipeline for ChIP-sequencing experiments. *Bioinformatics* **27**, 270–271 (2011).
42. Ciafre, S. A. & Galardi, S. MicroRNAs and RNA-binding proteins: a complex network of interactions and reciprocal regulations in cancer. *RNA Biol.* **10**, 935–942 (2013).
43. Ohrt, T. et al. Fluorescence correlation spectroscopy and fluorescence cross-correlation spectroscopy reveal the cytoplasmic origination of loaded nuclear RISC in vivo in human cells. *Nucleic Acids Res.* **36**, 6439–6449 (2008).
44. Weinmann, L. et al. Importin 8 is a gene silencing factor that targets argonaute proteins to distinct mRNAs. *Cell* **136**, 496–507 (2009).
45. Kataoka, N. et al. Pre-mRNA splicing imprints mRNA in the nucleus with a novel RNA-binding protein that persists in the cytoplasm. *Mol. Cell* **6**, 673–682 (2000).
46. Leung, A. K. et al. Genome-wide identification of Ago2 binding sites from mouse embryonic stem cells with and without mature microRNAs. *Nat. Struct. Mol. Biol.* **18**, 237–244 (2011).
47. Riemondy, K., Wang, X. J., Torchia, E. C., Roop, D. R. & Yi, R. MicroRNA-203 represses selection and expansion of oncogenic Hras transformed tumor initiating cells. *Elife* **4**, e07004 (2015).
48. Klass, D. M. et al. Quantitative proteomic analysis reveals concurrent RNA-protein interactions and identifies new RNA-binding proteins in *Saccharomyces cerevisiae*. *Genome Res.* **23**, 1028–1038 (2013).
49. Shevchenko, A., Wilm, M., Vorm, O. & Mann, M. Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal. Chem.* **68**, 850–858 (1996).
50. Boyle, E. I. et al. GO::TermFinder—open source software for accessing gene ontology information and finding significantly enriched gene ontology terms associated with a list of genes. *Bioinformatics* **20**, 3710–3715 (2004).
51. Letunic, I., Doerks, T. & Bork, P. SMART: recent updates, new developments and status in 2015. *Nucleic Acids Res.* **43**, D257–D260 (2014).
52. Moore, M. J. et al. Mapping Argonaute and conventional RNA-binding protein interactions with RNA at single-nucleotide resolution using HITS-CLIP and CIMS analysis. *Nat. Protoc.* **9**, 263–293 (2014).
53. Wang, Z., Tollervey, J., Briese, M., Turner, D. & Ule, J. CLIP: construction of cDNA libraries for high-throughput sequencing from RNAs cross-linked to proteins in vivo. *Methods* **48**, 287–293 (2009).
54. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2012).
55. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
56. Wang, T., Xie, Y. & Xiao, G. dCLIP: a computational approach for comparative CLIP-seq analyses. *Genome Biol.* **15**, R11 (2014).
57. Althammer, S., Gonzalez-Vallinas, J., Ballare, C., Beato, M. & Eyras, E. Pyicos: a versatile toolkit for the analysis of high-throughput sequencing data. *Bioinformatics* **27**, 3333–3340 (2011).
58. Favorov, A. et al. Exploring massive, genome scale datasets with the GenometriCorr package. *PLoS Comput. Biol.* **8**, e1002529 (2012).
59. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
60. Bailey, T. L. DREME: motif discovery in transcription factor ChIP-seq data. *Bioinformatics* **27**, 1653–1659 (2011).
61. Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017–1018 (2011).
62. Muller, S. et al. omiRas: a web server for differential expression analysis of miRNAs derived from small RNA-Seq data. *Bioinformatics* **29**, 2651–2652 (2013).
63. Wu, J. et al. SpliceTrap: a method to quantify alternative splicing under single cellular conditions. *Bioinformatics* **27**, 3010–3016 (2011).
64. Xia, Z. et al. Dynamic analyses of alternative polyadenylation from RNA-seq reveal a 3'-UTR landscape across seven tumour types. *Nat. Commun.* **5**, 5274 (2014).
65. Supek, F., Bosnjak, M., Skunca, N. & Smuc, T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS ONE* **6**, e21800 (2011).
66. Cock, P. J. et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**, 1422–1423 (2009).
67. Gagnon, K. T., Li, L., Janowski, B. A. & Corey, D. R. Analysis of nuclear RNA interference in human cells by subcellular fractionation and Argonaute loading. *Nat. Protoc.* **9**, 2045–2060 (2014).
68. Repetto, E. et al. Let-7b/c enhance the stability of a tissue-specific mRNA during mammalian organogenesis as part of a feedback loop involving KSRP. *PLoS Genet.* **8**, e1002823 (2012).
69. Trabucchi, M. et al. The RNA-binding protein KSRP promotes the biogenesis of a subset of microRNAs. *Nature* **459**, 1010–1014 (2009).
70. Roberts, T. C. et al. Quantification of nascent transcription by bromouridine immunocapture nuclear run-on RT-qPCR. *Nat. Protoc.* **10**, 1198–1211 (2015).

## Acknowledgements

We are indebted to Drs D. Baltimore, J.G. Belasco, K.W. Lynch, C.S. Bond, and S. Sacconi for reagents and to Mr. G. Bottini for the user-friendly interface website. We thank Drs D. Alcor for imaging facility and D. Byrne for protein expression facility. This work has been supported by the AVENIR program of INSERM and Marie Curie CIG to M.T.,



ANR through the “Investments for the Future” #ANR-11-LABX-0028-01 (LABEX SIGNALIFE) to M.T. and P.B., ITMO cancer of Avesian within the framework Plan Cancer 2009–2013 (ProstaMir P029393) to M.T., P.B., and M.B. N.T. was supported by INSERM-PACA region fellowship, M.T. by FRM (grant #DEQ20140329551), P.B. by FRM (grant #DEQ20130326464), and E.R. by FRM (ING20140129224). Marseille Proteomics (IBiSA) is supported by Institut Paoli-Calmettes and Canceropôle PACA. PICMI, the IRCAN’s Imaging facility is supported by ARC, IBiSA and Conseil General 06 de la Région PACA. Sequencing was performed by the UCA GenomiX platform, supported by ANR (ANR-11-LABX-0028-01, ANR-10-INBS-09-03, and ANR-10-INBS-09-02) and Canceropole PACA.

### Author contributions

M.T. conceived and coordinated the study, and designed the experiments. N.H.-T. performed the majority of the experiments. R.M., E.R. and M.T. performed the experiments. L.-E.Z. and P.B. performed high-throughput sequencing. S.A. performed the mass-spectrometry experiments. S.P. performed the AFM experiments. S.B., E.R. and M. T. analyzed the data. V.G., M.B. and C.M. participated in data analysis. S.B. performed the bioinformatics analyses. M.T. wrote the manuscript. All authors read and approved the final manuscript.

### Additional information

**Supplementary Information** accompanies this paper at doi:10.1038/s41467-017-01126-x.

**Competing interests:** The authors declare no competing financial interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**Publisher’s note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017

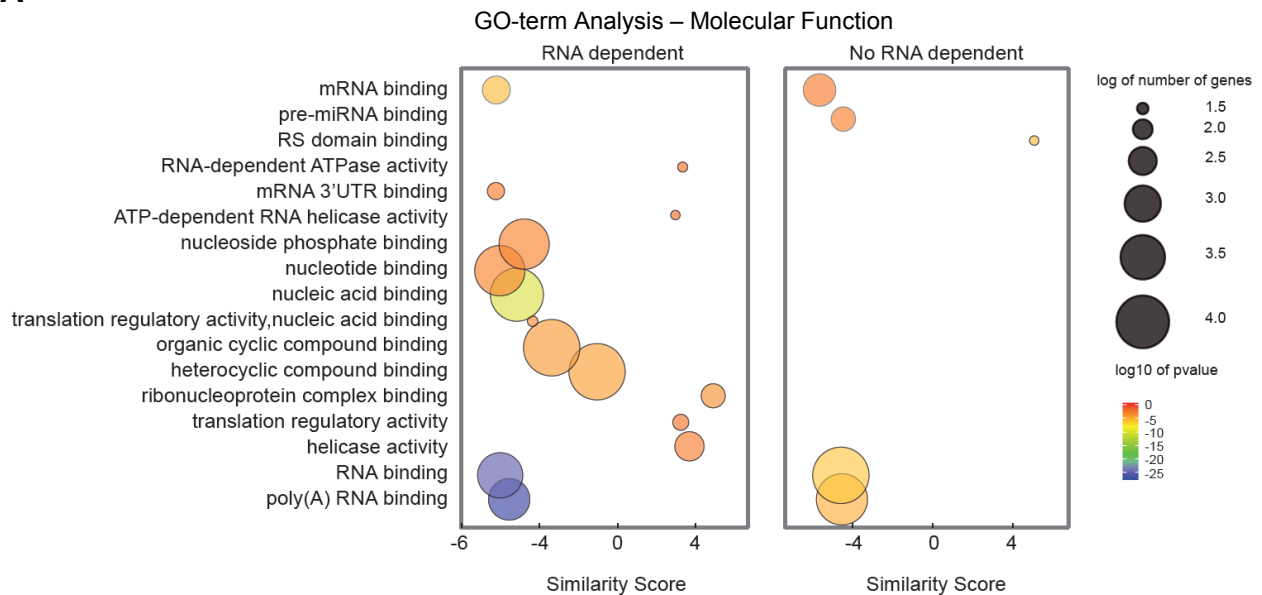
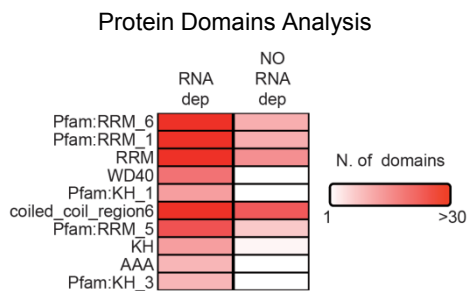
## Supplementary Information

### Supplementary Table 1

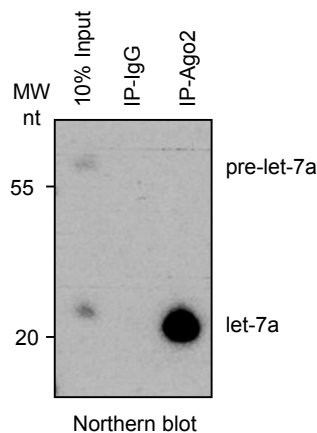
Oligonucleotides used for quantitative RT-PCR (RT-qPCR), cloning, EMSA, and Northern blot analyses.

	Forward primer	Reverse primer
let-7a (qPCR)	5'-TGAGGTAGTAGGTTGTATAGTT-3'	
miR-24 (qPCR)	5'-TGGCTCAGTTCAGCAGGAACAG -3'	
miR-125b (qPCR)	5'-TCCCTGAGACCCUAACTTGTGA -3'	
universal primer (qPCR)		Universal primer from N-code kit (Life-Technology)
Mouse U2 snRNA(qPCR)	5'-GAAGTAGGAGTTGGAATAGGA -3'	5'-ACCGTTCCTGGAGGACTG -3'
Human/Mouse U6 snRNA (qPCR)	5'- CGCTTCGGCAGCACATATAC -3'	5'- AAATATGGAACGCTTCACGA -3'
Human U2 snRNA(qPCR)	5'- GAGCAGGGAGATGGAATAGGA -3'	5'-ACCGTTCCTGGAGGACTG -3'
Human Phc3 - miR-302b binding site (qPCR)	5'- TGTTGTCCCCACTCAAACA-3'	5'-TGTGATAGCAAACAAGCCATTT-3'
mouse Phc3 - miR-302b binding site (qPCR)	5'- CCTCCAGCAATAATGCTTCC -3'	5'- CACGATGGGCTGGTTTTAGT -3'
Human Phlpp2 - miR-302b binding site (qPCR)	5'- TGAGAGCACTTTTCATTGACTG -3'	5'- GATTGGCTCGAGTGCAGATT -3'
mouse Phlpp2 - miR-302b binding site (qPCR)	5'- CAGAACCAATCAAGCAGGTG -3'	5'- TCCGAGTGCAGATTCAAAAA -3'
Human Crtc2 - miR-302b binding site (qPCR)	5'- GGAGGGCCTAAAGCACTTGT -3'	5'- TCCAGACAGACGGTTCAAAG -3'
mouse Crtc2 - miR-302b binding site (qPCR)	5'- GGAGGGCCTAAAAGCACTTG -3'	5'- AGGCTCTGACCTCCAGACAG -3'
mouse Hic2 (qPCR)	5'- GCAAGGCCTGTAAAGACTGG -3'	5'- GGAAGACCTTGGTCCCTTTC -3'
mouse Mier2 (qPCR)	5'- TCCATGCCAGTACCACATA -3'	5'- CTCCACCTCCAAGTGCTTA -3'
mouse Lin28a – let-7a binding site (qPCR)	5'- TGAGTATGGGCCTCAGTGG -3'	5'- ACACACCACCCAATGTGTTC-3'
mouse Igf2bp1 – let-7a binding site (qPCR)	5'- GGATGAACTACCTCAGTCCTC -3'	5'- AGGCATGAGTCCCTGCCCTC -3'
mouse Dusp1 – let-7a binding site (qPCR)	5'- AAGAACCAAATACCTCAATT -3'	5'- GGGTCTATTTACAAGATAGT -3'
mouse Atf6b – let-7a binding site (qPCR)	5'- CCCGAGGGTACCTCGTGAGG -3'	5'- GATAAAGGAGAGAATGAAGA -3'

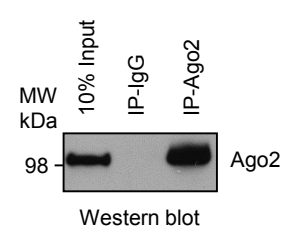
human Lin28a – let-7a binding site (qPCR)	5'- GCACAGCCTATTGAACTACCTC -3'	5'- AGAAAAGCCAGCTCTTATTGG -3'
mouse beta-2 microglobulin (qPCR)	5'- CTGCTACGTAACACAGTTCCACCC -3'	5'-CATGATGCTTGATCACATG-3'
mouse actin b (qPCR)	5'- CGGTTCCGATGCCCTGAGGCTCTT -3'	5'- CGTCACACTTCATGATGGAATTGA -3'
GFP (qPCR)	5'- AGGTGAAGTTCGAGGGCGAC -3'	5'- TTGTACTCCAGTTGTGCCC -3'
firefly (qPCR)	5'- GGAACAACCTTTACCGACCGC -3'	5'- TTCCATCTTCCAGGGATACG -3'
Mouse Lin28 mRNA (RT-qPCR and NRO)	5'- GTTCGGCTTCCTGTCTATGA -3'	5'- GTTGTAGCACCTGTCTCCTT -3'
Sfpq binding sites Motif#1 and #2 n1 (EMSA)	5'- CUGUACUGUA -3'	
Sfpq binding sites Motif#1 n2 (EMSA)	5'- CUGACUGA -3'	
Sfpq binding sites Motif#2 n1 (EMSA)	5'- UUGUAUUGUA -3'	
Sfpq binding sites Motif#2 n2 (EMSA)	5'- AUGUAAUGUA -3'	
Sfpq binding sites Motif#2 n3 (EMSA)	5'- GUGUAGUGUA -3'	
Sfpq negative binding site RNA oligo (EMSA)	5'- CCCC GCCCCG -3'	
Human Sfpq- RRM1 and 2 domains (cloning)	5'- GGAGAGAAA ACTTACACAC -3'	5'- CTAATGATGATGATGATGATG-3'
let-7a (Northern blot)		5'-AACTATAACAACCTACTACCTCA -3'
miR-23b (Northern blot)		5'- GGTAATCCCTGGCAATGTGAT-3'

**A****B****C**

RNA-IP from RAW 264.7 cells

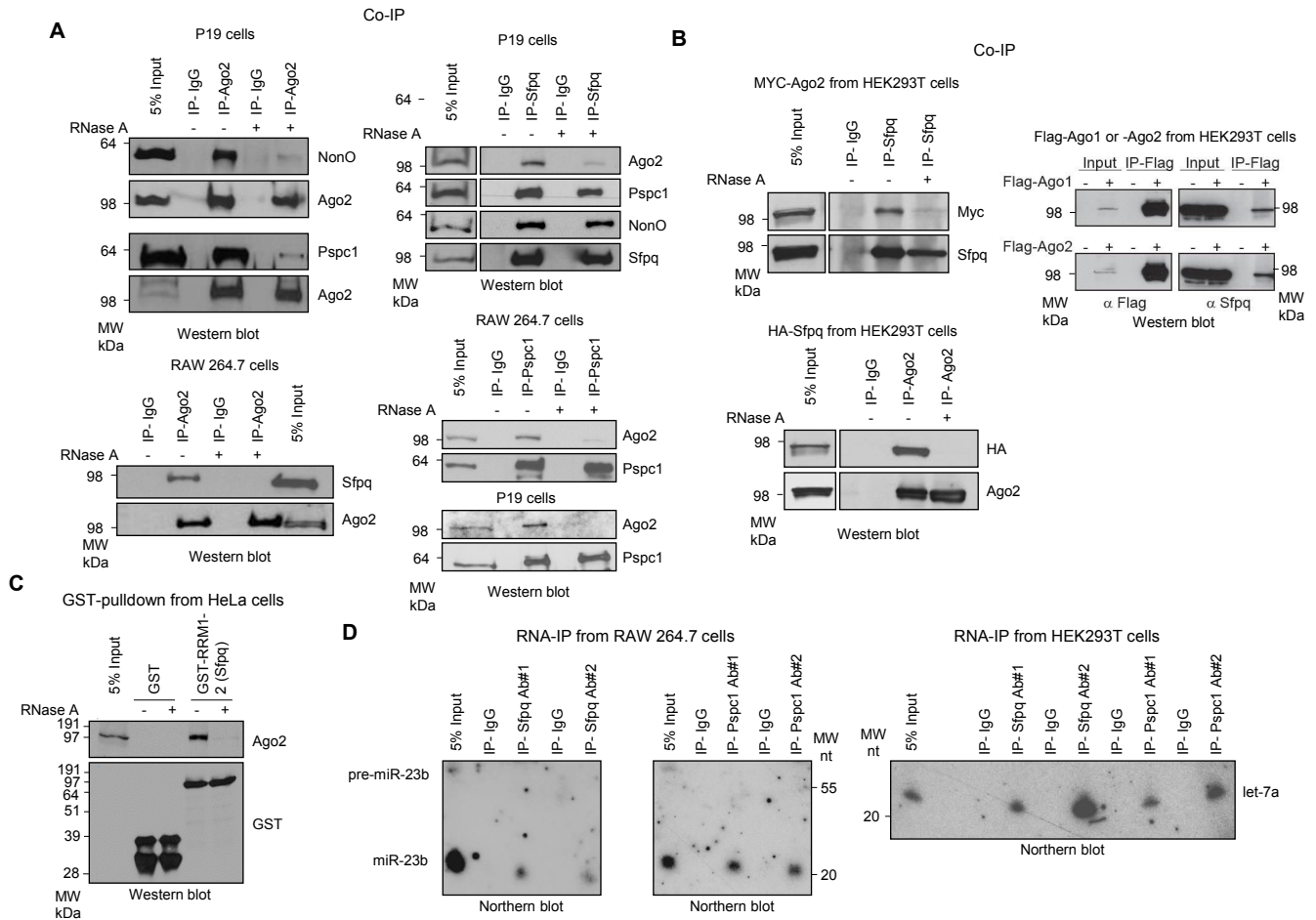
**D**

Co-IP from RAW 264.7 cells

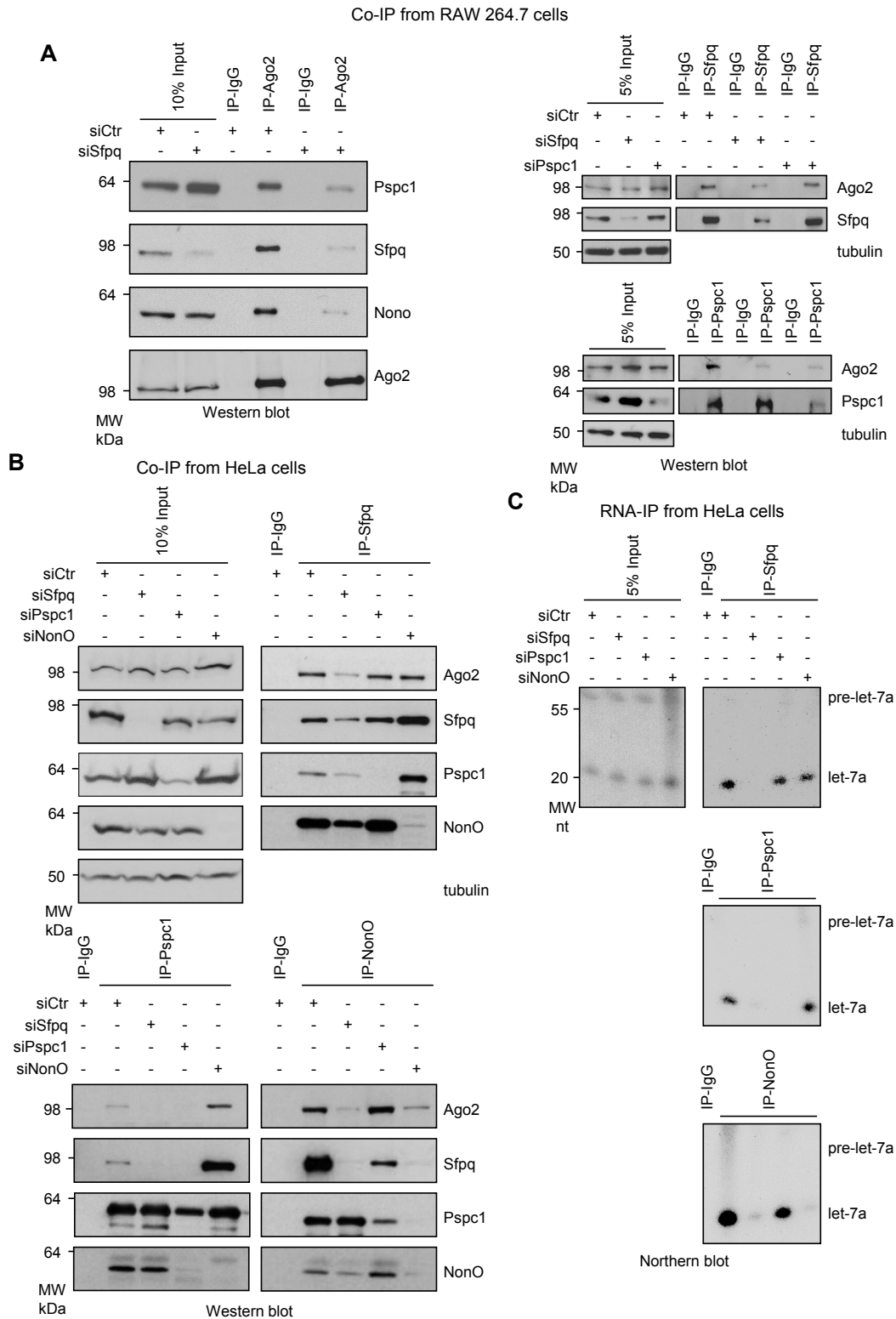


### Supplementary Figure 1. Proteomic analysis to identify RNA-dependent proteins

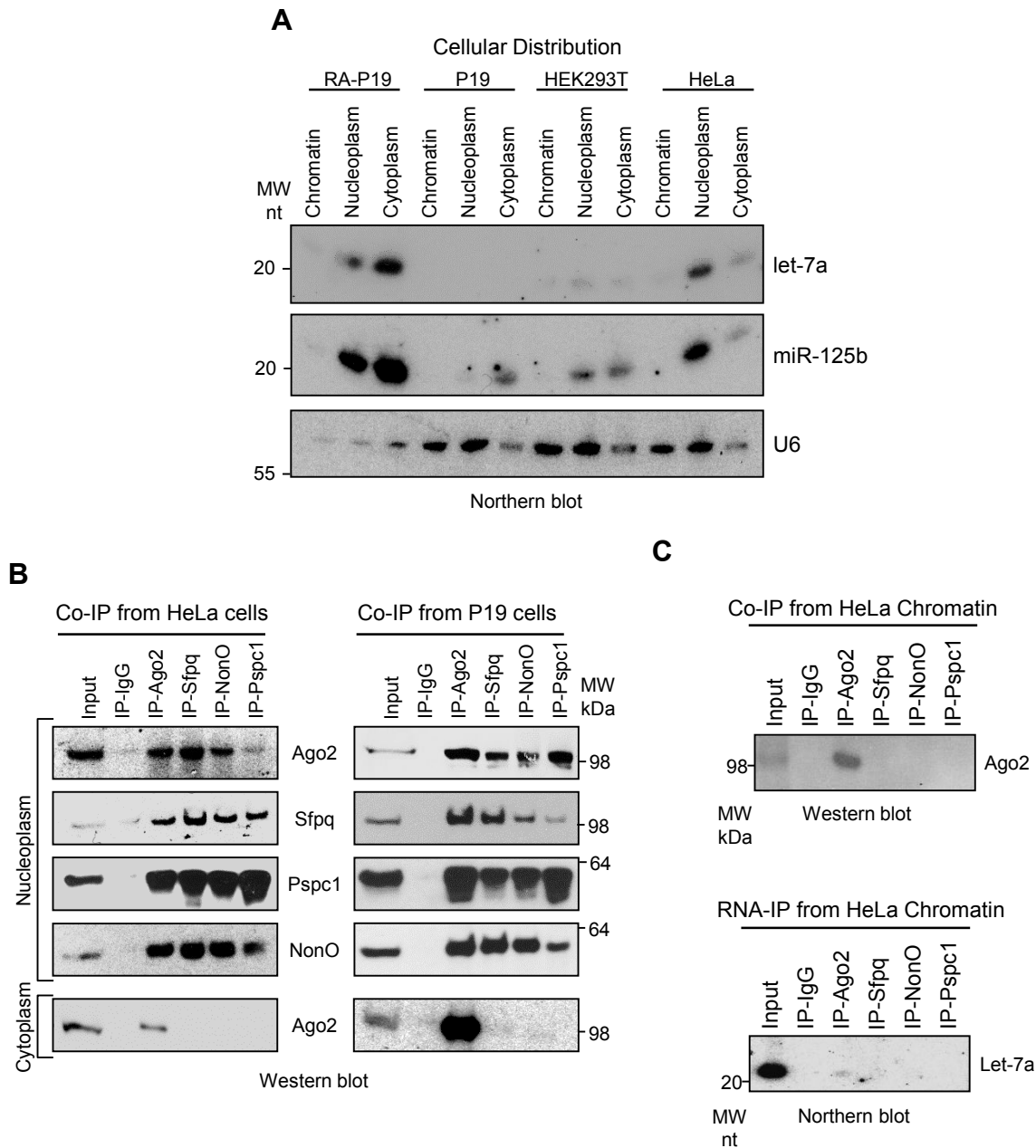
**interacting to Ago2.** (A) GO Term analysis of RNA–dependent and –independent interactors of Ago2 found by MS analysis. Scores among the enriched GO TERM were calculated by REVIGO<sup>1</sup>. Color code and different radius of circle meaning are indicated. (B) Protein domain enrichment on RNA–dependent and –independent interactors of Ago2 found by MS analysis. Color code is indicated. (C) RNA-IP of let-7a with Ago2. RAW 264.7 cell extracts were immunoprecipitated with anti-Ago2 antibody, RNA was purified, and analyzed by Northern blotting. (D) Immunoprecipitation of endogenous Ago2 in RAW 264.7 cells and immunoblotted with anti-Ago2 antibody.



