

PROJET DE RECHERCHE

présenté devant

L'UNIVERSITÉ TOULOUSE III — PAUL SABATIER

pour obtenir une

HABILITATION À DIRIGER LES RECHERCHES

Discipline : Biologie moléculaire

présenté et soutenu

par

Hervé SEITZ

Petits ARN régulateurs chez les Animaux

Soutenance publique dans la salle de conférences du bâtiment IBCG, le 3 novembre 2009
à 14h30 devant la commission d'examen :

CAVAILLÉ Jérôme, Directeur de recherche, CNRS, Toulouse	Directeur des recherches
ANTONIEWSKI Christophe, Directeur de recherche, CNRS, Paris	Rapporteur
CAVALLI Giacomo, Directeur de recherche, CNRS, Montpellier	Rapporteur
VOINNET Olivier, Directeur de recherche, CNRS, Strasbourg	Rapporteur
BUCHETON Alain, Directeur de recherche, CNRS, Montpellier	Examineur
PRATS Hervé, Professeur des universités, UPS, Toulouse	Président du jury

Abréviations

ADN :	acide désoxyribonucléique
ARN :	acide ribonucléique
ARN _m :	ARN messenger
cDNA :	ADN copie (issu de la reverse-transcription d'un ARN)
kb (ou : kpb) :	kilo-paire de bases
Mb (ou : Mpb) :	méga-paire de bases
miRNA :	microARN (voir page 19)
nt :	nucléotide
pb :	paire de bases
RdRP :	polymérase à ARN dépendant de l'ARN (voir page 13)
RISC :	complexe effecteur du RNAi (voir page 11)
RNAi :	interférence à ARN (voir page 11)
RNase :	ribonucléase
siRNA :	petit ARN interférent (voir page 11)
UTR :	région non traduite d'un ARN messenger

Notes

Dans tout le manuscrit, les « ou » sont des « ou » inclusifs, susceptibles d'être remplacés par des « et/ou ».

Le terme « transposon » est ici utilisé dans le sens : « élément génétique mobile », qui inclut les transposons à ADN, les rétrotransposons et les héliçons (certains auteurs réservent le nom de « transposons » aux transposons à ADN ; ce ne sera pas le cas ici).

Table des matières

1	Introduction	9
1.1	Avertissement	10
1.2	L'interférence à ARN, une répression post-transcriptionnelle	11
1.2.1	Présentation générale	11
1.2.2	Biogenèse des siRNA	12
1.2.3	Les protéines Argonaute, partenaires des petits ARN	14
1.2.4	Activité biologique des siRNA	16
1.3	Les microARN, des régulateurs post-transcriptionnels endogènes	19
1.3.1	Présentation générale	19
1.3.2	Biogenèse des microARN	20
1.3.3	Modes d'action des microARN	22
2	Des siRNA endogènes chez la Drosophile ?	29
2.1	Contexte : quel rôle endogène pour le RNAi ?	30
2.2	Article n°1 : une nouvelle classe de petits ARN	31
2.3	Discussion : finalement, existe-t-il des siRNA endogènes chez la Drosophile ?	36
2.4	Article n°2 : découverte des siRNA endogènes de Drosophile	38
2.5	Discussion : une nouvelle classe de régulateurs	43
3	La biogenèse des microARN : mécanismes de contrôle	45
3.1	Contexte : hétérogénéité des extrémités de microARN	46
3.2	Article n°3 : un contrôle de qualité des microARN	47
3.3	Discussion : une nouvelle étape de contrôle de la biogenèse des microARN	52
3.4	Revue : structure secondaire et contrôle du clivage par Drosha	54
3.5	Discussion : la structure secondaire des pri-miRNA, un module régulateur	57
3.6	Article n°4 : caractéristiques du chargement des microARN	58
3.7	Discussion : caractérisation biochimique d'une machine moléculaire inconnue	67
4	La biogenèse des piRNA	69
4.1	Contexte : qu'est-ce qui fait un piRNA ?	70
4.2	Article n°5 : deux voies de biogenèse pour les piRNA	73
4.3	Discussion : qu'est-ce qui fait un locus de piRNA ?	86
4.4	Article n°6 : biogenèse des piRNA guidée par des modifications de la chromatine	89
4.5	Discussion : un système de défense réactif face aux transposons	102
5	Quel est le rôle biologique des piRNA ?	103
5.1	Contexte : un rôle redondant avec celui des siRNA ?	104
5.2	Article en préparation : piRNA, microARN et siRNA au cours de l'évolution des Animaux	106
5.3	Discussion : une fonction supplémentaire inconnue	112

6	Quel est le rôle biologique des microARN ?	113
6.1	Contexte : le dogme actuel et ses limites	114
6.2	Article n°7 : une nouvelle définition des cibles de microARN	115
6.3	Discussion : un nouveau cadre théorique	119
7	Projet de recherche	121
7.1	Quelle est la fonction des piRNA ?	122
7.1.1	Objectif	122
7.1.2	Démarche expérimentale	122
7.2	Qu'est-ce qu'une cible de microARN ?	127
7.2.1	Objectif	127
7.2.2	Démarche expérimentale	127
7.2.3	Expériences supplémentaires	130
7.3	Régulations post-transcriptionnelles de l'expression des microARN	131
7.3.1	Objectif	131
7.3.2	Démarche expérimentale	132
	Bibliographie	133
	Index	148

Table des figures

1.1	Les protéines Argonaute	15
1.2	Le RNAi chez les Animaux	18
1.3	La biogenèse des microARN chez les Animaux	21
3.1	Drosha et Dicer possèdent deux domaines RNase III chacune	52
4.1	Le modèle du ping-pong pour la biogenèse des piRNA	71
5.1	Arbre phylogénétique des Animaux	105
5.2	Le cycle de <i>Nematostella vectensis</i>	107
5.3	Banques de petits ARN de <i>Nematostella</i>	108
5.4	Deux catégories de piRNA chez <i>Nematostella</i>	109
5.5	Des siRNA endogènes chez <i>Nematostella</i>	111
7.1	Les protéines Argonaute de <i>Nematostella vectensis</i>	123
7.2	Répartition géographique de <i>Nematostella vectensis</i>	124

Liste des tableaux

1.1	Recherches de cibles de microARN par des approches expérimentales	26
2.1	De probables siRNA endogènes chez la Drosophile et l'Homme	36
5.1	Protéines associées aux petits ARN chez <i>Nematostella</i>	106

Chapitre 1

Introduction

1.1 Avertissement

Ce chapitre d'introduction présentera les connaissances relatives aux microARN et au RNAi nécessaires à la compréhension des chapitres suivants. Les données les plus anciennes (antérieures à la fin 2004) seront simplement résumées, et le lecteur est invité à consulter la section « 3. siRNA, microARN et leur métabolisme » de ma thèse ¹ pour un exposé plus complet. Seuls les résultats plus récents seront ici décrits dans le détail.

¹Disponible notamment sur
<http://tel.archives-ouvertes.fr/docs/00/04/74/70/PDF/tel-00007781.pdf>
et sur
<http://www.normalesup.org/~seitz/These.pdf>

1.2 L'interférence à ARN, une répression post-transcriptionnelle

1.2.1 Présentation générale

L'injection d'ARN antisens d'un ARN_m est quelque fois utilisée pour réprimer un gène-cible de manière spécifique ; on considère alors que l'appariement entre les deux ARN empêche le ribosome d'accéder à l'ARN_m cible, et en réprime donc la traduction. Mais deux observations curieuses s'accordaient mal avec cette interprétation chez le Nématode *Caenorhabditis elegans* : d'une part, des préparations d'ARN sens pouvaient réprimer la cible aussi efficacement que l'ARN antisens ; d'autre part, la répression de la cible pouvait être transmise à la génération suivante, alors que les ARN sont rapidement dégradés dans l'embryon de Nématode. Ces contradictions ont amené Fire *et al.*, 1998 à proposer que l'effet répressif n'était pas dû à l'ARN antisens proprement dit, mais aux traces d'ARN double-brin qui contaminent aussi bien les préparations d'ARN sens qu'antisens. Ils ont effectivement montré que la répression était due à l'ARN double-brin, et qu'elle disparaissait quand l'ARN simple-brin était soigneusement purifié. Ce phénomène, baptisé « interférence à ARN », et abrégé en : RNAi (*RNA interference*), est observé chez la plupart des Animaux, des Plantes et des Champignons (Cogoni et Macino, 2000).

Les années suivantes ont assisté à la compréhension progressive des mécanismes moléculaires de cette répression. L'ARN double-brin déclencheur est clivé en duplex d'une vingtaine de paires de bases, appelés « siRNA » (*small interfering RNAs*, petits ARN interférents) (Yang *et al.*, 2000 ; Zamore *et al.*, 2000 ; Parrish *et al.*, 2000) par une enzyme de la famille RNase III, appelée « Dicer » (Bernstein *et al.*, 2001 ; Knight et Bass, 2001). Ces petits duplex sont chargés sur une protéine de la famille Argonaute (voir leur description détaillée ci-dessous, sous-section 1.2.3, page 14) ; l'un des deux brins du siRNA est ensuite dégradé par la protéine Argonaute (Matranga *et al.*, 2005 ; Rand *et al.*, 2005), et le complexe ribonucléique mature (appelé RISC, pour : *RNA-induced silencing complex*) contient donc un unique brin d'ARN. Par convention, on appelle « brin guide » le brin d'ARN qui reste stablement associé à la protéine Argonaute, et « brin passager » le brin rapidement dégradé.

Le complexe RISC clive les ARN complémentaires de son ARN guide, par une activité endonucléolytique portée par la protéine Argonaute (toutes les protéines de la famille Argonaute ne sont toutefois pas capables de catalyser ce clivage : voir sous-section 1.2.3, page 14 ; mais les siRNA sont généralement chargés sur les membres de la famille qui possèdent cette activité endonucléolytique : voir « Chargement sur Argonaute : le triage des petits ARN », dans la sous-section 1.3.2, page 21).

Le RNAi est rapidement devenu un outil populaire pour réprimer des gènes d'intérêt de manière simple, chez de nombreux organismes modèles. L'introduction d'ARN double-brin entraîne, chez certaines espèces, une répression traductionnelle non-spécifique appelée « réponse interféron » (c'est le cas des Mammifères) ; dans ce cas, l'expérimentateur traite les cellules-cibles avec des siRNA, au lieu du long ARN double-brin (Elbashir *et al.*, 2001a) : les siRNA synthétiques sont efficacement reconnus par la machinerie de chargement des protéines Argonaute, et ils ne déclenchent pas de réponse interféron². Une autre source de siRNA est fréquemment utilisée pour réprimer des gènes-cibles *in vivo* : les « shRNA » (*short hairpin RNAs*, ARN provenant de courtes tiges-boucles) (Hannon et Conklin, 2004). Ce sont des ARN qui se replient en tiges-boucles parfaites d'environ 70 nt, transcrits à partir de constructions transgéniques, et qui sont digérés par Dicer *in vivo* pour générer un petit duplex parfaitement apparié (donc finalement, un siRNA) ; ces constructions permettent donc de réprimer stablement un gène d'intérêt, éventuellement dans certains tissus uniquement (selon la nature du promoteur du shRNA), dans des animaux transgéniques.

²Il existe toutefois une controverse : Sledz *et al.*, 2003 ont montré que les siRNA provoquaient une réponse interféron dans une variété de lignées cellulaires et de cellules primaires, même à de faibles doses de siRNA.

1.2.2 Biogenèse des siRNA

siRNA d'origine exogène

Deux sources de siRNA exogènes sont connues chez les Animaux : les ARN viraux (la réplication des virus à ARN implique un intermédiaire double-brin) et les ARN double-brin synthétiques, introduits par un expérimentateur.

Ces longs ARN double-brin sont clivés par Dicer, une endonucléase de la famille RNase III (Bernstein *et al.*, 2001 ; Knight et Bass, 2001). Les Vertébrés et le Nématode ne possèdent qu'un gène *Dicer* chacun ; la Drosophile en possède deux (appelés *Dicer-1* et *Dicer-2*) : les siRNA sont produits par Dicer-2 (Lee *et al.*, 2004b). Chez les organismes modèles les mieux étudiés, Dicer interagit physiquement avec un partenaire protéique : RDE-4 chez *C. elegans* (Tabara *et al.*, 2002) ; R2D2, partenaire de Dicer-2, chez *Drosophila melanogaster* (Liu *et al.*, 2003) ; et les protéines TRBP (Chendrimada *et al.*, 2005 ; Haase *et al.*, 2005) et PACT (Lee *et al.*, 2006) chez les Mammifères. Toutes ces protéines sont homologues entre elles, et possèdent des domaines dsRBD (*double-stranded RNA-binding domain*, domaine de liaison à l'ARN double-brin). Il est donc probable que ces protéines jouent un rôle similaire : deux d'entre elles, au moins, stabilisent la protéine Dicer (R2D2 : Liu *et al.*, 2003, et PACT : Lee *et al.*, 2006)³ ; certaines n'ont pas d'effet sur l'activité endonucléolytique de Dicer (R2D2 : Liu *et al.*, 2003 ; et TRBP : Chendrimada *et al.*, 2005) alors que d'autres semblent nécessaires à la production de siRNA (RDE-4 : au moins quand l'ARN double-brin déclencheur n'est pas surexprimé : Parrish et Fire, 2001 ; Habig *et al.*, 2008 ; et PACT : Lee *et al.*, 2006). Mais ces deux dernières observations pourraient simplement découler d'une déstabilisation de Dicer en l'absence de son partenaire protéique (il est possible que les molécules de Dicer qui persistent en l'absence de RDE-4 ou PACT soient aussi actives qu'en leur présence — elles seraient simplement moins nombreuses).

Les siRNA ont les caractéristiques typiques des produits de clivage de la famille RNase III : ils présentent des extrémités sortantes en 3' longues de 2 nt, leurs extrémités 5' sont phosphorylées et leurs extrémités 3' sont hydroxyles (Elbashir *et al.*, 2001b) — ils peuvent toutefois subir une modification chimique ultérieurement (voir ci-dessous). Ces duplex sont ensuite chargés sur une protéine de la famille Argonaute (voir sous-section 1.2.3, page 14, pour une description de ces protéines), par une machinerie de chargement spécialisée. Cette réaction a été le mieux décrite chez la Drosophile : les siRNA sont chargés sur la protéine Ago2 par un hétérodimère constitué des protéines Dicer-2 et R2D2 (Tomari *et al.*, 2004). Cet hétérodimère (appelé RLC, *RISC-loading complex*, complexe de chargement de RISC) se fixe de manière asymétrique sur le duplex (R2D2 interagit principalement avec l'extrémité la plus stablement appariée, alors que Dicer-2 interagit avec l'autre extrémité du duplex), et charge le siRNA sur la protéine Ago2.

Bien entendu, les deux brins d'ARN occupent des positions différentes dans la protéine Ago2 : l'extrémité 5' du brin guide se niche dans une cavité du domaine Piwi de la protéine, et son extrémité 3' interagit avec le domaine Paz (voir sous-section 1.2.3, page 14, pour la description de ces domaines) ; c'est la position du petit ARN guide dans le complexe RISC mature. Le brin passager, quant à lui, occupe la position des futurs ARN-cibles de ce guide (son extrémité 3' fait face à l'extrémité 5' du guide, et vice versa). Le brin passager est ensuite clivé par une activité endonucléolytique portée par Ago2 (l'activité qui lui permet aussi de détruire les futurs ARN-cibles), et les deux fragments de clivage sont dégradés (Matranga *et al.*, 2005 ; Rand *et al.*, 2005). Finalement, c'est donc le complexe RLC qui, en orientant le duplex d'ARN, détermine la position des deux brins d'ARN dans la protéine Argonaute, donc leur destinée : le brin dont l'extrémité 5' interagissait avec Dicer-2 sera le brin guide, qui restera stablement associé à la protéine Argonaute et la guidera vers les ARN qui lui sont complémentaires ; le brin dont

³Chendrimada *et al.*, 2005 ont également proposé que TRBP stabilisait Dicer chez les Mammifères, mais Haase *et al.*, 2005 signalent que le protocole utilisé ne permet pas de conclure sans ambiguïté — et ils n'observent, quant à eux, pas d'effet de TRBP sur la stabilité de Dicer.

l'extrémité 5' interagissait avec R2D2 sera le brin passager, la première molécule de cible pour le complexe RISC, détruite après le chargement du duplex.

Dicer-2 remplit donc deux rôles distincts dans la biogenèse des siRNA de *Drosophila* : d'une part, elle les génère (par digestion endonucléolytique de l'ARN double-brin déclencheur), puis d'autre part, elle participe (avec R2D2) à leur chargement sur Ago2, déterminant ainsi quel brin du duplex servira de guide au complexe RISC. Il est d'ailleurs possible de découpler ces deux fonctions génétiquement : la mutation G31R de Dicer-2 (une substitution dans le domaine d'interaction à l'ATP) abolit l'activité de génération des siRNA, sans modifier l'activité de chargement de Dicer-2 (Lee *et al.*, 2004b).

Chez la *Drosophila*, le brin guide, débarrassé de son passager, est ensuite méthylé sur l'oxygène en 2' de son dernier nucléotide (Pélisson *et al.*, 2007 ; Horwich *et al.*, 2007). Cette modification chimique avait échappé aux premières analyses, pourtant destinées à trouver des modifications chimiques aux extrémités des siRNA (Elbashir *et al.*, 2001b). L'enzyme responsable de cette modification est la 2'-O-méthylase Hen1 (Horwich *et al.*, 2007). On ignore encore la nature chimique exacte de l'extrémité 3' des siRNA chez les autres Animaux : les Vertébrés expriment une protéine Hen1, mais on ne peut pas en conclure qu'elle modifie les siRNA (on lui connaît déjà une autre fonction potentielle, celle de méthyler une autre classe de petits ARN : Kirino et Mourelatos, 2007) ; quant au Nématode, il ne possède pas d'orthologue évident pour Hen1.

siRNA endogènes chez le Nématode

Une autre classe de petits ARN, de nature complètement différente, est appelée « siRNA secondaires ». Ces petits ARN simple-brin, observés chez le Nématode au cours de la réponse RNAi, sont polymérisés par une RdRP (*RNA-dependent RNA polymerase*, polymérase à ARN dépendant de l'ARN), sans amorce, et dont la matrice est un ARN_m ciblé par le RNAi (Pak et Fire, 2007 ; Sijen *et al.*, 2007 ; Aoki *et al.*, 2007). En tant que transcrits primaires, ces petits ARN portent un triphosphate sur leur extrémité 5' ; leur extrémité 3' porte un 2'-OH et un 3'-OH. Ils sont spécifiquement antisens aux ARN_m ciblés par les siRNA « primaires » (qui dérivent du double-brin déclencheur), et ils sont complémentaires d'une région du voisinage de la région ciblée par les siRNA primaires. Pour la plupart des siRNA secondaires (93 % d'après Aoki *et al.*, 2007), le premier nucléotide est une guanosine, ce qui pourrait indiquer une préférence de séquence de la RdRP.

Les siRNA secondaires sont nettement plus abondants que les siRNA primaires (environ 200 fois d'après Pak et Fire, 2007), et ils sont responsables de la plus grande part de la répression de la cible (Sijen *et al.*, 2007 ; Aoki *et al.*, 2007). Ils interagissent avec des protéines Argonaute spécifiques, CSR-1 (Aoki *et al.*, 2007), SAGO-1 et SAGO-2 (Yigit *et al.*, 2006) (les siRNA primaires interagissent essentiellement avec la protéine Argonaute RDE-1 : Yigit *et al.*, 2006).

La réponse RNAi se déroule donc en plusieurs étapes chez le Nématode : les siRNA primaires, issus du clivage de l'ARN double-brin déclencheur par Dicer, ne sont finalement pas directement responsables de la destruction de l'ARN-cible. Ils ne déstabilisent pas, ou très peu, l'ARN-cible (qui est suffisamment stable après avoir été reconnu par le siRNA primaire, pour servir de matrice à la polymérisation de siRNA secondaires : certains siRNA secondaires chevauchent même le site du clivage guidé par le siRNA primaire : Sijen *et al.*, 2007). L'activité endonucléolytique de RDE-1, Argonaute partenaire des siRNA primaires, est même dispensable à la répression de la cible, pour peu que le brin passager des siRNA primaires soit dégradé par un autre mécanisme (Steiner *et al.*, 2009). La production de siRNA secondaires, complémentaires à l'ARN cible, amplifie la répression, et ce sont les siRNA secondaires qui guident la destruction de la cible ⁴.

⁴Certaines observations plus anciennes restent toutefois mystérieuses : en particulier, Tijsterman *et al.*, 2002 ont mis en évidence un mode de répression, spécifique de certains gènes dans les gonades, après injection de

1.2.3 Les protéines Argonaute, partenaires des petits ARN

La famille protéique Argonaute (parfois appelée PPD, pour : *Paz and Piwi domain proteins*, protéines aux domaines Paz et Piwi) est caractérisée par les domaines Paz (qui interagit avec les ARN simple-brin et les extrémités 3' sortantes des produits de clivage par les RNase III : Song *et al.*, 2003) et Piwi (apparenté structurellement au domaine RNase H : Song *et al.*, 2004 ; il est toujours situé en aval du domaines Paz).

Elle regroupe deux sous-familles :

- les protéines Ago (qui possèdent en outre un troisième domaine reconnaissable : DUF1785 (*domain of unknown function* n°1785 ; sa fonction éventuelle est inconnue), en amont du domaines Paz (voir figure 1.1, panneau A) ;
- les protéines Piwi.

Il faut se méfier de la nomenclature de cette famille et de ses sous-familles : de nombreux auteurs appellent également « Argonaute » la sous-famille Ago (et la famille complète est donc homonyme à l'une de ses sous-familles). J'ai choisi d'adopter la convention de Carmell *et al.*, 2002, qui baptise cette sous-famille « Ago » ; cette convention est peu répandue, mais elle a le mérite d'éviter les confusions.

La nomenclature de ces protéines recèle une autre difficulté : le nom de « Piwi » a été donné à trois entités différentes ; il est important de distinguer la sous-famille Piwi (voir ci-dessus) de la protéine Piwi (l'un des trois membres de cette sous-famille chez la Drosophile ; c'est elle qui lui a donné son nom), du domaine protéique Piwi. Ce domaine, donc, n'est pas spécifique de la sous-famille homonyme, puisqu'il est partagé avec la sous-famille Ago ...

Le domaine Paz est également retrouvé dans les protéines Dicer, mais dans aucune autre protéine connue, ce qui (avec l'apport d'expériences de biochimie sur l'affinité de ce domaine pour les acides nucléiques) suggère un rôle de ce domaine dans la reconnaissance spécifique des extrémités 3' des petits ARN régulateurs (Carmell et Hannon, 2004). Le domaine Piwi, quant à lui, se replie d'une manière similaire au domaine RNase H, ce qui suggérerait qu'il portait lui-même une activité RNase (Song *et al.*, 2004) — et effectivement, des mutations ponctuelles dans le domaine Piwi abolissent l'activité endonucléolytique de la protéine Ago2 de Souris, sans affecter son affinité pour l'ARN-guide (Liu *et al.*, 2004). Les deux positions mutées appartenaient à la triade catalytique (trois acides aminés à groupement carboxylate, soit des acides aspartiques soit des acides glutamiques) qui, dans le domaine RNase H, chélatent un cation métallique divalent qui participe au clivage du pont phosphodiester ciblé. Le domaine Piwi des Argonaute ne contient que deux de ces trois carboxylates (les deux premiers, deux acides aspartiques) : le troisième partenaire du cation métallique est une histidine (Rivas *et al.*, 2005). La triade catalytique des Argonaute (DDH, contre DDE chez les RNases H) n'est pas retrouvée dans tous les membres de la famille (ainsi, les protéines humaines Ago1 et Ago4 ne possèdent pas l'histidine), ce qui explique que toutes les Argonaute ne sont pas capables de cliver un ARN-cible.

Il faut noter que la plupart des structures protéiques disponibles pour ces domaines proviennent de l'analyse de protéines procaryotiques. En particulier, toutes les structures résolues de protéines entières (en association ou non avec un acide nucléique) concernent des protéines d'Eubactéries ou d'Archæ : la protéine Ago de *Pyrococcus furiosus* (Song *et al.*, 2004) ; la protéine Piwi (qui ne possède pas de domaine Paz, et qui a une meilleure affinité pour les petits ADN guides que pour les petits ARN guides) d'*Archaeoglobus fulgidus* (Parker *et al.*, 2004 ; Parker *et al.*, 2005 ; Ma *et al.*, 2005) ; la protéine Ago (qui, également, interagit plus stablement avec

petits ARN antisens ; ces ARN peuvent être relativement longs (40 nt), mais ils ne répriment pas leur cible si leur extrémité 3' porte une modification chimique ou si Dicer est mutée, ce qui semblait montrer que ces ARN servent d'amorces pour la polymérisation d'un long ARN double-brin, ensuite digéré par Dicer. À ma connaissance, ces observations n'ont toujours pas été expliquées.

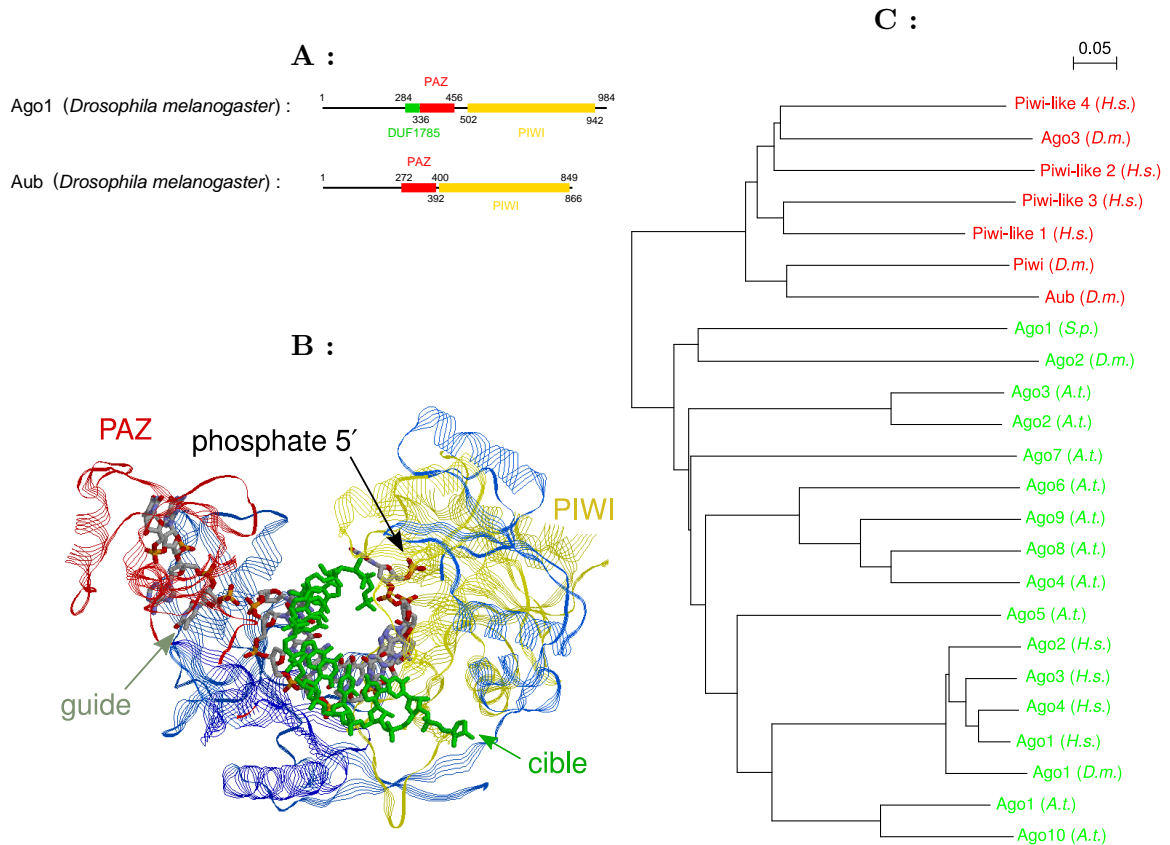


FIG. 1.1 – **Les protéines Argonaute.** **A :** Architecture d'une protéine Argonaute représentative (Ago1 de *D. melanogaster* et l'isoforme A d'Aubergine, Aub, de *D. melanogaster*). Les nombres en noir indiquent les positions des acides aminés qui délimitent les domaines. **B :** Structure de la protéine Argonaute de *Thermus thermophilus* complexée à un ADN-guide (représenté en gris, bleu, orange et rouge) et un ARN-cible (représenté en vert) (structure 3F73 de la Protein Data Bank ; décrite dans Wang *et al.*, 2008a). La protéine est représentée par les rubans rouge, bleu et jaune (en rouge : domaines Paz, acides aminés 176 à 276 ; en jaune : domaine Piwi, acides aminés 405 à 671). Le phosphate 5' de l'ADN-guide est indiqué. **C :** L'alignement des séquences des domaines Paz des protéines Argonaute distingue clairement les protéines Piwi (en rouge) des protéines Ago (en vert) chez quelques espèces représentatives (*H.s.* : *Homo sapiens* ; *D.m.* : *Drosophila melanogaster* ; *S.p.* : *Schizosaccharomyces pombe* ; *A.t.* : *Arabidopsis thaliana*). Alignement réalisé avec le programme `clustalw 2.0.9` ; la barre d'échelle indique la longueur de branche pour une fréquence de substitution de 0.05 par acide aminé.

les petits ADN guides qu'avec les petits ARN guides) d'*Aquifex aeolicus* (Yuan *et al.*, 2005) et la protéine Ago de *Thermus thermophilus*, qui n'a pu être cristallisée en complexe binaire qu'avec un ADN guide, pas avec un ARN guide (Wang *et al.*, 2008b). Chez les Eucaryotes, seuls les domaines Paz, isolés, des protéines Ago1 et Ago2 de Drosophile, et Ago1 humaine, ont été résolus, par RMN ou cristallographie, en association ou non avec un acide nucléique (Yan *et al.*, 2003 ; Song *et al.*, 2003 ; Lingel *et al.*, 2003 ; Lingel *et al.*, 2004 ; Ma *et al.*, 2004).

Or la fonction éventuelle de ces protéines Argonaute procaryotiques n'est pas connue, et rien n'indique qu'elles interagissent avec des petits ARN régulateurs *in vivo* (aucun de ces Procaryotes ne possède d'enzyme Dicer, et aucun petit ARN similaire aux petits ARN régulateurs eucaryotiques n'y a été décrit). L'affinité préférentielle pour les ADN, chez plusieurs de ces protéines, pourrait même indiquer qu'elles interagissent naturellement avec des ADN-guides

plutôt que des ARN-guides (elles pourraient alors remplir une fonction de RNase H *in vivo*).

Malgré ces réserves, il est couramment admis que les structures des protéines Argonaute eucaryotiques sont suffisamment semblables à celles des protéines cristallisées, et de nombreuses caractéristiques biochimiques des protéines Argonaute eucaryotiques peuvent être ainsi interprétées grâce à ces structures :

- l'enfouissement du phosphate 5' du brin guide dans une poche du domaine Piwi (voir panneau **B** de la figure 1.1, page 15) expliquerait à la fois l'importance de ce phosphate dans le positionnement du site de clivage de la cible, et, plus généralement, dans l'efficacité de clivage (voir Nykänen *et al.*, 2001 ; Schwarz *et al.*, 2002 ; Martinez *et al.*, 2002 pour les données de biochimie, et Parker *et al.*, 2005 ; Ma *et al.*, 2005 pour leur interprétation structurale) ;
- la distorsion du premier nucléotide du brin guide (qui n'est pas empilé avec les nucléotides suivants, donc incapable de s'apparier avec un nucléotide de la cible : voir panneau **B** de la figure 1.1) expliquerait pourquoi une complémentarité entre le premier nucléotide du guide et la cible n'améliore pas l'activité de RISC (voir Haley et Zamore, 2004 ; Lewis *et al.*, 2005 pour les données biochimiques et statistiques, et Parker *et al.*, 2005 ; Ma *et al.*, 2005 ; Wang *et al.*, 2008b pour leur interprétation structurale) ;
- la géométrie du complexe ribonucléoprotéique expliquerait pourquoi la liaison phosphodiester clivée est celle qui relie les nucléotides qui font face aux nucléotides 10 et 11 du guide (voir Elbashir *et al.*, 2001b ; Elbashir *et al.*, 2001c ; Schwarz *et al.*, 2004 pour les données de biologie moléculaire et de biochimie, et Parker *et al.*, 2005 ; Ma *et al.*, 2005 pour leur interprétation structurale) ;
- elle expliquerait aussi pourquoi la « graine » (nucléotides 2 à 7, ou 2 à 8, selon les auteurs) des microARN aurait une influence disproportionnée dans la reconnaissance des cibles (voir « Identification des cibles de microARN », page 25) (voir Lai, 2002 ; Lewis *et al.*, 2003 ; Mallory *et al.*, 2004 ; Haley et Zamore, 2004 ; Krek *et al.*, 2005 pour les données de génétique, de statistiques, de biologie moléculaire et de biochimie, et Ma *et al.*, 2005 ; Parker *et al.*, 2009 ; Wang *et al.*, 2008b pour leur interprétation structurale) ;

Les nombres de protéines Argonaute sont assez variables chez les Animaux : alors que la Drosophile exprime deux Ago et trois Piwi ⁵, la Souris exprime quatre Ago et trois Piwi, l'Homme, quatre Ago et quatre Piwi, et le Nématode, vingt-cinq Ago et deux Piwi ⁶. On connaît encore mal les relations entre ces nombreuses protéines : alors que, chez la Drosophile, les deux Ago (Okamura *et al.*, 2004) et les trois Piwi (voir article n°5, page 73 : Li *et al.*, 2009a) remplissent, chacune, une fonction qui lui est propre, des Argonaute de Vertébrés et de Nématodes pourraient être redondantes (Yigit *et al.*, 2006 ; Landthaler *et al.*, 2008).

1.2.4 Activité biologique des siRNA

À l'échelle moléculaire, les siRNA guident le clivage des ARN qui leur sont complémentaires. Pour que cette réaction ait lieu, deux conditions doivent être remplies :

- il faut que la protéine Argonaute associée à ce siRNA soit compétente pour la réaction de clivage (en particulier, elle doit contenir une triade catalytique fonctionnelle : voir sous-

⁵Les deux protéines Ago sont appelées Ago1 et Ago2, et les trois protéines Piwi sont appelées : Piwi, Aubergine et Ago3. Voilà donc une nouvelle difficulté dans la nomenclature des protéines Argonaute : la protéine « Ago3 » n'est pas une Ago, mais une Piwi ...

⁶Cependant, le statut exact de ces vingt-sept gènes (Yigit *et al.*, 2006) n'est pas connu avec certitude : certains sont des pseudo-gènes, dont rien n'indique qu'ils sont effectivement exprimés ; et, si seulement deux Argonaute (PRG-1 et PRG-2) du Nématode sont manifestement des orthologues des protéines Piwi (Das *et al.*, 2008 ; Batista *et al.*, 2008), une autre protéine du Nématode, ERGO-1, semble plus proche des Piwi que des Ago (Yigit *et al.*, 2006).

- section 1.2.3, page 14) ; chez les Mammifères par exemple, une seule des quatre protéines Ago possède un site actif (la protéine appelée Ago2 ; Meister *et al.*, 2004 ; Liu *et al.*, 2004) ;
- il faut que le petit ARN et sa cible soient suffisamment complémentaires (notamment autour du phosphate scissile, entre les nucléotides qui font face aux nucléotides 10 et 11 du guide : des irrégularités dans la double hélice autour du site de clivage inhibent la réaction ; Hutvagner et Zamore, 2002 ; Chiu et Rana, 2003).

Les deux produits de clivage sont ensuite, vraisemblablement, dégradés par des exonucléases ; il est cependant possible de détecter le fragment 3' par une technique très sensible, le 5'-RACE (*rapid amplification of cDNA ends*, amplification rapide des extrémité des cDNA) (voir par exemple Frank-Kamenetsky *et al.*, 2008).

Chez les Plantes et les Champignons, les siRNA peuvent également guider des modifications de la chromatine, mais les données sont beaucoup moins convaincantes chez les Animaux (voir Djupedal et Ekwall, 2009 pour une revue ; voir également la sous-section « 1.4.5 RNAi et guidage de l'hétérochromatinisation » de ma thèse ⁷). Il faut d'ailleurs souligner que l'étude qui, en 2004, montrait les effets les plus spectaculaires des siRNA sur la méthylation de l'ADN (Kawasaki et Taira, 2004) a depuis été rétractée (Taira, 2006). Les seules données réellement convaincantes qui subsistent montrent que la protéine humaine Ago1 est recrutée sur des promoteurs méthylés sous l'effet de siRNA, et qu'elle est nécessaire à leur répression (Kim *et al.*, 2006).

Chez le Nématode, des mutations qui perturbent la biogenèse ou l'action des siRNA mobilisent des transposons (Ketting *et al.*, 1999 ; Tabara *et al.*, 1999 ; Vastenhouw *et al.*, 2003). Ces observations suggèrent donc qu'une des fonctions biologiques du RNAi chez le Nématode, est la répression des transposons. Il faut toutefois noter que tous les gènes impliqués dans le RNAi participent pas à la répression des transposons, et réciproquement : il semble donc que seule une sous-classe de siRNA réprime les transposons, et que d'autres mécanismes, indépendants du RNAi, contribuent à la mise en silence des éléments génétiques mobiles.

À l'échelle macroscopique, le RNAi participe à la défense antivirale chez la Drosophile (Wang *et al.*, 2006 ; Galiana-Arnoux *et al.*, 2006 ; van Rij *et al.*, 2006). Les mutants *dcr-2* et *ago2* sont plus sensibles que les mouches sauvages à l'infection par les virus FHV (*flock house virus*), CrPV (*cricket paralysis virus*), DCV (*Drosophila C virus*), et SINV (*Sindbis virus*), des virus à ARN (dont la réplication implique un intermédiaire ARN double-brin). La sensibilité des mutants *dcr-2* et *ago2* est spécifique de ces virus : la résistance aux autres pathogènes (bactéries : Wang *et al.*, 2006 ; Galiana-Arnoux *et al.*, 2006 et champignons : Galiana-Arnoux *et al.*, 2006) n'est pas affectée par ces mutations.

Lorsqu'elle a été mesurée, l'accumulation des ARN viraux était toujours plus élevée chez les mutants *dcr-2* et *ago2* que chez les sauvages, ce qui signifie probablement que les ARN double-brin viraux sont clivés en siRNA par Dicer-2, et que ces siRNA, chargés sur Ago2, guident la destruction des ARN simple-brin viraux. Les trois premiers de ces virus, au moins, expriment des inhibiteurs de RNAi qui leur permettent d'échapper partiellement à cette réponse immunitaire : l'inhibiteur exprimé par le FHV (Lu *et al.*, 2005) et celui exprimé par le DCV (van Rij *et al.*, 2006) se fixent sur l'ARN double-brin viral, empêchant Dicer de le cliver en siRNA ; un autre virus, le NoV (*Nodamura virus*), qui infecte insectes et mammifères, exprime également un inhibiteur qui se fixe aux ARN double-brin et bloque l'action de Dicer (Sullivan et Ganem, 2005).

⁷Disponible sur

<http://tel.archives-ouvertes.fr/docs/00/04/74/70/PDF/tel-00007781.pdf>

et sur

<http://www.normalesup.org/~seitz/These.pdf>.

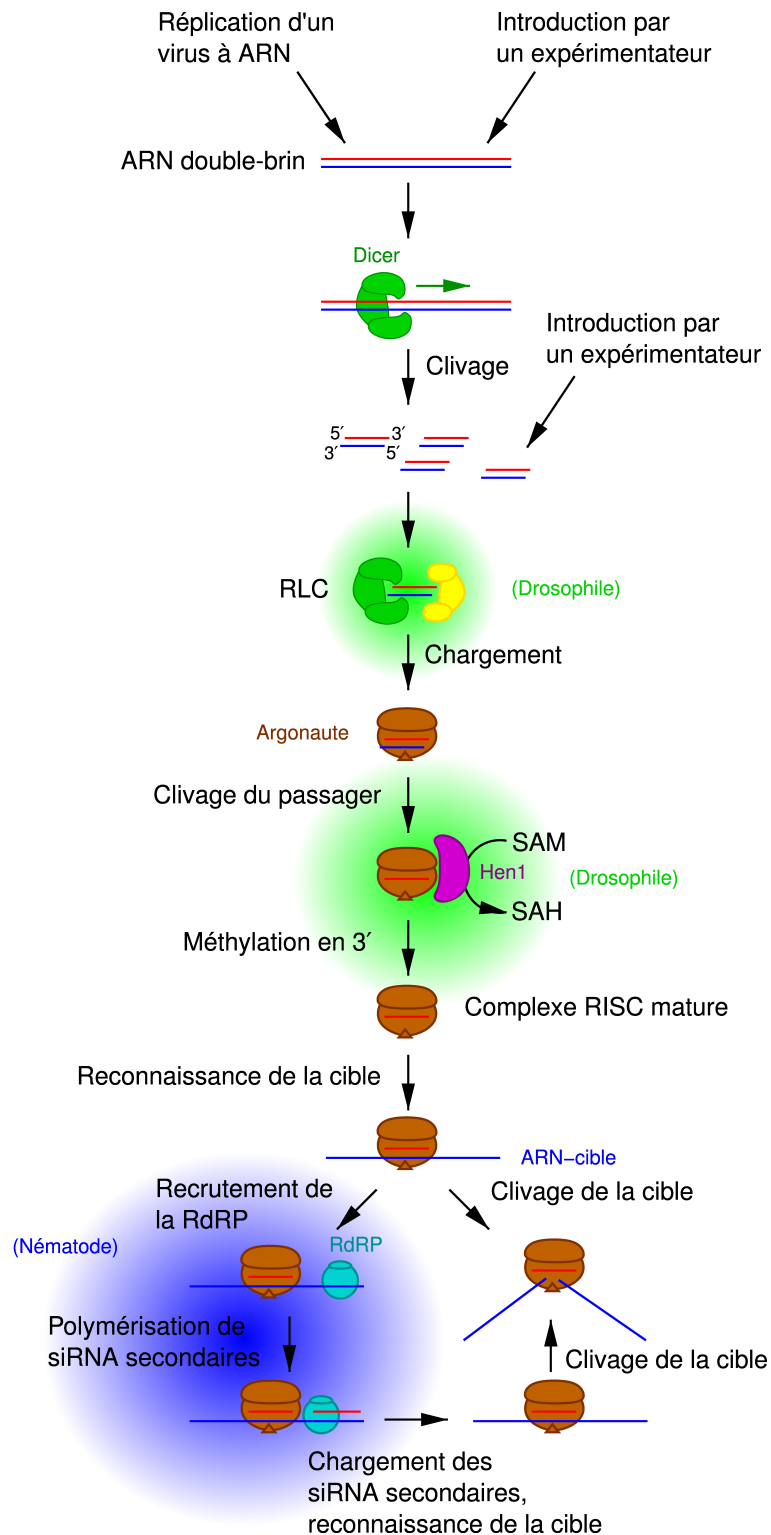


FIG. 1.2 – Le RNAi chez les Animaux. Certaines étapes n'ont été mises en évidence que chez une seule espèce : la Drosophile (disques dégradés verts) ou le Nématode (disque dégradé bleu). Voir texte pour le détail.

1.3 Les microARN, des régulateurs post-transcriptionnels endogènes

1.3.1 Présentation générale

Les microARN (miRNA) sont des petits ARN simple-brin, d'environ 22 nt (à 1 nt près en général). Ils sont transcrits sous la forme de précurseurs plus longs (appelés « pri-miRNA »), qui se replient en tiges-boucles imparfaites : chez les Animaux, cette tige-boucle est excisée du pri-miRNA dans le noyau par une endonucléase appelée « Drosha » (Lee *et al.*, 2003), et exportée dans le cytoplasme (Lee *et al.*, 2002). La tige-boucle (appelée « pre-miRNA ») y est clivée par une autre endonucléase, Dicer⁸, qui libère un petit duplex imparfaitement apparié, et la boucle apicale (voir figure 1.3).

L'un des deux brins du duplex est le miRNA ; l'autre est appelé « miRNA* ». Ce duplex est chargé sur une protéine de la sous-famille Ago de la famille Argonaute (voir page 14 pour une discussion sur cette nomenclature), mais un seul des deux brins reste stablement associé à la protéine, et s'accumule *in vivo* : le brin miRNA. L'autre brin est dissocié, et il est probablement rapidement dégradé par des nucléases cellulaires (voir l'article n°4, page 58, pour une démonstration expérimentale de la chronologie de ces événements).

Tout comme Dicer, Drosha appartient à la famille des RNase III. Elle clive donc des ARN double-brin, et génère des produits aux extrémités 3' sortantes, de deux nucléotides en général (voir figure 1.3). Le produit de clivage 3' porte un phosphate sur son extrémité 5', et le produit 5' porte une extrémité 3' hydroxyle. Finalement, après les clivages par Drosha et Dicer, le miRNA et le miRNA* sont, biochimiquement, très semblables : ce sont des petits ARN de taille similaire, phosphorylés en 5', et hydroxylés en 3'. À première vue, la seule différence entre ces deux petits ARN est leur abondance : par convention, on a appelé « miRNA » le brin qui s'accumule le plus, et « miRNA* », le brin qui s'accumule le moins (le miRNA* n'a d'ailleurs pas toujours été détecté).

Le miRNA mature guide la protéine Argonaute vers les ARN qui lui sont parfaitement ou imparfaitement complémentaires ; ces ARN-cibles sont généralement réprimés (voir sous-section 1.3.3, page 22, pour une discussion des mécanismes de cette répression), mais des données récentes suggèrent que, dans des circonstances bien particulières, les miRNA pourraient activer la traduction des ARN-cibles (voir « Régulation de l'activité des microARN », page 25).

Les règles de reconnaissance des cibles de miRNA ne sont pas encore complètement élucidées ; plusieurs programmes informatiques ont néanmoins été écrits, pour prédire les cibles de miRNA ; la plupart de ces programmes recherchent des complémentarités parfaites à la « graine » (nt 2 à 7, ou 2 à 8, selon les auteurs) du miRNA dans les 3' UTR des ARN messagers, et sélectionnent celles qui sont conservées phylogénétiquement. Ces programmes identifient en général des dizaines ou des centaines de cibles pour chaque miRNA (voir « Identification des cibles de microARN », page 24, pour une description détaillée de ces programmes et de leurs prédictions) ; selon le dogme le plus couramment admis, les miRNA régulent donc une grande part des gènes codants chez les Animaux (les ≈ 700 miRNA humains connus réguleraient ainsi 60 % des gènes codants humains : Friedman *et al.*, 2009). Ainsi, quel que soit le processus physiologique que l'on considère, au moins un de ses composants appartiendra certainement à ces 60 % — en d'autres termes : il semble que les miRNA contrôlent tous les processus physiologiques chez les Animaux.

⁸Chez la Drosophile, qui possède deux enzymes Dicer, cette réaction est catalysée par Dicer-1 (Lee *et al.*, 2004b), en association avec une protéine à domaines dsRBD, appelée Loquacious (Förstemann *et al.*, 2005 ; Saito *et al.*, 2005) ou R3D1 (Jiang *et al.*, 2005), selon les auteurs.

1.3.2 Biogenèse des microARN

Les pri-miRNA semblent transcrits par l'ARN polymérase II (« Pol II ») (Lee *et al.*, 2004a). Borchert *et al.*, 2006 ont proposé que certains pri-miRNA humains étaient transcrits par la Pol III, mais il s'agissait visiblement d'une erreur (Bortolin-Cavaillé *et al.*, 2009). Étant donné que plus de 80 % des miRNA humains sont hébergés dans des introns de gènes-hôtes, et au moins 35 % sont hébergés dans des introns de gènes-hôtes codants (Kim et Kim, 2007), le pri-miRNA est fréquemment aussi un ARN pré-messager.

La plupart des pre-miRNA sont excisés de leur pri-miRNA par Drosha, en association avec une protéine à dsRBD (*double-stranded RNA binding domain*, domaine de liaison à l'ARN double-brin), appelée DGCR8 chez les Vertébrés (Gregory *et al.*, 2004 ; Han *et al.*, 2004 ; Landthaler *et al.*, 2004) et Pasha chez la Drosophile (Denli *et al.*, 2004). Ce complexe protéique est appelé le « Microprocessor » ; on notera que les deux RNases III impliquées dans la biogenèse des petits ARN, Drosha et Dicer (voir sous-section 1.2.2, page 12), fonctionnent toutes les deux en association avec une protéine à dsRBD ; dans les deux cas, la RNase III est responsable de l'activité endonucléolytique, et le partenaire augmente l'affinité du complexe pour son substrat. DGCR8/Pasha joue au moins un autre rôle : en se fixant sur la base de la tige-boucle, elle impose la position du site de clivage par Drosha (dans la tige, à environ 11 bp de sa base) (Han *et al.*, 2006). En contraignant de la sorte la position du clivage, DGCR8/Pasha est donc responsable de la fidélité du clivage par Drosha.

Quelques pre-miRNA ne sont pas excisés par Drosha : ce sont des pre-miRNA introniques, libérés de leur pri-miRNA par la réaction d'épissage et le débranchement du lasso de l'intron (l'intron est alors très précisément le pre-miRNA) (Ruby *et al.*, 2007 ; Okamura *et al.*, 2007 ; Berezikov *et al.*, 2007) ; les miRNA issus de cette voie de maturation sont appelés « mirtrons ».

Le Microprocessor (comme la machinerie d'épissage) agit très précocément sur le pri-miRNA : il est recruté par le transcrit naissant, encore associé à la chromatine, et il libère les pre-miRNA introniques avant que leur intron ne soit épissé (Morlando *et al.*, 2008). L'excision du pre-miRNA n'inhibe toutefois pas l'épissage de l'intron-hôte : l'intron clivé est correctement épissé après l'action du Microprocessor, ce qui signifie probablement que les extrémités de l'intron ont été correctement reconnues, et qu'elles ont été engagées dans les premières étapes de la réaction d'épissage, au moment où l'intron est clivé par Drosha (Kim et Kim, 2007).

Le pre-miRNA est ensuite exporté dans le cytoplasme par l'exportine-5 (Yi *et al.*, 2003 ; Lund *et al.*, 2004 ; Bohnsack *et al.*, 2004). Il y est clivé par la RNase III Dicer (l'unique Dicer, chez les Vertébrés et le Nématode : Grishok *et al.*, 2001 ; Hutvagner *et al.*, 2001 ; Ketting *et al.*, 2001 ; l'enzyme Dicer-1, chez la Drosophile : Lee *et al.*, 2004b). Elle libère un petit duplex d'ARN, dont les deux brins sont longs d'environ 22 nt, l'extrémité 5' est phosphorylée et l'extrémité 3' est hydroxyle. Elle libère aussi, théoriquement, un ARN simple brin issu de la boucle du pre-miRNA, et dont les extrémités doivent avoir la même nature chimique : ces ARN de boucle sont probablement rapidement dégradés, mais les plus abondants sont désormais détectables par les nouvelles techniques de séquençage à haut débit (voir par exemple les données supplémentaires de Ghildiyal *et al.*, 2008 : dans un échantillon d'ARN de têtes de drosophiles, nous avons détecté 569 molécules de l'ARN de boucle de pre-miR-34, pour 188 574 molécules de miRNA et 2 101 molécules de miRNA* ; échantillon n^oGSM278707 de la banque de données GEO : <http://www.ncbi.nlm.nih.gov/projects/geo/>).

Le duplex miRNA/miRNA* est alors chargé sur une protéine de la sous-famille Ago. Les deux brins de ce duplex sont ensuite dissociés (voir article n^o4, page 58, pour les caractéristiques biochimiques de ce phénomène sur la protéine Ago1 de Drosophile), et le brin passager (le brin miRNA* pour la majorité des duplex) est perdu, et probablement rapidement dégradé par des nucléases cellulaires. De la même manière que pour les siRNA (voir sous-section 1.2.2, page 12), le brin du duplex miRNA/miRNA* préférentiellement choisi comme brin guide est habituellement

celui qui est le moins stablement apparié sur son extrémité 5' (Schwarz *et al.*, 2003).

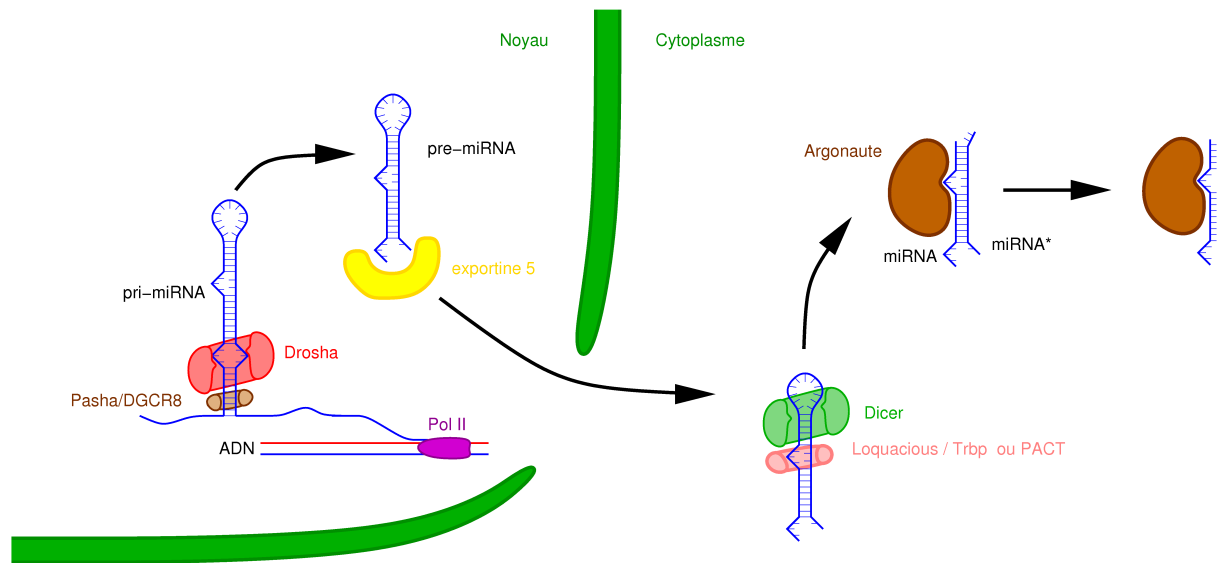


FIG. 1.3 – La biogenèse des microARN chez les Animaux. Voir texte pour le détail.

Chargement sur Argonaute : le triage des petits ARN

Chez la *Drosophile*, la plupart des miRNA sont principalement chargés sur Ago1 (Okamura *et al.*, 2004), mais quelques-uns sont chargés sur Ago2, le partenaire habituel des siRNA (Förstemann *et al.*, 2007). C'est la structure secondaire du duplex entre le miRNA et le miRNA* qui détermine l'identité de la protéine Ago sur laquelle il sera chargé : un duplex dont les positions centrales (autour du neuvième nucléotide du brin guide) sont mésappariées sera principalement chargé sur Ago1, alors qu'il sera surtout chargé sur Ago2 si ces positions sont appariées (Tomari *et al.*, 2007). La plupart des duplex miRNA/miRNA* de *Drosophile*, mésappariés sur ces positions, sont donc chargés sur Ago1.

À la différence de la machinerie de chargement de la protéine Ago2 (voir sous-section 1.2.2, page 12), la machinerie de chargement d'Ago1 de *Drosophile* est encore inconnue. Mais il est déjà clair que cette machinerie a une préférence pour les duplex mésappariés en positions centrales, de même que l'hétérodimère R2D2/Dicer-2 (machinerie de chargement d'Ago2) a une préférence pour les duplex appariés sur ces positions : le choix de la protéine Ago de destination est déterminé par les affinités relatives des deux machineries de chargement pour le duplex, sachant que chacune reconnaît la géométrie rejetée par l'autre, et qu'aucune n'est choisie par défaut, sans préférence de structure (Tomari *et al.*, 2007).

Chez le Nématode, un mécanisme similaire semble diriger les duplex de siRNA (appariés sur toute leur longueur) vers RDE-1, et les duplex miRNA/miRNA* (imparfaitement appariés) vers ALG-1 (Steiner *et al.*, 2007 ; Jannot *et al.*, 2008). Chez l'Homme, les machineries de chargement des quatre protéines Ago ont visiblement des préférences très similaires, et les miRNA chargés sur chacune sont souvent identiques (Landthaler *et al.*, 2004) ; quelques miRNA sont cependant chargés plus fréquemment sur certaines Ago (Azuma-Mukai *et al.*, 2008).

L'identité de la protéine Ago sur laquelle un miRNA est chargé influencera son activité biologique :

- chez la *Drosophile*, Ago1 (qui possède la triade catalytique capable de cliver un ARN-cible ; voir sous-section 1.2.4, page 16) est en réalité une très mauvaise nucléase, dont la lenteur

semble peu compatible avec une véritable activité *in vivo*, alors qu'Ago2 est capable de cliver ses cibles dans des conditions physiologiques (Förstemann *et al.*, 2007) ; ces deux protéines sont également capables de réprimer la traduction de leurs ARN_m-cibles, mais par des mécanismes différents (Iwasaki *et al.*, 2009 ; voir sous-section 1.3.3, ci-dessous) ;

- chez le Nématode, les miRNA chargés sur RDE-1 guideront le clivage des ARN-cibles suffisamment complémentaires, alors que les miRNA chargés sur ALG-1 en seront incapables (Steiner *et al.*, 2007) ;
- chez les Mammifères, seule Ago2 peut cliver ses ARN-cibles (voir sous-section 1.2.4, page 16).

1.3.3 Modes d'action des microARN

Une diversité de mécanismes

Lorsqu'il est associé à une protéine Ago compétente pour la réaction de clivage (voir ci-dessus) et suffisamment complémentaire à son ARN-cible (en particulier autour du phosphate scissile, comme pour les siRNA : voir sous-section 1.2.4, page 16), un miRNA guide le clivage de l'ARN-cible. Ces conditions sont rarement remplies chez les Animaux, et on n'y connaît que deux cibles de miRNA dont l'ARN_m est clivé : *HOXB8* (Yekta *et al.*, 2004) et *Rtl1* (Davis *et al.*, 2005) chez les Mammifères.

Les autres cibles connues sont réprimées par un autre mécanisme ; on a longtemps considéré qu'elles étaient réprimées uniquement traductionnellement (ce qui signifierait que l'accumulation des l'ARN_m-cible n'est pas affectée, seulement sa faculté à être traduit en protéine) (Bartel, 2004). Des données plus récentes ont cependant montré que l'ARN_m-cible pouvait également être dégradé (par un mécanisme exonucléolytique, distinct du clivage endonucléolytique des ARN_m de *HOXB8* et *Rtl1*), dans des cellules humaines en culture (Lim *et al.*, 2005 ; Wu *et al.*, 2006 ; Schmitter *et al.*, 2006), chez le Nématode (Bagga *et al.*, 2005), la Drosophile (Behm-Ansmant *et al.*, 2006) et le Poisson-Zèbre (Giraldez *et al.*, 2006). Dans ces deux derniers cas en particulier, et dans les cellules humaines d'après Wu *et al.*, 2006, la déstabilisation de l'ARN_m semble due à la perte de sa queue poly(A), provoquée par les miRNA. Cette déstabilisation n'est toutefois pas toujours suffisante pour expliquer la diminution de production de protéine, et il existe donc bien un mécanisme de répression traductionnelle ; la répression se distribue donc en deux composantes, déstabilisation de l'ARN_m et inhibition de sa traduction, dont l'ampleur relative dépend au moins de la séquence et de la structure du milieu du duplex entre le petit ARN et sa cible (Alemán *et al.*, 2007 ; cette étude a utilisé un système artificiel, où le petit ARN est synthétique, et transfecté dans des cellules en culture).

La déstabilisation de l'ARN_m-cible a été confirmée depuis, dans de nombreux modèles expérimentaux ; elle a été mise à profit dans plusieurs analyses, qui ont recherché les ARN_m dont le niveau d'accumulation était influencé par un miRNA (il est en effet plus facile de quantifier un grand nombre d'ARN, par puce à ADN, qu'un grand nombre de protéines) (voir par exemple Linsley *et al.*, 2007 ; Grimson *et al.*, 2007). Cette méthode présente l'inconvénient de n'identifier que les ARN_m dont l'accumulation est affectée par le miRNA, et de négliger ceux qui sont réprimés uniquement au niveau traductionnel.

La recherche du mécanisme de la répression traductionnelle a suscité de nombreuses études, dont les résultats étaient parfois contradictoires :

- Humphreys *et al.*, 2005 ont montré que la répression d'une cible artificielle nécessite une coiffe sur son extrémité 5' (une cible traduite à partir d'un site interne d'entrée du ribosome, IRES, et dépourvue de coiffe, est insensible au miRNA). De la même manière, une cible privée de queue poly(A) est moins sensible qu'une cible polyadénylée. Les auteurs en concluent donc que le complexe RISC inhibe l'initiation de la traduction dépendant des facteurs d'initiation eIF4E (recruté par la coiffe lors de l'initiation de la traduction) et eIF4G (recruté par eIF4E et une protéine qui reconnaît la queue poly(A), la PABP).

- Pillai *et al.*, 2005 proposent également que les miRNA répriment l'initiation de la traduction (des ARN_m traduits à partir d'une IRES, donc de manière indépendante de la coiffe, sont insensibles à la répression), mais ils ne rapportent aucun effet de la queue poly(A).
- De manière plus précise, Kiriakidou *et al.*, 2007 ont montré que la protéine Ago2 humaine pouvait interagir avec un analogue de la coiffe des ARN_m (le m⁷GTP), et qu'un double mutant ponctuel (qui n'interagit pas avec cet analogue) ne pouvait pas réprimer une cible artificielle. Ces observations suggéraient que les protéines Ago pourraient entrer en compétition avec eIF4E pour la fixation à la coiffe de l'ARN_m, et donc, inhiber l'initiation de la traduction.
- Kinch et Grishin, 2009 affirment cependant (en se basant sur les structures cristallographiques de protéines Argonaute archéobactériennes) que le site d'interaction avec la coiffe proposé par Kiriakidou *et al.*, 2007 ne peut pas fixer une coiffe d'ARN_m, et que les mutations ponctuelles utilisées dans leur étude pourraient modifier profondément la structure d'Ago2, et donc en inhiber l'activité de manière non spécifique.
- Wakiyama *et al.*, 2007 ont mis au point un système *in vitro* qui récapitule la répression guidée par les miRNA (à partir d'extraits de cellules humaines en culture qui surexpriment plusieurs protéines impliquées dans la biogenèse ou le mode d'action des miRNA). La répression guidée par un miRNA, dans ce système, est sensible à l'absence de la coiffe, et très sensible à l'absence de queue poly(A) ; la fixation du complexe RISC entraîne de plus une déadénylation de l'ARN_m-cible, et les auteurs suggèrent que la perte de queue poly(A) est responsable de la perte de traductibilité de l'ARN_m-cible (en plus d'un éventuel effet direct du complexe RISC sur la traduction).
- Alors que Chendrimada *et al.*, 2007 ont étudié la même protéine Ago que Kiriakidou *et al.*, 2007 (la protéine Ago2 humaine), ils lui ont trouvé une autre fonction : elle participe à un complexe multiprotéique qui contient également eIF6, un facteur connu pour inhiber le recrutement de la grande sous-unité ribosomique sur la petite. Ces auteurs proposent donc que la protéine Ago réprime la traduction de ses cibles en empêchant l'assemblage d'un ribosome fonctionnel sur le codon d'initiation. Ils montrent, en outre, que la protéine eIF6 est nécessaire à la répression guidée par un miRNA, dans des cellules humaines en culture et *in vivo* chez le Nématode, et suggèrent donc un rôle universel de eIF6 dans la répression des cibles de miRNA.
- Ces conclusions sont toutefois contredites par Eulalio *et al.*, 2008, qui montrent que eIF6 est dispensable à la répression traductionnelle chez la Drosophile. Ils mettent également en doute le mécanisme proposé par Kiriakidou *et al.*, 2007 (voir plus haut), puisque la protéine Ago1 de Drosophile reste active quand le site présumé d'interaction avec la coiffe est partiellement muté — et effectivement, l'affinité de mutants de Ago1 pour l'analogue de coiffe m⁷GTP ne corrèle pas avec leur activité de répression. En revanche, les mutations ponctuelles qui abolissent l'activité d'Ago1 empêchent également son association avec GW182, une protéine nécessaire à la répression guidée par les miRNA (Behm-Ansmant *et al.*, 2006) : cette association elle-même est nécessaire à la répression, et il semble donc, finalement, qu'Ago1 ne soit qu'un intermédiaire qui recrute GW182 (le véritable répresseur) sur les ARN_m-cibles.
- L'autre protéine Ago de Drosophile, Ago2, est également capable de réprimer la traduction de ses ARN_m-cibles, mais par un mécanisme différent : à la différence d'Ago1, elle ne nécessite pas GW182, elle ne provoque pas de déadénylation de la cible, et elle ne réprime que la traduction dépendante de la coiffe ⁹ (Iwasaki *et al.*, 2009). Ces auteurs montrent

⁹Le mécanisme de cette répression apparaît cependant différent de celui proposé par Kiriakidou *et al.*, 2007 pour Ago2 humaine : Ago2 de Drosophile n'est pas en compétition avec eIF4E pour la reconnaissance de la coiffe. Elle semble plutôt entrer en compétition avec eIF4G pour la fixation à eIF4E : Ago2 interagit physiquement avec eIF4E, et son affinité pour ce facteur augmente fortement quand Ago2 se lie à un ARN-cible, ce qui suggère que son domaine d'interaction est inaccessible dans la protéine Ago2 libre (qui serait alors incapable de titrer

également qu'Ago1 réprime ses cibles par deux mécanismes : en provoquant la dégradation de la queue poly(A), ce qui inhibe le recrutement des facteurs d'initiation sur la coiffe ; mais également à une étape ultérieure, qui reste à identifier.

Enfin, quelques auteurs ont observé l'association du complexe RISC avec des polysomes (donc : avec des ARN_m qui sont visiblement activement traduits) ; cette observation les a amenés à proposer un autre modèle : RISC n'inhiberait pas l'initiation de la traduction, ni même son élongation, mais il provoquerait la dégradation co-traductionnelle des peptides naissants (Olsen et Ambros, 1999 ; Nelson *et al.*, 2004 ; Maroney *et al.*, 2006 ; Nottrott *et al.*, 2006). Ce modèle est très controversé (l'activité protéolytique postulée n'a jamais été observée, et les spécialistes ne s'accordent pas sur l'interprétation des différences, parfois minimes, entre gradients de polysomes).

Identification des cibles de microARN

L'identification des cibles de miRNA a également fait l'objet de nombreuses études, dès 2003. Quelques stratégies expérimentales ont été mises au point ces dernières années, mais la plupart des efforts se sont concentrés sur la prédiction de cibles par des moyens informatiques. On est pourtant loin de comprendre les règles de reconnaissance des cibles par les miRNA (et on en savait encore moins en 2003), si bien que les algorithmes de prédiction reposent parfois sur des hypothèses discutables ; les programmes les plus anciens utilisaient d'ailleurs des critères dont on sait, maintenant, qu'ils étaient fantaisistes :

- la stabilité prédite de l'interaction entre le miRNA et l'ARN-cible (il est en effet possible de prédire l'enthalpie libre de repliement d'une séquence d'ARN, par des programmes informatiques qui se basent sur des règles empiriques) : Enright *et al.*, 2003 ; les analyses statistiques des sites de complémentarité conservés aux miRNA montrent maintenant, au contraire, que les séquences-cibles les plus crédibles sont riches en A/U (dont les appariements sont moins stables que les appariements des régions riches en G/C) (Grimson *et al.*, 2007) ;
- la conservation phylogénétique des complémentarités aux miRNA : Enright *et al.*, 2003 ; Lewis *et al.*, 2003 ; Stark *et al.*, 2003 ; il a, depuis, été démontré que les sites non conservés sont également fonctionnels (Grimson *et al.*, 2007 ; Kim et Bartel, 2009) ;
- la recherche de sites uniquement dans les 3' UTR des ARN_m : Enright *et al.*, 2003 ; Lewis *et al.*, 2003 ; Stark *et al.*, 2003 ; Kiriakidou *et al.*, 2004 ; John *et al.*, 2004 ; Lewis *et al.*, 2005 ; Krek *et al.*, 2005 (ainsi que d'autres programmes, d'usage moins répandu) ; on sait maintenant que des petits ARN synthétiques (Saxena *et al.*, 2003) et des miRNA endogènes (Kloosterman *et al.*, 2004 ; Duursma *et al.*, 2008 ; Forman *et al.*, 2008) peuvent réprimer des ARN_m -cibles en s'appariant imparfaitement à leur séquence codante ou leur 5' UTR.

Deux programmes récents utilisent des stratégies différentes :

- le programme rna22 de Miranda *et al.*, 2006 n'utilise ni la stabilité prédite de l'appariement entre le miRNA et sa cible, ni la conservation phylogénétique de l'interaction entre les deux ARN, et il peut analyser les 5' UTR et les séquences codantes : il recherche des motifs (éventuellement dégénérés, de longueur variable, et de position variable dans le miRNA) partagés par des miRNA connus, puis recherche des concentrations de segments complémentaires à ces motifs dans les séquences-cibles ;
- le programme PITA de Kertesz *et al.*, 2007 tient compte d'un critère négligé par les autres programmes : l'accessibilité de la séquence-cible (qui est probablement mal reconnue par le miRNA si elle participe à une structure secondaire stable de l' ARN_m) ; mais les au-

non-spécifiquement le eIF4E cellulaire) et qu'il se découvre quand Ago2 est recrutée sur une cible.

teurs n'ont analysé que les 3' UTR, et l'ampleur de sa tâche (il doit prédire la structure secondaire de toutes les cibles potentielles) rend son exécution très lente.

Cependant, des comparaisons récentes ont montré que ces deux programmes prédisent très mal quels ARN_m sont réprimés par des miRNA (Selbach *et al.*, 2008 ; Baek *et al.*, 2008). D'après ces études, les programmes qui prédisent le mieux quels ARN_m sont réprimés sont PicTar (Selbach *et al.*, 2008) et TargetScan (Baek *et al.*, 2008)¹⁰.

Ces deux programmes, PicTar et TargetScan, sont probablement les plus utilisés par la communauté. Leurs algorithmes sont très similaires : ils recherchent les 3' UTR qui présentent une complémentarité parfaite à la graine (nucléotides 2 à 7, ou 2 à 8, selon les auteurs) du miRNA, et qui soit phylogénétiquement conservée (Krek *et al.*, 2005 ; Lewis *et al.*, 2005) ; PicTar tolère aussi, sous certaines conditions, des mésappariements à la graine (Krek *et al.*, 2005). Le programme TargetScan a été révisé plus récemment : il prend également en compte la richesse locale de la 3' UTR en nucléotides A et U, la proximité de sites de reconnaissance d'autres miRNA, l'appariement du miRNA en-dehors de la graine, et la position du site de complémentarité dans la 3' UTR ; des graines couvrant les nucléotides 3 à 8 sont désormais tolérées, et la conservation phylogénétique de la complémentarité peut être négligée (Grimson *et al.*, 2007 ; Friedman *et al.*, 2009).

Ces dernières années, plusieurs laboratoires ont cherché à identifier les cibles de miRNA par des approches expérimentales (voir table 1.1, page 26). La plupart de ces travaux ont montré que les miRNA répriment l'expression de nombreux gènes-cibles (quelques dizaines ou centaines par miRNA), et ils ont tous montré que les 3' UTR des ARN_m de ces cibles sont enrichies en séquences complémentaires à la graine des miRNA testés.

Enfin, il est apparu que des motifs de séquence dans la 3' UTR pouvaient affecter l'efficacité des sites de reconnaissance des miRNA : chez le Nématode, la 3' UTR de l' ARN_m de *cog-1* possède deux sites de complémentarité à la graine du miRNA *lsy-6*. Chacun de ces sites est nécessaire à la répression, mais des séquences supplémentaires sont également requises : les 25 nt immédiatement en aval du deuxième site de reconnaissance de *lsy-6* (qui ne semblent reconnus par aucun miRNA), et la vingtaine de nucléotides qui séparent les deux sites de reconnaissance de *lsy-6* (Didiano et Hobert, 2008). La séquence de 25 nt peut être déplacée dans la 3' UTR, et son activité est peu affectée du moment qu'elle est placée dans un contexte riche en A et U — mais une mutation de sa séquence réduit son activité ; le séparateur d'une vingtaine de nucléotides, quant à lui, peut être muté, mais il ne peut pas être trop raccourci.

Ces résultats montrent que la simple présence d'une séquence complémentaire à la graine d'un miRNA ne préjuge pas de l'activité du miRNA sur un ARN_m : des éléments de séquence supplémentaires peuvent moduler son activité.

Régulation de l'activité des microARN

Vasudevan *et al.*, 2007 ont montré que le miRNA humain miR-369-3p active la traduction de l' ARN_m TNF α dans des cellules en culture, mais uniquement lorsqu'elles sont privées de sérum. Cet ARN_m présente deux sites de complémentarité à la graine de miR-369-3p, qui sont requises pour cette activation, et le recrutement d'Ago2 sur un ARN_m rapporteur de manière indépendante des miRNA¹¹, ou guidée par un autre miRNA, suffit à activer la traduction

¹⁰Il faut toutefois remarquer que ces comparaisons ont été réalisées par les équipes qui avaient écrit les programmes PicTar et TargetScan ; et, curieusement, alors qu'ils utilisent pratiquement le même test pour leurs comparaisons, les auteurs de PicTar estiment que PicTar est le plus performant des deux, alors que les auteurs de TargetScan estiment que TargetScan est le meilleur ...

¹¹Ago2 est fusionnée au peptide λN , et l' ARN_m rapporteur contient cinq motifs « B-box », qui interagissent avec λN .

Principe	Étude	Description
Immunoprécipitation de RISC	Easow <i>et al.</i> , 2007	Immunoprécipitation des complexes associés à une protéine Ago1 étiquetée avec un épitope HA, dans des cellules de <i>Drosophila</i> en culture ; analyse des ARN _m immunoprécipités par puce à ADN.
	Karginov <i>et al.</i> , 2007	Immunoprécipitation des complexes associés à une protéine Ago2 étiquetée avec un épitope c-myc, dans des cellules humaines en culture ; analyse des ARN _m immunoprécipités par puce à ADN.
	Zhang <i>et al.</i> , 2007	Immunoprécipitation des complexes associés aux protéines GW182 (l'une étiquetée avec un épitope GFP, l'autre native), dans un lysat de nématodes ; analyse des ARN _m immunoprécipités par puce à ADN et séquençage à haut débit.
	Beitzinger <i>et al.</i> , 2007	Immunoprécipitation des complexes associés aux protéines Ago1 et Ago2 natives, dans des cellules humaines en culture ; identification des ARN _m immunoprécipités, par clonage.
	Hendrickson <i>et al.</i> , 2008	Immunoprécipitation des complexes associés à une protéine Ago2 étiquetée avec un épitope FLAG, dans des cellules humaines en culture ; analyse des ARN _m immunoprécipités par puce à ADN.
	Landthaler <i>et al.</i> , 2008	Immunoprécipitation des complexes associés aux protéines Ago étiquetées avec les épitopes FLAG et HA, dans des cellules humaines en culture ; analyse des ARN _m immunoprécipités par puce à ADN.
Quantification de l'accumulation de protéines en fonction de l'accumulation de miRNA	Nakahara <i>et al.</i> , 2005	Quantification de 1003 protéines par gel bidimensionnel et identification par spectrométrie de masse, dans des ovocytes de <i>Drosophila</i> sauvages et mutants pour Dicer.
	Vinther <i>et al.</i> , 2006	Quantification de 504 protéines par spectrométrie de masse, en surexpression de miRNA dans des cellules humaines en culture.
	Selbach <i>et al.</i> , 2008	Quantification de plus de 3000 protéines par spectrométrie de masse, en présence, absence ou surexpression de miRNA, dans des cellules humaines en culture.
	Baek <i>et al.</i> , 2008	Quantification de 1544 à 3819 protéines par spectrométrie de masse, en présence, absence ou surexpression de miRNA, dans des cellules humaines en culture ou dans des cellules primaires de Souris.

TAB. 1.1 – **Recherches de cibles de microARN par des approches expérimentales.** Outre les analyses par puce à ADN, qui identifient les ARN_m dont l'accumulation est réduite sous l'effet d'un miRNA (voir « Une diversité de mécanismes », page 22), plusieurs stratégies expérimentales ont été développées pour identifier des cibles de miRNA.

du rapporteur, pourvu que les cellules soient privées de sérum. Les auteurs proposent donc que les miRNA puissent se comporter en activateurs ou en inhibiteurs selon les conditions de croissance des cellules : seul l'effet inhibiteur aurait été observé jusqu'alors, parce que les autres expérimentateurs utilisent des cellules en phase de croissance.

Ces résultats, qui n'ont pas encore été reproduits par d'autres laboratoires, pourraient signifier que l'activité des miRNA *in vivo* est régulée par les conditions de croissance des cellules.

Des facteurs *trans* peuvent aussi affecter l'efficacité des miRNA :

- chez l'Homme, la protéine HuR (connue pour interagir avec les ARE, *AU-rich elements*, éléments riches en A et U dans les 3' UTR) peut déréguler une cible de miRNA, très certainement en chassant le miRNA de la 3' UTR en s'y appariant (Bhattacharyya *et al.*, 2006) ;
- chez l'Homme et le Poisson-Zèbre, la protéine « Dead end 1 » (Dnd1), en se fixant sur la 3' UTR d'ARN_m ciblés par des miRNA, peut les masquer et empêcher leur interaction avec RISC (Kedde *et al.*, 2007).

Chapitre 2

Des siRNA endogènes chez la Drosophile ?

2.1 Contexte : quel rôle endogène pour le RNAi ?

Le RNAi a été remarquablement bien conservé au cours de l'évolution : observé chez la plupart des Plantes, des Animaux, et de nombreux Champignons, il semble avoir été préservé depuis la divergence des trois règnes, il y a plus d'un milliard d'années¹. Une telle conservation indique vraisemblablement que le RNAi remplit une fonction biologique, et que cette fonction a été conservée.

Pourtant, parmi les Animaux, on ne lui connaissait de fonction que chez la Drosophile et le Nématode (voir Introduction, sous-section 1.2.4, page 16 : le RNAi participe à la défense antivirale chez la Drosophile, et à la répression des transposons chez le Nématode). Ces deux fonctions ne semblent pas conservées chez les Mammifères : les ARN double-brin (en particulier, ceux d'origine virale) y déclenchent la réponse interféron, une répression traductionnelle non-spécifique, qui rendrait même inutile le RNAi dans la défense antivirale (Fire, 1999) ; quant à la répression des transposons par RNAi endogène, observée chez le Nématode (voir sous-section 1.2.4, page 16), elle n'a pas été confirmée chez les autres Animaux. Ces considérations suggèrent que le RNAi pourrait jouer d'autres rôles biologiques, qui, eux, seraient conservés chez les Mammifères.

On avait cru découvrir cette fonction en 2001, quand des petits ARN, similaires aux siRNA, ont été mis en évidence chez la Drosophile (Aravin *et al.*, 2001). Ces petits ARN endogènes provenaient de séquences répétées dans le génome de la Drosophile (le gène *suppressor of Stellate*, *Su(Ste)*, répété en tandem une cinquantaine de fois sur le chromosome Y), et ils avaient donc été baptisés « rasiRNAs » (*repeat-associated siRNAs*, siRNA associés aux répétitions) : les auteurs pensaient avoir découvert des siRNA endogènes, qui proviendraient de ces séquences répétées.

Si effectivement il existait des siRNA endogènes, et si leur rôle biologique était conservé parmi les Animaux, alors la conservation du RNAi chez les Mammifères serait plus facile à expliquer : des siRNA endogènes, une nouvelle classe de régulateurs chez les Animaux, justifieraient la conservation du RNAi.

Mais notre ré-analyse de ces petits ARN a montré que les rasiRNA ne sont pas des siRNA : à la différence d'authentiques siRNA, ils ne proviennent pas du clivage processif, par Dicer, de longs ARN double-brin, et ils ne sont pas chargés sur les protéines de la sous-famille Ago. Ils sont chargés sur les protéines de la sous-famille Piwi, dont la fonction moléculaire était restée mystérieuse jusqu'alors (voir Introduction, sous-section 1.2.3, page 14, pour une description de ces sous-familles).

Ces résultats sont présentés dans l'article qui suit (Vagin *et al.*, 2006).

¹La datation de ces divergences fait l'objet de controverses (Roger et Hug, 2006) : d'après Hedges *et al.*, 2004, Animaux et Champignons auraient divergé il y a $\approx 1,5$ milliard d'années, et Plantes et (Animaux et Champignons) auraient divergé il y a $\approx 1,6$ milliards d'années. ; d'après Douzery *et al.*, 2004, Animaux et Champignons auraient divergé il y a ≈ 1 milliard d'années, et Plantes et (Animaux et Champignons) auraient divergé il y a $\approx 1,1$ milliard d'années.

2.2 Article n^o1 : une nouvelle classe de petits ARN

RESEARCH ARTICLES

A Distinct Small RNA Pathway Silences Selfish Genetic Elements in the Germline

Vasily V. Vagin,^{1,2*} Alla Sigova,^{1*} Chengjian Li,¹ Hervé Seitz,¹ Vladimir Gvozdev,² Phillip D. Zamore^{1†}

In the *Drosophila* germline, repeat-associated small interfering RNAs (rasiRNAs) ensure genomic stability by silencing endogenous selfish genetic elements such as retrotransposons and repetitive sequences. Whereas small interfering RNAs (siRNAs) derive from both the sense and antisense strands of their double-stranded RNA precursors, rasiRNAs arise mainly from the antisense strand. rasiRNA production appears not to require Dicer-1, which makes microRNAs (miRNAs), or Dicer-2, which makes siRNAs, and rasiRNAs lack the 2',3' hydroxy termini characteristic of animal siRNA and miRNA. Unlike siRNAs and miRNAs, rasiRNAs function through the Piwi, rather than the Ago, Argonaute protein subfamily. Our data suggest that rasiRNAs protect the fly germline through a silencing mechanism distinct from both the miRNA and RNA interference pathways.

In plants and animals, RNA silencing pathways defend against viruses (1–3), regulate endogenous gene expression (4), and protect the genome against selfish genetic elements such as retrotransposons and repetitive sequences (5). Common to all RNA silencing pathways are RNAs 19 to 30 nucleotides (nt) long that specify the target RNAs to be repressed. In RNA interference (RNAi) (6), siRNAs are produced from long exogenous double-stranded RNA (dsRNA). In contrast, ~22-nt miRNAs are endonucleolytically processed from endogenous RNA polymerase II transcripts. Dicer ribonuclease III (RNase III) enzymes produce both siRNAs and miRNAs. In flies, Dicer-2 (Dcr-2) generates siRNAs, whereas the Dicer-1 (Dcr-1)–Loquacious (Loqs) complex produces miRNAs (7–11). After their production, small silencing RNAs bind Argonaute proteins to form the functional RNA silencing effector complexes believed to mediate all RNA silencing processes.

Phased sense and antisense siRNAs in vivo.

In *Drosophila*, processive dicing of long dsRNA and the accumulation of sense and antisense siRNAs without reference to the orientation of the target mRNA are hallmarks of RNAi in vitro (12, 13) and in vivo (Fig. 1). We prepared total small RNA from the heads of adult males expressing a dsRNA hairpin (fig. S1A) that silences the *white* gene via the RNAi pathway (14). *white* silencing requires Dcr-2 (7), R2D2 (9), and Ago2. siRNAs were detected with a microarray containing T_M (melting temperature)–normalized probes, 22 nt long, for all sense

and antisense siRNAs that theoretically can be produced by dicing the *white* exon 3 hairpin (Fig. 1A). Both sense and antisense *white* siRNAs were detected in wild-type flies but not in *dcr-2^{LS118X}* homozygous mutant flies. The Dcr-2–dependent siRNAs were produced with a periodicity of ~22 nt (Fig. 1B), consistent with the phased processing of the dsRNA hairpin from the end formed by the 6-nt loop predicted to remain after splicing of its intron-containing primary transcript (fig. S1B).

***Su(Ste)* rasiRNAs.** *Drosophila* repeat-associated small interfering RNAs (rasiRNAs) can be distinguished from siRNAs by their longer length, 24 to 29 nt (15, 16). rasiRNAs have been proposed to

be diced from long dsRNA triggers (5, 16), such as the ~50 copies of the bidirectionally transcribed *Suppressor of Stellate* [*Su(Ste)*] locus on the Y chromosome (17) (fig. S2B) that in testes silences the ~200 copies of the protein-coding gene *Stellate* (*Ste*) found on the X chromosome.

Microarray analysis of total small RNA isolated from fly testes revealed that *Su(Ste)* rasiRNAs detectably accumulate only from the antisense strand (Fig. 2A), with little or no phasing (fig. S2A). As expected, *Su(Ste)* rasiRNAs were not detected in testes from males lacking the *Su(Ste)* loci (*cry^Y*) (Fig. 2A). *Su(Ste)* rasiRNAs were also absent from *armitage* (*armi*) mutant testes (Fig. 2A), which fail to silence *Ste* and do not support RNAi in vitro (18). *armi* encodes a non-DEAD-box helicase (19) homologous to the *Arabidopsis thaliana* protein SDE3, which is required for RNA silencing triggered by transgenes and some viruses (20), and depletion by RNAi of the mammalian *Armi* homolog *Mov10* blocks siRNA-directed RNAi in cultured human cells (21). Normal accumulation of *Su(Ste)* rasiRNA and robust *Ste* silencing also require the putative helicase Spindle-E (*Spn-E*), a member of the DExH family of adenosine triphosphatases (ATPases) (16, 22) (Fig. 2B and fig. S2B).

The accumulation in vivo of only antisense rasiRNAs from *Su(Ste)* implies that sense *Su(Ste)* rasiRNAs either are not produced or are selectively destroyed. Either process would make *Ste* silencing mechanistically different from RNAi. In support of this view, mutations in the central components of the *Drosophila* RNAi pathway—*dcr-2*, *r2d2*, and *ago2*—did not diminish *Su(Ste)* rasiRNA accumulation (Fig. 2B). Deletion of the *Su(Ste)* silencing trigger (*cry^Y*) caused a factor

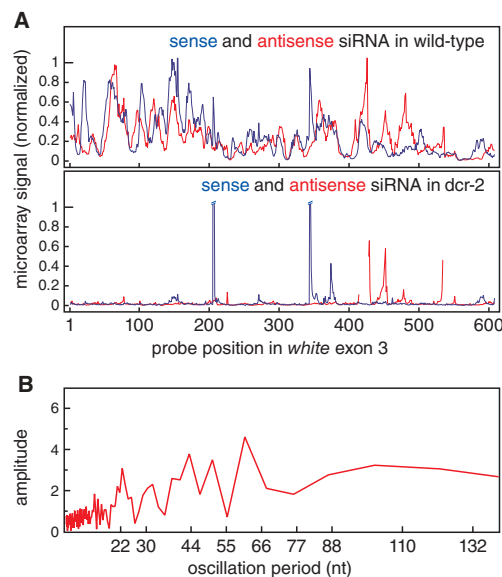


Fig. 1. Sense and antisense siRNAs accumulate during RNAi in vivo. (A) Microarray analysis of the siRNAs derived from the *white* exon 3 hairpin RNAi trigger. (B) Fourier transform analysis of the Dcr-2–dependent siRNAs in (A).

¹Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA 01605, USA. ²Department of Animal Molecular Genetics, Institute of Molecular Genetics, Moscow 123182, Russia.

*These authors contributed equally to this work.

†To whom correspondence should be addressed. E-mail: phillip.zamore@umassmed.edu

of ~65 increase in *Ste* mRNA (Fig. 3A), but null or strong hypomorphic mutations in the three key RNAi proteins did not (Fig. 3B).

Fly Argonaute proteins can be subdivided into the Ago (Ago1 and Ago2) and Piwi [Aubergine (Aub), Piwi, and Ago3] subfamilies. Unlike *ago1* and *ago2*, the *aub*, *piwi*, and *ago3* mRNAs are enriched in the germline (23–25). *Aub* is required for *Ste* silencing (16) and *Su(Ste)* rasiRNA accumulation (26). In *aub^{HN2}/aub^{OC42}* trans-heterozygous mutants, *Su(Ste)* rasiRNAs were not detected by microarray (fig. S2B) or Northern analysis (Fig. 2B), and *Su(Ste)*-triggered silencing of *Ste* mRNA was lost completely (Fig. 3B). Even *aub^{HN2}*+ heterozygotes accumulated less of the most abundant *Su(Ste)* rasiRNA than did the wild type (Fig. 2B). That the Ago subfamily protein Ago2 is not required for *Ste* silencing, whereas the Piwi subfamily protein *Aub* is essential for it, supports the view that *Ste* is silenced by a pathway distinct from RNAi. Intriguingly, *Su(Ste)* rasiRNAs hyperaccumulated in *piwi* mutant testes, where *Ste* is silenced normally (Figs. 2B and 3B and fig. S2B).

Mutations in *aub* also cause an increase in sense, but not antisense, *Su(Ste)* RNA (16); our results suggest that antisense *Su(Ste)* rasiRNAs can silence both *Ste* mRNA and sense *Su(Ste)* RNA, but that no *Su(Ste)* rasiRNAs exist that can target the antisense *Su(Ste)* transcript. Our finding that *Su(Ste)* rasiRNAs are predominantly or exclusively antisense is essentially in agreement with the results of small RNA cloning experiments, in which four of five *Su(Ste)* rasiRNAs sequenced were in the antisense orientation (15), but is at odds with earlier reports detecting both sense and antisense *Su(Ste)* rasiRNAs by non-quantitative Northern hybridization (16).

A third RNA silencing pathway in flies. Is germline RNA silencing of selfish genetic elements generally distinct from the RNAi and miRNA pathways? We examined the expression of a panel of germline-expressed selfish genetic elements—three long terminal repeat (LTR)-containing retrotransposons (*roo*, *mdg1*, and

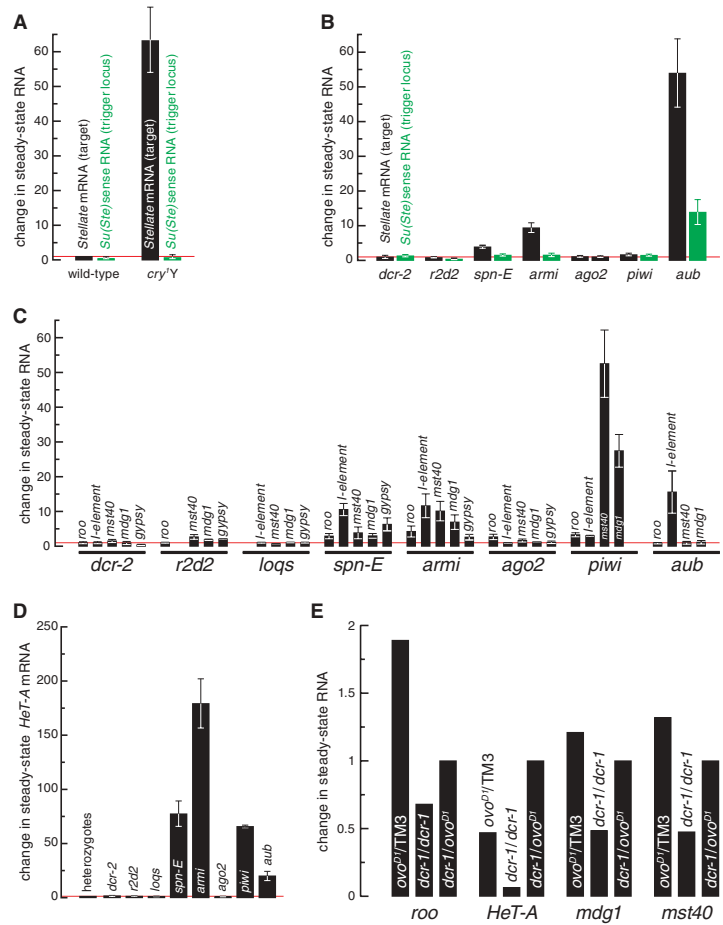


Fig. 3. (A to D) RNA expression from selfish genetic elements was measured in homozygous mutants relative to heterozygotes for *Ste* silencing in testes [(A) and (B)] and for the repeated locus *mst40*, the LTR retrotransposons *roo*, *mdg1*, and *gypsy*, and the non-LTR retrotransposons *l-element* and *HeT-A* in ovaries [(C) and (D)]. (E) RNA expression from selfish genetic elements in *dcr-1^{Q1147X}* null mutant clones generated by mitotic recombination in the ovary.

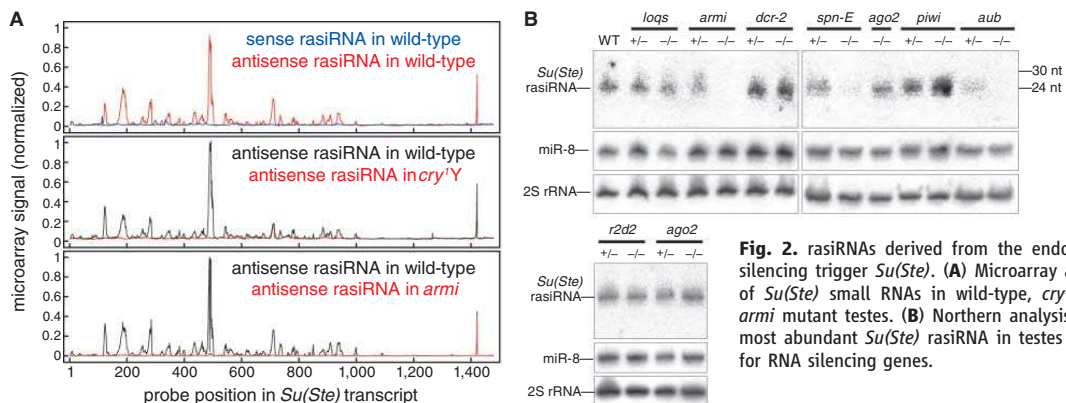


Fig. 2. rasiRNAs derived from the endogenous silencing trigger *Su(Ste)*. (A) Microarray analysis of *Su(Ste)* small RNAs in wild-type, *cry^{1Y}*, and *armi* mutant testes. (B) Northern analysis of the most abundant *Su(Ste)* rasiRNA in testes mutant for RNA silencing genes.

RESEARCH ARTICLES

gypsy), two non-LTR retrotransposons (*I-element* and *HeT-A*, a component of the *Drosophila* telomere), and a repetitive locus (*mst40*)—in mutants defective for eight RNA silencing proteins. All selfish genetic elements tested behaved like *Ste*: Loss of the RNAi proteins Dcr-2, R2D2, or Ago2 had little or no effect on retrotransposon or repetitive element silencing (Fig. 3, C and D). Instead, silencing required the putative helicases Spn-E and Armi plus one or both of the Piwi subfamily Argonaute proteins, Aub and Piwi. Silencing did not require Loqs, the dsRNA-binding protein required to produce miRNAs (Fig. 3, C and D).

The null allele *dcr-1^{Q1147X}* is homozygous lethal, making it impossible to procure *dcr-1* mutant ovaries from *dcr-1^{Q1147X}/dcr-1^{Q1147X}* adult females (7). Therefore, we generated clones of *dcr-1^{Q1147X}/dcr-1^{Q1147X}* cells in the ovary by mitotic recombination in flies heterozygous for the dominant female-sterile mutation *ovo^{D1}* (27). We measured RNA levels, relative to *rp49* mRNA, for three retrotransposons (*roo*, *HeT-A*, and *mdg1*) and one repetitive sequence (*mst40*) in *dcr-1/dcr-1* recombinant ovary clones and in *ovo^{D1}/TM3* and *dcr-1/ovo^{D1}* nonrecombinant ovaries. The *ovo^{D1}* mutation blocks oogenesis at stage 4, after the onset of *HeT-A* (28, 29) and *roo* rasiRNA production. Retrotransposon or repetitive sequence transcript abundance was unaltered or decreased in *dcr-1/dcr-1* relative to *ovo^{D1}/TM3* and *dcr-1/ovo^{D1}* controls (Fig. 3E). We conclude that Dcr-1 is dispensable for silencing these selfish genetic elements in the *Drosophila* female germline.

roo is the most abundant LTR retrotransposon in flies. We analyzed *roo* silencing in the female germline with the use of microarrays con-

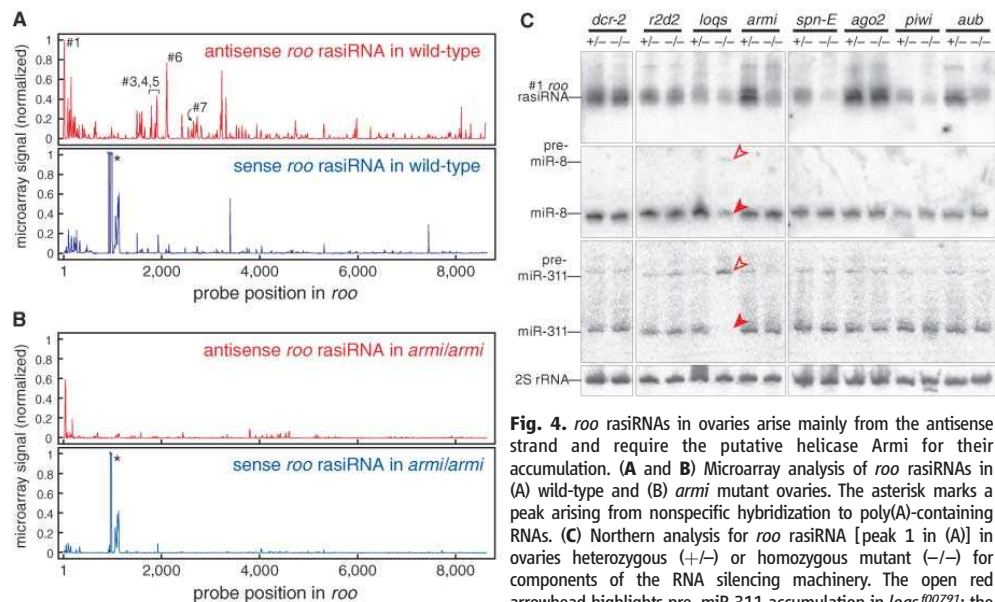
taining 30-nt probes, tiled at 5-nt resolution, for all ~18,000 possible *roo* rasiRNAs (Fig. 4, A and B); we corroborated the data at 1-nt resolution for those rasiRNAs derived from LTR sequences (fig. S3A). As observed for *Su(Ste)* but not for *white* RNAi, *roo* rasiRNAs were nonhomogeneously distributed along the *roo* sequence and accumulated primarily from the antisense strand (Fig. 4A and fig. S3A). In fact, the most abundant sense rasiRNA peak (asterisk in Fig. 4, A and B) corresponded to a set of probes containing 16 contiguous uracil residues, which suggests that these probes nonspecifically detected fragments of the mRNA polyadenylate [poly(A)] tail. Most of the remaining sense peaks were unaltered in *armi* mutant ovaries, in which *roo* expression is increased; this result implies that they do not contribute to *roo* silencing (Figs. 3C and 4A). We detected no phasing in the distribution of *roo* rasiRNAs (fig. S3B).

As for *Su(Ste)*, wild-type accumulation of antisense *roo* rasiRNA required the putative helicases Armi and Spn-E and the Piwi subfamily Argonaute proteins Piwi and Aub, but not the RNAi proteins Dcr-2, R2D2, and Ago2 (Fig. 4, B and C). Moreover, accumulation of *roo* rasiRNA was not measurably altered in *loqs^{f00791}*, an allele that strongly disrupts miRNA production in the female germline (Fig. 4C).

Are *roo* rasiRNAs not made by dicing? Loss of Dcr-2 or Dcr-1 did not increase retrotransposon or repetitive element expression, which suggests that neither enzyme acts in rasiRNA-directed silencing. Moreover, loss of Dcr-2 had no detectable effect on *Su(Ste)* rasiRNA in testes or *roo* rasiRNA in ovaries (Figs. 2B and 4C). We measured the amount of *roo* rasiRNA and miR-311

in *dcr-1/dcr-1* ovary clones generated by mitotic recombination. Comparison of recombinant (*dcr-1/dcr-1*) and nonrecombinant (*ovo^{D1}/TM3* and *dcr-1/ovo^{D1}*) ovaries by Northern analysis revealed that *roo* rasiRNA accumulation was unperturbed by the null *dcr-1^{Q1147X}* mutation (Fig. 5A and fig. S4). Pre-miR-311 increased and miR-311 declined by a factor of ~3 in the *dcr-1/dcr-1* clones (Fig. 5B and fig. S4), consistent with about two-thirds of the tissue corresponding to mitotic *dcr-1/dcr-1* recombinant cells. Yet, although most of the tissue lacked *dcr-1* function, we observed improved, rather than diminished, silencing for the four selfish genetic elements examined (Fig. 3E). Moreover, the dsRNA-binding protein Loqs, which acts with Dcr-1 to produce miRNAs, was also dispensable for *roo* rasiRNA production and selfish genetic element silencing (Fig. 3, C and D, and Fig. 4C). Although we cannot exclude the possibility that *dcr-1* and *dcr-2* can fully substitute for each other in the production of rasiRNA in the ovary, previous biochemical evidence suggests that none of the three RNase III enzymes in flies—Dcr-1, Dcr-2, and Drosha—can cleave long dsRNA into small RNAs 24 to 30 nt long (10, 30, 31).

Animal siRNA and miRNA contain 5' phosphate and 2',3' hydroxy termini (32, 33). We used enzymatic and chemical probing to infer the terminal structure of *roo* and *Su(Ste)* rasiRNAs. RNA from ovaries or testes was treated with calf intestinal phosphatase (CIP) or CIP followed by polynucleotide kinase plus ATP. CIP treatment caused *roo* (Fig. 6A) and *Su(Ste)* (fig. S5) rasiRNA to migrate more slowly in polyacrylamide gel electrophoresis, consistent with the loss



of one or more terminal phosphate groups. Subsequent incubation with polynucleotide kinase and ATP restored the original gel mobility of the rasiRNAs, indicating that they contained a single 5' or 3' phosphate before CIP treatment. The *roo* rasiRNA served as a substrate for ligation of a 23-nt 5' RNA adapter by T4 RNA ligase, a process that requires a 5' phosphate; pretreatment with CIP blocked ligation (Fig. 6B), thus establishing that the monophosphate lies at the 5' end. The rasiRNA must also contain at least one terminal hydroxyl group, because it could be joined by T4 RNA ligase to a preadenylated 17-nt 3' RNA adapter (Fig. 6B). Notably, the 3' ligation reaction was less efficient for the *roo* rasiRNA than for a miRNA in the same reaction (22% versus 50% conversion to ligated product).

RNA from ovaries or testes was reacted with NaIO_4 , then subjected to β -elimination, to determine whether the rasiRNA had either a single 2'

or 3' terminal hydroxy group or had terminal hydroxy groups at both the 2' and 3' positions, as do animal siRNA and miRNA. Only RNAs containing both 2' and 3' hydroxy groups react with NaIO_4 ; β -elimination shortens NaIO_4 -reacted RNA by one nucleotide, leaving a 3' monophosphate terminus, which adds one negative charge. Consequently, NaIO_4 -reacted, β -eliminated RNAs migrate faster in polyacrylamide gel electrophoresis than does the original unreacted RNA. Both *roo* (Fig. 6A) and *Su(Ste)* (fig. S5) rasiRNA lack either a 2' or a 3' hydroxyl group, because they failed to react with NaIO_4 ; miRNAs in the same samples reacted with NaIO_4 . Together, our results show that rasiRNAs contain one modified and one unmodified hydroxyl. Because T4 RNA ligase can make both 3'-5' and 2'-5' bonds (34), we cannot currently determine the blocked position. Some plant small silencing RNAs contain a 2'-O-methyl modification at their 3' terminus (34).

Fig. 5. (A) Northern analysis of mitotic recombinant *dcr-1^{Q1147X}* homozygous mutant ovaries and nonrecombinant controls. *roo* rasiRNAs were detected with a mixture of five hybridization probes. Arrowheads are as in Fig. 4C. **(B)** Quantification of the data in (A), normalized to the 2S rRNA loading control.

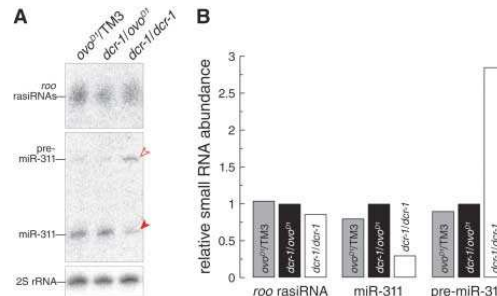
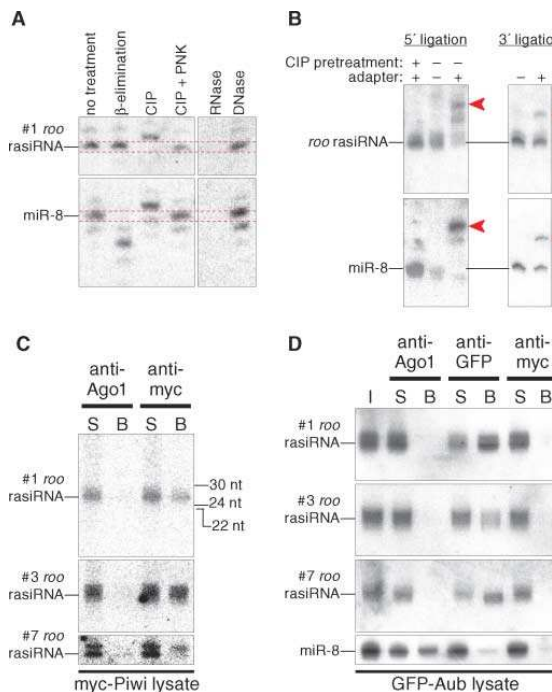


Fig. 6. *roo* rasiRNAs are modified at their 3' terminus and associate with Piwi subfamily Argonaute proteins. (A) Chemical and enzymatic probing of *roo* rasiRNA structure. *roo* rasiRNAs (peak 1 in Fig. 4A) were detected by Northern hybridization. The membrane was then stripped and reprobed for miR-8. **(B)** *roo* rasiRNAs can serve as a 3' or a 5' substrate for T4 RNA ligase. Solid arrowheads, 5' ligation products; open arrowheads, 3' ligation products. **(C and D)** *roo* rasiRNAs associate with myc-tagged Piwi (C) and GFP-tagged Aub protein (D), but not with Ago1. I, input; S, supernatant; B, bound.



rasiRNAs bind Piwi and Aub. *Drosophila* and mammalian siRNA and miRNA function through members of the Ago subfamily of Argonaute proteins, but *Su(Ste)* and *roo* rasiRNAs require at least one member of the Piwi subfamily for their function and accumulation. To determine whether *roo* rasiRNAs physically associate with Piwi and Aub, we prepared ovary lysate from wild-type flies or transgenic flies expressing either myc-tagged Piwi or green fluorescent protein (GFP)-tagged Aub protein (25, 35); immunoprecipitated them with monoclonal antibodies (mAbs) to myc, GFP, or Ago1; and then analyzed the supernatant and antibody-bound small RNAs by Northern blotting (Fig. 6, C and D, and fig. S6). We analyzed six different *roo* rasiRNAs. All were associated with Piwi but not with Ago1, the *Drosophila* Argonaute protein typically associated with miRNAs (36); miR-8 (Fig. 6D and fig. S6C), miR-311, and *bantam* immunoprecipitated with Ago1 mAb. No rasiRNAs immunoprecipitated with the myc mAb when we used lysate from flies lacking the myc-Piwi transgene (fig. S6B).

Although *aub* mutant ovaries silenced *roo* mRNA normally, they showed reduced accumulation of *roo* rasiRNA relative to *aub/+* heterozygotes (Fig. 4C), which suggests that *roo* rasiRNAs associate with both Piwi and Aub. We analyzed the supernatant and antibody-bound small RNAs after GFP-mAb immunoprecipitation of ovary lysate from GFP-Aub transgenic flies and flies lacking the transgene. *roo* rasiRNA was recovered only when the immunoprecipitation was performed with the GFP mAb in ovary lysate from GFP-Aub transgenic flies (Fig. 6D and fig. S6D). The simplest interpretation of our data is that *roo* rasiRNAs physically associate with both Piwi and Aub, although it remains possible that the *roo* rasiRNAs are loaded only into Piwi and that Aub associates with Piwi in a stable complex. The association of *roo* rasiRNA with both Piwi and Aub suggests that *piwi* and *aub* are partially redundant, as does the modest reduction in *roo* silencing—by a factor of 3.3 ± 0.6 (average \pm SD, $n = 3$)—in *piwi* but not in *aub* mutants (Fig. 3C). Alternatively, *roo* silencing might proceed through Piwi alone, but the two proteins could function in the same pathway to silence selfish genetic elements.

Our data suggest that in flies, rasiRNAs are produced by a mechanism that requires neither Dcr-1 nor Dcr-2, yet the patterns of rasiRNAs that direct *roo* and *Ste* silencing are as stereotyped as the distinctive siRNA population generated from the *white* hairpin by Dcr-2 (Fig. 1A) or the unique miRNA species made from each pre-miRNA by Dcr-1. A key challenge for the future will be to determine what enzyme makes rasiRNAs and what sequence or structural features of the unknown rasiRNA precursor lead to the accumulation of a stereotyped pattern of predominantly antisense rasiRNAs.

References and Notes

1. D. Galiana-Arnoux, C. Dostert, A. Schneemann, J. A. Hoffmann, J.-L. Imler, *Nat. Immunol.* **7**, 590 (2006).

RESEARCH ARTICLES

2. X.-H. Wang *et al.*, *Science* **312**, 452 (2006); published online 22 March 2006 (10.1126/science.1125694).
3. R. A. Zambon, V. N. Vakharia, L. P. Wu, *Cell. Microbiol.* **8**, 880 (2006).
4. T. Du, P. D. Zamore, *Development* **132**, 4645 (2005).
5. A. I. Kalmykova, M. S. Klenov, V. A. Gvozdev, *Nucleic Acids Res.* **33**, 2052 (2005).
6. A. Fire *et al.*, *Nature* **391**, 806 (1998).
7. Y. S. Lee *et al.*, *Cell* **117**, 69 (2004).
8. N. T. Hoa, K. M. Keene, K. E. Olson, L. Zheng, *Insect Biochem. Mol. Biol.* **33**, 949 (2003).
9. K. Forstemann *et al.*, *PLoS Biol.* **3**, e236 (2005).
10. K. Saito, A. Ishizuka, H. Siomi, M. C. Siomi, *PLoS Biol.* **3**, e235 (2005).
11. F. Jiang *et al.*, *Genes Dev.* **19**, 1674 (2005).
12. D. S. Schwarz *et al.*, *Cell* **115**, 199 (2003).
13. P. D. Zamore, T. Tuschl, P. A. Sharp, D. P. Bartel, *Cell* **101**, 25 (2000).
14. Y. S. Lee, R. W. Carthew, *Methods* **30**, 322 (2003).
15. A. A. Aravin *et al.*, *Dev. Cell* **5**, 337 (2003).
16. A. A. Aravin *et al.*, *Curr. Biol.* **11**, 1017 (2001).
17. P. Tritto *et al.*, *Genetica* **117**, 247 (2003).
18. Y. Tomari *et al.*, *Cell* **116**, 831 (2004).
19. H. A. Cook, B. S. Koppetsch, J. Wu, W. E. Theurkauf, *Cell* **116**, 817 (2004).
20. T. Dalmay, R. Horsefield, T. H. Braunstein, D. C. Baulcombe, *EMBO J.* **20**, 2069 (2001).
21. G. Meister *et al.*, *Curr. Biol.* **15**, 2149 (2005).
22. W. Stapleton, S. Das, B. D. McKee, *Chromosoma* **110**, 228 (2001).
23. R. W. Williams, G. M. Rubin, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 6889 (2002).
24. D. N. Cox *et al.*, *Genes Dev.* **12**, 3715 (1998).
25. A. N. Harris, P. M. Macdonald, *Development* **128**, 2823 (2001).
26. A. A. Aravin *et al.*, *Mol. Cell. Biol.* **24**, 6742 (2004).
27. I. Sahut-Barnola, D. Pauli, *Development* **126**, 1917 (1999).
28. V. V. Vagin *et al.*, *RNA Biol.* **1**, 54 (2004).
29. M. Savitsky, D. Kwon, P. Georgiev, A. Kalmykova, V. Gvozdev, *Genes Dev.* **20**, 345 (2006).
30. E. Bernstein, A. A. Caudy, S. M. Hammond, G. J. Hannon, *Nature* **409**, 363 (2001).
31. Q. Liu *et al.*, *Science* **301**, 1921 (2003).
32. S. M. Elbashir, W. Lendeckel, T. Tuschl, *Genes Dev.* **15**, 188 (2001).
33. G. Hutvagner *et al.*, *Science* **293**, 834 (2001); published online 12 July 2001 (10.1126/science.1062961).
34. Z. Yang, Y. W. Ebright, B. Yu, X. Chen, *Nucleic Acids Res.* **34**, 667 (2006).
35. D. N. Cox, A. Chao, H. Lin, *Development* **127**, 503 (2000).
36. K. Okamura, A. Ishizuka, H. Siomi, M. C. Siomi, *Genes Dev.* **18**, 1655 (2004).
37. We thank T. Du, M. Ghildiyal, and A. Boucher for assistance with fly husbandry, and J. Birchler, R. Carthew, F. Gao, R. Lehmann, H. Lin, P. MacDonald, M. Siomi, H. Siomi, D. Smith, E. Sontheimer, H. Ruohola-Baker, and M. Van Doren for fly stocks. P.D.Z. is a W.M. Keck Foundation Young Scholar in Medical Research. Supported in part by NIH grants GM62862 and GM65236 (P.D.Z.) and postdoctoral fellowships from the European Molecular Biology Organization (EMBO) and the Human Frontier Science Program (H.S.).

Supporting Online Material

www.sciencemag.org/cgi/content/full/1129333
Materials and Methods

Figs. S1 to S7

Table S1

References

1 May 2006; accepted 19 June 2006

Published online 29 June 2006;

10.1126/science.1129333

Include this information when citing this paper.

α -Synuclein Blocks ER-Golgi Traffic and Rab1 Rescues Neuron Loss in Parkinson's Models

Antony A. Cooper,^{1*} Aaron D. Gitler,^{2*} Anil Cashikar,^{2‡} Cole M. Haynes,^{1§} Kathryn J. Hill,^{1†} Bhupinder Bhullar,^{2,3} Kangning Liu,^{4,5} Kexiang Xu,⁴ Katherine E. Strathearn,⁶ Fang Liu,⁶ Joshua Cao,⁷ Kim A. Caldwell,⁷ Guy A. Caldwell,⁷ Gerald Marsischky,³ Richard D. Kolodner,⁸ Joshua LaBaer,³ Jean-Christophe Rochet,⁶ Nancy M. Bonini,^{4,5} Susan Lindquist^{2,9||}

Alpha-synuclein (α Syn) misfolding is associated with several devastating neurodegenerative disorders, including Parkinson's disease (PD). In yeast cells and in neurons α Syn accumulation is cytotoxic, but little is known about its normal function or pathobiology. The earliest defect following α Syn expression in yeast was a block in endoplasmic reticulum (ER)-to-Golgi vesicular trafficking. In a genomewide screen, the largest class of toxicity modifiers were proteins functioning at this same step, including the Rab guanosine triphosphatase Ypt1p, which associated with cytoplasmic α Syn inclusions. Elevated expression of Rab1, the mammalian YPT1 homolog, protected against α Syn-induced dopaminergic neuron loss in animal models of PD. Thus, synucleinopathies may result from disruptions in basic cellular functions that interface with the unique biology of particular neurons to make them especially vulnerable.

Parkinson's disease (PD) is the second most common neurodegenerative disorder (1, 2). Accruing evidence points to a causative role for the presynaptic protein alpha-synuclein (α Syn) in PD pathogenesis. α Syn is a major constituent of Lewy Bodies—cellular inclusions that are the hallmark pathological feature of PD and other neurodegenerative disorders collectively referred to as synucleinopathies (3). Moreover, missense mutations in the α Syn gene (A53T, A30P, E46K) (4–6) and duplication or triplication of the locus cause PD (7–9). In mouse, rat, fly, and nematode models of PD, increased levels of α Syn lead to neurodegeneration (10–13). Elucidating the mechanisms underlying the cytotoxic effects of α Syn will be essential for the development of treatments to ameliorate or prevent the synucleinopathies.

Despite extensive study, little is known about α Syn's normal function or how α Syn contributes

to disease. Many cellular defects have been implicated in the etiology of synucleinopathies, including impairment of the ubiquitin-proteasome system, mitochondrial dysfunction, accumulation of lipid droplets, production of reactive oxygen species (ROS), and stress within the ER (14). A yeast PD model, with dosage sensitivity for α Syn expression, recapitulates many of these defects (15). But which are cause and which effect remain unclear. Here, two independent approaches, genetic and cell biological, converged to identify inhibition of ER-Golgi trafficking as a major component of synuclein-dependent toxicity.

α Syn accumulation causes ER stress. An increase in α Syn gene dosage in yeast from one copy (no growth defect) to two copies results in growth arrest and cell death (15) (Fig. 1A). To investigate the earliest defects caused by α Syn, we took advantage of the ability to rapidly and

synchronously induce its expression from a galactose-inducible promoter. A slight decline in viability was observed after 4 hours of induction, and 60% of cells lost colony-forming ability by 8 hours (Fig. 1, A and B). ER stress, measured by a reporter for the unfolded protein response, appeared earlier. Expression of wild-type α Syn (α Syn-WT) or disease-associated α Syn (α Syn-A53T) caused a fourfold increase in ER stress relative to control cells after 6 hours (Fig. 1C).

α Syn accumulation impairs degradation of selective ERAD substrates. ER stress typically results from the accumulation of misfolded proteins within the ER. Such malformed proteins are retrotranslocated from the ER to the cytoplasm for degradation by the proteasome through a process termed ERAD (endoplasmic reticulum associated degradation) (16). Misfolded cytosolic α Syn might impair the proteasome's capacity for protein degradation and so cause an accumulation of misfolded proteins in the ER and associated ER stress. To investigate this possi-

¹School of Biological Sciences, University of Missouri–Kansas City, Kansas City, MO 64110, USA. ²Whitehead Institute for Biomedical Research, Cambridge, MA 02142, USA. ³Harvard Institute of Proteomics, 320 Charles Street, Cambridge, MA 02141, USA. ⁴Department of Biology, University of Pennsylvania, Philadelphia, PA 19104, USA. ⁵Howard Hughes Medical Institute, Philadelphia, PA 19104, USA. ⁶Department of Medicinal Chemistry and Molecular Pharmacology, Purdue University, West Lafayette, IN 47907–2091, USA. ⁷Department of Biological Sciences, University of Alabama, Tuscaloosa, AL 35487, USA. ⁸Ludwig Institute for Cancer Research, University of California, San Diego, School of Medicine, La Jolla, CA 92093, USA. ⁹Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, MA 02142, USA.

*These authors contributed equally to this work.

†Present address: Garvan Institute of Medical Research, Sydney, Australia.

‡Present address: Medical College of Georgia, Augusta, GA, USA.

§Present address: New York University, New York, NY, USA.

||To whom correspondence should be addressed. E-mail: lindquist_admin@wi.mit.edu

2.3 Discussion : finalement, existe-t-il des siRNA endogènes chez la Drosophile ?

Plusieurs autres laboratoires faisaient simultanément la même découverte : des petits ARN, plus longs que les miRNA ou les brins de siRNA, interagissent avec les protéines Piwi chez la Drosophile, le Rat et la Souris (Saito *et al.*, 2006 ; Lau *et al.*, 2006 ; Girard *et al.*, 2006 ; Aravin *et al.*, 2006 ; Grivna *et al.*, 2006) ou proviennent de séquences répétées (Watanabe *et al.*, 2006) (le partenaire protéique des petits ARN décrits dans ce dernier article n'a pas été identifié — mais ces petits ARN partagent toutes les autres caractéristiques des ARN décrits dans les autres analyses). Cette nouvelle classe de petits ARN a donc été baptisée « piRNA », pour : *Piwi-interacting RNAs* (ARN interagissant avec les protéines Piwi).

La biogenèse et le mode d'action de ces petits ARN ont été partiellement élucidés dans les années qui ont suivi : voir le chapitre 4 (page 69) pour le détail de leur biogenèse, et le chapitre 5 (page 103) pour une réflexion sur leur rôle biologique.

Quoiqu'il en soit, la démonstration que les rasiRNA ne sont pas des siRNA nous ramène à la question initiale : existe-t-il des siRNA endogènes, qui expliqueraient la conservation du RNAi parmi les Animaux ?

Quelques observations suggèrent qu'effectivement, le RNAi est actif de manière naturelle chez la Drosophile et chez l'Homme :

- des petits ARN qui semblent être des siRNA ont été observés chez la Drosophile et dans des cellules humaines en culture (voir table 2.1) ;
- chez la Drosophile, des mutations dans la voie du RNAi peuvent avoir des conséquences visibles même en l'absence d'ARN double-brin exogène : la mutation de *Lip* (également appelé *Dmp68* ; ce gène code une hélicase à ARN nécessaire au RNAi dans des cellules en culture, sans que sa fonction moléculaire soit connue : Ishizuka *et al.*, 2002) déréprime plusieurs transposons (Csink *et al.*, 1994) ;
- chez la Drosophile, le rétrotransposon *Idefix* est réprimé par des piRNA dans les gonades, mais par un autre mécanisme dans le reste du corps (Desset *et al.*, 2008).

Espèce	Origine des siRNA	Référence
<i>Drosophila melanogaster</i>	Transposon à ADN 1360	Haynes <i>et al.</i> , 2006
<i>Drosophila melanogaster</i>	Rétrotransposon <i>suffix</i>	Tchurikov et Kretova, 2007
<i>Homo sapiens</i>	Rétrotransposon LINE-1	Yang et Kazazian, 2006

TAB. 2.1 – **De probables siRNA endogènes chez la Drosophile et l'Homme.** Haynes *et al.*, 2006 ont détecté de petits ARN d'environ 23 nt dans des cellules Kc en culture, qui pourraient ne pas être des produits de dégradation d'ARN plus longs ; ils proviennent d'un transposon à ADN appelé « 1360 ». 1360, et deux autres transposons, sont déréprimés quand Dicer-1 et Dicer-2 sont réprimés simultanément par RNAi. De même, Tchurikov et Kretova, 2007 ont observé de petits ARN issus du rétrotransposon *suffix*, qui pourraient ne pas être des produits de dégradation. Yang et Kazazian, 2006 ont détecté des petits ARN d'environ 21 nt dans des cellules humaines en culture ; ces petits ARN proviennent d'une région du rétrotransposon LINE-1 encadrée par deux promoteurs convergents, et l'expression de LINE-1 augmente lorsque Dicer est réprimé par RNAi, ce qui suggère que ces deux promoteurs convergents guident la transcription d'un long ARN double-brin, clivé par Dicer en siRNA, qui répriment ensuite LINE-1 par RNAi.

Nous avons donc entrepris de rechercher d'éventuels siRNA endogènes chez la Drosophile par une technique très sensible, le séquençage à haut débit, et de confirmer leur nature de siRNA

2.3. DISCUSSION : FINALEMENT, EXISTE-T-IL DES siRNA ENDOGÈNES CHEZ LA DROSOPHILE ? 3

en vérifiant qu'ils dépendent de Dicer-2 et d'Ago2, et qu'ils sont méthylés sur l'oxygène 2' de leur dernier ribose, comme les siRNA exogènes. Ces travaux sont présentés dans l'article qui suit (Ghildiyal *et al.*, 2008).

2.4 Article n°2 : découverte des siRNA endogènes de Drosophile

LMA and T_b were placed into a food-restriction paradigm under DD conditions, we found that they maintained the rhythm that had been entrained by light (high-amplitude, free-running period of ~23.7 hours) and never showed an increase in LMA or T_b in anticipation of the food presentation (fig. S3). Hence, although a BMAL1-based clock is necessary to support food entrainment, restoration of clock function in the SCN alone is not able to rescue this behavior.

To test the hypothesis that the BMAL1-based clock induced in the DMH during restricted feeding might drive circadian entrainment, we performed stereotaxic bilateral delivery of AAV-BMAL1 (the same construct and vector as used in the SCN) into the DMH of *Bmal1*^{-/-} mice. Mice who sustained bilateral DMH injections of the AAV-BMAL1 did not demonstrate entrainment to a 12:12 LD cycle or free-running rhythms of T_b or LMA in DD (Fig. 3A). By contrast, under conditions of food restriction in DD, they exhibited a clear anticipatory increase in T_b and LMA before food presentation (Fig. 2C and Fig. 3B). Each individual mouse showed very little day-to-day variation in the timing of the increase in T_b and LMA under DD (i.e., the phase angle of entrainment was stable). Finally, the increase in T_b and LMA before the predicted period of food presentation persisted during a 24-hour fast at the end of restricted feeding (arrow in Fig. 3B), demonstrating the circadian nature of the response.

In both our study and the study by Mieda *et al.* (8), clock gene expression in the DMH was largely restricted to cells in the compact part of the nucleus, which consists of small, closely packed neurons that are highly reminiscent of the SCN itself. These neurons appear mainly to have local connections with the adjacent output zones of the DMH (23), suggesting that the timing signal from the compact part of the DMH may impinge upon the same output neurons in the remainder of the DMH as are used to control light-entrained rhythms directed by the SCN. This relationship may explain how the DMH clock is able to override the SCN clock input during conditions of food entrainment in an intact animal. It is unlikely that feedback from the DMH alters activity in the SCN in any major way, because the SCN remains phase-locked to the LD cycle for many weeks during food entrainment (as long as the animals are not also hypocaloric). These observations also raise the interesting possibility that the DMH may form the neuroanatomic basis of the so-called methamphetamine-sensitive circadian oscillator (MASCO), which also operates independent of the SCN and does not entrain to light [for a review, see (24)].

Our data indicate that there is an inducible clock in the DMH that can override the SCN and drive circadian rhythms when the animal is faced with limited food availability. Thus, under restricted feeding conditions, the DMH clock can assume an executive role in the temporal regulation of behavioral state. For a small mammal, finding food on a daily basis is a critical mission.

Even a few days of starvation, a common threat in natural environments, may result in death. Hence, it is adaptive for animals to have a secondary “master clock” that can allow the animal to switch its behavioral patterns rapidly after a period of starvation to maximize the opportunity of finding food sources at the same time on following days.

In an intact animal, peripheral oscillators in many tissues in the body, including the stomach and liver, as well as elsewhere in the brain, may contribute to food entrainment of circadian rhythms (25, 26). Consequently, it has been difficult to dissect this system by using lesions of individual components of the pathway (3, 9, 10). However, by starting with a genetically arrhythmic mouse and using focal genetic rescue in the brain, we have identified the SCN molecular clock as sufficient for light but not food entrainment of T_b and LMA rhythms in mice, and the DMH as sufficient for food but not light entrainment of circadian rhythms of T_b and LMA. These results demonstrate the power of viral-based gene replacement in the central nervous system to dissect complex neural functions.

References and Notes

1. M. H. Hastings, A. B. Reddy, E. S. Maywood, *Nat. Rev. Neurosci.* **4**, 649 (2003).
2. C. B. Saper, J. Lu, T. C. Chou, J. Gooley, *Trends Neurosci.* **28**, 152 (2005).
3. J. J. Gooley, A. Schomer, C. B. Saper, *Nat. Neurosci.* **9**, 398 (2006).
4. F. K. Stephan, J. M. Swann, C. L. Sisk, *Behav. Neural Biol.* **25**, 346 (1979).
5. P. L. Lowrey, J. S. Takahashi, *Annu. Rev. Genomics Hum. Genet.* **5**, 407 (2004).

6. A. Kramer *et al.*, *Science* **294**, 2511 (2001).
7. F. K. Stephan, *Physiol. Behav.* **46**, 489 (1989).
8. M. Mieda *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 12150 (2006).
9. G. J. Landry, M. M. Simons, I. C. Webb, R. E. Mistlberger, *Am. J. Physiol.* **290**, 6 (2006).
10. J. J. Gooley, C. B. Saper, *J. Biol. Rhythms* **22**, 479 (2008).
11. N. Gekakis *et al.*, *Science* **280**, 1564 (1998).
12. B. Zheng *et al.*, *Cell* **105**, 683 (2001).
13. M. K. Bunker *et al.*, *Cell* **103**, 1009 (2000).
14. E. L. McDearmon *et al.*, *Science* **314**, 1304 (2006).
15. M. K. Bunker *et al.*, *Genes Dev.* **41**, 122 (2005).
16. A. Laposky *et al.*, *Sleep* **28**, 395 (2005).
17. S. Shimba *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 12071 (2005).
18. R. D. Rudic *et al.*, *PLoS Biol.* **2**, e377 (2004).
19. Materials and methods are available as supporting material on Science Online.
20. T. C. Chou *et al.*, *J. Neurosci.* **23**, 1069 (2003).
21. C. A. Fuller *et al.*, *Am. J. Physiol.* **241**, 5 (1981).
22. D. T. Krieger, H. Hauser, L. C. Krey, *Science* **197**, 398 (1977).
23. R. H. Thompson, N. S. Canteras, L. W. Swanson, *Brain Res. Brain Res. Rev.* **27**, 89 (1998).
24. T. Hiroshige, K. Honma, S. Honma, *Brain Res. Bull.* **27**, 441 (1991).
25. K. A. Stokkan *et al.*, *Science* **291**, 490 (2001).
26. F. Damiola *et al.*, *Genes Dev.* **14**, 2950 (2000).
27. We thank Q. Ha and M. Ha for technical work. Support was provided by U.S. Public Health Service grants HL60292, NS33987, NS051609, NS057119, and HL07901-08.

Supporting Online Material

www.sciencemag.org/cgi/content/full/320/5879/1074/DC1
Materials and Methods
Figs. S1 to S4
References
26 November 2007; accepted 14 April 2008
10.1126/science.1153277

Endogenous siRNAs Derived from Transposons and mRNAs in *Drosophila* Somatic Cells

Megha Ghildiyal,^{1*} Hervé Seitz,^{1*} Michael D. Horwich,¹ Chengjian Li,¹ Tingting Du,¹ Soohyun Lee,² Jia Xu,³ Ellen L.W. Kittler,⁴ Maria L. Zapp,⁴ Zhiping Weng,⁵ Phillip D. Zamore^{1†}

Small interfering RNAs (siRNAs) direct RNA interference (RNAi) in eukaryotes. In flies, somatic cells produce siRNAs from exogenous double-stranded RNA (dsRNA) as a defense against viral infection. We identified endogenous siRNAs (endo-siRNAs), 21 nucleotides in length, that correspond to transposons and heterochromatic sequences in the somatic cells of *Drosophila melanogaster*. We also detected endo-siRNAs complementary to messenger RNAs (mRNAs); these siRNAs disproportionately mapped to the complementary regions of overlapping mRNAs predicted to form double-stranded RNA *in vivo*. Normal accumulation of somatic endo-siRNAs requires the siRNA-generating ribonuclease Dicer-2 and the RNAi effector protein Argonaute2 (*Ago2*). We propose that endo-siRNAs generated by the fly RNAi pathway silence selfish genetic elements in the soma, much as Piwi-interacting RNAs do in the germ line.

Three RNA-silencing pathways have been identified in flies and mammals: RNA interference (RNAi), guided by small interfering RNAs (siRNAs) derived from exogenous double-stranded RNA (dsRNA); the microRNA (miRNA) pathway, in which endogenous small RNAs repress partially complementary mRNAs;

and the Piwi-interacting RNA (piRNA) pathway, whose small RNAs repress transposons in the germ line (1–3) and can activate transcription in heterochromatin (4).

Endogenous siRNAs (endo-siRNAs) silence retrotransposons in plants (5, 6), and siRNAs corresponding to the L1 retrotransposon have

panel) and in vitro (18), nearly all siRNAs produced by Dcr-2 from exogenous dsRNA are 21 nucleotides (nt) in length.

We characterized the somatic small RNA content of S2 cells (19) and of heads expressing an RNA hairpin silencing the *white* gene by

RNAi (20). To identify endo-siRNA candidates, we analyzed two types of RNA libraries. For total 18- to 29-nt RNA libraries, 89% (S2 cells) and 96% (heads) mapped to annotated miRNA loci. In contrast, libraries enriched for small RNAs bearing a 3'-terminal, 2'-*O*-methyl modification (21) were depleted of miRNAs: Only 19% (S2 cells) and 49% (heads) of reads and 2.4% (S2 cells; 58,681 reads; 12,036 sequences) and 12% (heads; 22,685 reads; 2929 sequences) of unique sequences mapped to miRNA loci.

Figure 1 shows the length distribution and sequence composition of the four libraries. The total RNA samples were predominantly miRNAs, a bias reflected in their modal length (22 nt) and pronounced tendency to begin with uracil. Exclusion of miRNAs revealed a class of small RNAs with a narrow length distribution and no tendency to begin with uracil. Except for an unusual cluster of X-chromosome small RNAs (fig. S1) and a miRNA-like sequence with an unusual putative precursor on chromosome 2 (fig. S2), few of these small RNAs are likely to correspond to novel miRNAs: None lie in the arms of hairpins predicted to be as thermodynamically stable as most pre-miRNAs (i.e., < -15 kcal/mol).

After excluding known miRNAs, 64% (heads) (Fig. 1A) and 78% (S2 cells) (Fig. 1B) of sequences in the libraries enriched for 3'-terminally modified small RNAs—that is, those likely to be Ago2-associated—were 21 nt long. For fly heads, 37% (8404 reads) derived from the *white* dsRNA hairpin. The abundance of these exo-siRNAs can be estimated by comparing them to the number of reads for individual miRNAs in the total small RNA library, where 1.6% (660 antisense and 491 sense reads) were 21-nt oligomers (21-mers) and matched the *white* sequences in the dsRNA-expressing transgene. The collective abundance of all *white* exo-siRNAs was less than the individual abundance of the 10 most abundant miRNAs in this sample; the median abundance of any one exo-siRNA species was two reads. The *white*-inverted repeat (IR) transgene phenocopies a nearly null mutation in *white*, yet the sequence of the most abundant exo-siRNA was read just 37 times.

In heads, the sequence composition of the 21-nt, 3'-terminally modified small RNAs closely resembled that of exo-siRNAs, which tended to begin and end with cytosine. In heads and S2 cells, the 21-mers lacked the sequence features of piRNAs, which either begin with uracil (Aub- and Piwi-bound) or contain an adenine at position 10 (Ago3-bound) and are 23 to 29 nt long (1, 2). These data suggest that the 21-nt small RNAs are somatic endo-siRNAs.

In S2 cells, endo-siRNAs mapped largely to transposons (86%); in fly heads, they mapped about equally to transposons, intergenic and unannotated sequences, and mRNAs. The finding that 41% of endo-siRNAs mapped to mRNAs without mapping to transposons suggests that endo-siRNAs may regulate mRNA

Table 1. Endo-siRNAs preferentially map to overlapping, complementary mRNAs.

Sample	Enrichment	Enrichment after randomization		Z score	P
		Mean	SD		
Fly heads	10.9	1.0	0.38	26.1	7.9×10^{-151}
S2 cells	12.3	1.1	0.42	27.0	5.2×10^{-161}

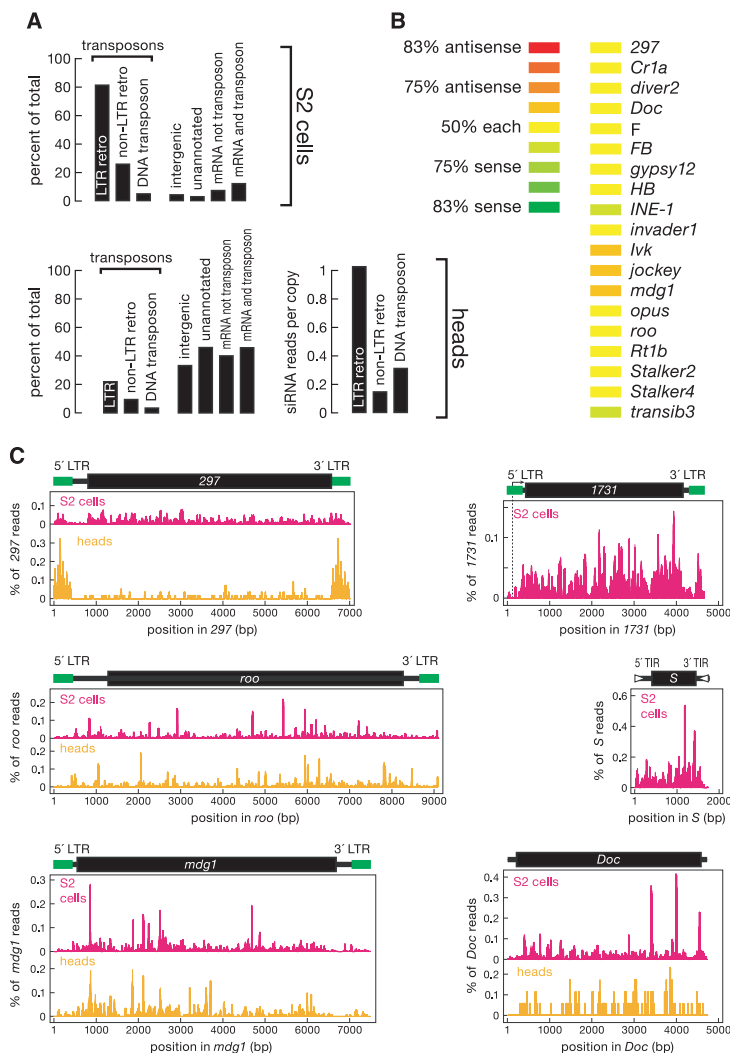


Fig. 2. Endo-siRNAs correspond to transposons. (A) Distribution of annotations for the genomic matches of endo-siRNA sequences. Bars total more than 100% because some siRNAs match both LTR and non-LTR retrotransposons or match both mRNA and transposons. (B) Transposon-derived siRNAs with more than 50 21-nt reads mapped about equally to sense and antisense orientations. (C) Alignment of endo-siRNA sequences to *Drosophila* transposons. The abundance of each sequence is shown as a percentage of all transposon-matching siRNA sequences. LTR, long terminal repeat; TIR, terminal inverted repeat. Here and in subsequent figures, data from high-throughput pyrosequencing and sequencing-by-synthesis were pooled for wild-type heads.

REPORTS

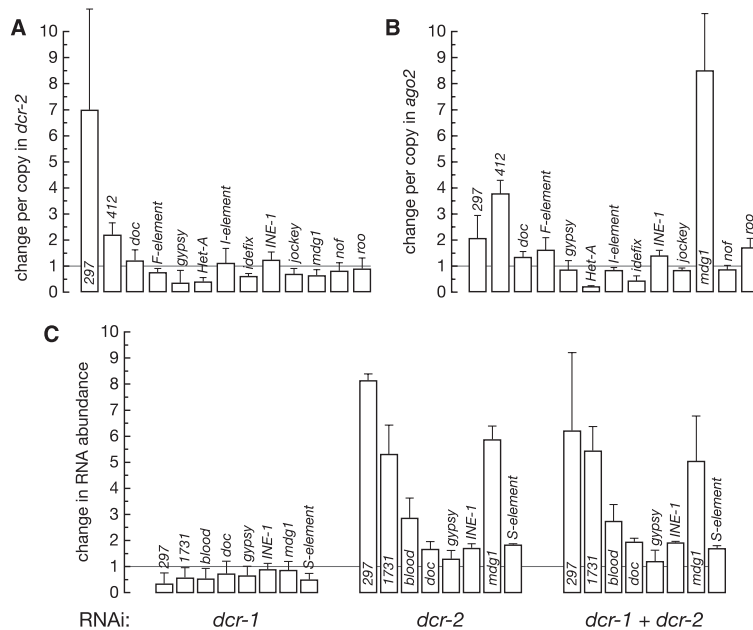


Fig. 3. Transposon silencing requires Dcr-2 and Ago2, but not Dcr-1. (A and B) The change in mRNA expression (mean \pm SD, $N = 3$) for each transposon between *dcr-2*^{L811fSX} (A) or *ago2*⁴¹⁴ (B) heterozygous and homozygous heads was measured by quantitative reverse transcription polymerase chain reaction. The data were corrected for differences in transposon copy number between the paired genotypes. (C) The change in transposon expression (mean \pm SD, $N = 3$) in S2 cells was measured for the indicated RNAi depletion relative to a control dsRNA.

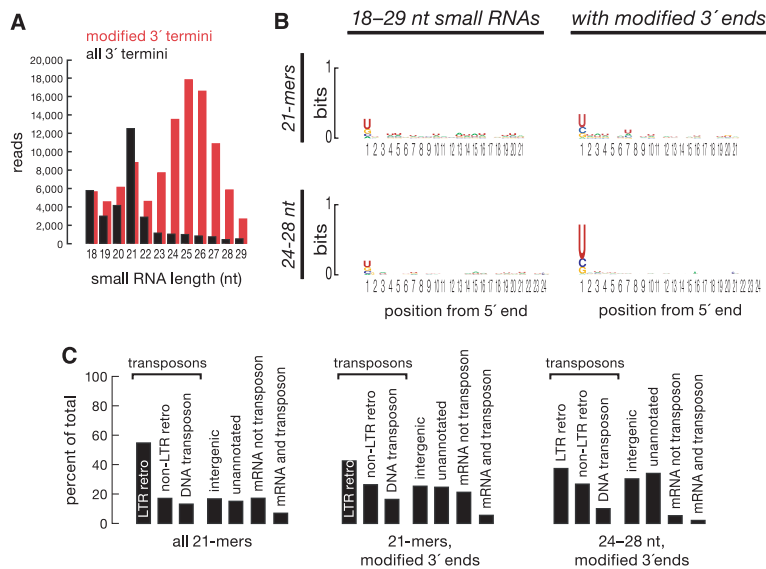


Fig. 4. The composition of somatic small RNAs is altered in the absence of Ago2. (A and B) Size distribution (A) and sequence composition (B) of sequences from a library of total 18- to 29-nt RNA from the heads of *ago2* null mutant flies or a library enriched for 3'-terminally modified RNAs. Reads matching pre-miRNA sequences were removed. (C) Distribution of annotations for the genomic matches of small RNA sequences from the two *ago2* libraries.

expression. Endo-siRNAs mapping to mRNAs were likelier by a factor of >10 than expected by chance ($5.22 \times 10^{-161} < P < 8 \times 10^{-151}$) to derive from genomic regions annotated to produce overlapping, complementary transcripts (Table 1 and table S1). These data suggest that such overlapping, complementary transcripts anneal in vivo to form dsRNA that is diced into endo-siRNAs. We note that among the mRNAs for which we detected complementary 21-mers was *ago2* itself.

Endo-siRNAs mapped to all three large chromosomes (figs. S3 to S5). siRNAs corresponding to the three transposon types in *Drosophila* were detected, but long terminal repeat (LTR) retrotransposons, the dominant class of selfish genetic elements in flies, were overrepresented even after accounting for their abundance in the genome (Fig. 2A and table S2). Unlike piRNAs, which are disproportionately antisense to transposons, but like siRNAs derived from exogenous dsRNA, about equal numbers of sense and antisense transposon-matching endo-siRNAs were detected (Fig. 2B and fig. S6) (1–3, 22). Like piRNAs, endo-siRNAs map to large genomic clusters (table S3). Of 172 endo-siRNA clusters in S2 cells, four coincided with previously identified piRNA clusters (cluster 1, at 42A of chromosome 2R; clusters 7 and 10 in unassembled genomic sequence; and cluster 15 in the chromosome 3L heterochromatin). In heads, we detected 17 clusters; five corresponded to clusters found in S2 cells, but only one was shared with the germline piRNAs: the *flamenco* locus, consistent with recent genetic evidence that a Piwi-independent but *flamenco*-dependent pathway represses the *Idefix* and *ZAM* transposons in the soma (23). That both endo-siRNAs and piRNAs can arise from the same region suggests either that a single transcript can be a substrate for both piRNA and siRNA production or that distinct classes of transcripts arise from a single locus. The abundance and distribution of endo-siRNAs across the sequences of individual transposon species reflected the natural history of when the elements entered the fly genome, but not their mechanism of transposition (Fig. 2C) (24).

Statistically significant reductions in siRNA abundance were observed in *dcr-2*^{L811fSX} null mutant heads relative to heads from heterozygous siblings for 38 transposons (fig. S7 and table S4). Normalized for sequencing depth, sequencing results from homozygous *dcr-2* mutant heads yielded fewer 21-mers overall (by a factor of 3.1) and fewer 21-mers corresponding to transposons (by a factor of 6.3) than did their heterozygous siblings ($P < 2.2 \times 10^{-16}$; χ^2 test). In contrast, overall miRNA abundance—normalized to sequencing depth—was essentially unchanged between *dcr-2* heterozygotes and homozygotes (fig. S7 and table S5). These data suggest that endo-siRNAs are produced by Dcr-2, but we do not yet know why some endo-siRNAs persist in *dcr-2*^{L811fSX} mutants.