**Supplementary Figure 2. Sfpq, Pspc1, and NonO interact with the miRISC.** (A) Co-IP of the indicated endogenous proteins in the indicated cell lines. (B) Co-IP of the indicated proteins from HEK293T cells. Cells were transfected with the indicated tagged-proteins. (C) GST-pulldown experiment from 500  $\mu$ g proteins from HeLa cell extract. Pulldown was performed with either the two RRM domains from human Sfpq fused with GST or GST alone as control. (D) RNA-IP of the indicated miRNAs with the endogenous Sfpq and Pspc1 from the indicated cell lines. Cell extracts were immunoprecipitated with two different antibodies and RNA was purified and analyzed by Northern blotting. When indicated, cell lysates were incubated at room temperature with RNase A ( $10 \text{ mg ml}^{-1}$ ) for 30min.

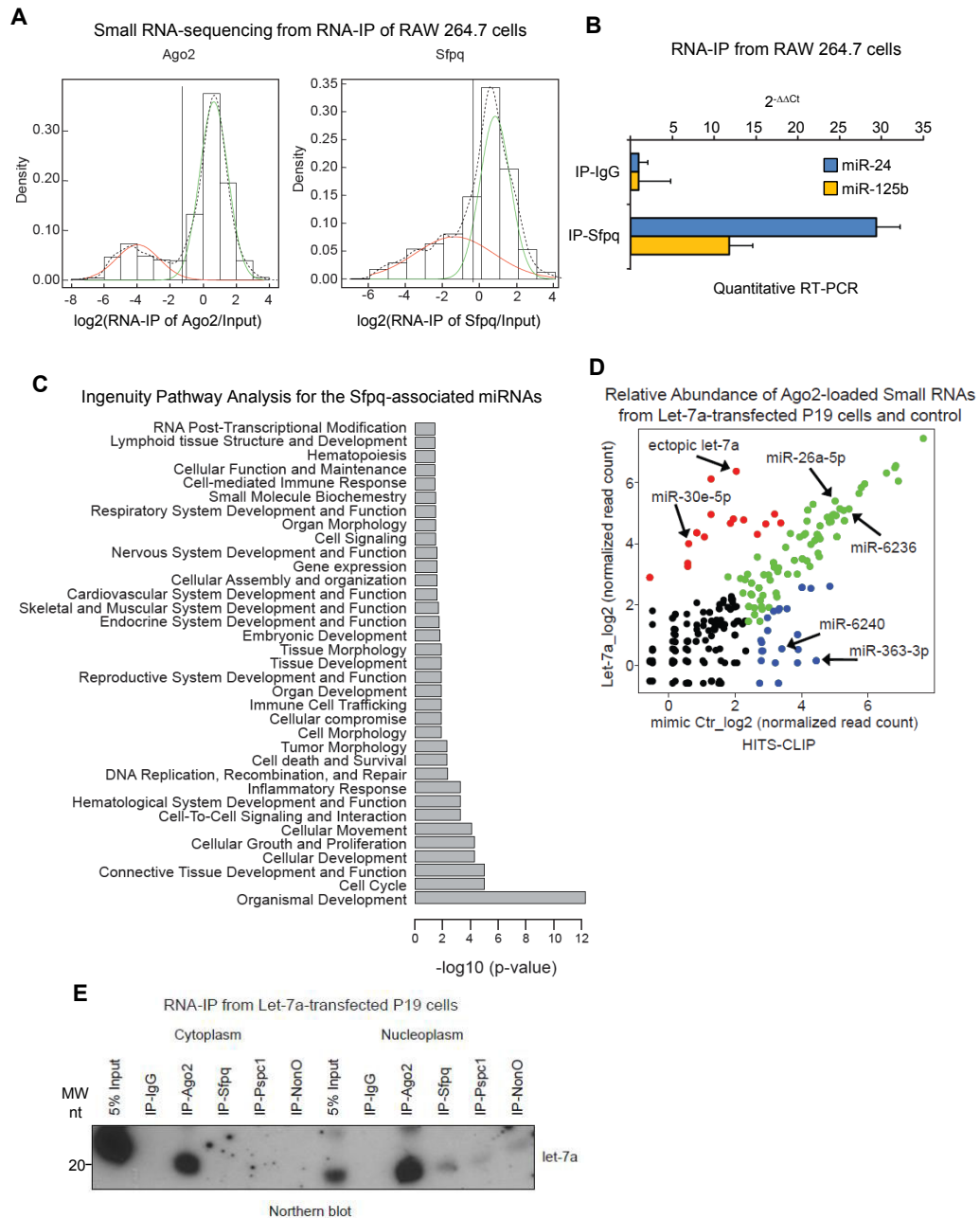


**Supplementary Figure 3. Sfpq mediates the interaction between the miRISC and Pspc1 or NonO.** Co-IP of the indicated endogenous proteins from (A) RAW 264.7 or (B) HeLa cells. Cells were transfected with the indicated siRNAs. Tubulin served as control for the knockdown. (C) RNA-IP of let-7a with the endogenous Sfpq, Pspc1, or NonO. HeLa cells were transfected with the indicated siRNAs. Cell extracts were immunoprecipitated with the indicated antibodies, RNA was purified and analyzed by Northern blotting.



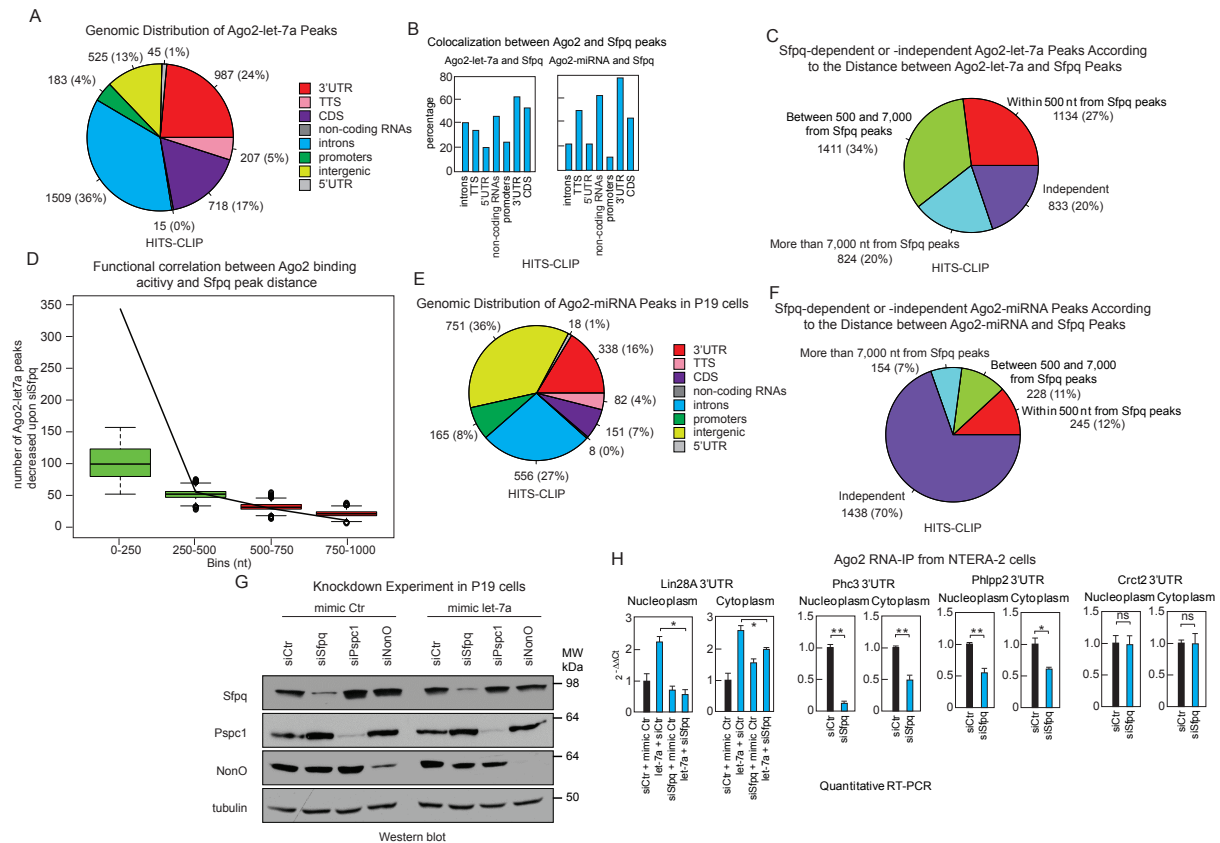
**Supplementary Figure 4. Sfpq, Pspc1 and NonO interact with miRISC in the nucleoplasm.**

(A) Northern blot analysis of chromatin, nucleoplasm, or cytoplasm from HeLa, HEK293T, P19, or retinoic acid (RA)-treated P19 cells using the indicated probes. (B) Co-IP of the indicated endogenous proteins and the indicated proteins from nucleoplasmic or cytoplasmic extracts of HeLa (left panel) or P19 cells (right panel). (C) Co-IP (upper panel) and RNA-IP (lower panel) of the indicated endogenous proteins and let-7a from chromatin extracts of HeLa cells.

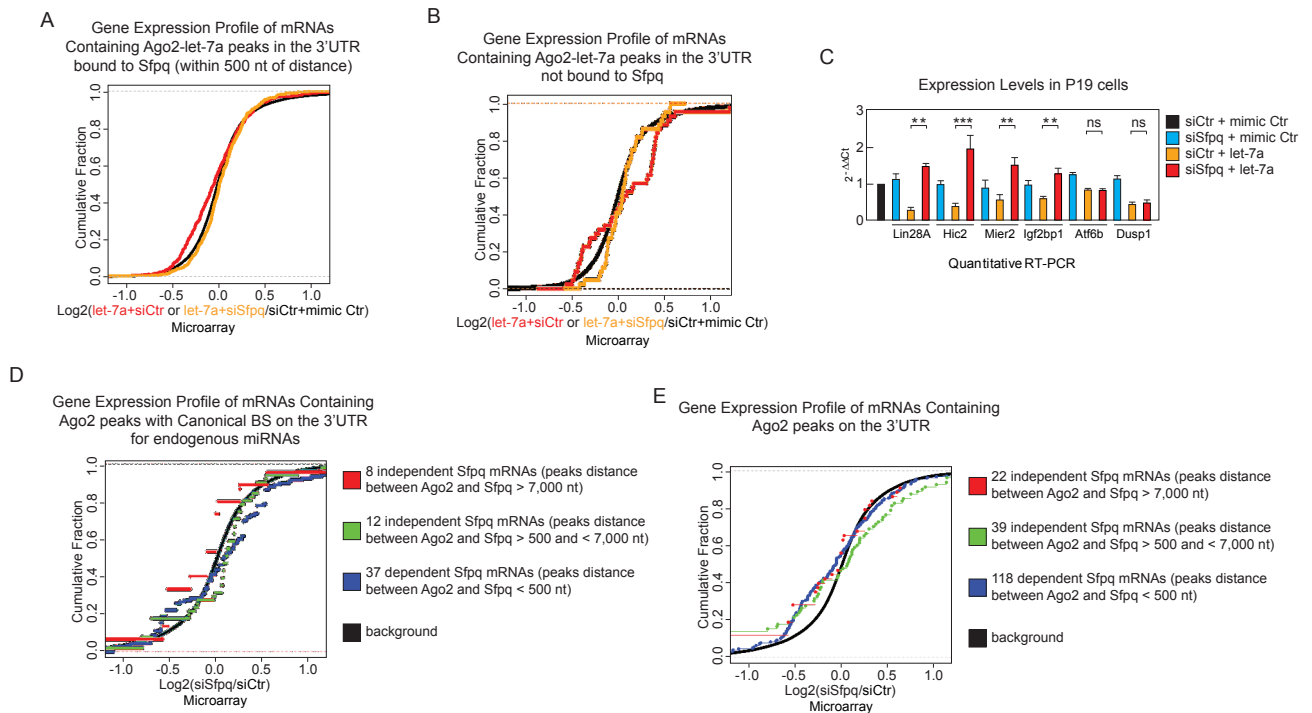


**Supplementary Figure 5. Sfpq associates to miRNAs.** (A) Small RNA  $\log_2(\text{RNA-IP Input}^{-1})$  distribution (dotted line) from small RNA sequencing of Ago2 (left panel) or Sfpq (right panel) RNA-IP from RAW 264.7 cells. Main distributions are composed by two Gaussians: the green line represents small RNA mostly expressed in the RNA-IPs whereas the red line represents the small RNAs mostly expressed in the inputs. A LOD approach was set up to find a threshold to discriminate the two populations (vertical solid line). (B) RNA-IP of Sfpq and miR-24 or miR-125b in RAW 264.7 cells. Total cell extracts were immunoprecipitated and RNA in the immunocomplexes was analyzed by quantitative RT-PCR. Data are normalized with U2 snRNA and presented as the mean  $\pm$  s.e.m. ( $n = 4$ ). (C) P-values of the Ingenuity Pathway Analysis for the Sfpq-associated miRNAs. (D) Relative abundance of small RNAs found in Ago2 HITS-CLIP experiments in let-7a-transfected P19 cells and control. Red: small RNAs enriched in let-7a condition. Blue: small RNAs enriched in mimic control condition (Ctr). Green: small RNAs found in equal levels in both conditions. Black: not significantly enriched small RNAs. (E) RNA-IP of let-7a with Ago2, Sfpq, Pspc1, or NonO in let-7a-transfected P19 cells. Cell extracts were immunoprecipitated from fractionated nucleoplasm or cytoplasm with the indicated antibodies, RNA was purified and analyzed by Northern blotting.

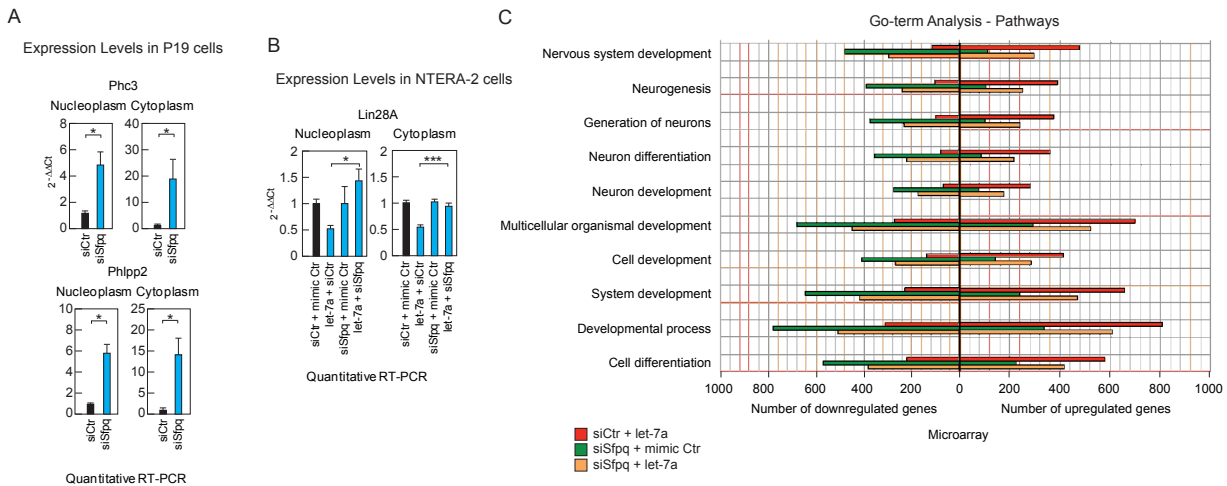




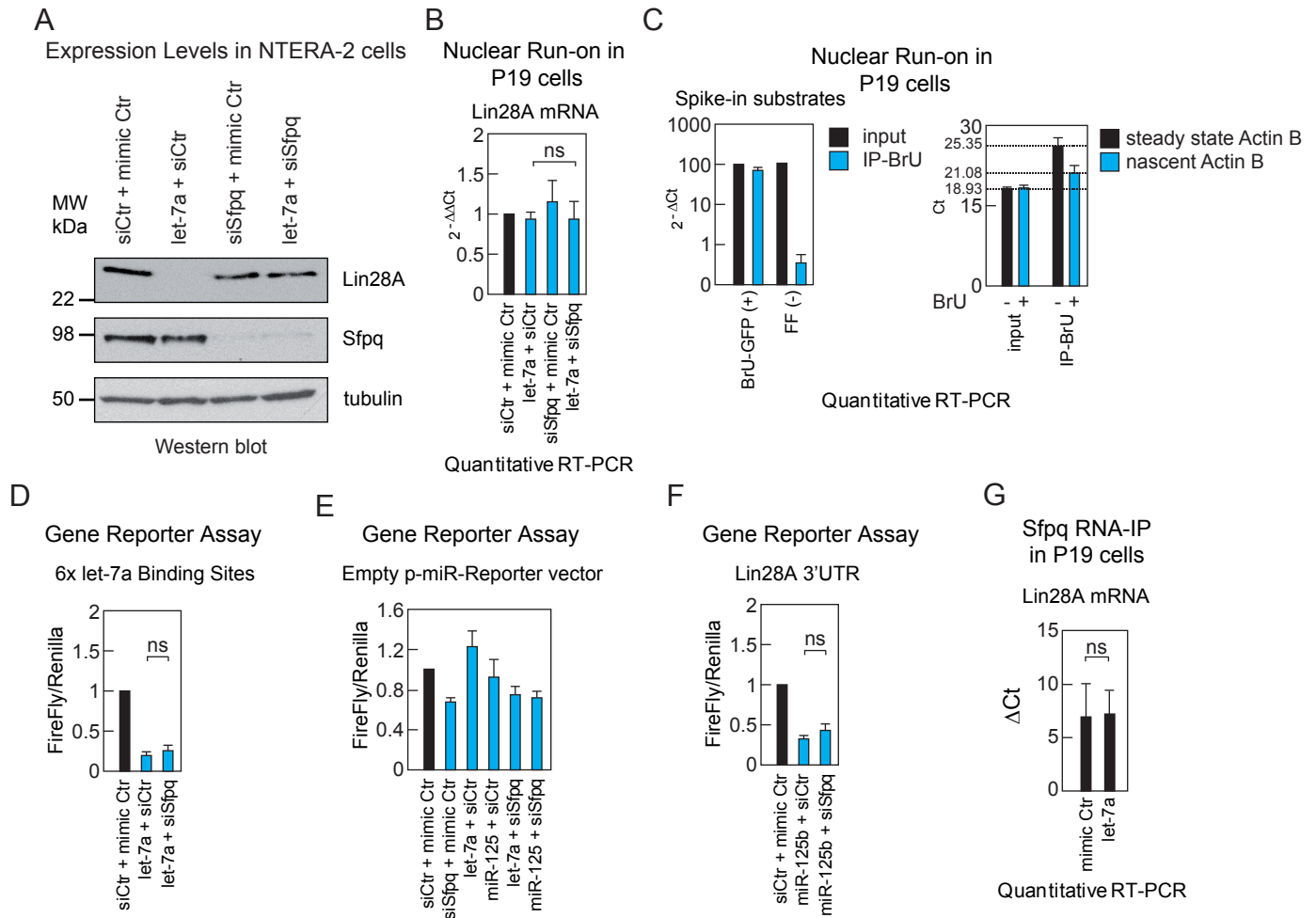
**Supplementary Figure 6. Global analysis of Sfpq-dependent miRNA targeting by HITS-CLIP analysis in P19 cells.** (A) Ago2-let-7a peaks from dCLIP analysis of HITS-CLIP data grouped in different genomic regions. HITS-CLIP experiments were performed in let-7a-transfected P19 cells and mimic control. (B) Genomic distribution of Ago2-let-7a or Ago2-miRNA peaks that colocalize with Sfpq peaks identified by HITS-CLIP analysis. (C) Percentage of the Sfpq-dependent or independent Ago2-let-7a peaks. The Sfpq-dependent Ago2-let-7a peaks were further grouped considering their position to the closest Sfpq peak determined by HITS-CLIP analysis in P19 cells. The groups considered are the following: (i) close distance to Sfpq peaks (less than 500 nt); (ii) far distance to Sfpq peaks (between 500 and 7,000 nt); and (iii) very far distance (more than 7,000 nt). (D) Number of reduced Ago2-let-7a peaks upon Sfpq knockdown in the 3'UTR divided in bins of 250 nt from Sfpq Peaks (black solid line). In the box-plots are represented the 10,000 times shuffling of the distance between Sfpq and the reduced Ago2-let-7a peaks upon Sfpq knockdown in each bin. The colors in the plot indicate whether the Z score was positive (green) or negative (red). (E) Genomic distribution of endogenously expressed miRNA binding sites (Ago2-miRNA peaks) according to Bottini et al.<sup>2</sup>. (F) Genomic distribution of the Sfpq-dependent or independent endogenous Ago2-miRNA peaks. The Sfpq-dependent Ago2-miRNA peaks were further grouped considering their position to the closest Sfpq peak determined by HITS-CLIP analysis in P19 cells. (G) Immunoblot showing the effectiveness and the specificity of the knockdown for Sfpq, Pspc1, and NonO in P19 cells transfected with mimic let-7a or mimic control. (H) NTERA-2 cells were transfected with the indicated molecules before lysate them by fractionating nucleoplasm and cytoplasm. Before Ago2 IP the lysates were partially digested with 10 $\mu$ g ml<sup>-1</sup> RNase A for 30 min at room temperature. RNA was purified from the immunocomplexes and from 5% of input, and analyzed by RT-qPCR using oligonucleotide probes surrounding let-7a binding site for Lin28A 3'UTR and the miR-302b binding sites for Phc3, Phlpp2 or Crct2 3'UTRs, identified by HITS-CLIP analysis. Data are normalized with their own inputs and presented as the mean  $\pm$  s.e.m. (n = 3). Student's t-test for Phc3, Phlpp2 or Crct2 3'UTRs and one-way ANOVA followed by Tukey's *post hoc* test for Lin28A 3'UTR: \*P<0.05, \*\*P<0.01, ns (not significant).



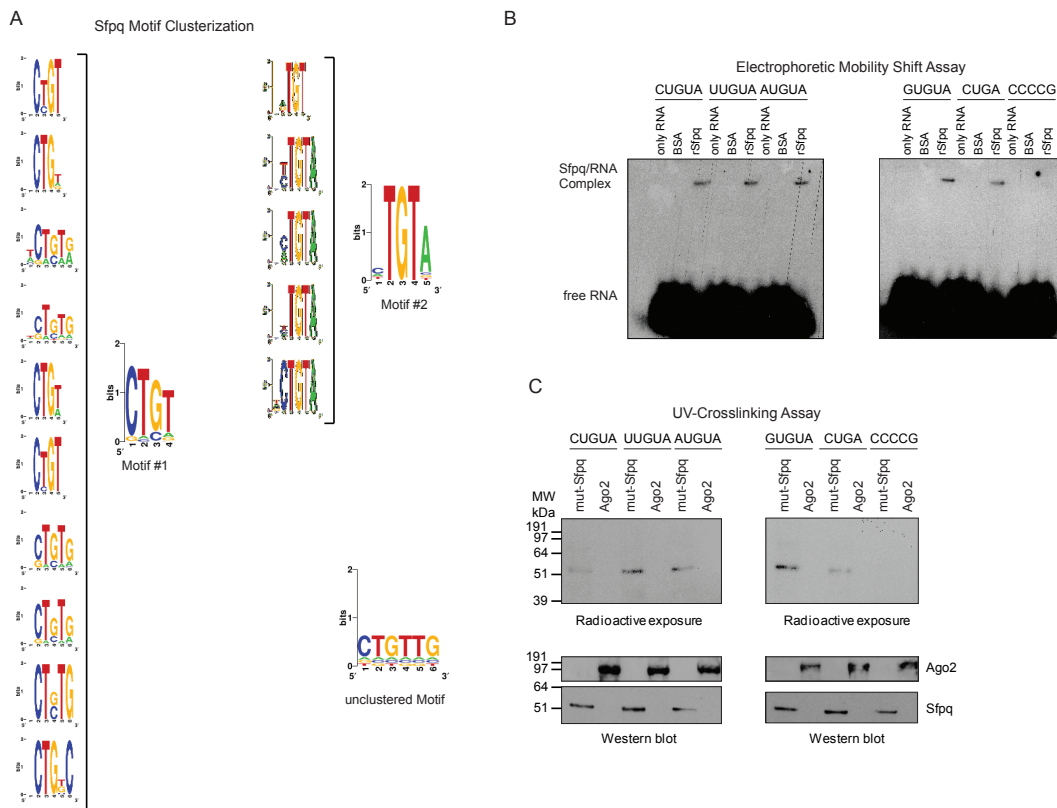
**Supplementary Figure 7. Sfpq promotes post-transcriptional silencing mediated by miRNAs.** Gene expression differences in let-7a-transfected P19 cells upon Sfpq knockdown and control. Expression is plotted for mRNAs containing Ago2-let-7a peaks in the 3'UTR localized in a (A) closely distance to Sfpq peaks (less than 500 nt) or (B) very far distance-unbound to Sfpq peaks (more than 7,000 nt), and upon Sfpq knockdown. (C) Sfpq knockdown inhibits the let-7a-dependent downregulation of selected direct target mRNAs. mRNA levels were estimated by RT-qPCR in let-7a-transfected P19 cells. Data are normalized with U2 snRNA and presented as the mean  $\pm$  s.e.m. ( $n = 6$ ). Gene expression differences in siSfpq-transfected P19 cells and control. Differential expression is plotted for mRNAs containing Ago2 peaks in the 3'UTR (D) bearing only the canonical binding sites (BS) for the 20 most expressed miRNAs or (E) the all set. One-way ANOVA followed by Tukey's *post hoc* test for C: \*\* $P < 0.01$ , \*\*\* $P < 0.001$ , ns (not significant).