Transposon expression in the soma reflects both the silencing of transposons—potentially by either or both posttranscriptional and transcriptional mechanisms—and the tissue specificity of transposon promoters. *Drosophila* somatic cells may contain siRNAs targeting transposons that would not be highly expressed even in the absence of those siRNAs, because the promoters of those transposons are not active in some or all somatic tissues or because they are repressed by additional mechanisms. We analyzed the expression of a panel of transposons in heads from *ago2* and *dcr-2* mutants and in S2 cells depleted of Dcr-1, Dcr-2, or Ago2 by RNAi (Fig. 3 and fig. S8). We found that the steady-state abundance of RNA from the LTR retrotransposons 297 and 412 increased in heads from *dcr-2^{Δ115X}* null mutants (Fig. 3A). Similarly, the steady-state abundance of RNA from the LTR retrotransposons 297, 412, *mdg1*, and *roo*, the non-LTR retrotransposon *F-element*, and the SINE-like element *INE-1* increased in *ago2^{Δ14}* mutant heads (Fig. 3B).

In S2 cells, RNA expression from the LTR retrotransposons 297, 1731, *mdg1*, *blood*, and *gypsy* and from the DNA transposon *S-element* all increased significantly ($0.00001 < P < 0.002$) when Dcr-2 was depleted or when both Dcr-2 and Dcr-1 were depleted, but not when Dcr-1 alone was depleted (Fig. 3C). Similarly, *ago2(RNAi)* in S2 cells desilenced transposons, including nine LTR and non-LTR retrotransposons and the DNA transposon *S-element* (fig. S8).

Is Ago2 required for the production or accumulation of endo-siRNAs? We sequenced 18- to 29-nt small RNAs from *ago2^{Δ14}* homozygous fly heads and from the same small RNA sample treated to enrich for 3'-terminally modified RNAs. After computationally removing miRNAs, the sequences from the untreated library contained a prominent 21-nt peak (Fig. 4A) that predominantly began with uracil (Fig. 4B), much like miRNAs and unlike siRNAs in wild-type heads, which often began with cytosine (Fig. 1A). Perhaps in the absence of Ago2, only a subpopulation of endo-siRNAs that can bind Ago1 accumulates. The small RNAs from the *ago2^{Δ14}* library enriched for 3'-terminally modified sequences were predominantly 24 to 27 nt long and often began with uracil—a length distribution and sequence bias characteristic of piRNAs, which, like siRNAs, are 2'-O-methylated at their 3' ends. Both the 21-nt small RNAs and the piRNA-like RNAs in the *ago2* mutant heads mapped to transposons, unannotated heterochromatic and unassembled sequences, but the piRNA-like sequences mapped to mRNAs far less frequently than did either the 21-mers or wild-type endo-siRNAs (Fig. 4C). How these piRNA-like small RNAs are generated and whether they contribute to transposon silencing in the fly soma remain unknown.

Note added in proof: The loci described here in figs. S1 and S2 correspond to endo-siRNA-generating hairpins recently identified in (25–27).

References and Notes

1. L. S. Gunawardane *et al.*, *Science* **315**, 1587 (2007); published online 21 February 2007 (10.1126/science.1140494).
2. J. Brennecke *et al.*, *Cell* **128**, 1089 (2007).
3. V. V. Vagin *et al.*, *Science* **313**, 320 (2006); published online 28 June 2006 (10.1126/science.1129333).
4. H. Yin, H. Lin, *Nature* **450**, 304 (2007).
5. A. Hamilton, O. Voinnet, L. Chappell, D. Baulcombe, *EMBO J.* **21**, 4671 (2002).
6. R. Sunkar, T. Girke, J. K. Zhu, *Nucleic Acids Res.* **33**, 4443 (2005).
7. N. Yang, H. H. J. Kazazian, *Nat. Struct. Mol. Biol.* **13**, 763 (2006).
8. A. Ishizuka, M. C. Siomi, H. Siomi, *Genes Dev.* **16**, 2497 (2002).
9. A. K. Csink, R. Linsk, J. A. Birchler, *Genetics* **138**, 153 (1994).
10. J. Rehwinkel *et al.*, *Mol. Cell Biol.* **26**, 2965 (2006).
11. K. A. Haynes, A. A. Caudy, L. Collins, S. C. Elgin, *Curr. Biol.* **16**, 2222 (2006).
12. M. Pal-Bhadra, U. Bhadra, J. A. Birchler, *Mol. Cell* **9**, 315 (2002).
13. N. A. Tchurikov, O. V. Kretova, *PLoS ONE* **2**, e476 (2007).
14. M. D. Horwich *et al.*, *Curr. Biol.* **17**, 1265 (2007).
15. A. Pelisson, E. Sarot, G. Payen-Groschene, A. Bucheton, *J. Virol.* **81**, 1951 (2007).
16. K. Saito *et al.*, *Genes Dev.* **21**, 1603 (2007).
17. K. Okamura, A. Ishizuka, H. Siomi, M. C. Siomi, *Genes Dev.* **18**, 1655 (2004).
18. A. Nykanen, B. Haley, P. D. Zamore, *Cell* **107**, 309 (2001).
19. *Drosophila* RNAi Screening Center at Harvard Medical School (http://flyrnai.org/cgi-bin/RNAi_FAQ_lines.pl).
20. Y. S. Lee, R. W. Carthew, *Methods* **30**, 322 (2003).
21. H. Seitz, M. Ghildiyal, P. D. Zamore, *Curr. Biol.* **18**, 147 (2008).
22. P. D. Zamore, T. Tuschl, P. A. Sharp, D. P. Bartel, *Cell* **101**, 25 (2000).
23. S. Dessel, N. Buchon, C. Meignin, M. Coiffet, C. Vaury, *PLoS ONE* **3**, e1526 (2008).
24. See supporting material on Science Online.
25. B. Czech *et al.*, *Nature* 10.1038/nature07007 (2008).
26. Y. Kawamura *et al.*, *Nature* 10.1038/nature06938 (2008).
27. K. Okamura *et al.*, *Nature* 10.1038/nature07015 (2008).
28. We thank A. Boucher and S. Ma for technical assistance; G. Farley for encouragement, support, and technical assistance; and Roche Applied Science for high-throughput sequencing. P.D.Z. is a W. M. Keck Foundation Young Scholar in Medical Research. Supported by NIH grants GM62862 and GM65236 (P.D.Z.), GM080625 (J.X. and Z.W.), and HG003367 (S.L.); EMBO long-term (LT00575/2005-L) fellowships (H.S.); and a National Research Service Award predoctoral MD/PhD fellowship from the National Institute on Aging (F30AG030283) (M.D.H.). NCBI Gene Expression Omnibus accession numbers for sequence and abundance data are GSE9389 and GSE11019, respectively.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1157396/DC1
Materials and Methods
Figs. S1 to S8
Tables S1 to S7
References

5 March 2008; accepted 31 March 2008

Published online 10 April 2008;

10.1126/science.1157396

Include this information when citing this paper.

Resource Partitioning and Sympatric Differentiation Among Closely Related Bacterioplankton

Dana E. Hunt,^{1*} Lawrence A. David,^{2*} Dirk Gevers,^{1,3,4} Sarah P. Preheim,¹ Eric J. Alm,^{1,5†} Martin F. Polz^{1†}

Identifying ecologically differentiated populations within complex microbial communities remains challenging, yet is critical for interpreting the evolution and ecology of microbes in the wild. Here we describe spatial and temporal resource partitioning among *Vibrionaceae* strains coexisting in coastal bacterioplankton. A quantitative model (AdaptML) establishes the evolutionary history of ecological differentiation, thus revealing populations specific for seasons and life-styles (combinations of free-living, particle, or zooplankton associations). These ecological population boundaries frequently occur at deep phylogenetic levels (consistent with named species); however, recent and perhaps ongoing adaptive radiation is evident in *Vibrio splendidus*, which comprises numerous ecologically distinct populations at different levels of phylogenetic differentiation. Thus, environmental specialization may be an important correlate or even trigger of speciation among sympatric microbes.

Microbes dominate biomass and control biogeochemical cycling in the ocean, but we know little about the mechanisms and dynamics of their functional differentiation in the environment. Culture-independent analysis typically reveals vast microbial diversity, and although some taxa and gene families are differentially distributed among environments (1, 2), it is not clear to what extent coexisting genotypic diversity can be divided into functionally cohesive populations (1, 3). First, we lack broad surveys of nonpathogenic free-living bacte-

ria that establish robust associations of individual strains with spatiotemporal conditions (4, 5); second, it remains controversial what level of genetic diversification reflects ecological differentiation. Phylogenetic clusters have been proposed to correspond to ecological populations that arise by neutral diversification after niche-specific selective sweeps (6). Clusters are indeed observed among closely related isolates (e.g., when examined by multilocus sequence analysis) (7) and in culture-independent analyses of coastal bacterioplankton (8). Yet recent theoretical studies suggest

2.5 Discussion : une nouvelle classe de régulateurs

Nos travaux ont donc mis en évidence la grande variété des siRNA endogènes chez la *Drosophile*, et vérifié qu'ils partageaient toutes les caractéristiques connues des siRNA exogènes. À nouveau, plusieurs autres laboratoires faisaient la même découverte simultanément :

- Okamura *et al.*, 2008b ont rapporté l'observation, chez la *Drosophile*, de petits ARN issus de précurseurs qui se replient en longues tiges-boucles, et qui dépendent de Dicer-2 et d'Ago2. Certains d'entre eux répriment *mus308* (un gène non transposable) par RNAi. Les mêmes auteurs décrivent également l'autre classe de siRNA endogènes, qui provient d'ARN double-brin bimoléculaires (dus à une transcription convergente), et dont la biogenèse nécessite également Dicer-2 (Okamura *et al.*, 2008a). Ils soulignent que de nombreux siRNA endogènes répriment les transposons *in vivo* (Chung *et al.*, 2008).
- Czech *et al.*, 2008 ont séquencé les petits ARN qui interagissent avec Ago2 *in vivo* chez la *Drosophile*, et montré qu'un quart (dans les cellules S2) à la moitié (dans les ovaires) de ces siRNA proviennent de transposons, qu'ils répriment par RNAi. Les siRNA endogènes peuvent également provenir d'ARN repliés en longues tiges-boucles, et réprimer d'autres cibles (ces auteurs montrent également que *mus308* est réprimé par ces siRNA). Les siRNA issus de ces longues tiges-boucles sont perdus en l'absence de Dicer-2 et Ago2, mais aussi, de façon inattendue, en l'absence de Loquacious, partenaire connu de Dicer-1 (voir page 19). Il semble que, sur les tiges-boucles les moins parfaitement appariées, Dicer-2 s'associe avec Loquacious au lieu de R2D2 (son partenaire prépondérant sur les ARN double-brin parfaitement appariés). Les siRNA issus d'autres loci (unités de transcription convergentes, et séquences répétées) sont quant à eux insensibles à la perte de Loquacious.
- Kawamura *et al.*, 2008 ont également séquencé les petits ARN qui interagissent avec Ago2 (dans des cellules de *Drosophile* en culture), et montré que leur accumulation dépendait de Dicer-2. Ces auteurs remarquent aussi que ces siRNA endogènes proviennent fréquemment de rétrotransposons, ou d'ARN qui se replient en longues tiges-boucles.
- Dans les ovocytes de Souris, des siRNA endogènes sont détectés par Tam *et al.*, 2008 et Watanabe *et al.*, 2008. Ces siRNA proviennent fréquemment des loci (riches en transposons) qui expriment également des piRNA, mais la proportion relative des piRNA et des siRNA est variable d'un transposon à l'autre. Une autre famille de siRNA provient de gènes non transposables, et, le plus souvent, leurs brins sens sont issus du gène fonctionnel et leurs brins antisens, d'un pseudogène homologue : ces siRNA semblent donc dériver d'ARN double-brin issus de l'appariement d'un ARN_m avec l'ARN antisens d'un pseudogène (qui se serait inséré dans le génome en aval d'un promoteur, en orientation antisens). Les cibles de ces siRNA pseudogéniques sont dérégulées dans des ovocytes mutants pour Dicer, ce qui confirme l'effet de ces siRNA *in vivo*.

Collectivement, ces travaux ont donc mis en évidence une nouvelle classe d'ARN régulateurs chez des organismes dépourvus de polymérase à ARN dépendant de l'ARN (RdRP), où les mécanismes de biogenèse d'éventuels siRNA endogènes n'apparaissent pas évidents. Dans le cas des Mammifères, on ne sait pas encore si les longs ARN double-brin (dont dérivent certains siRNA endogènes) déclenchent une réponse interféron quand ils sont transcrits, ou si un mécanisme additionnel les soustrait au système de détection des ARN double-brin viraux.

Ces siRNA endogènes remplissent apparemment plusieurs fonctions : ils répriment aussi bien des transposons que des gènes non transposables, dans la lignée germinale et dans le soma. Ils attribuent également une fonction aux pseudogènes, au moins chez la Souris : en permettant la production de siRNA spécifiques de leurs gènes ancestraux, des pseudogènes semblent participer à leur régulation.

Chapitre 3

La biogenèse des microARN : mécanismes de contrôle

3.1 Contexte : hétérogénéité des extrémités de microARN

Dans l'article n°2 (page 38), des expériences de séquençage à haut débit nous ont permis de découvrir des siRNA endogènes chez la Drosophile. Nous avons pris soin de nous affranchir des deux autres classes de petits ARN, piRNA et miRNA, pour pouvoir les détecter : c'est ainsi que nous avons séquençé des préparations d'ARN à partir de têtes de Drosophile, et de cellules S2 en culture (pour éviter les piRNA, spécifiques des gonades) ; il n'était pas possible d'éliminer les miRNA (exprimés dans tous les tissus) par voie expérimentale, c'est pourquoi nous avons dû les éliminer informatiquement de nos jeux de données.

Mais ces séquences contaminantes pouvaient présenter un intérêt : nous avons séquençé des centaines de milliers de molécules de miRNA et de miRNA*, ce qui allait nous permettre d'étudier la séquence de molécules uniques de ces petits ARN, sur des nombres suffisamment grands pour pouvoir faire des analyses statistiques rigoureuses.

Ces jeux de données montrent que la position des extrémités des miRNA présente une grande hétérogénéité, qui n'est pas reflétée dans la banque de données miRBase ¹. Nous avons identifié deux sources d'hétérogénéité : d'une part, des additions extramatricielles (addition de nucléotides sur la molécule mature ; la séquence du petit ARN modifié diffère alors de la séquence de ses précurseurs, sauf coïncidence, qui aurait ajouté au petit ARN le nucléotide qui le suivait dans la séquence du précurseur) ; d'autre part, une imprécision de la position des sites de clivage du précurseur, qui déterminent la position des extrémités du petit ARN.

Ces deux types d'hétérogénéité sont distribués de façon très asymétrique : ils affectent très peu l'extrémité 5' (pour un miRNA typique, environ 95 % de ses molécules commencent sur le même nucléotide), mais sont très fréquents sur l'extrémité 3' (pour un miRNA typique, 40 % de ses molécules ont une extrémité 3' différente de l'extrémité 3' la plus abondante). Des expériences supplémentaires nous ont permis de montrer que le chargement des miRNA (et, ce qui n'était pas attendu : de quelques miRNA*) sur la protéine Ago2 filtre les produits de Dicer, dont les extrémités 5' sont plus hétérogènes que celles des petits ARN chargés sur Ago2. La machinerie de chargement d'Ago2 homogénéise donc les extrémités 5' des petits ARN, en chargeant préférentiellement une unique isoforme 5' pour chaque miRNA (en revanche, elle ne semble pas avoir de préférence sur l'extrémité 3').

Ces résultats sont présentés dans l'article qui suit (Seitz *et al.*, 2008).

¹Cette banque de données (anciennement « microRNA registry ») est accessible sur <http://microrna.sanger.ac.uk/>. Elle recense tous les miRNA détectés expérimentalement, et leurs orthologues dans d'autres espèces. Une seule séquence y est donnée pour chaque miRNA (en général, la première séquence à avoir été clonée), même quand plusieurs isoformes ont été décrites.

3.2 Article n°3 : un contrôle de qualité des microARN

Current Biology 18, 147–151, January 22, 2008 ©2008 Elsevier Ltd All rights reserved DOI 10.1016/j.cub.2007.12.049

Report

Argonaute Loading Improves the 5' Precision of Both MicroRNAs and Their miRNA* Strands in Flies

Hervé Seitz,¹ Megha Ghildiyal,¹ and Phillip D. Zamore^{1,*}¹Department of Biochemistry and Molecular Pharmacology
University of Massachusetts Medical School
Worcester, Massachusetts 01605

Summary

MicroRNAs (miRNAs) are short regulatory RNAs that direct repression of their mRNA targets. The miRNA “seed”—nucleotides 2–7—establishes target specificity by mediating target binding [1–5]. Accurate processing of the miRNA 5' end is thought to be under strong selective pressure [6, 7] because a shift by just one nucleotide in the 5' end of a miRNA alters its seed sequence, redefining its repertoire of targets (Figure 1). Animal miRNAs are produced by the sequential cleavage of partially double-stranded precursors by the RNase III endonucleases Droscha and Dicer, thereby generating a transitory double-stranded intermediate comprising the miRNA paired to its partially complementary miRNA* strand [8, 9]. Here, we report that in flies, the 5' ends of miRNAs and miRNA* strands are typically more precisely defined than their 3' ends. Surprisingly, the precision of the 5' ends of both miRNA and miRNA* sequences increases after Argonaute2 (Ago2) loading. Our data imply that either many miRNA* sequences are under evolutionary pressure to maintain their seed sequences—that is, they have targets—or that secondary constraints, such as the sequence requirements for loading small RNAs into functional Argonaute complexes, narrow the range of miRNA and miRNA* 5' ends that accumulate in flies.

Results and Discussion

We used high-throughput pyrosequencing of 18–30 nt RNAs to identify miRNAs expressed in *Drosophila melanogaster* heads and in cultured *Drosophila* S2 cells. Among the 120,896 miRNA reads (66,377 from fly heads; 54,519 from S2 cells), we observed two sources of heterogeneity for the ends of fly miRNAs: the addition of nucleotides not present in the gene from which the miRNA is transcribed (nontemplated nucleotides) and inaccurate or alternative cleavage by Droscha or Dicer. Approximately 5% of the reads for a typical miRNA contained nontemplated nucleotides on at least one end (Figure 2A and Figure S1 available online), most frequently the addition of a single uridine or adenosine to the 3' end, but longer extensions were also observed, both on the 5' and the 3' ends (Table S1). Interestingly, longer extensions were also U- and A-rich at the 3' end, whereas at the 5' end, the 3'-most nontemplated nucleotide was frequently a cytidine, and other added nucleotides were typically uridines. This observation could prove to be useful for the identification of the 5'-elongating enzymatic activity. The nontemplated addition of nucleotides, especially uridines, to the 3' ends of miRNAs has been reported

previously in wild-type *Caenorhabditis elegans* [6] and *hen1* mutant *Arabidopsis thaliana* [10]. Overall, the addition of nontemplated nucleotides to the 5' end of miRNAs was rarer (~1%; Figure 2A and Table S1).

We also observed a second, more frequent type of heterogeneity: variability in the position of the miRNA 5' and 3' ends within the sequence of the miRNA precursors (Figure 2B). Nontemplated nucleotides fortuitously matching the templated sequence are predicted to occur much less often than the heterogeneity we observe (Table S2). Similar terminal heterogeneity has been noted for the 3' ends of *C. elegans* [6] and the 5' and 3' ends of mouse [11] miRNAs. The aberrant miRNA termini we observe are likely to reflect imprecision in precursor cleavage by Droscha and Dicer. They are unlikely to correspond to degradation products because we recorded nearly as many miRNA reads that were longer than the dominant species as were shorter (Figure S2) and because 93% (S2 cells) and 99% (fly heads) of sequences of the fly-specific 30 nt 2S ribosomal RNA (rRNA)—whose termini are expected to be single-stranded—were full length (Supplemental Discussion). 3' degradation was slightly more common than 5' degradation: We detected 3' degradation for 1010 reads versus 5' degradation for 201 reads among the 33,505 total 2S rRNA reads from S2 cells and fly heads combined; five reads corresponded to 2S rRNA trimmed from both ends.

The 5' ends of miRNAs were more precisely defined than their 3' ends, irrespective of whether the miRNA originated from the 5' or 3' arm of the pre-miRNA (Figure 3A). Thus, the difference in cleavage accuracy between the 5' and 3' ends cannot be attributed to an intrinsic difference in fidelity between Droscha and Dcr-1. We expected that the 3' ends of miRNA* strands would be precisely defined because they are created by the pair of cuts that generates the 5' ends of miRNA and that the 5' ends of miRNA* strands would be imprecisely determined because they are created by the pair of cleavages that generates the highly heterogeneous 3' ends of miRNA. Instead, we found that the 5' end of a strand (for example, the miRNA) was more accurate than the 3' end of the adjacent strand (in this example, the miRNA*; Figure 3B); these two extremities are produced by a pair of cuts catalyzed by the same enzyme.

Current dogma holds that the local sequence or structure of miRNA precursors is under strong selective pressure to generate accurate 5' ends because a precise miRNA 5' end directly establishes the seed sequence and hence the targets of the miRNA. Because we observe that, in flies, the 5' ends of both the miRNA and the miRNA* are more precisely determined than the 3' ends of either strand, this explanation implies that miRNA* sequences are under selective pressure to establish a unique seed sequence, implying that they, too, have regulatory targets.

It is also possible that both Droscha and Dcr-1—whose active sites are homologous—may also be intrinsically more precise in 5' cleavage than in 3' cutting. A third alternative is that 5' and 3' ends might be generated with similar, imperfect accuracy, but subsequent constraints in RISC loading or stability select for those small RNAs that begin with

*Correspondence: phillip.zamore@umassmed.edu

Current Biology Vol 18 No 2
148

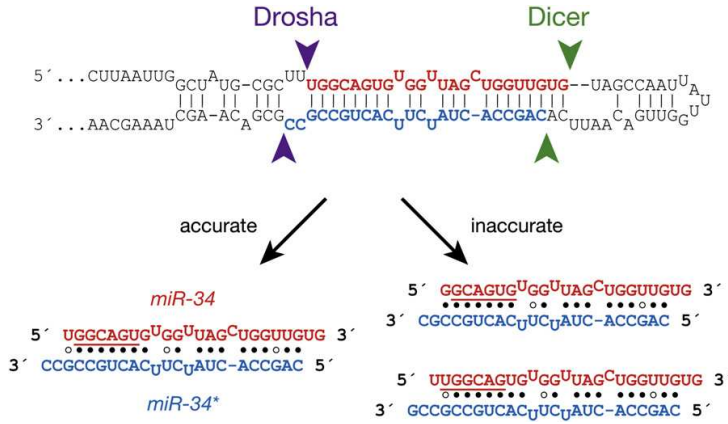


Figure 1. Inaccurate Processing of the 5' End of a miRNA Alters its Seed Sequence
miRNA precursors are cleaved by two RNase III enzymes, Drosha and Dicer, liberating a short duplex: In this duplex, the mature miRNA (red) is paired to a partially complementary small RNA, the miRNA* (blue), derived from the opposite arm of the pre-miRNA stem. Inaccurate cleavage of the miRNA 5' end changes its seed sequence (underlined).

a particular nucleotide or sequence. The subsequent destruction of miRNAs without these 5' features would increase the apparent accuracy of miRNA 5' ends and retain miRNA 3' heterogeneity. To test this idea, we separately sequenced small RNAs containing modified 3' termini (Table S3). In flies, the 3' termini of small RNAs that are loaded into Ago2 [12], but not those bound to Argonaute1 [13], are 2'-O-methylated

by *Drosophila* Hen1 as the last step in Ago2-RISC maturation [14]. To sequence small RNAs bearing 2'-O-methylated 3' ends, we treated the total small RNA with NaO₄ followed by β-elimination; this method blocks ligation of adapters to small RNAs bearing 2',3' hydroxy termini, preventing them from being sequenced.

To determine whether the greater accuracy of miRNA and miRNA* 5' versus 3' ends reflects the constraints of RISC assembly or stability, rather than more accurate 5' versus 3' cleavage by Drosha and Dicer, we compared the terminal heterogeneity of miRNA and miRNA* reads from the 3' modified population to the heterogeneity of the total miRNA and miRNA* population. As a control, we compared the 3' heterogeneity between the two populations. For both analyses, we only considered miRNA or miRNA* strands displaying some heterogeneity in the total population. For both fly heads and S2 cells, we observed a dramatic increase in the precision of the 5'—but not the 3'—ends of miRNAs and miRNA* strands upon loading into Ago2 (Figure 4). We also performed the analysis for those small RNAs that both had heterogeneous 5' termini and were specifically enriched in the β-eliminated sequences relative to the non-β-eliminated set. For the 13 small RNAs (four miRNAs and nine miRNA*s) meeting these criteria, the 5' ends in the subpopulation of miRNA and miRNA* sequences loaded into Ago2—i.e., those that were 2'-O-methylated—were again more precisely defined than the 5' ends of the

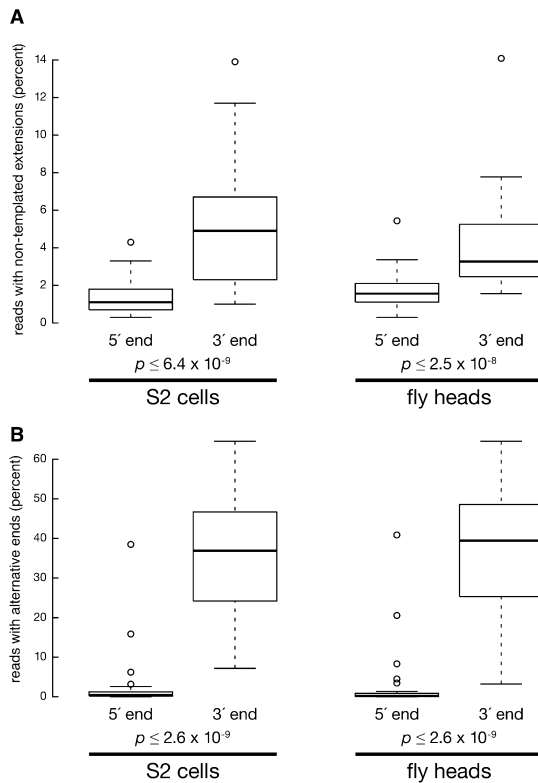


Figure 2. Cleavage Inaccuracies Are More Frequent than Nontemplated Additions

(A) The percentage of reads with nontemplated 5' or 3' extensions was evaluated for each miRNA whose sequence was read at least 100 times.
(B) The most abundant 5' and 3' ends were identified for each miRNA, and all other ends corresponding to the sequence of the primary miRNA transcript were flagged as "alternative." The percentage of reads with alternative ends was then determined for each miRNA read at least 100 times. Note the difference in the y axis scales in (A) and (B). Box plots follow Tukey's standard conventions: A rectangle encloses all data from the first to the third quartiles, a bold horizontal line reports the median, whiskers connected to the rectangle indicate the largest and smallest nonoutlier data, and outliers (values distant from the box by more than 1.5× the interquartile range) are displayed as open circles.

Argonaute Loading Improves MicroRNA 5' Accuracy

149

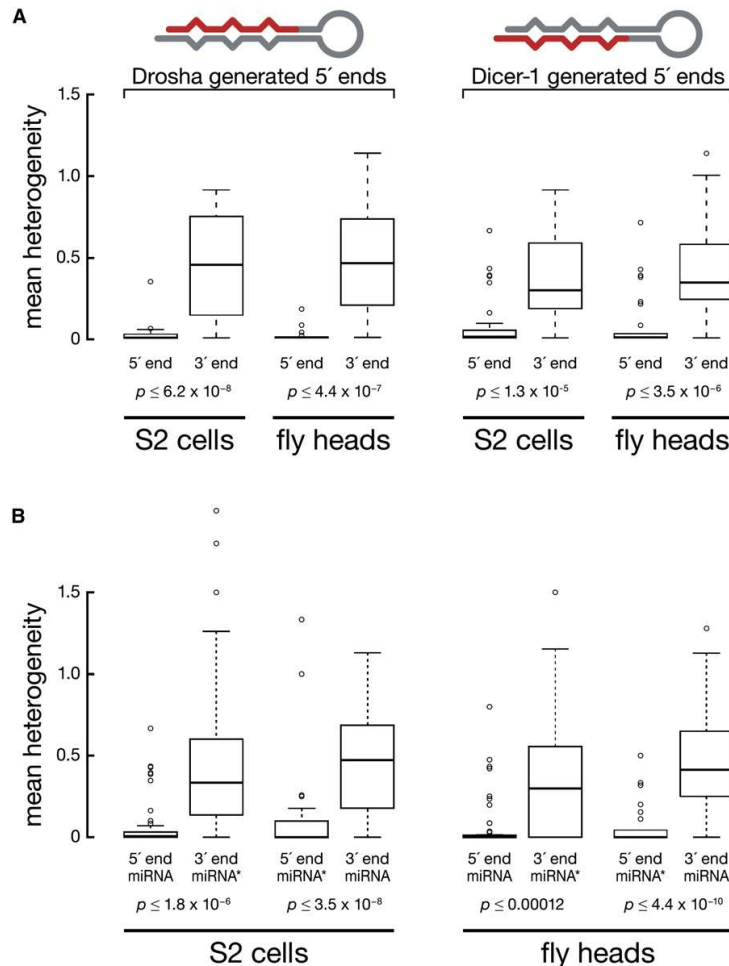


Figure 3. miRNA and miRNA* 5' Ends Are More Precisely Defined than Their 3' Ends

(A) miRNAs originating from the 5' (left panels) or 3' (right panels) arm of their pre-miRNAs were analyzed separately. For each miRNA, the heterogeneity of its termini was calculated as the mean of the absolute values of the distance between the 5' or 3' extremity of an individual templated read and the most abundant 5' or 3' end for that miRNA. Sequences read from RNA isolated from fly heads and cultured S2 cells were analyzed separately.

(B) Box plots show the distribution of mean heterogeneity for the 5' and 3' ends of miRNA and miRNA* sequences.

[16, 17] in nested, prechilled sieves (U.S.A. standard sieve, Humboldt MFG, Chicago, IL), allowing the heads to pass through the top sieve (No. 25), and collecting them on the bottom sieve (No. 40). S2 cell RNA was prepared from a clonal line containing the stably integrated GFP transgene (pKF63) and was transiently transfected with a double-stranded RNA against GFP [18].

RNA Preparation

A amount of 400 μ g total RNA was extracted with the mirVana kit (Ambion), then 18- to 30 nt long RNAs were gel purified. 2S rRNA was depleted by hybridization to immobilized DNA oligonucleotide (5'-biotin-TCA ATG TCG ATA CAA CCC TCA ACC ATA TGT AGT CCA AGC A-3'). A total of 1.6 nmol of the biotinylated oligonucleotide was bound to 32 mg M270 Streptabeads (Dyna, Norway) in 3.2 ml 0.5 \times SSC for 30 min on ice, and then the beads were washed with ice-cold 0.5 \times SSC, resuspended in 8 ml 0.5 \times SSC, and incubated 5 min at 65 $^{\circ}$ C. Gel-purified RNAs were diluted with 7 volumes 0.5 \times SSC to a final volume of 160 μ l and denatured at 80 $^{\circ}$ C for 5 min; then they were added to the bead suspension and incubated 1 hr at 50 $^{\circ}$ C. Beads were magnetically captured for 1 min at room temperature, and then the 2S rRNA-depleted supernatant

collected and precipitated with absolute ethanol. Greater than 99% of the 2S rRNA was routinely removed without measurably altering miRNA concentration; without the depletion step, nearly all the small RNA reads would correspond to 2S rRNA. Half the sample was then β -eliminated as described [19], and half was subject to the same treatment, except that sodium periodate was omitted.

Amplification and Pyrosequencing

Adapters were ligated to the small RNA sample, and the resulting library was amplified by PCR as described [20], except that a truncation mutant of RNA ligase 2 [Rnl2(1-249)] (see [21]) was used for the 3' ligation step; T4 RNA ligase (Ambion) was used for 5' ligation. The 5' adaptor was 5'-dAdTdC dGdTrA rGrGrC rArCrC rUrGrA rArA-3' (Dharmacon, Lafayette, CO); 3' "pre-adenylated" adapters were 5'-rAppdCdA dCdTdC dGdGdG dCdAdC dCdAdA dGdGdA ddC-3' for fly head and 5'-rAppdTdT dTAdA dCdCdG dCdGdA dAdTdT dCdCdA dGddC-3' for S2 cell RNA (IDT DNA, Coralville, IA). After adaptor addition, the RNA was amplified by PCR with DNA primers corresponding to the adapters. This PCR pool was gel purified (4% Meta-phor Agarose, Cambrex, East Rutherford, NJ) with Qiaex II (QIAGEN, Valencia, CA), then reamplified by PCR (common 5' primer, 5'-GCC TCC CTC GCG CCA TCA GAT CGT AGG CAC CTG AAA-3'; 3' primer for fly heads, 5'-GCC TTG CCA GCC CGC TCA GTC CTT GGT GCC CGA GTG-3'; 3' primer for S2 cells, 5'-GCC TTG CCA GCC CGC TCA GCT GGA ATT CGC GGT TAA A-3'). The PCR-amplified libraries were pyrosequenced by Roche Applied Science (Branford, CT).

same small RNA sequences in the total small RNA population (Figure S3). We conclude that loading or stabilization of miRNAs in Ago2, and perhaps Argonaute proteins in general, imposes a purifying selection on their 5' ends.

The mechanism responsible for the homogenization of 5' ends remains to be determined. We can imagine that the efficiency of Argonaute loading is affected by the nature of the 5' end of a small RNA, much as the stability of its pairing to the other strand influences this process [15]. The 5' sequence itself may also play a role in RISC assembly, with some miRNA variants loaded more efficiently than others, according to the identity of their 5' nucleotide(s). Alternatively, some Argonaute complexes might be selectively stabilized after their assembly, for example, by the presence of a target RNA whose binding stabilizes those RISCs containing miRNA isoforms with a complementary seed sequence.

Experimental Procedures

Biological Sources

We isolated fly heads by vigorously shaking liquid nitrogen-frozen flies expressing a long double-stranded hairpin RNA corresponding to *white*

Current Biology Vol 18 No 2
150

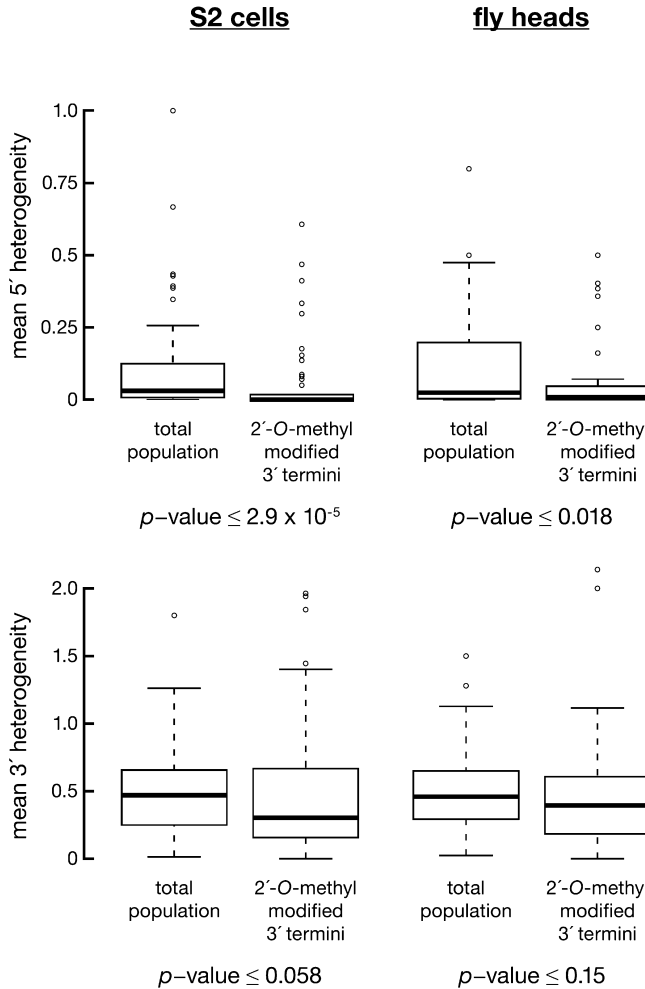


Figure 4. Ago2 Loading, as Evidenced by 3' Terminal 2'-O-Methylation, Refines miRNA and miRNA* 5' Ends

On average, the 5' ends of the miRNAs and miRNA* strands in the 2'-O-methylated populations from both fly heads and S2 cells were more precisely defined than in the total population. We observed no statistically significant increase in the precision of the 3' ends of the 3' modified miRNAs and miRNA* strands.

Acknowledgments

We thank Stewart Shuman and Greg Hannon for providing the truncated Rnl 2 expression plasmid, Alicia Boucher for assistance with fly husbandry, Gwen Farley for preparing RNA ligase, and members of the Zamore lab for advice, suggestions, and critical comments on the text. We are especially grateful to Roche Applied Science for high-throughput sequencing. P.D.Z. is a W.M. Keck Foundation Young Scholar in Medical Research. This work was supported in part by grants from the National Institutes of Health to P.D.Z. (GM62862 and GM65236) and EMBO long-term (ALTF 910-2004) and HFSP (LT00575/2005-L) fellowships to H.S. P.D.Z. is a founder of Alnylam Pharmaceuticals and a member of its scientific advisory board. He is also a member of the scientific advisory board of Regulus Therapeutics.

Received: November 1, 2007

Revised: December 18, 2007

Accepted: December 19, 2007

Published online: January 17, 2008

References

- Lai, E.C. (2002). MicroRNAs are complementary to 3' UTR sequence motifs that mediate negative post-transcriptional regulation. *Nat. Genet.* 30, 363–364.
- Lewis, B.P., Burge, C.B., and Bartel, D.P. (2005). Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* 120, 15–20.
- Lewis, B.P., Shih, I.H., Jones-Rhoades, M.W., Bartel, D.P., and Burge, C.B. (2003). Prediction of mammalian microRNA targets. *Cell* 115, 787–798.
- Haley, B., and Zamore, P.D. (2004). Kinetic analysis of the RNAi enzyme complex. *Nat. Struct. Mol. Biol.* 11, 599–606.
- Krek, A., Grun, D., Poy, M.N., Wolf, R., Rosenberg, L., Epstein, E.J., MacMenamin, P., da Piedade, I., Gunsalus, K.C., Stoffel, M., et al. (2005). Combinatorial microRNA target predictions. *Nat. Genet.* 37, 495–500.
- Ruby, J.G., Jan, C., Player, C., Axtell, M.J., Lee, W., Nusbaum, C., Ge, H., and Bartel, D.P. (2006). Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in *C. elegans*. *Cell* 127, 1193–1207.
- Farh, K.K., Grimson, A., Jan, C., Lewis, B.P., Johnston, W.K., Lim, L.P., Burge, C.B., and Bartel, D.P. (2005). The widespread impact of mammalian MicroRNAs on mRNA repression and evolution. *Science* 310, 1817–1821.
- Du, T., and Zamore, P.D. (2005). microPrimer: The biogenesis and function of microRNA. *Development* 132, 4645–4652.
- Kim, V.N. (2005). MicroRNA biogenesis: Coordinated cropping and dicing. *Nat. Rev. Mol. Cell Biol.* 6, 376–385.
- Li, J., Yang, Z., Yu, B., Liu, J., and Chen, X. (2005). Methylation protects miRNAs and siRNAs from a 3'-end uridylation activity in *Arabidopsis*. *Curr. Biol.* 15, 1501–1507.
- Wu, H., Neilson, J.R., Kumar, P., Manocha, M., Shankar, P., Sharp, P.A., and Manjunath, N. (2007). miRNA profiling of naive, effector and memory CD8 T cells. *PLoS ONE* 2, e1020.
- Pelisson, A., Sarot, E., Payen-Groschene, G., and Bucheton, A. (2007). A novel repeat-associated small interfering RNA-mediated silencing

Computational Methods

Eighteen-to-thirty-nucleotide-long reads were mapped to the *Drosophila melanogaster* genome (FlyBase assemblyR5.1; <http://flybase.org/>) and to the *D. melanogaster* "stem-loops" (which include the pre-miRNA sequences, usually extended by a few nucleotides) listed in miRBase (<http://microrna.sanger.ac.uk/sequences/>; version 10.0, August 2007). To identify nontemplated microRNA additions, we iteratively trimmed nongene matching sequences by one to three nucleotides on either the 5' or the 3' end and mapped them to stem-loops.

Among stem-loop-matching reads, miRNA-matching and miRNA*-matching reads were identified, with either the experimentally detected miRNA* sequence (when it was available in the miRBase records) or the product of conceptual dicing of the hairpin [15]. So that reads that showed extremities different from those annotated in miRBase could be included, a distance of as many as nine nucleotides 5' or 3' from the annotated miRNA or miRNA* sequence was tolerated. Statistical calculations were made with the R statistical package; p values were calculated with the Wilcoxon test.

Supplemental Data

Additional Discussion, three figures, and three tables are available at <http://www.current-biology.com/cgi/content/full/18/2/147/DC1/>.

Argonaute Loading Improves MicroRNA 5' Accuracy
151

- pathway downregulates complementary sense gypsy transcripts in somatic cells of the *Drosophila* ovary. *J. Virol.* *81*, 1951–1960.
13. Saito, K., Sakaguchi, Y., Suzuki, T., Suzuki, T., Siomi, H., and Siomi, M.C. (2007). Pimet, the *Drosophila* homolog of HEN1, mediates 2'-O-methylation of Piwi-interacting RNAs at their 3' ends. *Genes Dev.* *21*, 1603–1608.
 14. Horwich, M.D., Li, C., Matranga, C., Vagin, V., Farley, G., Wang, P., and Zamore, P.D. (2007). The *Drosophila* RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC. *Curr. Biol.* *17*, 1265–1272.
 15. Schwarz, D.S., Hutvagner, G., Du, T., Xu, Z., Aronin, N., and Zamore, P.D. (2003). Asymmetry in the assembly of the RNAi enzyme complex. *Cell* *115*, 199–208.
 16. Lee, Y.S., Nakahara, K., Pham, J.W., Kim, K., He, Z., Sontheimer, E.J., and Carthew, R.W. (2004). Distinct roles for *Drosophila* Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. *Cell* *117*, 69–81.
 17. Lee, Y.S., and Carthew, R.W. (2003). Making a better RNAi vector for *Drosophila*: Use of intron spacers. *Methods* *30*, 322–329.
 18. Förstemann, K., Tomari, Y., Du, T., Vagin, V.V., Denli, A.M., Bratu, D.P., Klattenhoff, C., Theurkauf, W.E., and Zamore, P.D. (2005). Normal microRNA maturation and germ-line stem cell maintenance requires Loquacious, a double-stranded RNA-binding domain protein. *PLoS Biol.* *3*, e236.
 19. Vagin, V.V., Sigova, A., Li, C., Seitz, H., Gvozdev, V., and Zamore, P.D. (2006). A distinct small RNA pathway silences selfish genetic elements in the germline. *Science* *313*, 320–324.
 20. Lau, N.C., Lim, L.P., Weinstein, E.G., and Bartel, D.P. (2001). An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science* *294*, 858–862.
 21. Ho, C.K., Wang, L.K., Lima, C.D., and Shuman, S. (2004). Structure and mechanism of RNA ligase. *Structure* *12*, 327–339.

Accession Numbers

The GEO (<http://www.ncbi.nlm.nih.gov/geo/>) accession number for sequence and abundance data reported in this paper is GSE9389.

3.3 Discussion : une nouvelle étape de contrôle de la biogenèse des microARN

Ces travaux montrent que le chargement des miRNA et miRNA* sur Ago2 contribue à la précision de leur extrémité 5'. L'homogénéité de cette extrémité présente un avantage fonctionnel : il semble que la « graine » (nucléotides 2 à 7, ou 2 à 8, selon les auteurs) d'un miRNA est le déterminant principal de la reconnaissance de ses cibles (voir « Identification des cibles de microARN », page 25). La position de cette fenêtre de six (ou sept) nucléotides est comptée à partir de l'extrémité 5' : si cette extrémité était imprécise sur le miRNA mature, cette fenêtre de six (ou sept) nucléotides serait décalée, donc sa séquence serait modifiée, et le miRNA ne reconnaîtrait plus les mêmes ARN-cibles. *A priori*, une telle contrainte n'existe pas sur l'extrémité 3' : la machinerie de chargement des miRNA participerait donc au contrôle de la fonctionnalité des petits ARN en sélectionnant les isoformes qui présentent une extrémité 5' bien précise ; ce criblage des produits de Dicer constituerait donc un contrôle de leur qualité, avant chargement sur la protéine effectrice.

La purification des produits de Dicer que nous avons mise en évidence n'explique cependant pas toute l'homogénéité des extrémités 5' des miRNA : même les petits ARN non chargés semblent avoir une extrémité 5' plus précise que leur extrémité 3'. Cette observation suggère que les enzymes Drosha et Dicer elles-mêmes clivent les extrémités 5' des petits ARN de manière plus précise que leurs extrémités 3'. On pourrait imaginer deux interprétations à ce phénomène :

- la structure secondaire des pri-miRNA, qui contraint certainement la position des sites de clivage, aurait été sélectionnée au cours de l'évolution, de manière à ce que les imperfections de la double hélice de la tige-boucle guident les clivages par Drosha et Dicer ; puisque la précision de l'extrémité 5' apporte un avantage fonctionnel, la sélection naturelle aurait conservé les structures secondaires qui garantissent l'homogénéité du 5' des miRNA (et elle n'aurait pas sélectionné de motifs de guidage du clivage de leur 3', qui n'aurait aucun effet sur la fonction des miRNA) ; cette hypothèse avait déjà été proposée par Ruby *et al.*, 2006 ;
- Drosha et Dicer possèdent chacune deux domaines RNase III, responsables du clivage de leurs substrats ; chacun catalyse l'un des deux événements de clivage (voir figure 3.1) ². Il est donc possible que, pour chacune de ces deux enzymes, le domaine qui clive les extrémités 5' des petits ARN soit intrinsèquement plus précis que l'autre domaine.

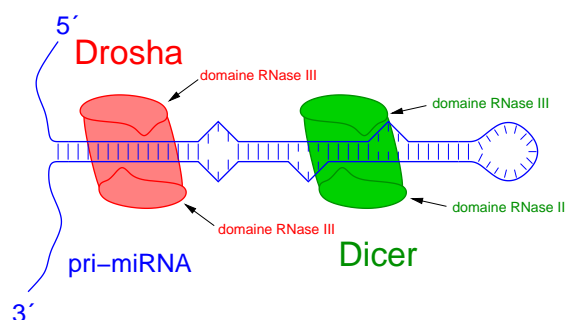


FIG. 3.1 – Drosha et Dicer possèdent deux domaines RNase III chacune. Sur ce schéma, le domaine RNase IIIb de Dicer est représenté au-dessus de la tige-boucle, et le domaine RNase IIIa en-dessous (Zhang *et al.*, 2004 ont montré que le premier générait une extrémité 3' du petit duplex, et que le deuxième générait une extrémité 5').

²Ceci n'a toutefois été démontré que pour Dicer (Zhang *et al.*, 2004), et, même si Drosha possède également deux domaines RNase III fonctionnels, il reste formellement possible qu'un seul des deux catalyse les deux clivages.

La première hypothèse suppose que des irrégularités de l'ARN double-brin affectent la position du clivage par Drosha et Dicer. Han *et al.*, 2006 ont recherché les déterminants de la position des sites de clivage par le Microprocessor (Drosha et son partenaire, appelé DGCR8 chez les Vertébrés et Pasha chez la Drosophile) : ce complexe reconnaît la base de la tige-boucle du pre-miRNA dans le pri-miRNA, et clive la tige à environ 11 bp de distance de son pied. Si le Microprocessor se fixait sur le sommet de la tige, il cliverait une autre position ; la meilleure affinité du complexe pour la base que pour le sommet garantit que seule la base sera reconnue. Les données de Han *et al.*, 2006 suggèrent la possibilité d'un système additionnel, dans la sélection du site de clivage : un mésappariement introduit à proximité du site de clivage inhibe la réaction. Cette observation montre donc que des irrégularités dans la structure de l'ARN double-brin peuvent affecter la position du site de clivage par Drosha.

L'article suivant (Seitz et Zamore, 2006) est une revue de présentation du contexte et des résultats de l'article de Han *et al.*, 2006, sur demande de l'éditeur.

3.4 Revue : structure secondaire et contrôle du clivage par Drosha

Leading Edge
Previews

Cell

Rethinking the Microprocessor

Hervé Seitz¹ and Phillip D. Zamore^{1,*}

¹Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA 01532, USA

*Contact: phillip.zamore@umassmed.edu

DOI 10.1016/j.cell.2006.05.018

MicroRNAs (miRNAs) are tiny regulators of gene expression that are processed from longer primary transcripts. In this issue, Han et al. (2006) report some of the structural features of the primary transcript that ensure that the Drosha-DGCR8 enzyme complex liberates precisely the correct precursor sequence, enabling production of a fully functional miRNA.

MicroRNAs (miRNAs) are single-stranded, 21–23 nucleotide RNAs that are able to repress specific target genes. They do this by base-pairing to target mRNAs, and then either accelerating degradation of the mRNA or inhibiting its translation. Base-pairing of miRNAs to their mRNA targets is often imperfect because miRNA nucleotides 2 through 7 have a disproportionate influence on target RNA selection. This “seed sequence” typically is exactly complementary to the target RNA, whereas the rest of the duplex tolerates imperfections in pairing. Because miRNA function requires such a remarkably small amount of complementarity to a target RNA, each miRNA species may regulate hundreds of distinct mRNA sequences. Given that the human genome may contain genes producing ~1,000 different miRNAs (Berezikov et al., 2005), it is possible that a majority of human genes are regulated posttranscriptionally by miRNAs (Lewis et al., 2005). Not surprisingly, miRNAs regulate a wide range of biological processes, including developmental timing, differentiation, apoptosis, insulin secretion, and even innate immunity against viruses.

Like mRNAs, miRNA genes are initially transcribed by RNA polymerase II as long primary transcripts (pri-miRNAs) that require subsequent processing to yield a functional mature miRNA (see Figure 1). In animals, pri-miRNAs are processed in the nucleus by the RNase III enzyme Drosha, acting with its double-stranded RNA binding partner protein DGCR8 (in verte-

brates) or Pasha (in invertebrates). This RNase III enzyme/dsRNA binding protein partnership converts pri-miRNAs into small stem-loop structures called precursor miRNAs (pre-miRNAs), which are then processed further by a second RNase III enzyme/dsRNA binding protein

duo into mature miRNAs. New work by Han et al. (2006) in this issue of *Cell* reveals some of the structural features of pri-miRNAs that encourage the Drosha-DGCR8 complex to liberate precisely the correct pre-miRNA, hence ensuring production of the correct miRNA sequence.

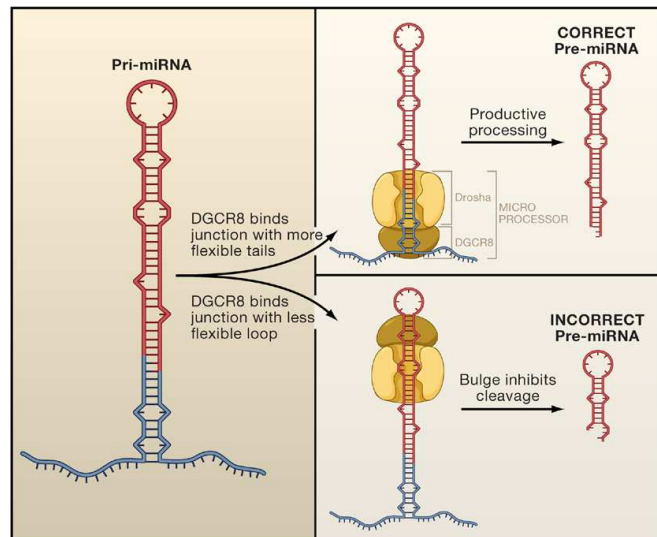


Figure 1. Pri-miRNA Processing by Drosha and DGCR8, Components of the Microprocessor

Structural features of pri-miRNAs promote their accurate processing into pre-miRNAs by the RNase III enzyme Drosha and its double-stranded RNA binding protein partner, DGCR8 (Pasha in invertebrates). DGCR8 is thought to bind more favorably to the junction between the rigid double-stranded stem and the 5' and 3' flexible, single-stranded segments of the pri-miRNA than to the junction between the stem and the considerably more constrained loop. Correct binding of DGCR8 to the base of the stem is proposed to position the processing center of Drosha ~11 bp up along the stem, where it makes a staggered pair of breaks in the RNA to create the ~65 nucleotide-long pre-miRNA. Binding of DGCR8 at the loop end of the stem positions Drosha inappropriately. Unpaired or weakly paired nucleotides at this site serve to discourage such unproductive cleavage, reducing the number of abortive Drosha products and favoring accurate pre-miRNA production.

The sequence of the miRNA is embedded in one of the two arms of the stem of a stem-loop structure within the pri-miRNA. The Drosha-DGCR8 heterodimer, a component of a large protein complex dubbed the "Microprocessor," cuts the stem loop containing the future miRNA out of the pri-miRNA (Denli et al., 2004; Gregory et al., 2004; Han et al., 2004; Landthaler et al., 2004). The resulting stem-loop structure, the pre-miRNA, is about ~65 nucleotides long and is exported to the cytoplasm. There, another RNase III enzyme, Dicer, aided by its own double-stranded RNA binding partner, cleaves the pre-miRNA approximately two helical turns, that is, ~21 base pairs, from the site of Drosha cleavage. Dicer cleavage liberates an imperfect duplex comprising the mature miRNA paired to the miRNA* strand, which derives from the other arm of the stem of the pre-miRNA. The arms of both the pri-miRNA and pre-miRNA stem-loop structures are imperfectly base-paired, containing G:U wobble pairs, single-nucleotide insertions, and outright mismatches. Consequently, the miRNA:miRNA* duplex contains imperfections in its pairing. These imperfections cause one strand of the duplex to be less stably paired at its 5' end (Khvorova et al., 2003; Schwarz et al., 2003)—this is the miRNA strand. In the process of loading the miRNA into the RNA binding pocket of a member of the Argonaute family of proteins, the miRNA is separated from the miRNA* strand and the miRNA* is degraded. At last, a functional miRNA-Argonaute protein complex is born.

The precision of Drosha-DGCR8 cleavage is crucial for the fidelity of miRNA maturation: if the position of the Drosha cut is shifted by a single nucleotide on the pri-miRNA, then Dicer cleavage, too, will be shifted, and the final miRNA will have different 5' and 3' ends. Now imagine that the single-nucleotide shift inverts the relative stabilities of the 5' end of the miRNA and of the miRNA* strand. In this event, the wrong strand may be chosen as the mature miRNA. Even if the relative stabi-

ties are preserved and the correct strand is loaded into a functional protein-RNA complex, the shift in the 5' end of the miRNA will reposition the seed sequence, redefining the set of mRNAs it regulates.

How then do Drosha and DGCR8-Pasha collaborate to excise pre-miRNA so precisely from pri-miRNA transcripts? As there is no strong sequence bias in pri-miRNAs, some structural feature of the RNA must determine the site of Drosha cleavage. The two most obvious distinguishing features are the extremities of the miRNA stem, that is, the flanking single-stranded RNA segments at the base of the hairpin and the terminal loop at its top.

Initially, Drosha was thought to cut the stem by measuring two helical turns from the loop (Zeng et al., 2005). Enter Han et al. (2006) in this issue with their report demonstrating that the terminal loop is unlikely to be the reference point for the molecular ruler that positions the site of cleavage. The terminal loop can be replaced by single-stranded RNA with no major effect on pri-miRNA processing, but the single-stranded RNA segments flanking the base of the stem are indispensable for Drosha cleavage. Han and colleagues show that deleting these single-stranded regions or converting them to double-stranded RNA by annealing a synthetic oligonucleotide to them greatly impairs the conversion of pri-miRNA to pre-miRNA (see also Zeng and Cullen, 2005). Modifying the length of the base of the stem also shifts the cleavage site. So it seems that the molecular ruler is anchored by the junction between the 5' and 3' single-stranded segments and the base of the double-stranded stem. Drosha or DGCR8-Pasha must then recognize this junction of single-stranded and double-stranded RNA and count up ~11 bp, one helical RNA turn, to the scissile phosphodiester bond. As DGCR8, but not Drosha, can be crosslinked to pri-miRNA, DGCR8-Pasha is the better candidate for the molecular ruler.

Yet, when Han et al. fed the Microprocessor an "inverted hairpin" in

which the single-stranded RNA segments at the base of the stem were replaced by a loop and the loop was replaced by two single-stranded RNA segments, they observed that the correct cleavage site was nonetheless selected. For this pri-miRNA, the Microprocessor appears to measure from the loop. What precisely does DGCR8-Pasha see if both single-stranded tails and a loop suffice to anchor the complex? One explanation is that the terminal loop and the basal single-stranded RNA segments are both unpaired, predicting that a highly structured loop—such as a GNRA tetraloop—would not anchor accurate processing. For a more open loop, DGCR8 may bind to the loop-to-stem junction positioning the Drosha processing center ~11 bp away.

How can DGCR8-Pasha differentiate between binding to the junction of the single-stranded RNA segments and the base of the stem junction, thereby accurately defining one end of the miRNA, and counterproductive binding to the junction of the loop and stem, which will promote abortive processing? Two determinants seem to favor productive binding. First, DGCR8 prefers the junction between flexible single-stranded RNA and a double-stranded stem; a small loop will always be more constrained than single-stranded RNA. Second, the authors' large-scale computational analysis of human and *Drosophila* pri-miRNAs suggests that most pri-miRNAs contain internal bulges or weakly paired bases ~11 bp from the loop-to-stem junction, that is, at abortive Drosha cleavage sites. Such sites may act to deter inappropriate cleavage by Drosha.

These results have several implications. The seemingly dual abilities of DGCR8 to bind to both single-stranded RNA and double-stranded RNA may allow it to bind cooperatively to the pri-miRNA. Cleavage by Drosha, which separates the two single-stranded segments from the double-stranded stem of the pre-miRNA, might then decrease the affinity of DGCR8 for all three reaction products, facilitating their release.



Computationally, our new understanding of how DGCR8 positions Drosha to promote productive pri-miRNA cleavage should help in the search for new miRNA genes and in the design of artificial miRNA genes. To date, successful algorithms for finding miRNAs have relied on phylogenetic conservation because simply searching the genome of a plant or animal for 65 nucleotide-long hairpins yields mainly false-positive results. The phylogenetic conservation approach is powerful but cannot find species-specific miRNAs and perhaps may not even find primate-specific miRNAs. The search for miRNAs that have not been well conserved through evolution may be facilitated by seeking sequences capable of folding into a structure

predicted to be bound by DGCR8-Pasha and to promote Drosha cleavage ~11 bp from the junction of a stem with single-stranded RNA tails (while discouraging Drosha cleavage ~11 bp from a terminal loop). That kind of algorithm may finally allow us to ask of the genomes of the Earth's animals: Are any of miRNAs different from yours?

REFERENCES

Berezikov, E., Guryev, V., van de Belt, J., Wienholds, E., Plasterk, R.H., and Cuppen, E. (2005). *Cell* 120, 21–24.

Denli, A.M., Tops, B.B., Plasterk, R.H., Ketting, R.F., and Hannon, G.J. (2004). *Nature* 432, 231–235.

Gregory, R.I., Yan, K.P., Amuthan, G., Chendrimada, T., Doratotaj, B., Cooch, N., and Shiekhattar, R. (2004). *Nature* 432, 235–240.

Han, J., Lee, Y., Yeom, K.H., Kim, Y.K., Jin, H., and Kim, V.N. (2004). *Genes Dev.* 18, 3016–3027.

Han, J., Lee, Y., Yeom, K.H., Kim, Y.K., Jin, H., Nam, J.W., Heo, I., Rhee, J.K., Sohn, S.Y., Cho, Y., et al. (2006). *Cell* 125, this issue.

Khvorovova, A., Reynolds, A., and Jayasena, S.D. (2003). *Cell* 115, 209–216.

Landthaler, M., Yalcin, A., and Tuschl, T. (2004). *Curr. Biol.* 14, 2162–2167.

Lewis, B.P., Burge, C.B., and Bartel, D.P. (2005). *Cell* 120, 15–20.

Schwarz, D.S., Hutvagner, G., Du, T., Xu, Z., Aronin, N., and Zamore, P.D. (2003). *Cell* 115, 199–208.

Zeng, Y., and Cullen, B.R. (2005). *J. Biol. Chem.* 280, 27595–27603.

Zeng, Y., Yi, R., and Cullen, B.R. (2005). *EMBO J.* 24, 38–48.

TFIIIC Boxes in the Genome

Lori L. Wallrath¹ and Pamela K. Geyer^{1,*}

¹Department of Biochemistry, University of Iowa, Iowa City, IA 52242, USA

*Contact: pamela-geyer@uiowa.edu

DOI 10.1016/j.cell.2006.05.016

In this issue of *Cell*, Noma et al. (2006) show that *B*-boxes and TFIIIC limit the spread of heterochromatin at the silent *mat* region in the fission yeast genome. Global analysis of TFIIIC distribution revealed dispersed sites of association that coalesce at the nuclear periphery, suggesting that TFIIIC may act as a barrier throughout the genome.

Eukaryotic chromosomes are divided into domains with distinct structural features (Wallrath et al., 2004). Heterochromatic domains are required for chromosome segregation and telomere maintenance as well as for suppressing recombination between repetitive elements. These domains encompass chromosomal regions that have few genes and are assembled into hypoacetylated, regularly spaced nucleosomal arrays containing the epigenetic mark of methylated lysine 9 of histone H3 (H3K9me). Heterochromatic histone modifications generate a condensed chromatin structure that limits the access

of transcription factors to target sequences. Interspersed with these transcriptionally repressive domains are euchromatic regions that are gene rich and organized into hyperacetylated irregular arrays of nucleosomes enriched in histone H3 methylated at lysine 4 (H3K4me). Such euchromatic modifications establish chromatin packaging that is accessible to transcription factors. The ability of heterochromatin to propagate in *cis* implies that mechanisms are needed for limiting the spread of silent chromatin. A class of DNA elements, known as insulators, has been implicated in defining the junc-

tions between structural domains (Kuhn and Geyer, 2003). Insulators are found in most eukaryotes, suggesting that these elements have a conserved role in organizing transcriptional domains. Two classes of insulators have been identified that differentially affect transcriptional processes (Kuhn and Geyer, 2003). Enhancer blockers are insulators that prevent enhancer-dependent transcription when placed between an enhancer and promoter. In contrast, barriers are insulators that impede the spread of heterochromatin emanating from an initiation site. In this issue of *Cell*, Noma et al. (2006)

3.5 Discussion : la structure secondaire des pri-miRNA, un module régulateur

Les travaux de Han *et al.*, 2006 montrent donc que la structure secondaire des pri-miRNA détermine le site de clivage par Drosha, en recrutant le Microprocessor préférentiellement à la base de la tige-boucle, et éventuellement, en modulant son activité par des mésappariements.

Les données de Förstemann *et al.*, 2007 et Tomari *et al.*, 2007 montrent quant à elles que la structure secondaire du produit de Dicer (le duplex miRNA/miRNA*) affecte également les étapes tardives de la biogenèse chez la Drosophile : un mésappariement autour du neuvième nucléotide du brin guide dirige le duplex vers la machinerie de chargement d'Ago1, alors qu'un appariement de ces positions le dirige vers la machinerie de chargement d'Ago2 (voir « Chargement sur Argonaute : le triage des petits ARN », page 21).

L'analyse biochimique du chargement du duplex sur Ago1 permet de distinguer deux effets de sa structure secondaire : d'une part, le mésappariement autour du nucléotide 9 facilite la prise en charge du duplex par la machinerie de chargement ; d'autre part, des mésappariements additionnels, dans la graine ou dans la moitié 3' du guide, permettent l'ouverture du duplex et la perte du brin passager (un processus appelé « maturation » du complexe RISC). L'ouverture du duplex est en effet un mécanisme passif, qui ne nécessite ni hydrolyse d'ATP, ni activité endonucléolytique de la protéine Ago1.

Ces résultats sont présentés dans l'article qui suit (Kawamata *et al.*, 2009).

3.6 Article n°4 : caractéristiques du chargement des microARN

ARTICLES

nature
structural &
molecular biology

Structural determinants of miRNAs for RISC loading and slicer-independent unwinding

Tomoko Kawamata^{1,2}, Hervé Seitz^{3,4} & Yukihide Tomari^{1,2,5}

MicroRNAs (miRNAs) regulate expression of their target mRNAs through the RNA-induced silencing complex (RISC), which contains an Argonaute (Ago) family protein as a core component. In *Drosophila melanogaster*, miRNAs are generally sorted into Ago1-containing RISC (Ago1-RISC). We established a native gel system that can biochemically dissect the Ago1-RISC assembly pathway. We found that miRNA-miRNA* duplexes are loaded into Ago1 as double-stranded RNAs in an ATP-dependent fashion. In contrast, unexpectedly, unwinding of miRNA-miRNA* duplexes is a passive process that does not require ATP or slicer activity of Ago1. Central mismatches direct miRNA-miRNA* duplexes into pre-Ago1-RISC, whereas mismatches in the seed or guide strand positions 12–15 promote conversion of pre-Ago1-RISC into mature Ago1-RISC. Our findings show that unwinding of miRNAs is a precise mirror-image process of target recognition, and both processes reflect the unique geometry of RNAs in Ago proteins.

At the core of RISCs are Ago family proteins^{1–3}. Each small RNA species often binds to a specific Ago protein. In *Drosophila*, miRNAs and siRNAs are actively sorted into Ago1 and Ago2 complexes, respectively^{4–6}. The position of mismatches is a key determinant for this small-RNA partitioning. Typical miRNAs, which usually have mismatches in the central region of their miRNA-miRNA* duplexes, are primarily loaded into Ago1, whereas highly complementary exogenous and endogenous siRNAs primarily bind Ago2 (refs. 4–8). A small-RNA sorting system may exist in mammals, although it seems to be less stringent than those in other animals or plants^{9,10}. The association of a small RNA with a specific Ago protein dictates its function. Of the four mammalian Ago-subfamily proteins (Ago1–4), only Ago2 possesses target cleavage ('slicer') activity^{11,12}. In flies, both Ago1 and Ago2 have slicer activity; however, Ago1 is a much weaker endonuclease than is Ago2 (ref. 5).

RISC assembly follows an ordered, complex pathway². In flies, the RISC-loading complex, which contains Dicer-2 (Dcr-2) and R2D2, has an essential role in loading small RNA duplexes into Ago2-RISC^{13–15}. Dcr-2–R2D2 also acts as a gatekeeper in the Ago2 pathway, incorporating siRNA-like perfectly complementary small RNA duplexes and excluding miRNA-like small RNA duplexes⁶. In contrast, how the Ago1-RISC is assembled remains largely unknown. Dicer-1 (Dcr-1) and its partner protein Loquacious (Loqs)^{16,17} might have a role in Ago1-RISC loading, by analogy with the function of Dcr-2–R2D2 in Ago2-RISC assembly. In fact, lysate from *dcr-1* null eggs (but not lysate from *loqs*^{KO} eggs) lacks Ago1-mediated target cleavage activity¹⁸. However, *dcr-1* lysate cannot be rescued by recombinant Dcr-1 or Dcr-1–Loqs heterodimer¹⁸, raising the possibility that the defect in the *dcr-1* lysate results indirectly from the wide-ranging developmental

abnormalities of *dcr-1* eggs¹⁹. Regardless of a potential role for Dcr-1 in Ago1-RISC loading, it has been implied that an unidentified independent mechanism, in parallel with Dcr-2–R2D2, favors assembly of miRNA-like duplexes into Ago1-RISC⁶.

Both the miRNA-miRNA* duplex and siRNA duplex need to be unwound to act as a single-stranded guide in RISC to recognize their target mRNAs. It was originally proposed that an ATP-dependent helicase ('unwindase') separates the two small RNA strands, and the resulting single-stranded guide is then loaded into Ago proteins²⁰. However, it was later shown that *Drosophila* Ago2 (refs. 21–23), as well as human Ago2 (ref. 24), directly receive double-stranded small RNA from the RISC-loading complex. Ago2 then cleaves the passenger strand, thereby liberating the single-stranded guide to form mature Ago2-RISC. However, the cleavage activity of fly Ago1 is very inefficient⁵, and mammalian Ago1, Ago3 and Ago4 show no cleavage activity^{11,12}. Thus, it remains unknown when (before or after RISC loading) and, more importantly, how the two strands are separated in such 'slicer-independent' RISC assembly pathways.

Here we describe a native gel system that can biochemically dissect the fly Ago1-RISC assembly pathway. We show that miRNA-miRNA* duplexes are loaded into Ago1 as double-stranded RNAs in an ATP-dependent fashion. In contrast, unwinding requires neither ATP nor the slicer activity of Ago1. Moreover, we identified the structural determinants of small RNA duplexes for Ago1-RISC loading and subsequent separation of the two strands. Our data collectively show that slicer-independent unwinding is a precise mirror-image process of target recognition, and both processes reflect the unique geometry of RNAs in Ago proteins.

¹Institute of Molecular and Cellular Biosciences, The University of Tokyo, Japan. ²Department of Medical Genome Sciences, The University of Tokyo, Bunkyo-ku, Tokyo, Japan. ³Université de Toulouse, UPS, Laboratoire de Biologie Moléculaire Eucaryote, Toulouse, France. ⁴CNRS, Laboratoire de Biologie Moléculaire Eucaryote, Toulouse, France. ⁵PRESTO, Japan Science and Technology Agency, Kawaguchi-shi, Saitama, Japan. Correspondence should be addressed to Y.T. (tomari@iam.u-tokyo.ac.jp).

Received 23 February; accepted 3 June; published online 16 August 2009; doi:10.1038/nsmb.1630



ARTICLES

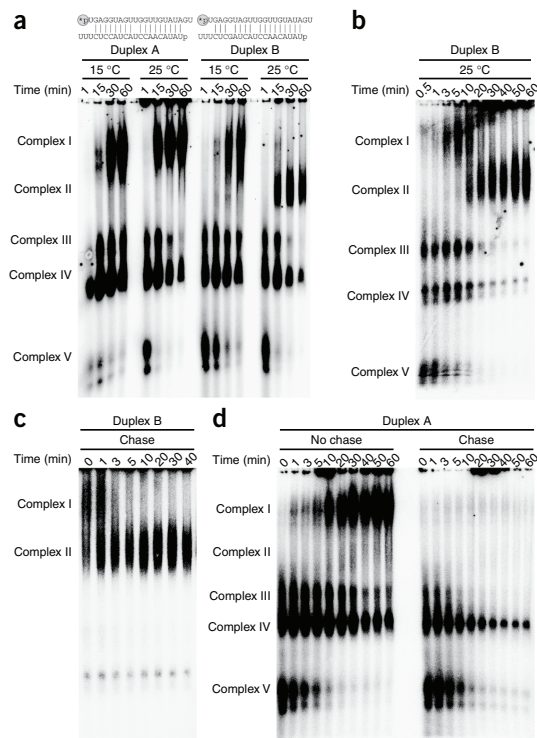


Figure 1 Identification of small RNA-protein complexes in *dcr-2* embryo lysate. **(a)** Native gel analysis in *dcr-2* embryo lysate revealed five complexes (I–V). Lysate and 5' guide-radiolabeled duplexes were incubated at 15 °C or 25 °C together with nonradiolabeled target mRNA. Complexes assembled at the indicated time points were then separated on a vertical agarose native gel. Structure of duplexes A and B are shown; both duplexes contained the identical guide sequence (above). Asterisk denotes a 5' ³²P group. **(b)** Time course of the complexes formed with duplex B at 25 °C, as in **a**. **(c)** Complex I is a precursor of complex II. Duplex B was incubated with lysate at 15 °C for 60 min, 10-fold excess of nonradiolabeled duplex B was added (time = 0), and 'chase' of complex I into complex II was monitored. **(d)** Complexes III, IV and V are not directly involved in the formation of complexes I and II. Duplex A was preincubated with lysate at 25 °C for 1 min, 10-fold excess of nonradiolabeled duplex A was added (time = 0), and complex formation was monitored.

in lysate for 60 min at 15 °C. Subsequently, we added a 10-fold excess of nonradiolabeled duplex B to prevent further incorporation of the radiolabeled duplex B into complex I. At the same time, we shifted the temperature to 25 °C and monitored complex formation. Within 5 min, complex I was completely chased into complex II (Fig. 1c), indicating that complex I is a precursor of complex II. We next formed complexes III–V by incubating radiolabeled duplex A for only 1 min at 25 °C and then adding a 10-fold excess of nonradiolabeled duplex. Complexes III–V were not chased into complex I or II (Fig. 1d), suggesting that complexes III–V are irrelevant to complex I or II. Single-stranded *let-7*, which was previously tested in a native gel system using *Caenorhabditis elegans* lysate²⁶, formed complex V at the earliest time point but immediately degraded without forming any other complexes (Supplementary Fig. 1).