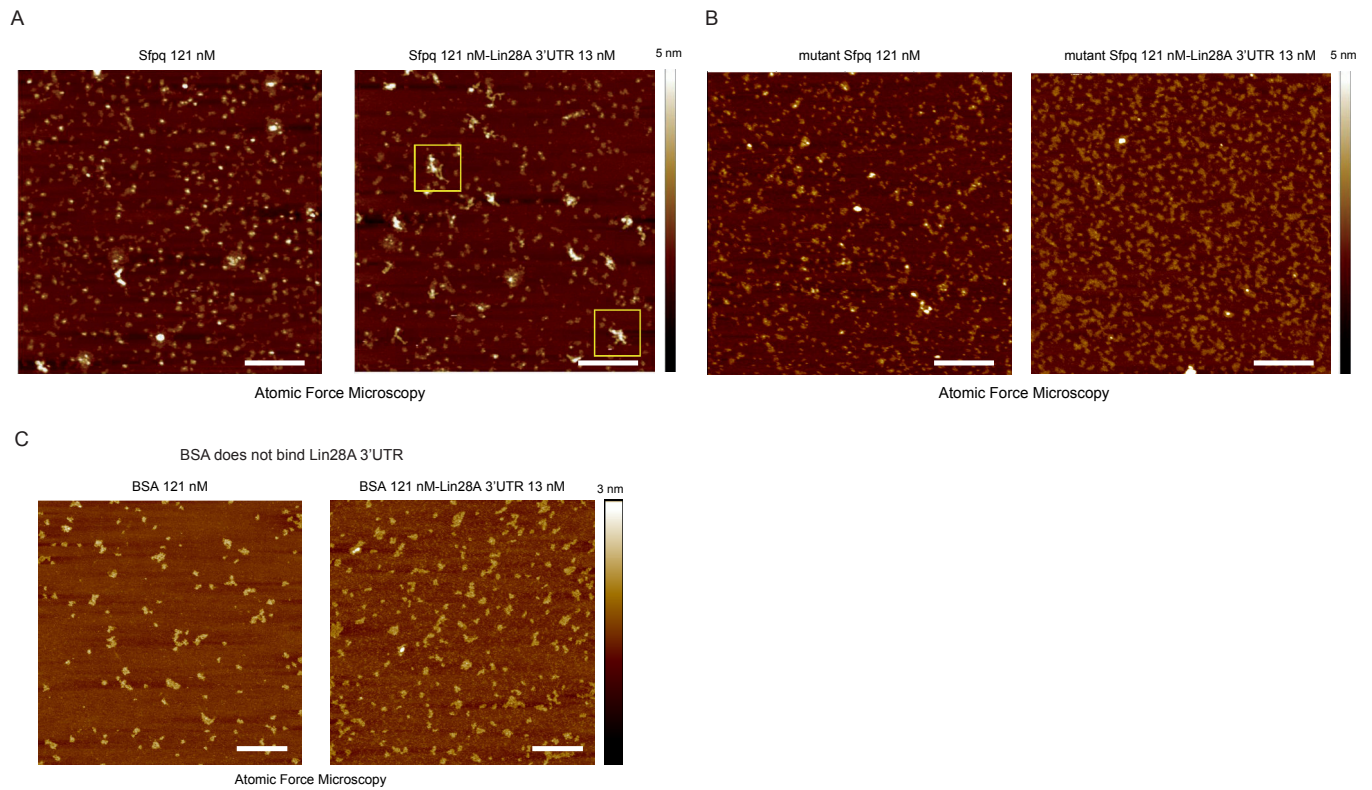
**Supplementary Figure 8. Sfpq promotes post-transcriptional silencing mediated by miRNAs in both nucleoplasm and cytoplasm to regulate gene expression programs.** (A) P19 or (B) NTERA-2 cells were transfected with the indicated molecules before lysate them by fractionating nucleoplasm and cytoplasm to measure the indicated miR-302b-target mRNAs and the let-7a-target Lin28A mRNA, respectively. RNA was purified and analyzed by RT-qPCR. Data are normalized with U2 snRNA and presented as the mean  $\pm$  s.e.m. ( $n = 3$ ). (C) Number of downregulated or upregulated transcripts in the top 10 enriched pathways from GO Term analysis from microarray data of the indicated conditions in P19 cells. Student's t-test for A and one-way ANOVA followed by Tukey's *post hoc* test for B: \* $P < 0.05$ , \*\* $P < 0.01$ , ns (not significant).



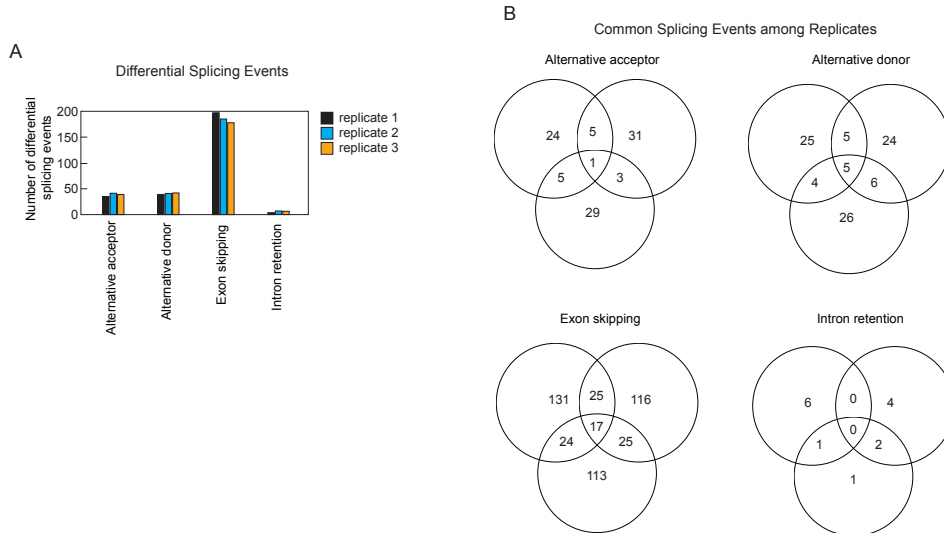
**Supplementary Figure 9. Sfpq promotes post-transcriptional silencing of Lin28A mRNA mediated by let-7a.** (A) Immunoblot analysis of Lin28A, Sfpq, and tubulin in NTERA-2 cells transfected with mimic let-7a and or siSfpq. RT-qPCR analysis of the (B) nascent Lin28A mRNA, (C left panel) spike-in negative (Fire Fly or FF) and positive (GFP) RNA controls, or (C right panel) actin B as endogenously expressed control. Data are presented as the mean  $\pm$  s.e.m. ( $n = 3$ ), and normalized with actin B for Lin28A mRNA, or with the input for both spike-in RNAs. (D) Relative luciferase activity of reporter constructs containing only the sequence of 6 canonical binding sites for let-7a from mouse Lin28a 3'UTR in HEK293T cells transfected with mimic let-7a or siSfpq, as indicated. The data were normalized using Renilla activity. Relative luciferase activity of (E) reporter empty vector sequence or (F) the construct containing the mouse Lin28a 3'UTR in HEK293T cells transfected with the indicated molecules. (G) RNA-IP of Sfpq and Lin28A mRNA in let-7a-transfected P19 cells and control. Total cell extracts were immunoprecipitated and RNA in the immunocomplexes was analyzed by quantitative RT-PCR. Data are normalized with the inputs and presented as the mean  $\pm$  s.e.m. ( $n = 4$ ). Student's t-test for G and one-way ANOVA followed by Tukey's *post hoc* test for B, D and F: ns (not significant).



**Supplementary Figure 10. Search and validation of Sfpq binding motif(s).** (A) Output results of the Sfpq motif clusterization with the program Gimmemotifs. (B) EMSA of the indicated Sfpq-binding RNA sequence and the full length of human recombinant wild-type Sfpq (100 nM). Recombinant protein and  $^{32}\text{P}$ -labeled RNA substrates were incubated at room temperature for 30 min and resolved on 10% polyacrylamide native gel. (C) UV-crosslinking assay to analyze the interaction of the recombinant Sfpq-214–598 quadruple mutant (L535A, L539A, L546A, and M549A) (100 nM) or the full length of human recombinant Ago2 (100 nM) with the indicated  $^{32}\text{P}$ -labeled RNA oligonucleotides containing two copies of the indicated Sfpq-binding RNA sequence or negative control sequence. After radioactive exposure the membrane were subject of immunoblot analyses with the indicated antibodies.

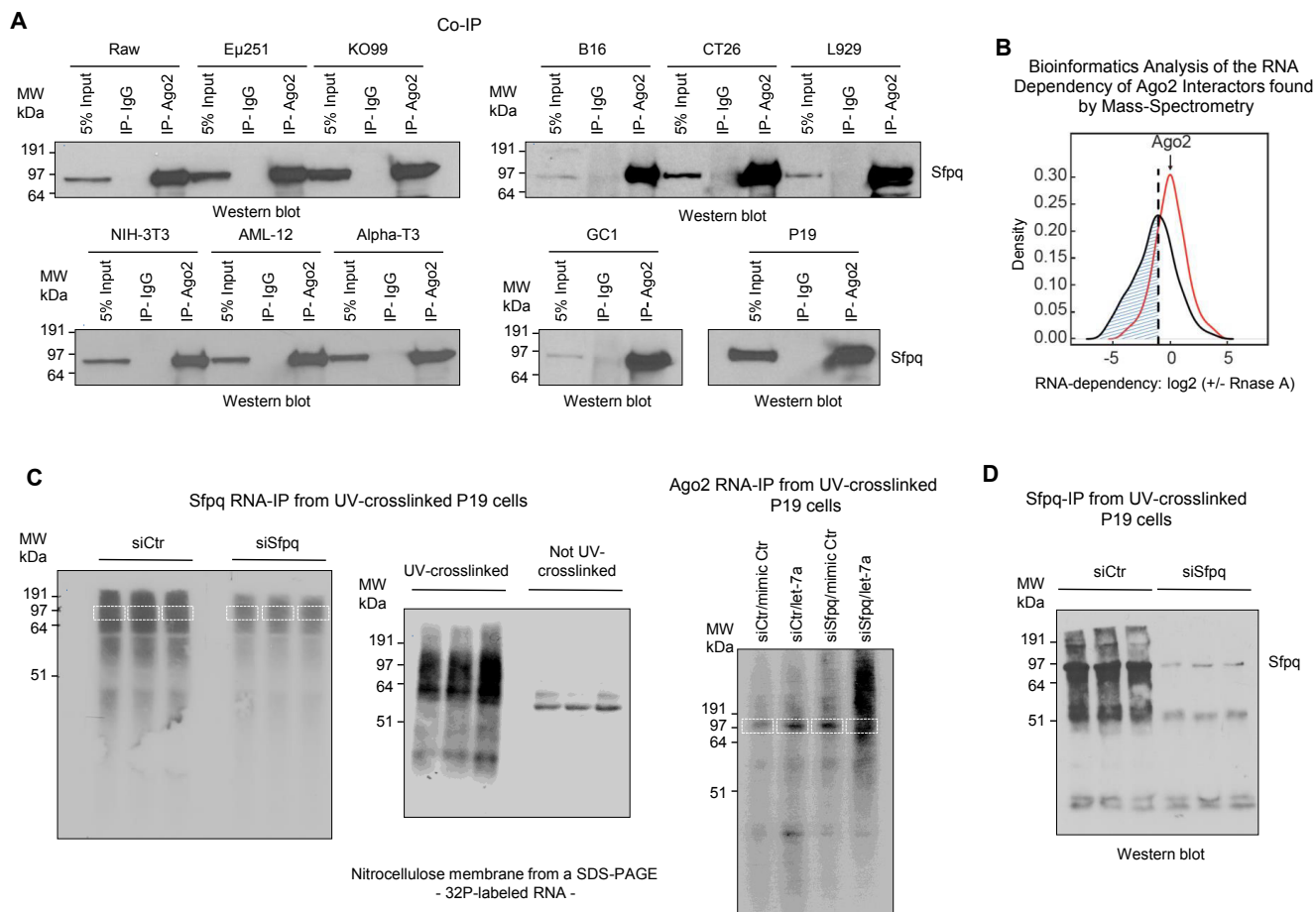


**Supplementary Figure 11. Sfpq aggregates onto Lin28A 3'UTR.** Top view of topographic Atomic Force Microscope (AFM) images of (A) wild-type Sfpq protein (left panel) and wild-type Sfpq-Lin28A 3'UTR RNA complexes (right panel - in yellow squares some complexes are shown) or (B) Sfpq-214–598 quadruple mutant (L535A, L539A, L546A, and M549A) protein (left panel) and mutant Sfpq-Lin28A 3'UTR RNA complexes (right panel) adsorbed on untreated mica surface. The color bar on the right of each image represents the height scale with a maximum of 5nm. Scale bar corresponds to 200 nm. (C) Top view of topographic AFM images of BSA (left panel) and BSA previously incubated with Lin28 3'UTR (right panel). The color bar on the right represents the height scale with a maximum of 3nm. Scale bar corresponds to 200 nm.



**Supplementary Figure 12. Sfpq does not control alternative splicing on the same set of 3'UTRs on which it controls miRNA targeting.** (A) Histogram shows the number of differentially splicing events obtained by “SpliceTrap” program comparing RNA-sequencing samples of siSfpq-transfected P19 cells to siCtr-transfected P19 cells from each replicates, by RNA-sequencing analysis. (B) Venn diagrams show the number of differentially splicing events in common between the three replicates for each event category.





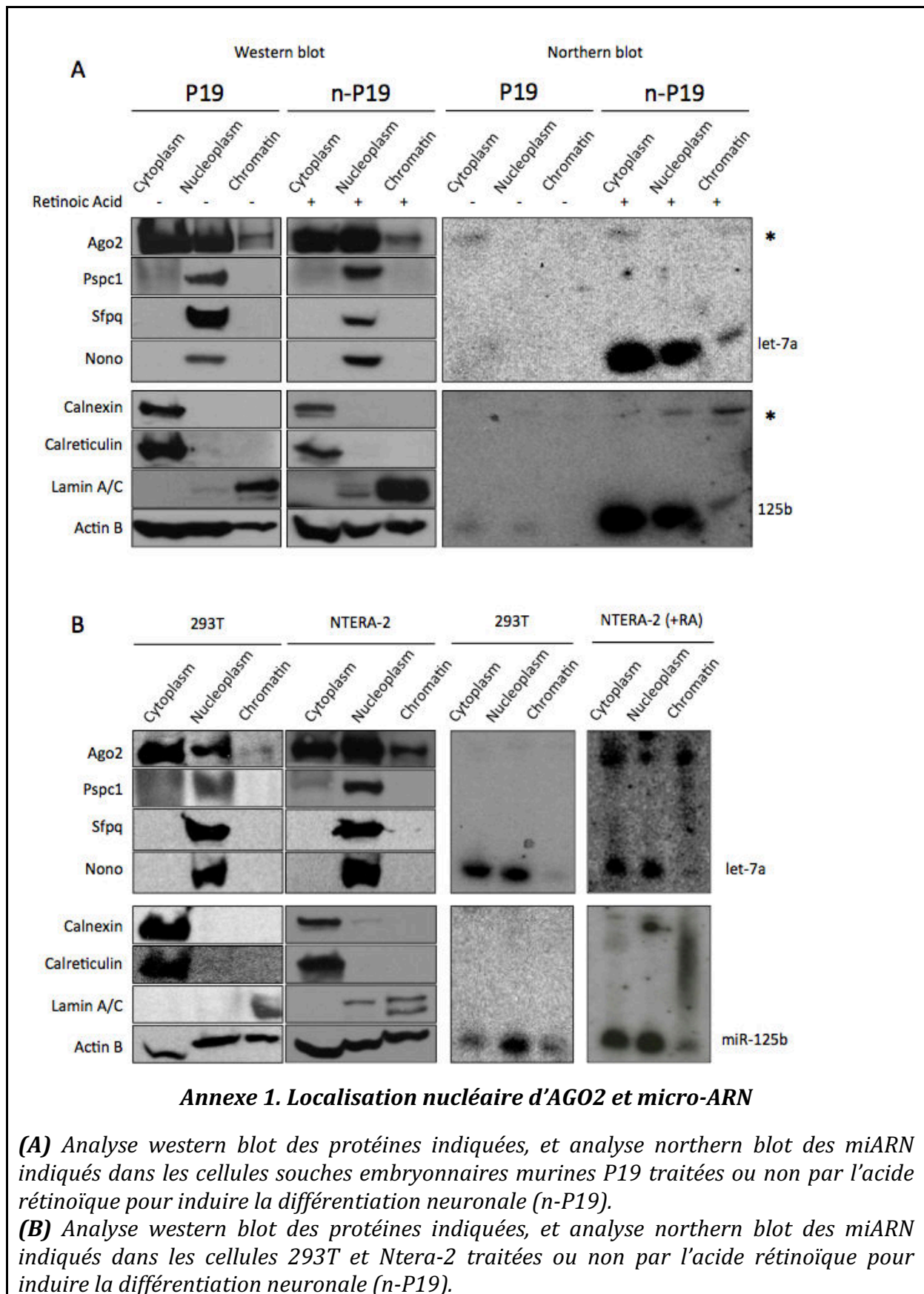
**Supplementary Figure 13.** (A) Sfpq and Ago2 interaction is conserved in various mouse cell types. Co-IP of endogenous Sfpq and Ago2 in the indicated cell lines. RAW 264.7 cells are macrophages; Eμ251 and KO99 cells are lymphocytes; B16 cells are melanocytes; CT26 cells are colon carcinoma cells; L929 and NIH-3T3 cells are fibroblasts; AML-12 cells are hepatocytes; alpha-T3 are from pituitary; GC1 are germ cells; P19 cells are embryonic stem cells. (B) Bioinformatics analysis about the RNA-dependency of Ago2 interactors from the MS analysis. Density plot of the log base 2 (-/+ RNase ratios (abundance scores) from all the data (shown in black) and the null distribution (shown in red). The vertical dashed black line indicates the 10% LOD cutoff that was used to classify a protein as RNA-dependent interactor (on the left of the dashed line). See the data analysis section of methods for an explanation of how the null distribution was calculated. (C) IP of either Sfpq-RNA complex (left panel) or Ago2-RNA complex (right panel) labeled with <sup>32</sup>P from P19 cells. The protein-RNA complex in dashed box was purified and used to prepare libraries for HITS-CLIP analysis. Illustrative images are shown. As control experiment for Sfpq HITS-CLIP experiment, we ran a protein gel of either UV-crosslinked or not UV-crosslinked Sfpq immunoprecipitated samples labeled with <sup>32</sup>P from P19 cells (middle panel). (D) As a further control for Sfpq HITS-CLIP experiment, we also immunoprecipitated endogenous Sfpq from UV-crosslinked P19 cells upon Sfpq knockdown or control followed by Sfpq immunoblotting.

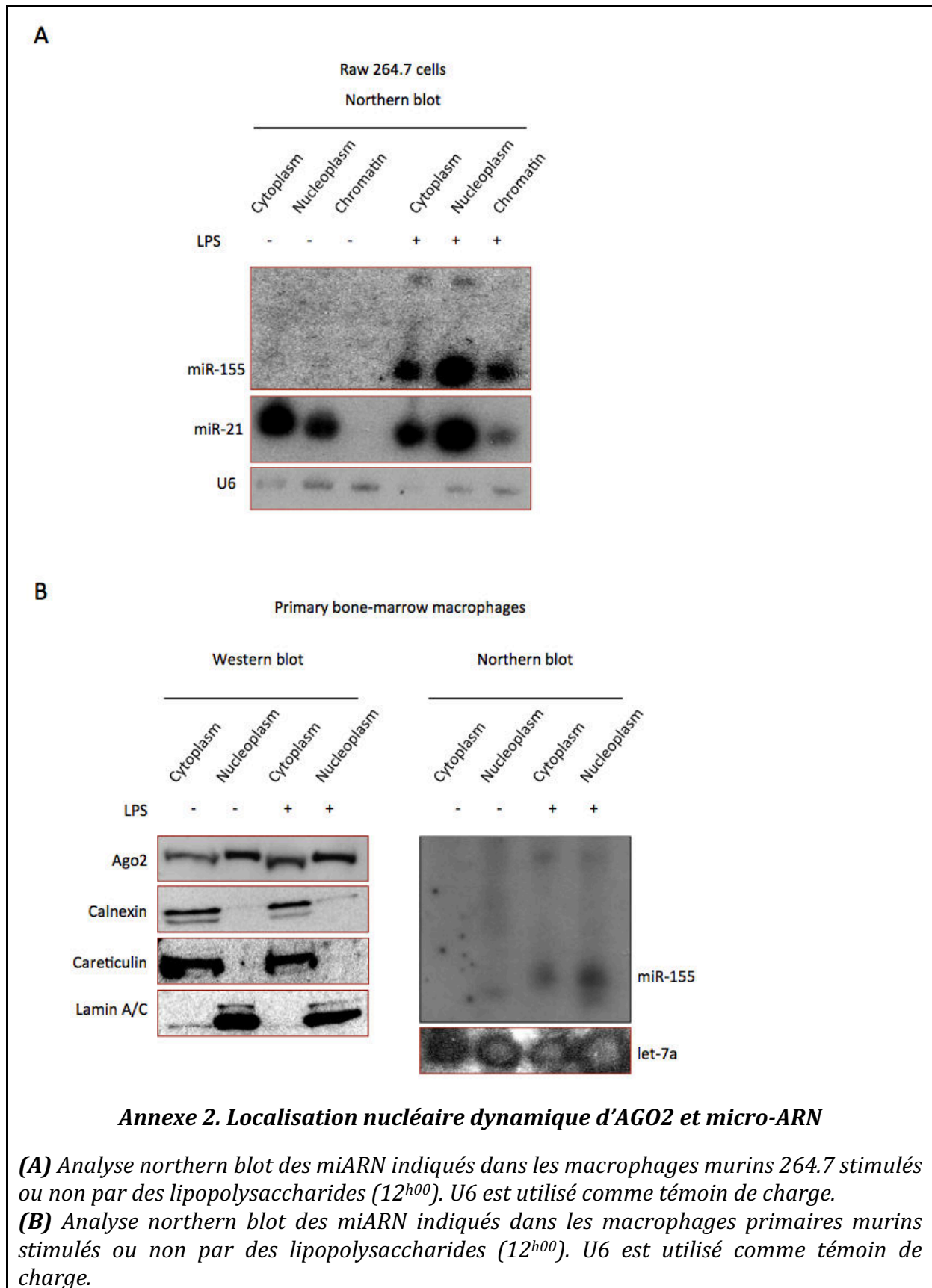


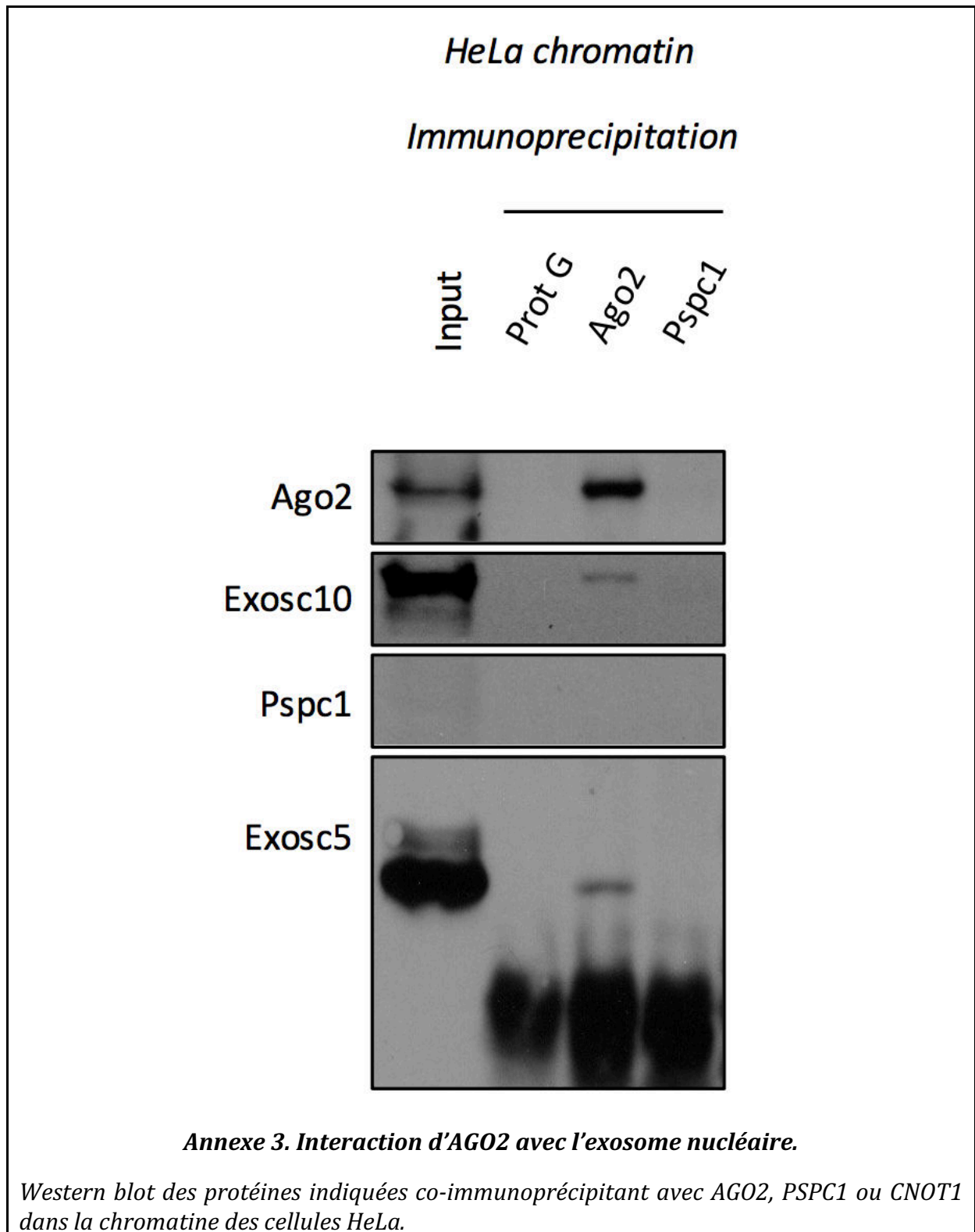
**Supplementary References**

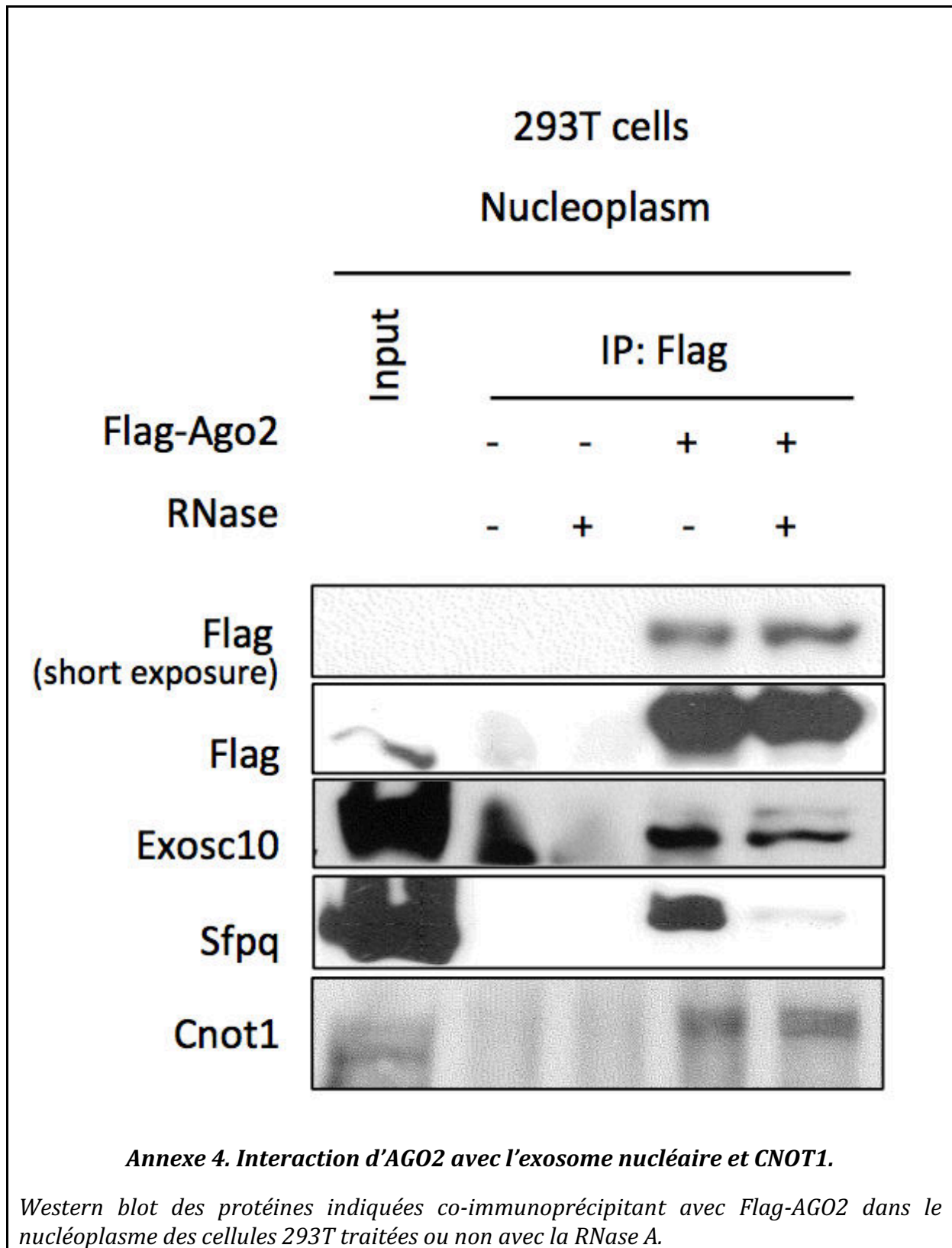
1. Supek F, Bosnjak M, Skunca N, Smuc T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* **6**, e21800 (2011).
2. Bottini S, *et al.* From benchmarking HITS-CLIP peak detection programs to a new method for identification of miRNA-binding sites from Ago2-CLIP data. *Nucleic Acids Res* **45**, e71 (2017).

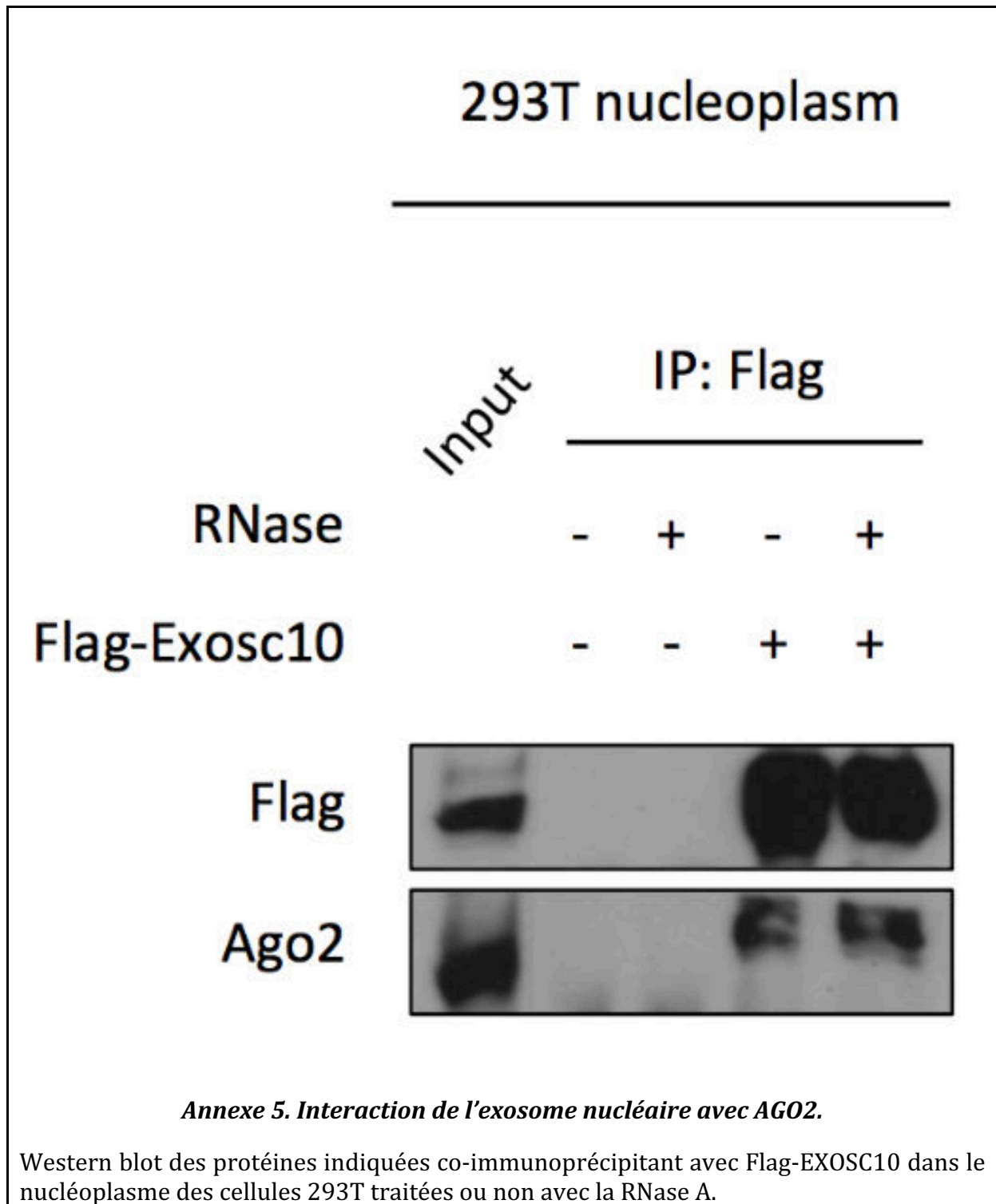
## 8.4 Annexes











## 9 Discussion

Au cours de ce travail de thèse, nous nous sommes intéressés à l'activité de dégradation induite par les miARN nucléaires sur des cibles partiellement complémentaires.

Dans la première partie du projet, nous avons montré que l'activité de dégradation de l'ARNm induite par les miARN est préservée dans le noyau chez les mammifères, et peut être modulé par la protéine de liaison à l'ARN SFPQ (**Article 1. Post-transcriptional gene silencing mediated by microRNAs is controlled by nucleoplasmic SFPQ**).

Dans la seconde partie du projet nous nous sommes intéressés au mécanisme d'action des miARN nucléaires, et nous avons identifié une interaction directe des protéines AGO2 avec CNOT1 et l'exosome dans le noyau, qui est indépendante de l'ARN.

### 9.1 Première partie du projet

#### 9.1.1 Confrontation des résultats avec les données de la littérature (1)

Plusieurs études tendent à confirmer les résultats obtenus pour l'interaction ARN-dépendante de AGO2 avec les protéines DBHS. En effet, la présence des protéines DBHS a été observée dans certaines spectrométries de masse de la protéine AGO2<sup>314,393,401</sup>. En particulier, une étude basée sur la méthode de biotinylation en culture (BioID et Split-BioID) a identifié la protéine NONO parmi les quelques protéines proximales de la protéine appât AGO2 fusionnée au domaine biotine-ligase bactérien<sup>402</sup>.

#### 9.1.2 Fonctions connues des protéines DBHS et paraspeckles

Les protéines de la famille DBHS, dont le membre fondateur est la protéine NONO-1, sont conservées jusqu'aux bilatériens<sup>403</sup>. Chez les vertébrés, la famille est composée de trois paralogues soit NONO, SFPQ et PSPC1<sup>404</sup>.

Les protéines DBHS sont constituées d'une partie N-terminale comprenant deux domaines de liaison à l'ARN (RRM2), un domaine central NonA/paraspeckle (NOPS), un domaine coiled-coil qui médie la dimérisation ou l'oligomérisation/agrégation des protéines et un signal de localisation nucléaire du côté C-terminal<sup>405</sup>.

La dimérisation des protéines DBHS, soit par homodimérisation ou hétérodimérisation, est essentielle pour leur fonction et leur intégrité dans la cellule<sup>406</sup>, et dépendrait de leur concentration relative<sup>407</sup>.

Les protéines DBHS sont localisées dans les corps nucléaires appelés paraspeckles, dont le membre fondateur est la protéine PSPC1 (marqueur spécifique du paraspeckle) <sup>408</sup>. Cependant, c'est l'expression des protéines SFPQ et NONO qui est essentielle pour l'intégrité des paraspeckles, ainsi que l'ARN non-codant NEAT1<sup>409-411</sup>.

Les protéines DBHS sont multifonctionnelles, et impliquées dans plusieurs processus nucléaires comme l'initiation de la transcription, la terminaison de la transcription, l'épissage alternatif, la réparation de l'ADN et le contrôle du rythme circadien<sup>412-422</sup>.

Dans certains contextes neuronaux, les protéines SFPQ ou NONO auraient également une localisation et des fonctions cytoplasmiques<sup>412,423-427</sup>.

### **9.1.3 Interaction entre les protéines DBHS et le complexe miRISC**

Les résultats de notre spectrométrie de masse indiquent que les protéines DBHS figurent parmi les partenaires ARN-dépendants les plus abondants du complexe miRISC. Dans ce cadre nous pensons que la propriété d'agrégation fonctionnelle des protéines SFPQ, NONO et PSPC1 participe à un mécanisme de programmation de la structure du 3'UTR dans le noyau, ce qui module le ciblage par les miARN. De plus, différentes combinaisons de dimères des protéines DBHS pourraient influencer le type d'agrégation et moduler la programmation du ciblage par les miARN<sup>404,428</sup>.

### **9.1.4 Programmation du ciblage par les miARN dans le noyau**

De plus, nous observons également l'action de SFPQ sur les ARNm cytoplasmiques bien que SFPQ soit uniquement nucléaire, ce qui suggère que la liaison et l'agrégation de la protéine SFPQ programme le ciblage du 3'UTR par les miARN dans le noyau, d'une manière structurale qui semble préservée dans le cytoplasme.

Ce concept de détermination nucléaire d'événements cytoplasmiques est observé pour certaines protéines du complexe exon-jonction (EJC) comme Aly/REF ou Y14/Magoh qui sont principalement nucléaires. En effet, l'export et la destination cytoplasmique des ARN messagers est stimulée par le dépôt des protéines Aly/REF ou Y14/Magoh du complexe EJC sur les ARN messagers épissés<sup>429</sup>.



### 9.1.5 Fonction des protéines DBHS et paraspeckles dans la biogénèse des miARN

Les protéines DBHS et l'ARN non-codant NEAT1 constituent un sous-compartiment nucléaire appelé paraspeckles contenant >60 protéines, et est important pour la rétention nucléaire d'ARN hyper-édités (A->I), en particulier ceux qui contiennent des rétrotransposons de type Alu<sup>430</sup>.

L'identification des protéines DBHS dans notre spectrométrie de masse de la protéine AGO2 a soulevé un intérêt quant au rôle des paraspeckles dans la biologie des miARN.

Dans ce cadre, un article récent montre que la protéine FUS, une des protéines essentielles pour l'intégrité des paraspeckles<sup>431</sup>, semble importante pour l'activité des miARN. En effet, la délétion du gène FUS compromet le ciblage d'un gène rapporteur par un siARN, mais aussi le ciblage de miARN comme miR-200c. Enfin, une version mutée de la séquence NLS de la protéine FUS qui empêche sa translocation dans le noyau compromet le ciblage par les miARN<sup>432</sup>. L'ensemble de ces résultats suggère que la protéine FUS participe au mécanisme d'action des miARN.

Récemment, une étude a proposé une nouvelle fonction pour les corps nucléaires paraspeckles<sup>97</sup>. En effet, les composants des paraspeckles (en particulier SFPQ, NONO) co-immunoprécipitent avec les facteurs de biogénèse DGCR8/DROSHA d'une manière ARN-dépendante et NEAT1-dépendante. De plus, la déplétion de SFPQ, NONO ou NEAT1 entraîne un défaut de processing de la majorité des pri-miARN exprimés dans les cellules HeLa. Enfin, le complexe microprocesseur semble s'associer à la périphérie des paraspeckles, d'une manière qui semble favorisée par la structure en pri-miARN de l'extrémité 3' de l'ARN non-codant NEAT1. D'après les résultats de cette étude, il semble que les paraspeckles constituent un foyer d'assemblage de facteurs de biogénèse et de stimulation du processing des pri-miARN<sup>97</sup>. En effet, d'autres facteurs impliqués dans la biogénèse des miARN comme TDP-43<sup>433</sup> sont localisés à l'enveloppe externe des paraspeckles<sup>96,434</sup>.

## 9.2 Deuxième partie du projet

### 9.2.1 Confrontation des résultats avec les données de la littérature (2)

Plusieurs études tendent à confirmer les résultats obtenus pour l'interaction de AGO2 avec l'exosome nucléaire. En effet, il a été montré précédemment que la protéine AGO2 co-immunoprécipite spécifiquement avec les protéines EXOSC10 et DIS3<sup>435</sup>. De plus, l'interaction du complexe RISC avec l'exosome semble indépendante de l'ARN comme il est possible de l'observer dans certaines spectrométries de masse de la protéine GW182 à partir de lysat nucléaire traité à la RNase<sup>314</sup>. Enfin, il a aussi été observé que l'exosome nucléaire<sup>i</sup> co-immunoprécipite avec une version biotinylée de let-7d, dans des cellules épithéliales de poumon de souris<sup>436</sup>.

### 9.2.2 Une interaction potentielle de l'exosome avec le complexe CCR4-NOT1

Dans la seconde partie du projet nous nous sommes intéressés au mécanisme d'action des miARN nucléaires. Sur la base de nos résultats et des données actuelles, nous pensons que AGO2 est connectée à l'exosome par l'intermédiaire du complexe CCR4-NOT1<sup>ii</sup>.

En effet, les interactions des protéines AGO2 nucléaires avec le complexe CCR4-NOT1, les protéines TNRC6, DICER ou HSP90 apparaissent préservées et stables dans le noyau<sup>313,314</sup>. Les interactions avec le complexe CCR4-NOT1 sont notoirement moins stables<sup>402</sup>, bien que le complexe soit un partenaire connu du RISC. La nature modulaire du complexe CCR4-NOT1 rend compte de plusieurs activités telles que l'ubiquitination de certaines protéines<sup>438</sup>, et a été initialement caractérisé dans un processus de régulation de la transcription chez la levure<sup>396</sup>, puis impliqué dans le catabolisme de l'ARN<sup>439</sup> avant d'être relié au mécanisme d'action des miARN<sup>164</sup>.

Nous souhaitons mettre en avant l'identification de l'exosome nucléaire (**Boîte 1**) en tant que nouveau composant intégral du complexe RISC.

En effet, une interaction directe entre le complexe CCR4-NOT1 et l'exosome a été rapportée chez la levure bourgeonnante<sup>437</sup>, bien que le modèle proposé soit insatisfaisant en l'absence d'investigations complémentaires.

---

<sup>i</sup> Contenant aussi la protéine C1D, un des co-facteurs de l'exosome

<sup>ii</sup> Comme proposé par Collart et al., 2009 (réf <sup>437</sup>), développé en 9.2.2.1.

Nous pensons que l'exosome est directement recruté par le complexe CCR4-NOT1 via le module CNOT2-CNOT3 (**Figure 14**). Ci-dessous, nous développons brièvement une argumentation soutenant cette hypothèse de travail.

### 9.2.2.1 Des indices pour le rôle du module CNOT2-CNOT3

Les gènes du complexe Not (negative on tata-less) ont été initialement caractérisés chez la levure *Saccharomyces Cerevisiae*, à partir de souches mutantes (sensibles à la température) qui activent la transcription de manière globale après avoir été placées à la température restrictive<sup>395,396,440</sup>.

Le phénotype Not1 résulte de mutations dans le gène Not1 qui compromettent la fonction du domaine C-terminal de la protéine<sup>440,441</sup>. Il a été montré par la suite que ce domaine recrute normalement le module Not2-Not3<sup>399</sup>, qui induit la dégradation de l'ARN dans les expériences d'ancrage à un gène rapporteur<sup>442</sup>, et apparaît important pour l'intégrité de complexes CCR4-NOT1 à haut poids moléculaire sur gradients de densité ou des analyses de filtration sur gel dans certaines études<sup>443,444</sup>.

Le phénotype Not2 résulte de mutations dans le gène Not2 qui déstabilisent l'ancrage à la protéine Not1, ou de mutations dans le domaine C-terminal appelé Not-box et qui est caractéristique du gène Not2 et du gène Not3<sup>398</sup>.

Il a été proposé que le module Not2-Not3 intervient dans la dégradation de l'ARN, soit par la modulation de l'activité des autres composants du complexe CCR4-NOT1<sup>399</sup> ou bien en portant une fonction propre de recrutement d'autres facteurs<sup>445</sup>.

Dans ce cadre, certaines propriétés du complexe CCR4-NOT1-CAF1 compliquent son étude fonctionnelle. En effet, l'expression des gènes *Not* est coregulée<sup>398,444,446-448</sup>, tel que dans plusieurs études l'inhibition par exemple du gène *Cnot2* par ARNi entraîne en réaction l'inhibition du gène *Cnot1*, *Cnot3*<sup>399</sup>, mais aussi *Cnot7*, *Cnot9*<sup>444,447</sup>. De fait, il est difficile dans un système cellulaire d'isoler la contribution d'une seule des protéines de l'ensemble des fonctions du complexe CCR4-NOT1, comme par exemple l'activité de déadénylation (CCR4 ou CNOT7/CAF1) ou de décoiffage en 5' de l'ARN (domaine MIF4G de CNOT1 et interaction avec DDX6)<sup>i</sup>. De plus, en raison de la nature modulaire du complexe CCR4-NOT, chaque protéine Not peut potentiellement nucléer l'assemblage entier du complexe, ce qui complique les expériences d'ancrage à un gène rapporteur<sup>442</sup>.

---

<sup>i</sup> Voir section 7.6.3.2.

Toutefois, une étude fonctionnelle permet d'avancer l'hypothèse selon laquelle le module Not2-Not3 joue un rôle spécifique dans la dégradation de l'ARN. En effet, le fragment C-terminal de CNOT1, qui soutient uniquement l'assemblage du module Not2-Not3, induit la dégradation de l'ARN dans les expériences d'ancrage à un gène rapporteur<sup>442</sup>. Accessoirement, les mutants tronqués de CNOT1 sans domaine de liaison au module CNOT2-CNOT3 échouent à compléter la diminution de la dégradation d'un gène rapporteur dans les cellules où le gène CNOT1 a été éliminé par ARNi, sans compromettre pour autant l'intégrité de la déadénylase CNOT7 dans le complexe<sup>399</sup>.

Au niveau structurel, la diffraction aux rayons X d'un crystal du complexe CNOT1-CNOT2-CNOT3 révèle une structure en feuillet- $\beta$  de type SH3 du domaine Not-box des protéines Not2-Not3, argumentant pour un rôle adaptateur d'interactions protéine-protéine de ce domaine (**Figure 15**). Les deux gènes Not-box (Not2, Not3) sont conservés chez tous les eucaryotes<sup>i</sup>, tandis que les levures possèdent un gène Not-box supplémentaire appelé Not5, qui peut dimériser avec Not2 pour former des complexes Not1-Not2-Not5 de structure similaire au module Not1-Not2-Not3.

Parmi les gènes Not-box, c'est le gène CNOT2 qui semble le plus critique pour la fonction du module Not, en raison de l'observation que le gène CNOT3 a fait l'objet de réarrangements dans le génome des oiseaux et a été entièrement perdu<sup>ii</sup>, à l'exception d'une espèce de moineau japonais<sup>449</sup>. Chez certains oiseaux, il subsiste uniquement un fragment résiduel du domaine N-terminal coiled-coil spécifique au gène CNOT3. Toutefois, le gène CNOT3 est essentiel chez les mammifères<sup>450</sup>. Ces observations suggèrent que cet épisode de réarrangement génomique a été précoce dans la phylogénie aviaire, et que chez la plupart des oiseaux l'intégrité du module Not repose sur l'unique gène Not-box restant (CNOT2), probablement par homodimérisation de la protéine CNOT2. En effet, les protéines humaines CNOT2 ou CNOT3 forment des homodimères en solution individuelles, mais en présence de l'une et l'autre la combinaison qui est favorisée est l'hétérodimère, dont la conformation est stabilisée par la présence de CNOT1<sup>399</sup>.

---

<sup>i</sup> Le choanoflagellé *Nebria brevicollis* a perdu Not2 et Not3, voir BLAST.

<sup>ii</sup> D'après des recherches avec l'algorithme BLAST et la base de données aviaire.

Certaines observations suggèrent que le gène CNOT3 participe au mécanisme de recrutement de l'exosome. Chez la levure, l'ancrage de la protéine Not3 à un gène rapporteur induit la dégradation de l'ARN d'une manière indépendante de Ccr4<sup>445</sup>. Cependant, parmi les gènes du complexe CCR4-NOT1, c'est la délétion du gène Not2 qui semble causer la dérégulation du plus grand nombre de gènes sur puce microarray (898 gènes augmentés, 354 gènes diminués) avec une augmentation de l'ensemble des gènes de l'exosome, mais uniquement pendant la phase de croissance exponentielle<sup>451</sup>.

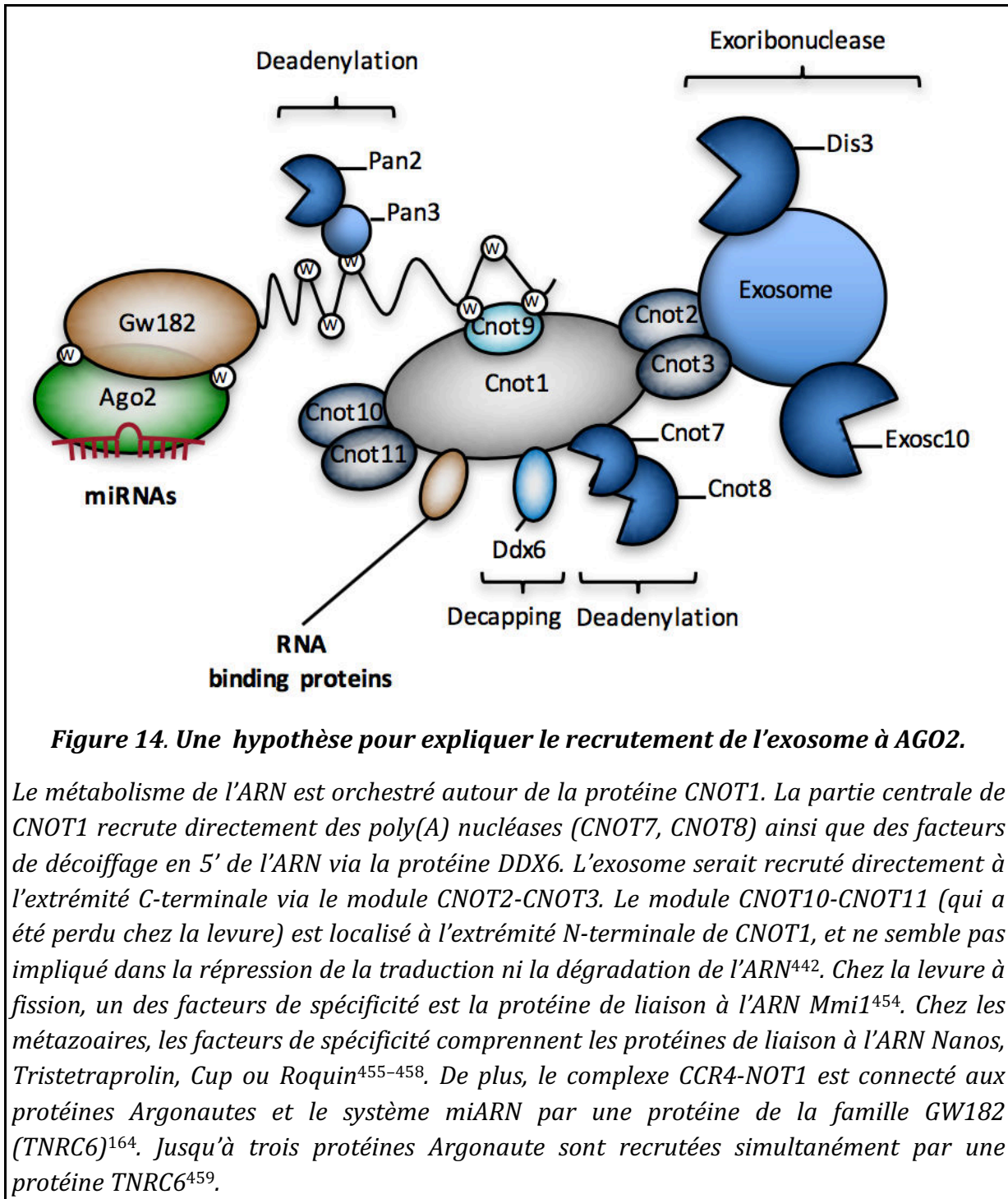
Dans le cadre de l'hypothèse d'un mécanisme relié à la régulation de la transcription<sup>i</sup>, il a été observé que la déplétion du gène CNOT3 (mais aussi CNOT2 ou CNOT1) dans des cellules humaines entraîne l'augmentation de la transcription de certains gènes comme ceux du complexe majeur d'histocompatibilité<sup>452</sup>. Accessoirement, des mutations dans le domaine N-terminal<sup>ii</sup> ont été observées dans des cas de leucémies lymphoïdes aiguës de type T<sup>453</sup>.