Complexes I and II contain Ago1

To identify the proteins bound to the duplexes, we incubated duplexes A or B at 25 °C and photo-cross-linked the reaction mixture with 254-nm UV light. We separated each complex on a native agarose gel, excised it from the gel, and analyzed the proteins covalently linked to the radiolabeled duplexes by SDS-PAGE (Fig. 2a,b). We detected only a ~110-kDa protein in complexes I and II (Fig. 2b), and we specifically immunoprecipitated this cross-linked band using antibody to Ago1 (ref. 6; data not shown). Moreover, in a *dcr-2* lysate where ~99% of endogenous Ago1 was immunodepleted, complexes I and II were barely detectable, confirming that complexes I and II contain Ago1 (Fig. 2c). We also observed ~250-kDa and ~80-kDa proteins in complex IV (Fig. 2b), which we did not attempt to identify because complex IV is not involved in the assembly of the Ago1-containing complexes (Fig. 1d).

Complexes I and II are pre- and mature Ago1-RISC, respectively

As both complexes I and II contain Ago1, we hypothesized that complex I represents pre-Ago1-RISC, which contains double-stranded small RNA, and that complex II represents mature Ago1-RISC, which contains single-stranded small RNA. Supporting this hypothesis, when we radiolabeled the passenger strand instead of the guide strand, we detected complex I but not complex II (Supplementary Fig. 2a). Moreover, when we incubated guide-radiolabeled duplex A or duplex B in lysate, immunoprecipitated Ago1 and analyzed the Ago1-bound small RNAs, we observed that duplex A, which mainly forms complex I, remained essentially double-stranded, whereas duplex B, which mainly forms complex II, became considerably more single-stranded (Fig. 2d).

To directly determine whether the complexes contain double-stranded or single-stranded small RNAs, we separated and excised complexes from native agarose gels. We then extracted the complex-bound RNA by deproteinizing each gel slice using proteinase K,

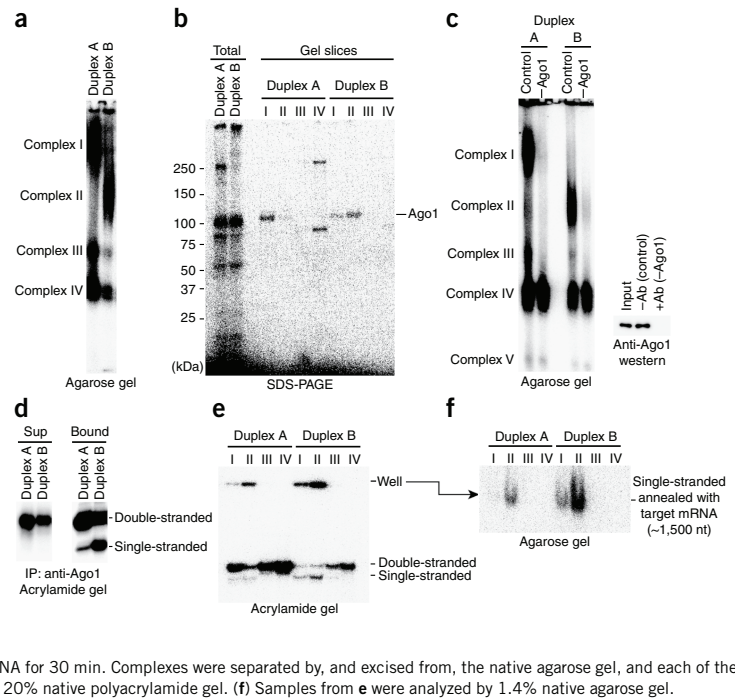
RESULTS

Intermediate complexes in the Ago1-RISC assembly pathway

To better understand Ago1-RISC assembly, we first sought to biochemically identify complexes formed in the Ago1-RISC assembly pathway. *Dcr-2*, which is essential for Ago2-RISC loading, is dispensable in the Ago1-RISC assembly pathway^{16,17,25}. We therefore used an embryo lysate prepared from *dcr-2*-null mutant flies. We prepared two small RNA duplexes, duplex A and duplex B, which were derived from the natural *let-7-let-7** miRNA duplex (Fig. 1a). Both duplexes contained an identical guide-strand sequence and a central mismatch at position 10. Duplex B bore an additional mismatch in the passenger strand across from guide position 5 in the seed region, reminiscent of typical miRNA-miRNA* duplexes. We incubated the duplexes radiolabeled with ³²P at the 5' end with lysate at 25 °C, together with a nonradiolabeled ~1,500-nucleotide (nt) target mRNA harboring target sites complementary to the guide strand with a central bulge. We then separated the complexes on a vertical agarose native gel (Fig. 1a and Supplementary Fig. 1). We detected five distinct ribonucleoprotein complexes, which we tentatively named complexes I–V. For both of the duplexes, complexes III–V appeared almost immediately and then slowly diminished. At later time points (>15 min), duplex A formed only complex I, whereas duplex B formed mostly complex II. At 15 °C, both duplex A and duplex B formed only complex I (Fig. 1a).

To determine the relationship between these complexes, we monitored the kinetics of complex formation more intensively. We detected complex I even with duplex B, at very early time points (<10 min) at 25 °C (Fig. 1b). Complex II began to form as complex I diminished. These observations imply that complex I is a precursor of complex II. To directly test this hypothesis, we performed 'chase' experiments. We first assembled complex I by incubating radiolabeled duplex B

Figure 2 Complex I is pre-Ago1-RISC, and complex II is mature Ago1-RISC. (a) Lysate and 5'-radiolabeled duplex A or duplex B were incubated at 25 °C for 30 min as in **Figure 1a**, photo-cross-linked with 254-nm UV light for 5 min and separated on a native agarose gel. (b) Each complex (I–IV) was excised from the gel. The proteins photo-cross-linked with duplex A or B were eluted by boiling in SDS loading buffer for 5 min and subjected to 5–20% SDS-PAGE. The ~110-kDa protein detected in complexes I and II was determined to be Ago1. 'Total' indicates photo-cross-linked samples before separation by the native gel. (c) Depletion of Ago1 eliminated the formation of complexes I and II. Native gel analysis of the complexes formed in Ago1-immunodepleted, *dcr-2* embryo lysate is shown. Input, supernatant after Ago1 immunodepletion (+Ab) and control (-Ab) are indicated. Efficiency of Ago1 immunodepletion was estimated by western blotting. (d–f) Complex I contains double-stranded small RNA, whereas complex II contains single-stranded small RNA. (d) Lysate and 5'-radiolabeled duplex A or B were incubated without target mRNA. Ago1 was immunoprecipitated, and Ago1-bound small RNAs were analyzed by 15% polyacrylamide native gel. Sup, supernatant after Ago1 immunoprecipitation; bound, immunoprecipitated Ago1 beads. (e) Lysate and 5'-radiolabeled duplex A or B were incubated together with target mRNA for 30 min. Complexes were separated by, and excised from, the native agarose gel, and each of the complex-bound RNAs was extracted and separated by 20% native polyacrylamide gel. (f) Samples from **e** were analyzed by 1.4% native agarose gel.



and we separated the RNAs using native polyacrylamide gels. Complex I mostly contained double-stranded small RNA, whereas complex II contained single-stranded small RNA together with another band that barely migrated in the gel (**Fig. 2e**). To analyze this slow-migrating band, we separated the same RNA samples on agarose gels. We detected a band in complex II that corresponded to the single-stranded guide annealed to the target mRNA (**Fig. 2f**). Target RNAs of various lengths affected the mobility of complex II but not that of complex I (**Supplementary Fig. 2b**), suggesting that only complex II can recognize the target RNA. We concluded that complex I is pre-Ago1-RISC, and complex II is mature Ago1-RISC.

Dcr-1 and GW182 are dispensable for Ago1-RISC assembly

Dcr-1, which processes pre-miRNA into the miRNA-miRNA* duplex²⁵, and GW182, a P-body component required for miRNA-directed translational repression^{27,28}, are known to associate with Ago1. We assessed whether these proteins are required for Ago1-RISC assembly. We depleted >95% of endogenous Dcr-1 or GW182 in S2 cells by RNA interference (**Supplementary Fig. 3a**). Dcr-1 knockdown abolished pre-miRNA processing activity (**Supplementary Fig. 3b**), and GW182 knockdown markedly impaired Ago1-mediated translational repression (**Supplementary Fig. 3c**). However, depletion of Dcr-1 or GW182 did not affect Ago1-mediated target cleavage (**Supplementary Fig. 3d**) or formation of pre- and mature Ago1-RISC (**Supplementary Fig. 3e**). We concluded that these proteins are dispensable for Ago1-RISC assembly *per se*, although upstream and downstream steps require Dcr-1 and GW182, respectively.

Ago1-RISC loading, but not unwinding, requires ATP

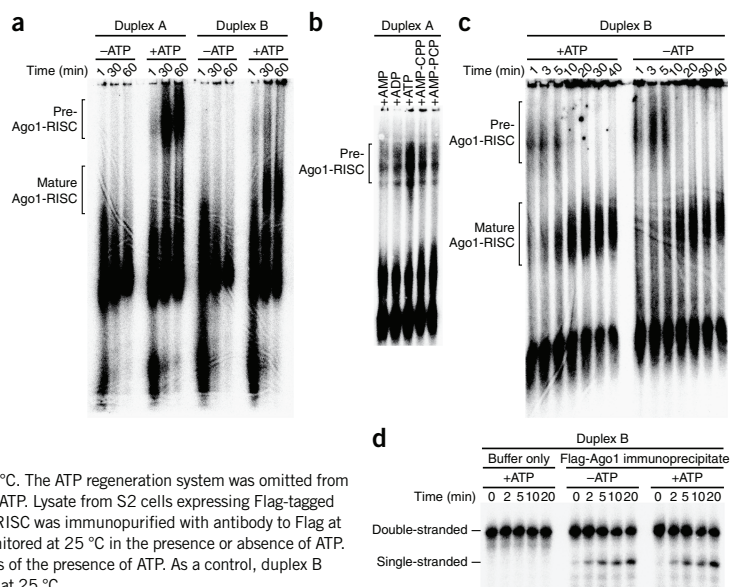
We next determined whether ATP is required for the formation of pre-Ago1-RISC or mature Ago1-RISC. When we depleted endogenous

ATP using hexokinase and glucose, we observed neither pre-Ago1-RISC nor mature Ago1-RISC (**Fig. 3a**), indicating that loading of small RNA duplexes into pre-Ago1-RISC requires ATP. Addition of AMP, ADP or nonhydrolyzable ATP analogs did not promote pre-Ago1-RISC formation, indicating that hydrolysis of ATP is essential for Ago1-RISC loading (**Fig. 3b**). We then incubated lysates with radiolabeled duplex B at 15 °C to form pre-Ago1-RISC, depleted ATP at 15 °C, added 20-fold excess nonradiolabeled duplex B and shifted the reaction temperature to 25 °C. Pre-Ago1-RISC was chased into mature Ago1-RISC in the absence of ATP, much as in the presence of ATP (**Fig. 3c**). The identical ATP-depletion protocol at 15 °C completely blocked pre-Ago1-RISC formation of duplex A (**Supplementary Fig. 4**). These observations indicate that formation of pre-Ago1-RISC requires ATP, whereas its conversion into mature Ago1-RISC occurs independent of ATP.

To confirm this, we prepared lysate from S2 cells expressing Flag-tagged Ago1, incubated the lysate with duplex B at 15 °C in the presence of ATP, shifted the temperature to 4 °C to block RISC loading and unwinding, and immunoprecipitated pre-Ago1-RISC using an antibody to Flag. We washed the immunoprecipitate intensively at 4 °C using a high-salt (800 mM NaCl) buffer to remove ATP and most of the Ago1-associated proteins, then monitored unwinding at 25 °C in the presence or absence of ATP. Regardless of the presence of ATP, duplex B in pre-Ago1-RISC was efficiently unwound on the beads, whereas the duplex remained completely double-stranded in lysis buffer only, confirming that ATP is dispensable for unwinding (**Fig. 3d**). These data suggest that Ago1 itself can unwind small RNA duplexes in an ATP-independent fashion, although we cannot exclude the possibility that protein(s) coimmunoprecipitated with Ago1 facilitate unwinding.

ARTICLES

Figure 3 ATP is essential for Ago1-RISC loading but dispensable for unwinding. **(a)** Depletion of ATP blocks small-RNA loading into Ago1. Endogenous ATP was depleted by 0.15 U μl^{-1} hexokinase and 20 mM glucose at 25 °C for 20 min (-ATP, <100 nM ATP). In the mock depletion, no hexokinase was added (+ATP, 1 mM ATP). RISC assembly was done as in **Figure 1a**. **(b)** Hydrolysis of ATP is essential for pre-Ago1-RISC formation. RISC assembly was done using duplex A as in **Figure 1a**, at 25 °C in the presence of 2.5 mM AMP, ADP, ATP, AMP-CPP or AMP-PCP. A basal level of endogenous ATP in *dcr-2* embryo lysate (~10 μM) was present. **(c)** Conversion of pre-Ago1-RISC to mature Ago1-RISC does not require ATP. RISC assembly was done as in **Figure 1a**, at 15 °C for 30 min in the presence of 1 mM ATP to permit pre-Ago1-RISC formation. ATP was depleted from the reaction at 15 °C for 30 min by 0.3 U μl^{-1} hexokinase and 20 mM glucose. A 20-fold excess of nonradiolabeled duplex B was added, the reaction temperature was immediately shifted to 25 °C (time = 0) and complex formation was monitored at 25 °C. The ATP regeneration system was omitted from the assays in **a-c**. **(d)** Unwinding requires Ago1 but not ATP. Lysate from S2 cells expressing Flag-tagged Ago1 was incubated with duplex B at 15 °C, pre-Ago1-RISC was immunopurified with antibody to Flag at 4 °C, and unwinding in the immunoprecipitate was monitored at 25 °C in the presence or absence of ATP. Unwound single-stranded guide was detected regardless of the presence of ATP. As a control, duplex B stayed completely double-stranded in lysis buffer alone at 25 °C.



Central mismatches direct Ago1-RISC loading

As shown above, although both duplexes A and B can enter pre-Ago1-RISC, only duplex B can form mature Ago1-RISC. This prompted us to examine how the structure of duplexes affects pre- and mature Ago1-RISC formation. We prepared 17 small RNA duplexes: a functionally asymmetric *let-7* siRNA duplex, in which the guide and passenger strands were fully paired except at guide position 1 (duplex mm1) and 16 duplex derivatives bearing one additional mismatch at every guide position (duplexes mm2 to mm17; **Supplementary Table 1**). At 15 °C (at which little mature Ago1-RISC is formed), only duplexes mm8 to mm12, peaking at mm10, efficiently formed pre-Ago1-RISC (**Fig. 4a,b**), indicating that central mismatches are essential for the formation of pre-Ago1-RISC.

Seed and 3'-mid mismatches promote unwinding

We next prepared another series of 17 duplexes (duplexes MM10/mm1 to MM10/mm17; **Supplementary Table 1**) that were identical to duplexes mm1 to mm17, except that an additional mismatch was introduced at guide position 10 to promote pre-Ago1-RISC formation. At 25 °C, pre-Ago1-RISC containing MM10/mm1 and MM10/mm9 to MM10/mm11 remained largely unconverted, whereas pre-Ago1-RISC containing MM10/mm2 to MM10/mm8 and MM10/mm12 to MM10/mm15 were efficiently converted into mature Ago1-RISC (**Fig. 4c-e**). Thus, seed and 3'-mid (guide positions 12–15) mismatches promote unwinding. We observed that mismatches near the 3' end of the radiolabeled MM10 strand (MM10/mm16 and MM10/mm17) inverted the strand asymmetry of the small RNA duplexes^{29,30} and directed the nonradiolabeled strand into mature Ago1-RISC. A target cleavage assay measuring the amount of mature Ago1-RISC in *dcr-2* lysate confirmed the preference of the seed and 3'-mid mismatches in the (original) mm and MM10/mm series (**Supplementary Fig. 5a,b**). Notably, in the mm series, duplexes with a central mismatch (mm9 to mm11) formed more pre-Ago1-RISC (**Fig. 4a**), yet their target cleavage activity was lower than that of duplexes with a mismatch in the seed or 3'-mid regions (**Supplementary Fig. 5a**), which

reinforces the requirement of seed or 3'-mid mismatches for Ago1-RISC maturation. Coincidentally, seed matches are not always sufficient for miRNA-mediated gene regulation, and contiguous pairing in the 3'-mid region (positions 12–16) substantially enhances the efficacy³¹ (**Fig. 4f**).

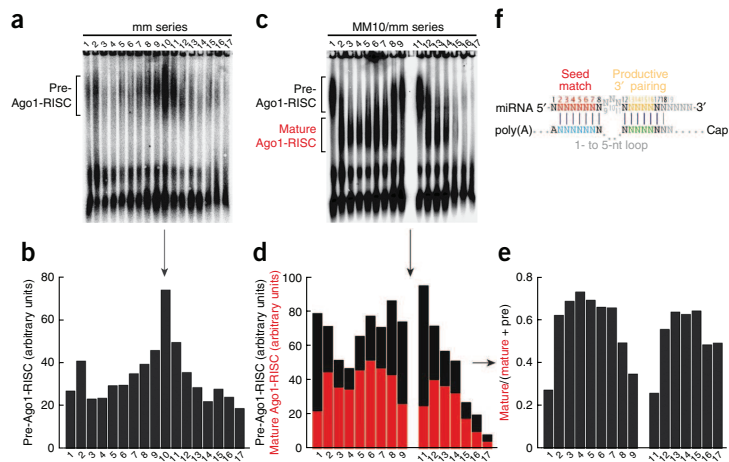
Position of mismatches is measured from 5' end of the guide

Given that the position of a mismatch is crucial to Ago1-RISC loading and maturation, we sought to determine from where the position is measured. To that end, we prepared a series of longer (25-nt) duplexes (**Supplementary Table 1**, 25-nt mm and MM10/mm series) and assayed for Ago1-RISC loading and unwinding. As in the 21-nt duplex series, mismatches at positions 7–11 of the 25-nt duplexes directed the formation of pre-Ago1-RISC (**Supplementary Fig. 6a,b**), and mismatches in the seed or 3'-mid regions promoted unwinding (**Supplementary Fig. 6c-e**). Target cleavage assays (**Supplementary Fig. 5c,d**) also confirmed these observations. We concluded that the position of mismatch is measured from the 5' end of the guide, not the 'center' of the duplex or the 3' end of the guide, for both Ago1-RISC loading and maturation.

G:U wobbles behave like mismatches in Ago1-RISC assembly

G:U wobbles between the seed region of an miRNA and its target mRNA significantly hinder miRNA-mediated translational repression, even though they are as thermodynamically favorable as the A:U base-pair^{31–33}. We introduced a G:U wobble at guide positions 5, 9 or 15 of a perfectly complementary duplex (GU5, GU9 or GU15; **Supplementary Table 1**). The G:U wobble in the central region (GU9), but not those in the seed or 3'-mid regions (GU5 or GU15), enhanced Ago1-RISC formation, similar to the mismatch (**Supplementary Fig. 7a**). We then prepared MM10 duplex derivatives containing an A:U base pair, a G:U wobble, or a U:U mismatch at position 5 or 13 (**Supplementary Table 1**). A G:U wobble in the seed or 3'-mid region promoted the conversion of pre-Ago1-RISC into mature Ago1-RISC, similar to a mismatch (**Supplementary Fig. 7b**). These results indicate

Figure 4 Central mismatches direct small RNA duplexes into pre-Ago1-RISC, whereas seed or 3'-mid mismatches promote conversion of pre-Ago1-RISC to mature Ago1-RISC. (a) The 21-nt mm series (mm1 to mm17) was used to assemble pre-Ago1-RISC at 15 °C. (b) Quantification of pre-Ago1-RISC in a. (c) The 21-nt MM10-mm series (MM10-mm1 to MM10-mm17) was used to monitor the conversion from pre-Ago1-RISC to mature Ago1-RISC at 25 °C. (d) Quantification of pre-Ago1-RISC and mature Ago1-RISC in c. (e) Efficiency of strand separation was calculated as (mature Ago1-RISC) / (mature Ago1-RISC + pre-Ago1-RISC). These native gel experiments were repeated five times with excellent reproducibility, and representative data are shown. (f) Representation of determinants for efficient target recognition by miRNAs, as reported in ref. 31. Productive 3' pairing, in addition to seed pairing, promotes efficient target recognition³¹. Preference for mismatches in the seed or 3'-mid regions for Ago1-RISC maturation observed in c–e is the reverse of the preference for complete base-pairing in those regions for target recognition by miRNAs.



that G:U wobble base pairs within miRNA duplexes behave similar to mismatches, for both Ago1-RISC loading and unwinding.

Unwinding does not require the slicer activity of Ago1

Ago1 retains slicer activity⁴, albeit far less efficient than that of Ago2 (ref. 5). In the Ago2 pathway, cleavage of the passenger strand facilitates the maturation of Ago2-RISC (refs. 21–23). We examined whether the slicer activity of Ago1 is required for Ago1-RISC maturation. Slicer-competent Ago proteins contain the highly conserved Asp-Asp-His (DDH) motif that is required for endonucleolytic activity^{11,34,35}. We expressed Flag-tagged Ago1 proteins, in which each residue in the DDH motif was replaced with alanine, in S2 cells. Any mutation in this motif abolished the target cleavage activity of Ago1 (Fig. 5a). In contrast, when we immunoprecipitated the tagged Ago1 and separated

the Ago1-bound small RNAs on the gel, we detected the same amount of single-stranded guide strand in the mutants as in the wild type (Fig. 5b). We then conducted native gel analysis of lysates prepared from cells overexpressing wild-type or mutant Flag-tagged Ago1, in which *dcr-2* was knocked down to eliminate the Ago2 pathway. We detected the same amount of mature Ago1-RISC in the mutants as in the wild type (Fig. 5c,d). We concluded that unwinding is independent of the slicer activity of Ago1.

Structures of natural miRNA-miRNA* duplexes

Given the identified structural determinants for Ago1-RISC loading and Ago1-RISC maturation, natural miRNA-miRNA* duplexes should meet such criteria. We used publicly available deep-sequencing datasets of fly miRNAs^{36–39}, which accurately describe the real miRNA sequences, and predicted the sequences of the miRNA-miRNA* duplexes by 'conceptual dicing' (see Methods). We counted the miRNAs with at least one mismatch in a sliding 3-nt window and observed a significant tendency of mismatches at the

© 2009 Nature America, Inc. All rights reserved.

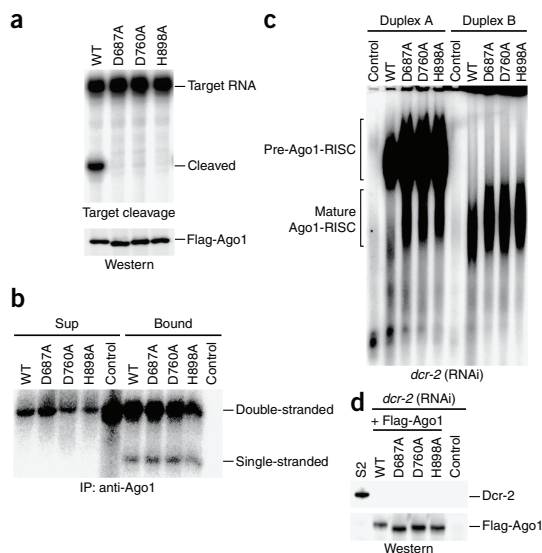
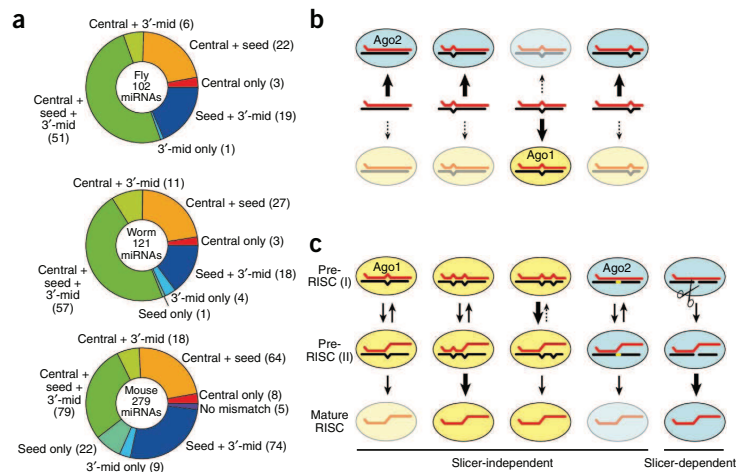


Figure 5 Strand separation of small RNA duplex does not require the slicer activity of Ago1. (a) Mutation in the DDH motif abolish target cleavage activity of Ago1. Lysates from S2 cells expressing Flag-tagged wild-type (WT) Ago1 or its mutant derivatives were incubated with nonradiolabeled duplex B. Flag-tagged Ago1 was then immunoprecipitated with antibody to Flag. After extensive washing, target cleavage was monitored. Expression of Flag-tagged Ago1 proteins was confirmed by western blotting (below) using antibody to Flag. (b) Mutations in the DDH motif do not affect strand separation of the small RNA duplex. The 5'-radiolabeled duplex B was incubated with lysates from S2 cells expressing Flag-tagged WT Ago1 or its mutant derivatives, and Flag-Ago1 was immunoprecipitated. Ago1-bound RNA was extracted and separated by 20% native polyacrylamide gel. S2 cells that did not express Flag-Ago1 were used as a control. Sup, supernatant after Ago1 immunoprecipitation; bound, immunoprecipitated Ago1-beads. (c) Ago1 DDH mutants formed both pre-Ago1-RISC and mature Ago1-RISC. Native gel analysis was done using lysate from S2 cells in which Dcr-2 was knocked down, and WT or mutant Flag-Ago1 proteins were expressed. Lysate from S2 cells in which Dcr-2 was knocked down but no Flag-Ago1 was expressed was used as a control. (d) Expression of Flag-tagged Ago1 proteins and Dcr-2 knockdown in lysates used in c were confirmed by western blotting.

ARTICLES

Figure 6 Structural determinants for Ago1-RISC loading and strand separation. (a) Structural signature of natural miRNA-miRNA* duplexes in *D. melanogaster*, *C. elegans* and *M. musculus*. Most miRNA-miRNA* duplexes show mismatches in the central region as well as in the seed and/or 3'-mid regions, which ensure efficient RISC loading and unwinding. miRNAs were categorized according to the region(s) where at least one mismatch exists. Seed, guide positions 2–8; central, guide positions 9–11; 3'-mid, guide positions 12–15. (b) Efficient Ago1-RISC loading requires central mismatches, which complements the rejection of central mismatches in Ago2-RISC loading⁶. Guide strands are shown in red; passenger strands are shown in black. Solid and dashed arrows indicate strong and weak bindings, respectively. (c) Seed and 3'-mid mismatches promote conversion of pre-Ago1-RISC into mature Ago1-RISC, which can be explained by postulating two states of pre-Ago1-RISC: pre-RISC (I), where both the seed and 3' half



regions of the guide can anneal to the passenger, and pre-RISC (II), where the base-pairings are limited to the seed region of the guide (see Discussion). Ago2 usually liberates the guide strand by cleavage of the passenger strand. However, when passenger cleavage is blocked by chemical modification (shown in yellow), a slower 'bypass' mechanism still dissociates the passenger strand²², which we assume is reminiscent of the slicer-independent unwinding by Ago1. Thick solid, thin solid and dashed arrows indicate fast, moderate and slow reactions, respectively.

5' end, as well as in positions 9–12 (**Supplementary Fig. 8a**), indicating that natural miRNA-miRNA* duplexes tend to have mismatches in the central region (**Fig. 4a,b**). We similarly analyzed miRNA-miRNA* duplexes predicted from *C. elegans*^{40,41} and *Mus musculus*^{42–45} miRNA deep-sequencing datasets. Whereas worm miRNA-miRNA* duplexes showed a marked frequency of central mismatches (**Supplementary Fig. 8b**), mouse miRNA-miRNA* duplexes showed only a modest central peak (**Supplementary Fig. 8c**), which probably reflects the looseness of small-RNA sorting in mammals^{3,10}. We then categorized each miRNA-miRNA* duplex according to the region(s) where at least one mismatch exists (**Fig. 6a**). Of 102 fly miRNAs, 82 (80%) have at least one mismatch in the central region, among which 79 (96%) also have at least one mismatch either in the seed or in the 3'-mid region. Similarly, 81% of worm miRNAs and 61% of mouse miRNAs have at least one mismatch in the central region; among those, 97% and 58%, respectively, also have at least one mismatch either in the seed or the 3'-mid region. Thus, structures of natural miRNA-miRNA* duplexes tend to be optimal for both RISC loading and unwinding.

DISCUSSION

Our data indicate that pre-Ago1-RISC formation requires central mismatches, whereas Ago1-RISC maturation requires mismatches in the seed or 3'-mid regions (**Fig. 4**). Supporting this, our computational analyses show that natural miRNAs tend to have optimal secondary structures of their miRNA-miRNA* duplexes for both RISC loading and maturation (**Fig. 6a** and **Supplementary Fig. 8**). Notably, the preference for central mismatches in Ago1-RISC loading is a perfect complement to the rejection of central mismatches in Ago2-RISC loading⁶, thus ensuring accurate sorting of small RNAs in flies (**Fig. 6b**).

Dicer contains a DExD/H-box helicase domain, and genetics and biochemistry have implicated many helicases in the RISC assembly pathway whose functions remain obscure^{14,46–51}. Thus, it was plausible that a helicase domain directly unwinds the small RNA duplexes using the energy of ATP when the passenger-strand cleavage pathway is not available. However, our data indicate that unwinding is ATP

independent, but the preceding RISC-loading step is ATP dependent (**Fig. 3**). Recent crystal structures of *Thermus thermophilus* Ago protein (TtAgo) bound to DNA guide, or both DNA guide and target RNA, revealed remarkable flexibility of the Ago structure^{52,53}. Given that animal Ago proteins are loaded with double-stranded small RNAs from RISC loading machinery, ATP might be used to trigger dynamic conformational opening of Ago proteins to accept RNA duplexes, presumably through interaction between Ago and the RISC loading machinery. We show that at least Dcr-1 and GW182 are dispensable for Ago1-RISC loading (**Supplementary Fig. 3d,e**). Further study is warranted to determine the identity of the Ago1-RISC loading machinery.

Our data show that, in contrast to RISC loading, subsequent strand separation does not require the slicer activity of Ago1 or ATP (**Figs. 3** and **5**). Thus slicer-independent unwinding is an unexpectedly passive process. Notably, the preference for mismatches in the seed or 3'-mid regions for Ago1-RISC maturation is the reverse of the previously reported preference for base-pairing in those regions for target recognition by miRNAs³¹ (**Fig. 4f**). G:U wobbles are favored for unwinding (**Supplementary Fig. 7b**) and disfavored for target recognition^{31–33}, both at a level disproportionate to their thermodynamic stability. Both unwinding and target recognition measure the position of mismatches from the 5' end of the guide^{54,55} (**Fig. 4** and **Supplementary Figs. 5** and **6**). Moreover, both unwinding and target recognition and cleavage are, *per se*, ATP independent^{27,56} (**Fig. 3c,d**). These data indicate that unwinding is a precise mirror-image process of target recognition.

The biased positional effect in base-pairing for unwinding and target recognition might also be explained by the structures of Ago proteins. In the structures of TtAgo, the seed region of the guide strand is preorganized in an A-form helix-like geometry, whereas the 3' half of the guide strand is disturbed. Although base-pairing in the 3'-half region is dispensable for target cleavage activity of TtAgo^{52,53}, supplemental base-pairings in 3'-mid regions enhance target recognition in animals^{31,56–58}. This contribution of the 3'-mid base-pairings for target recognition can be explained by a so-called

'two-state' model^{2,3,59}. In the first state, only the seed region of the guide strand is paired to the target mRNA, whereas the 3' half of the guide is organized in a way that prevents base-pairing to the target. In the second state, both the seed and 3' half regions of the guide can anneal to the target mRNA. Because target recognition and cleavage are ATP independent^{27,56}, we envision that these two states are near-isoenergetic and can rapidly interconvert. Given this model, we reason that pre-RISC also oscillates between the two states—pre-RISC (I), where both the seed and 3' half regions of the guide can anneal to the passenger, and pre-RISC (II), where only the seed region of the guide is allowed to pair with the passenger (Fig. 6c). In these two states, only pre-RISC (II) is favorable for strand separation. In this model, seed mismatches promote conversion of pre-RISC (II) into mature RISC, whereas 3'-mid mismatches shift the equilibrium between pre-RISC (I) and pre-RISC (II) toward pre-RISC (II) (Fig. 6c).

Taken together, our analysis of slicer-independent unwinding provides a biochemical explanation for the recognition of target mRNAs by RISC, with both processes being subjected to the identical, unique structural constraints on RNAs by Ago proteins. Because the structural signature of natural miRNA-miRNA* duplexes is conserved among species (Fig. 6a and Supplementary Fig. 8), our findings for *Drosophila* Ago1 may also explain how non-slicer Ago proteins, such as mammalian Ago1, Ago3 and Ago4, generally separate the two small RNA strands. Moreover, our findings explain why miRNA genes evolutionally retain so many mismatches between their miRNA and miRNA* strands.

METHODS

Methods and any associated references are available in the online version of the paper at <http://www.nature.com/nsmb/>.

Note: Supplementary information is available on the Nature Structural & Molecular Biology website.

ACKNOWLEDGMENTS

We thank M. Siomi and H. Siomi (Keio University) for the antibody to Ago1, E. Izaurralde (Max Planck Institute) for the antibody to GW182, R. Carthew (Northwestern University) for the *dcr-2^{LS1H5X}* flies, and M. Horwich and P. Zamore (University of Massachusetts) for the pENTR-Ago1 construct. We also thank M. Mitomi and A. Maruyama for technical assistance, S. Iwasaki and members of the Tomari laboratory for constant and helpful discussions, and H. Sasaki and P. Zamore for stimulating discussions. This work was supported in part by a Grant-in-Aid for Young Scientists (A) from the Japan Ministry of Education, Culture, Sports, Science and Technology, and a Carrier Development Award from The International Human Frontier Science Program Organization. Y.T. is a Japan Science and Technology Agency PRESTO researcher.

AUTHOR CONTRIBUTIONS

T.K. and Y.T. conducted biochemical experiments, H.S. conducted bioinformatic analyses, Y.T. supervised the study, and all authors discussed the results and wrote the manuscript.

Published online at <http://www.nature.com/nsmb/>.
Reprints and permissions information is available online at <http://npg.nature.com/reprintsandpermissions/>.

- Bartel, D.P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281–297 (2004).
- Tomari, Y. & Zamore, P.D. Perspective: machines for RNAi. *Genes Dev.* **19**, 517–529 (2005).
- Filipowicz, W. RNAi: the nuts and bolts of the RISC machine. *Cell* **122**, 17–20 (2005).
- Okamura, K., Ishizuka, A., Siomi, H. & Siomi, M.C. Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways. *Genes Dev.* **18**, 1655–1666 (2004).
- Förstemann, K., Horwich, M.D., Wee, L., Tomari, Y. & Zamore, P.D. *Drosophila* microRNAs are sorted into functionally distinct Argonaute complexes after production by Dicer-1. *Cell* **130**, 287–297 (2007).
- Tomari, Y., Du, T. & Zamore, P.D. Sorting of *Drosophila* small silencing RNAs. *Cell* **130**, 299–308 (2007).
- Kawamura, Y. *et al.* *Drosophila* endogenous small RNAs bind to Argonaute2 in somatic cells. *Nature* **453**, 793–797 (2008).
- Ghildiyal, M. *et al.* Endogenous siRNAs derived from transposons and mRNAs in *Drosophila* somatic cells. *Science* **320**, 1077–1081 (2008).
- Azuma-Mukai, A. *et al.* Characterization of endogenous human Argonautes and their miRNA partners in RNA silencing. *Proc. Natl. Acad. Sci. USA* **105**, 7964–7969 (2008).
- Su, H., Trombly, M.I., Chen, J. & Wang, X. Essential and overlapping functions for mammalian Argonautes in microRNA silencing. *Genes Dev.* **23**, 304–317 (2009).
- Liu, J. *et al.* Argonaute2 is the catalytic engine of mammalian RNAi. *Science* **305**, 1437–1441 (2004).
- Meister, G. *et al.* Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Mol. Cell* **15**, 185–197 (2004).
- Liu, Q. *et al.* R2D2, a bridge between the initiation and effector steps of the *Drosophila* RNAi pathway. *Science* **301**, 1921–1925 (2003).
- Tomari, Y. *et al.* RISC assembly defects in the *Drosophila* RNAi mutant *armitage*. *Cell* **116**, 831–841 (2004).
- Pham, J.W., Pellino, J.L., Lee, Y.S., Carthew, R.W. & Sontheimer, E.J.A. Dicer-2-dependent 80S complex cleaves targeted mRNAs during RNAi in *Drosophila*. *Cell* **117**, 83–94 (2004).
- Förstemann, K. *et al.* Normal microRNA maturation and germ-line stem cell maintenance requires Loquacious, a double-stranded RNA-binding domain protein. *PLoS Biol.* **3**, e236 (2005).
- Saito, K., Ishizuka, A., Siomi, H. & Siomi, M.C. Processing of pre-microRNAs by the Dicer-1-Loquacious complex in *Drosophila* cells. *PLoS Biol.* **3**, e235 (2005).
- Liu, X. *et al.* Dicer-1, but not Loquacious, is critical for assembly of miRNA-induced silencing complexes. *RNA* **13**, 2324–2329 (2007).
- Jin, Z. & Xie, T. Dcr-1 maintains *Drosophila* ovarian stem cells. *Curr. Biol.* **17**, 539–544 (2007).
- Nykänen, A., Haley, B. & Zamore, P.D. ATP requirements and small interfering RNA structure in the RNA interference pathway. *Cell* **107**, 309–321 (2001).
- Rand, T.A., Petersen, S., Du, F. & Wang, X. Argonaute2 cleaves the anti-guide strand of siRNA during RISC activation. *Cell* **123**, 621–629 (2005).
- Matranga, C., Tomari, Y., Shin, C., Bartel, D.P. & Zamore, P.D. Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes. *Cell* **123**, 607–620 (2005).
- Miyoshi, K., Tsukumo, H., Nagami, T., Siomi, H. & Siomi, M.C. Slicer function of *Drosophila* Argonautes and its involvement in RISC formation. *Genes Dev.* **19**, 2837–2848 (2005).
- Leuschner, P.J., Ameres, S.L., Kuegn, S. & Martinez, J. Cleavage of the siRNA passenger strand during RISC assembly in human cells. *EMBO Rep.* **7**, 314–320 (2006).
- Lee, Y.S. *et al.* Distinct roles for *Drosophila* Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. *Cell* **117**, 69–81 (2004).
- Chan, S.P., Ramaswamy, G., Choi, E.Y. & Slack, F.J. Identification of specific *let-7* microRNA binding complexes in *Caenorhabditis elegans*. *RNA* **14**, 2104–2114 (2008).
- Iwasaki, S., Kawamata, T. & Tomari, Y. *Drosophila* Argonaute1 and Argonaute2 employ distinct mechanisms for translational repression. *Mol. Cell* **34**, 58–67 (2009).
- Eulalio, A., Huntzinger, E. & Izaurralde, E. GW182 interaction with Argonaute is essential for miRNA-mediated translational repression and mRNA decay. *Nat. Struct. Mol. Biol.* **15**, 346–353 (2008).
- Khvorova, A., Reynolds, A. & Jayasena, S.D. Functional siRNAs and miRNAs exhibit strand bias. *Cell* **115**, 209–216 (2003).
- Schwarz, D.S. *et al.* Asymmetry in the assembly of the RNAi enzyme complex. *Cell* **115**, 199–208 (2003).
- Grimson, A. *et al.* microRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol. Cell* **27**, 91–105 (2007).
- Doench, J.G., Petersen, C.P. & Sharp, P.A. siRNAs can function as miRNAs. *Genes Dev.* **17**, 438–442 (2003).
- 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).
- 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).
- Rivas, F.V. *et al.* Purified Argonaute2 and an siRNA form recombinant human RISC. *Nat. Struct. Mol. Biol.* **12**, 340–349 (2005).
- Ruby, J.G. *et al.* Evolution, biogenesis, expression, and target predictions of a substantially expanded set of *Drosophila* microRNAs. *Genome Res.* **17**, 1850–1864 (2007).
- Seitz, H., Ghildiyal, M. & Zamore, P.D. Argonaute loading improves the 5' precision of both microRNAs and their miRNA* strands in flies. *Curr. Biol.* **18**, 147–151 (2008).
- Lu, J. *et al.* The birth and death of microRNA genes in *Drosophila*. *Nat. Genet.* **40**, 351–355 (2008).
- Chung, W.J., Okamura, K., Martin, R. & Lai, E.C. Endogenous RNA Interference provides a somatic defense against *Drosophila* transposons. *Curr. Biol.* **18**, 795–802 (2008).

ARTICLES

40. Ruby, J.G. *et al.* Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in *C. elegans*. *Cell* **127**, 1193–1207 (2006).
41. Batista, P.J. *et al.* PRG-1 and 21U-RNAs interact to form the piRNA complex required for fertility in *C. elegans*. *Mol. Cell* **31**, 67–78 (2008).
42. Calabrese, J.M., Seila, A.C., Yeo, G.W. & Sharp, P.A. RNA sequence analysis defines Dicer's role in mouse embryonic stem cells. *Proc. Natl. Acad. Sci. USA* **104**, 18097–18102 (2007).
43. Tam, O.H. *et al.* Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. *Nature* **453**, 534–538 (2008).
44. Baek, D. *et al.* The impact of microRNAs on protein output. *Nature* **455**, 64–71 (2008).
45. Babiarz, J.E., Ruby, J.G., Wang, Y., Bartel, D.P. & Blelloch, R. Mouse ES cells express endogenous shRNAs, siRNAs, and other Microprocessor-independent, Dicer-dependent small RNAs. *Genes Dev.* **22**, 2773–2785 (2008).
46. Ishizuka, A., Siomi, M.C. & Siomi, H. A *Drosophila* fragile X protein interacts with components of RNAi and ribosomal proteins. *Genes Dev.* **16**, 2497–2508 (2002).
47. Robb, G.B. & Rana, T.M. RNA helicase A interacts with RISC in human cells and functions in RISC loading. *Mol. Cell* **26**, 523–537 (2007).
48. Meister, G. *et al.* Identification of novel Argonaute-associated proteins. *Curr. Biol.* **15**, 2149–2155 (2005).
49. Höck, J. *et al.* Proteomic and functional analysis of Argonaute-containing mRNA-protein complexes in human cells. *EMBO Rep.* **8**, 1052–1060 (2007).
50. Landthaler, M. *et al.* Molecular characterization of human Argonaute-containing ribonucleoprotein complexes and their bound target mRNAs. *RNA* **14**, 2580–2596 (2008).
51. Zhou, R. *et al.* Comparative analysis of Argonaute-dependent small RNA pathways in *Drosophila*. *Mol. Cell* **32**, 592–599 (2008).
52. Wang, Y., Sheng, G., Juranek, S., Tuschl, T. & Patel, D.J. Structure of the guide-strand-containing Argonaute silencing complex. *Nature* **456**, 209–213 (2008).
53. Wang, Y. *et al.* Structure of an Argonaute silencing complex with a seed-containing guide DNA and target RNA duplex. *Nature* **456**, 921–926 (2008).
54. Elbashir, S.M., Martinez, J., Patkaniowska, A., Lendeckel, W. & Tuschl, T. Functional anatomy of siRNAs for mediating efficient RNAi in *Drosophila melanogaster* embryo lysate. *EMBO J.* **20**, 6877–6888 (2001).
55. Elbashir, S.M., Lendeckel, W. & Tuschl, T. RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes Dev.* **15**, 188–200 (2001).
56. Haley, B. & Zamore, P.D. Kinetic analysis of the RNAi enzyme complex. *Nat. Struct. Mol. Biol.* **11**, 599–606 (2004).
57. Doench, J.G. & Sharp, P.A. Specificity of microRNA target selection in translational repression. *Genes Dev.* **18**, 504–511 (2004).
58. Brennecke, J., Stark, A., Russell, R.B. & Cohen, S.M. Principles of microRNA-target recognition. *PLoS Biol.* **3**, e85 (2005).
59. Mallory, A.C. *et al.* microRNA control of PHABULOSA in leaf development: importance of pairing to the microRNA 5' region. *EMBO J.* **23**, 3356–3364 (2004).

ONLINE METHODS

General methods. Preparation of *Drosophila melanogaster* overnight embryo or S2 cell lysates by Dounce homogenization, 40× reaction mix (containing ATP, ATP regeneration system and RNase inhibitor), and lysis buffer (30 mM HEPES (pH 7.4), 100 mM KOAc and 2 mM Mg(OAc)₂) has been described in detail⁶⁰. The *in vitro* RISC assembly assay typically contained 5 μl of lysate, 3 μl of 40× reaction mix, 1 μl of small RNA duplex and 1 μl of target RNA in a total of 10 μl. The *in vitro* target cleavage assay was done with 50 nM small RNA duplex and 1 nM target RNA cap-radiolabeled with ³²P. The unwinding assay using 20% native acrylamide gels^{6,20}, S2 cell culture and knockdown⁵, small-RNA radiolabeling with T4 polynucleotide kinase (Takara)^{20,60}, 254-nm UV cross-linking⁶, cap-labeling of target mRNA with guanylyltransferase (Epicentre or Ambion)^{6,60} and *in vitro* analysis of Ago1-mediated translational repression²⁷ were done as described previously.

Preparation of the target mRNAs. A DNA fragment containing the gene encoding *Remilla* luciferase (RL) and four *let-7* target sites in its 3' untranslated region was amplified by PCR from psiCHECK2-*let-7* 4× (ref. 27). Each target site contained a central bulge to prevent endonucleolytic cleavage by Ago1-RISC. RL 4× mRNA was transcribed using an mScript mRNA production system (Epicentre). We confirmed that mature RISC could be detected using RL 1× mRNA (which contains only one *let-7* target site) with efficiency equal to that of RL 4× mRNA. For the experiments in **Supplementary Figure 2a**, RL 1× mm1 mRNA (which has one target site complementary to the mm1 passenger strand with a central bulge) and RL 1× mm5 mRNA (which has one target site complementary to the mm5 passenger strand with a central bulge) were used. For the experiment in **Supplementary Figure 7b** (left), RL 1× MM10/(U:G)5 mRNA (which has one target site complementary to the MM10/(U:G)5 guide strand with a central bulge) was used.

Native gel analysis. Native gels 1.5–2 mm thick, containing 1.4% (w/v) agarose (Low Range Ultra Agarose, Bio-Rad Laboratories), were cast vertically between glass plates with 0.5-mm-thick bottom spacers (16 cm × 16 cm). To detect intermediate complexes in the Ago1-RISC assembly pathway, 10–20 nM of ³²P-radiolabeled duplex and 10 nM of a nonradiolabeled ~1,500-nt target mRNA were incubated under standard RISC assembly assay conditions (see above) at 25 °C or 15 °C. After incubation, the samples were separated by vertical native agarose gel electrophoresis at 300 V for 1.5 h in ice-cold 0.5× Tris-borate-EDTA buffer. Complexes were detected by PhosphorImager (BAS-1500 or FLA-7000 image analyzer, Fujifilm) and quantified using Multi Gauge software (Fujifilm).

Detection of RNA in the complexes. ³²P-radiolabeled duplex A or duplex B was incubated with lysate and RL 4× mRNA for 30 min and separated by native gel electrophoresis, after which the complexes were excised from the gel. The gel slices were deproteinized using 1 mg ml⁻¹ proteinase K in 2× PK buffer (200 mM Tris (pH 7.5), 300 mM NaCl and 2% (w/v) SDS) at room temperature for 18 h. The eluted RNA was precipitated with ice-cold ethanol and then dissolved in native loading dye and analyzed by electrophoresis.

ATP depletion. ATP was depleted by preincubating lysate with 20 mM glucose and 0.15 or 0.3 U μl⁻¹ hexokinase (Sigma) for 30 min at 25 °C (**Fig. 3a**) or 15 °C (**Fig. 3c**). More than 99% of endogenous ATP was depleted (final <100 nM), as measured by ENLITEN luciferase/luciferin reagent (Promega). The ATP regeneration system (creatine kinase and creatine monophosphate) was omitted in the experiments in **Figure 3** and **Supplementary Figure 4**.

Plasmid construction. pENTR-Ago1 was a kind gift from M. Horwich and P. Zamore. Mutations in the *ago1* gene (D687A, D760A, or H898A) were intro-

duced using the QuikChange site-directed mutagenesis kit (Stratagene) to generate the 'entry' clones. The entry clones were recombined with the pAFW destination vector (*Drosophila* Gateway vector collection, Actin5C promoter and N-terminal 3× Flag-tag) using Gateway LR Clonase enzyme mix (Invitrogen).

Transfection. S2 cells (1 × 10⁶ cells ml⁻¹) were transfected with 10 μg pAFW-Ago1 vectors (or their mutant derivatives) per 10 ml culture using FuGENE HD transfection reagent (Roche).

Immunoprecipitation and immunodepletion. Immunoprecipitation was done as previously described^{6,27}. The immune complexes were washed three times with lysis buffer containing 0.1% (v/v) Triton X-100 and washed again with 1× lysis buffer. In the experiment in **Figure 3d**, 800 mM NaCl was supplemented in the wash buffer. Ago1 immunodepletion was carried out as described previously⁶.

Western blotting. Antibodies to Flag (1:5,000; Sigma), Ago1 (1:10,000; ref. 23), GW182 (1:1,000; ref. 28), Dcr-1 (1:1,000; ref. 16) and Dcr-2 (1:5,000; ab4732, Abcam) were used as the primary antibodies. Chemiluminescence was induced by SuperSignal West Dura substrate (Pierce), and images were acquired by a LAS-3000 imaging system and quantified using Multi Gauge software (Fujifilm Life Sciences).

In silico conceptual dicing. Pre-miRNA sequences were downloaded from miRBase (version 12.0), then extended by 20 nt at each extremity based on the current genome assemblies (April 2006 release of the *D. melanogaster* genome, July 2007 release of the *M. musculus* genome and May 2008 release of the *C. elegans* genome). Pre-miRNA sequences were folded into an unbranched hairpin, using RNAsubopt from the Vienna RNA package (<http://www.tbi.univie.ac.at/RNA/>) to select the most stable unbranched hairpin within 8 kcal mol⁻¹ of the most stable structure. miRNA sequences were extracted from the following GEO data sets: *D. melanogaster*: GSM180329, GSM180330, GSM180332, GSM180333, GSM180335, GSM180336 (ref. 36), GSM239041, GSM239050, GSM239051, GSM239052, GSM239053, GSM239054, GSM239055, GSM239056 (ref. 37), GSM246084 (ref. 38), GSM272651, GSM272652, GSM272653 and GSM275691 (ref. 39); *M. musculus*: GSM237107, GSM237110 (ref. 42), GSM261957, GSM261959 (ref. 43), GSM304914 (ref. 44), GSM314552 and GSM314558 (ref. 45); *C. elegans*: GSM139137 (ref. 40), GSM297743, GSM297744, GSM297746, GSM297747, GSM297750, GSM297751 and GSM297757 (ref. 41). To include every alternative processing isoform, all of the libraries of a given species were pooled, and up to 9 nt of distance between the observed isoforms and the miRBase-annotated isoform was tolerated. miRNA* sequences were extracted from the folded pre-miRNA hairpins, considering that a miRNA* pairs to the miRNA, with each strand showing a 2-nt 3'-overhang.

Analysis of structural signature of natural miRNA-miRNA* duplexes. miRNAs with at least 100 reads in the pooled deep-sequencing libraries were selected. For each miRNA, the most abundant isoform was selected from the pooled deep-sequencing libraries, and its most stable miRNA-miRNA* duplex structure was predicted using RNAfold from the Vienna RNA package. For the analyses in **Supplementary Figure 8**, a 3-nt window centered at each position was slid across the miRNA strand (guide positions 2–17). For each window, we then calculated the mean and standard error of the random variable *x*, defined by *x* = 1 if there is at least one mismatch in the window, and *x* = 0 if there is no mismatch. For the analyses in **Figure 6a**, positions of mismatches in each miRNA-miRNA* duplex were recorded and classified as described in Results.

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

3.7 Discussion : caractérisation biochimique d'une machine moléculaire inconnue

Ces travaux ont donc montré que le chargement, puis l'ouverture, d'un duplex sur Ago1, présentent des caractéristiques bien distinguables :

- le chargement nécessite de l'ATP, alors que (et de manière contre-intuitive *a priori*) l'ouverture du duplex n'en nécessite pas ;
- le chargement est facilité par la présence d'un mésappariement central dans le duplex, alors que l'ouverture du duplex requiert des mésappariements dans la graine ou la moitié 3' du guide.

Les pre-miRNA naturels ont précisément optimisé ces paramètres, et, chez chacune des trois espèces animales analysées, la plupart des duplex miRNA/miRNA* qui présentent un mésappariement central (donc, chargés sur Ago1 ou un orthologue) sont également mésappariés dans la graine ou la moitié 3' de leur miRNA.

Ces mésappariements facilitent vraisemblablement l'ouverture du duplex, qui ne fait pas intervenir d'hélicase ATP-asiq. La maturation de RISC sur Ago1 est donc un processus passif, qui dépend de l'ouverture spontanée d'un duplex imparfaitement apparié. L'irréversibilité du processus (le brin passager est probablement rapidement dégradé une fois libéré dans le cytoplasme) déplace alors l'équilibre dans le sens de la maturation.

Il faut souligner que la machinerie de chargement sur Ago1 n'est pas encore connue (alors qu'elle est connue depuis plusieurs années pour Ago2 : voir sous-section 1.2.2, page 12). La découverte de ses caractéristiques (préférence pour les duplex mésappariés autour du neuvième nucléotide du guide, activité dépendant de l'ATP) devrait aider à l'identifier, de même que le système d'analyse sur gels natifs, développé dans le laboratoire de Yukihide Tomari.

Chapitre 4

La biogenèse des piRNA

4.1 Contexte : qu'est-ce qui fait un piRNA ?

Dès la découverte des piRNA (voir chapitre 2, page 29), il est apparu que ces petits ARN sont, individuellement, très peu abondants (la plupart d'entre eux ont été séquencés une seule fois chacun, dans des expériences de séquençage à haut débit où un miRNA typique est séquencé plusieurs centaines de fois), mais qu'ils sont très divers en séquence. Cette grande variété de séquence soulève un problème intéressant : comment les piRNA sont-ils fabriqués ? Qu'est-ce qui fait qu'une séquence génomique sera exprimée en piRNA, et pas une autre ? À part une haute fréquence d'uridine sur l'extrémité 5', les piRNA ne présentent pas de motif de séquence clair, sur la séquence du piRNA mature ou sur les séquences génomiques flanquantes, qui pourrait recruter une machinerie de maturation ¹.

Chez la Drosophile, la plupart des piRNA proviennent de transposons. Le séquençage de piRNA associés aux différentes protéines Piwi de Drosophile a révélé à Gunawardane *et al.*, 2007 et Brennecke *et al.*, 2007 que les piRNA chargés sur Aub et Piwi sont fréquemment antisens aux transposons, alors que les piRNA chargés sur Ago3 sont généralement dans l'orientation sens. Et plus précisément, les piRNA chargés sur Aub ou Piwi sont fréquemment complémentaires des piRNA chargés sur Ago3, sur leurs dix premiers nucléotides (par conséquent, les piRNA chargés sur Ago3 ont fréquemment une adénosine en position 10, complémentaire de l'uridine en position 1 des piRNA chargés sur Aub ou Piwi). Cette complémentarité de 10 pb ne concerne pas que les piRNA matures, mais également leurs précurseurs présumptifs : ces deux classes de piRNA proviennent souvent des mêmes loci génomiques, et leurs extrémités 5' sont séparées de 9 pb sur le génome (les piRNA matures se recouvrent donc de 9+1=10 nt).

Ces observations, et les caractéristiques des protéines Argonaute capables de cliver des ARN-cibles (le pont phosphodiester clivé est celui qui sépare les nucléotides de la cible qui font face aux dixième et onzième nucléotide du guide : voir sous-section 1.2.3, page 14), ont amené Gunawardane *et al.*, 2007 et Brennecke *et al.*, 2007 à proposer un modèle pour la biogenèse des piRNA : l'une des deux classes de piRNA (par exemple, les piRNA chargés sur Aub ou Piwi) guide le clivage de l'extrémité 5' des piRNA de l'autre classe (dans cet exemple : les piRNA chargés sur Ago3), et réciproquement. Chaque piRNA proviendrait d'un précurseur simple brin, qui serait reconnu par un piRNA mature, lequel guiderait le clivage (par la protéine Argonaute à laquelle il est associé ²) qui définirait l'extrémité 5' du piRNA mûré. Son extrémité 3' serait définie par un autre mécanisme. Ce modèle (appelé « modèle du ping-pong ») est schématisé en figure 4.1, page 71.

Quantitativement, les différentes classes de piRNA ne se valent pas : d'une part, la population totale de piRNA est essentiellement constituée de piRNA chargés sur Aub ou Piwi, puisqu'elle présente une haute fréquence d'uridine en position 1 et pas d'enrichissement notable en adénosine en position 10 ; d'autre part, les piRNA chargés sur Ago3 semblent s'apparier plus fréquemment aux piRNA chargés sur Aub que sur Piwi (et des appariements entre piRNA chargés sur Aub, et entre piRNA chargés sur Ago3, sont rares, mais observés : Brennecke *et al.*, 2007).

Le rôle particulier d'Ago3 dans ce modèle (seule protéine de la sous-famille Piwi à interagir

¹À l'exception des piRNA de Nématode, dont le contexte génomique porte systématiquement un motif de séquence (CTGTTCA, dans un contexte riche en A et T) environ 45 pb en amont du premier nucléotide du petit ARN (Ruby *et al.*, 2006). Les piRNA de Nématode présentent deux particularités supplémentaires, qui les distinguent des piRNA connus chez les autres Animaux : ils sont plus courts (21 nt chez le Nématode, entre 24 et 30 nt chez la Drosophile et les Vertébrés) et très homogènes en taille, et ils proviennent d'un unique chromosome. Ces petits ARN n'ont donc pas immédiatement été reconnus comme des piRNA, et ils avaient été baptisés les « 21U RNA » ; ce n'est que deux ans plus tard que la communauté a compris que ces 21U RNA sont les piRNA du Nématode (Das *et al.*, 2008 ; Batista *et al.*, 2008).

²Bien que chacune des trois protéines de la sous-famille Piwi chez la Drosophile possède la triade catalytique nécessaire au clivage d'un ARN-cible (voir sous-section 1.2.3, page 14), leur activité endonucléolytique n'a été démontrée que pour Ago3 et Aub, guidées par un petit ARN de 21 ou 26 nt de long (Gunawardane *et al.*, 2007).

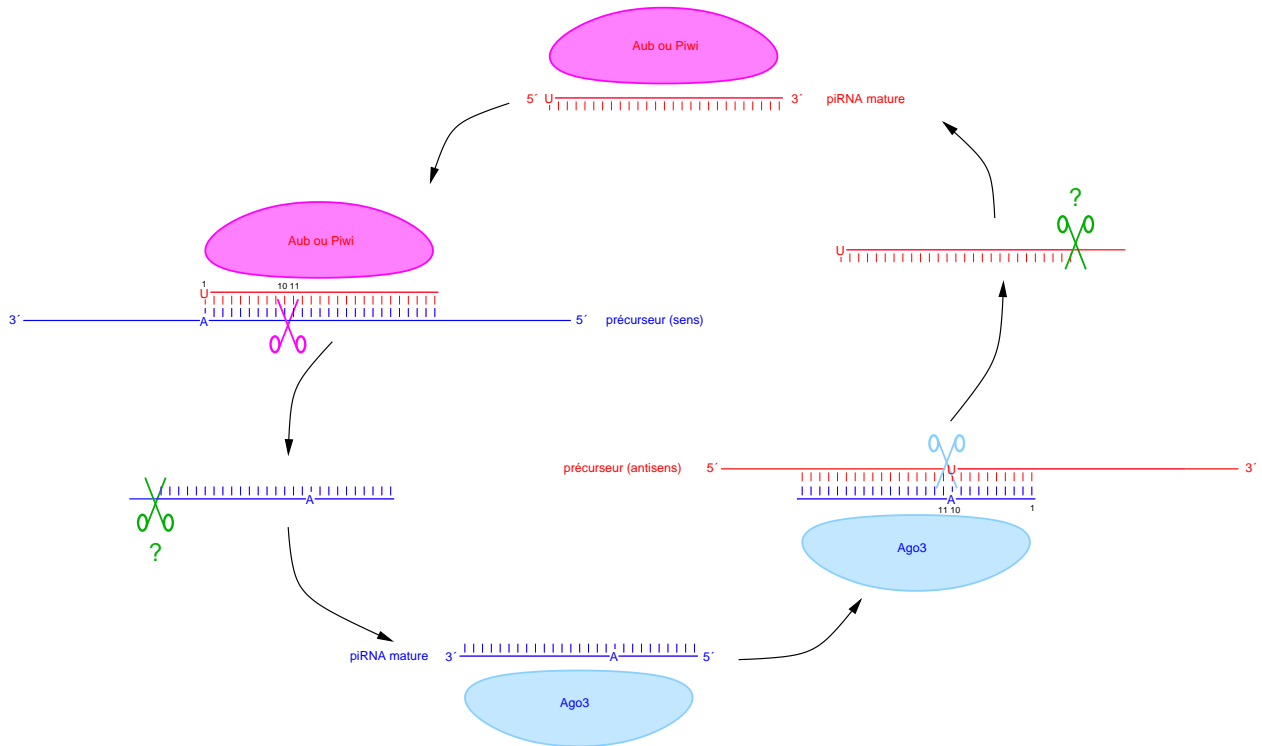


FIG. 4.1 – **Le modèle du ping-pong pour la biogenèse des piRNA.** Les piRNA chargés sur Aub ou Piwi (représentés en rouge) sont fréquemment antisens aux transposons, et commencent souvent par une uridine. Ils guideraient le clivage d'ARN simple-brin qui leur sont complémentaires (en orientation sens par rapport aux transposons), générant ainsi l'extrémité 5' des piRNA qui seront chargés sur Ago3 (représentés en bleu). L'extrémité 3' serait générée par une nucléase non identifiée (en vert). Réciproquement, les piRNA chargés sur Ago3 (fréquemment en orientation sens par rapport aux transposons, et portant une adénosine en position 10) pourraient guider le clivage de l'extrémité 5' des piRNA qui seront chargés sur Aub ou Piwi.

préférentiellement avec les piRNA sens, et supposée responsable de la maturation de l'extrémité 5' de la plupart des piRNA chargés sur les deux autres protéines) rendait particulièrement attractive l'étude de mutants *ago3*. Mais aucun mutant n'était référencé pour ce gène (localisé dans l'hétérochromatine péricentromérique, il est peu accessible à la mutagenèse par mobilisation d'éléments P) ; il a donc fallu recourir à la mutagenèse chimique pour isoler trois allèles qui introduisent un codon stop prématuré dans *ago3*, et une déficience qui élimine le gène *ago3* et les gènes voisins. De manière contre-intuitive, les piRNA les plus affectés par la perte d'Ago3 ne sont pas les piRNA qui (chez le sauvage) s'associent à Ago3 : ce sont les piRNA qui s'associaient à Aub. Cette expérience permet donc de tester le modèle du ping-pong, qui prévoyait effectivement qu'Ago3 est nécessaire à la production des piRNA chargés sur Aub (et en l'absence d'Ago3, les piRNA qui y étaient habituellement chargés sont probablement chargés sur une autre protéine — par exemple, Aub). Cette analyse a également mis en évidence trois groupes de transposons, au comportement très variable face à la mutation d'*ago3* : pour les transposons du premier groupe (le plus nombreux), les piRNA chargés sur Ago3 sont dans l'orientation sens, et les piRNA chargés sur Aub (dont la biogenèse nécessite Ago3) sont dans l'orientation antisens ; les quelques transposons du deuxième groupe ont des caractéristiques exactement opposées (les piRNA chargés sur Ago3 sont en orientation antisens, les piRNA chargés sur Aub sont en orientation sens) ; et les transposons du troisième groupe sont réprimés par des piRNA peu sensibles à la perte d'Ago3 ou d'Aub : ces piRNA sont chargés sur Piwi, et ils répriment les transposons du

troisième groupe dans les cellules folliculaires (des cellules somatiques en contact avec l'ovocyte). Ces résultats sont présentés dans l'article qui suit (Li *et al.*, 2009a).

4.2 Article n°5 : deux voies de biogenèse pour les piRNA

Cell

Collapse of Germline piRNAs in the Absence of Argonaute3 Reveals Somatic piRNAs in Flies

Chengjian Li,^{1,9} Vasily V. Vagin,^{1,9,10} Soohyun Lee,^{2,5,9} Jia Xu,^{2,6,9} Shengmei Ma,¹ Hualin Xi,^{5,7} Hervé Seitz,¹ Michael D. Horwich,¹ Monika Syrzycka,⁸ Barry M. Honda,⁸ Ellen L.W. Kittler,³ Maria L. Zapp,³ Carla Klattenhoff,⁴ Nadine Schulz,⁴ William E. Theurkauf,⁴ Zhiping Weng,^{2,*} and Phillip D. Zamore^{1,*}

¹Department of Biochemistry and Molecular Pharmacology and Howard Hughes Medical Institute

²Program in Bioinformatics and Integrative Biology

³Program in Molecular Medicine and Center for AIDS Research (CFAR)

⁴Program in Molecular Medicine

University of Massachusetts Medical School, Worcester, Massachusetts 01605, USA

⁵Bioinformatics Program

⁶Biomedical Engineering

Boston University, Boston, MA 02215, USA

⁷Computational Sciences Center of Emphasis, Pfizer, Inc., 620 Memorial Drive, Cambridge, MA 02139, USA

⁸Molecular Biology and Biochemistry, Simon Fraser University, Burnaby BC V5A 1S6, Canada

⁹These authors contributed equally to this work

¹⁰Present address: Cold Spring Harbor Laboratories, Cold Spring Harbor, New York 11724, USA

*Correspondence: zhiping.weng@umassmed.edu (Z.W.), phillip.zamore@umassmed.edu (P.D.Z.)

DOI 10.1016/j.cell.2009.04.027

SUMMARY

Piwi-interacting RNAs (piRNAs) silence transposons in animal germ cells. piRNAs are thought to derive from long transcripts spanning transposon-rich genomic loci and to direct an autoamplification loop in which an antisense piRNA, bound to Aubergine or Piwi protein, triggers production of a sense piRNA bound to the PIWI protein Argonaute3 (Ago3). In turn, the new piRNA is envisioned to produce a second antisense piRNA. Here, we describe strong loss-of-function mutations in *ago3*, allowing a direct genetic test of this model. We find that Ago3 acts to amplify piRNA pools and to enforce on them an antisense bias, increasing the number of piRNAs that can act to silence transposons. We also detect a second, Ago3-independent piRNA pathway centered on Piwi. Transposons targeted by this second pathway often reside in the *flamenco* locus, which is expressed in somatic ovarian follicle cells, suggesting a role for piRNAs beyond the germline.

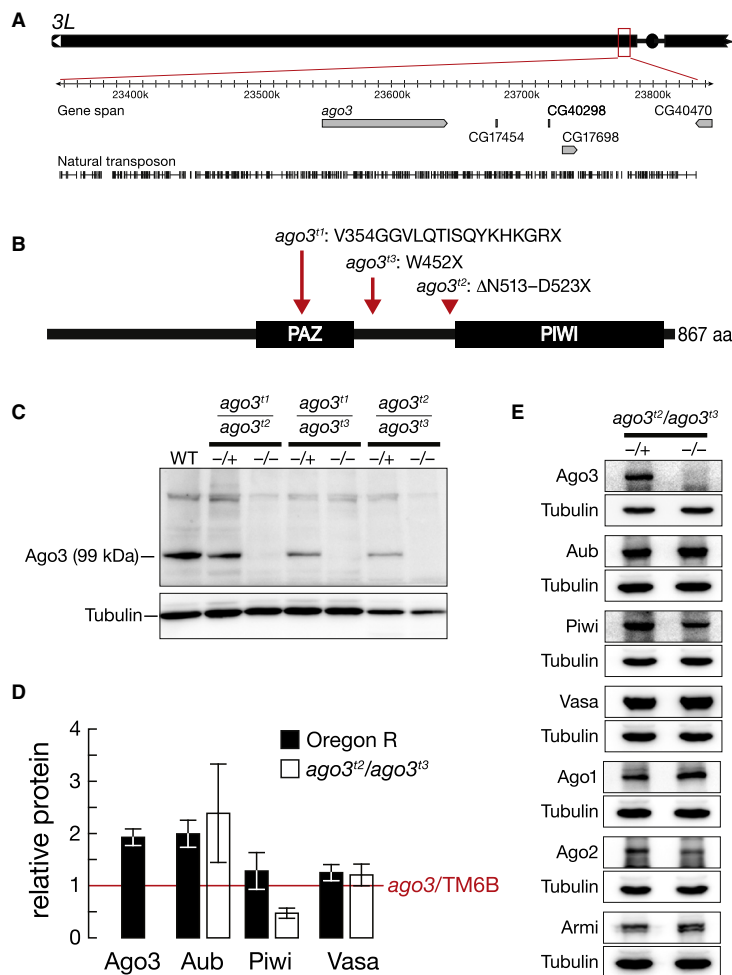
INTRODUCTION

The ability to tame transposons while retaining them in the genome is a particular specialty of eukaryotes. Transposons, repetitive sequences, and other forms of “selfish” DNA comprise as much as 42% of the human genome and nearly 30% of the genome of *Drosophila melanogaster*. In metazoa, transposons

are silenced by the piRNA pathway, which is guided by 23–30 nt RNAs (Vagin et al., 2006; Brennecke et al., 2007).

The piRNA pathway is distinct from other RNA silencing pathways in that its small RNA guides are not produced by dicing. In contrast, both small interfering RNAs (siRNAs) and microRNAs (miRNAs) are cleaved by double-stranded RNA-specific endonucleases, Dicers, to yield double-stranded intermediates—siRNA or miRNA/miRNA* duplexes—that are loaded into members of the Argonaute family of proteins (reviewed in Ghildiyal and Zamore, 2009). piRNAs, too, act as guides for Argonaute proteins, but they appear not to exist as stable double-stranded intermediates at any point in their biogenesis (Vagin et al., 2006; Houwing et al., 2007). piRNAs bind PIWI proteins, a subfamily of Argonaute proteins that are expressed in germline cells (Aravin et al., 2006; Girard et al., 2006; Lau et al., 2006; Vagin et al., 2006; Brennecke et al., 2007; Gunawardane et al., 2007; Batista et al., 2008; Das et al., 2008; Wang and Reinke, 2008). PIWI proteins were first identified by their roles in maintaining (Cox et al., 1998, 2000) and patterning (Wilson et al., 1996; Harris and Macdonald, 2001) *Drosophila* germ cells. The defects in the organization of embryonic pattern in piRNA pathway mutations are likely an indirect consequence of their larger role in maintaining genomic stability (Klattenhoff et al., 2007). The three *Drosophila* PIWI proteins, Piwi, Aubergine (Aub), and Argonaute3 (Ago3), are expressed in the male and female germline cells (Wilson et al., 1996; Cox et al., 1998, 2000; Williams and Rubin, 2002; Brennecke et al., 2007; Gunawardane et al., 2007; Nishida et al., 2007).

The prevailing model for piRNA biogenesis—the “ping-pong” model—reflects the discovery that the first 10 nt of piRNAs bound to Aub or Piwi, which are largely antisense and typically

**Figure 1. ago3 Mutants**

(A) The *ago3* gene resides in pericentromeric heterochromatin on the left arm of chromosome 3. (B) *ago3* mutant alleles were identified by TILLING. (C) Full-length Ago3 protein was not detected in transheterozygous *ago3* ovaries, but was readily detected in heterozygotes and wild-type ovaries. (D) Protein levels in wild-type and *ago3* ovaries, relative to *ago3*/TM6B (red line). The average \pm standard deviation for at least three independent biological samples is shown. (E) Representative data for Aub, Piwi, Vasa, Argonaute1 (Ago1), Argonaute2 (Ago2), and Armitage (Armi) in *ago3* ovaries. Tubulin served as a loading control.

begin with uridine, are often complementary to the first 10 nt of piRNAs bound to Ago3, which are largely sense and typically bear an adenosine at position 10 (Brennecke et al., 2007; Gunawardane et al., 2007). Many Argonaute proteins can act as RNA-guided RNA endonucleases, and all such Argonautes cut their target RNAs 5' to the base that pairs with the tenth nucleotide of the small RNA guide; all three fly PIWI proteins retain their endonuclease activity (Saito et al., 2006; Gunawardane et al., 2007). Thus, the observed 10 nt 5' complementarity between piRNAs suggests that the 5' ends of piRNAs bound to Aub or Piwi are defined by Ago3-catalyzed cleavage, and, reciprocally, that the 5' ends of piRNAs bound to Ago3 are defined by piRNAs bound to Aub or Piwi. The ping-pong model seeks to explain these observations, as well as the role of piRNA cluster transcripts in piRNA biogenesis, the function of piRNAs in silencing transposons, and the extraordinary antisense bias of piRNAs

identified as a repressor of transposon expression in somatic follicle cells (Pelisson et al., 1994).