L'ensemble de ces données suggère que le module Not2-Not3 porte une fonction conservée essentielle, qui semble impliquée dans la dégradation de l'ARN. Il est tentant de spéculer que la fonction de ce module est le recrutement direct de l'exosome par le complexe CCR4-NOT1, sur lequel reposerait donc le mécanisme de ciblage de l'exosome. Dans un premier temps, il serait intéressant de valider ces résultats et hypothèses par des méthodes supplémentaires d'étude d'interaction protéine-protéine ainsi que des études fonctionnelles ciblées sur le module CNOT2-CNOT3 et l'exosome nucléaire (abordé dans la section 8.2.4). Le cas échéant, il serait intéressant de caractériser la structure du complexe.

---

<sup>i</sup> Hypothèse développée dans le paragraphe suivant, et représenté **Figure 16**.

<sup>ii</sup> Le domaine coiled-coil de fonction inconnue spécifique au gène CNOT3.



### **Boîte 1. L'exosome: Structure, co-facteurs et cibles**

L'exosome est un complexe à activité 3'→5' exoribonucléase essentiel et conservé chez tous les eucaryotes et certaines archaebactéries. L'exosome des eucaryotes est composé de 9 sous-unités structurales, (EXOSC1 à EXOSC9) qui s'assemblent en une structure sphérique chapeautée par deux sous-unités à activité exoribonucléase, EXOSC10 et DIS3 (**Figure 15**). L'ARN monocaténaire pénètre dans l'exosome par le chapeau K-homology KH (EXOSC1 à EXOSC3), et est dégradé par l'activité distributive de EXOSC10. Alternativement, l'ARN progresse davantage dans la structure en anneau pleckstrin-homology PH (EXOSC4 à EXOSC9) pour atteindre l'activité processive de DIS3<sup>460,461</sup>.

L'activité de l'exosome est stimulée par une hélicase appelée SKIV2L2 (Mtr4), qui facilite le dépliement des structures secondaires de l'extrémité 3' des ARN et leur insertion dans l'exosome<sup>462</sup>. SKIV2L2 (Mtr4) fait partie du complexe de polyadénylation TRAMP qui contient la protéine à doigt de zinc Air2p et la poly(A) polymérase Trf4p. Le complexe TRAMP agit comme co-facteur de l'exosome et stimule la dégradation de l'ARN, y compris l'ARN ribosomal<sup>462</sup>. Le recrutement de l'hélicase SKIV2L2 est stimulé par les protéines NOP53 et Utp18, qui facilitent la maturation de l'ARN ribosomal et la dégradation des séquences 5' externes transcrites espacées (ETS1)<sup>463</sup>.

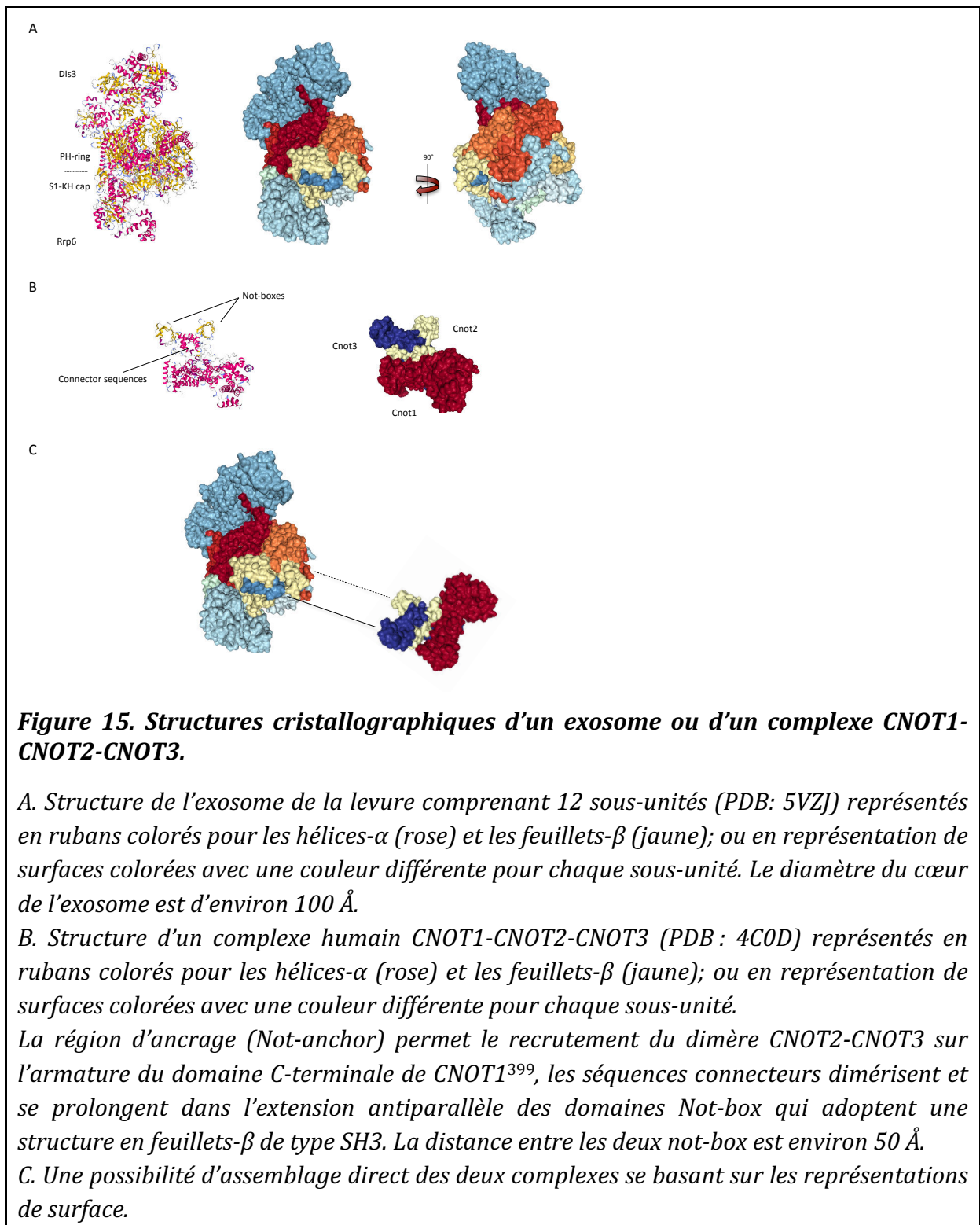
Chez *Saccharomyces Pombe*, le gène Rrp6 (EXOSC10) interagit avec le facteur d'interception et dégradation des transcrits méiotiques Mmi1<sup>464,465</sup>, une protéine de liaison à l'ARN qui interagit aussi avec le complexe CCR4-NOT1<sup>446,454,464-466</sup><sup>466465464463462</sup>. Chez *Saccharomyces Cerevisiae*, un complexe protéique d'interaction avec l'ARN composé de Nrd1-Nab3-Sen1 (NNS) recrute l'exosome de manière co-transcriptionnelle en interagissant avec le complexe TRAMP, et induit la dégradation de transcrits codants et non-codants<sup>429</sup>. Le complexe NNS est aussi impliqué dans la surveillance transcriptionnelle et la terminaison de la transcription dans les cas où l'ARN Pol II saute le signal de polyadénylation<sup>468</sup>.

Chez l'humain, le complexe de ciblage de l'exosome nucléoplasmique (NEXT) comprend les protéines de liaison à l'ARN RBM7 et ZCCHC8<sup>469</sup>, cependant la contribution de ces protéines à la spécificité de ciblage de l'exosome semble limitée<sup>470</sup>. La protéine DGCR8 a été impliquée dans le ciblage de l'exosome sur des petits ARN nucléolaires (snoARN), mais le mécanisme de recrutement de DGCR8 n'est pas encore connu<sup>471</sup>.

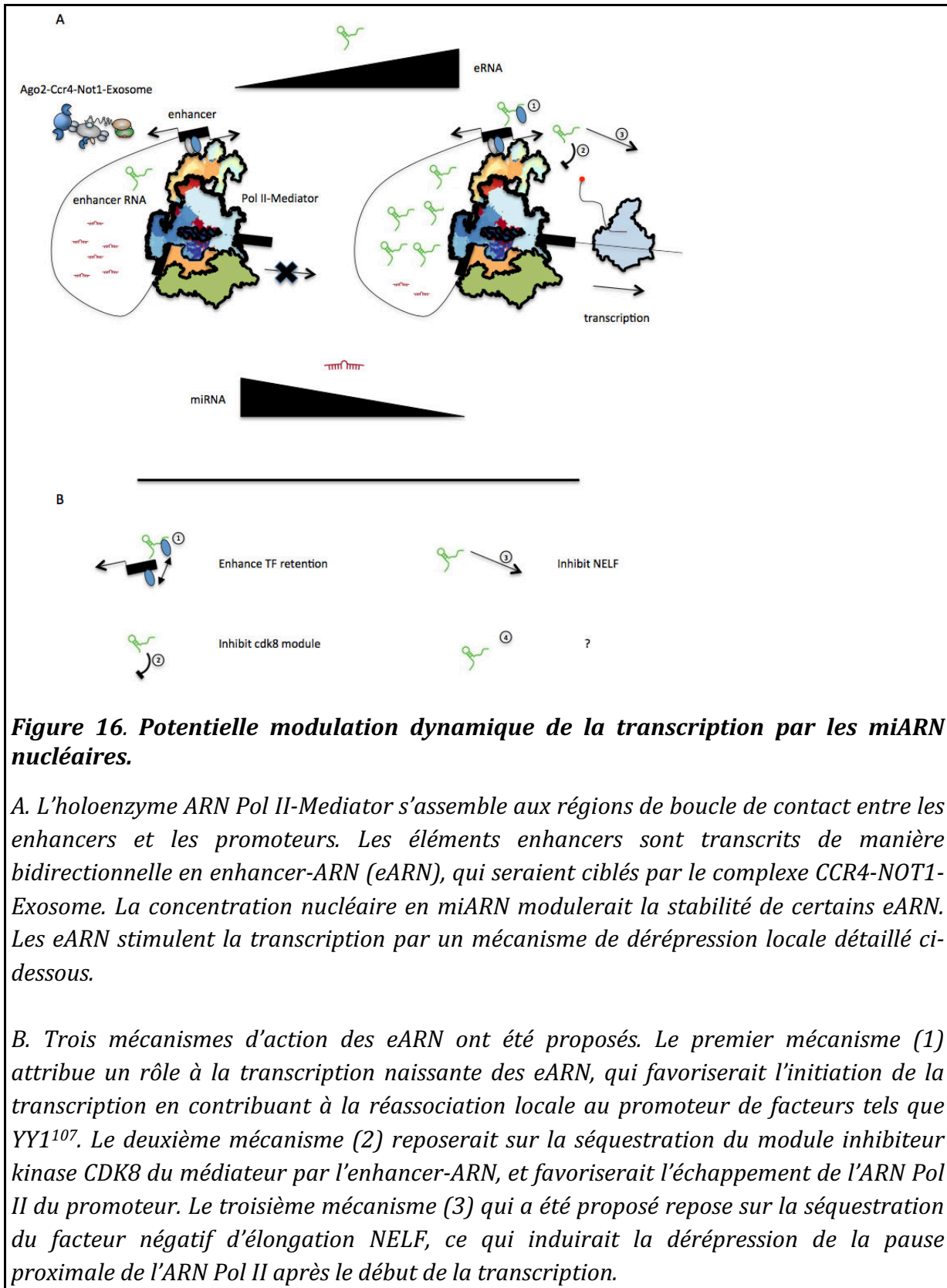
Les substrats de l'exosome incluent donc virtuellement tous les ARN de la cellule, y compris les transcrits cryptiques instables (CUTs), la transcription antisens en amont des promoteurs (PROMPTs), et les enhancer-ARN (eARN).

Les eARN sont centraux dans l'activation de la transcription chez les métazoaires<sup>472</sup>, et sont stabilisés quand l'activité de l'exosome est compromise<sup>473</sup>. Une relation linéaire a été observée entre l'expression des eARN et l'expression correspondante de l'ARNm des gènes cibles<sup>474</sup>. Dans le mécanisme d'augmentation de la transcription par les éléments de type enhancers, la contribution des transcrits eARN a été évaluée à 50-70% d'activation supplémentaire<sup>i</sup> sur la base d'expériences avec des plasmides et gènes rapporteurs, ce qui se traduit par une augmentation de 1.5-fold à 1.7-fold de l'expression génique en présence des transcrits eARN<sup>475,476</sup>. Cependant, les éléments de séquences fonctionnels des eARN restent inconnus à ce stade.

<sup>i</sup> Dans les cas où les transcrits eARN sont fonctionnels.







## 10 Conclusions

Au cours de cette thèse, nous nous sommes intéressés à l'activité nucléaire des miARN et aux facteurs moléculaires requis pour cette activité.

Par des approches de fractionnement cellulaire et des expériences standard, nous avons démontré que l'activité des micro-ARN est préservée dans le noyau, et peut être modulée par certaines protéines nucléaires de liaison à l'ARN comme SFPQ. Au niveau du mécanisme d'action, nous avons découvert que la protéine AGO2 interagit avec l'exosome nucléaire, d'une manière indépendante de l'ARN.

Nos résultats ouvrent un nouveau champ de recherche dans le domaine de la biologie des miARN, et suggèrent que ceux-ci s'intègrent dans une boucle supplémentaire de modulation de l'expression des gènes au niveau transcriptionnel dans le noyau.

Enfin, l'ensemble de ces résultats et observations vont dans le sens d'un mécanisme de recrutement et de ciblage de l'exosome qui repose sur le complexe RISC, dont nous proposons un modèle de travail basé sur le complexe CCR4-NOT1. Nous pensons que le recrutement de l'exosome par le complexe CCR4-NOT1 est direct via le module CNOT2-CNOT3.

## 11 Bibliographie

1. Watson, J. D. & Crick, F. H. C. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **171**, 737–738 (1953).
2. Jacob, F. & Monod, J. Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology* **3**, 318–356 (1961).
3. Makarova, K. S., Wolf, Y. I., van der Oost, J. & Koonin, E. V. Prokaryotic homologs of Argonaute proteins are predicted to function as key components of a novel system of defense against mobile genetic elements. *Biol. Direct* **4**, 29 (2009).
4. Swarts, D. C. *et al.* The evolutionary journey of Argonaute proteins. *Nat Struct Mol Biol* **21**, 743–753 (2014).
5. Crick, F. H. The origin of the genetic code. *J. Mol. Biol.* **38**, 367–379 (1968).
6. Horning, D. P. & Joyce, G. F. Amplification of RNA by an RNA polymerase ribozyme. *PNAS* **113**, 9786–9791 (2016).
7. Wu, X. & Sharp, P. A. Divergent transcription: a driving force for new gene origination? *Cell* **155**, 990–996 (2013).
8. Venter, J. C. *et al.* The sequence of the human genome. *Science* **291**, 1304–1351 (2001).
9. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
10. Biémont, C. & Vieira, C. Genetics: Junk DNA as an evolutionary force. *Nature* (2006). doi:10.1038/443521a
11. Eddy, S. R. Non-coding RNA genes and the modern RNA world. *Nature Reviews Genetics* **2**, 919–929 (2001).
12. Mattick, J. S. & Makunin, I. V. Non-coding RNA. *Hum Mol Genet* **15**, R17–R29 (2006).
13. Saccani, S. & Trabucchi, M. Regulation of stimulus-inducible gene expression in myeloid cells. *Seminars in Immunology* **27**, 33–43 (2015).
14. Chi, S. W., Hannon, G. J. & Darnell, R. B. An alternative mode of microRNA target recognition. *Nat Struct Mol Biol* **19**, 321–327 (2012).
15. Helwak, A., Kudla, G., Dudnakova, T. & Tollervey, D. Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. *Cell* **153**, 654–665 (2013).
16. Bottini, S. *et al.* From benchmarking HITS-CLIP peak detection programs to a new method for identification of miRNA-binding sites from Ago2-CLIP data. *Nucleic Acids Res.* **45**, e71 (2017).
17. Friedländer, M. R. *et al.* Evidence for the biogenesis of more than 1,000 novel human microRNAs. *Genome Biol.* **15**, R57 (2014).
18. Bartel, D. P. Metazoan MicroRNAs. *Cell* **173**, 20–51 (2018).
19. Ruggiero, T. *et al.* LPS induces KH-type splicing regulatory protein-dependent processing of microRNA-155 precursors in macrophages. *The FASEB Journal* **23**, 2898–2908 (2009).
20. Trabucchi, M. *et al.* How to control miRNA maturation? Co-activators and co-repressors take the stage. *RNA Biology* **6**, 536–540 (2009).
21. Hu, G., Drescher, K. M. & Chen, X.-M. Exosomal miRNAs: Biological Properties and Therapeutic Potential. *Front Genet* **3**, (2012).
22. Pasquinelli, A. E. MicroRNAs and their targets: recognition, regulation and an emerging reciprocal relationship. *Nat. Rev. Genet.* **13**, 271–282 (2012).
23. Repetto, E. *et al.* Let-7b/c enhance the stability of a tissue-specific mRNA during mammalian organogenesis as part of a feedback loop involving KSRP. *PLoS Genet.* **8**, e1002823 (2012).

24. Di Leva, G., Garofalo, M. & Croce, C. M. MicroRNAs in cancer. *Annu Rev Pathol* **9**, 287–314 (2014).
25. Repetto, E. *et al.* RNY-derived small RNAs as a signature of coronary artery disease. *BMC Medicine* **13**, 259 (2015).
26. Bruscella, P. *et al.* Viruses and miRNAs: More Friends than Foes. *Front Microbiol* **8**, (2017).
27. Hizir, Z., Bottini, S., Grandjean, V., Trabucchi, M. & Repetto, E. RNY (YRNA)-derived small RNAs regulate cell death and inflammation in monocytes/macrophages. *Cell Death Dis* **8**, e2530 (2017).
28. Mello, C. C. & Jr, D. C. Revealing the world of RNA interference. *Nature* (2004). doi:10.1038/nature02872
29. Czech, B. *et al.* An endogenous small interfering RNA pathway in *Drosophila*. *Nature* **453**, 798–802 (2008).
30. Ghildiyal, M. *et al.* Endogenous siRNAs Derived from Transposons and mRNAs in *Drosophila* Somatic Cells. *Science* **320**, 1077–1081 (2008).
31. Kawamura, Y. *et al.* *Drosophila* endogenous small RNAs bind to Argonaute 2 in somatic cells. *Nature* **453**, 793–797 (2008).
32. Okamura, K. *et al.* The *Drosophila* hairpin RNA pathway generates endogenous short interfering RNAs. *Nature* **453**, 803–806 (2008).
33. Bohmert, K. *et al.* AGO1 defines a novel locus of *Arabidopsis* controlling leaf development. *EMBO J.* **17**, 170–180 (1998).
34. Elmayan, T. *et al.* *Arabidopsis* mutants impaired in cosuppression. *Plant Cell* **10**, 1747–1758 (1998).
35. Napoli, C., Lemieux, C. & Jorgensen, R. Introduction of a Chimeric Chalcone Synthase Gene into *Petunia* Results in Reversible Co-Suppression of Homologous Genes in trans. *Plant Cell* **2**, 279–289 (1990).
36. Van Blokland, R., Van der Geest, N., Mol, J. n. m. & Kooter, J. m. Transgene-mediated suppression of chalcone synthase expression in *Petunia hybrida* results from an increase in RNA turnover. *The Plant Journal* **6**, 861–877 (1994).
37. Hamilton, A. J. & Baulcombe, D. C. A Species of Small Antisense RNA in Posttranscriptional Gene Silencing in Plants. *Science* **286**, 950–952 (1999).
38. Cogoni, C. & Macino, G. Gene silencing in *Neurospora crassa* requires a protein homologous to RNA-dependent RNA polymerase. *Nature* **399**, 166–169 (1999).
39. Timmons, L. & Fire, A. Specific interference by ingested dsRNA. *Nature* **395**, 854 (1998).
40. Tabara, H., Grishok, A. & Mello, C. C. RNAi in *C. elegans*: soaking in the genome sequence. *Science* **282**, 430–431 (1998).
41. Tabara, H. *et al.* The *rde-1* gene, RNA interference, and transposon silencing in *C. elegans*. *Cell* **99**, 123–132 (1999).
42. Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**, 806–811 (1998).
43. Smardon, A. *et al.* EGO-1 is related to RNA-directed RNA polymerase and functions in germ-line development and RNA interference in *C. elegans*. *Curr. Biol.* **10**, 169–178 (2000).
44. Sijen, T. *et al.* On the role of RNA amplification in dsRNA-triggered gene silencing. *Cell* **107**, 465–476 (2001).
45. Grishok, A., Tabara, H. & Mello, C. C. Genetic Requirements for Inheritance of RNAi in *C. elegans*. *Science* **287**, 2494–2497 (2000).
46. Clark, S. G., Lu, X. & Horvitz, H. R. The *Caenorhabditis Elegans* Locus *Lin-15*, a Negative Regulator of a Tyrosine Kinase Signaling Pathway, Encodes Two Different

Proteins. *Genetics* **137**, 987–997 (1994).

47. Montgomery, M. K., Xu, S. & Fire, A. RNA as a target of double-stranded RNA-mediated genetic interference in *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 15502–15507 (1998).
48. Zamore, P. D., Tuschl, T., Sharp, P. A. & Bartel, D. P. RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell* **101**, 25–33 (2000).
49. Elbashir, S. M., Lendeckel, W. & Tuschl, T. RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes Dev.* **15**, 188–200 (2001).
50. Martinez, J. & Tuschl, T. RISC is a 5' phosphomonoester-producing RNA endonuclease. *Genes Dev.* **18**, 975–980 (2004).
51. Barber, G. N. Host defense, viruses and apoptosis. *Cell Death Differ.* **8**, 113–126 (2001).
52. Czech, B. & Hannon, G. J. One Loop to Rule Them All: The Ping-Pong Cycle and piRNA-Guided Silencing. *Trends Biochem Sci* **41**, 324–337 (2016).
53. Lee, Y. S., Shibata, Y., Malhotra, A. & Dutta, A. A novel class of small RNAs: tRNA-derived RNA fragments (tRFs). *Genes Dev.* **23**, 2639–2649 (2009).
54. Sharma, U. *et al.* Biogenesis and function of tRNA fragments during sperm maturation and fertilization in mammals. *Science* **351**, 391–396 (2016).
55. Schorn, A. J., Gutbrod, M. J., LeBlanc, C. & Martienssen, R. LTR-Retrotransposon Control by tRNA-Derived Small RNAs. *Cell* **170**, 61–71.e11 (2017).
56. Goll, M. G. *et al.* Methylation of tRNA<sup>Asp</sup> by the DNA methyltransferase homolog Dnmt2. *Science* **311**, 395–398 (2006).
57. Tuorto, F. *et al.* RNA cytosine methylation by Dnmt2 and NSun2 promotes tRNA stability and protein synthesis. *Nat. Struct. Mol. Biol.* **19**, 900–905 (2012).
58. Kiani, J. *et al.* RNA-mediated epigenetic heredity requires the cytosine methyltransferase Dnmt2. *PLoS Genet.* **9**, e1003498 (2013).
59. Zhang, Y. *et al.* Dnmt2 mediates intergenerational transmission of paternally acquired metabolic disorders through sperm small non-coding RNAs. *Nat. Cell Biol.* **20**, 535–540 (2018).
60. Grandjean, V. *et al.* RNA-mediated paternal heredity of diet-induced obesity and metabolic disorders. *Scientific Reports* **5**, 18193 (2015).
61. Gapp, K. & Miska, E. A. tRNA fragments: novel players in intergenerational inheritance. *Cell Research* **26**, 395–396 (2016).
62. Kuscu, C. *et al.* tRNA fragments (tRFs) guide Ago to regulate gene expression post-transcriptionally in a Dicer independent manner. *RNA* (2018). doi:10.1261/rna.066126.118
63. Hendrick, J. P., Wolin, S. L., Rinke, J., Lerner, M. R. & Steitz, J. A. Ro small cytoplasmic ribonucleoproteins are a subclass of La ribonucleoproteins: further characterization of the Ro and La small ribonucleoproteins from uninfected mammalian cells. *Mol. Cell. Biol.* **1**, 1138–1149 (1981).
64. Kowalski, M. P. & Krude, T. Functional roles of non-coding Y RNAs. *The International Journal of Biochemistry & Cell Biology* **66**, 20–29 (2015).
65. Christov, C. P., Gardiner, T. J., Szüts, D. & Krude, T. Functional Requirement of Noncoding Y RNAs for Human Chromosomal DNA Replication. *Mol Cell Biol* **26**, 6993–7004 (2006).
66. Stein, A. J., Fuchs, G., Fu, C., Wolin, S. L. & Reinisch, K. M. Structural insights into RNA quality control: the Ro autoantigen binds misfolded RNAs via its central cavity. *Cell* **121**, 529–539 (2005).
67. Hung, T. *et al.* The Ro60 autoantigen binds endogenous retroelements and regulates inflammatory gene expression. *Science* **350**, 455–459 (2015).

68. Fuchs, G., Stein, A. J., Fu, C., Reinisch, K. M. & Wolin, S. L. Structural and biochemical basis for misfolded RNA recognition by the Ro autoantigen. *Nat. Struct. Mol. Biol.* **13**, 1002–1009 (2006).
69. Hogg, J. R. & Collins, K. Human Y5 RNA specializes a Ro ribonucleoprotein for 5S ribosomal RNA quality control. *Genes Dev* **21**, 3067–3072 (2007).
70. Collart, C., Christov, C. P., Smith, J. C. & Krude, T. The midblastula transition defines the onset of Y RNA-dependent DNA replication in *Xenopus laevis*. *Mol. Cell. Biol.* **31**, 3857–3870 (2011).
71. Krude, T. Initiation of Human DNA Replication in Vitro Using Nuclei from Cells Arrested at an Initiation-competent State. *J. Biol. Chem.* **275**, 13699–13707 (2000).
72. Nicolas, F. E., Hall, A. E., Csorba, T., Turnbull, C. & Dalmay, T. Biogenesis of Y RNA-derived small RNAs is independent of the microRNA pathway. *FEBS Lett.* **586**, 1226–1230 (2012).
73. Chen, C.-J. & Heard, E. Small RNAs derived from structural non-coding RNAs. *Methods* **63**, 76–84 (2013).
74. Rutjes, S. A., van der Heijden, A., Utz, P. J., van Venrooij, W. J. & Pruijn, G. J. Rapid nucleolytic degradation of the small cytoplasmic Y RNAs during apoptosis. *J. Biol. Chem.* **274**, 24799–24807 (1999).
75. Lee, R. C., Feinbaum, R. L. & Ambros, V. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* **75**, 843–854 (1993).
76. Wightman, B., Ha, I. & Ruvkun, G. Posttranscriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal pattern formation in *C. elegans*. *Cell* **75**, 855–862 (1993).
77. Lau, N. C., Lim, L. P., Weinstein, E. G. & Bartel, D. P. An Abundant Class of Tiny RNAs with Probable Regulatory Roles in *Caenorhabditis elegans*. *Science* **294**, 858–862 (2001).
78. Cullen, B. R. Transcription and Processing of Human microRNA Precursors. *Molecular Cell* **16**, 861–865 (2004).
79. de Rie, D. *et al.* An integrated expression atlas of miRNAs and their promoters in human and mouse. *Nat. Biotechnol.* **35**, 872–878 (2017).
80. Reinhart, B. J. *et al.* The 21-nucleotide *let-7* RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature* **403**, 901–906 (2000).
81. Pasquinelli, A. E. *et al.* Conservation of the sequence and temporal expression of *let-7* heterochronic regulatory RNA. *Nature* **408**, 86–89 (2000).
82. Roush, S. & Slack, F. J. The *let-7* family of microRNAs. *Trends in Cell Biology* **18**, 505–516 (2008).
83. Fromm, B. *et al.* A Uniform System for the Annotation of Vertebrate microRNA Genes and the Evolution of the Human microRNAome. *Annu. Rev. Genet.* **49**, 213–242 (2015).
84. Kozomara, A. & Griffiths-Jones, S. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res* **42**, D68–D73 (2014).
85. Chiang, H. R. *et al.* Mammalian microRNAs: experimental evaluation of novel and previously annotated genes. *Genes Dev.* **24**, 992–1009 (2010).
86. Tuschl, T., Zamore, P. D., Lehmann, R., Bartel, D. P. & Sharp, P. A. Targeted mRNA degradation by double-stranded RNA in vitro. *Genes Dev.* **13**, 3191–3197 (1999).
87. Lee, R. C. & Ambros, V. An extensive class of small RNAs in *Caenorhabditis elegans*. *Science* **294**, 862–864 (2001).
88. Krol, J., Loedige, I. & Filipowicz, W. The widespread regulation of microRNA biogenesis, function and decay. *Nat. Rev. Genet.* **11**, 597–610 (2010).
89. Lee, Y., Jeon, K., Lee, J.-T., Kim, S. & Kim, V. N. MicroRNA maturation: stepwise

processing and subcellular localization. *EMBO J.* **21**, 4663–4670 (2002).

90. Borchert, G. M., Lanier, W. & Davidson, B. L. RNA polymerase III transcribes human microRNAs. *Nat. Struct. Mol. Biol.* **13**, 1097–1101 (2006).

91. Lee, Y. *et al.* MicroRNA genes are transcribed by RNA polymerase II. *EMBO J* **23**, 4051–4060 (2004).

92. Lee, Y. *et al.* The nuclear RNase III Drosha initiates microRNA processing. *Nature* **425**, 415–419 (2003).

93. Gregory, R. I. *et al.* The Microprocessor complex mediates the genesis of microRNAs. *Nature* **432**, 235–240 (2004).

94. Fang, W. & Bartel, D. P. The Menu of Features that Define Primary MicroRNAs and Enable De Novo Design of MicroRNA Genes. *Molecular Cell* **60**, 131–145 (2015).

95. Han, J. *et al.* Molecular Basis for the Recognition of Primary microRNAs by the Drosha-DGCR8 Complex. *Cell* **125**, 887–901 (2006).

96. West, J. A. *et al.* Structural, super-resolution microscopy analysis of paraspeckle nuclear body organization. *J. Cell Biol.* **214**, 817–830 (2016).

97. Jiang, L. *et al.* NEAT1 scaffolds RNA-binding proteins and the Microprocessor to globally enhance pri-miRNA processing. *Nat. Struct. Mol. Biol.* **24**, 816–824 (2017).

98. Yi, R., Qin, Y., Macara, I. G. & Cullen, B. R. Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev.* **17**, 3011–3016 (2003).

99. Ruby, J. G., Jan, C. H. & Bartel, D. P. Intronic microRNA precursors that bypass Drosha processing. *Nature* **448**, 83–86 (2007).

100. Berezikov, E., Chung, W.-J., Willis, J., Cuppen, E. & Lai, E. C. Mammalian mirtron genes. *Mol. Cell* **28**, 328–336 (2007).

101. Ladewig, E., Okamura, K., Flynt, A. S., Westholm, J. O. & Lai, E. C. Discovery of hundreds of mirtrons in mouse and human small RNA data. *Genome Res.* **22**, 1634–1645 (2012).

102. Bernstein, E., Caudy, A. A., Hammond, S. M. & Hannon, G. J. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409**, 363–366 (2001).

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

104. Chendrimada, T. P. *et al.* TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. *Nature* **436**, 740–744 (2005).

105. Liu, Q. *et al.* R2D2, a bridge between the initiation and effector steps of the Drosophila RNAi pathway. *Science* **301**, 1921–1925 (2003).

106. Iwasaki, S. *et al.* Hsc70/Hsp90 chaperone machinery mediates ATP-dependent RISC loading of small RNA duplexes. *Mol. Cell* **39**, 292–299 (2010).

107. Landgraf, P. *et al.* A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell* **129**, 1401–1414 (2007).

108. Calabrese, J. M., Seila, A. C., Yeo, G. W. & Sharp, P. A. RNA sequence analysis defines Dicer's role in mouse embryonic stem cells. *PNAS* **104**, 18097–18102 (2007).

109. Suzuki, H. I., Young, R. A. & Sharp, P. A. Super-Enhancer-Mediated RNA Processing Revealed by Integrative MicroRNA Network Analysis. *Cell* **168**, 1000-1014.e15 (2017).

110. O'Donnell, K. A., Wentzel, E. A., Zeller, K. I., Dang, C. V. & Mendell, J. T. c-Myc-regulated microRNAs modulate E2F1 expression. *Nature* **435**, 839–843 (2005).

111. Han, L., Witmer, P. D., Casey, E., Valle, D. & Sukumar, S. DNA methylation regulates MicroRNA expression. *Cancer Biol. Ther.* **6**, 1284–1288 (2007).

112. Hnisz, D. *et al.* Super-Enhancers in the Control of Cell Identity and Disease. *Cell* **155**, 934–947 (2013).

113. Ma, L. *et al.* miR-9, a MYC/MYCN-activated microRNA, regulates E-cadherin and cancer metastasis. *Nat Cell Biol* **12**, 247–256 (2010).
114. He, L. *et al.* A microRNA component of the p53 tumour suppressor network. *Nature* **447**, 1130–1134 (2007).
115. Conaco, C., Otto, S., Han, J.-J. & Mandel, G. Reciprocal actions of REST and a microRNA promote neuronal identity. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 2422–2427 (2006).
116. Kim, J. *et al.* A MicroRNA feedback circuit in midbrain dopamine neurons. *Science* **317**, 1220–1224 (2007).
117. Huang, X. *et al.* Hypoxia-inducible mir-210 regulates normoxic gene expression involved in tumor initiation. *Mol. Cell* **35**, 856–867 (2009).
118. Tessel, M. A., Krett, N. L. & Rosen, S. T. Steroid receptor and microRNA regulation in cancer. *Curr Opin Oncol* **22**, 592–597 (2010).
119. Ben-Ami, O., Pencovich, N., Lotem, J., Levanon, D. & Groner, Y. A regulatory interplay between miR-27a and Runx1 during megakaryopoiesis. *PNAS* **106**, 238–243 (2009).
120. Zhao, H., Kalota, A., Jin, S. & Gewirtz, A. M. The c-myc proto-oncogene and microRNA-15a comprise an active autoregulatory feedback loop in human hematopoietic cells. *Blood* **113**, 505–516 (2009).
121. Roush, S. F. & Slack, F. J. Transcription of the *C. elegans* let-7 microRNA is temporally regulated by one of its targets, hbl-1. *Dev Biol* **334**, 523–534 (2009).
122. Li, X. & Carthew, R. W. A microRNA Mediates EGF Receptor Signaling and Promotes Photoreceptor Differentiation in the *Drosophila* Eye. *Cell* **123**, 1267–1277 (2005).
123. Viswanathan, S. R., Daley, G. Q. & Gregory, R. I. Selective blockade of microRNA processing by Lin-28. *Science* **320**, 97–100 (2008).
124. Trabucchi, M. *et al.* The RNA-binding protein KSRP promotes the biogenesis of a subset of microRNAs. *Nature* **459**, 1010–1014 (2009).
125. Chang, H.-M., Triboulet, R., Thornton, J. E. & Gregory, R. I. A role for the Perlman syndrome exonuclease Dis3l2 in the Lin28–let-7 pathway. *Nature* **497**, 244–248 (2013).
126. Fukuda, T. *et al.* DEAD-box RNA helicase subunits of the Drosha complex are required for processing of rRNA and a subset of microRNAs. *Nat. Cell Biol.* **9**, 604–611 (2007).
127. Suzuki, H. I. *et al.* Modulation of microRNA processing by p53. *Nature* **460**, 529–533 (2009).
128. Heo, I. *et al.* Lin28 mediates the terminal uridylation of let-7 precursor MicroRNA. *Mol. Cell* **32**, 276–284 (2008).
129. Davis, B. N., Hilyard, A. C., Lagna, G. & Hata, A. SMAD proteins control DROSHA-mediated microRNA maturation. *Nature* **454**, 56–61 (2008).
130. Guil, S. & Cáceres, J. F. The multifunctional RNA-binding protein hnRNP A1 is required for processing of miR-18a. *Nat. Struct. Mol. Biol.* **14**, 591–596 (2007).
131. Sakamoto, S. *et al.* The NF90-NF45 complex functions as a negative regulator in the microRNA processing pathway. *Mol. Cell. Biol.* **29**, 3754–3769 (2009).
132. Olsen, P. H. & Ambros, V. The lin-4 Regulatory RNA Controls Developmental Timing in *Caenorhabditis elegans* by Blocking LIN-14 Protein Synthesis after the Initiation of Translation. *Developmental Biology* **216**, 671–680 (1999).
133. Zeng, Y., Wagner, E. J. & Cullen, B. R. Both Natural and Designed Micro RNAs Can Inhibit the Expression of Cognate mRNAs When Expressed in Human Cells. *Molecular Cell* **9**, 1327–1333 (2002).
134. PILLAI, R. S., ARTUS, C. G. & FILIPOWICZ, W. Tethering of human Ago proteins to mRNA mimics the miRNA-mediated repression of protein synthesis. *RNA* **10**, 1518–1525 (2004).
135. Pillai, R. S. *et al.* Inhibition of Translational Initiation by Let-7 MicroRNA in Human



- Cells. *Science* **309**, 1573–1576 (2005).
136. Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. S. & Weissman, J. S. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* **324**, 218–223 (2009).
137. Guo, H., Ingolia, N. T., Weissman, J. S. & Bartel, D. P. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* **466**, 835–840 (2010).
138. Bagga, S. *et al.* Regulation by let-7 and lin-4 miRNAs Results in Target mRNA Degradation. *Cell* **122**, 553–563 (2005).
139. Zeng, Y., Yi, R. & Cullen, B. R. MicroRNAs and small interfering RNAs can inhibit mRNA expression by similar mechanisms. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 9779–9784 (2003).
140. Wu, L. & Belasco, J. G. Micro-RNA regulation of the mammalian lin-28 gene during neuronal differentiation of embryonal carcinoma cells. *Mol. Cell. Biol.* **25**, 9198–9208 (2005).
141. Wu, L., Fan, J. & Belasco, J. G. MicroRNAs direct rapid deadenylation of mRNA. *PNAS* **103**, 4034–4039 (2006).
142. Mathonnet, G. *et al.* MicroRNA Inhibition of Translation Initiation in Vitro by Targeting the Cap-Binding Complex eIF4F. *Science* **317**, 1764–1767 (2007).
143. Fabian, M. R. *et al.* Mammalian miRNA RISC Recruits CAF1 and PABP to Affect PABP-Dependent Deadenylation. *Molecular Cell* **35**, 868–880 (2009).
144. Jackson, A. L. *et al.* Expression profiling reveals off-target gene regulation by RNAi. *Nature Biotechnology* **21**, 635–637 (2003).
145. Lim, L. P. *et al.* Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* **433**, 769–773 (2005).
146. Krützfeldt, J. *et al.* Silencing of microRNAs in vivo with ‘antagomirs’. *Nature* **438**, 685–689 (2005).
147. Meister, G. *et al.* Identification of novel argonaute-associated proteins. *Curr. Biol.* **15**, 2149–2155 (2005).
148. Cikaluk, D. E. *et al.* GERp95, a membrane-associated protein that belongs to a family of proteins involved in stem cell differentiation. *Mol. Biol. Cell* **10**, 3357–3372 (1999).
149. Drinnenberg, I. A. *et al.* RNAi in budding yeast. *Science* **326**, 544–550 (2009).
150. Chen, G. R., Sive, H. & Bartel, D. P. A Seed Mismatch Enhances Argonaute2-Catalyzed Cleavage and Partially Rescues Severely Impaired Cleavage Found in Fish. *Molecular Cell* **68**, 1095–1107.e5 (2017).
151. Kennerdell, J. R. & Carthew, R. W. Use of dsRNA-mediated genetic interference to demonstrate that frizzled and frizzled 2 act in the wingless pathway. *Cell* **95**, 1017–1026 (1998).
152. Hammond, S. M., Bernstein, E., Beach, D. & Hannon, G. J. An RNA-directed nuclease mediates post-transcriptional gene silencing in *Drosophila* cells. *Nature* **404**, 293–296 (2000).
153. Mourelatos, Z. *et al.* miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs. *Genes Dev.* **16**, 720–728 (2002).
154. Lin, H. & Spradling, A. C. A novel group of pumilio mutations affects the asymmetric division of germline stem cells in the *Drosophila* ovary. *Development* **124**, 2463–2476 (1997).
155. Song, J.-J., Smith, S. K., Hannon, G. J. & Joshua-Tor, L. Crystal Structure of Argonaute and Its Implications for RISC Slicer Activity. *Science* **305**, 1434–1437 (2004).
156. Meister, G. *et al.* Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Mol. Cell* **15**, 185–197 (2004).
157. Liu, J. *et al.* Argonaute2 Is the Catalytic Engine of Mammalian RNAi. *Science* **305**, 1437–1441 (2004).
158. Ma, J.-B. *et al.* Structural basis for 5'-end-specific recognition of guide RNA by the A.

- fulgidus Piwi protein. *Nature* **434**, 666–670 (2005).
159. Schirle, N. T., Sheu-Gruttadauria, J. & MacRae, I. J. Structural basis for microRNA targeting. *Science* **346**, 608–613 (2014).
160. Frank, F., Sonenberg, N. & Nagar, B. Structural basis for 5'-nucleotide base-specific recognition of guide RNA by human AGO2. *Nature* **465**, 818–822 (2010).
161. Nykänen, A., Haley, B. & Zamore, P. D. ATP requirements and small interfering RNA structure in the RNA interference pathway. *Cell* **107**, 309–321 (2001).
162. Martinez, J., Patkaniowska, A., Urlaub, H., Lührmann, R. & Tuschl, T. Single-Stranded Antisense siRNAs Guide Target RNA Cleavage in RNAi. *Cell* **110**, 563–574 (2002).
163. Filipowicz, W., Bhattacharyya, S. N. & Sonenberg, N. Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nat. Rev. Genet.* **9**, 102–114 (2008).
164. Behm-Ansmant, I. *et al.* mRNA degradation by miRNAs and GW182 requires both CCR4:NOT deadenylase and DCP1:DCP2 decapping complexes. *Genes Dev.* **20**, 1885–1898 (2006).
165. Till, S. *et al.* A conserved motif in Argonaute-interacting proteins mediates functional interactions through the Argonaute PIWI domain. *Nature Structural & Molecular Biology* **14**, 897–903 (2007).
166. Mathys, H. *et al.* Structural and biochemical insights to the role of the CCR4-NOT complex and DDX6 ATPase in microRNA repression. *Mol. Cell* **54**, 751–765 (2014).
167. Chen, Y. *et al.* A DDX6-CNOT1 Complex and W-Binding Pockets in CNOT9 Reveal Direct Links between miRNA Target Recognition and Silencing. *Molecular Cell* **54**, 737–750 (2014).
168. Sgromo, A. *et al.* Drosophila Bag-of-marbles directly interacts with the CAF40 subunit of the CCR4-NOT complex to elicit repression of mRNA targets. *RNA* (2017). doi:10.1261/rna.064584.117
169. Eulalio, A., Helms, S., Fritsch, C., Fauser, M. & Izaurralde, E. A C-terminal silencing domain in GW182 is essential for miRNA function. *RNA* **15**, 1067–1077 (2009).
170. Lazzaretti, D., Tournier, I. & Izaurralde, E. The C-terminal domains of human TNRC6A, TNRC6B, and TNRC6C silence bound transcripts independently of Argonaute proteins. *RNA* **15**, 1059–1066 (2009).
171. Zekri, L., Huntzinger, E., Heimstädt, S. & Izaurralde, E. The silencing domain of GW182 interacts with PABPC1 to promote translational repression and degradation of microRNA targets and is required for target release. *Mol. Cell. Biol.* **29**, 6220–6231 (2009).
172. Jinek, M., Fabian, M. R., Coyle, S. M., Sonenberg, N. & Doudna, J. A. Structural insights into the human GW182-PABC interaction in microRNA-mediated deadenylation. *Nat. Struct. Mol. Biol.* **17**, 238–240 (2010).
173. Huntzinger, E., Braun, J. E., Heimstädt, S., Zekri, L. & Izaurralde, E. Two PABPC1-binding sites in GW182 proteins promote miRNA-mediated gene silencing. *EMBO J.* **29**, 4146–4160 (2010).
174. Fabian, M. R. *et al.* miRNA-mediated deadenylation is orchestrated by GW182 through two conserved motifs that interact with CCR4–NOT. *Nat Struct Mol Biol* **18**, 1211–1217 (2011).
175. Chekulaeva, M. *et al.* miRNA repression involves GW182-mediated recruitment of CCR4–NOT through conserved W-containing motifs. *Nat Struct Mol Biol* **18**, 1218–1226 (2011).
176. Braun, J. E., Huntzinger, E., Fauser, M. & Izaurralde, E. GW182 Proteins Directly Recruit Cytoplasmic Deadenylase Complexes to miRNA Targets. *Molecular Cell* **44**, 120–133 (2011).

177. Chu, C. & Rana, T. M. Translation Repression in Human Cells by MicroRNA-Induced Gene Silencing Requires RCK/p54. *PLOS Biology* **4**, e210 (2006).
178. Collier, J. M., Tucker, M., Sheth, U., Valencia-Sanchez, M. A. & Parker, R. The DEAD box helicase, Dhh1p, functions in mRNA decapping and interacts with both the decapping and deadenylase complexes. *RNA* **7**, 1717–1727 (2001).
179. Hata, H. *et al.* Dhh1p, a putative RNA helicase, associates with the general transcription factors Pop2p and Ccr4p from *Saccharomyces cerevisiae*. *Genetics* **148**, 571–579 (1998).
180. Uetz, P. *et al.* A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* **403**, 623–627 (2000).
181. Collier, J. & Parker, R. General Translational Repression by Activators of mRNA Decapping. *Cell* **122**, 875–886 (2005).
182. Eulalio, A. *et al.* Target-specific requirements for enhancers of decapping in miRNA-mediated gene silencing. *Genes Dev.* **21**, 2558–2570 (2007).
183. Sharif, H. *et al.* Structural analysis of the yeast Dhh1–Pat1 complex reveals how Dhh1 engages Pat1, Edc3 and RNA in mutually exclusive interactions. *Nucleic Acids Res* **41**, 8377–8390 (2013).
184. Ozgur, S. *et al.* Structure of a Human 4E-T/DDX6/CNOT1 Complex Reveals the Different Interplay of DDX6-Binding Proteins with the CCR4-NOT Complex. *Cell Reports* **13**, 703–711 (2015).
185. Schraivogel, D. *et al.* Importin- $\beta$  facilitates nuclear import of human GW proteins and balances cytoplasmic gene silencing protein levels. *Nucleic Acids Res* **43**, 7447–7461 (2015).
186. Cougot, N., Babajko, S. & Séraphin, B. Cytoplasmic foci are sites of mRNA decay in human cells. *J Cell Biol* **165**, 31–40 (2004).
187. Youn, J.-Y. *et al.* High-Density Proximity Mapping Reveals the Subcellular Organization of mRNA-Associated Granules and Bodies. *Molecular Cell* **69**, 517–532.e11 (2018).
188. Eulalio, A., Behm-Ansmant, I., Schweizer, D. & Izaurralde, E. P-body formation is a consequence, not the cause, of RNA-mediated gene silencing. *Mol. Cell. Biol.* **27**, 3970–3981 (2007).
189. Hubstenberger, A. *et al.* P-Body Purification Reveals the Condensation of Repressed mRNA Regulons. *Molecular Cell* **68**, 144–157.e5 (2017).
190. Bhattacharyya, S. N., Habermacher, R., Martine, U., Closs, E. I. & Filipowicz, W. Relief of microRNA-Mediated Translational Repression in Human Cells Subjected to Stress. *Cell* **125**, 1111–1124 (2006).
191. Kedde, M. *et al.* RNA-Binding Protein Dnd1 Inhibits MicroRNA Access to Target mRNA. *Cell* **131**, 1273–1286 (2007).
192. Edbauer, D. *et al.* Regulation of synaptic structure and function by FMRP-associated microRNAs miR-125b and miR-132. *Neuron* **65**, 373–384 (2010).
193. Kim, H. H. *et al.* HuR recruits let-7/RISC to repress c-Myc expression. *Genes Dev* **23**, 1743–1748 (2009).
194. Kedde, M. *et al.* A Pumilio-induced RNA structure switch in p27-3' UTR controls miR-221 and miR-222 accessibility. *Nature Cell Biology* **12**, 1014–1020 (2010).
195. Van Etten, J. *et al.* Human Pumilio Proteins Recruit Multiple Deadenylases to Efficiently Repress Messenger RNAs. *J Biol Chem* **287**, 36370–36383 (2012).
196. Xue, Y. *et al.* Direct Conversion of Fibroblasts to Neurons by Reprogramming PTB-Regulated microRNA Circuits. *Cell* **152**, 82–96 (2013).
197. Dalmay, T., Hamilton, A., Rudd, S., Angell, S. & Baulcombe, D. C. An RNA-Dependent RNA Polymerase Gene in Arabidopsis Is Required for Posttranscriptional Gene Silencing Mediated by a Transgene but Not by a Virus. *Cell* **101**, 543–553 (2000).

198. Tomari, Y. *et al.* RISC Assembly Defects in the Drosophila RNAi Mutant armitage. *Cell* **116**, 831–841 (2004).
199. Cook, H. A., Koppetsch, B. S., Wu, J. & Theurkauf, W. E. The Drosophila SDE3 Homolog armitage Is Required for oskar mRNA Silencing and Embryonic Axis Specification. *Cell* **116**, 817–829 (2004).
200. Banerjee, S., Neveu, P. & Kosik, K. S. A coordinated local translational control point at the synapse involving relief from silencing and MOV10 degradation. *Neuron* **64**, 871–884 (2009).
201. Svitkin, Y. V. *et al.* Control of Translation and miRNA-Dependent Repression by a Novel Poly(A) Binding Protein, hnRNP-Q. *PLOS Biology* **11**, e1001564 (2013).
202. Srikantan, S. *et al.* Translational Control of TOP2A Influences Doxorubicin Efficacy<sup>▽</sup>. *Mol Cell Biol* **31**, 3790–3801 (2011).
203. Tominaga, K. *et al.* Competitive regulation of nucleolin expression by HuR and miR-494. *Mol. Cell. Biol.* **31**, 4219–4231 (2011).
204. Young, L. E., Moore, A. E., Sokol, L., Meisner-Kober, N. & Dixon, D. A. The mRNA stability factor HuR inhibits microRNA-16 targeting of COX-2. *Mol. Cancer Res.* **10**, 167–180 (2012).
205. Epis, M. R., Barker, A., Giles, K. M., Beveridge, D. J. & Leedman, P. J. The RNA-binding protein HuR opposes the repression of ERBB-2 gene expression by microRNA miR-331-3p in prostate cancer cells. *J. Biol. Chem.* **286**, 41442–41454 (2011).
206. Kim, Y. *et al.* Posttranscriptional Regulation of the Inflammatory Marker C-Reactive Protein by the RNA-Binding Protein HuR and MicroRNA 637. *Mol Cell Biol* **35**, 4212–4221 (2015).
207. Glorian, V. *et al.* HuR-dependent loading of miRNA RISC to the mRNA encoding the Ras-related small GTPase RhoB controls its translation during UV-induced apoptosis. *Cell Death Differ.* **18**, 1692–1701 (2011).
208. Kedde, M. *et al.* RNA-binding protein Dnd1 inhibits microRNA access to target mRNA. *Cell* **131**, 1273–1286 (2007).
209. Miles, W. O., Tschöp, K., Herr, A., Ji, J.-Y. & Dyson, N. J. Pumilio facilitates miRNA regulation of the E2F3 oncogene. *Genes Dev* **26**, 356–368 (2012).
210. Makeyev, E. V., Zhang, J., Carrasco, M. A. & Maniatis, T. The MicroRNA miR-124 Promotes Neuronal Differentiation by Triggering Brain-Specific Alternative Pre-mRNA Splicing. *Mol Cell* **27**, 435–448 (2007).
211. Elcheva, I., Goswami, S., Noubissi, F. K. & Spiegelman, V. S. CRD-BP protects the coding region of betaTrCP1 mRNA from miR-183-mediated degradation. *Mol. Cell* **35**, 240–246 (2009).
212. Goswami, S. *et al.* MicroRNA-340-mediated degradation of microphthalmia-associated transcription factor (MITF) mRNA is inhibited by coding region determinant-binding protein (CRD-BP). *J. Biol. Chem.* **290**, 384–395 (2015).
213. Léveillé, N. *et al.* Selective inhibition of microRNA accessibility by RBM38 is required for p53 activity. *Nature Communications* **2**, 513 (2011).
214. Srivastava, M. *et al.* Roquin binds microRNA-146a and Argonaute2 to regulate microRNA homeostasis. *Nature Communications* **6**, 6253 (2015).
215. Doench, J. G. & Sharp, P. A. Specificity of microRNA target selection in translational repression. *Genes Dev* **18**, 504–511 (2004).
216. Lai, E. C. Micro RNAs are complementary to 3' UTR sequence motifs that mediate negative post-transcriptional regulation. *Nature Genetics* **30**, 363–364 (2002).
217. Chandradoss, S. D., Schirle, N. T., Szczepaniak, M., MacRae, I. J. & Joo, C. A Dynamic Search Process Underlies MicroRNA Targeting. *Cell* **162**, 96–107 (2015).
218. Giraldez, A. J. *et al.* Zebrafish MiR-430 Promotes Deadenylation and Clearance of

Maternal mRNAs. *Science* **312**, 75–79 (2006).