RESULTS

Loss-of-Function ago3 Alleles

The *Drosophila ago3* gene resides in pericentromeric heterochromatin on the left arm of chromosome 3 (Figure 1A). Heterochromatin is refractory to *P* element or *piggyBac* transposon insertion and to targeted recombination, complicating isolation of *ago3* mutants. We used TILLING (Cooper et al., 2008) to identify three mutant alleles (*t1*, *t2*, and *t3*) that create premature stop codons in *ago3* (Figure 1B) and obtained one chromosomal deficiency [*Df(3L)TTT*] that deletes at least six genes, including *ago3* (Figure S1A available online). One homozygous (*ago3^{t3}/ago3^{t3}*) and three transheterozygous (*ago3^{t1}/ago3^{t2}*, *ago3^{t1}/ago3^{t3}*,

generally. At its core, the model proposes that piRNAs participate in an amplification loop in which transposon sense transcripts (e.g., transposon mRNAs) trigger the production of new, antisense piRNAs. Ago3, guided by sense piRNAs, lies at the heart of the amplification loop.

To test the ping-pong model, we isolated strong loss-of-function mutations in *ago3*. Here, we report the detailed analysis of *ago3* and *aub* mutant flies. Our data provide strong support for an amplification cycle in which Ago3 amplifies piRNA pools and enforces on them a strong antisense bias, increasing the number of piRNAs that can act to destroy transposon mRNAs. Moreover, we detect a second, perhaps somatic, piRNA pathway centered on Piwi and functioning without benefit of Ago3-catalyzed amplification. Most of the transposons targeted by this second pathway reside in the *flamenco* piRNA cluster, which was first

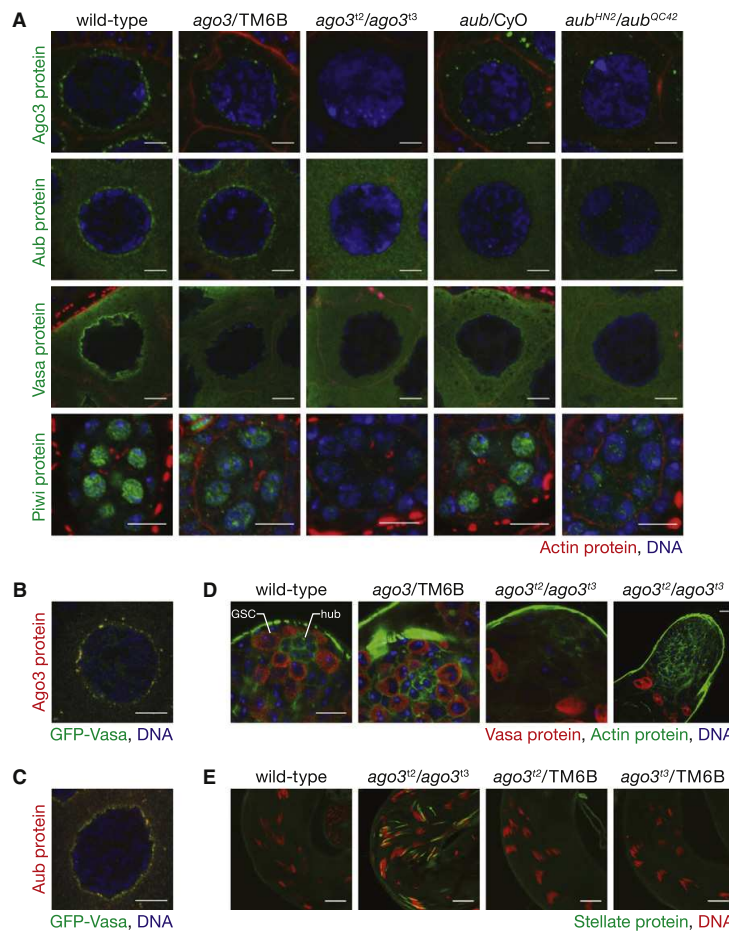


Figure 2. Consequences of Loss of Ago3 in the *Drosophila* Germline

(A) Mutually interdependent incorporation of Ago3 and Aub into nuage in ovaries. Images correspond to single confocal sections (63× magnification); scale bars represent 5 μm (Ago3, Aub, and Vasa panels) or 10 μm (Piwi).

(B and C) Vasa localization within nuage is only partially congruent with that of Ago3 and Aub. Images correspond to single confocal sections of stage 4/5 egg chambers (63× lens, 4× zoom). Scale bars represent 10 μm.

(D) In 5- to 7-day-old male adults, Vasa-expressing germline stem cells (GSC), which normally surround the somatic hub cells (hub), were not detected in *ago3* testes. Since Vasa expression increases in *ago3*, the two images from *ago3* mutants were acquired with reduced gain relative to wild-type and heterozygous testes. Scale bars represent 10 μm (first three images) or 20 μm (fourth).

(E) *Stellate* silencing in testes requires Ago3. Scale bars represent 20 μm.

and *ago3^{t1}/ago3^{t3}*) combinations, as well as each mutant allele in *trans* to *Df(3L)TTT*, were viable. Full-length Ago3 protein was essentially undetectable in all seven allelic combinations (Figure 1C and Figure S1B), and *ago3^{t2}/TM6B* or *ago3^{t3}/TM6B* heterozygous ovaries contained half as much Ago3 protein as Oregon R wild-type ovaries (Figure 1D), suggesting that the *ago3* alleles correspond to strong loss-of-function mutations. We refer to *ago3^{t2}/ago3^{t3}* transheterozygotes as *ago3* for brevity.

Mutually Interdependent Localization of PIWI Proteins

The intracellular localization of each PIWI protein appears to require other PIWI proteins. In the *Drosophila* female germline, Piwi is largely in the nucleus, whereas Aub and Ago3 are cytoplasmic and concentrated in “nuage,” a ring around the cytoplasmic face of the nuclei of the transcriptionally active germline nurse cells (Findley et al., 2003; Brennecke et al., 2007; Pane et al., 2007). The putative RNA helicase Vasa, which marks the

germline in most animals, is also in nuage (Liang et al., 1994). Mutations that disrupt piRNA biogenesis, but not those that block siRNA or miRNA production, disrupt localization of Aub to nuage (Lim and Kai, 2007).

Ago3 is also required for Aub to localize to nuage (Figure 2A). However, Ago3 is not required for Aub expression or stability: the concentration of Aub protein and the amount of Aub in the cytoplasm increased in *ago3* ovaries relative to their heterozygous siblings (Figures 1D and 1E). Reciprocally, Aub plays a role in the localization of Ago3 to nuage, although some Ago3 persisted in nuage in an *aub* hypomorphic mutant allelic combination (*aub^{HN2}/aub^{QC42}*). The function of nuage is unknown, but it may have a complex substructure, because the localization of Vasa within nuage only partially coincides with that of Ago3 (Figure 2B) and Aub (Figure 2C).

The interdependence of Ago3 and Aub for their localization to nuage likely reflects an underlying requirement for these two proteins for nuage assembly: the perinuclear localization of Vasa, a marker for nuage, was absent in both *ago3* and *aub* mutants, although the abundance of Vasa was unaltered in *ago3* (Figures 1D and 1E). The abundance of Armitage, Argonaute1, and Argonaute2, proteins required for small silencing RNA biogenesis or function, were similarly unaltered in *ago3* mutants (Figure 1E). In *ago3/TM6B* heterozygotes—which produce half as much Ago3 protein as the wild-type—and in *aub/CyO* heterozygotes, less Aub, Ago3, or Vasa was present in nuage than in the wild-type, suggesting that assembly of nuage is exquisitely sensitive to the amount of Ago3 and Aub (Figure 2A).

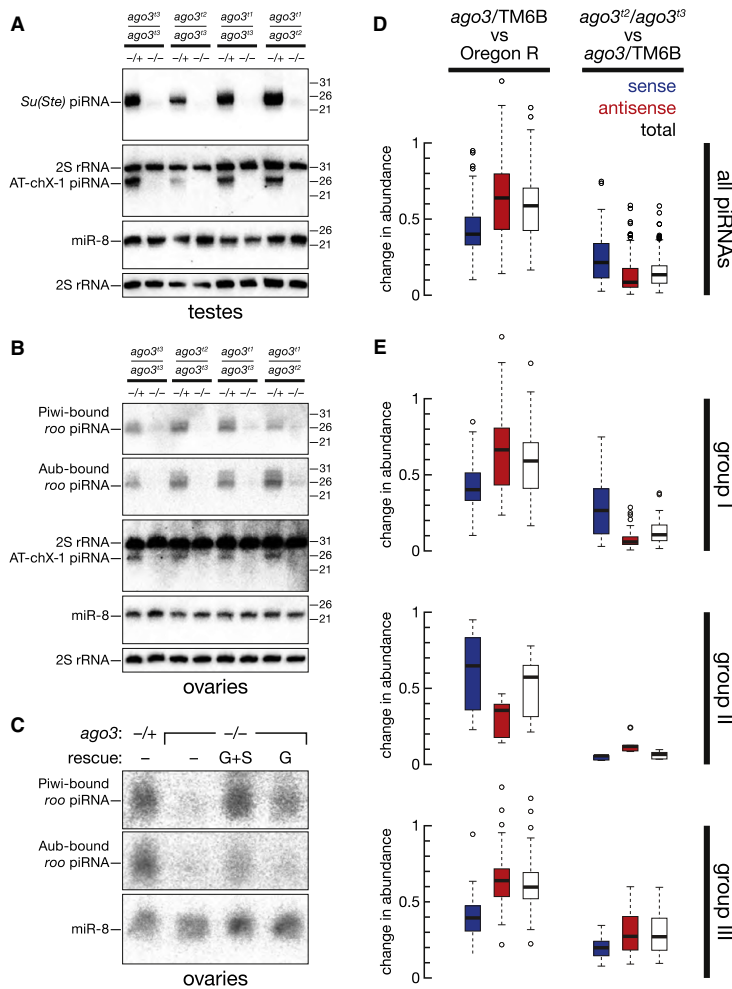


Figure 3. Aub- and Piwi-bound piRNAs Disappear without Ago3

(A) Accumulation of *Su(Ste)* and *AT-chX-1* piRNAs in testes required Ago3.

(B) Accumulation of *roo* antisense piRNAs, as well as the *AT-chX-1* piRNA, in ovaries required Ago3.

(C) Accumulation of Piwi- and Aub-bound *roo* antisense piRNAs in the ovary was rescued in *ago3* mutants by a single-copy transgene expressing Ago3 in the germline (G) or in both the germline and the soma (G+S).

(D) Box plots illustrating the change in abundance of piRNAs, analyzed by transposon family, in *ago3/TM6B* versus Oregon R (left panel) and *ago3* versus *ago3/TM6B* (right panel) ovaries.

(E) Box plots illustrating the change in abundance for piRNAs for each of 95 transposon families, separated by group.

ago3 Mutations Affect Fecundity

ago3 females are sterile. *ago3* females initially laid far fewer eggs than their heterozygous siblings and by day 10 stopped laying altogether (Figure S2A). The egg-laying deficit was rescued by a single-copy transgene expressing Ago3 in the germline or in both the germline and the soma. Of those eggs laid, none of the embryos hatched, compared with ~96% for the wild-type (Table S1). For 39%–67% of the embryos produced by *ago3* mothers, the dorsal appendages were fused or absent (Table S1), indicating a maternally inherited patterning defect, as has been reported for other piRNA pathway mutations (Schupbach and Wieschaus, 1991; Wilson et al., 1996; Cook et al., 2004; Klattenhoff et al., 2007; Pane et al., 2007). Loss of Ago3 in the germline, rather than a secondary mutation present in the *ago3* mutants, caused the patterning defects, as a single-copy transgene expressing Ago3 from the germline-specific *nanos* promoter rescued the dorsal appendage defects to essentially wild-type rates (Table S1).

In males, Ago3 is required to maintain germline stem cells (Figure 2D). We examined testes from 5- to 7-day-old wild-type, *ago3/TM6B*, and *ago3* males, using Vasa protein expression to identify germline cells. In the wild-type and *ago3/TM6B*, Vasa-expressing cells were present at the tip of the testes, in a ring around the hub cells, a group of somatic cells that support the adjacent germline stem cells (Fuller, 1993). In contrast, the corresponding region in *ago3* was devoid of Vasa-expressing cells, suggesting that germline stem cells are not properly maintained in the absence of Ago3. Consistent with a failure to maintain germline stem cells, *ago3* males are semifertile, siring fewer offspring than do males with the same genetic background (Figure S2B).

The peak of nuclear-localized Piwi protein expression occurs earlier in oogenesis than when Ago3 and Aub are maximally perinuclear (Cox et al., 2000). Less Piwi is found in the nucleoplasm of *ago3/TM6B* or *aub/CyO* heterozygotes than in the wild-type, and little or no Piwi is present in the nuclei of the *ago3* or *aub* mutants (Figure 2A); *ago3* ovaries contain about half as much Piwi as their heterozygous siblings (Figure 1E). (The localization of Aub and Ago3 cannot be studied in *piwi*, because *piwi* mutants arrest early in oogenesis.) We conclude that the correct intracellular localization of Ago3, Aub, and Piwi is interdependent.

As reported previously for *aub* testes, Vasa protein expression increased in male germ cells in *ago3* testes, relative to the wild-type or *ago3/TM6B* (data not shown), consistent with the proposal that the *AT-chX-1* and *AT-chX-2* piRNAs downregulate *vasa* expression (Nishida et al., 2007); *AT-chX-1* piRNA levels were dramatically reduced in *ago3* testes (Figure 3A).

Silencing Selfish Genetic Elements in Germline Requires Ago3

In males, antisense piRNAs derived from the *Suppressor of Stellate* [*Su(Ste)*] locus on the Y chromosome silence the X-linked *Stellate* locus during spermatogenesis (Balakireva et al., 1992; Palumbo et al., 1994; Bozzetti et al., 1995; Aravin et al., 2001, 2003; Vagin et al., 2006). *Su(Ste)* piRNAs were the first piRNAs to be identified (Aravin et al., 2001), and, in *Su(Ste)* mutants, piRNAs targeting *Stellate* are lost, causing *Stellate* protein crystals to form in primary spermatocytes (Livak, 1984; Pimpinelli et al., 1985; Livak, 1990). *Stellate* silencing requires the piRNA pathway genes, *aub*, *spindle E*, *armitage*, *squash*, and *zucchini* (Schmidt et al., 1999; Stapleton et al., 2001; Tomari et al., 2004; Pane et al., 2007). *Stellate* protein crystals form in primary spermatocytes in *ago3* testes (Figure 2E).

Virtually all *Su(Ste)* piRNAs are antisense (Vagin et al., 2006) and bound to Aub, not Piwi (Nishida et al., 2007). *Su(Ste)* piRNAs fail to accumulate in *aub* mutants, but accumulate to higher than normal levels in *piwi* testes (Vagin et al., 2006). We used northern hybridization to examine *Su(Ste)* piRNA production in the testes (Figure 3A). *Su(Ste)* piRNAs, as well as AT-chX-1 piRNA, were strongly reduced in *ago3* males, relative to their heterozygous siblings, for all seven allelic combinations examined (Figure 3A and data not shown). Thus, both Aub and Ago3 are required to silence *Stellate*, and both proteins are required to produce or stabilize those *Su(Ste)* piRNAs normally bound to Aub. We speculate that *Stellate*-derived, Ago3-bound sense piRNAs, while rare, amplify Aub-bound, antisense *Su(Ste)* piRNAs in testes.

piRNAs have been best characterized in the *Drosophila* ovary. piRNAs derived from *roo* LTR retrotransposons are among the most abundant ovarian piRNAs and are disproportionately antisense to *roo* coding sequences; these antisense piRNAs are bound to Aub and Piwi (Vagin et al., 2006; Gunawardane et al., 2007), but not to Ago3 (Brennecke et al., 2007). *roo* antisense piRNAs failed to accumulate in *ago3* ovaries, but were readily detectable in *ago3/TM6B* ovaries (Figure 3B). Production of *roo*-derived piRNAs was rescued by a single-copy transgene expressing Ago3 from the *actin5c* promoter and more weakly rescued by a single-copy transgene expressing Ago3 from the *nanos* promoter (Figure 3C).

Genome-wide piRNA Analysis

To obtain a broader view of the function of Ago3 in piRNA biogenesis, we sequenced piRNAs from wild-type (Oregon R), *ago3/TM6B*, *ago3¹²/ago3¹³*, *aub/CyO*, and *aub^{HN2}/aub^{QC42}* ovaries. Libraries were prepared from oxidized small RNAs to permit greater effective sequencing depth. In all, 3,282,391 genome-matching, non-ncRNA, non-miRNA, 23–29 nt long small RNAs were sequenced (Table S2). Despite the large number of sequences obtained from each genotype, most piRNA species were sequenced only once. This remained true even when all data sets were pooled, suggesting that piRNAs comprise the most diverse class of regulatory molecules in the fly. Together, our piRNA sequences from wild-type, *ago3/TM6B*, and *ago3* ovaries cover 10% of the fly genome.

In parallel, we immunoprecipitated Ago3, Aub, and Piwi from *ago3/TM6B* ovaries and Aub and Piwi from *ago3* ovaries and then constructed and sequenced libraries of the small RNAs

bound to each protein (Table S2). We do not know the extent to which PIWI proteins copurify, e.g., because they are bound to a common RNA or present in a common complex. To avoid potential misassignment of piRNAs to a specific PIWI protein, we analyzed only those piRNAs that associated uniquely with Ago3, Aub, or Piwi.

We calibrated the abundance of the piRNAs uniquely associated with Ago3, Aub, or Piwi such that the aggregate abundance of a subset of piRNAs in an immunoprecipitation data set equaled that in the total piRNA data set from the corresponding genotype. This subset of piRNAs was defined as those that mapped only once to the fly genome and were sequenced at least once in both of the two data sets. This strategy allows direct comparison of the relative abundance of piRNAs uniquely bound to one PIWI protein with those bound to another, as well as comparison of the uniquely bound piRNAs between *ago3* heterozygotes and mutants.

Ago3 Limits Sense piRNA Accumulation and Amplifies Antisense piRNAs

Both *ago3* and *ago3/TM6B* ovaries had fewer piRNAs than wild-type ovaries (Figure 3D). Ago3 was previously found to bind mainly piRNAs corresponding to the sense, mRNA strand of transposons. Our data suggest that the intracellular concentration of Ago3 limits the accumulation of sense piRNAs: the median abundance sense piRNAs, analyzed by transposon family, in *ago3/TM6B* heterozygotes was ~40% that of wild-type ovaries. Antisense piRNAs were less affected by halving of the amount of Ago3: the median abundance of antisense, transposon-mapping piRNAs in *ago3/TM6B* ovaries was ~64% that of the wild-type. While complete loss of Ago3 further depressed the abundance of transposon-mapping, sense piRNAs, antisense piRNA accumulation in *ago3* ovaries collapsed. The median abundance by transposon family for antisense piRNAs in *ago3* ovaries was less than a tenth of the median abundance in the heterozygotes, and less than one-twentieth the median abundance in wild-type ovaries. The data support the view that for most selfish genetic element families, Ago3, presumably guided by sense piRNAs, acts to amplify antisense piRNAs bound to Aub.

Immunoprecipitation data confirmed this idea. In *ago3/TM6B* ovaries, 71% of small RNAs uniquely bound to Aub were antisense, but in *ago3* mutants, only 41% of the remaining Aub-bound piRNAs were antisense (Table S3). Thus, the characteristic antisense bias of Aub-bound piRNAs is enforced by Ago3. The strand bias of Piwi-bound piRNAs was less affected by loss of Ago3: in the *ago3* heterozygotes, 73% of the piRNAs uniquely associated with Piwi were antisense; in the mutants, 62% of the remaining piRNAs uniquely bound to Piwi were antisense. That the strandedness of Piwi-bound piRNAs changes at all suggests that Ago3 plays a role in the production of at least some antisense, Piwi-bound piRNAs.

Three piRNA Groups

While antisense piRNAs are generally more abundant than sense piRNAs, individual transposon families have distinct ratios of sense and antisense piRNAs. To determine the role of Ago3 in establishing these ratios, we analyzed the fraction of sense

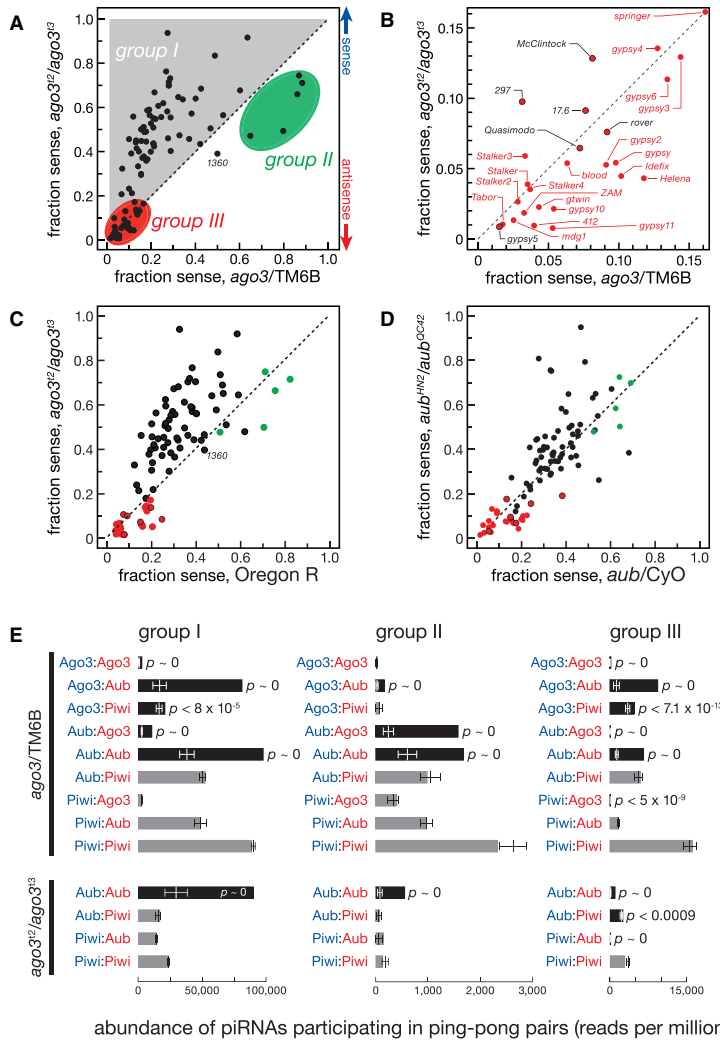


Figure 4. Three Distinct Groups of piRNAs

(A) The sense fraction of the piRNAs was compared for each transposon family between *ago3* and *ago3/TM6B* ovaries. (B) An enlargement of the lower-left corner of (A). Transposons labeled in red are present in the *flamenco* locus. Here and in (C) and (D), group III transposon families not present in *flamenco* are denoted with a red-filled black circle. (C) The sense fraction of piRNAs was compared between *ago3* and Oregon R ovaries. (D) The sense fraction of the piRNAs was compared between *aub* and *aub/CyO* ovaries. (E) The normalized abundance of ping-pong pairs detected for all nine (*ago3/TM6B*) or four (*ago3*) possible PIWI protein pairings for the piRNA species uniquely associated with a single PIWI protein. Blue, sense; red, antisense. Black bars, Bonferroni-corrected p value < 0.005 ; gray bars, p value > 0.005 . p values $< 10^{-30}$ are reported as " $p \sim 0$." Values are shown as average \pm two standard deviations.

R (Figure 4C). (For convenience, the enigmatic transposon family 1360 was included in group I in our subsequent analyses.)

Five transposon families compose group II. These all had more sense than antisense piRNAs in *ago3* heterozygotes, and four of the five had more sense than antisense piRNAs in wild-type ovaries, as previously noted (Brennecke et al., 2007). In *ago3* ovaries, both the fraction (Figures 4A, 4C, and S3) and the absolute amount (Figure 3E) of sense piRNAs declined for group II transposons, relative to heterozygous or wild-type ovaries.

piRNAs mapping to the 26 group III transposons were disproportionately antisense, ranging from 80% to nearly 100% antisense in *ago3* heterozygotes (Figures 4A and 4B). In *ago3* ovaries, group III transposons generally retained

their antisense piRNAs to a greater extent than those in groups I and II (Figures 3E and S3). Paradigmatic examples of each group—*HeT-A* for group I, *accord2* for group II, and *ZAM* for group III—are analyzed in greater detail in Figure 5. Figures S4–S6 present the corresponding data for each of the 95 transposons, as well as the tandem repeat *mst40*.

piRNAs (sense/[sense + antisense]) for each of the 95 transposon families for which we obtained ≥ 500 reads in *ago3/TM6B* ovaries. We compared the fraction of sense piRNAs in the heterozygotes to the fraction of sense piRNAs in *ago3* mutants (Figures 4A and 4B). We also compared the fraction of sense piRNAs in the heterozygotes to both the fraction of sense piRNAs in *ago3* mutants and the fold change in antisense piRNAs between the two genotypes (Figure S3).

We detected three groups of transposons. For 63 transposon families, the abundance of antisense piRNAs declined dramatically, causing an increase in the fraction of sense piRNAs (Figure S3). Among these 63 group I transposon families, 61 also had a decreased antisense bias comparing *ago3* to Oregon

Group I Transposons Require Ago3 for Antisense piRNA Amplification

HeT-A is the quintessential group I transposon (Figure 5). The piRNAs that map to group I transposons show a strong "ping-pong" signature that derives from the 10 nt overlap between antisense piRNAs bound to Aub and sense piRNAs bound to Ago3 and

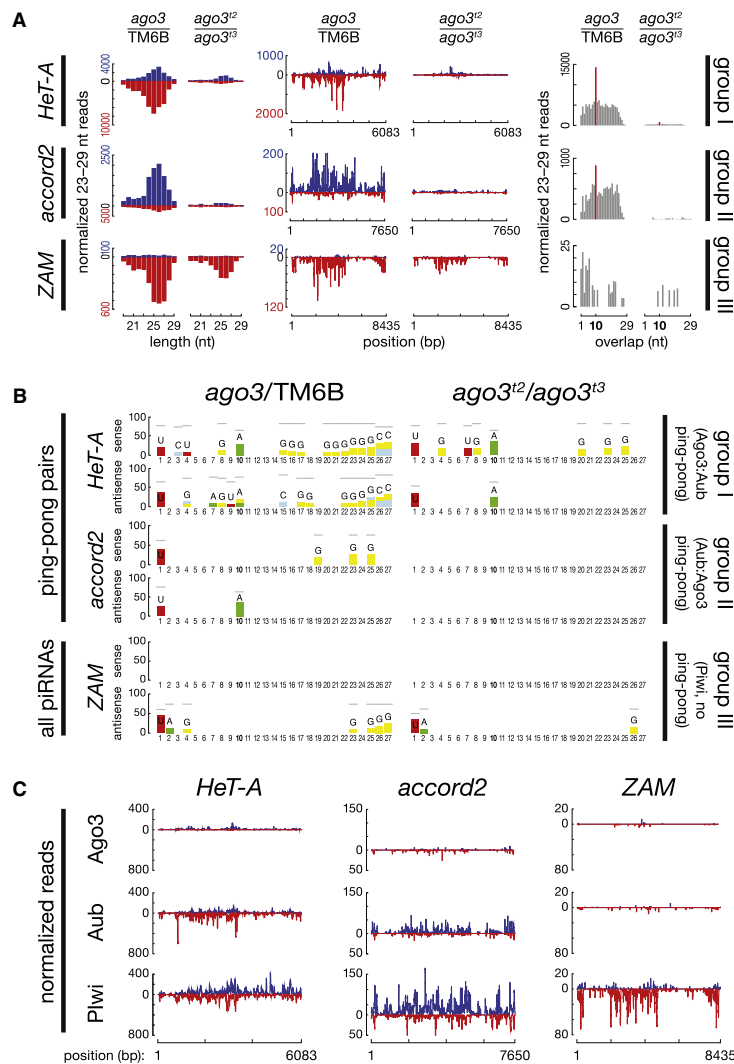


Figure 5. Paradigmatic Examples of Each Transposon Group

(A) piRNA length distribution, abundance relative to consensus position, and ping-pong pair abundance. Blue, sense piRNAs; red, antisense.

(B) Sequence-excess logos generated by subtracting the background from the relative frequency of each nucleotide at each position of the 23–29 nt RNAs in each sample. Only the nucleotide positions where foreground was significantly higher than background are shown (Bonferroni-corrected p values <0.001). Gray horizontal bars indicate the maximum possible value for each position.

(C) Normalized and calibrated abundance of piRNAs uniquely associated with Ago3, Aub, or Piwi in *ago3*/TM6B.

~35% that of the wild-type (Figure 3E). Sense piRNAs declined less when the abundance of Ago3 was halved. However, in the absence of Ago3, group II sense piRNAs declined ~17-fold from their levels in heterozygotes and ~27-fold from their levels in the wild-type. Group II antisense piRNAs also decreased, but less dramatically. The “backwards” behavior of group II piRNAs, relative to group I, suggests that the production or accumulation of Aub-bound piRNAs generally requires Ago3, irrespective of their sense or antisense identity.

piRNAs from group I and group II transposons normally partition among PIWI proteins according to their orientation and sequence bias (Figures S4–S6). Group I antisense (or group II sense) piRNAs typically begin with U (U1) and bind Aub, whereas group I sense (or group II antisense) piRNAs show less 5' nucleotide bias but typically bear an A at position 10 (A10) and bind Ago3 (e.g., *accord2* in Figure 5). In the absence of Ago3, both sense and antisense piRNAs remained associated with Aub

(Figures 4E). For group I transposons, although Ago3 associates almost exclusively with sense piRNAs, it is the Aub-bound antisense piRNAs that disappear in its absence (Figure S6). A central postulate of the ping-pong model is that the action of Ago3, rather than Aub, drives the production of more antisense than sense piRNAs (see Supplemental Discussion). Consistent with this proposal, the fraction of sense piRNAs for most group I transposons shows little overall change in *aub* ovaries (Figure 4D).

Group II Transposons Act “Backwards”

For group II transposons, antisense piRNA levels appear to reflect the abundance of Ago3: the median abundance by transposon family of group II antisense piRNAs in *ago3* heterozygotes was

(Figure S6). Some of these showed ping-pong pairing, but without respect to orientation, so that the U1 that normally characterizes antisense piRNAs and the A10 that normally characterizes sense piRNAs became conflated (Figures 5B and S4). In the absence of Ago3, those Aub-bound piRNAs with an A at position 10 began with U more often than expected by chance (Figure S7) and began with U about as often as did those Aub-bound piRNAs without an A at position 10. These observations suggest that a 5' U favors binding of a piRNA to Aub. Moreover, the data suggest that the A10 signature of Ago3-bound piRNAs is a consequence of the preponderance of Aub-bound piRNAs that start with U, consistent with the idea that Aub directs the production of Ago3-bound piRNAs.

piRNAs for Group III Transposons Are Produced by Both Aub- and Ago3-Dependent and Aub- and Ago3-Independent Pathways

In the absence of Ago3, the abundance of group III antisense piRNAs decreased far less than for group I (group I median = 0.06, group III median = 0.27; Figures 3E and S3). Antisense piRNAs from group III transposon families were mostly bound to Piwi: the median ratio between the amount of antisense piRNAs bound by Piwi and the amount bound by Aub in *ago3* ovaries was 31.3 for group III transposon families, a 5-fold increase from the median Piwi/Aub ratio for group III transposons in *ago3/TM6B* ovaries. This contrasts with the median Piwi/Aub ratios for group I (2.00) or II (2.94) transposon families (Figure S8A). Furthermore, loss of Ago3 caused group III antisense piRNAs bound to Piwi to decrease far less than antisense piRNAs bound to Aub; for group I, piRNAs bound to Piwi and Aub declined to a comparable extent (Figure S8B). Group III antisense piRNAs were also less affected by loss of Aub than the other two groups (Figures S9 and S10). Collectively, these observations suggest that group III transposons are predominantly silenced by a Piwi-dependent, Ago3-independent pathway.

While Piwi-bound piRNAs persisted in the absence of Ago3 to a greater extent than Aub-bound piRNAs, the absence of Ago3 clearly reduced the abundance of Piwi-bound piRNAs for all groups (Figure S8B). Perhaps some but not all Piwi-bound piRNAs are generated in the germline by an Ago3-directed amplification cycle. Supporting this view, we detected a statistically significant ($p < 7.1 \times 10^{-13}$) 10 nt overlap between sense Ago3-bound piRNAs and antisense Piwi-bound group III piRNAs (Figure 4E). Such group III Ago3 (sense):Piwi (antisense) piRNA ping-pong pairs were about half as abundant as Ago3:Aub pairs.

Most group III antisense piRNAs bound to Aub were eliminated in *ago3* mutants (Figure S8), suggesting that our group III piRNA data conflate at least two distinct pathways: a Piwi-dependent, Ago3- and Aub-independent pathway in which piRNA-directed piRNA amplification plays little if any role and an Aub-dependent pathway in which sense piRNAs bound to Ago3 act catalytically to amplify the antisense piRNAs associated with Aub. piRNAs acting in the Piwi-dependent pathway were disproportionately antisense, and this antisense bias required neither Ago3 nor Aub: production of antisense piRNAs for group III transposons remained essentially unchanged in *aub^{HN2}/aub^{QC42}* mutants (Figure 4D).

Group III Transposons Often Reside in the *flamenco* piRNA Cluster

Remarkably, of the 26 transposon families comprising group III, 20 are present in the *flamenco* locus (Pelisson et al., 1994; Brennecke et al., 2007). *flamenco* was originally identified as required for the silencing of *gypsy* (Prud'homme et al., 1995) and later shown to be required to silence *ZAM* and *Idefix* (Mével-Ninio et al., 2007; Desset et al., 2008). *gypsy*, *ZAM*, and *Idefix* are all group III transposons. Reporter experiments suggest that Piwi is required to silence these transposons (Sarot et al., 2004; Desset et al., 2008). Given that Piwi, but not Aub or Ago3, is readily detected in ovarian somatic follicle cells (Cox et al., 2000; Saito et al., 2006; Gunawardane et al., 2007; Nishida et al., 2007); that *gypsy* silencing requires *piwi* but not *aub* (Pelisson et al.,

2007); that the *gypsy* promoter drives expression of *gypsy* in the somatic follicle cells, which produce retrovirus-like particles that then infect the oocyte (Pelisson et al., 1994); that *ZAM* and *Idefix* are silenced in the somatic follicle cells by Piwi (Desset et al., 2008); and that follicle cell clones mutant for *piwi* or *flamenco* have remarkably similar defects (Cox et al., 1998; Mével-Ninio et al., 2007), it is tempting to speculate that group III transposons in general are repressed by a Piwi-dependent, Ago3- and Aub-independent pathway that operates in follicle cells. In this view, antisense, Piwi-bound piRNAs would provide the primary somatic cell defense against group III transposon expression, whereas an Aub- and Ago3-dependent pathway provides a secondary defense in the germline, the ultimate target of these transposons.

For groups I and II, loss of Ago3 switches the bias of both Aub and Piwi-bound piRNAs from strongly antisense-biased (or strongly sense-biased for group II) to slightly sense-biased (or antisense-biased for group II). That is, in wild-type flies, Ago3 acts to skew the strandedness of piRNA pools for group I and group II transposons. For group III transposons, however, loss of Ago3 slightly increased the antisense bias of piRNAs uniquely bound to Piwi: in *ago3/TM6B* ovaries, 88% of such piRNAs were antisense; in *ago3*, 90% were antisense (Table S3). We conclude that most Piwi-bound group III piRNAs are made directly from antisense transcripts such as the hypothesized precursor transcript that spans the *flamenco* locus. This explanation is consistent with the previous proposal that *flamenco* triggers silencing of transposons such as *gypsy*, *ZAM*, and *Idefix* in somatic follicle cells—a cell type that expresses little if any Ago3 or Aub.

Ago3 Amplifies piRNAs

Overall, piRNAs associated with Piwi were ~15-fold and piRNAs bound to Aub were ~6.4-fold more abundant than those bound to Ago3. The greater abundance of Piwi- and Aub-bound piRNAs is consistent both with the idea that Ago3 is less abundant than the other two PIWI proteins and with the proposal that Ago3 acts catalytically to amplify Piwi- and Aub-bound piRNAs. Consistent with the idea that group III piRNAs are largely Ago3-independent, the likelihood of a group III piRNA being associated with Piwi rather than Aub more than doubled in *ago3* ovaries: the Piwi-bound to Aub-bound piRNA ratio was 5.0 in *ago3/TM6B*, but 11.7 in *ago3*. This trend was not observed for group I piRNAs, where Piwi-bound piRNAs were twice as abundant as Aub-bound for *ago3/TM6B*, but only 1.6 times more abundant in *ago3* ovaries. Thus, for group III piRNAs, the absence of Ago3-catalyzed amplification shifts the piRNA pool toward the Piwi-dependent pathway.

Loss of Ago3 Increases Group I Transposon Expression

What is the molecular consequence of piRNA loss? We used whole-genome tiling microarrays to measure the effect of loss of Ago3 on gene and transposon expression (Figures 6A and S11) and quantitative RT-PCR of selected transposons to corroborate the microarray data (Figure 6B). The abundance of genic mRNA levels were generally unchanged in both *ago3* (Figure S11A) and *aub* (Figure S12A) ovaries, compared with wild-type controls (*w¹¹¹⁸*). In contrast, with a false-discovery

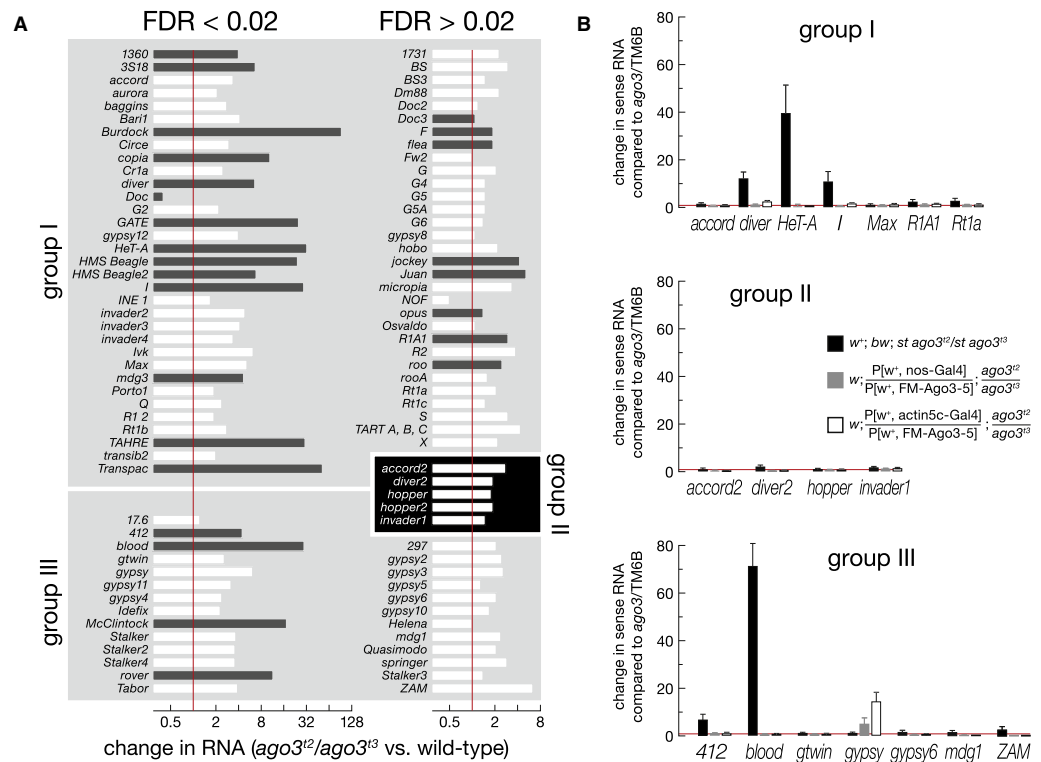


Figure 6. Loss of Ago3 Increases Expression of Some Group I and Group III, but not Group II, Transposons

(A) Expression of group I, group II, and group III transposon families in *ago3* ovaries, relative to wild-type (*w*¹¹¹⁸) ovaries, was assayed with whole-genome tiling microarrays. White bars show expression of the transposon family in both wild-type and *ago3* ovaries was less than that of the 50th percentile for expression of all mRNAs in wild-type, suggesting that expression change cannot be reliably quantified. Black bars show transposon families with expression greater than this threshold in one or both genotypes. Significant (false-discovery rate [FDR] <0.02) and nonsignificant data (FDR >0.02) are separated.

(B) Quantitative RT-PCR was used to assess transposon expression, relative to Actin, for *ago3* ovaries (black bars) and ovaries expressing one copy of a UAS-Ago3 transgene driven by *nanos*-Gal4 (gray bars) or *actin5c*-Gal4 (white bars), relative to *ago3*/TM6B ovaries. The figure reports mean \pm standard deviation for three independent biological samples.

rate (FDR) <0.02, the expression of 32 of the 64 group I transposons and 14 of the 26 group III transposons increased in the absence of Ago3 (Figures 6A and S11B). However, the level of expression for many of these in both wild-type and *ago3* ovaries was less than the 50th percentile of expression for mRNA in wild-type ovaries, making accurate quantification of their change in expression challenging. With this threshold for expression, 13 of the 64 group I and four of the 26 group III transposon families showed increased expression in the absence of Ago3 (Figure 6A).

Despite the loss of group II piRNAs in *ago3* mutants, expression of the five group II transposons was not significantly altered at a FDR <0.02 (Figure 6A). Perhaps the production of Aub-bound sense piRNAs for group II transposons is futile and does not silence these elements, since they were likewise not activated in an *aub* mutant (Figures S12B and S13).

Compared with wild-type ovaries, expression of 18 of the 64 group I transposon families increased in *aub* ovaries (eight of

these 18 had expression in both *w*¹¹¹⁸ and *ago3* greater than the 50th percentile for mRNA expression in the wild-type), including 14 that also increased in *ago3* ovaries (Figures S12B and S13). At a FDR <0.02, the expression of only two group III transposons—*rover* and *McClintock*—was increased; both were also desilenced in *ago3* ovaries. These data support the idea that group I transposons rely on Ago3 and Aub for silencing, whereas group III transposons are silenced by Piwi, with help from Ago3, likely in the germline only, for some transposons.

Germline Expression of Ago3 Rescues Group I Transposon Silencing

Our data suggest that Ago3 is required for piRNA production and silencing of group I transposons in the female *Drosophila* germline. As a final test of this hypothesis, we expressed Ago3 in *ago3* ovaries using the germline restricted *nanos* promoter. We used quantitative RT-PCR to measure the levels of mRNA for seven group I transposons, including five whose expression was

increased in the absence of Ago3: *HeT-A*, *Diver*, *I element*, *R1A1*, and *Rt1a*. For each, germline expression of Ago3 using the *nanos* promoter rescued silencing (Figure 6B).

Misexpression of Ago3 in the Soma Interferes with Group III Transposon Silencing

Argonaute proteins can compete for binding small RNAs. For example, decreasing the concentration of Ago1 in cultured S2 cells increases the loading of miRNAs into Ago2, whereas decreasing Ago2 increases miRNA loading into Ago1 (Förstmann et al., 2007; Horwich et al., 2007). Similarly, misexpression of Aub in the soma inhibits Ago2-mediated RNAi (Specchia et al., 2008). To test the role of Ago3 in silencing transposons in somatic ovarian cells, we used the *actin5c* promoter to express Ago3 in both the germline and the soma of *ago3* ovaries.

Surprisingly, ectopic expression of Ago3 in the soma increased the expression of the group III transposon *gypsy*, but enhanced silencing of the group I transposon *HeT-A* (Figure 6B). First, these data confirm our observation that increased Ago3 expression in the female germline increases silencing of group I transposons, perhaps because the concentration of Ago3 in the germline nurse cells limits piRNA amplification. Second, the data suggest that Ago3 cannot silence those transposons normally expressed in the somatic cells of the ovary. Perhaps Ago3 competes with Piwi for piRNAs, but, unlike Piwi, cannot act directly to silence transposons. Further supporting the idea that PIWI proteins compete for piRNAs, *gypsy* was silenced to an ~43-fold greater extent in *aub* ovaries than in *aub/CyO* (Figure S13). We speculate that when Aub levels are low, more *gypsy* piRNAs associate with Piwi, leading to enhanced silencing of *gypsy*.

DISCUSSION

Disentangling Multiple piRNA Pathways

Because ovaries contain both germline and somatic cells, our data conflate two distinct cell lineages. Combining our data with extensive genetic studies of *gypsy* and other transposon families represented in the *flamenco* locus, we have attempted to disentangle germline and somatic piRNA function (Figure 7). We propose that the somatic piRNA pathway is the more straightforward, involving only Piwi and not Ago3 or Aub. Hannon and coworkers (Malone et al., 2009) similarly deduce the existence of a somatic piRNA pathway by analyzing a broad panel of piRNA mutants and examining piRNAs from early embryos, which contain maternally deposited germline piRNAs but lack follicle cell-derived piRNAs. Thus, their study and ours infer from distinct data sets the existence of a somatic pathway in which primary piRNAs derived from *flamenco* are loaded directly into Piwi and not further amplified.

Our data suggest that Piwi cannot act alone to amplify piRNAs. We envision that Piwi-bound piRNAs in the soma are produced by a ribonuclease that randomly generates single-stranded guides that are subsequently loaded into Piwi and trimmed to length. Although Piwi-bound piRNAs generally begin with U and Piwi shows *in vitro* a preference for binding small RNA that begins with U (Yin and Lin, 2007), current evidence cannot distinguish between a putative piRNA-generating ribonuclease

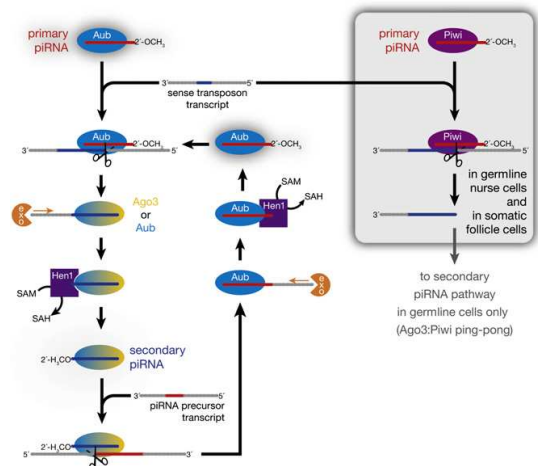


Figure 7. A Model for piRNA Biogenesis

The Aub- and Ago3-dependent piRNA amplification cycle is envisioned to operate only in the germline, whereas a Piwi-dependent, Aub- and Ago3-independent pathway is shown for somatic cells. In the germline, Piwi can also partner with Ago3 to amplify piRNAs.

cleaving mainly at U and Piwi selecting U1 piRNAs from a set of RNAs with all possible 5' nucleotides.

Without an amplification cycle to ensure an antisense bias, some other mechanism must operate to explain why Piwi-bound piRNAs are overwhelmingly antisense. A plausible but somewhat unsatisfying explanation comes from *flamenco* itself, whose constituent transposons are nearly all oriented in a single direction, so that the ~160 kb *flamenco* transcript is almost entirely antisense to the transposons. How such a nonrandom array of transposons could arise is unknown. Other nonrandomly oriented piRNA clusters may explain the smaller number of transposons in group III that are not present in *flamenco*.

The transposons in most piRNA clusters do not show such a pronounced nonrandom orientation. These probably act in the germline to produce primary piRNAs that load into Aub. The observed antisense bias of Aub-bound piRNAs arises subsequently, when Aub generates Ago3-bound secondary piRNAs and Ago3 acts, in turn, to produce Aub-bound secondary piRNAs. We propose that in the absence of Ago3, the sense/antisense ratio of Aub-bound piRNAs reverts to the inherent sense/antisense bias of the transposable element sequences present in the transcripts of piRNA clusters.

For this cycle to skew the Aub-bound piRNA population toward antisense, the substrate for cleavage by primary piRNA-bound Aub must be largely sense RNA. The best candidate for such sense RNA is mRNA derived from actively transcribed transposon copies. If such sense mRNA were largely found in the cytoplasm, it would be spatially segregated from the cluster transcripts, which we envision to be retained in the nucleus. Supporting this idea, sense transcripts from the group I transposon, *I element*, normally accumulate only in the nuclei of germline nurse cells (Chambeyron et al., 2008), likely because

they are destroyed in the cytoplasm by Aub-bound primary piRNAs and Aub-bound piRNAs produced by Ago3-dependent amplification. In the absence of Aub, these sense transposon transcripts accumulate in the cytoplasm instead, consistent with the strong desilencing of *I* element in *aub* and *ago3* mutants (Figures 6, S12, and S13) (Vagin et al., 2006).

Nuage and the Paradox of piRNA Production

The piRNA ping-pong hypothesis predicts a role for Ago3 in the production of Aub-bound antisense piRNAs, but our finding that loss of Ago3 also reduced the abundance of Piwi-bound antisense piRNAs was unexpected. The majority of Aub and Ago3 is found in nuage and in the cytoplasm, but Piwi is predominantly nuclear. How then can Ago3 direct the production of Piwi-bound piRNAs? Perhaps Piwi transits the nuage en route from its site of synthesis, the cytoplasm, to where it accumulates, the nucleus. In this view, cytoplasmic Piwi is predicted to lack a small RNA guide. Piwi would then acquire its small RNA guide in the nuage, through a process that requires Ago3. Loading a piRNA into Piwi might then license it for entry into the nucleus, where it could act posttranscriptionally or transcriptionally to silence transposon expression. In this view, mutations in genes required for nuage assembly or stability, such as *vasa*, as well as genes required for Piwi loading would reduce the amount of nuclear Piwi. A similar mechanism may operate in mammals, where the PIWI protein MILI is found in cytoplasmic granules, whereas MIWI2 is nuclear. In the absence of MILI, MIWI2 delocalizes from the nucleus to the cytoplasm, although MIWI2 is not required for the localization of MILI (Aravin et al., 2008).

Such a model cannot explain the loading of Piwi in the somatic follicle cells, which contain little or no Ago3 or Aub and which do not contain nuage. A simple but untested hypothesis for these cells is that in the absence of nuage, empty Piwi readily enters the nucleus, where it obtains its small RNA guide. We might reasonably expect that in germ cells the absence of nuage would impair the loading of Piwi by eliminating the Ago3-dependent, germline-specific Piwi-loading process, but also facilitate entry of some empty Piwi into the nucleus, where it could obtain small RNA guides. Consistent with this idea, we do detect some Piwi in the nucleus in *ago3* ovaries. The simplicity of this hypothesis, of course, belies the complexity of testing it.

Why Two Distinct piRNA Production Pathways?

Retrotransposons “reproduce” by producing sense RNA encoding transposases and other proteins that allow them to jump to new locations in the germ cell genome. The conservation of the piRNA ping-pong cycle in animals (Grimson et al., 2008) suggests that it is an ancient and conserved germline defense against retrotransposition. In flies, the *gypsy* family of retroelements appears to have moved its reproductive cycle to the somatic follicle cells adjacent to the germline, which it infects using retrovirus-like particles. *gypsy* thus appears to avoid germline piRNA surveillance by transcribing and packaging its RNA in the soma. Perhaps expression of Piwi in *Drosophila* follicle cells reflects an adaptive evolutionary counter move to the *gypsy* reproductive strategy. The simplicity of the direct loading of Piwi with antisense piRNAs derived from *flamenco* may have made this counter defense more evolutionarily accessible than

a strategy requiring expression of all the proteins needed for the Ago3:Aub ping-pong mechanism. In the future, more extensive analysis of the cellular and genetic requirements for ping-pong-independent and ping-pong-dependent piRNA mechanisms in *Drosophila melanogaster* and in more ancient animal species may provide a test for these ideas.

EXPERIMENTAL PROCEDURES

Detailed experimental and computational procedures are described in Supplemental Data.

Isolation of *ago3⁴¹*, *ago3⁴²*, and *ago3⁴³* Alleles

From the 6000 EMS-mutagenized fly lines (Koundakjian et al., 2004) screened by the Seattle TILLING project (<http://tilling.fhcr.org/>), candidate lines that contained mutations that induced premature stop codons in the *ago3* coding sequence were characterized by sequencing of genomic PCR amplicons.

Molecular Cloning and Generation of Transgenic Flies

Ago3 was cloned from Oregon R ovary complementary DNA (cDNA) by 3' RACE. Ovary RNA was reverse transcribed. First strand cDNA was treated with RNase H. Full-length Ago3 coding sequence was amplified by PCR, cloned, and recombined into the pPFMW vector. Insert junctions were confirmed by sequencing. Plasmid DNA was injected, and transgenic flies were identified.

Immunohistochemistry and Microscopy

Egg chamber fixation and whole-mount antibody labeling were performed as previously described (Theurkauf, 1994). Samples were analyzed with a Leica TCS-SP inverted laser-scanning microscope, with identical imaging conditions for each set of wild-type and mutant.

Immunoprecipitation

Immunoprecipitation was performed with anti-Ago3, anti-Aub, or anti-Piwi antibodies bound to protein G sepharose beads. Input, supernatant, and bound samples were subject to western blotting analysis to confirm immunoprecipitation.

Small RNA Cloning and Sequencing

Total RNA was isolated from manually dissected ovaries from 2- to 4-day-old flies. After 2S rRNA depletion, 18–29 nt small RNA was purified and oxidized, followed by ethanol precipitation. 3' ligated product was purified from a denaturing urea-polyacrylamide gel and then ligated to a 5' RNA adaptor. The small RNA library was amplified and then purified from an agarose gel. Purified libraries were sequenced with a Solexa Genome Analyzer (Illumina, San Diego, CA).

ACCESSION NUMBERS

Sequence data are available via the NCBI trace archives (<http://www.ncbi.nlm.nih.gov/Traces/>) with accession number SRP000458. Microarray data are available via the NCBI gene expression omnibus (<http://www.ncbi.nlm.nih.gov/geo/>) using accession number GSE114370.

SUPPLEMENTAL DATA

Supplemental Data include Supplemental Discussion, Supplemental Experimental Procedures, six tables, and 13 figures and can be found with this article online at [http://www.cell.com/supplemental/S0092-8674\(09\)00452-8](http://www.cell.com/supplemental/S0092-8674(09)00452-8).

ACKNOWLEDGMENTS

We thank Mikiko and Haruhiko Siomi and Greg Hannon for antibodies, Erica Selva for flies expressing *nanos*-Gal4 from a transgene on the second chromosome, Paul Lasko for GFP-Vasa flies, Bradley Till for assistance with tilling

primer design, Alicia Boucher for assistance with fly husbandry, Gwen Farley for technical assistance, and members of the Zamore and Weng laboratories for advice, suggestions, and critical comments on the text. This work was supported in part by grants from the National Institutes of Health to P.D.Z. (GM62862 and GM65236) and W.E.T. (GM050898 and HD049116) and from the Natural Sciences and Engineering Research Council of Canada and the Canadian Institutes of Health Research to B.M.H. H.X. was supported in part by Pfizer, Inc.

Received: January 20, 2009

Revised: February 24, 2009

Accepted: April 10, 2009

Published online: April 23, 2009

REFERENCES

- Aravin, A.A., Naumova, N.M., Tulin, A.V., Vagin, V.V., Rozovsky, Y.M., and Gvozdev, V.A. (2001). Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the *D. melanogaster* germline. *Curr. Biol.* **11**, 1017–1027.
- Aravin, A.A., Lagos-Quintana, M., Yalcin, A., Zavolan, M., Marks, D., Snyder, B., Gaasterland, T., Meyer, J., and Tuschl, T. (2003). The small RNA profile during *Drosophila melanogaster* development. *Dev. Cell* **5**, 337–350.
- Aravin, A., Gaidatzis, D., Pfeffer, S., Lagos-Quintana, M., Landgraf, P., Iovino, N., Morris, P., Brownstein, M.J., Kuramochi-Miyagawa, S., Nakano, T., et al. (2006). A novel class of small RNAs bind to MILI protein in mouse testes. *Nature* **442**, 203–207.
- Aravin, A.A., Sachidanandam, R., Bourc'his, D., Schaefer, C., Pezic, D., Toth, K.F., Bestor, T., and Hannon, G.J. (2008). A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. *Mol. Cell* **31**, 785–799.
- Balakireva, M.D., Shevelov, Y., Nurminsky, D.I., Livak, K.J., and Gvozdev, V.A. (1992). Structural organization and diversification of Y-linked sequences comprising *Su(Ste)* genes in *Drosophila melanogaster*. *Nucleic Acids Res.* **20**, 3731–3736.
- Batista, P.J., Ruby, J.G., Claycomb, J.M., Chiang, R., Fahlgren, N., Kasschau, K.D., Chaves, D.A., Gu, W., Vasale, J.J., Duan, S., et al. (2008). PRG-1 and 21U-RNAs interact to form the piRNA complex required for fertility in *C. elegans*. *Mol. Cell* **31**, 67–78.
- Bozzetti, M.P., Massari, S., Finelli, P., Meggio, F., Pinna, L.A., Boldyreff, B., Issinger, O.G., Palumbo, G., Ciriaco, C., Bonaccorsi, S., et al. (1995). The *Ste* locus, a component of the parasitic cry-Ste system of *Drosophila melanogaster*, encodes a protein that forms crystals in primary spermatocytes and mimics properties of the beta subunit of casein kinase 2. *Proc. Natl. Acad. Sci. USA* **92**, 6067–6071.
- Brennecke, J., Aravin, A.A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., and Hannon, G.J. (2007). Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell* **128**, 1089–1103.
- Chambeyron, S., Popkova, A., Payen-Groschene, G., Brun, C., Laouini, D., Pelisson, A., and Bucheton, A. (2008). piRNA-mediated nuclear accumulation of retrotransposon transcripts in the *Drosophila* female germline. *Proc. Natl. Acad. Sci. USA* **105**, 14964–14969.
- Cook, H.A., Koppetsch, B.S., Wu, J., and Theurkauf, W.E. (2004). The *Drosophila* SDE3 homolog armitage is required for oskar mRNA silencing and embryonic axis specification. *Cell* **116**, 817–829.
- Cooper, J.L., Till, B.J., and Henikoff, S. (2008). Fly-TILL: Reverse genetics using a living point mutation resource. *Fly (Austin)* **2**, 300–302.
- Cox, D.N., Chao, A., Baker, J., Chang, L., Qiao, D., and Lin, H. (1998). A novel class of evolutionarily conserved genes defined by *piwi* are essential for stem cell self-renewal. *Genes Dev.* **12**, 3715–3727.
- Cox, D.N., Chao, A., and Lin, H. (2000). *piwi* encodes a nucleoplasmic factor whose activity modulates the number and division rate of germline stem cells. *Development* **127**, 503–514.
- Das, P.P., Bagijn, M.P., Goldstein, L.D., Woolford, J.R., Lehrbach, N.J., Sapetschnig, A., Buhecha, H.R., Gilchrist, M.J., Howe, K.L., Stark, R., et al. (2008). Piwi and piRNAs act upstream of an endogenous siRNA pathway to suppress Tc3 transposon mobility in the *Caenorhabditis elegans* germline. *Mol. Cell* **31**, 79–90.
- Desset, S., Buchon, N., Meignin, C., Coiffet, M., and Vaury, C. (2008). In *Drosophila melanogaster* the COM locus directs the somatic silencing of two retrotransposons through both Piwi-dependent and -independent pathways. *PLoS ONE* **3**, e1526.
- Findley, S.D., Tamanaha, M., Clegg, N.J., and Ruohola-Baker, H. (2003). Maelstrom, a *Drosophila* spindle-class gene, encodes a protein that colocalizes with Vasa and RDE1/AGO1 homolog, Aubergine, in nuage. *Development* **130**, 859–871.
- Förstemann, K., Horwich, M.D., Wee, L.-M., Tomari, Y., and Zamore, P.D. (2007). *Drosophila* microRNAs are sorted into functionally distinct Argonaute protein complexes after their production by Dicer-1. *Cell* **130**, 287–297.
- Fuller, M.T. (1993). Spermatogenesis. In *The Development of Drosophila melanogaster*, M. Bate and A.A. Martinez, eds. (Plainview, N.Y.: Cold Spring Harbor Laboratory Press), pp. 71–148.
- Ghildiyal, M., and Zamore, P.D. (2009). Small silencing RNAs: an expanding universe. *Nat. Rev. Genet.* **10**, 94–108.
- Girard, A., Sachidanandam, R., Hannon, G.J., and Carmell, M.A. (2006). A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* **442**, 199–202.
- Grimson, A., Srivastava, M., Fahey, B., Woodcroft, B.J., Chiang, H.R., King, N., Degnan, B.M., Rokhsar, D.S., and Bartel, D.P. (2008). Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. *Nature* **455**, 1193–1197.
- Gunawardane, L.S., Saito, K., Nishida, K.M., Miyoshi, K., Kawamura, Y., Nagami, T., Siomi, H., and Siomi, M.C. (2007). A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in *Drosophila*. *Science* **315**, 1587–1590.
- Harris, A.N., and Macdonald, P.M. (2001). *aubergine* encodes a *Drosophila* polar granule component required for pole cell formation and related to eIF2C. *Development* **128**, 2823–2832.
- Horwich, M.D., Li, C., Matranga, C., Vagin, V., Farley, G., Wang, P., and Zamore, P.D. (2007). The *Drosophila* RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC. *Curr. Biol.* **17**, 1265–1272.
- Houwing, S., Kamminga, L.M., Berezikov, E., Cronembold, D., Girard, A., van den Elst, H., Filipov, D.V., Blaser, H., Raz, E., Moens, C.B., et al. (2007). A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish. *Cell* **129**, 69–82.
- Klattehoff, C., Bratu, D.P., McGinnis-Schultz, N., Koppetsch, B.S., Cook, H.A., and Theurkauf, W.E. (2007). *Drosophila* rasiRNA pathway mutations disrupt embryonic axis specification through activation of an ATR/Chk2 DNA damage response. *Dev. Cell* **12**, 45–55.
- Koundakjian, E.J., Cowan, D.M., Hardy, R.W., and Becker, A.H. (2004). The Zuker collection: a resource for the analysis of autosomal gene function in *Drosophila melanogaster*. *Genetics* **167**, 203–206.
- Lau, N.C., Seto, A.G., Kim, J., Kuramochi-Miyagawa, S., Nakano, T., Bartel, D.P., and Kingston, R.E. (2006). Characterization of the piRNA complex from rat testes. *Science* **313**, 363–367.
- Liang, L., Diehl-Jones, W., and Lasko, P. (1994). Localization of vasa protein to the *Drosophila* pole plasm is independent of its RNA-binding and helicase activities. *Development* **120**, 1201–1211.
- Lim, A.K., and Kai, T. (2007). Unique germ-line organelle, nuage, functions to repress selfish genetic elements in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* **104**, 6714–6719.
- Livak, K.J. (1984). Organization and mapping of a sequence on the *Drosophila melanogaster* X and Y chromosomes that is transcribed during spermatogenesis. *Genetics* **107**, 611–634.
- Livak, K.J. (1990). Detailed structure of the *Drosophila melanogaster stellate* genes and their transcripts. *Genetics* **124**, 303–316.

- Malone, C.D., Brennecke, J., Dus, M., Stark, A., McCombie, W.R., Sachidanandam, R., and Hannon, G.J. (2009). Specialized piRNA pathways act in germline and somatic tissues of the *Drosophila* ovary. *Cell* 137, in press. 10.1016/j.cell.2009.03.040.
- Mevel-Nirio, M., Pelisson, A., Kinder, J., Campos, A.R., and Bucheton, A. (2007). The flamenco locus controls the gypsy and ZAM retroviruses and is required for *Drosophila* oogenesis. *Genetics* 175, 1615–1624.
- Nishida, K.M., Saito, K., Mori, T., Kawamura, Y., Nagami-Okada, T., Inagaki, S., Siomi, H., and Siomi, M.C. (2007). Gene silencing mechanisms mediated by Aubergine piRNA complexes in *Drosophila* male gonad. *RNA* 13, 1911–1922.
- Palumbo, G., Bonaccorsi, S., Robbins, L.G., and Pimpinelli, S. (1994). Genetic analysis of *Stellate* elements of *Drosophila melanogaster*. *Genetics* 138, 1181–1197.
- Pane, A., Wehr, K., and Schupbach, T. (2007). zucchini and squash encode two putative nucleases required for rasiRNA production in the *Drosophila* germline. *Dev. Cell* 12, 851–862.
- Pelisson, A., Song, S.U., Prud'homme, N., Smith, P.A., Bucheton, A., and Corces, V.G. (1994). Gypsy transposition correlates with the production of a retroviral envelope-like protein under the tissue-specific control of the *Drosophila* flamenco gene. *EMBO J.* 13, 4401–4411.
- Pelisson, A., Sarot, E., Payen-Groschene, G., and Bucheton, A. (2007). A novel repeat-associated small interfering RNA-mediated silencing pathway downregulates complementary sense gypsy transcripts in somatic cells of the *Drosophila* ovary. *J. Virol.* 81, 1951–1960.
- Pimpinelli, S., Sullivan, W., Prout, M., and Sandler, L. (1985). On biological functions mapping to the heterochromatin of *Drosophila melanogaster*. *Genetics* 109, 701–724.
- Prud'homme, N., Gans, M., Masson, M., Terzian, C., and Bucheton, A. (1995). Flamenco, a gene controlling the gypsy retrovirus of *Drosophila melanogaster*. *Genetics* 139, 697–711.
- Saito, K., Nishida, K.M., Mori, T., Kawamura, Y., Miyoshi, K., Nagami, T., Siomi, H., and Siomi, M.C. (2006). Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the *Drosophila* genome. *Genes Dev.* 20, 2214–2222.
- Sarot, E., Payen-Groschene, G., Bucheton, A., and Pelisson, A. (2004). Evidence for a piwi-dependent RNA silencing of the gypsy endogenous retrovirus by the *Drosophila melanogaster* flamenco gene. *Genetics* 166, 1313–1321.
- Schmidt, A., Palumbo, G., Bozzetti, M.P., Tritto, P., Pimpinelli, S., and Schafer, U. (1999). 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.
- Schupbach, T., and Wieschaus, E. (1991). Female sterile mutations on the second chromosome of *Drosophila melanogaster*. II. Mutations blocking oogenesis or altering egg morphology. *Genetics* 129, 1119–1136.
- Specchia, V., Benna, C., Mazzotta, G.M., Piccin, A., Zordan, M.A., Costa, R., and Bozzetti, M.P. (2008). *aubergine* gene overexpression in somatic tissues of *aubergine* mutants interferes with the RNAi pathway of a *yellow* hairpin dsRNA in *Drosophila melanogaster*. *Genetics* 178, 1271–1282.
- Stapleton, W., Das, S., and McKee, B.D. (2001). A role of the *Drosophila* *homeless* gene in repression of *Stellate* in male meiosis. *Chromosoma* 110, 228–240.
- Theurkauf, W.E. (1994). Immunofluorescence analysis of the cytoskeleton during oogenesis and early embryogenesis. *Methods Cell Biol.* 44, 489–505.
- Tomari, Y., Du, T., Haley, B., Schwarz, D.S., Bennett, R., Cook, H.A., Koppetsch, B.S., Theurkauf, W.E., and Zamore, P.D. (2004). RISC assembly defects in the *Drosophila* RNAi mutant armitage. *Cell* 116, 831–841.
- Vagin, V.V., Sigova, A., Li, C., Seitz, H., Gvozdev, V., and Zamore, P.D. (2006). A distinct small RNA pathway silences selfish genetic elements in the germline. *Science* 313, 320–324.
- Wang, G., and Reinke, V. (2008). A *C. elegans* Piwi, PRG-1, regulates 21U-RNAs during spermatogenesis. *Curr. Biol.* 18, 861–867.
- Williams, R.W., and Rubin, G.M. (2002). ARGONAUTE1 is required for efficient RNA interference in *Drosophila* embryos. *Proc. Natl. Acad. Sci. USA* 99, 6889–6894.
- Wilson, J.E., Connell, J.E., and Macdonald, P.M. (1996). *aubergine* enhances *oskar* translation in the *Drosophila* ovary. *Development* 122, 1631–1639.
- Yin, H., and Lin, H. (2007). An epigenetic activation role of Piwi and a Piwi-associated piRNA in *Drosophila melanogaster*. *Nature* 450, 304–308.

4.3 Discussion : qu'est-ce qui fait un locus de piRNA ?

Ces travaux ont permis de tester expérimentalement le modèle du ping-pong pour la biogenèse des piRNA. Ils en ont précisé les règles et les limites : les piRNA effecteurs de la répression des transposons (et qui leur sont antisens) ségrègent en deux catégories, les piRNA chargés sur Aub (dont la biogenèse nécessite Ago3), dans l'ovocyte, et les piRNA chargés sur Piwi (dont la biogenèse dépend d'un ping-pong avec Ago3 dans une faible mesure, dans l'ovocyte ; la plupart de ces piRNA semblent produits par un mécanisme différent, dans les cellules folliculaires).

Cette étude a donc mis en évidence une activité des piRNA dans des cellules somatiques ; un autre laboratoire aboutissait à la même conclusion, par une approche différente (en comparant les piRNA d'ovaires entiers — qui contiennent ovocytes et cellules somatiques — aux piRNA d'œufs fraîchement déposés — qui reflètent probablement le contenu de l'ovocyte — Malone *et al.*, 2009 ont également montré que les piRNA du locus *flamenco* sont exprimés dans les cellules somatiques ; des données de génétique plus anciennes avaient déjà montré que *flamenco* réprime ses cibles principalement dans les cellules folliculaires : Péliesson *et al.*, 1994).

Ce locus semble spécialisé dans la répression des rétrotransposons de la famille *gypsy*, par des piRNA chargés sur Piwi et générés sans intervention d'Ago3. Les rétrotransposons *gypsy* de la *Drosophile* transposent par un mécanisme original : ils ne sont pas transcrits dans l'ovocyte, mais dans les cellules folliculaires, et infectent l'ovocyte par des particules semblables aux particules virales, bourgeonnées par les cellules folliculaires (Kim *et al.*, 1994 ; Péliesson *et al.*, 1994). Dans la « course à l'armement » qui oppose les transposons aux systèmes de défense chez les hôtes, la répression guidée par des piRNA issus de *flamenco*, maturés sans intervention du ping-pong, et chargés sur Piwi, aurait donc pu apparaître à la suite de cette innovation évolutive de la part des rétrotransposons ; l'apparition de cette deuxième catégorie de piRNA serait même récente, puisqu'un autre Insecte, le Lépidoptère *Bomby mori*, en est dépourvu (Kawaoka *et al.*, 2009).

Si le modèle du ping-pong explique comment les piRNA sont amplifiés, et comment ils sont enrichis en séquences antisens à la plupart des transposons (donc : susceptibles de les réprimer en guidant les protéines Piwi contre les ARN_m des transposons), il n'explique pas comment ce cycle est amorcé. D'après ce modèle, les piRNA chargés sur Aub sont maturés par Ago3, mais les piRNA chargés sur Ago3 sont maturés par Aub (il faudrait donc qu'une population de piRNA primaires puisse initier le cycle : d'après Brennecke *et al.*, 2008, ces piRNA amorceurs pourraient être déposés dans l'œuf par la mère — et l'origine exacte des premiers piRNA dans l'Histoire resterait mystérieuse). Alternativement, seule Ago3 participerait à la maturation des piRNA : Aub ne serait pas impliquée dans la biogenèse des piRNA chargés sur Ago3, même si la complémentarité de ses piRNA avec les précurseurs des piRNA chargés sur Ago3 semble le permettre. Les piRNA chargés sur Ago3 seraient alors générés par un mécanisme dédié, qui n'impliquerait pas Aub.

Ce modèle n'explique pas non plus ce qui distingue les loci de piRNA du reste du génome : il suppose que les précurseurs de piRNA sont transcrits à partir de loci spécifiques (les « clusters » de piRNA, des régions génomiques limitées d'où proviennent la plupart des piRNA) ainsi que des copies dispersées des transposons. Mais qu'est-ce qui détermine l'entrée d'un transcrit dans la voie des piRNA ? Pourquoi les longs transcrits des clusters de piRNA sont-ils orientés vers la machinerie de biogenèse des piRNA, alors que les transcrits du reste du génome y échappent ?

La protéine Rhino (un homologue de HP1, *heterochromatic protein 1*, protéine de l'hétérochromatine n°1) pourrait être un marqueur de loci à piRNA, dont la présence sur une région chromosomique pourrait adresser les transcrits vers la voie de biogenèse des piRNA. Cette protéine est particulièrement concentrée sur deux loci à piRNA dans l'ovocyte de *Drosophile*, et sa mutation perturbe l'accumulation des piRNA : elle abolit presque totalement l'expression des piRNA

issus des loci symétriques ³, alors qu'elle a un effet limité sur les loci dont les piRNA proviennent d'un seul brin. Ces deux types de loci expriment donc des piRNA selon des mécanismes différents, et Rhino, en tant que protéine de la chromatine, pourrait marquer spécifiquement l'un des deux types, et orienter ses transcrits vers la voie de biogenèse des piRNA.