219. Lewis, B. P., Shih, I. -hun., Jones-Rhoades, M. W., Bartel, D. P. & Burge, C. B. Prediction of mammalian microRNA targets. *Cell* **115**, 787–798 (2003).
220. Agarwal, V., Bell, G. W., Nam, J.-W. & Bartel, D. P. Predicting effective microRNA target sites in mammalian mRNAs. *Elife* **4**, (2015).
221. Chi, S. W., Zang, J. B., Mele, A. & Darnell, R. B. Ago HITS-CLIP decodes miRNA-mRNA interaction maps. *Nature* **460**, 479–486 (2009).
222. Lewis, B. P., Burge, C. B. & Bartel, D. P. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**, 15–20 (2005).
223. Krek, A. *et al.* Combinatorial microRNA target predictions. *Nat. Genet.* **37**, 495–500 (2005).
224. Krek, A. *et al.* Combinatorial microRNA target predictions. *Nature Genetics* **37**, 495–500 (2005).
225. Ha, I., Wightman, B. & Ruvkun, G. A bulged lin-4/lin-14 RNA duplex is sufficient for *Caenorhabditis elegans* lin-14 temporal gradient formation. *Genes Dev.* **10**, 3041–3050 (1996).
226. Stark, A., Brennecke, J., Bushati, N., Russell, R. B. & Cohen, S. M. Animal MicroRNAs Confer Robustness to Gene Expression and Have a Significant Impact on 3'UTR Evolution. *Cell* **123**, 1133–1146 (2005).
227. Friedman, R. C., Farh, K. K.-H., Burge, C. B. & Bartel, D. P. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res* **19**, 92–105 (2009).
228. Grimson, A. *et al.* MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol. Cell* **27**, 91–105 (2007).
229. Loeb, G. B. *et al.* Transcriptome-wide miR-155 binding map reveals widespread non-canonical microRNA targeting. *Mol Cell* **48**, 760–770 (2012).
230. Shin, C. *et al.* Expanding the MicroRNA Targeting Code: Functional Sites with Centered Pairing. *Mol Cell* **38**, 789–802 (2010).
231. Hafner, M. *et al.* Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141**, 129–141 (2010).
232. Gu, S., Jin, L., Zhang, F., Sarnow, P. & Kay, M. A. Biological basis for restriction of microRNA targets to the 3' untranslated region in mammalian mRNAs. *Nat. Struct. Mol. Biol.* **16**, 144–150 (2009).
233. Baek, D. *et al.* The impact of microRNAs on protein output. *Nature* **455**, 64–71 (2008).
234. Ule, J. *et al.* CLIP identifies Nova-regulated RNA networks in the brain. *Science* **302**, 1212–1215 (2003).
235. Hansen, T. B. *et al.* Natural RNA circles function as efficient microRNA sponges. *Nature* **495**, 384–388 (2013).
236. Yekta, S., Shih, I.-H. & Bartel, D. P. MicroRNA-directed cleavage of HOXB8 mRNA. *Science* **304**, 594–596 (2004).
237. Reinhart, B. J., Weinstein, E. G., Rhoades, M. W., Bartel, B. & Bartel, D. P. MicroRNAs in plants. *Genes Dev* **16**, 1616–1626 (2002).
238. Rhoades, M. W. *et al.* Prediction of Plant MicroRNA Targets. *Cell* **110**, 513–520 (2002).
239. Didiano, D. & Hobert, O. Perfect seed pairing is not a generally reliable predictor for miRNA-target interactions. *Nature Structural & Molecular Biology* **13**, 849–851 (2006).
240. Garcia, D. M. *et al.* Weak Seed-Pairing Stability and High Target-Site Abundance Decrease the Proficiency of *lscy-6* and Other miRNAs. *Nat Struct Mol Biol* **18**, 1139–1146 (2011).

241. Franco-Zorrilla, J. M. *et al.* Target mimicry provides a new mechanism for regulation of microRNA activity. *Nature Genetics* **39**, 1033–1037 (2007).
242. Kawahara, Y. *et al.* Frequency and fate of microRNA editing in human brain. *Nucleic Acids Res.* **36**, 5270–5280 (2008).
243. Ketting, R. F. *et al.* Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in *C. elegans*. *Genes Dev.* **15**, 2654–2659 (2001).
244. Knight, S. W. & Bass, B. L. A role for the RNase III enzyme DCR-1 in RNA interference and germ line development in *Caenorhabditis elegans*. *Science* **293**, 2269–2271 (2001).
245. Bernstein, E. *et al.* Dicer is essential for mouse development. *Nature Genetics* **35**, 215–217 (2003).
246. Cheloufi, S., Dos Santos, C. O., Chong, M. M. W. & Hannon, G. J. A Dicer-independent miRNA biogenesis pathway that requires Ago catalysis. *Nature* **465**, 584–589 (2010).
247. Zhao, Y. *et al.* Dysregulation of cardiogenesis, cardiac conduction, and cell cycle in mice lacking miRNA-1-2. *Cell* **129**, 303–317 (2007).
248. Heidersbach, A. *et al.* microRNA-1 regulates sarcomere formation and suppresses smooth muscle gene expression in the mammalian heart. *eLife* **2**, (2013).
249. Shibata, M., Nakao, H., Kiyonari, H., Abe, T. & Aizawa, S. MicroRNA-9 Regulates Neurogenesis in Mouse Telencephalon by Targeting Multiple Transcription Factors. *J. Neurosci.* **31**, 3407–3422 (2011).
250. Ventura, A. *et al.* Targeted deletion reveals essential and overlapping functions of the miR-17 through 92 family of miRNA clusters. *Cell* **132**, 875–886 (2008).
251. Han, Y.-C. *et al.* An allelic series of miR-17 ~ 92-mutant mice uncovers functional specialization and cooperation among members of a microRNA polycistron. *Nat. Genet.* **47**, 766–775 (2015).
252. Wang, H. *et al.* miR-219 Cooperates with miR-338 in Myelination and Promotes Myelin Repair in the CNS. *Developmental Cell* **40**, 566–582.e5 (2017).
253. Crowley, J. J. *et al.* Disruption of the MicroRNA 137 Primary Transcript Results in Early Embryonic Lethality in Mice. *Biological Psychiatry* **77**, e5–e7 (2015).
254. Henao-Mejia, J. *et al.* The MicroRNA miR-181 Is a Critical Cellular Metabolic Rheostat Essential for NKT Cell Ontogenesis and Lymphocyte Development and Homeostasis. *Immunity* **38**, 984–997 (2013).
255. Tsujimura, K. *et al.* miR-199a Links MeCP2 with mTOR Signaling and Its Dysregulation Leads to Rett Syndrome Phenotypes. *Cell Rep* **12**, 1887–1901 (2015).
256. Azevedo-Pouly, A. C. P. *et al.* miR-216 and miR-217 expression is reduced in transgenic mouse models of pancreatic adenocarcinoma, knockout of miR-216/miR-217 host gene is embryonic lethal. *Funct. Integr. Genomics* **17**, 203–212 (2017).
257. Amin, N. D. *et al.* Loss of motoneuron-specific microRNA-218 causes systemic neuromuscular failure. *Science* **350**, 1525–1529 (2015).
258. Tan, C. L. *et al.* MicroRNA-128 Governs Neuronal Excitability and Motor Behavior in Mice. *Science* **342**, 1254–1258 (2013).
259. Parchem, R. J. *et al.* miR-302 Is Required for Timing of Neural Differentiation, Neural Tube Closure, and Embryonic Viability. *Cell Reports* **12**, 760–773 (2015).
260. Sassi, Y. *et al.* Cardiac myocyte miR-29 promotes pathological remodeling of the heart by activating Wnt signaling. *Nature Communications* **8**, 1614 (2017).
261. Song, R. *et al.* miR-34/449 miRNAs are required for motile ciliogenesis by repressing *cp110*. *Nature* **510**, 115–120 (2014).
262. Sanuki, R. *et al.* miR-124a is required for hippocampal axogenesis and retinal cone survival through Lhx2 suppression. *Nature Neuroscience* **14**, 1125–1134 (2011).

263. Liu, N. *et al.* microRNA-133a regulates cardiomyocyte proliferation and suppresses smooth muscle gene expression in the heart. *Genes Dev.* **22**, 3242–3254 (2008).
264. Kuhnert, F. *et al.* Attribution of vascular phenotypes of the murine *Egfl7* locus to the microRNA miR-126. *Development* **135**, 3989–3993 (2008).
265. Wang, D. *et al.* MicroRNA-205 controls neonatal expansion of skin stem cells by modulating the PI(3)K pathway. *Nature Cell Biology* **15**, 1153–1163 (2013).
266. Medeiros, L. A. *et al.* Mir-290–295 deficiency in mice results in partially penetrant embryonic lethality and germ cell defects. *PNAS* **108**, 14163–14168 (2011).
267. Mencía, Á. *et al.* Mutations in the seed region of human miR-96 are responsible for nonsyndromic progressive hearing loss. *Nature Genetics* **41**, 609–613 (2009).
268. de Pontual, L. *et al.* Germline deletion of the miR-17-92 cluster causes growth and skeletal defects in humans. *Nat Genet* **43**, 1026–1030 (2011).
269. Saunders, M. A., Liang, H. & Li, W.-H. Human polymorphism at microRNAs and microRNA target sites. *Proc Natl Acad Sci U S A* **104**, 3300–3305 (2007).
270. Lin, S. & Gregory, R. I. MicroRNA biogenesis pathways in cancer. *Nat. Rev. Cancer* **15**, 321–333 (2015).
271. Calin, G. A. *et al.* Frequent deletions and down-regulation of micro-RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 15524–15529 (2002).
272. Calin, G. A. *et al.* Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. *Proc Natl Acad Sci U S A* **101**, 2999–3004 (2004).
273. Johnson, S. M. *et al.* RAS is regulated by the let-7 microRNA family. *Cell* **120**, 635–647 (2005).
274. Nigro, J. M. *et al.* Mutations in the p53 gene occur in diverse human tumour types. *Nature* **342**, 705–708 (1989).
275. Hollstein, M., Sidransky, D., Vogelstein, B. & Harris, C. C. p53 mutations in human cancers. *Science* **253**, 49–53 (1991).
276. He, L. *et al.* A microRNA polycistron as a potential human oncogene. *Nature* **435**, 828–833 (2005).
277. Lu, J. *et al.* MicroRNA expression profiles classify human cancers. *Nature* **435**, 834–838 (2005).
278. Kumar, M. S., Lu, J., Mercer, K. L., Golub, T. R. & Jacks, T. Impaired microRNA processing enhances cellular transformation and tumorigenesis. *Nat. Genet.* **39**, 673–677 (2007).
279. Worringer, K. A. *et al.* The let-7/LIN-41 Pathway Regulates Reprogramming to Human Induced Pluripotent Stem Cells by Controlling Expression of Pro-differentiation Genes. *Cell Stem Cell* **14**, 40–52 (2014).
280. Merritt, W. M. *et al.* Dicer, Drosha, and Outcomes in Patients with Ovarian Cancer. *New England Journal of Medicine* **359**, 2641–2650 (2008).
281. Heravi-Moussavi, A. *et al.* Recurrent Somatic DICER1 Mutations in Nonepithelial Ovarian Cancers. *New England Journal of Medicine* **366**, 234–242 (2012).
282. Walz, A. L. *et al.* Recurrent DGCR8, DROSHA, and SIX Homeodomain Mutations in Favorable Histology Wilms Tumors. *Cancer Cell* **27**, 286–297 (2015).
283. Rupaimoole, R. & Slack, F. J. MicroRNA therapeutics: towards a new era for the management of cancer and other diseases. *Nature Reviews Drug Discovery* **16**, 203–222 (2017).
284. Elmén, J. *et al.* LNA-mediated microRNA silencing in non-human primates. *Nature* **452**, 896–899 (2008).
285. Janssen, H. L. A. *et al.* Treatment of HCV infection by targeting microRNA. *N. Engl. J. Med.* **368**, 1685–1694 (2013).

286. Shimakami, T. *et al.* Stabilization of hepatitis C virus RNA by an Ago2–miR-122 complex. *PNAS* **109**, 941–946 (2012).
287. Beg, M. S. *et al.* Phase I study of MRX34, a liposomal miR-34a mimic, administered twice weekly in patients with advanced solid tumors. *Invest New Drugs* **35**, 180–188 (2017).
288. van Zandwijk, N. *et al.* P1.02MesomiR 1: A Phase I study of TargomiRs in patients with refractory malignant pleural mesothelioma (MPM) and lung cancer (NSCLC). *Ann Oncol* **26**, ii16–ii16 (2015).
289. Reid, G. *et al.* Clinical development of TargomiRs, a miRNA mimic-based treatment for patients with recurrent thoracic cancer. *Epigenomics* **8**, 1079–1085 (2016).
290. Stalder, L. *et al.* The rough endoplasmatic reticulum is a central nucleation site of siRNA-mediated RNA silencing. *EMBO J.* **32**, 1115–1127 (2013).
291. Li, S. *et al.* MicroRNAs inhibit the translation of target mRNAs on the endoplasmic reticulum in Arabidopsis. *Cell* **153**, 562–574 (2013).
292. Barman, B. & Bhattacharyya, S. N. mRNA Targeting to Endoplasmic Reticulum Precedes Ago Protein Interaction and MicroRNA (miRNA)-mediated Translation Repression in Mammalian Cells. *J Biol Chem* **290**, 24650–24656 (2015).
293. Valadi, H. *et al.* Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. *Nat. Cell Biol.* **9**, 654–659 (2007).
294. Li, Y. *et al.* Circular RNA is enriched and stable in exosomes: a promising biomarker for cancer diagnosis. *Cell Research* **25**, 981–984 (2015).
295. Mitchell, P. S. *et al.* Circulating microRNAs as stable blood-based markers for cancer detection. *Proc Natl Acad Sci U S A* **105**, 10513–10518 (2008).
296. Chim, S. S. C. *et al.* Detection and Characterization of Placental MicroRNAs in Maternal Plasma. *Clinical Chemistry* **54**, 482–490 (2008).
297. Zhou, W. *et al.* Cancer-Secreted miR-105 Destroys Vascular Endothelial Barriers to Promote Metastasis. *Cancer Cell* **25**, 501–515 (2014).
298. Kren, B. T. *et al.* MicroRNAs identified in highly purified liver-derived mitochondria may play a role in apoptosis. *RNA Biol* **6**, 65–72 (2009).
299. Bian, Z. *et al.* Identification of mouse liver mitochondria-associated miRNAs and their potential biological functions. *Cell Research* **20**, 1076–1078 (2010).
300. Bandiera, S. *et al.* Nuclear Outsourcing of RNA Interference Components to Human Mitochondria. *PLOS ONE* **6**, e20746 (2011).
301. Das, S. *et al.* Nuclear miRNA Regulates the Mitochondrial Genome in the Heart Novelty and Significance. *Circulation Research* **110**, 1596–1603 (2012).
302. Zhang, X. *et al.* MicroRNA directly enhances mitochondrial translation during muscle differentiation. *Cell* **158**, 607–619 (2014).
303. Bandiera, S., Matégot, R., Girard, M., Demongeot, J. & Henrion-Caude, A. MitomiRs delineating the intracellular localization of microRNAs at mitochondria. *Free Radic. Biol. Med.* **64**, 12–19 (2013).
304. ZENG, Y. & CULLEN, B. R. Sequence requirements for micro RNA processing and function in human cells. *RNA* **9**, 112–123 (2003).
305. Eis, P. S. *et al.* Accumulation of miR-155 and BIC RNA in human B cell lymphomas. *Proc Natl Acad Sci U S A* **102**, 3627–3632 (2005).
306. Robb, G. B., Brown, K. M., Khurana, J. & Rana, T. M. Specific and potent RNAi in the nucleus of human cells. *Nat. Struct. Mol. Biol.* **12**, 133–137 (2005).
307. Gagnon, K. T., Li, L., Chu, Y., Janowski, B. A. & Corey, D. R. RNAi Factors are Present and Active in Human Cell Nuclei. *Cell Rep* **6**, 211–221 (2014).
308. Hwang, H.-W., Wentzel, E. A. & Mendell, J. T. A hexanucleotide element directs microRNA nuclear import. *Science* **315**, 97–100 (2007).
309. Ritland Politz, J. C., Zhang, F. & Pederson, T. MicroRNA-206 colocalizes with



- ribosome-rich regions in both the nucleolus and cytoplasm of rat myogenic cells. *Proc Natl Acad Sci U S A* **103**, 18957–18962 (2006).
310. Roberts, T. C. The MicroRNA Biology of the Mammalian Nucleus. *Molecular Therapy - Nucleic Acids* **3**, e188 (2014).
311. Ameyar-Zazoua, M. *et al.* Argonaute proteins couple chromatin silencing to alternative splicing. *Nat. Struct. Mol. Biol.* **19**, 998–1004 (2012).
312. Weinmann, L. *et al.* Importin 8 is a gene silencing factor that targets argonaute proteins to distinct mRNAs. *Cell* **136**, 496–507 (2009).
313. Kalantari, R. *et al.* Stable association of RNAi machinery is conserved between the cytoplasm and nucleus of human cells. *RNA* **22**, 1085 (2016).
314. Hicks, J. A. *et al.* Human GW182 Paralogs Are the Central Organizers for RNA-Mediated Control of Transcription. *Cell Reports* **20**, 1543–1552 (2017).
315. Morris, K. V., Chan, S. W.-L., Jacobsen, S. E. & Looney, D. J. Small Interfering RNA-Induced Transcriptional Gene Silencing in Human Cells. *Science* **305**, 1289–1292 (2004).
316. Kim, D. H., Villeneuve, L. M., Morris, K. V. & Rossi, J. J. Argonaute-1 directs siRNA-mediated transcriptional gene silencing in human cells. *Nat. Struct. Mol. Biol.* **13**, 793–797 (2006).
317. Li, L.-C. *et al.* Small dsRNAs induce transcriptional activation in human cells. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 17337–17342 (2006).
318. Janowski, B. A. *et al.* Activating gene expression in mammalian cells with promoter-targeted duplex RNAs. *Nat. Chem. Biol.* **3**, 166–173 (2007).
319. Han, J., Kim, D. & Morris, K. V. Promoter-associated RNA is required for RNA-directed transcriptional gene silencing in human cells. *Proc Natl Acad Sci U S A* **104**, 12422–12427 (2007).
320. Schwartz, J. C. *et al.* Antisense transcripts are targets for activating small RNAs. *Nat. Struct. Mol. Biol.* **15**, 842–848 (2008).
321. Napoli, S., Pastori, C., Magistri, M., Carbone, G. M. & Catapano, C. V. Promoter-specific transcriptional interference and c-myc gene silencing by siRNAs in human cells. *EMBO J.* **28**, 1708–1719 (2009).
322. Wang, D. *et al.* Quantitative functions of Argonaute proteins in mammalian development. *Genes Dev* **26**, 693–704 (2012).
323. Chu, Y., Yue, X., Younger, S. T., Janowski, B. A. & Corey, D. R. Involvement of argonaute proteins in gene silencing and activation by RNAs complementary to a non-coding transcript at the progesterone receptor promoter. *Nucleic Acids Res* **38**, 7736–7748 (2010).
324. Kim, D. H., Sætrom, P., Snøve, O. & Rossi, J. J. MicroRNA-directed transcriptional gene silencing in mammalian cells. *PNAS* **105**, 16230–16235 (2008).
325. Matsui, M. *et al.* Promoter RNA links transcriptional regulation of inflammatory pathway genes. *Nucleic Acids Res.* **41**, 10086–10109 (2013).
326. Benhamed, M., Herbig, U., Ye, T., Dejean, A. & Bischof, O. Senescence is an Endogenous Trigger for microRNA-Directed Transcriptional Gene Silencing in Human Cells. *Nat Cell Biol* **14**, 266–275 (2012).
327. Place, R. F., Li, L.-C., Pookot, D., Noonan, E. J. & Dahiya, R. MicroRNA-373 induces expression of genes with complementary promoter sequences. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 1608–1613 (2008).
328. Majid, S. *et al.* MicroRNA-205 directed transcriptional activation of tumor suppressor genes in prostate cancer. *Cancer* **116**, 5637–5649 (2010).
329. Luco, R. F. *et al.* Regulation of Alternative Splicing by Histone Modifications. *Science* **327**, 996–1000 (2010).
330. Alló, M. *et al.* Control of alternative splicing through siRNA-mediated transcriptional

- gene silencing. *Nat. Struct. Mol. Biol.* **16**, 717–724 (2009).
331. Hansen, T. B. *et al.* miRNA-dependent gene silencing involving Ago2-mediated cleavage of a circular antisense RNA. *EMBO J.* **30**, 4414–4422 (2011).
332. Leucci, E. *et al.* microRNA-9 targets the long non-coding RNA *MALAT1* for degradation in the nucleus. *Scientific Reports* **3**, 2535 (2013).
333. Tang, R. *et al.* Mouse miRNA-709 directly regulates miRNA-15a/16-1 biogenesis at the posttranscriptional level in the nucleus: evidence for a microRNA hierarchy system. *Cell Res.* **22**, 504–515 (2012).
334. Mochizuki, K., Fine, N. A., Fujisawa, T. & Gorovsky, M. A. Analysis of a piwi-Related Gene Implicates Small RNAs in Genome Rearrangement in Tetrahymena. *Cell* **110**, 689–699 (2002).
335. Malone, C. D., Anderson, A. M., Motl, J. A., Rexer, C. H. & Chalker, D. L. Germ Line Transcripts Are Processed by a Dicer-Like Protein That Is Essential for Developmentally Programmed Genome Rearrangements of Tetrahymena thermophila. *Mol. Cell. Biol.* **25**, 9151–9164 (2005).
336. Mochizuki, K. & Gorovsky, M. A. Conjugation-specific small RNAs in Tetrahymena have predicted properties of scan (scn) RNAs involved in genome rearrangement. *Genes Dev* **18**, 2068–2073 (2004).
337. Taverna, S. D., Coyne, R. S. & Allis, C. D. Methylation of Histone H3 at Lysine 9 Targets Programmed DNA Elimination in Tetrahymena. *Cell* **110**, 701–711 (2002).
338. Holoch, D. & Moazed, D. RNA-mediated epigenetic regulation of gene expression. *Nat Rev Genet* **16**, 71–84 (2015).
339. Wassenegger, M., Heimes, S., Riedel, L. & Sanger, H. L. RNA-directed de novo methylation of genomic sequences in plants. *Cell* **76**, 567–576 (1994).
340. Mette, M. F., van der Winden, J., Matzke, M. A. & Matzke, A. J. Production of aberrant promoter transcripts contributes to methylation and silencing of unlinked homologous promoters in trans. *EMBO J* **18**, 241–248 (1999).
341. Mette, M. F., Aufsatz, W., van der Winden, J., Matzke, M. A. & Matzke, A. J. M. Transcriptional silencing and promoter methylation triggered by double-stranded RNA. *EMBO J* **19**, 5194–5201 (2000).
342. Xie, Z. *et al.* Genetic and functional diversification of small RNA pathways in plants. *PLoS Biol.* **2**, E104 (2004).
343. Chan, S. W.-L. *et al.* RNA Silencing Genes Control de Novo DNA Methylation. *Science* **303**, 1336–1336 (2004).
344. Cao, X. & Jacobsen, S. E. Role of the arabidopsis DRM methyltransferases in de novo DNA methylation and gene silencing. *Curr. Biol.* **12**, 1138–1144 (2002).
345. Jones, L., Ratcliff, F. & Baulcombe, D. C. RNA-directed transcriptional gene silencing in plants can be inherited independently of the RNA trigger and requires Met1 for maintenance. *Current Biology* **11**, 747–757 (2001).
346. Zilberman, D., Cao, X. & Jacobsen, S. E. ARGONAUTE4 Control of Locus-Specific siRNA Accumulation and DNA and Histone Methylation. *Science* **299**, 716–719 (2003).
347. Law, J. A. *et al.* Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1. *Nature* **498**, 385–389 (2013).
348. Herr, A. J., Jensen, M. B., Dalmay, T. & Baulcombe, D. C. RNA polymerase IV directs silencing of endogenous DNA. *Science* **308**, 118–120 (2005).
349. Yu, B. *et al.* Methylation as a crucial step in plant microRNA biogenesis. *Science* **307**, 932–935 (2005).
350. Pontier, D. *et al.* Reinforcement of silencing at transposons and highly repeated sequences requires the concerted action of two distinct RNA polymerases IV in Arabidopsis. *Genes Dev* **19**, 2030–2040 (2005).

351. He, X.-J. *et al.* An effector of RNA-directed DNA methylation in Arabidopsis is an ARGONAUTE 4- and RNA-binding protein. *Cell* **137**, 498–508 (2009).
352. Li, C. F. *et al.* An ARGONAUTE4-Containing Nuclear Processing Center Colocalized with Cajal Bodies in Arabidopsis thaliana. *Cell* **126**, 93–106 (2006).
353. El-Shami, M. *et al.* Reiterated WG/GW motifs form functionally and evolutionarily conserved ARGONAUTE-binding platforms in RNAi-related components. *Genes Dev* **21**, 2539–2544 (2007).
354. Matzke, M. A. & Mosher, R. A. RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. *Nature Reviews Genetics* **15**, 394–408 (2014).
355. Hall, I. M. *et al.* Establishment and maintenance of a heterochromatin domain. *Science* **297**, 2232–2237 (2002).
356. Volpe, T. A. *et al.* Regulation of Heterochromatic Silencing and Histone H3 Lysine-9 Methylation by RNAi. *Science* **297**, 1833–1837 (2002).
357. Thon, G. & Verhein-Hansen, J. Four chromo-domain proteins of Schizosaccharomyces pombe differentially repress transcription at various chromosomal locations. *Genetics* **155**, 551–568 (2000).
358. Verdel, A. *et al.* RNAi-Mediated Targeting of Heterochromatin by the RITS Complex. *Science* **303**, 672–676 (2004).
359. Irvine, D. V. *et al.* Argonaute Slicing Is Required for Heterochromatic Silencing and Spreading. *Science* **313**, 1134–1137 (2006).
360. Motamedi, M. R. *et al.* Two RNAi complexes, RITS and RDRC, physically interact and localize to noncoding centromeric RNAs. *Cell* **119**, 789–802 (2004).
361. Buker, S. M. *et al.* Two different Argonaute complexes are required for siRNA generation and heterochromatin assembly in fission yeast. *Nat. Struct. Mol. Biol.* **14**, 200–207 (2007).
362. Holoch, D. & Moazed, D. Small-RNA loading licenses Argonaute for assembly into a transcriptional silencing complex. *Nat Struct Mol Biol* **22**, 328–335 (2015).
363. Bühler, M., Verdel, A. & Moazed, D. Tethering RITS to a Nascent Transcript Initiates RNAi- and Heterochromatin-Dependent Gene Silencing. *Cell* **125**, 873–886 (2006).
364. Moazed, D. Small RNAs in transcriptional gene silencing and genome defence. *Nature* **457**, 413–420 (2009).
365. Noma, K. *et al.* RITS acts in *cis* to promote RNA interference-mediated transcriptional and post-transcriptional silencing. *Nature Genetics* **36**, 1174–1180 (2004).
366. Bühler, M., Haas, W., Gygi, S. P. & Moazed, D. RNAi-dependent and -independent RNA turnover mechanisms contribute to heterochromatic gene silencing. *Cell* **129**, 707–721 (2007).
367. Chen, C.-C. G. *et al.* A member of the polymerase beta nucleotidyltransferase superfamily is required for RNA interference in *C. elegans*. *Curr. Biol.* **15**, 378–383 (2005).
368. Bühler, M., Spies, N., Bartel, D. P. & Moazed, D. TRAMP-mediated RNA surveillance prevents spurious entry of RNAs into the Schizosaccharomyces pombe siRNA pathway. *Nat Struct Mol Biol* **15**, 1015–1023 (2008).
369. Brennecke, J. *et al.* Discrete Small RNA-Generating Loci as Master Regulators of Transposon Activity in Drosophila. *Cell* **128**, 1089–1103 (2007).
370. Schmidt, A. *et al.* Genetic and molecular characterization of sting, a gene involved in crystal formation and meiotic drive in the male germ line of Drosophila melanogaster. *Genetics* **151**, 749–760 (1999).
371. Stapleton, W., Das, S. & McKee, B. D. A role of the *Drosophila* *homeless* gene in repression of *Stellate* in male meiosis. *Chromosoma* **110**, 228–240 (2001).
372. Aravin, A. A. *et al.* Double-stranded RNA-mediated silencing of genomic tandem

- repeats and transposable elements in the *D. melanogaster* germline. *Curr. Biol.* **11**, 1017–1027 (2001).
373. Vagin, V. V. *et al.* A Distinct Small RNA Pathway Silences Selfish Genetic Elements in the Germline. *Science* **313**, 320–324 (2006).
374. Dorer, D. R. & Henikoff, S. Expansions of transgene repeats cause heterochromatin formation and gene silencing in *Drosophila*. *Cell* **77**, 993–1002 (1994).
375. Pal-Bhadra, M. *et al.* Heterochromatic Silencing and HP1 Localization in *Drosophila* Are Dependent on the RNAi Machinery. *Science* **303**, 669–672 (2004).
376. Pane, A., Wehr, K. & Schüpbach, T. zucchini and squash encode two putative nucleases required for rasiRNA production in the *Drosophila* germline. *Dev. Cell* **12**, 851–862 (2007).
377. Malone, C. D. *et al.* Specialized piRNA Pathways Act in Germline and Somatic Tissues of the *Drosophila* Ovary. *Cell* **137**, 522–535 (2009).
378. Li, C. *et al.* Collapse of germline piRNAs in the absence of Argonaute3 reveals somatic piRNAs in flies. *Cell* **137**, 509–521 (2009).
379. Saito, K. *et al.* Pimet, the *Drosophila* homolog of HEN1, mediates 2'-O-methylation of Piwi-interacting RNAs at their 3' ends. *Genes Dev.* **21**, 1603–1608 (2007).
380. Gunawardane, L. S. *et al.* A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in *Drosophila*. *Science* **315**, 1587–1590 (2007).
381. Girard, A., Sachidanandam, R., Hannon, G. J. & Carmell, M. A. A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* **442**, 199–202 (2006).
382. Brennecke, J. *et al.* An Epigenetic Role for Maternally Inherited piRNAs in Transposon Silencing. *Science* **322**, 1387–1392 (2008).
383. Ipsaro, J. J., Haase, A. D., Knott, S. R., Joshua-Tor, L. & Hannon, G. J. The structural biochemistry of Zucchini implicates it as a nuclease in piRNA biogenesis. *Nature* **491**, 279–283 (2012).
384. Tang, W., Tu, S., Lee, H.-C., Weng, Z. & Mello, C. C. The ribonuclease PARN-1 trims piRNA 3' ends to promote transcriptome surveillance in *C. elegans*. *Cell* **164**, 974–984 (2016).
385. Hayashi, R. *et al.* Genetic and mechanistic diversity of piRNA 3'-end formation. *Nature* **539**, 588–592 (2016).
386. Izumi, N. *et al.* Identification and Functional Analysis of the Pre-piRNA 3' Trimmer in Silkworms. *Cell* **164**, 962–973 (2016).
387. Ding, D. *et al.* PNLDC1 is essential for piRNA 3' end trimming and transposon silencing during spermatogenesis in mice. *Nature Communications* **8**, 819 (2017).
388. Czech, B. & Hannon, G. J. A Happy 3' Ending to the piRNA Maturation Story. *Cell* **164**, 838–840 (2016).
389. Tóth, K. F., Pezic, D., Stuwe, E. & Webster, A. The piRNA Pathway Guards the Germline Genome Against Transposable Elements. *Adv Exp Med Biol* **886**, 51–77 (2016).
390. Sienski, G., Dönertas, D. & Brennecke, J. Transcriptional silencing of transposons by Piwi and maelstrom and its impact on chromatin state and gene expression. *Cell* **151**, 964–980 (2012).
391. Muerdter, F. *et al.* A genome-wide RNAi screen draws a genetic framework for transposon control and primary piRNA biogenesis in *Drosophila*. *Mol. Cell* **50**, 736–748 (2013).
392. Czech, B., Preall, J. B., McGinn, J. & Hannon, G. J. A transcriptome-wide RNAi screen in the *Drosophila* ovary reveals factors of the germline piRNA pathway. *Mol. Cell* **50**, 749–761 (2013).
393. Gagnon, K. T., Li, L., Janowski, B. A. & Corey, D. R. Analysis of nuclear RNA interference in human cells by subcellular fractionation and Argonaute loading. *Nat Protoc* **9**,

2045–2060 (2014).

394. Minajigi, A. *et al.* Chromosomes. A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation. *Science* **349**, (2015).
395. Reed, S. I. The Selection of *S. Cerevisiae* Mutants Defective in the Start Event of Cell Division. *Genetics* **95**, 561–577 (1980).
396. Collart, M. A. & Struhl, K. CDC39, an essential nuclear protein that negatively regulates transcription and differentially affects the constitutive and inducible HIS3 promoters. *EMBO J.* **12**, 177–186 (1993).
397. Lee, T. I. *et al.* Interplay of Positive and Negative Regulators in Transcription Initiation by RNA Polymerase II Holoenzyme. *Mol Cell Biol* **18**, 4455–4462 (1998).
398. Russell, P., Benson, J. D. & Denis, C. L. Characterization of Mutations in NOT2 Indicates that it Plays an Important Role in Maintaining the Integrity of the CCR4–NOT Complex. *Journal of Molecular Biology* **322**, 27–39 (2002).
399. Boland, A. *et al.* Structure and assembly of the NOT module of the human CCR4–NOT complex. *Nature Structural & Molecular Biology* **20**, 1289 (2013).
400. Greimann, J. C. & Lima, C. D. Reconstitution of RNA exosomes from human and *Saccharomyces cerevisiae*: Cloning, expression, purification, and activity assays. *Methods Enzymol* **448**, 185–210 (2008).
401. Zhang, H. *et al.* PABPC1 interacts with AGO2 and is responsible for the microRNA mediated gene silencing in high grade hepatocellular carcinoma. *Cancer Lett.* **367**, 49–57 (2015).
402. Schopp, I. M. *et al.* Split-BioID a conditional proteomics approach to monitor the composition of spatiotemporally defined protein complexes. *Nature Communications* **8**, ncomms15690 (2017).
403. Knott, G. J., Lee, M., Passon, D. M., Fox, A. H. & Bond, C. S. *Caenorhabditis elegans* NONO-1: Insights into DBHS protein structure, architecture, and function. *Protein Sci* **24**, 2033–2043 (2015).
404. Knott, G. J., Bond, C. S. & Fox, A. H. The DBHS proteins SFPQ, NONO and PSPC1: a multipurpose molecular scaffold. *Nucleic Acids Res* **44**, 3989–4004 (2016).
405. Fox, A. H. & Lamond, A. I. Paraspeckles. *Cold Spring Harb Perspect Biol* **2**, (2010).
406. Passon, D. M. *et al.* Structure of the heterodimer of human NONO and paraspeckle protein component 1 and analysis of its role in subnuclear body formation. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 4846–4850 (2012).
407. Kuwahara, S. *et al.* PSPC1, NONO, and SFPQ are expressed in mouse Sertoli cells and may function as coregulators of androgen receptor-mediated transcription. *Biol. Reprod.* **75**, 352–359 (2006).
408. Fox, A. H. *et al.* Paraspeckles: a novel nuclear domain. *Curr. Biol.* **12**, 13–25 (2002).
409. Sunwoo, H. *et al.* 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 (2009).
410. Sasaki, Y. T. F., Ideue, T., Sano, M., Mituyama, T. & Hirose, T. MENε/β noncoding RNAs are essential for structural integrity of nuclear paraspeckles. *PNAS* **106**, 2525–2530 (2009).
411. Clemson, C. M. *et al.* An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles. *Mol. Cell* **33**, 717–726 (2009).
412. Mircsof, D. *et al.* Mutations in NONO lead to syndromic intellectual disability and inhibitory synaptic defects. *Nat Neurosci* **18**, 1731–1736 (2015).
413. Patton, J. G., Porro, E. B., Galceran, J., Tempst, P. & Nadal-Ginard, B. Cloning and characterization of PSF, a novel pre-mRNA splicing factor. *Genes Dev.* **7**, 393–406 (1993).
414. Yang, Y. S. *et al.* NonO, a non-POU-domain-containing, octamer-binding protein, is

- the mammalian homolog of *Drosophila* nonAdiss. *Mol. Cell. Biol.* **13**, 5593–5603 (1993).
415. Peng, R. *et al.* PSF and p54nrb bind a conserved stem in U5 snRNA. *RNA* **8**, 1334–1347 (2002).
416. Kameoka, S., Duque, P. & Konarska, M. M. p54nrb associates with the 5' splice site within large transcription/splicing complexes. *EMBO J* **23**, 1782–1791 (2004).
417. Kaneko, S., Rozenblatt-Rosen, O., Meyerson, M. & Manley, J. L. The multifunctional protein p54nrb/PSF recruits the exonuclease XRN2 to facilitate pre-mRNA 3' processing and transcription termination. *Genes Dev.* **21**, 1779–1789 (2007).
418. Ito, T. *et al.* Brm transactivates the telomerase reverse transcriptase (TERT) gene and modulates the splicing patterns of its transcripts in concert with p54(nrb). *Biochem. J.* **411**, 201–209 (2008).
419. Salton, M., Lerenthal, Y., Wang, S.-Y., Chen, D. J. & Shiloh, Y. Involvement of Matrin 3 and SFPQ/NONO in the DNA damage response. *Cell Cycle* **9**, 1568–1576 (2010).
420. Luisier, R. *et al.* Intron retention and nuclear loss of SFPQ are molecular hallmarks of ALS. *Nat Commun* **9**, 2010 (2018).
421. Brown, S. A. *et al.* PERIOD1-associated proteins modulate the negative limb of the mammalian circadian oscillator. *Science* **308**, 693–696 (2005).
422. Ke, Y. D. *et al.* Tau-mediated nuclear depletion and cytoplasmic accumulation of SFPQ in Alzheimer's and Pick's disease. *PLoS ONE* **7**, e35678 (2012).
423. Kanai, Y., Dohmae, N. & Hirokawa, N. Kinesin transports RNA: isolation and characterization of an RNA-transporting granule. *Neuron* **43**, 513–525 (2004).
424. Furukawa, M. T., Sakamoto, H. & Inoue, K. Interaction and colocalization of HERMES/RBPMS with NonO, PSF, and G3BP1 in neuronal cytoplasmic RNP granules in mouse retinal line cells. *Genes Cells* **20**, 257–266 (2015).
425. Lukong, K. E., Huot, M.-E. & Richard, S. BRK phosphorylates PSF promoting its cytoplasmic localization and cell cycle arrest. *Cell. Signal.* **21**, 1415–1422 (2009).
426. Cosker, K. E., Fenstermacher, S. J., Pazyra-Murphy, M. F., Elliott, H. L. & Segal, R. A. The RNA-binding protein SFPQ orchestrates an RNA regulon to promote axon viability. *Nat. Neurosci.* **19**, 690–696 (2016).
427. Thomas-Jinu, S. *et al.* Non-nuclear Pool of Splicing Factor SFPQ Regulates Axonal Transcripts Required for Normal Motor Development. *Neuron* **94**, 322–336.e5 (2017).
428. Fox, A. H., Bond, C. S. & Lamond, A. I. P54nrb forms a heterodimer with PSP1 that localizes to paraspeckles in an RNA-dependent manner. *Mol. Biol. Cell* **16**, 5304–5315 (2005).
429. Dreyfuss, G., Kim, V. N. & Kataoka, N. Messenger-RNA-binding proteins and the messages they carry. *Nature Reviews Molecular Cell Biology* **3**, 195–205 (2002).
430. Chen, L.-L. & Carmichael, G. G. 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 (2009).
431. Yamazaki, T. & Hirose, T. The building process of the functional paraspeckle with long non-coding RNAs. *Front Biosci (Elite Ed)* **7**, 1–41 (2015).
432. Zhang, T. *et al.* FUS Regulates Activity of MicroRNA-Mediated Gene Silencing. *Mol. Cell* **69**, 787–801.e8 (2018).
433. Kawahara, Y. & Mieda-Sato, A. TDP-43 promotes microRNA biogenesis as a component of the Drosha and Dicer complexes. *Proc Natl Acad Sci U S A* **109**, 3347–3352 (2012).
434. Hu, S.-B., Yao, R.-W. & Chen, L.-L. Shedding light on paraspeckle structure by super-resolution microscopy. *J Cell Biol* jcb.201609008 (2016). doi:10.1083/jcb.201609008
435. Liu, X. *et al.* A MicroRNA Precursor Surveillance System in Quality Control of MicroRNA Synthesis. *Mol Cell* **55**, 868–879 (2014).

436. Singh, I. *et al.* MiCEE is a ncRNA-protein complex that mediates epigenetic silencing and nucleolar organization. *Nature Genetics* **1** (2018). doi:10.1038/s41588-018-0139-3
437. Azzouz, N., Panasenko, O. O., Colau, G. & Collart, M. A. The CCR4-NOT Complex Physically and Functionally Interacts with TRAMP and the Nuclear Exosome. *PLoS ONE* **4**, (2009).
438. Albert, T. K. *et al.* Identification of a ubiquitin–protein ligase subunit within the CCR4–NOT transcription repressor complex. *The EMBO Journal* **21**, 355–364 (2002).
439. Collart, M. A. & Panasenko, O. O. The Ccr4-Not Complex: Architecture and Structural Insights. *Subcell. Biochem.* **83**, 349–379 (2017).
440. Collart, M. A. & Struhl, K. NOT1(CDC39), NOT2(CDC36), NOT3, and NOT4 encode a global-negative regulator of transcription that differentially affects TATA-element utilization. *Genes Dev.* **8**, 525–537 (1994).
441. Maillet, L., Tu, C., Hong, Y. K., Shuster, E. O. & Collart, M. A. The essential function of not1 lies within the Ccr4-not complex1. *Journal of Molecular Biology* **303**, 131–143 (2000).
442. Bawankar, P., Loh, B., Wohlbold, L., Schmidt, S. & Izaurralde, E. NOT10 and C2orf29/NOT11 form a conserved module of the CCR4-NOT complex that docks onto the NOT1 N-terminal domain. *RNA Biol* **10**, 228–244 (2013).
443. Bai, Y. *et al.* The CCR4 and CAF1 Proteins of the CCR4-NOT Complex Are Physically and Functionally Separated from NOT2, NOT4, and NOT5. *Mol. Cell. Biol.* **19**, 6642–6651 (1999).
444. Ito, K. *et al.* CNOT2 depletion disrupts and inhibits the CCR4–NOT deadenylase complex and induces apoptotic cell death. *Genes to Cells* **16**, 368–379 (2011).
445. Alhusaini, N. & Collier, J. The deadenylase components Not2p, Not3p, and Not5p promote mRNA decapping. *RNA* **22**, 709–721 (2016).
446. Temme, C. *et al.* Subunits of the Drosophila CCR4-NOT complex and their roles in mRNA deadenylation. *RNA* **16**, 1356–1370 (2010).
447. Ito, K., Takahashi, A., Morita, M., Suzuki, T. & Yamamoto, T. The role of the CNOT1 subunit of the CCR4-NOT complex in mRNA deadenylation and cell viability. *Protein Cell* **2**, 755–763 (2011).
448. Mauxion, F., Prève, B. & Séraphin, B. C2ORF29/CNOT11 and CNOT10 form a new module of the CCR4-NOT complex. *RNA Biol* **10**, 267–276 (2013).
449. Jarvis, E. D. *et al.* Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* **346**, 1320–1331 (2014).
450. Zheng, X. *et al.* CNOT3-Dependent mRNA Deadenylation Safeguards the Pluripotent State. *Stem Cell Reports* **7**, 897 (2016).
451. Azzouz, N. *et al.* Specific roles for the Ccr4-Not complex subunits in expression of the genome. *RNA* **15**, 377–383 (2009).
452. Rodríguez-Gil, A. *et al.* The CCR4-NOT complex contributes to repression of Major Histocompatibility Complex class II transcription. *Scientific Reports* **7**, 3547 (2017).
453. De Keersmaecker, K. *et al.* Exome sequencing identifies mutation in CNOT3 and ribosomal genes RPL5 and RPL10 in T-cell acute lymphoblastic leukemia. *Nat. Genet.* **45**, 186–190 (2013).
454. Stowell, J. A. W. *et al.* A low-complexity region in the YTH domain protein Mmi1 enhances RNA binding. *J. Biol. Chem.* (2018). doi:10.1074/jbc.RA118.002291
455. Suzuki, A., Igarashi, K., Aisaki, K.-I., Kanno, J. & Saga, Y. NANOS2 interacts with the CCR4-NOT deadenylation complex and leads to suppression of specific RNAs. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 3594–3599 (2010).
456. Sandler, H., Kreth, J., Timmers, H. T. M. & Stoecklin, G. Not1 mediates recruitment of the deadenylase Caf1 to mRNAs targeted for degradation by tristetraprolin. *Nucleic Acids*

- Res* **39**, 4373–4386 (2011).
457. Igreja, C. & Izaurralde, E. CUP promotes deadenylation and inhibits decapping of mRNA targets. *Genes Dev.* **25**, 1955–1967 (2011).
458. Leppik, K. *et al.* Roquin promotes constitutive mRNA decay via a conserved class of stem-loop recognition motifs. *Cell* **153**, 869–881 (2013).
459. Elkayam, E. *et al.* Multivalent Recruitment of Human Argonaute by GW182. *Molecular Cell* **67**, 646–658.e3 (2017).
460. Schneider, C. & Tollervey, D. Threading the barrel of the RNA exosome. *Trends in Biochemical Sciences* **38**, 485–493 (2013).
461. Kilchert, C., Wittmann, S. & Vasiljeva, L. The regulation and functions of the nuclear RNA exosome complex. *Nature Reviews Molecular Cell Biology* **17**, 227–239 (2016).
462. LaCava, J. *et al.* RNA degradation by the exosome is promoted by a nuclear polyadenylation complex. *Cell* **121**, 713–724 (2005).
463. Thoms, M. *et al.* The Exosome Is Recruited to RNA Substrates through Specific Adaptor Proteins. *Cell* **162**, 1029–1038 (2015).
464. Harigaya, Y. *et al.* Selective elimination of messenger RNA prevents an incidence of untimely meiosis. *Nature* **442**, 45–50 (2006).
465. Kilchert, C. *et al.* Regulation of mRNA Levels by Decay-Promoting Introns that Recruit the Exosome Specificity Factor Mmi1. *Cell Rep* **13**, 2504–2515 (2015).
466. Stowell, J. A. W. *et al.* Reconstitution of Targeted Deadenylation by the Ccr4-Not Complex and the YTH Domain Protein Mmi1. *Cell Rep* **17**, 1978–1989 (2016).
467. Tudek, A. *et al.* Molecular Basis for Coordinating Transcription Termination with Noncoding RNA Degradation. *Molecular Cell* **55**, 467–481 (2014).
468. Webb, S., Hector, R. D., Kudla, G. & Granneman, S. PAR-CLIP data indicate that Nrd1-Nab3-dependent transcription termination regulates expression of hundreds of protein coding genes in yeast. *Genome Biol.* **15**, R8 (2014).
469. Lubas, M. *et al.* Interaction profiling identifies the human nuclear exosome targeting complex. *Mol. Cell* **43**, 624–637 (2011).
470. Lubas, M. *et al.* The Human Nuclear Exosome Targeting Complex Is Loaded onto Newly Synthesized RNA to Direct Early Ribonucleolysis. *Cell Reports* **10**, 178–192 (2015).
471. Macias, S., Cordiner, R. A., Gautier, P., Plass, M. & Cáceres, J. F. DGCR8 Acts as an Adaptor for the Exosome Complex to Degrade Double-Stranded Structured RNAs. *Mol Cell* **60**, 873–885 (2015).
472. Li, W., Notani, D. & Rosenfeld, M. G. Enhancers as non-coding RNA transcription units: recent insights and future perspectives. *Nature Reviews Genetics* **17**, 207–223 (2016).
473. Wyers, F. *et al.* Cryptic Pol II Transcripts Are Degraded by a Nuclear Quality Control Pathway Involving a New Poly(A) Polymerase. *Cell* **121**, 725–737 (2005).
474. Li, W. *et al.* Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature* **498**, 516–520 (2013).
475. Ørom, U. A. *et al.* Long noncoding RNAs with enhancer-like function in human cells. *Cell* **143**, 46–58 (2010).
476. Lam, M. T. Y. *et al.* Rev-Erbs repress macrophage gene expression by inhibiting enhancer-directed transcription. *Nature* **498**, 511–515 (2013).
477. Sigova, A. A. *et al.* Transcription factor trapping by RNA in gene regulatory elements. *Science* **350**, 978–981 (2015).



## 12 Note biographique

Nom: Raphaël Matégot

Date de naissance: 2 Juin 1990

Lieu de naissance: Paris, France

Nationalité : Française

### Education

- 2004-2007 Lycée Paul-Bert, Paris, France
- 2007-2011 Université Paris Descartes, Paris  
Licence de Biologie, Juin 2011
- 2012-2013 Université Sorbonne, Ecole Pratique des Hautes Etudes, Paris  
Master en Biologie, Juin 2013
- 2014-2018 Ecole doctorale Université Nice Côte d'Azur, Nice Sophia Antipolis, Nice  
Doctorat en Biologie Moléculaire, 21 Septembre 2018

### Recherche

- 2011-2012 Inserm U978 - Health, Medicine and Human biology Paris 13 University, Signaling adaptor proteins in Hematology group, avec Dr. Lionel Guittat
- 2012-2013 Inserm U781 - Imagine Institute, Hôpital Necker-enfants malades, Genetics and epigenetics of metabolic, neurosensorial and developmental disorders group, avec Dr. Alexandra Henrion-Caude
- 2014-2018 Inserm U1065 - Centre de Médecine Moléculaire Méditerranéen, Thèse de doctorat dans le groupe Control of gene expression, avec Dr. Michele Trabucchi

### Publications

1. Bandiera, S., Matégot, R., Girard, M., Demongeot, J. & Henrion-Caude, A. MitomiRs delineating the intracellular localization of microRNAs at mitochondria. *Free Radic. Biol. Med.* **64**, 12–19 (2013).
2. Girard, M. *et al.* MYO5B and bile salt export pump contribute to cholestatic liver disorder in microvillous inclusion disease. *Hepatology* **60**, 301–310 (2014).
3. Bottini, S. *et al.* Post-transcriptional gene silencing mediated by microRNAs is controlled by nucleoplasmic SFPQ. *Nat Commun* **8**, 1189 (2017).