Ces travaux sont présentés dans l'article qui suit (Klattenhoff *et al.*, 2009).

³Dans ces loci, les piRNA proviennent des deux brins génomiques.

4.4 Article n°6 : biogenèse des piRNA guidée par des modifications de la chromatine

Cell

The *Drosophila* HP1 Homolog Rhino Is Required for Transposon Silencing and piRNA Production by Dual-Strand Clusters

Carla Klattenhoff,¹ Hualin Xi,^{3,7,8} Chengjian Li,² Soohyun Lee,^{3,7} Jia Xu,^{3,6} Jaspreet S. Khurana,¹ Fan Zhang,¹ Nadine Schultz,¹ Birgit S. Koppetsch,¹ Anetta Nowosielska,¹ Herve Seitz,^{2,4,5} Phillip D. Zamore,^{2,*} Zhiping Weng,^{3,*} and William E. Theurkauf^{1,*}

¹Program in Molecular Medicine and Program in Cell Dynamics

²Department of Biochemistry and Molecular Pharmacology and Howard Hughes Medical Institute

³Program in Bioinformatics and Integrative Biology

University of Massachusetts Medical School, Worcester MA, 01605, USA

⁴Laboratoire de Biologie Moléculaire Eucaryote, Université de Toulouse, Université Paul Sabatier, F-31000 Toulouse, France

⁵Laboratoire de Biologie Moléculaire Eucaryote, Centre National de la Recherche Scientifique, F-31000 Toulouse, France

⁶Department of Biomedical Engineering

⁷Bioinformatics Program

Boston University, Boston, MA 02215, USA

⁸Computational Science Center of Emphasis, Pfizer Inc, 620 Memorial Drive, Cambridge, MA 02139, USA

*Correspondence: phillip.zamore@umassmed.edu (P.D.Z.), zhiping.weng@umassmed.edu (Z.W.), william.theurkauf@umassmed.edu (W.E.T.)
DOI 10.1016/j.cell.2009.07.014

SUMMARY

Piwi-interacting RNAs (piRNAs) silence transposons and maintain genome integrity during germline development. In *Drosophila*, transposon-rich heterochromatic clusters encode piRNAs either on both genomic strands (dual-strand clusters) or predominantly one genomic strand (uni-strand clusters). Primary piRNAs derived from these clusters are proposed to drive a ping-pong amplification cycle catalyzed by proteins that localize to the perinuclear nuage. We show that the HP1 homolog Rhino is required for nuage organization, transposon silencing, and ping-pong amplification of piRNAs. *rhi* mutations virtually eliminate piRNAs from the dual-strand clusters and block production of putative precursor RNAs from both strands of the major 42AB dual-strand cluster, but not of transcripts or piRNAs from the uni-strand clusters. Furthermore, Rhino protein associates with the 42AB dual-strand cluster, but does not bind to uni-strand cluster 2 or *flamenco*. Rhino thus appears to promote transcription of dual-strand clusters, leading to production of piRNAs that drive the ping-pong amplification cycle.

INTRODUCTION

Mutations in the *Drosophila* piwi-interacting RNA (piRNA) pathway disrupt transposon silencing, cause DNA break accumulation during female germline development, and lead to defects in posterior and dorsoventral axis specification (Brennecke et al., 2007; Chambeyron et al., 2008; Klattenhoff et al.,

2007; Vagin et al., 2006). The axis specification defects associated with piRNA pathway mutations are dramatically suppressed by mutations in *mnk* and *mei-41*, which encode Chk2 and ATR kinase homologs that function in DNA damage signaling (Chen et al., 2007; Klattenhoff et al., 2007; Pane et al., 2007). The developmental defects linked to piRNA pathway mutations thus appear to be secondary to DNA damage, which may result from transposon mobilization. PIWI proteins bind piRNAs, and mutations in genes encoding mouse and Zebrafish *piwi* homologs lead to transposon overexpression and germline-specific apoptosis (Carmell et al., 2007; Houwing et al., 2007), which could be triggered by DNA damage. The piRNA pathway may therefore have a conserved function in transposon silencing and maintenance of germline genome integrity.

Drosophila piRNAs appear to be derived from transposon rich clusters, most of which are localized in pericentromeric and subtelomeric heterochromatin (Brennecke et al., 2007). The majority of clusters produce piRNAs from both genomic strands (dual-strand clusters). However, two major clusters on the X chromosome produce piRNAs predominantly from one genomic strand (uni-strand clusters) (Brennecke et al., 2007; Brennecke et al., 2008). One of these uni-strand clusters maps to *flamenco*, a locus required for transposon silencing in the somatic follicle cells (Brennecke et al., 2007; Mevel-Ninio et al., 2007; Pelisson et al., 2007; Pelisson et al., 1994; Prud'homme et al., 1995; Sarot et al., 2004). The *flamenco* cluster contains fragments of a number of transposons, including *Zam*, *idefix*, and *gypsy*, and *flamenco* mutations disrupt silencing of these transposons (Desset et al., 2008; Mevel-Ninio et al., 2007; Prud'homme et al., 1995). In addition, transgenes carrying fragments of transposons in this cluster show *flamenco*-dependent silencing (Sarot et al., 2004). These findings suggest that piRNAs encoded by *flamenco* trans-silence complementary transposons located outside this cluster (Brennecke et al., 2007).

The mechanism of *trans*-silencing by piRNA is not well understood. piRNA-Piwi protein complexes catalyze homology-dependent target cleavage, suggesting that target transposon mRNAs are cotranscriptionally or posttranscriptionally degraded (Gunawardane et al., 2007; Saito et al., 2006). However, several *Drosophila* piRNA pathway mutations have been reported to modify position effect variegation (PEV) (Brower-Toland et al., 2007; Pal-Bhadra et al., 2002; Pal-Bhadra et al., 2004), which is linked to spreading of transcriptionally silent heterochromatin from pericentric and telomeric regions (Girton and Johansen, 2008). Piwi protein also binds to heterochromatin in somatic cells, and interacts with Heterochromatin protein-1 (HP1) in yeast two-hybrid and immunoprecipitation assays (Brower-Toland et al., 2007). piRNA-Piwi protein complexes could therefore silence target transposons by directing assembly of heterochromatin-like domains. In fission yeast, which do not have piRNAs, small interfering RNAs (siRNAs) and Argonaute 1 (Ago1) appear to recognize nascent transcripts at the centromere, triggering both transcript destruction and HP1 recruitment and assembly of centromeric heterochromatin (Buhler et al., 2006; Verdell and Moazed, 2005). A similar combination of homology dependent cleavage and heterochromatin assembly could drive piRNA based silencing in the *Drosophila* germline.

The mechanism of piRNAs biogenesis also remains to be fully elucidated. Dicer endonucleases cleave double-stranded precursors to produce miRNAs and siRNAs (reviewed in Ghildiyal and Zamore, 2009), but piRNA production is Dicer independent (Houwing et al., 2007; Vagin et al., 2006). A subset of sense and antisense piRNAs overlap by 10 base pairs and show a strong bias toward an A at position 10 of the sense strand and a complementary U at the 5' end of the antisense strand, suggesting that positions 1 and 10 base pair (Brennecke et al., 2007; Gunawardane et al., 2007). As Argonautes cleave their targets between positions 10 and 11 of the guide strand (Gunawardane et al., 2007; Saito et al., 2006), these findings suggest that piRNAs are produced by a "ping-pong" amplification cycle in which antisense strand piRNAs bound to Argonaute proteins cleave complementary RNAs to produce the 5' end of sense piRNAs, which in turn direct a reciprocal reaction that generates the 5' end of antisense strand piRNAs (Brennecke et al., 2007; Gunawardane et al., 2007). However, most piRNAs cannot be assigned to ping-pong pairs, some clusters produce piRNAs from only one strand (Brennecke et al., 2007), and the mechanism of 3' end generation has not been determined. It is also unclear how ping-pong amplification is initiated, since the cycle depends on pre-existing primary piRNAs.

Here, we show that Rhino (Rhi), a member of the Heterochromatin Protein 1 (HP1) subfamily of chromo box proteins (Volpe et al., 2001), is required for transposon silencing, production of piRNAs by dual-strand heterochromatic clusters, and efficient ping-pong amplification. Significantly, Rhi protein associates with the 42AB dual-strand cluster and is required for production of longer RNAs from both strands of this cluster. Rhi thus appears to promote expression of trigger RNAs that are processed to form primary piRNAs that drive ping-pong amplification and transposon silencing. We also show that protein coding genes carrying transposons and transposon fragments within introns escape silencing, suggesting that piRNA silencing is

imposed after RNA processing. Furthermore, *rhi* mutations disrupt nuage, a perinuclear structure that is enriched in piRNA pathway components. We therefore speculate that the nuage functions as a perinuclear surveillance machine that scans RNAs exiting the nucleus and destroys transcripts with piRNA complementarity.

RESULTS

Drosophila piRNA pathway mutations lead to germline DNA damage and disrupt axis specification through activation of Chk2 and ATR kinases, which function in DNA damage signaling (Chen et al., 2007; Cook et al., 2004; Pane et al., 2007). Mutations in the *rhi* locus lead to very similar patterning defects (Volpe et al., 2001). The *mei-41* and *mnk* genes encode ATR and Chk2, respectively (Brodsky et al., 2004; Hari et al., 1995). To determine whether the axis specification defects associated with *rhi* result from damage signaling, we generated double mutants with *mnk* and *mei-41* and quantified axis specification by scoring for assembly of dorsal appendages, which are egg shell structures that form in response to dorsal signaling during oocyte development (Table S1 available online). Only 17% (n = 700) of embryos from *rhi^{KG}/rhi²* females had two wild-type appendages. However, 80% (n = 689) of embryos from *mnk;rhi^{KG}/rhi²* double-mutant females had two appendages (Table S1). In addition, 33% (n = 732) of embryos from *mei-41;rhi^{KG}/rhi²* double-mutant females had two appendages (Table S1). Consistent with these observations, *rhi* mutations disrupt dorsal localization of Gurken and posterior localization of Vasa in the oocyte, and localization of both proteins is restored in *mnk;rhi^{KG}/rhi²* double mutants (Figure 1).

Both ATM and ATR kinases have been reported to activate Chk2 (Wang et al., 2006). Mutations in the *Drosophila atm* gene are lethal, but caffeine inhibits ATM and to a lesser extent ATR (Sarkaria et al., 1999). Strikingly, 88% (n = 473) of embryos from *rhi* mutant mothers fed caffeine had wild-type dorsal appendages (Table S1). Similarly, only 2% (n = 277) of embryos from *armi* mutant females had two dorsal appendages, compared with 11% (n = 477) after caffeine treatment (Table S1). In addition, 56% (n = 575) of embryos from *mei^{41D3}/mei^{41D3}; armi^{72.1}/armi¹* females had wild-type appendages, but 83% (n = 226) of embryos from *mei^{41D3}/mei^{41D3}; armi^{72.1}/armi¹* double mutants fed with caffeine had two appendages (Table S1). Caffeine combined with *mei-41* mutations thus leads to levels of suppression that are similar to *mnk* single mutations, suggesting that ATM and ATR redundantly activate Chk2 in *armi* and *rhi* mutants.

The *mei-W68* locus encodes the *Drosophila* Spo11 homolog, which is required for meiotic double-strand break formation (McKim and Hayashi-Hagihara, 1998). However, *mei-W68* mutations fail to suppress the dorsal appendage defects associated with *rhi* (Table S1), indicating that DNA damage signaling in *rhi* mutants is not due to defects in meiotic break repair.

The phosphorylated form of the *Drosophila* histone H2AX (γ -H2Av) accumulates near DNA double-strand break sites (Gong et al., 2005; Modesti and Kanaar, 2001; Redon et al., 2002). In wild-type ovaries, γ -H2Av foci are generally restricted to region 2 of the germlarium, where meiotic double-strand breaks are formed (Figures 1D–1F) (Jang et al., 2003). As the

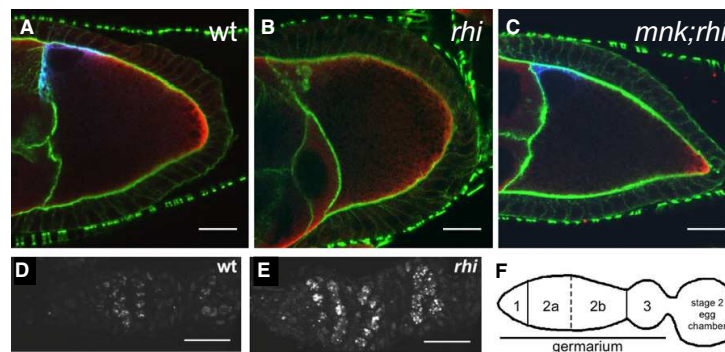


Figure 1. DNA Damage Signaling in *rhi* Mutants

(A–C) Mutations in *mnk*, which encodes the DNA damage signaling kinase Chk2, suppress the Gurken and Vasa protein localization defects in *rhi* mutants.

(A) In a stage 9 wild-type oocyte, Grk (blue) is localized at the dorsal anterior cortex near the oocyte nucleus and Vas (red) is localized at the posterior cortex. Actin filaments (green) mark the cell boundaries.

(B) In *rhi* egg chambers, this localization pattern is lost, with Grk and Vas dispersed throughout the oocyte.

(C) *mnk* suppresses the *rhi* phenotype, and rescues Grk and Vas localization during late oogenesis. Images were acquired under identical conditions. Projections of two serial 0.6 μm optical sections are shown. Scale bars represent 20 μm .

(D–F) *rhi* mutants have increased DNA damage in the germline.

(D) Foci of $\gamma\text{-H2Av}$ are observed in wild-type ovaries in region 2a and 2b of the germarium and correspond to the DSBs induced during meiotic recombination. (E) In *rhi* mutants, much larger foci also appear in region 2a of the germarium but persist in region 3 and the developing egg chambers. Samples were labeled and images were acquired under identical conditions. Projections of 5 serial 1 μm optical sections are shown. Posterior is oriented to the right. Scale bars represent 10 μm .

(F) A schematic representation of the regions of the germarium and a developing egg chamber.

cysts mature and pass through region 3 of the germarium, $\gamma\text{-H2Av}$ labeling is reduced. Stage 2 egg chambers, which bud from the germarium, show only low levels of $\gamma\text{-H2Av}$ labeling. In *rhi* mutants, prominent $\gamma\text{-H2Av}$ foci are present in germline cells of the germarium, and these foci persist and increase in intensity as cysts mature and bud to form stage 2 egg chambers (Figures 1D and 1E). *rhi* mutations thus appear to trigger germline-specific DNA breaks and damage signaling through ATM, ATR, and Chk2.

Transposon Silencing and Gene Expression

The piRNA pathway is required for transposon silencing in the *Drosophila* female germline (Vagin et al., 2006) but has also been implicated in heterochromatic gene silencing in somatic cells (Brower-Toland et al., 2007; Pal-Bhadra et al., 2002; Pal-Bhadra et al., 2004). We therefore assayed both transposon and protein-coding gene expression using whole-genome tiling arrays (Figure 2). In both *rhi* and *armi* mutants, most transposon families show a relatively modest 1.5- to 2-fold increase in expression, which is not statistically significant (false discovery rate [FDR] > 0.02). However, a subset of transposon families are dramatically overexpressed in both *rhi* and *armi* mutants (Figures 2B and 2C; blue points indicate FDR < 0.02). For example, *HeT-A* expression increased 70-fold in *rhino* and 117-fold in *armi* (Table S2). In total, 15 of 17 transposon families that are significantly overexpressed in *rhi* are also overexpressed in *armi* (Figure S1). 11 families are overexpressed with an FDR < 0.02 in *armi* mutants, but not in *rhi* (Figure S1). *Rhino* thus appears to silence a subset of the transposons silenced by *Armi*. This could reflect a role for *Armi* in transposon silencing in both somatic follicle cells and the germline (Klattenhoff et al., 2007), while *Rhi* appears to be restricted to the germline (see below).

Both *rhi* and *armi* mutations increased expression of long terminal repeat (LTR) elements, non-LTR retrotransposons,

and inverted repeat (IR) elements (Figure S2) (Vagin et al., 2006). Similar patterns of transposon overexpression are observed in *aub* and *ago3* mutants, which disrupt piRNA biogenesis (Li et al., 2009). Mutations in established piRNA pathway genes and in the *rhino* locus thus disrupt transposon silencing, independent of transposition mechanism.

piRNAs from the *suppressor of stellate* locus silence the *Stellate* gene during male germline development, and *Stellate* protein overexpression leads to *Stellate* crystal formation during spermatogenesis (Aravin et al., 2001; Bozzetti et al., 1995; Livak, 1984, 1990; Palumbo et al., 1994). However, *rhi* mutations do not lead to *Stellate* crystal formation or compromise male fertility (Figure S3 and data not shown).

HP1 and several genes in the piRNA pathway have been implicated in position effect variegation, which is linked to spreading of heterochromatin from centromeric and telomeric regions (Elgin and Grewal, 2003; Pal-Bhadra et al., 2004). However, neither *rhino* nor *armi* led to statistically significant changes in the expression of any protein coding genes, including the 613 annotated heterochromatic genes (Smith et al., 2007) (Figures 2D and 2E; green points indicate heterochromatic genes). piRNA pathway and *rhi* mutations thus do not produce changes in heterochromatin organization sufficient to alter protein coding gene expression during oogenesis.

Figure 2A shows a genome browser view of the region containing the heterochromatic gene *jing*. Expression of exons that comprise the mature *jing* transcript are essentially identical in *w¹¹¹⁸* and *rhi*, but expression of a *flea* transposon located in a major intron increases 7-fold (FDR < 0.02), and several transposons in the intergenic regions near *jing* are also overexpressed (Figure 2A, *rhino*, pink bars). The repeated nature of natural transposons and the design of the arrays makes it impossible to determine which specific transposon copy or copies are overexpressed, but we can conclude that at least one member of the

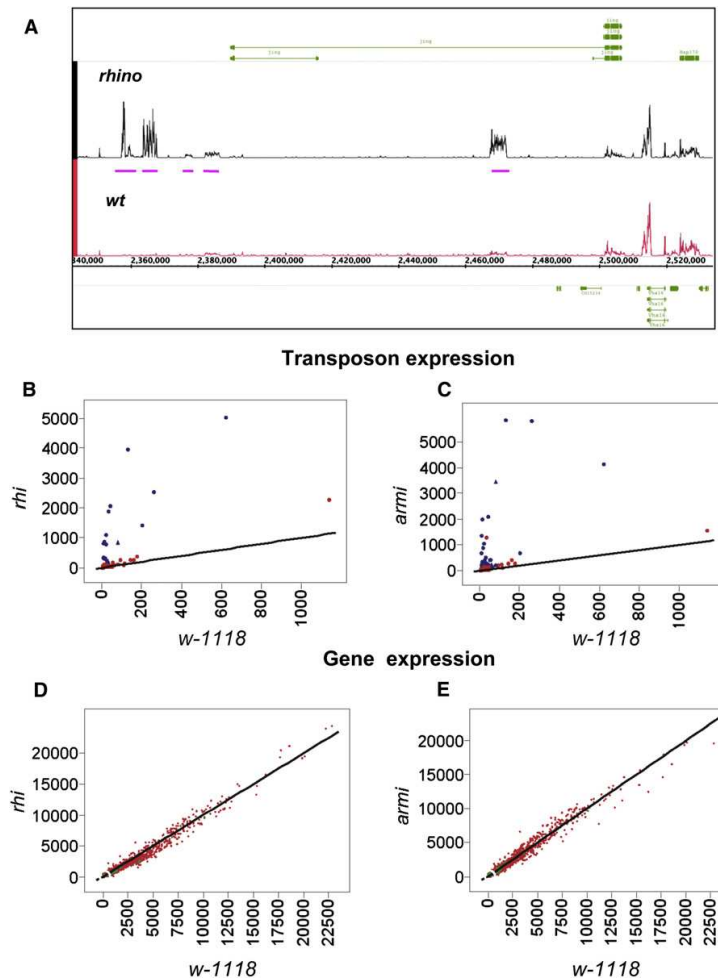


Figure 2. Gene and Transposon Expression in *rhi* Mutant Ovaries

(A) Genome browser view of tiling array data near *rhino*, a protein-coding gene in pericentromeric heterochromatin on chromosome 2R. Expression of *rhino* exons (green bars) is unaltered by *rhi* mutants. However, several intronic and extragenic transposons are significantly overexpressed (pink bars).

(B and C) Genome-wide analysis of transposon family expression in *rhi* and *armi* mutants. Tiling arrays were used to quantify expression of 95 transposon families in *rhi*, *armi*, and *w-1118* controls. Graphs show expression in *rhi* and *armi* plotted against expression in *w-1118*. The lines intercept the origin and have a slope of 1, and thus indicate equal expression in both genotypes. Significantly overexpressed transposon families are indicated by blue data points.

(D and E) Genome-wide comparison of protein coding gene expression in *rhi* and *armi* mutants, plotted against expression in *w-1118*. Heterochromatic genes are indicated by green data points, and euchromatic genes are indicated by red data points. Both classes cluster around the diagonal, indicating similar expression levels in mutant and controls.

transposon family is overexpressed. Over 1300 protein coding genes carry transposon insertions within introns, and thus have primary transcripts that could base pair with piRNAs. This includes *ago3*, which encodes an Arognate protein that is expressed in the female germline and is required for ping-pong amplification of piRNAs (Li et al., 2009). Our array studies show that expression of *ago3*, and the other protein coding genes carrying intronic transposon insertions, does not significantly change in *rhi* or *armi* mutants (Figures 2D and 2E and data not shown). These observations suggest that piRNA-dependent silencing may be imposed after splicing, which removed transposon homology from protein coding genes.

Rhi Localization

To define the subcellular distribution of Rhi, we generated a GFP-*rhi* transgene and raised anti-Rhi antibodies, which were used to

localize the protein in vivo and immunolabel whole-mount egg chambers. Both methods revealed germline-specific nuclear foci that are present throughout oogenesis (Figures 3A–3D). In addition, germline-specific expression of the GFP-Rhi fusion protein rescued fertility and axial patterning in *rhi* mutations (Table S1). Rhi thus appears to function specifically within the germline cells of the ovary.

To determine whether Rhi foci are associated with centromeres, we labeled for Rhi and CID, the *Drosophila* homolog of the centromere-specific, histone H3-like CENP-A (Blower and Karpen, 2001). Rhi accumulated in regions adjacent to most CID foci in germline nuclei, consistent with localization to pericentromeric heterochromatin (Figures 3E–3G). However, many Rhi foci were not obviously linked to CID. Some of these foci could be linked to telomeres or other chromatin domains. Resolving this question will require higher-resolution molecular approaches.

To determine whether Rhi localization depends on the piRNA pathway, we immunolabeled egg chambers mutant for *aub* and *armi*. Rhino localization to nuclear foci was not disrupted by either mutation (Figures 3A–3C). In striking contrast, *rhi* mutations disrupt localization of Aub and Ago3 to nuage, a perinuclear structure implicated in RNA processing (Figures 3H, 3I, 3L, and 3M). Vasa is a core component of nuage, and perinuclear localization of Vasa was also lost in *rhi* mutants (Figure S4). Piwi localizes to nuclei in both germline cells and the somatic follicle cells. In wild-type ovaries, Piwi is most abundant in

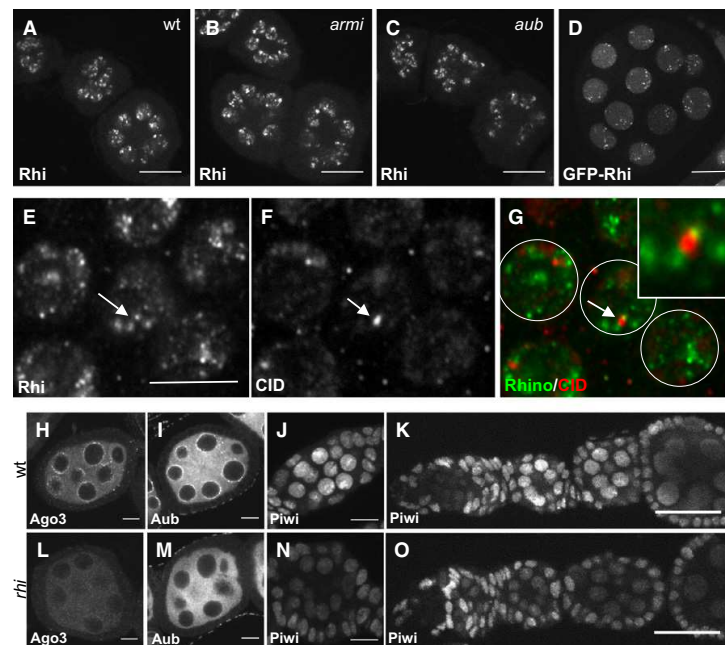


Figure 3. Rhi Localization Is Independent of the piRNA Pathway, but Localization of the PIWI Proteins Ago3 and Aub Requires Rhi

(A–C) Rhi localization appears similar in wild-type (A), *armi* (B), and *aub* (C) stage 2 to stage 4 egg chambers. Projections of five serial 1 μm optical sections are shown. Scale bars represent 20 μm . (D) GFP-Rhi transgene shows localization pattern similar to endogenous Rhi detected with anti-Rhi antiserum in the germline nuclei of stage 4–5 egg chambers. The scale bar represents 10 μm .

(E–G) Wild-type ovaries immunostained with (E and G) anti-Rhi antiserum and (F and G) anti-CID antibody show that some Rhi foci localize adjacent to CID foci (arrows) consistent with binding to peri-centromeric heterochromatin in some but not all chromosomes. The scale bar represents 5 μm .

(H–O) *rhi* mutation disrupts localization of PIWI class Argonautes. Stage 4–5 (H, I, L, and M), stage 2–3 (J and N), and germline to stage 3–4 egg chambers (K and O) of wild-type and *rhi* mutant ovaries were immunostained with corresponding antibodies. Projections of three serial 1 μm optical sections are shown. Scale bars represent 10 μm (H, I, J, L, M, and N) and 20 μm (J, K, and O). Wild-type localization of Ago3 and Aub proteins to perinuclear nuage is disrupted in *rhi* mutants. Piwi protein localizes to the nuclei of both germline and somatic cells in wild-type egg chambers. Only germline nuclear localization of Piwi in early stages is disrupted by mutations in *rhi*.

germline nuclei during early stages of oogenesis (Figure 3K). In *rhi* mutants, nuclear localization of Piwi is reduced during these early stages (Figures 3N and 3O). However, in later-stage egg chambers, which make up the bulk of the ovary, Piwi localization in *rhi* is similar to wild-type controls (Figures 3J, 3K, 3N, and 3O). These findings suggest *rhi* functions upstream of Ago3 and Aub, but may have a less critical role in Piwi-dependent processes.

piRNA Expression Is Ablated for Most Transposon Families in *rhino* Mutants

To determine whether Rhi is required for piRNA expression, we sequenced small RNAs from control and *rhi* mutant ovaries. Unlike miRNAs, piRNAs carry 2' methoxy, 3' hydroxy termini that render them resistant to oxidation and stabilize these RNAs in vivo (Vagin et al., 2006). To enrich for piRNAs and increase effective sequencing depth, we oxidized RNA samples prior to library construction and sequencing and normalized the data to surviving noncoding RNA fragments (Ghildiyal et al., 2008; Seitz et al., 2008) (see Table S3 for sequencing statistics). These studies indicate that *rhi* mutations reduce total piRNA abundance by approximately 80% (Figures 4A and 4B). Northern blotting for specific piRNAs and miRNAs support these findings (Figure S5). Defects in 3' modification destabilize piRNAs and would lead to preferential loss of piRNAs in oxidized samples. We therefore deep sequenced unoxidized RNAs and normalized piRNA abundance to miRNAs. These studies confirm that *rhi* mutations reduce piRNA abundance by 80%, and indicate that this reduction does not result from a defect in end modification (data not shown).

The majority of *Drosophila* piRNAs are derived from transposons and other repeated elements (Aravin et al., 2003; Brennecke et al., 2007). We analyzed the impact of *rhi* mutations on piRNA expression from 95 families with at least 500 matching reads in control samples (Table S3) (Li et al., 2009). *rhi* mutations lead to a 50% or greater reduction in antisense piRNA abundance for 83% of these transposon families, and a 98% reduction in antisense piRNAs for approximately 30% of these elements (Figure S7). For 66 of 95 families, both sense and antisense piRNAs are reduced. For example, *rhi* mutations nearly eliminate sense and antisense piRNAs from the telomeric transposon *HeT-A* (Figure 5A). Eight transposon families continue to express at least 50% of wild-type sense strand piRNAs but show an 80% or greater reduction in antisense piRNAs. The *jockey* element falls into this class. Mutations in *rhi* reduce sense strand piRNAs linked to *jockey* by only 10%, but antisense strand piRNAs are reduced by 95% (Figure 4B, *jockey*). For all of the transposon families that show reduced antisense piRNAs, including those that retain sense strand piRNAs, there is a clear reduction in opposite strand piRNAs that overlap by 10 nt, consistent with defects in ping-pong amplification (Figures 5Ac and 5Bc). A comparison of the p values for the 10 nt overlap bias across all transposon families confirms that the loss of ping-pong pairs in *rhi* is very highly significant (Figure S6, $p = 3e-10$). The loss of species that overlap by 10nt is also clear from an analysis of total piRNAs (Figures 4C and 4D). The *rhi* mutations thus lead to a near collapse of the ping-pong cycle amplification cycle.

Only 10 of 95 transposon families continue to express antisense piRNAs at or above 75% of wild-type levels in *rhi* mutants

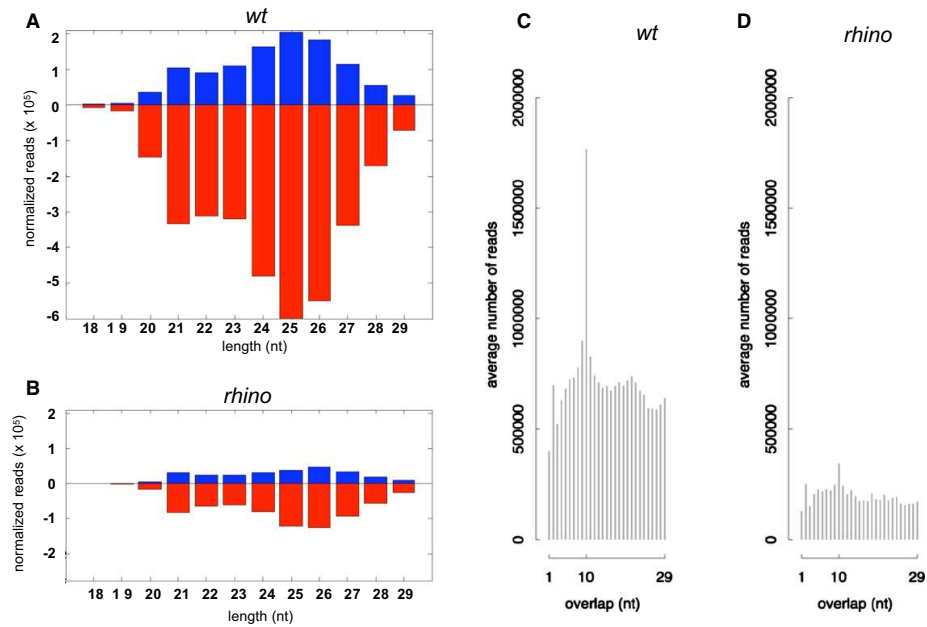


Figure 4. piRNA Production in *rho* Mutants

(A) Length histogram of piRNAs expressed in wild-type ovaries.

(B) Length histogram of piRNAs produced in *rho* mutants. Sense and antisense piRNAs are reduced by approximately 80%, and peak length shifts from 25 nt to 26 nt.

(C) Histogram of overlapping sense and antisense piRNA in wild-type ovaries, showing a pronounced peak at 10 nt, characteristic of ping-pong amplification.

(D) Histogram of overlapping sense and antisense piRNA in *rho* mutant ovaries. The 10 nt peak is nearly eliminated, suggesting a breakdown in the ping-pong amplification cycle.

(*blood*, *mdg-1*, *Tabor*, *Stalker*, *Stalker 2*, *Stalker3*, *Stalker4*, 412, 297, *gypsy 5*; Table S4). Eight of these families (*blood*, *mdg-1*, *Tabor*, *Stalker*, *Stalker 2*, *Stalker3*, *Stalker4*, 412) also show an increase in sense strand piRNAs (Figures 5Ca and S8). The sense strand piRNAs generally map to the same regions as peaks of antisense piRNAs (Figure 5Ca, *blood*; Figure S8). This pattern could indicate that antisense strand piRNA direct production of the sense strand piRNAs. Alternatively, specific regions within full-length elements or fragments of elements that lie within specific clusters may be preferentially utilized during piRNA production. The available data cannot distinguish between these alternatives.

An analysis of piRNAs encoded by the ten transposon families that show Rhi-independent piRNA production revealed three patterns with respect to overlapping sense and antisense species. The overlapping piRNAs encoded by *Stalker3* did not show a statistically significant ($p > 0.001$) 10 nt overlap bias in either wild-type or *rho* mutants, indicating that their production is independent of ping-pong amplification. However, six families showed a statistically significant 10 nt overlap peak in both wild-type and *rho* mutants, indicating that at least some of the piRNAs are produced by a ping-pong cycle that is independent of Rhi (*Tabor*, *Stalker*, *Stalker 2*, *Stalker4*, 412, 297; Figure S8). The final class of elements includes *blood*, *mdg1*, and *gypsy5*, which

show a statistically significant ping-pong peak in wild-type, but lose the 10 nt overlap bias in *rho* mutants (Figure 5Cc, *blood*; Figure S8). For this class, Rhi thus appears to promote production of only a subset of piRNAs through ping-pong amplification. Intriguingly, *rho* leads to a 10-fold increase in *blood* expression, suggesting the minor ping-pong pool of piRNAs may be critical to transposon silencing (Figure S9).

Overlapping ping-pong pairs show transposon family-specific nucleotide biases at positions 1 and 10 that appear to reflect the specific PIWI proteins that participate in the amplification cycle (Brennecke et al., 2007; Gunawardane et al., 2007). For example, elements for which sense strand piRNAs are primarily bound by Ago3 and antisense strand piRNAs are primarily bound by Aub, show an A bias at position 10 of the sense strand and a U bias at position 1 of the antisense strand (Brennecke et al., 2007; Gunawardane et al., 2007). Families that retain a statistically significant ping-pong peak generally retain the pattern of nucleotide bias observed in the wild-type (Figure S8), suggesting that *rho* reduces the efficiency of the ping-pong amplification but does not alter the specific PIWI proteins that participate in the cycle.

Antisense piRNAs can base pair with target RNAs and guide cleavage by PIWI proteins and are therefore presumed to be the effectors of transposon silencing. To determine whether loss of antisense piRNAs in *rho* mutants correlates with loss of

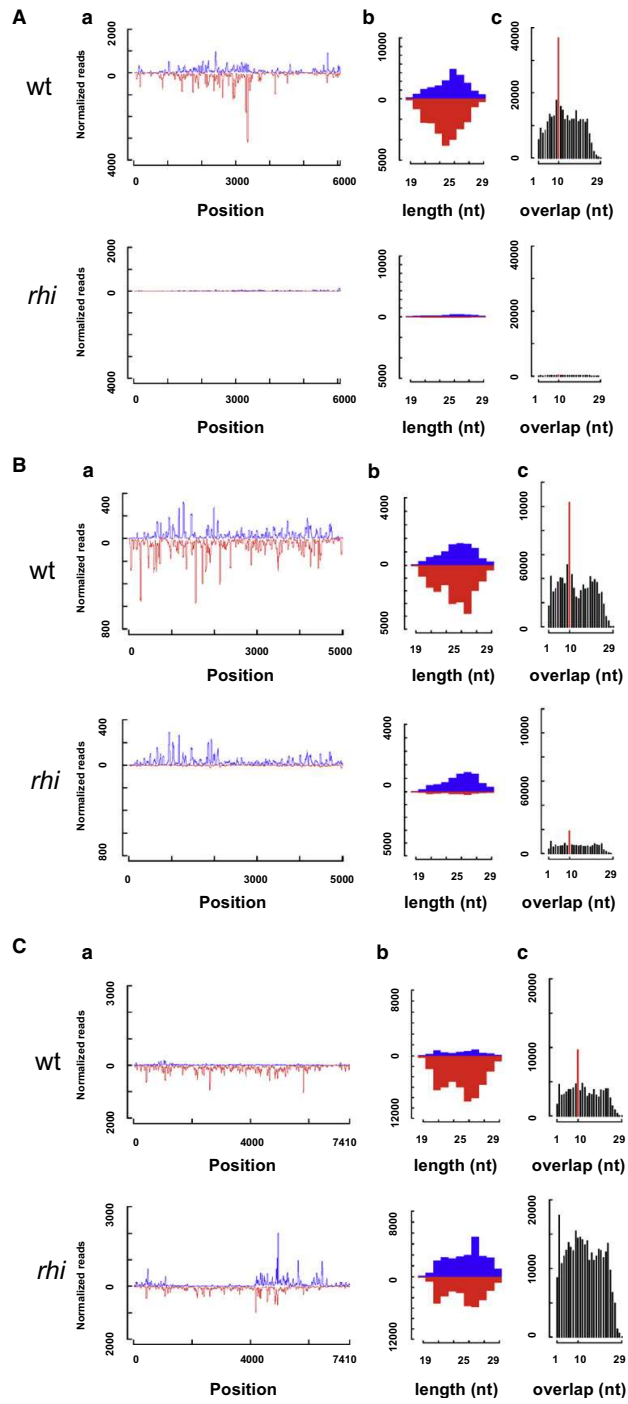


Figure 5. Transposon-Specific Changes in piRNA Abundance
 (A) Sense and antisense piRNA reads mapping to the consensus *Het-A* sequence (Aa), length histograms for all *HetA* piRNAs (Ab), and frequency distribution of overlapping *HetA* piRNA (Ac). (B) Sense and antisense piRNAs reads mapping to the consensus *jockey* sequence (Ba), length histograms for all *jockey* piRNAs (Bb), and frequency distribution of overlapping *jockey* piRNAs (Bc). (C) Sense and antisense piRNA reads mapping to the consensus *blood* sequence (Ca), length histograms for all *blood* piRNAs (Cb), and frequency distribution of overlapping *blood* piRNA (Cc). For the majority of transposons, including *HetA*, *rhi* mutations dramatically reduce sense and antisense piRNAs and nearly eliminate piRNA that overlap by 10 nt. For a subset of elements, represented by *jockey*, *rhi* leads to a loss of antisense piRNAs but no significant reduction in sense strand piRNAs. A very limited number of transposons, including *blood*, show no change or an increase in sense strand piRNAs in *rhi* mutants. Mutations in *rhi* reduce piRNAs with a 10 nt overlap, even for elements that show an increase in piRNAs from opposite strands (Cc, *blood*). In (Ac), (Bc) and (Cc), A statistically significant 10 nt bias is indicated by a red bar. This is characteristic of the ping-pong amplification cycle.

silencing, we plotted the fold change in transposon expression (*rhi*/wt) against the fold-change in antisense piRNAs (Figure S9). All of the transposon families that increased in expression by 20-fold or greater in *rhi* mutants also showed a 75% or greater reduction in antisense piRNA abundance. In addition, none of the families that retained antisense piRNA expression at 80% or higher levels were significantly overexpressed (FRD < 0.02; Figure S9). However, many transposon families that show a reduction in antisense piRNAs abundance of over 10-fold did not show a statistically significant increase in expression (Figure S9). These elements may be silenced by a piRNA independent mechanism. Alternatively, piRNAs linked to these elements could silence these elements, perhaps by inhibiting translation, without altering target transcript stability.

piRNA Clusters

The majority of piRNAs match transposons that are present in multiple copies in the genome and cannot be uniquely mapped. However, piRNAs encoded by polymorphic transposons, divergent transposon fragments, or other unique sequences can be mapped. Chromosome profiles of these “unique mappers” reveal dispersed piRNA peaks in the euchromatic chromosome arms and a limited number of prominent pericentromeric and subtelomeric clusters, which appear to be the source of the majority of piRNAs (Brennecke et al., 2007). *rhi* mutations essentially eliminate piRNAs mapping to pericentromeric heterochromatin on all of the autosomes (Figures S10 and 6). By contrast, piRNAs mapping to the pericentromeric region on the X are retained (Figures S10 and 6).

Most heterochromatic clusters produce piRNAs from both the plus and minus genomic strands (dual-strand clusters), but two major pericentromeric clusters on the X chromosome produce piRNA almost exclusively from one strand (uni-strand clusters) (Brennecke et al., 2007). We find that *rhi* reduces by 30- to 50-fold piRNAs from both strands of the top 11 dual-strand clusters (Figure S11, blue bars). For example, piRNA production from cluster 1/42AB, which is estimated to produce up to 30% of all piRNAs (Brennecke et al., 2007), is reduced by over 97% (Figures 6A–6C). In striking contrast, piRNAs encoded by the uni-strand clusters are only minimally impacted by *rhi* (Figure S11, red bars). As shown in Figures 6B and 6D, piRNAs from cluster 2 are derived almost exclusively from one strand, and production of these piRNAs is nearly unchanged in *rhi* mutants (Figures 6B and 6D). This does not appear to reflect expression of cluster 2 piRNAs exclusively in the somatic follicle cells, since Ago3 is germline specific and *ago3* mutations reduce piRNAs linked to this locus by close to 20-fold (Figure S12). In addition, unique piRNAs mapping to this cluster immunoprecipitate with the germline specific PIWI proteins Aub and Ago3 (Figure S12). Both dual-strand and uni-strand clusters thus appear to be expressed in the germline, but *rhi* mutations only disrupt piRNA production by the dual-strand clusters. Consistent with these findings, nine of the ten transposon families that continue to express high levels of antisense piRNAs in *rhi* (75% of wild-type or greater) have insertions in one or both of the major uni-strand clusters (Table S5).

To determine whether Rhino protein associates with clusters, we performed chromatin immunoprecipitation (ChIP) using anti-

GFP antibodies and flies expressing a functional Rhino-GFP transgene specifically in the germline. To control for nonspecific binding, precipitation was performed using nonimmune IgG on chromatin from ovaries expressing the GFP-rhino transgene. Additionally, anti-GFP antibodies were used on chromatin isolated from wild-type flies that do not express the GFP fusion. Quantitative PCR (qPCR) assays showed only background signal in both of these control reactions (data not shown). The anti-GFP fractions were assayed for three regions of cluster 1/42AB, two regions of cluster 2, two regions in the heterochromatic protein coding genes *jing* and *pld* that flank the 42A/B cluster, and the euchromatic protein coding genes *rp49* and *ry* (Figure 6G). GFP-Rhi binding, measured as a fraction of input chromatin, was enriched at all three sites in the dual-strand cluster relative to the euchromatic protein coding genes (Figure 6F). By contrast, the two sites in uni-strand cluster 2 showed no enrichment relative to *rp49* or *ry* controls (Figure 6F). Regions in the two heterochromatic genes immediately flanking the 42A/B cluster showed binding that was 3- to 4-fold lower than the peak region in the cluster (1A) and approximately 3-fold higher than binding to euchromatic genes. Rhino thus appears to be enriched at dual-strand heterochromatic clusters and may spread somewhat beyond the computationally defined limits of these clusters.

Cluster Transcription

The piRNA clusters are proposed to produce long precursor RNAs that are processed to form primary piRNAs, which in turn trigger the ping-pong amplification cycle by targeting sense strand transposon transcripts. To determine whether Rhino is required for RNA production by clusters, we used quantitative reverse transcriptase (RT)-PCR to assay RNAs derived from both strands of cluster 1/42AB, cluster 2, and *flam*. Reactions without RT produced no significant signal, and the low level of signal obtained in the absence of the strand specific RT primers was subtracted from the signal obtained with the strand specific primers. Consistent with production of piRNAs from both genomic strands, we detected longer RNAs from both strands at two independent locations in cluster 1/42AB (Figure 6E, blue bars). Significantly, RNAs from both strands were nearly eliminated in *rhi* homozygous mutants (Figure 6E, red bars). At cluster 1 and *flam*, which produce piRNA almost exclusively from the plus strand (Figure 6D) (Brennecke et al., 2007), RT-PCR detected RNA from only the plus strands (Figure 6F, blue bars). In striking contrast to cluster 1/42AB, *rhi* mutations led to a slight increase in plus strand transcript from cluster 2 and only a modest decrease in plus strand RNA from *flam*. These observations suggest that Rhino promotes production of precursor RNAs from dual-strand cluster 1/42AB, and possibly all dual-strand clusters.

Brennecke et al. (2007) proposed that antisense piRNAs derived from the clusters initiate ping-pong amplification by cleaving sense strand transcripts from target transposons. However, unique piRNAs derived from opposite strands of cluster 1/42AB show a strong 10 nt overlap bias (Figure S13), indicating that they are produced by ping-pong processing of precursor RNAs derived from the cluster. Antisense piRNAs derived through cluster based ping-pong amplification thus

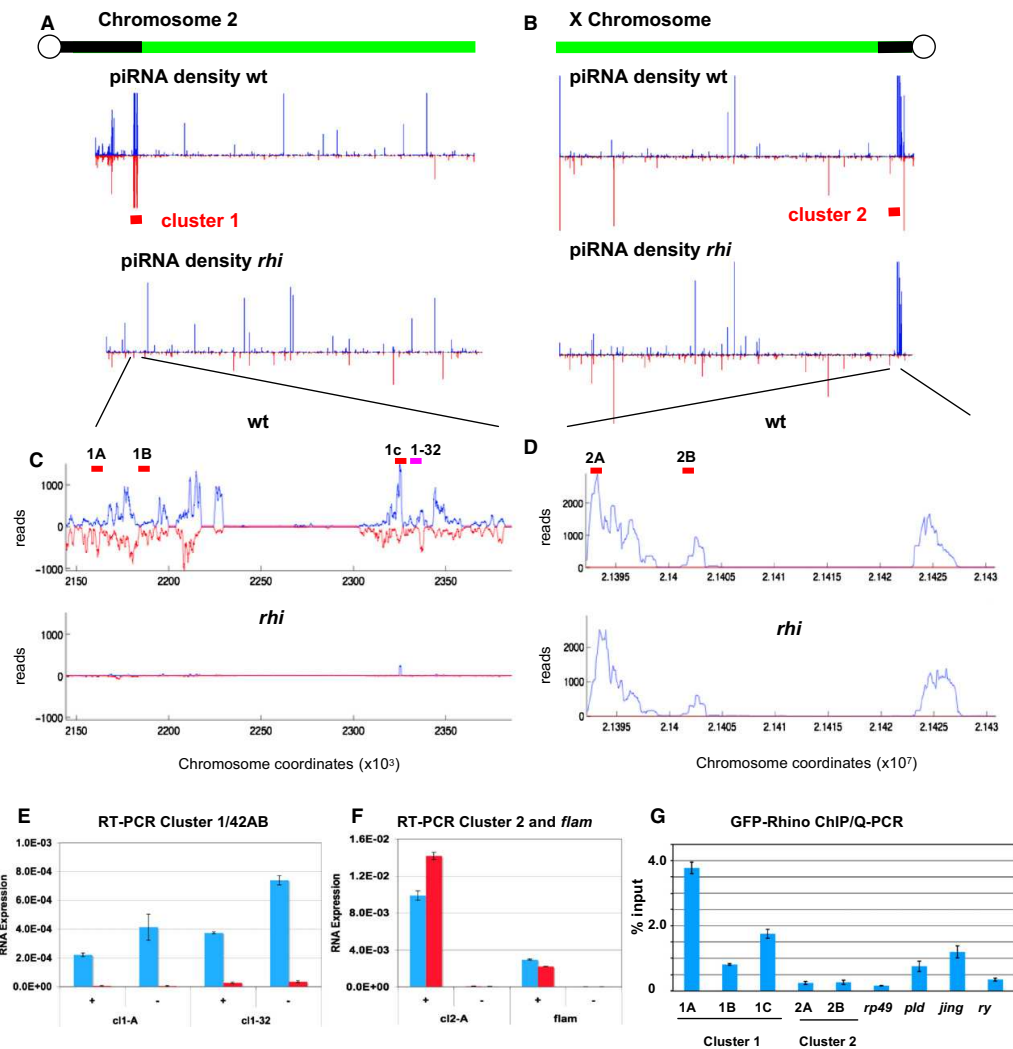


Figure 6. Rhino Is Required for piRNA Production by Dual-Strand Heterochromatic Clusters

(A and B) Chromosome 2R and X density profiles of uniquely mapping plus (blue) and minus (red) strand piRNAs in wild-type and *rho* mutants. piRNAs map to dispersed loci on the chromosome arms and prominent heterochromatic clusters. Pericentromeric piRNAs from chromosome 2R (A), and all other autosomes (Figure S10) are dramatically reduced in *rho* mutants. Pericentromeric piRNAs on the X chromosome show relatively little change.

(C and D) Higher-resolution maps of clusters 1 and 2, which map to the indicated regions on 2R and X, respectively. Mutations in *rho* nearly eliminated piRNAs encoded by cluster 1, which is the major dual-strand cluster, but have little impact on piRNAs from cluster 2, which is the major uni-strand cluster.

(E and F) Quantitative strand-specific RT-PCR for RNA derived from dual-strand cluster 1 (E) and uni-strand clusters 2 and *flam* (F). In wild-type ovaries, RNA is detected from both the plus (+) and minus (-) strands of cluster 1, at two independent locations (F, cl1-A and cl1-32, blue bars). RNAs from both strands of cluster 1 are dramatically reduced in *rho* mutants (red bars). Significant levels of RNA are only detected from the plus strand of cluster 2 and *flam* (F, blue bars), and *rho* does not block expression of these RNAs (F, red bars).

(G) Chromatin immunoprecipitation/quantitative PCR analysis of Rhino binding to cluster 1/42AB, the euchromatic genes *rp49* and *ry*, and the heterochromatic genes flanking cluster 1 show intermediate levels of binding, suggesting that Rhino may spread to regions flanking the dual-strand clusters. Anti-GFP antibodies were used to precipitate Rhino-GFP from crosslinked ovary chromatin fractions.

The approximate positions of the qPCR primer pairs used in ChIP and RT-PCR reactions are indicated by the red bars in (C) and (D). The pink bar in (C) indicates the approximate position of an additional primer pair used in RT-PCR.

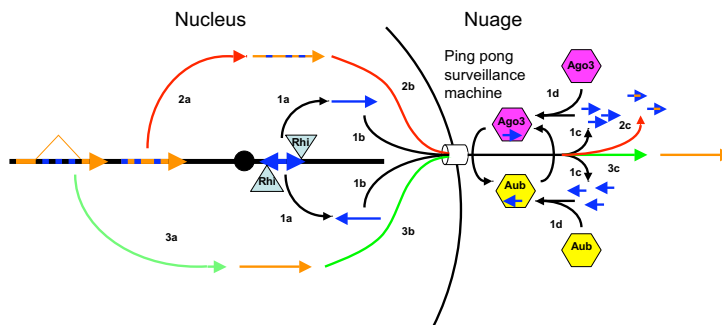


Figure 7. Model for Rhino-Dependent Transposon Silencing

Step 1 (black arrows): Rhino binds to dual-strand clusters and promotes production of RNAs from both genomic strands (1a), which are exported from the nucleus (1b) and processed into piRNAs by a ping-pong cycle driven by Ago3 (pink hexagon) and Aub (yellow hexagon), localized to the perinuclear nuage (1c).

Step 2 (red arrows): Transposons carrying piRNA homology are transcribed (2a), exported from the nucleus (2b), and degraded as they encounter the "ping-pong surveillance machine" within the nuage (2c).

Step 3 (green arrows): Protein-coding genes with intronic transposon insertions are spliced, which removes piRNA homology (3a). These transcripts are exported from the nucleus (3b) escape recognition by the surveillance system, and are translated (3c). Sequences matching piRNAs are indicated by blue. Other transcribed regions are in orange.

appear to target sense strand RNAs derived from functional transposons located throughout the genome.

DISCUSSION

piRNAs encoded by transposon-rich heterochromatic clusters have been proposed to initiate a ping-pong cycle that amplifies the piRNA pool and mediates transposon silencing (Brennecke et al., 2007; Lin, 2007; O'Donnell and Boeke, 2007). However, the mechanisms of piRNA biogenesis and silencing are not well understood, and it is unclear how the piRNA clusters are differentiated from other chromatin domains. We show that the HP1 homolog Rhino is required for production of piRNAs from dual-strand clusters and associates with the major 42AB cluster by ChIP. Significantly, we also identify putative piRNA precursor RNAs from both strands of the 42AB cluster and show that Rhino is required for production of these RNAs. These findings lead us to propose that Rhi binding promotes transcription of dual-strand clusters, and that the resulting RNAs are processed to form primary piRNAs that drive the ping-pong amplification cycle and transposon silencing (Figure 7, black pathway).

While Rhino protein appears to be restricted to germline nuclei, *rhi* mutations disrupt perinuclear localization of Ago3 and Aub (Figure 3), which catalyze the ping-pong amplification cycle (Li et al., 2009). Mutations in *krimper*, which encodes a component of the perinuclear nuage, also disrupt transposon silencing and piRNA production (Lim and Kai, 2007). piRNA silencing and nuage assembly thus appear to be codependent processes. These observations, with our finding that protein coding genes carrying piRNA homology within introns escape silencing by the piRNA pathway (Figure 2A), suggest that transcripts are scanned for piRNA homology within the nuage, after splicing and nuclear export. Mature protein coding mRNAs thus pass through the nuage and are translated because piRNA homology has been removed by splicing. By contrast, mature transposon transcripts carry piRNA complementarity are recognized by the perinuclear ping-pong machine, leading to destruction. Interestingly, mutations in the mouse *maelstrom* gene disrupt nuage and lead to

male sterility and significant overexpression of LINE-1 elements (Soper et al., 2008). Nuage may therefore have a conserved function in transposon RNA surveillance and silencing.

In *S. pombe*, siRNAs bound to Ago1 appear to recruit HP1 to centromeres through interactions with nascent transcripts, thus triggering heterochromatin assembly and transcriptional silencing. Our data indicate that the HP1 homolog Rhino is required for transposon silencing, but this process appears to be mechanistically distinct from centromeric heterochromatin silencing in yeast. For example, localization of the Rhino HP1 homolog to nuclear foci is independent of piRNA production, and Rhino binding appears to promote transcription of heterochromatic clusters. This in turn generates piRNAs that may direct silencing through posttranscriptional target cleavage. However, piRNAs bound to PIWI proteins have been implicated in heterochromatin assembly in somatic cells, and this process could be related evolutionarily to heterochromatin assembly in fission yeast.

Intriguingly, *rhi* is a rapidly evolving gene, and all three Rhi protein domains (chromo, chromo shadow, and hinge) show evidence of strong positive selection (Vermaak et al., 2005). On the basis of these observations, Vermaak et al. (2005) proposed that *rhino* is involved in a genetic conflict within the germline. The observations reported here suggest that the conflict between transposon propagation and maintenance of germline DNA integrity drives *rhi* evolution, and that the heterochromatic dual-strand clusters have a key role in this battle. Rhino appears to define heterochromatic domains that produce transposon silencing piRNAs. Rhino could therefore have evolved to bind transposon integration proteins, which would promote transposition into clusters and production of *trans*-silencing piRNAs. In this model, the transposon integration machinery would evolve to escape Rhino binding and silencing. The rapid pace of *rhino* evolution makes identification of homologs in other species difficult (Vermaak et al., 2005), but the conserved role for piRNAs in germline development suggests that HP1 variants may have critical roles in the conflict between selfish elements and genome integrity in other species, including humans.

EXPERIMENTAL PROCEDURES

Drosophila Stocks

All animals were raised at 25°C. Oregon R, *w¹¹¹⁸* and *cn¹*; *ry⁵⁰⁶* were used as controls, as noted. The following alleles were used: *mnk^{P6}* (Brodsky et al., 2004; Takada et al., 2003); *rhl^{KG00910}* (*rhl^{KG}*) and *rhl^{D2086}* (*rhl^D*) (Volpe et al., 2001); *armi^{72.1}* and *armi¹* (Cook et al., 2004); *mei41^{D3}* (Hari et al., 1995; Hawley and Tartof, 1983); and P[*lacW*]/*mei-W68^{K05803}*, *mei-W68¹* (McKim and Hayashi-Hagihara, 1998). The *mnk^{P6}* allele was kindly provided by M. Brodsky (Brodsky et al., 2004). All other stocks were obtained from the Bloomington Stock Center, Bloomington, Indiana (<http://flystocks.bio.indiana.edu/>). Standard genetic procedures were used to generate double mutant combinations.

Immunohistochemistry

Antibody production is described in the Supplemental Experimental Procedures. Egg chamber fixation and whole-mount antibody labeling were performed as previously described (Theurkauf, 1994). Vas protein was labeled with rabbit polyclonal anti-Vas antibody (Liang et al., 1994) at 1:1000. Gurken protein was labeled with mouse monoclonal anti-Gurken antibody (obtained from the Developmental Studies Hybridoma Bank, University of Iowa) at 1:10. Rhl protein was labeled with a guinea pig polyclonal anti-Rhl antiserum developed by our group (see above) at 1:2000. Piwi, Aub, and Ago3 were labeled with rabbit polyclonal anti-Piwi, anti-Aub and anti-Ago3 antibodies developed for this study (see above) at 1:1000. Antibody against γ -H2Av was kindly provided by K. McKim (Gong et al., 2005) and egg chambers were labeled as described previously (Belmont et al., 1989). CID was labeled with an affinity-purified chicken anti-CID antibody provided by G. Karpen at 1:100 (Blower and Karpen, 2001). HOAP was labeled with a polyclonal rabbit anti-Hoap antibody generated by our group (see above) at 1:1000. Rhodamine-conjugated phalloidin (Molecular Probes) was used at 1:100 to stain F-Actin, and TOTO3 (Molecular Probes) was used at 1:500 (0.2 mM final concentration) to visualize DNA.

Labeled tissue was mounted and analyzed with a Leica TCS-SP inverted laser-scanning microscope as described previously (Cha et al., 2001).

GFP-Rhino Transgene

The GFP-Rhino transgene was generated by recombining the Rhl-DONR (see above) construct with a modified pCasper vector containing the GFP sequence and Gateway cloning cassette B (Invitrogen). The resulting vector contained GFP fused in frame to the N terminus of Rhino under the control of the Gal4 promoter. Transgenic animals were generated using standard embryo microinjection techniques at Genetic Services.

RNA Isolation and Tiling Array Hybridization

Total RNA from was isolated from manually dissected ovaries from 2- to 4-day-old flies with RNeasy (QIAGEN) according to the manufacturer's instructions. The RNA was quantified by absorbance at 260 nm. Three independent RNA isolates from each genotype was then assayed as follows: Double-stranded cDNA was prepared with the GeneChip WT Amplified Double-Stranded cDNA Synthesis Kit (Affymetrix). DNA was labeled with the GeneChip WT Double-Stranded DNA Terminal Labeling Kit (Affymetrix). Labeled DNA was hybridized to GeneChip *Drosophila* Tiling 2.0R Arrays (Affymetrix) with the GeneChip Hybridization, Wash, and Stain Kit (Affymetrix) at the University of Massachusetts Medical School genomic core facility.

To determine whether genetic background or DNA damage significantly alters gene or transposon expression, we assayed ovarian RNA isolated from two common laboratory strains (*w-1118* and *cn,bw*), the meiotic repair mutant *okra*, the DNA damage signaling mutant *mnk*, and *mnk;okra* double mutants. Pairwise comparisons show little difference in genome wide patterns of gene or transposon expression in any of these five strains (Figure S12). The background for the *rhl* heteroallelic combination used here is *cn/+*; *ry/+*, which is genetically wild-type. Since the *armi* allelic combination used here is in a homozygous *w-1118* background, this genotype was used a control in our array studies.

The tiling array data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus (Edgar et al., 2002) and are accessible

through GEO Series accession number GSE14370 (<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE14370>).

Small RNA isolation, oxidation, and sequencing were performed as described elsewhere (Li et al., 2009). Bioinformatics methods, chromatin immunoprecipitation and strand specific RT-PCR procedures are described in the Supplemental Experimental Procedures.

SUPPLEMENTAL DATA

Supplemental Data include Supplemental Experimental Procedures, 14 figures, and six tables and can be found with this article online at [http://www.cell.com/supplemental/S0092-8674\(09\)00853-8](http://www.cell.com/supplemental/S0092-8674(09)00853-8).

ACKNOWLEDGMENTS

We thank Celeste Berg for *rhl* mutant stocks, Maria Zapp and Ellie Kittler in the UMass Deep Sequencing Core and Phyllis Spartick in the UMass Genomics core for expert assistance with small RNA sequencing and tiling array analyses, and members of the Zamore, Weng, and Theurkauf labs for critical discussions during the course of these studies. The Genomics Core was supported by the Diabetes and Endocrinology research Center grant DK032502. This work was supported in part by grants from the National Institutes of Health to W.E.T. (HD049116) and P.D.Z. (GM62862 and GM65236). H.X. was supported in part by Pfizer.

Received: February 20, 2009

Revised: May 7, 2009

Accepted: July 2, 2009

Published online: September 3, 2009

REFERENCES

- Aravin, A.A., Naumova, N.M., Tulin, A.V., Vagin, V.V., Rozovsky, Y.M., and Gvozdev, V.A. (2001). Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the *D. melanogaster* germline. *Curr. Biol.* 11, 1017–1027.
- Aravin, A.A., Lagos-Quintana, M., Yalcin, A., Zavolan, M., Marks, D., Snyder, B., Gaasterland, T., Meyer, J., and Tuschl, T. (2003). The small RNA profile during *Drosophila melanogaster* development. *Dev. Cell* 5, 337–350.
- Belmont, A.S., Braunfeld, M.B., Sedat, J.W., and Agard, D.A. (1989). Large-scale chromatin structural domains within mitotic and interphase chromosomes in vivo and in vitro. *Chromosoma* 98, 129–143.
- Blower, M.D., and Karpen, G.H. (2001). The role of *Drosophila* CID in kinetochore formation, cell-cycle progression and heterochromatin interactions. *Nat. Cell Biol.* 3, 730–739.
- Bozzetti, M.P., Massari, S., Finelli, P., Meggio, F., Pinna, L.A., Boldyreff, B., Issinger, O.G., Palumbo, G., Ciriaco, C., Bonaccorsi, S., et al. (1995). The Ste locus, a component of the parasitic cry-Ste system of *Drosophila melanogaster*, encodes a protein that forms crystals in primary spermatocytes and mimics properties of the beta subunit of casein kinase 2. *Proc. Natl. Acad. Sci. USA* 92, 6067–6071.
- Brennecke, J., Aravin, A.A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., and Hannon, G.J. (2007). Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell* 128, 1089–1103.
- Brennecke, J., Malone, C., Aravin, A., Sachidanandam, R., Stark, A., and Hannon, G. (2008). An epigenetic role for maternally inherited piRNAs in transposon silencing. *Science* 322, 1387–1392.
- Brodsky, M.H., Weinert, B.T., Tsang, G., Rong, Y.S., McGinnis, N.M., Golic, K.G., Rio, D.C., and Rubin, G.M. (2004). *Drosophila melanogaster* MNK/Chk2 and p53 regulate multiple DNA repair and apoptotic pathways following DNA damage. *Mol. Cell Biol.* 24, 1219–1231.
- Brower-Toland, B., Findley, S.D., Jiang, L., Liu, L., Yin, H., Dus, M., Zhou, P., Elgin, S.C., and Lin, H. (2007). *Drosophila* PIWI associates with chromatin and interacts directly with HP1a. *Genes Dev.* 21, 2300–2311.

- Buhler, M., Verdel, A., and Moazed, D. (2006). Tethering RITS to a nascent transcript initiates RNAi- and heterochromatin-dependent gene silencing. *Cell* 125, 873–886.
- Carmell, M.A., Girard, A., van de Kant, H.J., Bourc'his, D., Bestor, T.H., de Rooij, D.G., and Hannon, G.J. (2007). MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. *Dev. Cell* 12, 503–514.
- Cha, B.-J., Koppetsch, B., and Theurkauf, W.E. (2001). *In vivo* analysis of *bicoid* mRNA localization reveals a novel microtubule-dependent anterior axis specification pathway. *Cell* 106, 35–46.
- Chambeyron, S., Popkova, A., Payen-Groschene, G., Brun, C., Laouini, D., Pelisson, A., and Bucheton, A. (2008). piRNA-mediated nuclear accumulation of retrotransposon transcripts in the *Drosophila* female germline. *Proc. Natl. Acad. Sci. USA* 105, 14964–14969.
- Chen, Y., Pane, A., and Schubach, T. (2007). Cutoff and aubergine mutations result in retrotransposon upregulation and checkpoint activation in *Drosophila*. *Curr. Biol.* 17, 637–642.
- Cook, H.A., Koppetsch, B.S., Wu, J., and Theurkauf, W.E. (2004). The *Drosophila* SDE3 homolog armitage is required for oskar mRNA silencing and embryonic axis specification. *Cell* 116, 817–829.
- Desset, S., Buchon, N., Meignin, C., Coiffet, M., and Vaury, C. (2008). In *Drosophila melanogaster* the COM locus directs the somatic silencing of two retrotransposons through both Piwi-dependent and -independent pathways. *PLoS ONE* 3, e1526.
- Edgar, R., Domrachev, M., and Lash, A.E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* 30, 207–210.
- Elgin, S.C., and Grewal, S.I. (2003). Heterochromatin: silence is golden. *Curr. Biol.* 13, R895–R898.
- Ghildiyal, M., and Zamore, P. (2009). Small silencing RNAs: an expanding universe. *Nat. Rev. Genet.* 10, 94–108.
- Ghildiyal, M., Seitz, H., Horwich, M.D., Li, C., Du, T., Lee, S., Xu, J., Kittler, E.L., Zapp, M.L., Weng, Z., and Zamore, P.D. (2008). Endogenous siRNAs derived from transposons and mRNAs in *Drosophila* somatic cells. *Science* 320, 1077–1081.
- Girton, J.R., and Johansen, K.M. (2008). Chromatin structure and the regulation of gene expression: the lessons of PEV in *Drosophila*. *Adv. Genet.* 61, 1–43.
- Gong, W.J., McKim, K.S., and Hawley, R.S. (2005). All paired up with no place to go: pairing, synapsis, and DSB formation in a balancer heterozygote. *PLoS Genet.* 1, e67.
- Gunawardane, L.S., Saito, K., Nishida, K.M., Miyoshi, K., Kawamura, Y., Nagami, T., Siomi, H., and Siomi, M.C. (2007). A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in *Drosophila*. *Science* 315, 1587–1590.
- Hari, K.L., Santerre, A., Sekelsky, J.J., McKim, K.S., Boyd, J.B., and Hawley, R.S. (1995). The mei-41 gene of *D. melanogaster* is a structural and functional homolog of the human ataxia telangiectasia gene. *Cell* 82, 815–821.
- Hawley, R.S., and Tartof, K.D. (1983). The effect of mei-41 on rDNA redundancy in *Drosophila melanogaster*. *Genetics* 104, 63–80.
- Houwing, S., Kamminga, L.M., Berezikov, E., Cronembold, D., Girard, A., van den Elst, H., Filippov, D.V., Blaser, H., Raz, E., Moens, C.B., et al. (2007). A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in *Zebrafish*. *Cell* 129, 69–82.
- Jang, J.K., Sherizen, D.E., Bhagat, R., Manheim, E.A., and McKim, K.S. (2003). Relationship of DNA double-strand breaks to synapsis in *Drosophila*. *J. Cell Sci.* 116, 3069–3077.
- Klattenhoff, C., Bratu, D.P., McGinnis-Schultz, N., Koppetsch, B.S., Cook, H.A., and Theurkauf, W.E. (2007). *Drosophila* rasiRNA pathway mutations disrupt embryonic axis specification through activation of an ATR/Chk2 DNA damage response. *Dev. Cell* 12, 45–55.
- Li, C., Vagin, V.V., Lee, S., Xu, J., Ma, S., Xi, H., Seitz, H., Horwich, M.D., Syrzycka, M., Honda, B.M., et al. (2009). Collapse of germline piRNAs in the absence of Argonaute3 reveals somatic piRNAs in flies. *Cell* 137, 509–521.
- Liang, L., Diehl-Jones, W., and Lasko, P. (1994). Localization of vasa protein to the *Drosophila* pole plasm is independent of its RNA-binding and helicase activities. *Development* 120, 1201–1211.
- Lim, A.K., and Kai, T. (2007). Unique germ-line organelle, nuage, functions to repress selfish genetic elements in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* 104, 6714–6719.
- Lin, H. (2007). piRNAs in the germ line. *Science* 316, 397.
- Livak, K.J. (1984). Organization and mapping of a sequence on the *Drosophila melanogaster* X and Y chromosomes that is transcribed during spermatogenesis. *Genetics* 107, 611–634.
- Livak, K.J. (1990). Detailed structure of the *Drosophila melanogaster* stellate genes and their transcripts. *Genetics* 124, 303–316.
- McKim, K.S., and Hayashi-Hagihara, A. (1998). mei-W68 in *Drosophila melanogaster* encodes a Spo11 homolog: evidence that the mechanism for initiating meiotic recombination is conserved. *Genes Dev.* 12, 2932–2942.
- Mevel-Ninio, M., Pelisson, A., Kinder, J., Campos, A.R., and Bucheton, A. (2007). The flamenco locus controls the gypsy and ZAM retroviruses and is required for *Drosophila* oogenesis. *Genetics* 175, 1615–1624.
- Modesti, M., and Kanaar, R. (2001). DNA repair: spot(light)s on chromatin. *Curr. Biol.* 11, R229–R232.
- O'Donnell, K.A., and Boeke, J.D. (2007). Mighty Piwis defend the germline against genome intruders. *Cell* 129, 37–44.
- Pal-Bhadra, M., Bhadra, U., and Birchler, J.A. (2002). RNAi related mechanisms affect both transcriptional and posttranscriptional transgene silencing in *Drosophila*. *Mol. Cell* 9, 315–327.
- Pal-Bhadra, M., Leibovitch, B.A., Gandhi, S.G., Rao, M., Bhadra, U., Birchler, J.A., and Elgin, S.C. (2004). Heterochromatic silencing and HP1 localization in *Drosophila* are dependent on the RNAi machinery. *Science* 303, 669–672.
- Palumbo, G., Bonaccorsi, S., Robbins, L.G., and Pimpinelli, S. (1994). Genetic analysis of Stellate elements of *Drosophila melanogaster*. *Genetics* 138, 1181–1197.
- Pane, A., Wehr, K., and Schubach, T. (2007). zucchini and squash encode two putative nucleases required for rasiRNA production in the *Drosophila* germline. *Dev. Cell* 12, 851–862.
- Pelisson, A., Song, S.U., Prud'homme, N., Smith, P.A., Bucheton, A., and Corces, V.G. (1994). Gypsy transposition correlates with the production of a retroviral envelope-like protein under the tissue-specific control of the *Drosophila* flamenco gene. *EMBO J.* 13, 4401–4411.
- Pelisson, A., Payen-Groschene, G., Terzian, C., and Bucheton, A. (2007). Restrictive flamenco alleles are maintained in *Drosophila melanogaster* population cages, despite the absence of their endogenous gypsy retroviral targets. *Mol. Biol. Evol.* 24, 498–504.
- Prud'homme, N., Gans, M., Masson, M., Terzian, C., and Bucheton, A. (1995). Flamenco, a gene controlling the gypsy retrovirus of *Drosophila melanogaster*. *Genetics* 139, 697–711.
- Redon, C., Pilch, D., Rogakou, E., Sedelnikova, O., Newrock, K., and Bonner, W. (2002). Histone H2A variants H2AX and H2AZ. *Curr. Opin. Genet. Dev.* 12, 162–169.
- Saito, K., Nishida, K.M., Mori, T., Kawamura, Y., Miyoshi, K., Nagami, T., Siomi, H., and Siomi, M.C. (2006). Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the *Drosophila* genome. *Genes Dev.* 20, 2214–2222.
- Sarkaria, J.N., Busby, E.C., Tibbetts, R.S., Roos, P., Taya, Y., Karnitz, L.M., and Abraham, R.T. (1999). Inhibition of ATM and ATR kinase activities by the radiosensitizing agent, caffeine. *Cancer Res.* 59, 4375–4382.
- Sarot, E., Payen-Groschene, G., Bucheton, A., and Pelisson, A. (2004). Evidence for a piwi-dependent RNA silencing of the gypsy endogenous retrovirus by the *Drosophila melanogaster* flamenco gene. *Genetics* 166, 1313–1321.

- Seitz, H., Ghildiyal, M., and Zamore, P.D. (2008). Argonaute loading improves the 5' precision of both MicroRNAs and their miRNA strands in flies. *Curr. Biol.* *18*, 147–151.
- Smith, C.D., Shu, S., Mungall, C.J., and Karpen, G.H. (2007). The Release 5.1 annotation of *Drosophila melanogaster* heterochromatin. *Science* *316*, 1586–1591.
- Soper, S.F., van der Heijden, G.W., Hardiman, T.C., Goodheart, M., Martin, S.L., de Boer, P., and Bortvin, A. (2008). Mouse maelstrom, a component of nuage, is essential for spermatogenesis and transposon repression in meiosis. *Dev. Cell* *15*, 285–297.
- Takada, S., Kelkar, A., and Theurkauf, W.E. (2003). *Drosophila* checkpoint kinase 2 couples centrosome function and spindle assembly to genomic integrity. *Cell* *113*, 87–99.
- Theurkauf, W.E. (1994). Immunofluorescence analysis of the cytoskeleton during oogenesis and early embryogenesis. *Methods Cell Biol.* *44*, 489–505.
- Vagin, V.V., Sigova, A., Li, C., Seitz, H., Gvozdev, V., and Zamore, P.D. (2006). A distinct small RNA pathway silences selfish genetic elements in the germline. *Science* *313*, 320–324.
- Verdel, A., and Moazed, D. (2005). RNAi-directed assembly of heterochromatin in fission yeast. *FEBS Lett.* *579*, 5872–5878.
- Vermaak, D., Henikoff, S., and Malik, H.S. (2005). Positive selection drives the evolution of rhino, a member of the heterochromatin protein 1 family in *Drosophila*. *PLoS Genet.* *1*, 96–108.
- Volpe, A.M., Horowitz, H., Grafer, C.M., Jackson, S.M., and Berg, C.A. (2001). *Drosophila* rhino encodes a female-specific chromo-domain protein that affects chromosome structure and egg polarity. *Genetics* *159*, 1117–1134.
- Wang, X.Q., Redpath, J.L., Fan, S.T., and Stanbridge, E.J. (2006). ATR dependent activation of Chk2. *J. Cell. Physiol.* *208*, 613–619.

4.5 Discussion : un système de défense réactif face aux transposons

La plupart des piRNA connus de *Drosophila* (Vagin *et al.*, 2006 ; Saito *et al.*, 2006), du Poisson-Zèbre (Houwing *et al.*, 2007), et une catégorie de piRNA de Souris (Aravin *et al.*, 2007) proviennent de séquences répétées (essentiellement des transposons) ; ils sont le plus souvent en orientation antisens par rapport aux ARN_m des transposons, et semblent donc capables d'en guider la répression post-transcriptionnelle, par les protéines Piwi. Effectivement, la plupart des mutations qui perturbent l'accumulation des piRNA dérèglent l'expression de transposons (voir par exemple les articles n°1, page 31, n°5, page 73 et n°6, page 89) : les piRNA sont donc des répresseurs des éléments génétiques mobiles dans la lignée germinale.

Le mode de biogenèse des piRNA, selon le modèle du ping-pong, garantit que les piRNA seront exprimés dès qu'un transposon est transcrit : son ARN_m (reconnu par un mécanisme qui reste à déterminer) alimentera la boucle d'amplification, qui convertira des précurseurs simple-brin antisens (issus de la transcription des clusters de piRNA) en piRNA antisens, qui guideront la dégradation des ARN_m du transposon, et pourraient provoquer sa mise en silence transcriptionnelle, par hétérochromatinisation (discuté dans Girard et Hannon, 2008).

Les clusters de piRNA jouent ici un rôle particulier : d'après le modèle, ils seraient transcrits, en permanence, sous forme de longs ARN contenant les séquences de tous les transposons réprimables ; seuls les transposons dont la séquence figure dans ces clusters pourront être réprimés, puisque ce sont les transcrits des clusters qui sont convertis en piRNA antisens, effecteurs de la répression. Ce modèle prévoit donc qu'un transposon inédit, qui envahirait une espèce pour la première fois, échapperait à la répression tant qu'une de ses copies n'est pas encore tombée dans l'un de ces clusters.

C'est probablement ce qui s'est passé au cours du XX^{ème} siècle dans les populations sauvages de *Drosophila melanogaster* : alors que leur génome contenait d'anciennes copies, dégénérées, du rétrotransposon « élément I », elles ont été envahies par des copies intactes de cet élément au cours du siècle (voir Busseau *et al.*, 1994 pour une revue ; les copies anciennes et nouvelles sont suffisamment différentes pour que les anciennes ne puissent pas cibler les nouvelles). Elles y sont devenues résistantes, et les souches de *Drosophila* prélevées dans la nature aujourd'hui expriment des piRNA dirigés contre cet élément ; en revanche, les souches de *Drosophila* prélevées au début du XX^{ème} siècle, puis entretenues en laboratoire pendant ces quelques décennies, ont été artificiellement préservées contre cet élément : leur génome ne contient donc aucune copie moderne de l'élément I, et par conséquent, elles ne peuvent pas le réprimer. Cette différence entre souches sauvages et souches de laboratoire se manifeste spectaculairement quand on les croise : alors que la descendance d'un croisement entre une femelle inductrice (dont le génome contient des copies modernes de l'élément I, et qui y est résistante) et un mâle sensible (dont le génome ne contient pas de copie moderne de l'élément I) est fertile, la descendance du croisement réciproque est stérile (ce phénomène, appelé « dysgénie des hybrides », a été observé à de multiples reprises ; la stérilité de la descendance est due à la dérèglement du transposon au cours du développement de sa lignée germinale ; d'autres transposons que l'élément I sont connus pour provoquer le même phénomène).

Si les clusters de piRNA sont nécessaires à la répression du transposon, ils ne sont pas suffisants : Brennecke *et al.*, 2008 ont montré que, dans un croisement dysgénique (σ inducteur \times φ sensible), les clusters de piRNA apportés par l'haplogénome paternel ne suffisent pas à réprimer les copies actives, également apportées par l'haplogénome paternel. Des piRNA déposés par la mère dans le cytoplasme de l'œuf (quand la mère est inductrice) suffisent en revanche à les réprimer. Il faut donc plusieurs générations après l'introduction paternelle d'un nouveau transposon, pour que l'œuf hérite à la fois de piRNA maternels, et de clusters contenant la séquence du transposon, pour réprimer efficacement les copies actives.

Chapitre 5

Quel est le rôle biologique des piRNA ?

5.1 Contexte : un rôle redondant avec celui des siRNA ?

La seule fonction connue des piRNA est de réprimer les séquences répétées dans la lignée germinale (et, chez la Drosophile : dans les cellules somatiques en contact avec les cellules germinales femelles) (voir chapitre 4, page 69). Deux ans après la découverte des piRNA, des siRNA endogènes ont été décrits chez la Drosophile et les Mammifères (voir chapitre 2, page 29), qui répriment séquences répétées et séquences non répétées, dans la lignée germinale et dans le soma.

Il semble donc que les piRNA sont dispensables, puisque les siRNA endogènes peuvent remplir leur fonction (et même davantage). Les piRNA et les siRNA sont tous deux nécessaires à la répression globale des séquences répétées, ce qui montre que chacun contribue partiellement à la répression (par exemple, les mutants de la voie des piRNA, qui n'affectent pas la voie des siRNA, dérèpriment certains transposons : pour ces transposons au moins, les siRNA seuls ne les répriment pas totalement). Mais si effectivement les siRNA peuvent réprimer les transposons, rien ne semble s'opposer à ce que les piRNA disparaissent, et que les siRNA acquièrent la capacité de les remplacer.

Il est donc possible que les piRNA soient finalement dispensables, et que, s'ils disparaissaient, les siRNA puissent rapidement les remplacer ; dans ce cas, les deux systèmes auraient divergé fortuitement. Alternativement, les piRNA pourraient avoir une fonction supplémentaire, encore inconnue, et que ne pourraient pas remplir les siRNA ; cette fonction les rendrait nécessaires dans la lignée germinale, même en présence de la voie des siRNA.

Une redondance entre les deux voies est tout à fait possible, mais elle sera moins crédible si on peut montrer que piRNA et siRNA ont coexisté pendant une longue période évolutive : la redondance aurait dû disparaître (par exemple, les nombres de gènes Argonaute sont très variables parmi les Animaux : voir sous-section 1.2.3, page 16 ; si une voie entière était dispensable, l'un au moins de ses composants aurait dû être perdu). Jusque très récemment, les siRNA et les piRNA n'ont été étudiés, chez les Animaux, que chez des Animaux à symétrie bilatérale (Vertébrés, Drosophiles, Nématodes ; voir figure 5.1, page 105). Pour déterminer plus précisément l'âge de la coexistence des piRNA et des siRNA, nous nous sommes intéressés au seul Cnidaire dont le génome était séquencé en 2007, l'Anémone de mer *Nematostella vectensis*¹.

Le génome de *Nematostella* code notamment trois protéines Ago (dont deux semblent fonctionnelles), trois Piwi, deux Dicer, une Drosha et une Hen1 : cette Anémone de mer semble donc posséder toute la machinerie associée aux petits ARN régulateurs, il est donc probable qu'elle exprime des petits ARN. Nous avons analysé les petits ARN de *Nematostella* par séquençage à haut débit dans deux stades développementaux, et montré qu'elle possède effectivement miRNA, piRNA et siRNA. Les piRNA et les siRNA proviennent essentiellement de transposons, comme chez la Drosophile : nous en concluons que les siRNA et les piRNA ont réprimé les transposons de manière concomitante pendant plus d'un milliard d'années, date estimée de la divergence des Cnidaires et des Animaux à symétrie bilatérale.

Ces résultats sont présentés dans l'article en préparation qui suit (Seitz *et al.*, prep).

¹Depuis, le génome d'un autre Cnidaire a été séquencé (celui de l'Hydre *Hydra magnipapillata*, rendu public en juillet 2008), ainsi que celui d'un Placozoaire, *Trichoplax adhaerens* (rendu public en juin 2008).

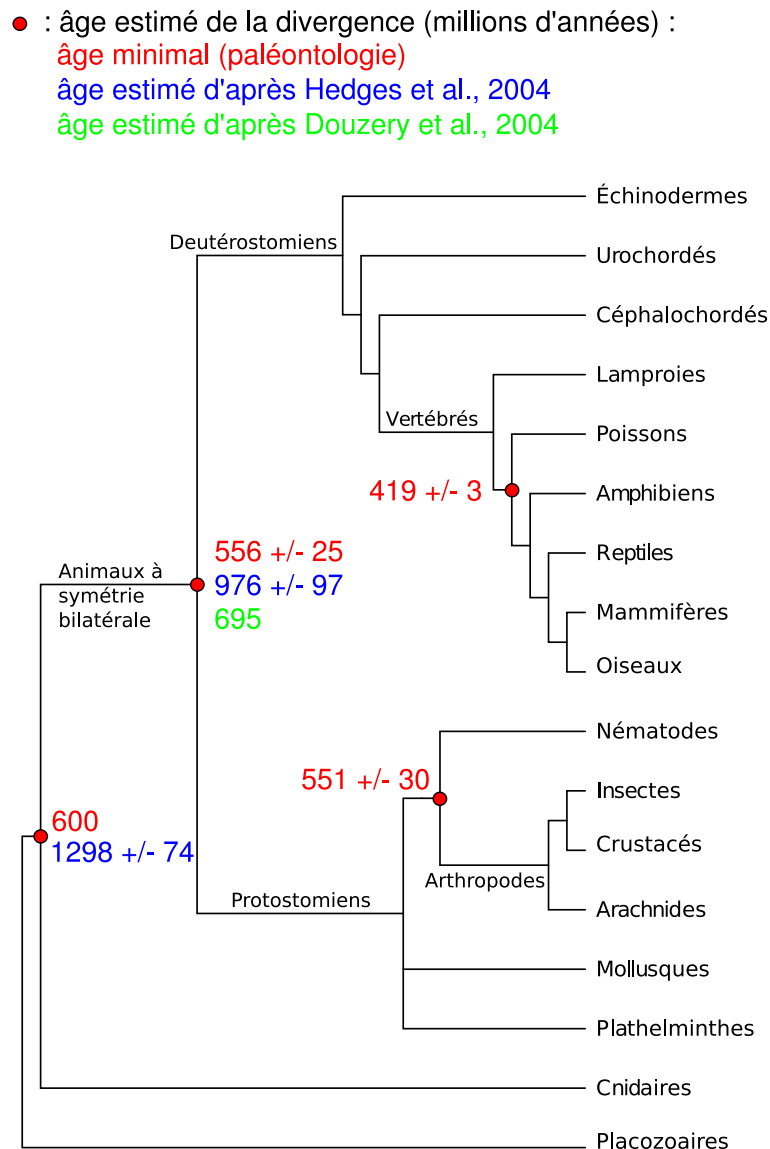


FIG. 5.1 – **Arbre phylogénétique des Animaux.** L'âge estimé de quelques divergences est indiqué. La paléontologie (âges indiqués en rouge) ne peut estimer que des âges minimaux de divergence (par exemple, l'observation de fossiles de Vertébrés terrestres vieux de 419 millions d'années signifie que les Vertébrés terrestres existaient déjà il y a 419 millions d'années — sans indiquer depuis combien de temps ils existaient à cette époque) ; données paléontologiques tirées de <http://www.fossilrecord.net/dateaclade/index.html> (décrit par Benton et Donoghue, 2007) et de Chen *et al.*, 2004. Les données paléontologiques sont peu fiables pour la divergence entre Animaux à symétrie bilatérale et Cnidaires (ces animaux, sans squelette ni carapace, ont laissé peu de fossiles, et l'interprétation des quelques fossiles disponibles est controversée : Bengtson et Budd, 2004). Les estimations de Hedges *et al.*, 2004 et Douzery *et al.*, 2004 reposent sur l'analyse des divergences des séquences géniques.

5.2 Article en préparation : piRNA, microARN et siRNA au cours de l'évolution des Animaux

Le génome de *Nematostella* code plusieurs protéines Argonaute, Dicer, Drosha et Hen1 (voir table 5.1). Cette observation suggérait que l'Anémone de mer exprimerait des miRNA et des piRNA, et éventuellement des siRNA. Nous avons donc préparé des banques de petits ARN (de 18 à 30 nt) pour séquençage à haut débit, à partir d'échantillons d'ARN totaux fournis par Fabian Rentsch (du Sars International Centre for Marine Molecular Biology, à Bergen en Norvège) : nous avons ainsi séquencé les petits ARN d'œufs non fécondés, de blastula, de gastrula, de planula précoce, de planula tardive, de polypes en métamorphose, de polypes primaires, de mâles adultes et de femelles adultes (voir le cycle de *Nematostella* en figure 5.2).

Protéines	Nombre et noms chez <i>Nematostella</i>	Fonction
Ago	Trois, dont deux fonctionnelles : estExt_GenewiseH_1.C_210299 (renommée « Ago1 ») et gw.117.88.1 (renommée « Ago2a »), et une tronquée : gw.117.7.1 (renommée « Ago2b »).	Interagissent avec les siRNA et les miRNA.
Piwi	Trois : e_gw.1.44.1 (renommée « Piwi1 »), e_gw.240.2.1 (renommée « Piwi2 ») et gw.115.8.1 (renommée « Piwi3 »).	Interagissent avec les piRNA.
Dicer	Deux : fgenesH1_pg.scaffold_161000019 et fgenesH1_pg.scaffold_186000022.	Biogenèse des miRNA et siRNA.
Drosha	Une : fgenesH1_pg.scaffold_302000002.	Biogenèse des miRNA.
Hen1	Une : e_gw.40.124.1.	Méthylation des piRNA.

TAB. 5.1 – **Protéines associées aux petits ARN chez *Nematostella*.** Une recherche des protéines à domaines Paz, Piwi et éventuellement DUF1785 (voir sous-section 1.2.3, page 14), à domaines RNase III, et des protéines possédant un domaine méthylase similaire à celui de Hen1 dans le protéome prédit de *Nematostella* révèle trois gènes *Ago*, trois gènes *Piwi*, deux gènes *Dicer*, un gène *Drosha* et un gène *Hen1* (recherche de motifs par HMMer 2.3.2, en définissant le profil de ces domaines à partir des protéines Argonaute, Dicer, Drosha et Hen1 de *Drosophila melanogaster*, *Homo sapiens*, *Schizosaccharomyces pombe* et *Arabidopsis thaliana*). *N.B.* : chez la Drosophile, Hen1 méthyle également les siRNA et certains miRNA (Horwich *et al.*, 2007).

Afin d'identifier d'éventuels petits ARN modifiés sur leur extrémité 3' (comme les piRNA chez les Animaux à symétrie bilatérale, qui portent une 2'-O-méthylation sur leur extrémité 3'), nous avons divisé en deux chaque préparation d'ARN, et nous avons séquencé une moitié sans traitement supplémentaire (banques « non traitées ») et l'autre moitié (banques « oxydées ») a subi une oxydation par le periodate de sodium, en présence d'acide borique (protocole décrit dans Vagin *et al.*, 2006, mais en omettant le traitement au glycérol, la lyophilisation et le traitement à pH basique, inutiles ici).

Les profils de distribution de taille des petits ARN détectés sont représentés figure 5.3. Tout au long du développement, la classe d'ARN la plus abondante est celle dont la taille est comprise entre 27 et 30 nt, la taille habituelle des piRNA chez les autres Animaux (à l'exception des piRNA du Nématode, les piRNA des Animaux à symétrie bilatérale sont longs de 24 à 30 nt; voir section 4.1, page 70). Ces ARN sont détectés dans les banques oxydées, ce qui indique qu'ils portent une modification chimique sur leur extrémité 3' — comme les piRNA des Animaux à symétrie bilatérale. Enfin, ils semblent participer au même mécanisme de biogenèse en ping-pong que les piRNA des Animaux à symétrie bilatérale (voir section 4.1, page 70), puisque ces ARN de *Nematostella* ségrègent en deux catégories : les plus longs (25 à 30 nt)

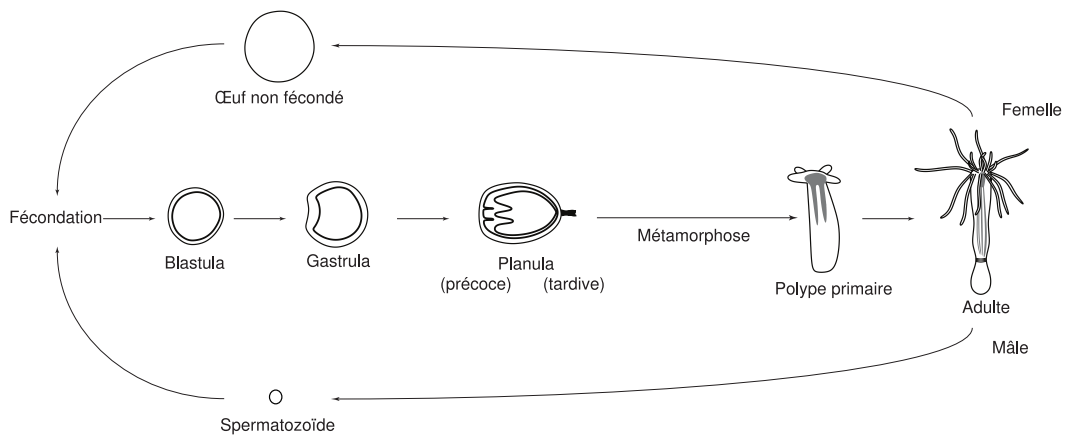


FIG. 5.2 – **Le cycle de *Nematostella vectensis*.** Les œufs fécondés se développent en quelques jours (à 20°C, les planula mobiles apparaissent au trois ou quatrième jour, et métamorphosent autour du septième jour); les polypes primaires grandissent et atteignent la maturité sexuelle pendant le troisième mois (Hand et Uhlinger, 1992).

portent fréquemment une uridine en position 1, alors que les plus courts (23 et 24 nt) portent fréquemment une adénosine en position 10 (voir panneau **A** de la figure 5.4, page 109). Ces deux catégories diffèrent également par leurs origines génomiques : alors que les ARN les plus longs proviennent d'un seul brin de leurs loci, les plus courts proviennent généralement des deux brins (voir panneau **B** de la figure 5.4, page 109) ; il est possible que la catégorie des ARN longs ait une taille très hétérogène, et qu'elle atteigne 23 nt, auquel cas les 23-mères analysés figure 5.4 proviendraient des deux catégories : il est donc tentant d'imaginer que la véritable catégorie des ARN courts provienne d'un seul brin de ses loci (le brin opposé au brin des ARN longs), ce qui expliquerait alors facilement comment ils pourraient être complémentaires des ARN longs, et participer à un mécanisme de biogenèse en ping-pong avec eux.

Ainsi, les ARN de 23 à 30 nt chez *Nematostella* partagent la plupart des caractéristiques des piRNA des Animaux à symétrie bilatérale, et ce sont donc certainement les piRNA de l'Anémone de mer.

De nombreux piRNA de *Nematostella* proviennent de transposons (30 % d'entre eux — alors que les transposons ne constituent que 10 % du génome de *Nematostella*), ce qui montre que, comme chez la *Drosophile* et les Vertébrés, les piRNA proviennent de loci riches en transposons, et semblent pouvoir les réprimer.

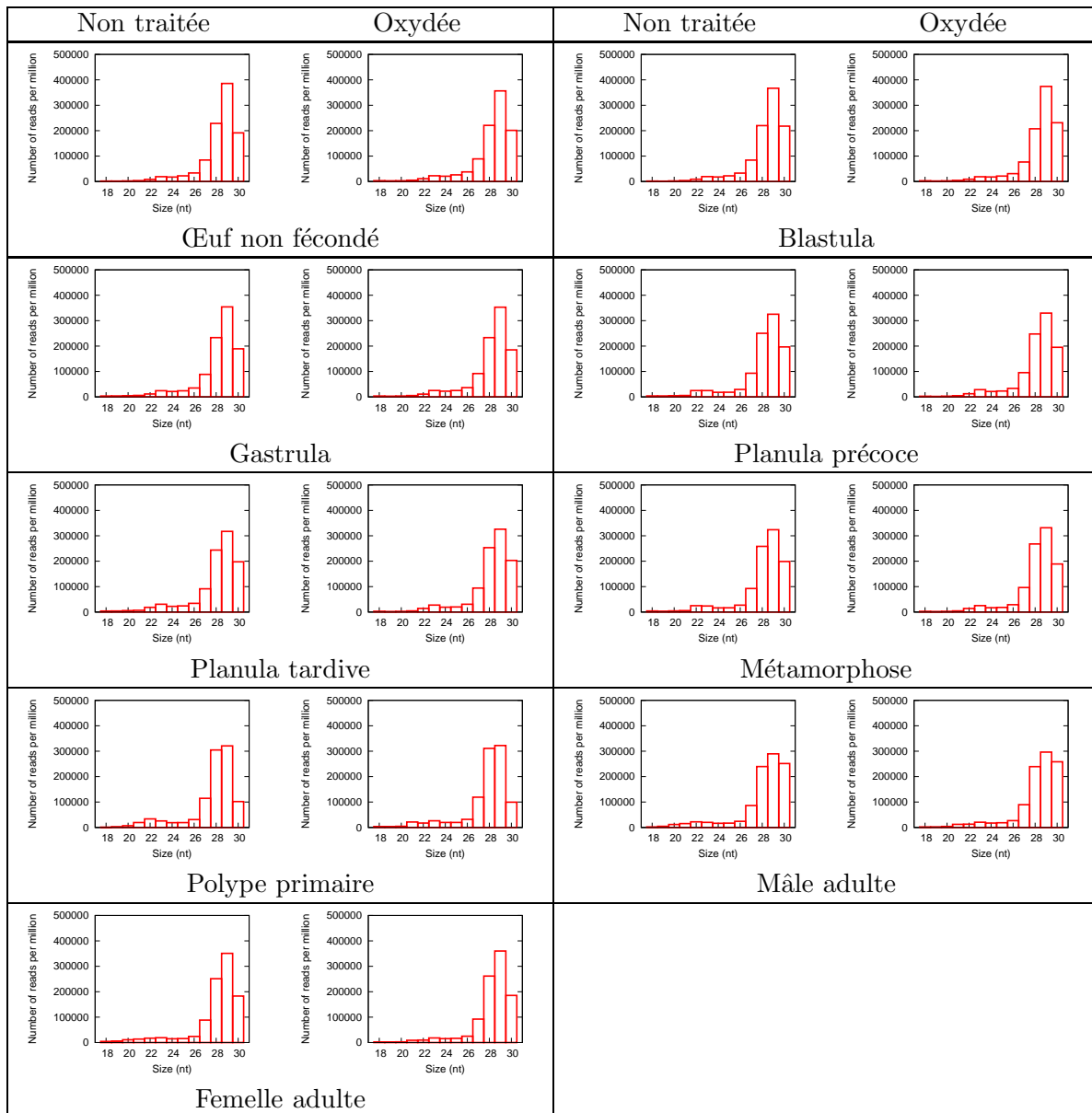


FIG. 5.3 – **Banques de petits ARN de *Nematostella*.** Ces histogrammes représentent la distribution des tailles des petits ARN détectés dans chacune des 18 banques (non traitées, et oxydées), entre 18 et 30 nt. L'axe des ordonnées donne le nombre de séquences lues dans chaque catégorie de taille pour un million de séquences lues dans la banque (dans toute l'analyse, seuls sont considérés les petits ARN dont la séquence est retrouvée dans le génome, et qui ne proviennent pas d'ARN non-codants abondants : ARN ribosomiques, ARN de transfert, petits ARN nucléolaires et autres petits ARN nucléaires).

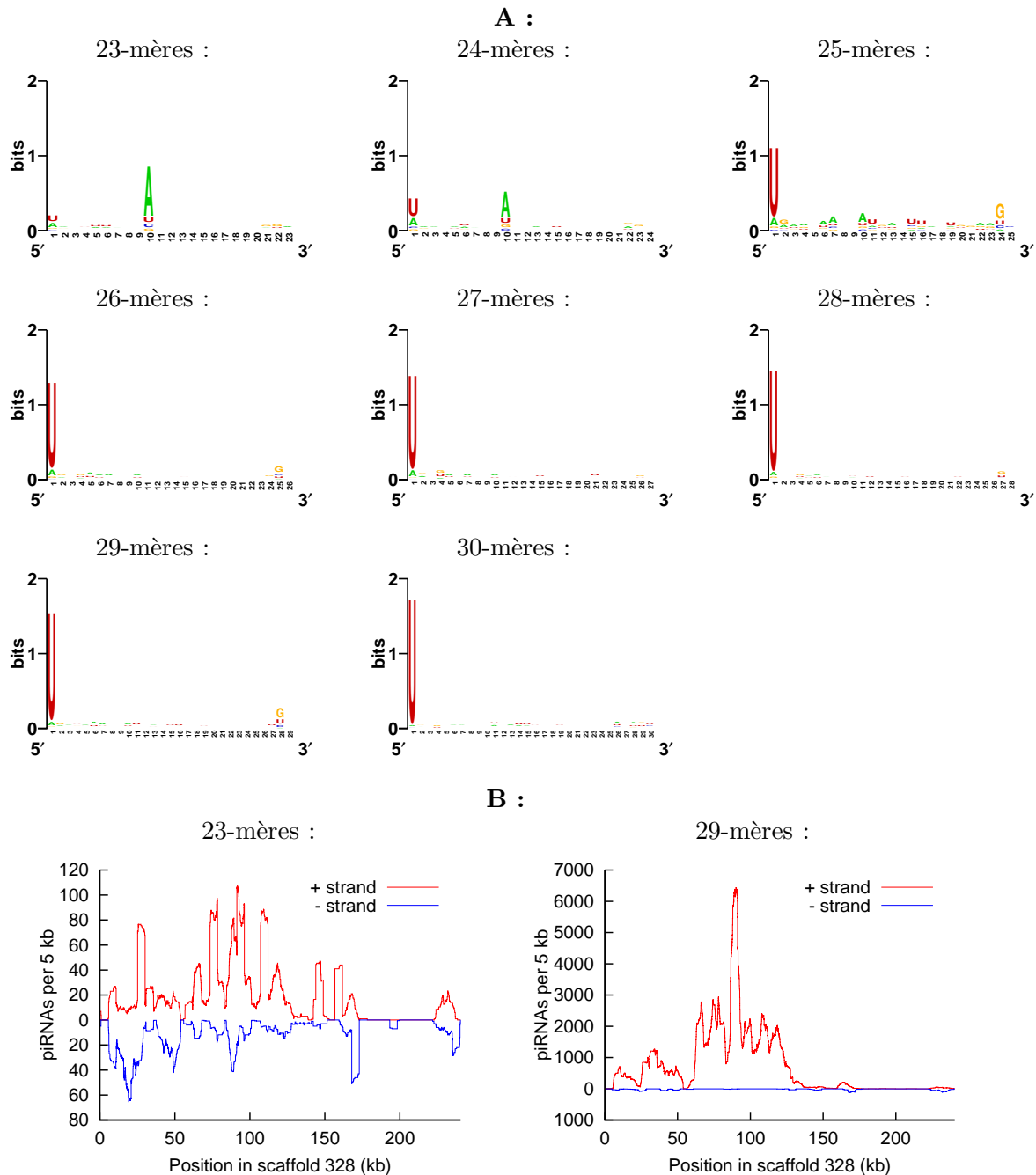


FIG. 5.4 – **Deux catégories de piRNA chez *Nematostella*.** **A :** Représentation en logo (Crooks *et al.*, 2004) des séquences les plus abondantes (5 % les plus abondants) de chaque classe de taille de petits ARN, entre 23 et 30 nt, chez les femelles adultes (banque oxydée). **B :** Les piRNA courts (ici : 23 nt chez les femelles adultes, banque oxydée) proviennent généralement des deux brins de leurs loci, alors que les piRNA longs (ici : 29 nt chez les femelles adultes, banque oxydée) proviennent généralement d'un seul brin. En rouge : densité de piRNA transcrits à partir du brin « + » du locus ; en bleu : densité de piRNA transcrits à partir du brin « - ».

Outre les piRNA, les banques de petits ARN de *Nematostella* révèlent la présence de miRNA (issus de la maturation de tiges-boucles de structure similaire à celle des pre-miRNA des Animaux à symétrie bilatérale), et de petits ARN qui dérivent d'ARN qui se replient en longues tiges-boucles. La figure 5.5 décrit les petits ARN qui proviennent de l'une de ces tiges-boucles : le panneau **A** montre que la plupart de ces ARN sont longs de 19 à 21 nt, une taille voisine de celle des siRNA de *Drosophile* (21 nt exactement, pour la plupart : voir article n°2, page 38) ; le panneau **B** montre que ces petits ARN proviennent des deux brins de la tige-boucle, et que leurs extrémités 5' sont parfois distants d'une vingtaine de nucléotides, ce qui pourrait suggérer que ces petits ARN sont issus d'un clivage processif de la tige-boucle par Dicer, qui cliverait tous les ≈ 20 pb.

Ces petits ARN partagent donc la plupart des caractéristiques des siRNA endogènes des Animaux à symétrie bilatérale (il faut toutefois noter que, à la différence des siRNA de *Drosophile*, ils ne portent pas de modification chimique sur leur extrémité 3'), et ce sont certainement des siRNA endogènes de l'Anémone de mer.

Environ la moitié de ces longues tiges-boucles sont hébergées par des transposons, et les siRNA qui en dérivent sont donc capables de réprimer ces transposons par RNAi : il semble donc que, chez *Nematostella* comme chez les Animaux à symétrie bilatérale, des siRNA endogènes répriment les transposons.

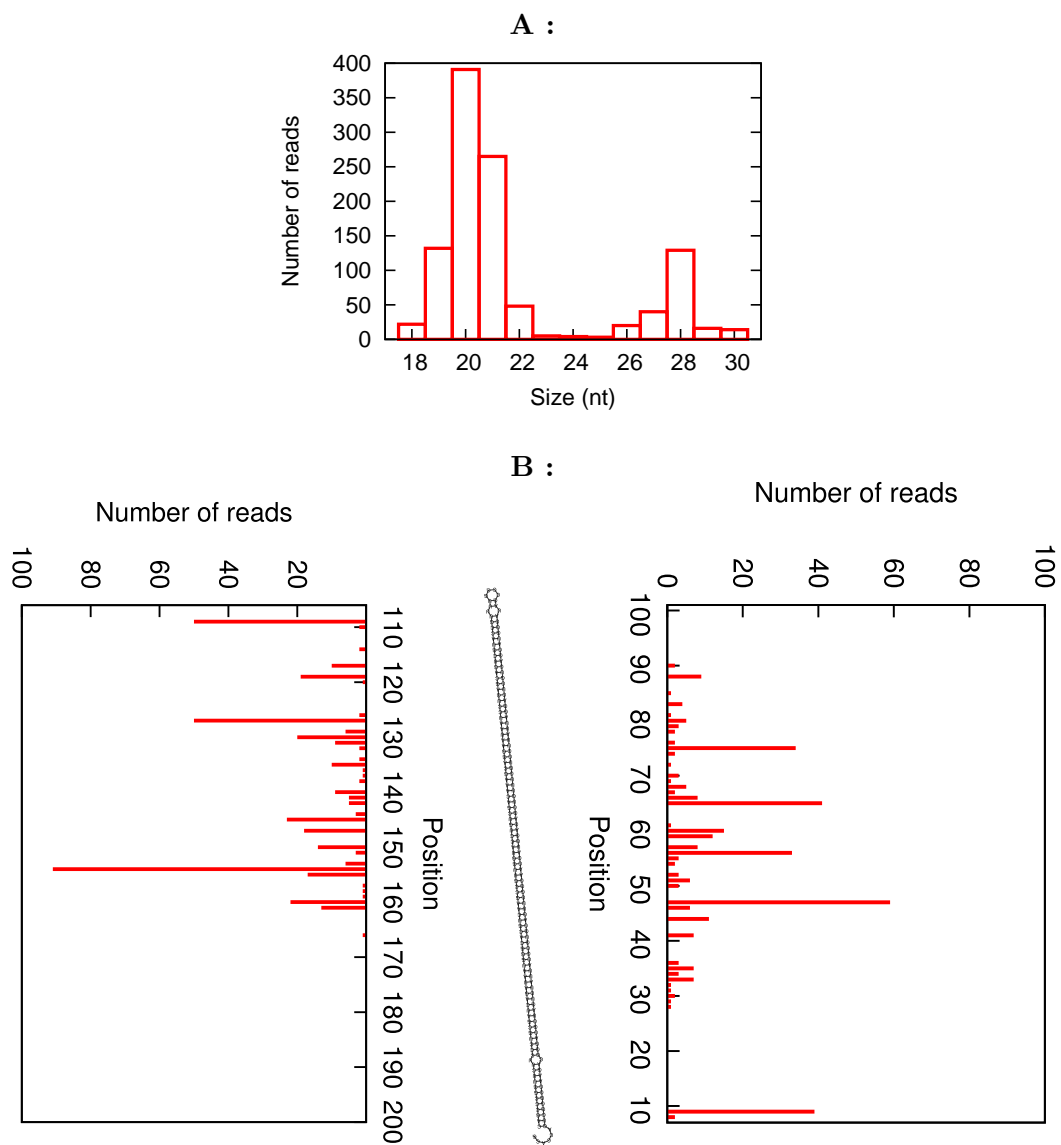


FIG. 5.5 – **Des siRNA endogènes chez *Nematostella*.** **A :** Distribution de tailles des petits ARN de mâle adulte (banque non traitée) qui proviennent d'une tige-boucle de 352 nt (séquence centrée sur une boucle de 4 nt, entre les pb 36021 et 36024 de la séquence « scaffold_480 » de la version 1.0 du génome de *Nematostella vectensis*, en orientation antisens). **B :** Position des extrémités 5' des ARN de 19 à 21 nt qui proviennent de cette tige-boucle (représentés sur les 200 nt centraux de la tige-boucle ; banque non traitée de mâle adulte).

5.3 Discussion : une fonction supplémentaire inconnue

Pendant le cours de cette étude, un autre laboratoire a publié une expérience similaire : Grimson *et al.*, 2008 ont séquencé les petits ARN de plusieurs animaux, dont *Nematostella*, et ils ont rapporté l'observation de miRNA et de piRNA (ils n'ont toutefois pas remarqué de siRNA).

Nos travaux montrent que, chez les Cnidaires comme chez les Animaux à symétrie bilatérale, les piRNA et les siRNA endogènes répriment les transposons. Il semble donc que ces deux classes de petits ARN aient rempli une fonction redondante depuis la divergence de ces deux clades, il y a plus d'un milliard d'années. Une coexistence aussi longue signifie probablement que ces deux classes de petits ARN ne sont pas réellement redondantes, et que chacune est nécessaire, même en présence de l'autre.

Les siRNA répriment également d'autres gènes, et dans les tissus somatiques : peut-être est-ce leur fonction supplémentaire. On ne connaît en revanche aucune fonction aux piRNA, qui ne soit pas réductible à la fonction des siRNA dans la lignée germinale ; la conservation de la voie des piRNA chez les Animaux suggère donc qu'ils jouent également un autre rôle, encore inconnu, et qui les rendrait nécessaires. Je propose de rechercher cette fonction inconnue (voir section 7.1, page 122).

Chapitre 6

Quel est le rôle biologique des microARN ?

6.1 Contexte : le dogme actuel et ses limites

D'après la conception la plus couramment admise, les miRNA régulent un grand nombre de gènes chez les Animaux (plus de la moitié des gènes codants humains : Friedman *et al.*, 2009) ; ils contrôleraient donc tous les processus physiologiques (qui, à une étape ou à une autre, feraient intervenir l'une de ces cibles de miRNA ; voir sous-section 1.3.1, page 19).

Ce dogme est rarement remis en question, alors qu'il ne parvient pas à résoudre trois paradoxes :

- Dans quelques cas bien étudiés *in vivo*, il est apparu que les miRNA analysés avaient un nombre très restreint de cibles (il est possible de corriger tous les défauts visibles du mutant du miRNA en mutant une unique cible — ce qui semble indiquer que, parmi toutes les régulations guidées par le miRNA, une seule aurait un effet macroscopique).
- La répression guidée par les miRNA a été quantifiée pour deux miRNA chez les Mammifères : elle ne dépasse jamais un facteur 2 (et le plus souvent, elle est de l'ordre de 1,1 à 1,2) ; une répression de 50 % est généralement insuffisante pour se traduire en différence observable à l'échelle physiologique : la structure des voies biologiques (intégrées, elles sont peu sensibles aux variations individuelles de l'expression de leurs composants), et les mécanismes d'homéostasie, tamponnent efficacement de telles fluctuations. Il faut sans doute y voir la raison pour laquelle la plupart des gènes chez les Animaux sont haplo-suffisants : une division par deux de leur transcription ne se traduit par aucun phénotype macroscopique ; de la même manière, les variations inter-individus d'expression des gènes dans des populations naturelles dépassent souvent un facteur 2 — alors que tous ces individus sont phénotypiquement sauvages. Il est donc difficile de comprendre comment les miRNA pourraient avoir un effet, si les répressions qu'ils guident sont plus faibles que les fluctuations intrinsèques de l'expression des gènes, facilement tamponnées par des mécanismes d'homéostasie.
- Alors que les miRNA peuvent être très conservés parmi les Animaux, leurs sites de complémentarité dans les 3' UTR des ARN_m le sont beaucoup moins : pourquoi le miRNA serait-il conservé, si ses fonctions régulatrices ne le sont pas ?

Une hypothèse pourrait résoudre ces trois paradoxes : il est probable que la plupart des « cibles » de miRNA identifiées informatiquement ne sont pas suffisamment réprimées pour que cette régulation ait un effet à l'échelle macroscopique ; il faudrait alors expliquer pourquoi leurs sites de complémentarité aux miRNA sont conservés (ils ne sont pas universellement conservés chez les Animaux, mais ils sont au moins conservés localement : parmi les Vertébrés, ou parmi les Nématodes, ou parmi les Insectes ; c'est d'ailleurs sur la base de cette conservation que plusieurs programmes de prédiction de cibles les identifient). Leur conservation phylogénétique signifie certainement que l'interaction entre le miRNA et ces ARN_m a une fonction biologique : cette fonction pourrait être la répression du miRNA. En fixant le miRNA, ces ARN_m (qu'on pourrait appeler « pseudo-cibles ») l'empêchent d'accéder à ses véritables cibles ; la traduction de ces pseudo-cibles serait probablement elle-même réprimée par cette interaction, mais l'activité de leur gène est peu sensible à ces petites variations. La fonction de ces interactions, et la raison de leur conservation phylogénétique, serait donc la modulation de l'activité du miRNA.

Ces trois paradoxes, et cette hypothèse, sont présentés dans l'article qui suit (Seitz, 2009).

6.2 Article n^o7 : une nouvelle définition des cibles de microARN

Current Biology 19, 870–873, May 26, 2009 ©2009 Elsevier Ltd All rights reserved DOI 10.1016/j.cub.2009.03.059

Report

Redefining MicroRNA Targets

Hervé Seitz^{1,2,3,*}¹Department of Biochemistry and Molecular Pharmacology
University of Massachusetts Medical School
364 Plantation Street
Worcester, MA 01605
USA

Summary

Animal microRNAs (miRNAs) guide proteins to repress the translation of target mRNAs via imperfect base pairing between the miRNA and the target. Computational analyses suggest that each miRNA regulates tens or hundreds of targets [1, 2], yet genetic studies usually show that the repression of a few targets plays a physiological role [3–5]. The extent of miRNA-mediated repression (which rarely exceeds 2-fold [1, 2]) is also surprisingly lower than most well-tolerated, intrinsic variations in gene expression [6, 7]. Although miRNA targets are well conserved among closely related species, they differ greatly between more distant animals [8]. The prevailing view is that miRNAs “tune” expression of most of their targets [1, 2]. Here, I propose an alternative hypothesis that could resolve these three paradoxes: many computationally identified miRNA targets may actually be competitive inhibitors of miRNA function, preventing miRNAs from binding their authentic targets by sequestering them. Depending on the prevalence of this type of miRNA:mRNA interaction, this new conception of miRNA regulation could have profound implications on our assumptions about miRNA function.

Results and Discussion

Functional complementarity between a microRNA (miRNA) and a messenger RNA can be imperfect—the best predictor seems to be complementarity between the “seed” of the miRNA (miRNA nucleotides 2 through 7) and the messenger RNA. Accordingly, mRNAs contain a significant proportion of phylogenetically conserved miRNA seed matches, likely reflecting selective pressure. This observation led to the implementation of miRNA target prediction algorithms that define miRNA targets as messenger RNAs containing evolutionarily conserved miRNA seed matches; such algorithms are the most efficient in identifying mRNAs that are actually repressed by miRNAs [1, 2].

Short complementarities between miRNAs and mRNAs are very common in transcriptomes, and efficient prediction programs typically identify tens or hundreds of targets for each miRNA. This observation is hard to reconcile with the few well-studied *in vivo* systems in which the many visible phenotypic defects of a miRNA mutant are rescued by the

mutation of a single target (see Table 1). Similarly, the loss of the miRNA maturation machinery usually results in a few specific deficiencies (discussed in [9]). One interpretation of this discrepancy is that the *in vivo* experiments overlook some subtle phenotypic changes due to the other miRNA targets. For instance, the regulation of *cog-1* by the miRNA *lcy-6* controls the differentiation of a single set of chemosensory neurons in worms [10]. For a given miRNA, if only one target controls an obvious phenotype whereas all of the others have such discreet consequences, then a defect in their regulation would remain unnoticed until specifically probed. It is also possible that many factors determine the functional outcome of specific miRNA:target mRNA interactions and that these *in vivo* studied miRNAs coincidentally have many fewer targets than other miRNAs. Alternatively, this discrepancy could mean that miRNA target prediction programs miss an essential feature of miRNA regulation.

A second apparent paradox is that miRNA-mediated regulation is surprisingly modest: miRNAs typically repress target protein expression by less than 2-fold [1, 2]. Intraindividual variation in protein expression is often greater yet is well tolerated in natural populations [6, 7] (Figure 1). Moreover, few animal genes are haploinsufficient: a 2-fold reduction in gene transcription usually yields no detectable defect [11]. These observations are consistent with the view that biological systems are robust, accommodating such variations in gene expression because most genes are integrated into complex systems that buffer natural fluctuations by negative feedback loops, by tolerance to changes in the concentration of components that are in excess, or by partial functional redundancies. It could be argued that subtly affecting many genes in a single pathway could result in a dramatic change in the final output. Such quantitative predictions on the behavior of complex systems are the goal of the emerging discipline of systems biology. Yet past studies have already shown that the intrinsic organization of biological pathways results in a robust response to such variations [12, 13].

It is thus likely that many cases of miRNA-mediated repression cannot lead to any consequence at a physiological level because their modest repressive effect is buffered by homeostatic mechanisms. Why, then, have tens or hundreds of complementary sites for each miRNA been conserved in evolution?

I propose that a large proportion of miRNA binding sites act as competitive inhibitors for the miRNA. Such a miRNA repressor has been found in plants [14], and in animals, artificial constructs can compete with endogenous mRNAs for miRNA binding [15]. These so-called “miRNA sponges” used in animal cells are coding, messenger RNAs that recapitulate all of the features of natural miRNA targets, and they repress miRNAs in a dose-dependent manner.

As miRNA modulators, these pseudotarget sites are also expected to be phylogenetically conserved—not because they regulate the mRNAs in which they reside, but because they regulate the miRNA that binds them by sequestering it. Such pseudotargets would have the central characteristics of miRNA targets recognized by the prediction algorithms: complementarity to miRNAs and phylogenetic conservation.

*Correspondence: seitz@ibcg.biotoul.fr

²Present address: Université de Toulouse, UPS, Laboratoire de Biologie Moléculaire Eucaryote (LBME), F-31000 Toulouse, France³Present address: CNRS, Laboratoire de Biologie Moléculaire Eucaryote (LBME), F-31000 Toulouse, France

Redefining MicroRNA Targets

871

Table 1. Discrepancy between Genetic and Computational miRNA Target Identification

miRNA	Species	Genetic Experiments		Number of Computationally Predicted Targets	
		Target	Reference	TargetScan	PicTar
<i>let-7</i>	<i>C. elegans</i>	<i>lin-41</i>	[3]	65	89
<i>lin-4</i>	<i>C. elegans</i>	<i>lin-14</i>	[4]	22	32
<i>let-7</i>	<i>D. melanogaster</i>	<i>abrupt</i>	[5]	50	68

miRNA:target mRNA interactions have been validated *in vivo* by suppressor experiments for only three cases. Each of these showed that the mutation of a single target mRNA corrected the pleiotropic defects conferred by the miRNA mutation, in clear contradiction to long lists of computationally predicted targets. Computational predictions were performed with TargetScan v4.2 [17] and the latest available version of PicTar [18] (at medium sensitivity and specificity for the *D. melanogaster* prediction).

MiRNA-sequestering pseudotargets do not need to be insensitive to miRNA-guided repression: after binding a miRNA, translation of these mRNAs would likely be repressed. But such modest repression would be buffered by the robustness of the biological pathways in which these mRNAs participate. In the end, the only difference between real, physiological targets and pseudotargets would be their sensitivity to an ≈ 2 -fold decrease in protein expression. Interestingly, the activity of some of the most validated miRNA targets (identified by *in vivo* studies) is indeed very sensitive to gene dosage (see Table 2). It is noteworthy that authentic targets can also be potent miRNA inhibitors, even though they are themselves repressed to a physiologically noticeable extent.

This hypothesis postulates that only genes whose activity is sensitive to a small reduction in protein expression could be authentic miRNA targets. Examples of such targets include *lin-14*, *lin-41*, and *cog-1* in worms; *abrupt* and *hid* in flies; and *Ptbp1* in mice (repressed by the miRNAs *lin-4*, *let-7*, *lisy-6*, *let-7*, *bantam*, and *miR-124*, respectively). Each gene that is repressed by these miRNAs and whose activity is robust to such a small reduction is proposed to act as a competitive inhibitor; its evolutionary conservation simply means that its expression pattern and the affinity of its mRNA for the miRNA provide a selective advantage by modulating miRNA activity in a cell-type-specific manner.

This idea would also resolve a third apparent paradox: although miRNA binding sites are typically well conserved

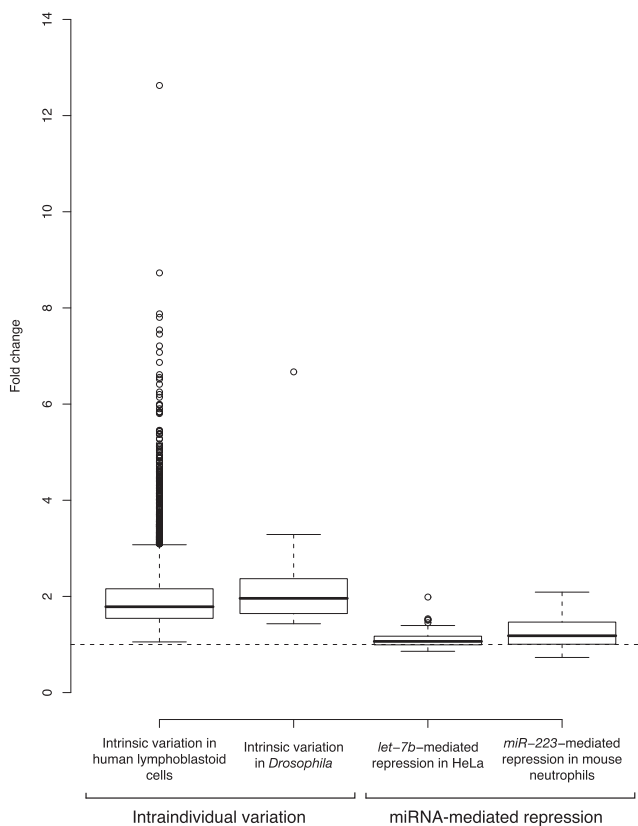


Figure 1. miRNA-Mediated Repression Is Typically Smaller Than Intraindividual Variability in Gene Expression

Two representative examples of expression polymorphism in natural animal populations are displayed here: the expression polymorphism of 5175 genes in human lymphoblastoid cells among 70 individuals (measured at the mRNA accumulation level [7]); the displayed values are the median fold changes relative to the geometric mean of the 70 individual expressions) and of 23 proteins and groups of proteins in *Drosophila* whole body (measured at the protein accumulation level [6]; the displayed values are the maximal observed variations across independent pools of 100 flies). This intraindividual variation is significantly larger than the repression contributed by miRNAs measured in HeLa cells [1] or in mouse neutrophils [2] (Wilcoxon test p value $\leq 1.0 \times 10^{-8}$ for each pairwise comparison). miRNA targets were retrieved with the best prediction algorithm for each data set (PicTar for HeLa cells; TargetScan v4.1 for mouse neutrophils). Gene expression polymorphism in human cells was evaluated for reliably quantified mRNAs (see Supplemental Data for details). The horizontal dashed line indicates the value for an absence of variation.

Current Biology Vol 19 No 10
872

Table 2. Dosage Sensitivity of miRNA Targets

Gene	Proposed Target for	Indication for Dosage Sensitivity
<i>C. elegans lin-14</i>	<i>lin-4</i> [19]	Formation of lateral alae at L4 molt is sensitive to <i>lin-14</i> dosage [20]
<i>C. elegans lin-41</i>	<i>let-7</i> [3]	Worm viability is sensitive to <i>lin-41</i> dosage [3]
<i>D. melanogaster abrupt</i>	<i>let-7</i> [5]	Dendritic branching in multiple dendritic sensory neurons is sensitive to <i>abrupt</i> dosage [21]
<i>C. elegans cog-1</i>	<i>lsy-6</i> [10]	Left/right ASE neuron asymmetry is sensitive to <i>cog-1</i> dosage [22]
<i>D. melanogaster hid</i>	<i>bantam</i> [23]	Survival of eye bristle cells is sensitive to <i>hid</i> dosage [24]
Mammalian myostatin gene (<i>GDF8</i>)	<i>miR-1</i> and <i>miR-206</i> in mutant sheep [25]	Muscle mass and racing performance correlate with gene dosage in dogs [26]
<i>M. musculus Ptbp1</i>	<i>miR-124</i> [27]	Alternative splicing is sensitive to <i>Ptbp1</i> expression level [27]

In addition to the three genetically validated miRNA:target mRNA interactions (first three rows), several interactions have been assessed in vivo (see [28] for review; these experiments typically measure the repression of a reporter gene bearing the target 3' untranslated region without assessing its physiological significance). The three genetically validated targets and four additional candidate targets exhibit dosage-dependent activity.

among closely related species, they differ between distant species [8]. This observation is predicted by the hypothesis. For miRNA pseudotargets, the only features that matter are the abundance and the affinity of the mRNA for a miRNA; the exact identity of these mRNAs does not have any influence. Thus, the repertoire of pseudotargets fluctuates among distantly related species, despite being quite consistent among closely related species. During a long evolutionary divergence, pseudotargets could be replaced by other mRNAs with similar expression patterns. It is tempting to speculate that the few deeply conserved miRNA binding sites identified computationally are enriched in real, physiological targets, whereas the less conserved sites would be mostly pseudotargets [8].

The hypothesis also predicts that the loss of miRNA-sequestering pseudotargets would increase the concentration

of free miRNA, hence enhancing the repression of authentic targets. This phenomenon would explain another puzzling observation: *let-7* mutant lethality is almost completely relieved when *daf-12*, *hbl-1*, *pha-4*, or *lin-41* is repressed by RNAi [16]. If the lethality of *let-7* mutants were due to the overexpression of these targets, then the inviability should be rescued only when all of these genes are repressed, not when only one is repressed. But if some of these genes are pseudotargets for *let-7*, the repression of any of these will release some free *let-7* miRNA (from the hypomorphic *let-7* allele or from other members of the *let-7* miRNA family) and improve the repression of the few authentic *let-7* targets, correcting the phenotype. This puzzling observation could be explained by other gene interaction models, in which the RNAi-targeted genes act downstream of the actual *let-7* targets. However, the pseudotarget hypothesis has the merit

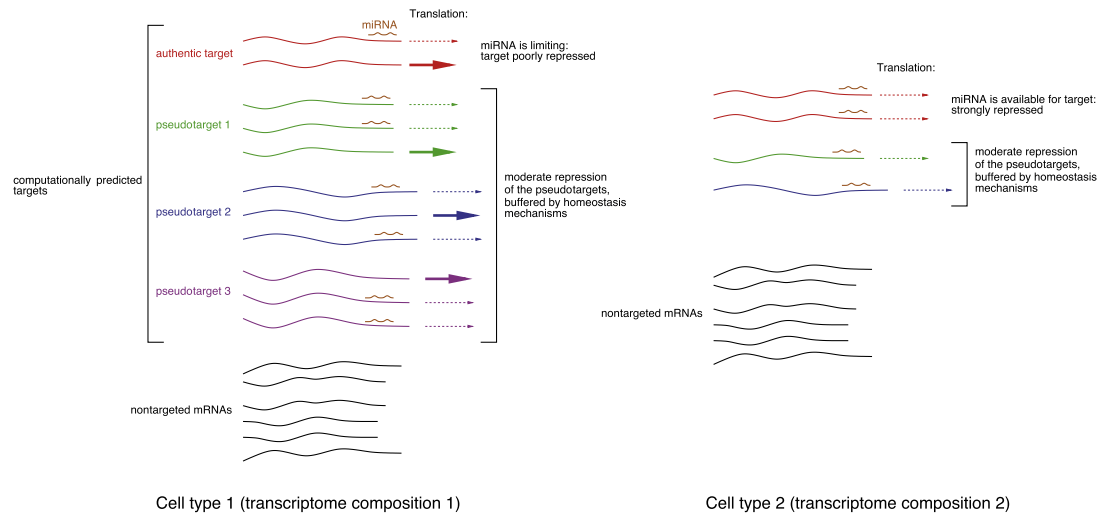


Figure 2. A Model for miRNA Regulation by Pseudotargets

Computationally predicted targets contain both authentic targets (in red) and pseudotargets (in green, blue, and magenta); all of these mRNAs exhibit phylogenetically conserved, complementary sites accessible to the miRNA. However, the moderate miRNA-guided repression of pseudotargets is buffered by the robustness of their biological pathways; the only physiological effects of the miRNA are due to the repression of authentic targets. In cell type 1, in which pseudotargets are abundantly expressed, the miRNA is titrated by pseudotargets; authentic targets are poorly repressed. In cell type 2, in which pseudotargets are sparse, the miRNA is available; authentic targets are strongly repressed.

Redefining MicroRNA Targets

873

of relying only on proven interactions between *let-7* and all of these genes.

In sum, according to this hypothesis, every phylogenetically conserved miRNA-complementary mRNA is not regulated by that miRNA; rather, most miRNA “targets” titrate miRNAs. That is, most miRNA “targets” regulate the miRNA, and not the reverse. Authentic miRNA regulatory targets need to be not only complementary and accessible to the miRNA but also sufficiently sensitive to a modest reduction in their protein production to benefit from miRNA regulation. Finally, although miRNAs have been known for some time to shape cell type identity, this hypothesis also predicts that each cell type, as a consequence of the composition of its transcriptome, modulates miRNA activity (see Figure 2). Such a system would ensure that the regulation of miRNA targets integrates expression information from hundreds of other genes.

Supplemental Data

The Supplemental Data include Supplemental Experimental Procedures and one figure and can be found with this article online at [http://www.cell.com/current-biology/supplemental/S0960-9822\(09\)00913-0](http://www.cell.com/current-biology/supplemental/S0960-9822(09)00913-0).

Acknowledgments

The author thanks Phillip Zamore for critical reading of the manuscript and outstanding mentorship and Victor Ambros for helpful discussions. The author was supported by EMBO Long-Term Fellowship ALTF 910-2004 and Human Frontier Science Program Long-Term Fellowship LT00575/2005-L.

Received: February 17, 2009

Revised: March 23, 2009

Accepted: March 26, 2009

Published online: April 16, 2009

References

- Selbach, M., Schwanhäusser, B., Thierfelder, N., Fang, Z., Khanin, R., and Rajewsky, N. (2008). Widespread changes in protein synthesis induced by microRNAs. *Nature* 455, 58–63.
- Baek, D., Villén, J., Shin, C., Camargo, F.D., Gygi, S.P., and Bartel, D.P. (2008). The impact of microRNAs on protein output. *Nature* 455, 64–71.
- Slack, F.J., Basson, M., Liu, Z., Ambros, V., Horvitz, H.R., and Ruvkun, G. (2000). The *lin-41* RBCC gene acts in the *C. elegans* heterochronic pathway between the *let-7* regulatory RNA and the LIN-29 transcription factor. *Mol. Cell* 5, 659–669.
- Ambros, V. (1989). A hierarchy of regulatory genes controls a larva-to-adult developmental switch in *C. elegans*. *Cell* 57, 49–57.
- Ambros, V., and Chen, X. (2007). The regulation of genes and genomes by small RNAs. *Development* 134, 1635–1641.
- Laurie-Ahlberg, C.C., Wilton, A.N., Curtsinger, J.W., and Emigh, T.H. (1982). Naturally occurring enzyme activity variation in *Drosophila melanogaster*. I. Sources of variation for 23 enzymes. *Genetics* 102, 191–206.
- Cheung, V.G., Conlin, L.K., Weber, T.M., Arcaro, M., Jen, K.Y., Morley, M., and Spielman, R.S. (2003). Natural variation in human gene expression assessed in lymphoblastoid cells. *Nat. Genet.* 33, 422–425.
- Chen, K., and Rajewsky, N. (2006). Deep conservation of microRNA-target relationships and 3' UTR motifs in vertebrates, flies, and nematodes. *Cold Spring Harb. Symp. Quant. Biol.* 71, 149–156.
- Carthew, R.W. (2006). Gene regulation by microRNAs. *Curr. Opin. Genet. Dev.* 16, 203–208.
- Johnston, R.J., and Hobert, O. (2003). A microRNA controlling left/right neuronal asymmetry in *Caenorhabditis elegans*. *Nature* 426, 845–849.
- Fisher, R.A. (1930). *The Genetical Theory of Natural Selection*, First Edition (Oxford: Oxford University Press), pp. 51–52. <http://openlibrary.org/details/geneticaltheory031631mbp>.
- Wright, S. (1934). Physiological and evolutionary theories of dominance. *Am. Nat.* 68, 24–53.
- Kacser, H., and Burns, J.A. (1981). The molecular basis of dominance. *Genetics* 97, 639–666.
- Franco-Zorrilla, J.M., Valli, A., Todesco, M., Mateos, I., Puga, M.I., Rubio-Somoza, I., Leyva, A., Weigel, D., García, J.A., and Paz-Ares, J. (2007). Target mimicry provides a new mechanism for regulation of microRNA activity. *Nat. Genet.* 39, 1033–1037.
- Ebert, M.S., Neilson, J.R., and Sharp, P.A. (2007). MicroRNA sponges: Competitive inhibitors of small RNAs in mammalian cells. *Nat. Methods* 4, 721–726.
- Grosshans, H., Johnson, T., Reinert, K.L., Gerstein, M., and Slack, F.J. (2005). The temporal patterning microRNA *let-7* regulates several transcription factors at the larval to adult transition in *C. elegans*. *Dev. Cell* 8, 321–330.
- Lewis, B.P., Burge, C.B., and Bartel, D.P. (2005). Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* 120, 15–20.
- Lall, S., Grün, D., Krek, A., Chen, K., Wang, Y.L., Dewey, C.N., Sood, P., Colombo, T., Bray, N., MacMenamin, P., et al. (2006). A genome-wide map of conserved microRNA targets in *C. elegans*. *Curr. Biol.* 16, 460–471.
- Wightman, B., Ha, I., and Ruvkun, G. (1993). Posttranscriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal pattern formation in *C. elegans*. *Cell* 75, 855–862.
- Ambros, V., and Horvitz, H.R. (1987). The *lin-14* locus of *Caenorhabditis elegans* controls the time of expression of specific postembryonic developmental events. *Genes Dev.* 1, 398–414.
- Li, W., Wang, F., Menut, L., and Gao, F.B. (2004). BTB/POZ-zinc finger protein abrupt suppresses dendritic branching in a neuronal subtype-specific and dosage-dependent manner. *Neuron* 43, 823–834.
- Chang, S., Johnston, R.J., and Hobert, O. (2003). A transcriptional regulatory cascade that controls left/right asymmetry in chemosensory neurons of *C. elegans*. *Genes Dev.* 17, 2123–2137.
- Brennecke, J., Hipfner, D.R., Stark, A., Russell, R.B., and Cohen, S.M. (2003). *bantam* encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene *hid* in *Drosophila*. *Cell* 113, 25–36.
- Grether, M.E., Abrams, J.M., Agapite, J., White, K., and Steller, H. (1995). The *head involution defective* gene of *Drosophila melanogaster* functions in programmed cell death. *Genes Dev.* 9, 1694–1708.
- Clop, A., Marcq, F., Takeda, H., Pirottin, D., Tordoir, X., Bibé, B., Bouix, J., Caiment, F., Elsen, J.M., Eychenne, F., et al. (2006). A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. *Nat. Genet.* 38, 813–818.
- Mosher, D.S., Quignon, P., Bustamante, C.D., Sutter, N.B., Mellersh, C.S., Parker, H.G., and Ostrander, E.A. (2007). A mutation in the myostatin gene increases muscle mass and enhances racing performance in heterozygote dogs. *PLoS Genet.* 3, e79.
- Makeyev, E.V., Zhang, J., Carrasco, M.A., and Maniatis, T. (2007). The microRNA miR-124 promotes neuronal differentiation by triggering brain-specific alternative pre-mRNA splicing. *Mol. Cell* 27, 435–448.
- Stefani, G., and Slack, F.J. (2008). Small non-coding RNAs in animal development. *Nat. Rev. Mol. Cell Biol.* 9, 219–230.

6.3 Discussion : un nouveau cadre théorique

Cette hypothèse propose donc de redéfinir les cibles de miRNA : une cible ne serait pas un ARN_m réprimé par un miRNA, ce serait : un ARN_m réprimé *suffisamment* par un miRNA ; suffisamment, pour que cette régulation ait un effet sur la physiologie de l'animal.

De ce point de vue, la plupart des « cibles » proposées par les programmes informatiques ne seraient pas de véritables cibles ; la conservation de leur interaction avec le miRNA signifierait simplement que, en le titrant dans les cellules où elles sont le plus abondamment exprimées, elles moduleraient son activité, et cette modulation présenterait un avantage sélectif.

La compétition entre ARN_m pour la fixation du miRNA n'avait jamais été envisagée ; un cas similaire avait toutefois été observé chez les Plantes : le gène non-codant *IPS1* d'*Arabidopsis thaliana* est exprimé sous la forme d'un ARN qui fixe le miRNA miR-399, et l'empêche de reconnaître ses véritables cibles (Franco-Zorrilla *et al.*, 2007). Rien ne s'oppose *a priori* à ce que les ARN_m eux-mêmes puissent jouer ce rôle : finalement, en considérant que toutes les interactions conservées entre ARN_m et miRNA étaient dues à la fonction répressive des miRNA sur les ARN_m, les auteurs des prédictions de cibles faisaient l'hypothèse implicite que les ARN_m ne pouvaient pas entrer en compétition pour le miRNA.

Si cette hypothèse était avérée, il faudrait donc revoir notre conception de la régulation guidée par les miRNA :

- d'une part, il faudrait que les prédictions de cibles tiennent compte de la sensibilité de l'activité des gènes à de petites fluctuations de leur expression, et qu'elles la compare à la répression guidée par les miRNA, ce qui leur permettrait de distinguer pseudo-cibles et cibles véritables ;
- d'autre part, il faudrait (dans l'estimation de l'ampleur de la répression guidée par le miRNA) tenir compte de toutes les pseudo-cibles exprimées dans le type cellulaire considéré : le résultat quantitatif de la répression ne dépendra pas uniquement du miRNA et de sa cible, mais également, de l'ensemble du transcriptome de la cellule.

Je propose de mettre cette hypothèse à l'épreuve, en testant plusieurs de ses prédictions (que ne prévoyait pas le dogme actuel), puis, si l'hypothèse s'avère plus crédible que le dogme actuel, je propose de contribuer à l'amélioration de la prédiction des cibles (voir section 7.2, page 127).

Chapitre 7

Projet de recherche

7.1 Quelle est la fonction des piRNA ?

7.1.1 Objectif

Le chapitre 5, page 103, a montré que les piRNA et les siRNA avaient conjointement réprimé les séquences répétées chez les Animaux pendant au moins un milliard d'années (l'âge approximatif de la divergence entre Cnidaires et Animaux à symétrie bilatérale), ce qui indiquait que les piRNA jouent probablement un rôle supplémentaire, non réductible à la répression post-transcriptionnelle guidée par les siRNA. Je propose de rechercher cette fonction additionnelle.

Étant donné que les piRNA sont spécifiques de la lignée germinale (et des cellules somatiques qui participent à l'ovogenèse : Li *et al.*, 2009a ; Malone *et al.*, 2009), cette fonction inconnue joue certainement un rôle dans la reproduction. Plus précisément, une série d'observations collectées chez la Drosophile suggère que les piRNA sont nécessaires au maintien de l'intégrité du génome de la lignée germinale : l'absence de piRNA active la voie ATR/Chk2 de réparation des cassures double-brin de l'ADN (voir Klattenhoff et Theurkauf, 2008 pour une revue). On ne sait pas encore si cet effet est dû à la mobilisation des transposons (de nombreux transposons sont déréprimés chez les mutants de la voie des piRNA, et leur insertion dans le génome implique une coupure d'ADN double-brin — et même trois coupures dans le cas des transposons à ADN) ou à un effet plus direct des piRNA sur la réparation des cassures double-brin. Si cet effet était dû à la mobilisation des transposons, alors à nouveau, il faudrait expliquer pourquoi les siRNA (qui répriment également les transposons) ne suffiraient pas à les réprimer.

Je compte utiliser une caractéristique de la biologie de l'Anémone de mer *Nematostella vectensis* pour découpler la fonction reproductrice des autres caractéristiques de la lignée germinale — pour déterminer auquel de ces phénomènes les piRNA sont nécessaires.

Nematostella se reproduit à la fois par reproduction sexuée et par reproduction asexuée (Hand et Uhlinger, 1992) :

- Dans la reproduction sexuée, une femelle adulte libère une masse d'œufs dans le milieu, et un mâle adulte émet des spermatozoïdes qui fécondent les œufs en moins de deux heures, entre 20 et 24 °C (Fritzenwanker et Technau, 2002).
- Dans la reproduction asexuée, une constriction apparaît dans la région aborale, qui se resserre jusqu'à séparer physiquement la partie orale de la partie aborale ; chacune régénère ensuite la partie manquante (Darling *et al.*, 2005). Un autre mode de fission, beaucoup plus rare, commence par faire apparaître un deuxième orifice oral à l'extrémité aborale, puis une constriction sépare les deux individus au milieu du corps (Darling *et al.*, 2005).

Donc chez *Nematostella*, le soma peut produire une descendance, par reproduction asexuée : je vais déterminer si le soma en cours de bourgeonnement exprime des piRNA (ce qui signifierait que l'expression des piRNA est toujours associée à la production d'une descendance, que ce soit par la lignée germinale ou par le soma). Si les piRNA ne sont pas exprimés par le soma pendant la fission, alors ces petits ARN seront spécifiques de la lignée germinale (et de ses cellules somatiques associées), indépendamment de sa fonction reproductrice.

7.1.2 Démarche expérimentale

Il est possible d'induire la reproduction asexuée chez *Nematostella* en imposant une constriction sur le milieu du corps en y nouant un fil (Putnam *et al.*, 2007). Plutôt que de détecter les piRNA eux-mêmes (difficiles à détecter par hybridation *in situ* : chaque piRNA individuel est peu exprimé, bien que, collectivement, ils soient très abondants ; voir chapitre 5, page 103 ; quant aux techniques de biologie moléculaire, comme le séquençage à haut débit, elles seront dangereusement sensibles à une contamination par les gonades), nous allons rechercher les ARN_m des protéines de la sous-famille Piwi (il en existe trois chez *Nematostella* : voir figure 7.1), par hybridation *in situ* au cours de la chronologie de la fission.

Un protocole d'hybridation *in situ* a été récemment mis au point pour l'Anémone de mer adulte (Burton et Finnerty, 2009) ; un protocole similaire a déjà été utilisé pour déterminer le patron d'expression d'un gène Piwi chez un autre Cnidaire, la Méduse *Podocoryne carnea* (cette expérience a d'ailleurs montré que ce gène, baptisé *Cniwi*, est exprimé principalement dans les ovogonies et les spermatogonies chez l'adulte) (Seipel *et al.*, 2004).

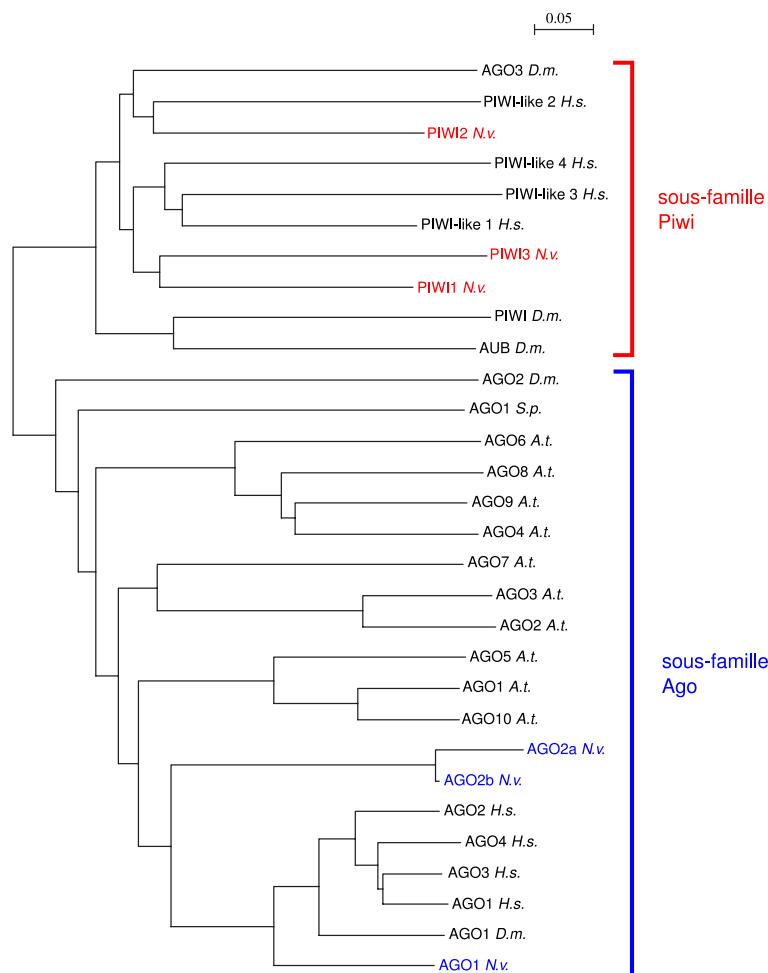


FIG. 7.1 – **Les protéines Argonaute de *Nematostella vectensis*.** Six protéines prédites de *N. vectensis* possèdent un domaine Piwi et un domaine Paz ; un alignement multiple de leurs domaines Piwi avec ceux d'autres protéines Argonaute connues (avec le programme `clustalw 2.0.9`) montre que trois d'entre elles (représentées en rouge) appartiennent à la sous-famille Piwi, et trois autres (en bleu) appartiennent à la sous-famille Ago (un alignement des domaines Paz donne le même résultat). *N.B.* : le gène que j'ai baptisé « *Ago2b* » chez *Nematostella* n'est pas fonctionnel (sa phase ouverte de lecture est interrompue par deux codons stop), il s'agit probablement d'un pseudogène de *Ago2a*. La barre d'échelle indique la longueur de branche pour une fréquence de substitution de 0.05 par acide aminé. *D.m.* : *Drosophila melanogaster* ; *H.s.* : *Homo sapiens* ; *N.v.* : *Nematostella vectensis* ; *S.p.* : *Schizosaccharomyces pombe* ; *A.t.* : *Arabidopsis thaliana*. Nomenclature des protéines Argonaute dans la banque de données <http://genome.jgi-psf.org/Nemve1/Nemve1.home.html> : Piwi1 est e_gw.1.44.1 ; Piwi2 est e_gw.240.2.1 ; Piwi3 est gw.115.8.1 ; Ago1 est estExt_GenewiseH_1.C_210299 ; Ago2a est gw.117.88.1 ; Ago2b est gw.117.7.1.

Cette expérience ne sera concluante que si la reproduction asexuée existe réellement dans

des populations sauvages (ce qui signifierait qu'elle est soumise à la sélection naturelle). Il est formellement possible que la reproduction asexuée soit en réalité un artefact des conditions de culture en laboratoire : elle pourrait n'être qu'une conséquence de l'activité mal contrôlée du système de régénération de *Nematostella* dans des conditions d'abondance de nourriture (en cultivant *Nematostella* pendant plus d'un an, je n'ai observé des événements de reproduction asexuée que sur les individus les plus grands, longs de plusieurs centimètres ; il est connu que *Nematostella* — moins de 2 cm de long, en général, dans son milieu naturel ¹ — atteint des longueurs très supérieures en laboratoire : Hand et Uhlinger, 1992). Les étonnantes capacités de régénération de *Nematostella* (capable de régénérer en quelques jours toute partie sectionnée : Darling *et al.*, 2005) démontrent l'existence d'une population de cellules totipotentes, ou de cellules capables de se différencier en cellules totipotentes, réparties dans tout l'animal ².

Mais Hand et Uhlinger, 1994 ont montré que la reproduction asexuée avait un réel impact écologique (elle n'est donc pas restreinte aux cultures de laboratoire) : certaines populations naturelles de *N. vectensis* sont constituées d'un seul sexe, mâle ou femelle (voir figure 7.2). Ces biotopes ont donc certainement été colonisés par un seul individu, ou un petit nombre d'individus, qui avaient tous le même sexe, et les populations qui y sont observées à présent ne peuvent donc provenir que d'une multiplication de ces quelques individus par reproduction asexuée. L'impact écologique évident de la reproduction asexuée chez *Nematostella* implique qu'elle est soumise à la sélection naturelle, donc si effectivement les piRNA étaient nécessaires à la production d'une descendance, l'expression somatique des piRNA au cours de la fission aura été conservée.

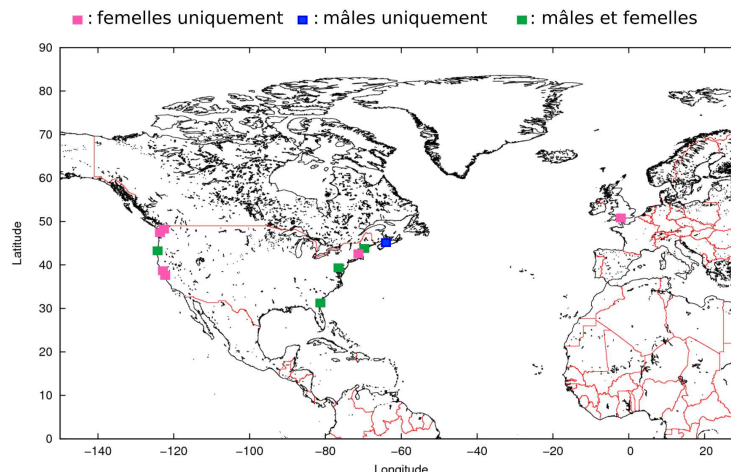


FIG. 7.2 – Répartition géographique de *Nematostella vectensis*. Seules les populations confirmées de *N. vectensis* (dont l'appartenance taxonomique a été vérifiée par croisements avec une souche de référence) sont représentées ici ; données tirées de Hand et Uhlinger, 1994. Certaines populations sont constituées de femelles uniquement (carrés roses), d'autres de mâles uniquement (carrés bleus), d'autres contiennent les deux sexes (carrés verts).

Nous déterminerons donc si les protéines Piwi sont exprimées dans le soma de *Nematostella* au cours de la fission :

1. S'il s'avère que les Piwi sont exprimés pendant la reproduction asexuée, alors la voie des piRNA sera systématiquement associée à la production d'une descendance. Pour déterminer

¹Les zones d'estuaire, sur les côtes des États-Unis et de l'Angleterre ; voir figure 7.2.

²Chez un autre Cnidaire, l'Hydre, il semblerait que toutes les parties du corps, sauf les tentacules, soient capables de régénérer un animal entier (Brigitte Galliot, communication personnelle).

l'effet des Piwi sur la reproduction, nous réprimerons les trois gènes Piwi (seuls ou en combinaison) par RNAi ; un protocole de RNAi a déjà été mis au point pour un autre Cnidaire, l'Hydre (Buzgariu *et al.*, 2008), et les nombreuses similitudes entre Hydre et Anémone de mer devraient faciliter l'adaptation de ce protocole à *Nematostella*. Cette expérience pourrait mettre en évidence une mobilisation des transposons en l'absence d'une (ou : des) protéine(s) Piwi. Nous traiterons alors ces anémones de mer avec des ARN double-brin dirigés contre les transposons déréprimés par la perte des Piwi, pour voir s'il est possible de compléter la disparition de la voie des piRNA par des siRNA. Simultanément, nous vérifierons la généralité de nos résultats dans un organisme-modèle où la génétique est plus facile, *Drosophila melanogaster*. Les loci de piRNA chez la Drosophile sont bien caractérisés, et la transgénèse y est facile.

- (a) Si les siRNA peuvent compléter la perte des piRNA, alors la coexistence des piRNA et des siRNA pendant un milliard d'années deviendra un problème fascinant. Nous étudierons les effets à long terme de la perte des piRNA, complétés par des siRNA, sur la santé des populations de Drosophile ; nous localiserons, par clonage, les sites de cassure double-brin de l'ADN dans les mouches non complétées, pour déterminer si ces cassures sont dues à la mobilisation des transposons.
 - (b) Sinon, il nous faudra comprendre pourquoi la répression des transposons par RNAi ne corrige pas le phénotype des mutants de la voie Piwi. Nous chercherons donc les différences de modes d'action entre les piRNA et les siRNA. Une explication possible pourrait tenir à la nature de la répression guidée par ces deux classes de petits ARN : alors que les siRNA semblent être des régulateurs uniquement post-transcriptionnels chez les Animaux (voir sous-section 1.2.4, page 16 : en dépit d'une recherche intense, il existe peu d'indications d'un effet des siRNA sur la structure chromatinienne chez les Animaux), les piRNA semblent pouvoir guider des modifications de la chromatine, qui pourraient réprimer les séquences-cibles transcriptionnellement (Aravin *et al.*, 2008). Nous utiliserons alors un autre système : deux lignées cellulaires d'Insectes, récemment décrites, expriment des piRNA et des protéines Piwi (Kawaoka *et al.*, 2009 ; Lau *et al.*, 2009). Nous les transfecterons avec des plasmides homologues aux loci de piRNA exprimés, pour déterminer par séquençage après traitement au bisulfite, si les piRNA peuvent guider une méthylation d'ADN en *trans*.
2. S'il s'avère au contraire que les piRNA restent spécifiques de la lignée germinale (et qu'ils ne sont pas exprimés dans le soma qui subit une reproduction asexuée), il nous faudra alors déterminer à quelle autre spécificité de la lignée germinale ils participent ³. La différence la plus remarquable entre soma et lignée germinale est la méiose : les cellules germinales sont les seules à subir une réduction de 2N à N chromosomes (que ne subissent pas non plus les cellules somatiques pendant la reproduction asexuée). Les piRNA pourraient donc être nécessaires à la méiose : plusieurs observations suggèrent un rôle des piRNA dans l'acquisition d'une structure chromosomique particulière (Lei et Corces, 2006 ; Brower-Toland *et al.*, 2007, et discutées dans Girard et Hannon, 2008) et dans l'assemblage du complexe synaptonémal et dans la progression de la méiose (Blumenstiel *et al.*, 2008). Chez les Mammifères, la classe de piRNA la plus abondante est exprimée au stade pachytène de

³Pour l'heure, des piRNA somatiques n'ont été détectés que dans l'ovaire de la Drosophile, où ils participent à la répression de rétrotransposons qui infecteraient l'ovocyte par des particules apparentées aux particules virales. Il semble donc que l'expression de piRNA somatiques soit une extension du système de piRNA germinaux, apparue en réponse au développement de ces transposons (Li *et al.*, 2009a). Les piRNA somatiques seraient donc apparus par une exaptation (acquisition d'une nouvelle fonction à partir d'une fonction ancienne, sous l'effet d'une pression de sélection) récente (un autre Insecte, le Lépidoptère *Bombyx mori*, possède un système orthologue aux piRNA germinaux de la Drosophile, mais aucun équivalent de la voie somatique : Kawaoka *et al.*, 2009 ; voir également la section 4.3, page 86), et il semble donc raisonnable de considérer que les fonctions germinales des piRNA sont responsables de leur conservation chez les Animaux.

la méiose (lorsque les chromosomes homologues sont appariés), et sa fonction est encore inconnue (ces piRNA ne semblent pas réprimer les séquences répétées) (Lau *et al.*, 2006 ; Aravin *et al.*, 2006 ; Girard *et al.*, 2006 ; Aravin *et al.*, 2007). Nous chercherons donc l'effet d'une répression par RNAi des gènes Piwi sur l'appariement des chromosomes homologues pendant la méiose chez *Nematostella* (par microscopie, après répression des gènes *Piwi* par RNAi), et nous vérifierons que les siRNA n'ont pas les mêmes effets (nous devrons d'abord déterminer laquelle des deux enzymes Dicer de *Nematostella* produit les siRNA, ou si les deux y participent, de manière à réprimer spécifiquement la voie des siRNA sans atteindre celle des miRNA).

7.2 Qu'est-ce qu'une cible de microARN ?

7.2.1 Objectif

Le chapitre 6, en page 113, propose une nouvelle définition des cibles de miRNA : une cible ne serait pas simplement un ARN_m réprimé par un miRNA, ce serait un ARN_m réprimé *suffisamment* par le miRNA, pour que cette régulation ait un effet physiologique. Les autres ARN_m réprimés (insuffisamment) par le miRNA pourraient alors être des pseudo-cibles, dont la conservation phylogénétique de l'interaction avec le miRNA signifierait simplement que leur fonction d'inhibiteur compétitif apporte un avantage sélectif, en modulant l'activité du miRNA.

Cette hypothèse implique plusieurs prédictions, que le dogme actuel ne prévoit pas. Nous allons donc éprouver l'hypothèse, en testant expérimentalement chacune des prédictions suivantes :

1. Les ARN_m sont en compétition pour s'apparier avec les miRNA (les miRNA ne sont donc pas en large excès devant leurs cibles).
2. Les pseudo-cibles les moins abondantes ont un effet modulateur plus faible que les pseudo-cibles les plus abondantes, et leur interaction avec les miRNA devrait donc être moins conservée dans l'évolution.
3. L'interaction des cibles authentiques (dont l'activité est suffisamment sensible à une répression d'un facteur $\approx 1,1$ à 2, guidée par les miRNA) avec les miRNA devrait être plus conservée que l'interaction des pseudo-cibles avec le miRNA (les pseudo-cibles peuvent être remplacées par d'autres ARN_m de patron d'expression similaire).
4. En revanche, l'abondance cumulée des pseudo-cibles (pondérée par leur affinité pour le miRNA) devrait être conservée, tissu par tissu : la grandeur soumise à une pression de sélection serait l'affinité globale du transcriptome pour un miRNA, et pas l'identité individuelle des pseudo-cibles.
5. La répression des pseudo-cibles, guidée par le miRNA, est tamponnée par des mécanismes d'homéostasie et par la robustesse intrinsèque des systèmes intégrés face aux fluctuations. Ses effets à l'échelle physiologique ne devraient donc pas excéder les variations inter-individus dans une population sauvage.
6. Les pseudo-cibles n'ont pas besoin d'être codantes, ni réprimées, pour remplir leur rôle de modulateurs de miRNA.

7.2.2 Démarche expérimentale

Nous allons tester la première prédiction de deux manières :

- En perturbant l'éventuelle compétition des ARN_m pour un miRNA, et en mesurant l'effet de cette perturbation sur la répression d'un rapporteur. Nous utiliserons ici des cellules S2 de Drosophile, stablement transfectées avec un gène GFP sous le contrôle du promoteur de l'ubiquitine, et de la 3' UTR du gène *hid* de Drosophile (*hid* est réprimé par le miRNA *bantam* chez la Drosophile : Brennecke *et al.*, 2003 ; une lignée cellulaire contrôle est stablement transfectée avec le même rapporteur, mais avec une 3' UTR insensible à *bantam*). Ces deux lignées ont été préparées par Klaus Förstemann (Université de Munich). Nous allons réprimer par RNAi les neuf cibles prédites pour *bantam* qui sont les plus fortement exprimées dans les cellules S2⁴ : l'hypothèse prédit que cette répression libérera du miRNA

⁴Ces neuf gènes (par ordre d'expression croissante : CG10741 ; *jumeau* ; CG32486 ; CG11593 ; CG9674 ; *split ends* ; *enabled* ; *cornichon* ; *pannier*) sont, parmi les 74 cibles prédites de *bantam* par le programme TargetScan 5.0 (<http://www.targetscan.org/fly/>), les plus exprimées dans des cellules S2 traitées avec un ARN double-brin contre le gène GFP (qui n'a pas d'homologue dans le génome de Drosophile) : voir numéros d'accension

bantam, qui réprimerait donc davantage le rapporteur GFP. Nous mesurerons l'activité de la GFP cellule par cellule, par cytométrie de flux. Nous vérifierons que les cellules traitées par RNAi contre ces neuf gènes n'expriment pas plus *bantam* que des cellules contrôle (traitées avec des ARN double-brin étrangers au génome de *Drosophila*), et nous nous assurerons que le RNAi contre ces neuf gènes particuliers n'entraîne pas d'autres effets sur l'intensité de la fluorescence (la lignée contrôle, dont la séquence codante de la GFP est fusionnée à une 3' UTR insensible à *bantam*, sera également traitée avec les mêmes ARN double-brin).

- En mesurant la quantité de quelques miRNA et de leurs cibles prédites dans l'embryon précoce de *Drosophila* (de 0 à 2 h d'âge). Nous mesurerons l'abondance des miRNA miR-1, miR-3, et la famille miR-310/miR-311/miR-312, par Northern blot quantitatif (ces miRNA sont abondants dans l'embryon précoce : Aravin *et al.*, 2003, et le Northern blot quantitatif, calibré avec des oligonucléotides synthétiques, répond linéairement sur 5 ordres de grandeur). Nous mesurerons l'abondance des ARN_m par puce à ADN, calibrée avec des transcrits *in vitro* pour convertir les intensités de signal en nombres de moles. Ces mesures nous permettront de comparer les abondances des miRNA et de leurs cibles prédites *in vivo*, dans un échantillon biologique homogène ; le nombre de sites de complémentarité prédits sur ces ARN_m nous renseignera sur la stœchiométrie des interactions entre miRNA et ARN_m.

La deuxième prédiction, formulée en termes plus précis, signifierait que, dans les types cellulaires où un miRNA est le plus abondamment exprimé, ses pseudo-cibles les plus abondantes auront un plus grand effet sur l'activité du miRNA, et donc, leurs complémentarités au miRNA devraient être les mieux conservées. Il faudra donc travailler sur un type cellulaire homogène, dont le répertoire de miRNA et d'ARN_m est bien connu :

- les hépatocytes constituent environ 80 % du volume du foie de Rat (Kmieć, 2001), et ils expriment très abondamment le miRNA miR-122 (au moins chez la Souris : Lagos-Quintana *et al.*, 2002 ; Landgraf *et al.*, 2007) ;
- les fibres musculaires squelettiques représentent la presque totalité de la masse d'un muscle squelettique de Souris, et elles expriment abondamment les miRNA miR-1 et miR-133 (Chen *et al.*, 2006).

Nous identifierons les ARN_m qui présentent une complémentarité à la graine de ces miRNA, soit dans leur 3' UTR seulement ⁵, soit dans leur séquence complète (voir plus bas), et nous analyserons la corrélation entre la conservation de ces sites de complémentarité, et l'abondance de ces ARN_m dans le foie et les muscles squelettiques de Souris. Plusieurs analyses exhaustives des ARN_m par puce à ADN sont disponibles dans la littérature pour ces deux organes (voir par exemple Miller *et al.*, 2007 ; Rodriguez *et al.*, 2007 ; Li *et al.*, 2009b).

D'après la troisième prédiction, les gènes présentant les complémentarités les plus conservées aux graines de miRNA, auront fréquemment une activité très sensible à leur niveau d'expression. Chen et Rajewsky, 2006 ont identifié des gènes dont l'interaction avec des miRNA est conservée entre Protostomiens et Deutérostomiens : l'hypothèse prédit que ces gènes seront enrichis en cibles authentiques, et que donc, l'activité de la plupart de ces gènes sera très sensible à leur niveau d'expression. Nous évaluerons la sensibilité de l'activité de ces gènes de deux manières :

- Par transgénèse chez la *Drosophila* : nous introduirons une copie supplémentaire (et une seule) des gènes testés (sous le contrôle de leur promoteur endogène, pour mimer le patron

GSM50397, GSM50398 et GSM50399 dans la banque de données GEO (<http://www.ncbi.nlm.nih.gov/geo/> ; l'expérience qui a généré ces données est décrite dans Page *et al.*, 2005). Collectivement, les ARN_m de ces neuf gènes sont responsables de 64 % du signal des 74 cibles prédites, sur la puce à ADN : réprimer totalement ces neuf gènes par RNAi devrait donc diviser par trois la quantité d'ARN_m endogènes interagissant avec *bantam*.

⁵Une étude récente montre que les ARN_m de Souris qui portent une complémentarité à leur graine sont effectivement réprimés *in vivo* par ces miRNA : Kim et Bartel, 2009.

d'expression de la copie endogène) ; à l'état hémizygote, ces constructions devraient donc augmenter l'expression du gène d'environ 50 %, et le doubler à l'état homozygote, ce qui reproduirait quantitativement l'effet de la perte de la régulation de ce gène par le miRNA. Nous mesurerons alors les conséquences phénotypiques de cette surexpression modérée.

- Par la quantification de la variation inter-individus de l'expression de ces gènes, chez le Nématode et chez la Drosophile. Si effectivement l'activité de ces gènes est particulièrement sensible à leur niveau d'expression, alors l'expression de ces gènes devrait être plus précisément régulée que celle des autres gènes, dans des animaux sauvages. Nous mesurerons donc l'accumulation des ARN_m de Nématode et de Drosophile, sur des animaux entiers, mais sur un grand nombre d'individus indépendants (au moins dix animaux seront analysés séparément).

Pour tester la quatrième prédiction, nous utiliserons la banque de données CGAP ⁶ qui recense, organe par organe, les niveaux d'expression de plusieurs milliers de gènes chez la Souris et chez l'Homme. L'hypothèse prédit que la somme des abondances des ARN_m, multipliés par leur nombre de sites pour les miRNA, sera mieux corrélée entre Homme et Souris, que les expressions individuelles des ARN_m. Si un atlas d'expression devient disponible pour d'autres espèces, nous les analyserons de la même manière.

Idéalement, on testerait la cinquième prédiction en mutant des sites de complémentarité à des miRNA, dans quelques gènes de fonction bien documentée, et on mesurerait l'effet de cette mutation ponctuelle sur la fonction physiologique de ces gènes (pour le comparer aux fluctuations naturelles de l'activité de cette fonction physiologique dans une population sauvage). Une telle expérience impliquerait de multiples mutagenèses dirigées, elle serait donc très longue et coûteuse.

Une stratégie plus simple consisterait à utiliser des polymorphismes naturels dans les sites de reconnaissance des miRNA : nous allons donc rechercher les SNP (*single nucleotide polymorphisms*, polymorphismes d'un nucléotide unique) documentées qui détruisent des sites de complémentarité conservée à des miRNA. La banque de données de SNP du NCBI ⁷ recense (en juin 2009) près de 120 000 SNP dans les 3' UTR de gènes codants de Souris ; les SNP ne sont pas encore aussi bien décrites chez les autres organismes modèles, mais elles sont en cours de recensement (Chen *et al.*, 2008 ; Flibotte *et al.*, 2009).

Nous rechercherons donc les cibles prédites de miRNA (qui seront des pseudo-cibles candidates) chez l'un de ces trois organismes modèles, dont le site de complémentarité est perdu chez un variant connu. Nous le croiserons alors avec la souche de référence sur plusieurs générations, pour homogénéiser le fond génétique (ces croisements prendront beaucoup plus de temps chez la Souris, dont le temps de génération est quatre fois plus long que celui de la Drosophile, et vingt fois plus long que celui du Nématode ; il ne sera probablement pas possible de purifier moins d'une dizaine de centimorgans en un temps raisonnable : nous compenserons alors le manque de pureté génétique en analysant simultanément une vingtaine de souris issues de croisement indépendants). Nous comparerons ensuite le phénotype des souris homozygotes pour la perte du site de reconnaissance du miRNA, au phénotype des souris de la souche de référence.

D'après la sixième prédiction, des ARN non-codants pourraient jouer le rôle de pseudo-cibles (à la manière de l'ARN_m IPS1 chez *Arabidopsis thaliana* : Franco-Zorrilla *et al.*, 2007). Nous rechercherons donc, dans les banques d'EST et de cDNA, les transcrits non-codants qui portent des complémentarités conservées à la graine de miRNA. Nous vérifierons ensuite leur expression par Northern blot. La sixième prédiction implique aussi que des sites de complémentarité qui ne répriment pas l'ARN_m qui les porte, pourraient être conservées malgré tout, pour leur rôle

⁶<http://cgap.nci.nih.gov/>

⁷<http://www.ncbi.nlm.nih.gov/sites/entrez?db=snp&cmd=search&term=>

d'inhibiteurs de miRNA. De tels sites pourraient se trouver dans les séquences codantes des ARN_m (il semble que les sites localisés dans les séquences codantes répriment moins l' ARN_m que ceux qui se trouvent dans les 3' UTR : Grimson *et al.*, 2007). Les séquences codantes sont généralement globalement conservées, mais il est possible d'identifier des segments particulièrement conservés, dont même la troisième position des codons mute peu. Nous pourrions donc rechercher des complémentarités conservées dans les séquences codantes de Vertébrés (la multitude de génomes séquencés chez les Vertébrés facilitera l'analyse). Nous mesurerons leur effet sur la traduction de l' ARN_m par mutagenèse dirigée, en utilisant la redondance du code génétique pour générer des séquences qui codent des protéines identiques, mais dont la complémentarité à la graine du miRNA sera perdue (ces constructions seront étiquetées avec un épitope facilement détectable par Western blot, et transfectées dans des cellules en culture).

7.2.3 Expériences supplémentaires

Si les expériences décrites ci-dessus confirment les prédictions de l'hypothèse présentée dans le chapitre 6, alors cette hypothèse deviendra plus crédible que le dogme actuel (présenté dans la sous-section 1.3.1, page 19). Il faudra alors réviser profondément notre conception de la régulation guidée par les miRNA : de nombreuses cibles prédites par les programmes de prédiction (et qui sont souvent considérées comme des cibles confirmées, par les utilisateurs de ces programmes) pourraient être déclassées en pseudo-cibles. Il faudrait améliorer la prédiction des cibles, en prenant en compte :

- la sensibilité de l'activité de chaque gène à son niveau d'expression (à comparer à l'ampleur de la répression guidée par le miRNA) ;
- l'abondance de toutes les pseudo-cibles de ce miRNA dans le type cellulaire considéré, et leur affinité pour le miRNA.

L'identification des cibles de miRNA deviendrait donc beaucoup plus compliquée que la simple recherche de complémentarités conservées dans les 3' UTR : l'effet d'un miRNA sur l'activité biologique d'une cible dépendrait de réseaux de relations impliquant des centaines de gènes (les pseudo-cibles, et les partenaires du gène-cible, dont la nature des interactions détermine la sensibilité du phénomène physiologique aux variations d'expression du gène-cible). La modélisation quantitative de réseaux d'interaction complexes est l'objectif principal d'une discipline nouvelle, appelée « biologie des systèmes ». La régulation guidée par les miRNA pourrait devenir un de ses premiers champs d'application : la biologie des systèmes devrait nous apporter des informations sur la structure des réseaux d'interaction entre les gènes, qui nous aideront à identifier les gènes dont l'activité est la plus sensible aux variations d'expression. Au besoin, nous pourrions contribuer à cette analyse, en quantifiant les effets physiologiques de mutations dans les gènes de miRNA ou dans leurs sites de complémentarité dans des cibles prédites.

L'abondance des pseudo-cibles et leur affinité pour le miRNA devraient être plus faciles à mesurer. Ces informations nous permettront de modéliser l'interaction entre le miRNA et les ARN_m , à partir de données de transcriptomique de la littérature (ou : produites par nous-mêmes, si les données disponibles ne suffisent pas). La quantification précise de l'accumulation des miRNA, et la prise en compte de tous les ARN_m compétiteurs, devraient nous permettre de prédire l'étendue de la répression guidée par le miRNA sur un gène donné. Cette information devrait être utile aux nombreux utilisateurs des programmes de prédiction actuels : l'information qu'un gène est réprimé par un miRNA est de peu d'utilité, si on ne peut pas prédire dans quelle mesure le gène est réprimé. La prise en compte des aspects quantitatifs de la répression guidée par les miRNA devrait donc nous permettre d'améliorer les prédictions de cibles — et ultimement, la biologie des systèmes devrait nous permettre de prédire les conséquences physiologiques de ces phénomènes microscopiques.

7.3 Régulations post-transcriptionnelles de l'expression des microARN

7.3.1 Objectif

L'expression des miRNA eux-mêmes est régulée : des sites de fixation pour des facteurs de transcription ont été identifiés dans les promoteurs des gènes de plusieurs miRNA (voir par exemple O'Donnell *et al.*, 2005 ; Zhao *et al.*, 2005 ; Chang *et al.*, 2008). Des régulations post-transcriptionnelles de leur expression sont aussi connues :

- Chez le Nématode, l'abondance du pre-miRNA de miR-38 est à peu près constante au cours du développement, mais le miRNA mature n'est détecté que dans l'embryon (Ambros *et al.*, 2003). La maturation du pre-miRNA par Dicer semble donc régulée au cours du développement.
- Chez la Souris, l'abondance du pre-miRNA de miR-138 est similaire dans plusieurs tissus et une lignée cellulaire, mais le miRNA mature n'est détecté que dans le cerveau (Obernosterer *et al.*, 2006). Ces différences sont dues à un répresseur (absent dans le cerveau, mais exprimé au moins dans la lignée cellulaire), spécifique de ce pre-miRNA.
- Chez la Souris, le rapport entre les abondances du miRNA et du miRNA* est variable, de tissu en tissu, pour plusieurs miRNA (Ro *et al.*, 2007).
- Chez la Souris, l'abondance du pri-miRNA et du pre-miRNA de *let-7g* sont à peu près constantes au cours de la différenciation de cellules-souches embryonnaires (cellules ES, *embryonic stem cells*), alors que le miRNA mature est beaucoup plus abondant à partir du dixième jour de différenciation (Viswanathan *et al.*, 2008). La protéine Lin28 inhibe le clivage des pri-miRNA de la famille *let-7* par le Microprocessor, et son expression est réprimée au cours de la différenciation des cellules ES : elle pourrait donc contribuer à la régulation des miRNA *let-7* (les auteurs n'ont apparemment pas mesuré l'effet de Lin28 sur le clivage du pre-miRNA en duplex miRNA/miRNA*, par Dicer ; la constance apparente des abondances du pri-miRNA et du pre-miRNA au cours de la différenciation plaident plutôt pour une régulation du clivage par Dicer). Cet effet est spécifique des miRNA de la famille *let-7* : Lin28 n'affecte pas la maturation de plusieurs miRNA contrôle. Le même groupe a ensuite montré que Lin28 reconnaissait la boucle apicale du pre-miRNA (puisque la séquence du pre-miRNA est incluse dans celle du pri-miRNA, les auteurs en déduisent que Lin28 interagit de la même manière avec le pri-miRNA, substrat de Drosha) (Piskounova *et al.*, 2008).
- Ces résultats sont confirmés par Newman *et al.*, 2008, qui montrent également que Lin28 se lie à la boucle apicale de la tige-boucle des pri-miRNA de la famille *let-7*.
- Wulczyn *et al.*, 2007 avaient quant à eux montré que la maturation des miRNA de la famille *let-7* est régulée à l'étape du clivage du pre-miRNA par Dicer, au cours de la différenciation des cellules ES de Souris. En recherchant le facteur responsable de cette régulation, ils ont montré que Lin28 était un candidat crédible : en se fixant sur la boucle apicale du pre-miRNA, il empêcherait Dicer de la cliver (Rybak *et al.*, 2008). Lin28 serait donc un répresseur des clivages par Drosha et par Dicer, des précurseurs de la famille *let-7*.
- Heo *et al.*, 2008 montrent également que Lin28 réprime la maturation des pre-*let-7* par Dicer (dans des cellules humaines en culture ; ils montrent aussi que le clivage par Drosha et l'export nucléo-cytoplasmique du pre-*let-7* ne sont pas affectés par Lin28, ce qui suggère que, dans ce système, Lin28 ne régule que la maturation du pre-*let-7* par Dicer), mais ils proposent un autre mécanisme : Lin28 provoque la poly-uridylation de l'extrémité 3' de pre-*let-7*, et cet ARN uridylylé (détectable expérimentalement), incapable d'être clivé par Dicer, est dégradé.

7.3.2 Démarche expérimentale

Toutes ces études ont été initiées par la découverte accidentelle de décorrélatons entre l'abondance de miRNA matures et celle de leurs précurseurs ou de leur miRNA*. Les techniques modernes d'analyse permettent désormais de les rechercher de manière active : il est possible de quantifier simultanément tous les miRNA et tous les miRNA*, et de quantifier simultanément tous leurs précurseurs ; en mesurant l'abondance de ces ARN dans une variété de tissus et de lignées cellulaires de Souris, par séquençage à haut débit et par hybridation sur des puces à ADN, nous serons en mesure d'identifier les distorsions dans les rapports entre miRNA et pre-miRNA, entre pre-miRNA et pri-miRNA, et entre miRNA et miRNA*, de tissu à tissu. Ces distorsions signaleront l'existence de régulations différentielles de la maturation des miRNA ; en regroupant les miRNA en fonction de leur comportement (par une analyse de « clustering » dont les deux variables seront l'identité du miRNA, et l'identité de l'échantillon biologique), nous définirons les familles de miRNA soumises aux mêmes régulations. Nous identifierons ensuite les motifs communs portés par ces précurseurs, responsables de ces régulations, par une recherche d'éléments de séquence ou de structure secondaire communs (nous utiliserons une recherche de modèles de Markov cachés, plus sensible que les autres algorithmes de recherche d'alignements, et facilement utilisable pour les recherches de motifs de structure secondaire).

La fonctionnalité des éléments identifiés sera ensuite vérifiée par transfection dans des cellules de Souris en culture (la maturation d'un pri-miRNA sauvage sera comparée à celle d'un pri-miRNA muté sur ces éléments), ou (si nous ne trouvons pas de système cellulaire qui reproduit ces régulations) par transgénèse chez la Souris.

Les motifs régulateurs confirmés seront ensuite utilisés comme appâts pour la purification par affinité des facteurs *trans* responsables de ces régulations, à partir de lysats cellulaires (lorsqu'un système cellulaire reproduit ces régulations) ou de broyats de tissus (dans les autres cas) ; les protéines purifiées seront identifiées par spectrométrie de masse, et leur rôle régulateur sera vérifié par des expériences de répression par des siRNA dans des cellules en culture (éventuellement en exprimant ces protéines à partir de plasmides transfectés, si elles ne sont pas naturellement exprimés dans les lignées cellulaires disponibles).

Bibliographie

- Alemán, L. M., Doench, J., et Sharp, P. A. (2007). Comparison of siRNA-induced off-target RNA and protein effects. *RNA*, 13(3) : 385–395.
- Ambros, V., Lee, R. C., Lavanway, A., Williams, P. T., et Jewell, D. (2003). MicroRNAs and other tiny endogenous RNAs in *C. elegans*. *Curr Biol*, 13(10) : 807–818.
- Aoki, K., Moriguchi, H., Yoshioka, T., Okawa, K., et Tabara, H. (2007). In vitro analyses of the production and activity of secondary small interfering RNAs in *C. elegans*. *EMBO J*, 26(24) : 5007–5019.
- Aravin, A., Gaidatzis, D., Pfeffer, S., Lagos-Quintana, M., Landgraf, P., Iovino, N., Morris, P., Brownstein, M. J., Kuramochi-Miyagawa, S., Nakano, T., Chien, M., Russo, J. J., Ju, J., Sheridan, R., Sander, C., Zavolan, M., et Tuschl, T. (2006). A novel class of small RNAs bind to MILI protein in mouse testes. *Nature*, 442(7099) : 203–207.
- Aravin, A. A., Lagos-Quintana, M., Yalcin, A., Zavolan, M., Marks, D., Snyder, B., Gaasterland, T., Meyer, J., et Tuschl, T. (2003). The small RNA profile during *Drosophila melanogaster* development. *Dev Cell*, 5(2) : 337–350.
- Aravin, A. A., Naumova, N. M., Tulin, A. V., Vagin, V. V., Rozovsky, Y. M., et Gvozdev, V. A. (2001). Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the *D. melanogaster* germline. *Curr Biol*, 11(13) : 1017–1027.
- Aravin, A. A., Sachidanandam, R., Bourc’his, D., Schaefer, C., Pezic, D., Toth, K. F., Bestor, T., et Hannon, G. J. (2008). A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. *Mol Cell*, 31(6) : 785–799.
- Aravin, A. A., Sachidanandam, R., Girard, A., Fejes-Toth, K., et Hannon, G. J. (2007). Developmentally regulated piRNA clusters implicate MILI in transposon control. *Science*, 316(5825) : 744–747.
- Azuma-Mukai, A., Oguri, H., Mituyama, T., Qian, Z. R., Asai, K., Siomi, H., et Siomi, M. C. (2008). Characterization of endogenous human Argonautes and their miRNA partners in RNA silencing. *Proc Natl Acad Sci USA*, 105(23) : 7964–7969.
- Baek, D., Villén, J., Shin, C., Camargo, F. D., Gygi, S. P., et Bartel, D. P. (2008). The impact of microRNAs on protein output. *Nature*, 455(7209) : 64–71.
- Bagga, S., Bracht, J., Hunter, S., Massirer, K., Holtz, J., Eachus, R., et Pasquinelli, A. E. (2005). Regulation by *let-7* and *lin-4* miRNAs results in target mRNA degradation. *Cell*, 122(4) : 553–563.
- Bartel, D. P. (2004). MicroRNAs : genomics, biogenesis, mechanism, and function. *Cell*, 116(2) : 281–297.
- Batista, P. J., Ruby, J. G., Claycomb, J. M., Chiang, R., Fahlgren, N., Kasschau, K. D., Chaves, D. A., Gu, W., Vasale, J. J., Duan, S., Conte, D. J., Luo, S., Schroth, G. P., Carrington, J. C., Bartel, D. P., et Mello, C. C. (2008). PRG-1 and 21U-RNAs interact to form the piRNA complex required for fertility in *C. elegans*. *Mol Cell*, 31(1) : 67–78.

- Behm-Ansmant, I., Rehwinkel, J., Doerks, T., Stark, A., Bork, P., et Izaurralde, E. (2006). mRNA degradation by miRNAs and GW182 requires both CCR4 :NOT deadenylase and DCP1 :DCP2 decapping complexes. *Genes Dev*, 20(14) : 1885–1898.
- Beitzinger, M., Peters, L., Zhu, J. Y., Kremmer, E., et Meister, G. (2007). Identification of human microRNA targets from isolated Argonaute protein complexes. *RNA Biol*, 4(2) : 76–84.
- Bengtson, S. et Budd, G. (2004). Comment on "Small bilaterian fossils from 40 to 55 million years before the Cambrian". *Science*, 306(5700) : 1291.
- Benton, M. J. et Donoghue, P. C. (2007). Paleontological evidence to date the tree of life. *Mol Biol Evol*, 24(1) : 26–53.
- Berezikov, E., Chung, W. J., Willis, J., Cuppen, E., et Lai, E. C. (2007). Mammalian mirtron genes. *Mol Cell*, 28(2) : 328–336.
- Bernstein, E., Caudy, A. A., Hammond, S. M., et Hannon, G. J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature*, 409(6818) : 363–366.
- Bhattacharyya, S. N., Habermacher, R., Martine, U., Closs, E. I., et Filipowicz, W. (2006). Relief of microRNA-mediated translational repression in human cells subjected to stress. *Cell*, 125(6) : 1111–1124.
- Blumenstiel, J. P., Fu, R., Theurkauf, W. E., et Hawley, R. S. (2008). Components of the RNAi machinery that mediate long-distance chromosomal associations are dispensable for meiotic and early somatic homolog pairing in *Drosophila melanogaster*. *Genetics*, 180(3) : 1355–1365.
- Bohnsack, M. T., Czaplinski, K., et Gorlich, D. (2004). Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. *RNA*, 10(2) : 185–191.
- Borchert, G. M., Lanier, W., et Davidson, B. L. (2006). RNA polymerase III transcribes human microRNAs. *Nat Struct Mol Biol*, 13(12) : 1097–1101.
- Bortolin-Cavaillé, M.-L., Dance, M., Weber, M., et Cavaillé, J. (2009). C19MC microRNAs are processed from introns of large Pol-II, non-protein-coding transcripts. *Nucleic Acids Res*, 37(10) : 3464–3473.
- Brennecke, J., Aravin, A. A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., et Hannon, G. J. (2007). Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell*, 128(6) : 1089–1103.
- Brennecke, J., Hipfner, D. R., Stark, A., Russell, R. B., et Cohen, S. M. (2003). *bantam* encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene *hid* in *Drosophila*. *Cell*, 113(1) : 25–36.
- Brennecke, J., Malone, C. D., Aravin, A. A., Sachidanandam, R., Stark, A., et Hannon, G. J. (2008). An epigenetic role for maternally inherited piRNAs in transposon silencing. *Science*, 322(5906) : 1387–1392.
- Brower-Toland, B., Findley, S. D., Jiang, L., Liu, L., Yin, H., Dus, M., Zhou, P., Elgin, S. C., et Lin, H. (2007). *Drosophila* PIWI associates with chromatin and interacts directly with HP1a. *Genes Dev*, 21(18) : 2300–2311.
- Burton, P. M. et Finnerty, J. R. (2009). Conserved and novel gene expression between regeneration and asexual fission in *Nematostella vectensis*. *Dev Genes Evol*, 219(2) : 79–87.
- Busseau, I., Chaboissier, M. C., Péliisson, A., et Bucheton, A. (1994). I factors in *Drosophila melanogaster* : transposition under control. *Genetica*, 93((1-3) :101-16.Links) : 1 : Genetica.
- Buzgariu, W., Chera, S., et Galliot, B. (2008). Methods to investigate autophagy during starvation and regeneration in hydra. *Methods Enzymol*, 451 : 409–437.
- Carmell, M. A. et Hannon, G. J. (2004). RNase III enzymes and the initiation of gene silencing. *Nat Struct Mol Biol*, 11(3) : 214–218.

- Carmell, M. A., Xuan, Z., Zhang, M. Q., et Hannon, G. J. (2002). The Argonaute family : tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis. *Genes Dev*, 16(21) : 2733–2742.
- Chang, T. C., Yu, D., Lee, Y. S., Wentzel, E. A., Arking, D. E., West, K. M., Dang, C. V., Thomas-Tikhonenko, A., et Mendell, J. T. (2008). Widespread microRNA repression by Myc contributes to tumorigenesis. *Nat Genet*, 40(1) : 43–50.
- Chen, D., Ahlford, A., Schnorrer, F., Kalchhauser, I., Fellner, M., Viràgh, E., Kiss, I., Syvänen, A. C., et Dickson, B. J. (2008). High-resolution, high-throughput SNP mapping in *Drosophila melanogaster*. *Nat Methods*, 5(4) : 323–329.
- Chen, J. F., Mandel, E. M., Thomson, J. M., Wu, Q., Callis, T. E., Hammond, S. M., Conlon, F. L., et Wang, D. Z. (2006). The role of microRNA-1 and microRNA-133 in skeletal muscle proliferation and differentiation. *Nat Genet*, 38(2) : 228–233.
- Chen, J. Y., Bottjer, D. J., Oliveri, P., Dornbos, S. Q., Gao, F., Ruffins, S., Chi, H., Li, C. W., et Davidson, E. H. (2004). Small bilaterian fossils from 40 to 55 million years before the cambrian. *Science*, 305(5681) : 218–222.
- Chen, K. et Rajewsky, N. (2006). Deep conservation of microRNA-target relationships and 3' UTR motifs in vertebrates, flies, and nematodes. *Cold Spring Harb Symp Quant Biol*, 71 : 149–156.
- Chendrimada, T. P., Finn, K. J., Ji, X., Baillat, D., Gregory, R. I., Liebhaber, S. A., Pasquinelli, A. E., et Shiekhattar, R. (2007). MicroRNA silencing through RISC recruitment of eIF6. *Nature*, 447(7146) : 823–828.
- Chendrimada, T. P., Gregory, R. I., Kumaraswamy, E., Norman, J., Cooch, N., Nishikura, K., et Shiekhattar, R. (2005). TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. *Nature*, 436(7051) : 740–744.
- Chiu, Y. L. et Rana, T. M. (2003). siRNA function in RNAi : a chemical modification analysis. *RNA*, 9(9) : 1034–1048.
- Chung, W. J., Okamura, K., Martin, R., et Lai, E. C. (2008). Endogenous RNA interference provides a somatic defense against *Drosophila* transposons. *Curr Biol*, 18(11) : 795–802.
- Cogoni, C. et Macino, G. (2000). Post-transcriptional gene silencing across kingdoms. *Curr Opin Genet Dev*, 10(6) : 638–643.
- Crooks, G. E., Hon, G., Chandonia, J. M., et Brenner, S. E. (2004). WebLogo : a sequence logo generator. *Genome Res*, 14(6) : 1188–1190.
- Csirik, A. K., Linsk, R., et Birchler, J. A. (1994). The *Lighten up (Lip)* gene of *Drosophila melanogaster*, a modifier of retroelement expression, position effect variegation and *white* locus insertion alleles. *Genetics*, 138(1) : 153–163.
- Czech, B., Malone, C. D., Zhou, R., Stark, A., Schlingeheyde, C., Dus, M., Perrimon, N., Kellis, M., Wohlschlegel, J. A., Sachidanandam, R., Hannon, G. J., et Brennecke, J. (2008). An endogenous small interfering RNA pathway in *Drosophila*. *Nature*, 453(7196) : 798–802.
- Darling, J. A., Reitzel, A. R., Burton, P. M., Mazza, M. E., Ryan, J. F., Sullivan, J. C., et Finnerty, J. R. (2005). Rising starlet : the starlet sea anemone, *Nematostella vectensis*. *Bioessays*, 27(2) : 211–221.
- Das, P. P., Bagijn, M. P., Goldstein, L. D., Woolford, J. R., Lehrbach, N. J., Sapetschnig, A., Buhecha, H. R., Gilchrist, M. J., Howe, K. L., Stark, R., Matthews, N., Berezikov, E., Ketting, R. F., Tavaré, S., et Miska, E. A. (2008). Piwi and piRNAs act upstream of an endogenous siRNA pathway to suppress Tc3 transposon mobility in the *Caenorhabditis elegans* germline. *Mol Cell*, 31(1) : 79–90.

- Davis, E., Caiment, F., Tordoir, X., Cavaillé, J., Ferguson-Smith, A., Cockett, N., Georges, M., et Charlier, C. (2005). RNAi-mediated allelic trans-interaction at the imprinted Rtl1/Peg11 locus. *Curr Biol*, 15(8) : 743–749.
- Denli, A. M., Tops, B. B., Plasterk, R. H., Ketting, R. F., et Hannon, G. J. (2004). Processing of primary microRNAs by the Microprocessor complex. *Nature*, 432(7014) : 231–235.
- Desset, S., Buchon, N., Meignin, C., Coiffet, M., et Vauray, C. (2008). In *Drosophila melanogaster* the COM locus directs the somatic silencing of two retrotransposons through both Piwi-dependent and -independent pathways. *PLoS One*, 3(2) : e1526.
- Didiano, D. et Hobert, O. (2008). Molecular architecture of a miRNA-regulated 3' UTR. *RNA*, 14(7) : 1297–1317.
- Djupedal, I. et Ekwall, K. (2009). Epigenetics : heterochromatin meets RNAi. *Cell Res*, 19(3) : 282–295.
- Douzery, E. J., Snell, E. A., Baptiste, E., Delsuc, F., et Philippe, H. (2004). The timing of eukaryotic evolution : does a relaxed molecular clock reconcile proteins and fossils? *Proc Natl Acad Sci USA*, 101(43) : 15386–15391.
- Duursma, A. M., Kedde, M., Schrier, M., le Sage, C., et Agami, R. (2008). miR-148 targets human DNMT3b protein coding region. *RNA*, 14(5) : 872–877.
- Easow, G., Teleman, A. A., et Cohen, S. M. (2007). Isolation of microRNA targets by miRNP immunopurification. *RNA*, 13(8) : 1198–1204.
- Elbashir, S. M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K., et Tuschl, T. (2001a). Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature*, 411(6836) : 494–498.
- Elbashir, S. M., Lendeckel, W., et Tuschl, T. (2001b). RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes Dev*, 15(2) : 188–200.
- Elbashir, S. M., Martinez, J., Patkaniowska, A., Lendeckel, W., et Tuschl, T. (2001c). Functional anatomy of siRNAs for mediating efficient RNAi in *Drosophila melanogaster* embryo lysate. *EMBO J*, 20(23) : 6877–6888.
- Enright, A. J., John, B., Gaul, U., Tuschl, T., Sander, C., et Marks, D. S. (2003). MicroRNA targets in *Drosophila*. *Genome Biol*, 5(1) : R1.
- Eulalio, A., Huntzinger, E., et Izaurralde, E. (2008). GW182 interaction with Argonaute is essential for miRNA-mediated translational repression and mRNA decay. *Nat Struct Mol Biol*, 15(4) : 346–353.
- Fire, A. (1999). RNA-triggered gene silencing. *Trends Genet*, 15(9) : 358–363.
- Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., et Mello, C. C. (1998). Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature*, 391(6669) : 806–811.
- Flibotte, S., Edgley, M. L., Maydan, J., Taylor, J., Zapf, R., Waterston, R., et Moerman, D. G. (2009). Rapid high resolution single nucleotide polymorphism-comparative genome hybridization mapping in *Caenorhabditis elegans*. *Genetics*, 181(1) : 33–37.
- Forman, J. J., Legesse-Miller, A., et Collier, H. A. (2008). A search for conserved sequences in coding regions reveals that the *let-7* microRNA targets Dicer within its coding sequence. *Proc Natl Acad Sci USA*, 105(39) : 14879–14884.
- Förstemann, K., Horwich, M. D., Wee, L., Tomari, Y., et Zamore, P. D. (2007). *Drosophila* microRNAs are sorted into functionally distinct Argonaute complexes after production by Dicer-1. *Cell*, 130(2) : 287–297.

- Förstemann, K., Tomari, Y., Du, T., Vagin, V. V., Denli, A. M., Bratu, D. P., Klattenhoff, C., Theurkauf, W. E., et Zamore, P. D. (2005). Normal microRNA maturation and germ-line stem cell maintenance requires Loquacious, a double-stranded RNA-binding domain protein. *PLoS Biol*, 3(7) : e236.
- Franco-Zorrilla, J. M., Valli, A., Todesco, M., Mateos, I., Puga, M. I., Rubio-Somoza, I., Leyva, A., Weigel, D., García, J. A., et Paz-Ares, J. (2007). Target mimicry provides a new mechanism for regulation of microRNA activity. *Nat Genet*, 39(8) : 1033–1037.
- Frank-Kamenetsky, M., Grefhorst, A., Anderson, N. N., Racie, T. S., Bramlage, B., Akinc, A., Butler, D., Charisse, K., Dorkin, R., Fan, Y., Gamba-Vitalo, C., Hadwiger, P., Jayaraman, M., John, M., Jayaprakash, K. N., Maier, M., Nechev, L., Rajeev, K. G., Read, T., Röhl, I., Soutschek, J., Tan, P., Wong, J., Wang, G., Zimmermann, T., de, Fougerolles, A., Vornlocher, H. P., Langer, R., Anderson, D. G., Manoharan, M., Koteliansky, V., Horton, J. D., et Fitzgerald, K. (2008). Therapeutic RNAi targeting PCSK9 acutely lowers plasma cholesterol in rodents and LDL cholesterol in nonhuman primates. *Proc Natl Acad Sci USA*, 105(33) : 11915–11920.
- Friedman, R. C., Farh, K. K., Burge, C. B., et Bartel, D. P. (2009). Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res*, 19(1) : 92–105.
- Fritzenwanker, J. H. et Technau, U. (2002). Induction of gametogenesis in the basal cnidarian *Nematostella vectensis* (Anthozoa). *Dev Genes Evol*, 212(2) : 99–103.
- Galiana-Arnoux, D., Dostert, C., Schneemann, A., Hoffmann, J. A., et Imler, J.-L. (2006). Essential function *in vivo* for Dicer-2 in host defense against RNA viruses in drosophila. *Nat Immunol*, 7(6) : 590–597.
- Ghildiyal, M., Seitz, H., Horwich, M. D., Li, C., Du, T., Lee, S., Xu, J., Kittler, E. L., Zapp, M. L., Weng, Z., et Zamore, P. D. (2008). Endogenous siRNAs derived from transposons and mRNAs in *Drosophila* somatic cells. *Science*, 320(5879) : 1077–1081.
- Giraldez, A. J., Mishima, Y., Rihel, J., Grocock, R. J., Van Dongen, S., Inoue, K., Enright, A. J., et Schier, A. F. (2006). Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. *Science*, 312(5770) : 75–79.
- Girard, A. et Hannon, G. J. (2008). Conserved themes in small-RNA-mediated transposon control. *Trends Cell Biol*, 18(3) : 136–148.
- Girard, A., Sachidanandam, R., Hannon, G. J., et Carmell, M. A. (2006). A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature*, 442(7099) : 199–202.
- Gregory, R. I., Yan, K. P., Amuthan, G., Chendrimada, T., Doratotaj, B., Cooch, N., et Shiekhattar, R. (2004). The Microprocessor complex mediates the genesis of microRNAs. *Nature*, 432(7014) : 235–240.
- Grimson, A., Farh, K. K., Johnston, W. K., Garrett-Engele, P., Lim, L. P., et Bartel, D. P. (2007). MicroRNA targeting specificity in mammals : determinants beyond seed pairing. *Mol Cell*, 27(1) : 91–105.
- Grimson, A., Srivastava, M., Fahey, B., Woodcroft, B. J., Chiang, H. R., King, N., Degan, B. M., Rokhsar, D. S., et Bartel, D. P. (2008). Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. *Nature*, 455(7217) : 1193–1197.
- Grishok, A., Pasquinelli, A. E., Conte, D., Li, N., Parrish, S., Ha, I., Baillie, D. L., Fire, A., Ruvkun, G., et Mello, C. C. (2001). Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control *C. elegans* developmental timing. *Cell*, 106(1) : 23–34.
- Grivna, S. T., Beyret, E., Wang, Z., et Lin, H. (2006). A novel class of small RNAs in mouse spermatogenic cells. *Genes Dev*, 20(13) : 1709–1714.

- Gunawardane, L. S., Saito, K., Nishida, K. M., Miyoshi, K., Kawamura, Y., Nagami, T., Siomi, H., et Siomi, M. C. (2007). A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in *Drosophila*. *Science*, 315(5818) : 1587–1590.
- Haase, A. D., Jaskiewicz, L., Zhang, H., Lainé, S., Sack, R., Gatignol, A., et Filipowicz, W. (2005). TRBP, a regulator of cellular PKR and HIV-1 virus expression, interacts with Dicer and functions in RNA silencing. *EMBO Rep*, 6(10) : 961–967.
- Habig, J. W., Aruscavage, P. J., et Bass, B. L. (2008). In *C. elegans*, high levels of dsRNA allow RNAi in the absence of RDE-4. *PLoS ONE*, 3(12) : e4052.
- Haley, B. et Zamore, P. D. (2004). Kinetic analysis of the RNAi enzyme complex. *Nat Struct Mol Biol*, 11(7) : 599–606.
- Han, J., Lee, Y., Yeom, K. H., Kim, Y. K., Jin, H., et Kim, V. N. (2004). The Drosha-DGCR8 complex in primary microRNA processing. *Genes Dev*, 18(24) : 3016–3027.
- Han, J., Lee, Y., Yeom, K. H., Nam, J. W., Heo, I., Rhee, J. K., Sohn, S. Y., Cho, Y., Zhang, B. T., et Kim, V. N. (2006). Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. *Cell*, 125(5) : 887–901.
- Hand, C. et Uhlinger, K. R. (1992). The culture, sexual and asexual reproduction, and growth of the sea anemone *Nematostella vectensis*. *Biol Bull*, 182 : 169–176.
- Hand, C. et Uhlinger, K. R. (1994). The unique, widely distributed, estuarine Sea Anemone, *Nematostella vectensis* Stephenson : a review, new facts, and questions. *Estuaries*, 17(2) : 501–508.
- Hannon, G. J. et Conklin, D. S. (2004). RNA interference by short hairpin RNAs expressed in vertebrate cells. *Methods Mol Biol*, 257 : 255–266.
- Haynes, K. A., Caudy, A. A., Collins, L., et Elgin, S. C. (2006). Element 1360 and RNAi components contribute to HP1-dependent silencing of a pericentric reporter. *Curr Biol*, 16(22) : 2222–2227.
- Hedges, S. B., Blair, J. E., Venturi, M. L., et Shoe, J. L. (2004). A molecular timescale of eukaryote evolution and the rise of complex multicellular life. *BMC Evol Biol*, 4 : 2.
- Hendrickson, D. G., Hogan, D. J., Herschlag, D., Ferrell, J. E., et Brown, P. O. (2008). Systematic identification of mRNAs recruited to Argonaute 2 by specific microRNAs and corresponding changes in transcript abundance. *PLoS One*, 3(5) : e2126.
- Heo, I., Joo, C., Cho, J., Ha, M., Han, J., et Kim, V. N. (2008). Lin28 mediates the terminal uridylation of let-7 precursor MicroRNA. *Mol Cell*, 32(2) : 276–284.
- Horwich, M. D., Li, C., Matranga, C., Vagin, V., Farley, G., Wang, P., et Zamore, P. D. (2007). The *Drosophila* RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC. *Curr Biol*, 17(14) : 1265–1272.
- Houwing, S., Kamminga, L. M., Berezikov, E., Cronembold, D., Girard, A., van den Elst, H., Philippov, D. V., Blaser, H., Raz, E., Moens, C. B., Plasterk, R. H., Hannon, G. J., Draper, B. W., et Ketting, R. F. (2007). A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish. *Cell*, 129(1) : 69–82.
- Humphreys, D. T., Westman, B. J., Martin, D. I., et Preiss, T. (2005). MicroRNAs control translation initiation by inhibiting eukaryotic initiation factor 4E/cap and poly(A) tail function. *Proc Natl Acad Sci USA*, 102(47) : 16961–16966.
- Hutvagner, G., McLachlan, J., Pasquinelli, A. E., Balint, E., Tuschl, T., et Zamore, P. D. (2001). A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science*, 293(5531) : 834–838.
- Hutvagner, G. et Zamore, P. D. (2002). A microRNA in a multiple-turnover RNAi enzyme complex. *Science*, 297(5589) : 2056–2060.

- Ishizuka, A., Siomi, M. C., et Siomi, H. (2002). A *Drosophila* fragile X protein interacts with components of RNAi and ribosomal proteins. *Genes Dev*, 16(19) : 2497–2508.
- Iwasaki, S., Kawamata, T., et Tomari, Y. (2009). *Drosophila* Argonaute1 and Argonaute2 employ distinct mechanisms for translational repression. *Mol Cell*, 34(1) : 58–67.
- Jannot, G., Boisvert, M. E., Banville, I. H., et Simard, M. J. (2008). Two molecular features contribute to the Argonaute specificity for the microRNA and RNAi pathways in *C. elegans*. *RNA*, 14(5) : 829–835.
- Jiang, F., Ye, X., Liu, X., Fincher, L., McKearin, D., et Liu, Q. (2005). Dicer-1 and R3D1-L catalyze microRNA maturation in *Drosophila*. *Genes Dev*, 19(14) : 1674–1679.
- John, B., Enright, A. J., Aravin, A., Tuschl, T., Sander, C., et Marks, D. S. (2004). Human MicroRNA targets. *PLoS Biol*, 2(11) : e363.
- Karginov, F. V., Conaco, C., Xuan, Z., Schmidt, B. H., Parker, J. S., Mandel, G., et Hannon, G. J. (2007). A biochemical approach to identifying microRNA targets. *Proc Natl Acad Sci USA*, 104(49) : 19291–19296.
- Kawamata, T., Seitz, H., et Tomari, Y. (2009). Structural requirements of miRNAs for RISC-loading and slicer-independent unwinding. *Nat Struct Mol Biol*. Sous presse.
- Kawamura, Y., Saito, K., Kin, T., Ono, Y., Asai, K., Sunohara, T., Okada, T. N., Siomi, M. C., et Siomi, H. (2008). *Drosophila* endogenous small RNAs bind to Argonaute 2 in somatic cells. *Nature*, 453(7196) : 793–797.
- Kawaoka, S., Hayashi, N., Suzuki, Y., Abe, H., Sugano, S., Tomari, Y., Shimada, T., et Katsuma, S. (2009). The Bombyx ovary-derived cell line endogenously expresses PIWI/PIWI-interacting RNA complexes. *RNA*, 15(7) : 1258–1264.
- Kawasaki, H. et Taira, K. (2004). Induction of DNA methylation and gene silencing by short interfering RNAs in human cells. *Nature*, 431(7005) : 211–217.
- Kedde, M., Strasser, M. J., Boldajipour, B., Oude, Vrieling, J. A., Slanchev, K., le Sage, C., Nagel, R., Voorhoeve, P. M., van Duijse, J., Orom, U. A., Lund, A. H., Perrakis, A., Raz, E., et Agami, R. (2007). RNA-binding protein Dnd1 inhibits microRNA access to target mRNA. *Cell*, 131(7) : 1273–1286.
- Kertesz, M., Iovino, N., Unnerstall, U., Gaul, U., et Segal, E. (2007). The role of site accessibility in microRNA target recognition. *Nat Genet*, 39(10) : 1278–1284.
- Ketting, R. F., Fischer, S. E. J., Bernstein, E., Sijen, T., Hannon, G. J., et Plasterk, R. H. A. (2001). Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in *C. elegans*. *Genes Dev*, 15(20) : 2654–2659.
- Ketting, R. F., Haverkamp, T. H., van, Luenen, H. G., et Plasterk, R. H. (1999). *mut-7* of *C. elegans*, required for transposon silencing and RNA interference, is a homolog of Werner syndrome helicase and RNaseD. *Cell*, 99(2) : 133–141.
- Kim, A., Terzian, C., Santamaria, P., Pélisson, A., Purd’homme, N., et Bucheton, A. (1994). Retroviruses in invertebrates : the *gypsy* retrotransposon is apparently an infectious retrovirus of *Drosophila melanogaster*. *Proc Natl Acad Sci USA*, 91(4) : 1285–1289.
- Kim, D. H., Villeneuve, L. M., Morris, K. V., et Rossi, J. J. (2006). Argonaute-1 directs siRNA-mediated transcriptional gene silencing in human cells. *Nat Struct Mol Biol*, 13(9) : 793–797.
- Kim, J. et Bartel, D. P. (2009). Allelic imbalance sequencing reveals that single-nucleotide polymorphisms frequently alter microRNA-directed repression. *Nat Biotechnol*, 27(5) : 472–477.
- Kim, Y. K. et Kim, V. N. (2007). Processing of intronic microRNAs. *EMBO J*, 26(3) : 775–783.

- Kinch, L. N. et Grishin, N. V. (2009). The human Ago2 MC region does not contain an eIF4E-like mRNA cap binding motif. *Biol Direct*, 4 : 2.
- Kiriakidou, M., Nelson, P. T., Kouranov, A., Fitziev, P., Bouyioukos, C., Mourelatos, Z., et Hatzigeorgiou, A. (2004). A combined computational-experimental approach predicts human microRNA targets. *Genes Dev*, 18(10) : 1165–1178.
- Kiriakidou, M., Tan, G. S., Lamprinaki, S., De, Planell-Saguer, M., Nelson, P. T., et Mourelatos, Z. (2007). An mRNA m7G cap binding-like motif within human Ago2 represses translation. *Cell*, 129(6) : 1141–1151.
- Kirino, Y. et Mourelatos, Z. (2007). The mouse homolog of HEN1 is a potential methylase for Piwi-interacting RNAs. *RNA*, 13(9) : 1397–1401.
- Klattenhoff, C. et Theurkauf, W. (2008). Biogenesis and germline functions of piRNAs. *Development*, 135(1) : 3–9.
- Klattenhoff, C., Xi, H., Li, C., Lee, S., Xu, J., Schultz, N., Koppetsch, B. S., Nowosielska, A., Seitz, H., Zamore, P. D., Weng, Z., et Theurkauf, W. E. (2009). The *Drosophila* HP1 homolog Rhino is required for transposon silencing and piRNA production by dual strand clusters. *Cell*. Sous presse.
- Kloosterman, W. P., Wienholds, E., Ketting, R. F., et Plasterk, R. H. (2004). Substrate requirements for *let-7* function in the developing zebrafish embryo. *Nucleic Acids Res*, 32(21) : 6284–6291.
- Kmieć, Z. (2001). *Cooperation of liver cells in health and disease.*, chapitre Introduction — Morphology of the liver lobule, page 5. Number 161 in *Adv Anat Embryol Cell Biol*. Springer.
- Knight, S. W. et Bass, B. L. (2001). A role for the RNase III enzyme DCR-1 in RNA interference and germ line development in *Caenorhabditis elegans*. *Science*, 293(5538) : 2269–2271.
- Krek, A., Grün, D., Poy, M. N., Wolf, R., Rosenberg, L., Epstein, E. J., MacMenamin, P., da Piedade, I., Gunsalus, K. C., Stoffel, M., et Rajewsky, N. (2005). Combinatorial microRNA target predictions. *Nat Genet*, 37(5) : 495–500.
- Lagos-Quintana, M., Rauhut, R., Yalcin, A., Meyer, J., Lendeckel, W., et Tuschl, T. (2002). Identification of tissue-specific microRNAs from mouse. *Curr Biol*, 12(9) : 735–739.
- Lai, E. C. (2002). MicroRNAs are complementary to 3' UTR sequence motifs that mediate negative post-transcriptional regulation. *Nat Genet*, 30 : 363–364.
- Landgraf, P., Rusu, M., Sheridan, R., Sewer, A., Iovino, N., Aravin, A., Pfeffer, S., Rice, A., Kamphorst, A. O., Landthaler, M., Lin, C., Socci, N. D., Hermida, L., Fulci, V., Chiaretti, S., Foà, R., Schliwka, J., Fuchs, U., Novosel, A., Müller, R. U., Schermer, B., Bissels, U., Inman, J., Phan, Q., Chien, M., Weir, D. B., Choksi, R., De Vita, G., Frezzetti, D., Trompeter, H. I., Hornung, V., Teng, G., Hartmann, G., Palkovits, M., Di Lauro, R., Wernet, P., Macino, G., Rogler, C. E., Nagle, J. W., Ju, J., Papavasiliou, F. N., Benzing, T., Lichter, P., Tam, W., Brownstein, M. J., Bosio, A., Borkhardt, A., Russo, J. J., Sander, C., Zavolan, M., et Tuschl, T. (2007). A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell*, 129(7) : 1401–1414.
- Landthaler, M., Gaidatzis, D., Rothballer, A., Chen, P. Y., Soll, S. J., Dinic, L., Ojo, T., Hafner, M., Zavolan, M., et Tuschl, T. (2008). Molecular characterization of human Argonaute-containing ribonucleoprotein complexes and their bound target mRNAs. *RNA*, 14(12) : 2580–2596.
- Landthaler, M., Yalcin, A., et Tuschl, T. (2004). The human DiGeorge syndrome critical region gene 8 and its *D. melanogaster* homolog are required for miRNA biogenesis. *Curr Biol*, 14(23) : 2162–2167.

- Lau, N., Robine, N., Martin, R., Chung, W. J., Niki, Y., Berezikov, E., et Lai, E. C. (2009). Abundant primary piRNAs, endo-siRNAs and microRNAs in a *Drosophila* ovary cell line. *Genome Res.*
- Lau, N. C., Seto, A. G., Kim, J., Kuramochi-Miyagawa, S., Nakano, T., Bartel, D. P., et Kingston, R. E. (2006). Characterization of the piRNA complex from rat testes. *Science*, 313(5785) : 363–367.
- Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Radmark, O., Kim, S., et Kim, V. N. (2003). The nuclear RNase III Drosha initiates microRNA processing. *Nature*, 425(6956) : 415–419.
- Lee, Y., Hur, I., Park, S. Y., Kim, Y. K., Suh, M. R., et Kim, V. N. (2006). The role of PACT in the RNA silencing pathway. *EMBO J*, 25(3) : 522–532.
- Lee, Y., Jeon, K., Lee, J. T., Kim, S., et Kim, V. N. (2002). MicroRNA maturation : stepwise processing and subcellular localization. *EMBO J*, 21(17) : 4663–4670.
- Lee, Y., Kim, M., Han, J., Yeom, K. H., Lee, S., Baek, S. H., et Kim, V. N. (2004a). MicroRNA genes are transcribed by RNA polymerase II. *EMBO J*, 23(20) : 4051–4060.
- Lee, Y. S., Nakahara, K., Pham, J. W., Kim, K., He, Z., Sontheimer, E. J., et Carthew, R. W. (2004b). Distinct roles for *Drosophila* Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. *Cell*, 117(1) : 69–81.
- Lei, E. P. et Corces, V. G. (2006). RNA interference machinery influences the nuclear organization of a chromatin insulator. *Nat Genet*, 38(8) : 936–941.
- Lewis, B. P., Burge, C. B., et Bartel, D. P. (2005). Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*, 120(1) : 15–20.
- Lewis, B. P., Shih, I. H., Jones-Rhoades, M. W., Bartel, D. P., et Burge, C. B. (2003). Prediction of mammalian microRNA targets. *Cell*, 115(7) : 787–798.
- Li, C., Vagin, V. V., Lee, S., Xu, J., Ma, S., Xi, H., Seitz, H., Horwich, M. D., Syrzycka, M., Honda, B. M., Kittler, E. L., Zapp, M. L., Klattenhoff, C., Schulz, N., Theurkauf, W. E., Weng, Z., et Zamore, P. D. (2009a). Collapse of germline piRNAs in the absence of Argonaute3 reveals somatic piRNAs in flies. *Cell*, 137(3) : 509–521.
- Li, T., Huang, J., Jiang, Y., Zeng, Y., He, F., Zhang, M. Q., Han, Z., et Zhang, X. (2009b). Multi-stage analysis of gene expression and transcription regulation in C57/B6 mouse liver development. *Genomics*, 93(3) : 235–242.
- Lim, L. P., Lau, N. C., Garrett-Engele, P., Grimson, A., Schelter, J. M., Castle, J., Bartel, D. P., Linsley, P. S., et Johnson, J. M. (2005). Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature*, 433(7027) : 769–773.
- Lingel, A., Simon, B., Izaurralde, E., et Sattler, M. (2003). Structure and nucleic-acid binding of the *Drosophila* Argonaute 2 PAZ domain. *Nature*, 426(6965) : 465–469.
- Lingel, A., Simon, B., Izaurralde, E., et Sattler, M. (2004). Nucleic acid 3'-end recognition by the Argonaute2 PAZ domain. *Nat Struct Mol Biol*, 11(6) : 576–577.
- Linsley, P. S., Schelter, J., Burchard, J., Kibukawa, M., Martin, M. M., Bartz, S. R., Johnson, J. M., Cummins, J. M., Raymond, C. K., Dai, H., Chau, N., Cleary, M., Jackson, A. L., Carleton, M., et Lim, L. (2007). Transcripts targeted by the microRNA-16 family cooperatively regulate cell cycle progression. *Mol Cell Biol*, 27(6) : 2240–2252.
- Liu, J., Carmell, M. A., Rivas, F. V., Marsden, C. G., Thomson, J. M., Song, J. J., Hammond, S. M., Joshua-Tor, L., et Hannon, G. J. (2004). Argonaute2 is the catalytic engine of mammalian RNAi. *Science*, 305(5689) : 1437–1441.

- Liu, Q., Rand, T. A., Kalidas, S., Du, F., Kim, H. E., Smith, D. P., et Wang, X. (2003). R2D2, a bridge between the initiation and effector steps of the *Drosophila* RNAi pathway. *Science*, 301(5641) : 1921–1925.
- Lu, R., Maduro, M., Li, F., Li, H. W., Broitman-Maduro, G., Li, W. X., et Ding, S. W. (2005). Animal virus replication and RNAi-mediated antiviral silencing in *Caenorhabditis elegans*. *Nature*, 436(7053) : 1040–1043.
- Lund, E., Güttinger, S., Calado, A., Dahlberg, J. E., et Kutay, U. (2004). Nuclear export of microRNA precursors. *Science*, 303(5654) : 95–98.
- Ma, J. B., Ye, K., et Patel, D. J. (2004). Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain. *Nature*, 429(6989) : 318–322.
- Ma, J. B., Yuan, Y. R., Meister, G., Pei, Y., Tuschl, T., et Patel, D. J. (2005). Structural basis for 5'-end-specific recognition of guide RNA by the *A. fulgidus* Piwi protein. *Nature*, 434(7033) : 666–670.
- Mallory, A. C., Reinhart, B. J., Jones-Rhoades, M. W., Tang, G., Zamore, P. D., Barton, M. K., et Bartel, D. P. (2004). MicroRNA control of PHABULOSA in leaf development : importance of pairing to the microRNA 5' region. *EMBO J*, 23(16) : 3356–3364.
- Malone, C. D., Brennecke, J., Dus, M., Stark, A., McCombie, W. R., Sachidanandam, R., et Hannon, G. J. (2009). Specialized piRNA pathways act in germline and somatic tissues of the *Drosophila* ovary. *Cell*, 137(3) : 522–535.
- Maroney, P. A., Yu, Y., Fisher, J., et Nilsen, T. W. (2006). Evidence that microRNAs are associated with translating messenger RNAs in human cells. *Nat Struct Mol Biol*, 13(12) : 1102–1107.
- Martinez, J., Patkaniowska, A., Urlaub, H., Lührmann, R., et Tuschl, T. (2002). Single-stranded antisense siRNAs guide target RNA cleavage in RNAi. *Cell*, 110(5) : 563–574.
- Matranga, C., Tomari, Y., Shin, C., Bartel, D. P., et Zamore, P. D. (2005). Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes. *Cell*, 123(4) : 607–620.
- Meister, G., Landthaler, M., Patkaniowska, A., Dorsett, Y., Teng, G., et Tuschl, T. (2004). Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Mol Cell*, 15(2) : 185–197.
- Miller, B. H., McDearmon, E. L., Panda, S., Hayes, K. R., Zhang, J., Andrews, J. L., Antoch, M. P., Walker, J. R., Esser, K. A., Hogenesch, J. B., et Takahashi, J. S. (2007). Circadian and CLOCK-controlled regulation of the mouse transcriptome and cell proliferation. *Proc Natl Acad Sci USA*, 104(9) : 3342–3347.
- Miranda, K. C., Huynh, T., Tay, Y., Ang, Y. S., Tam, W. L., Thomson, A. M., Lim, B., et Rigoutsos, I. (2006). A pattern-based method for the identification of microRNA binding sites and their corresponding heteroduplexes. *Cell*, 126(6) : 1203–1217.
- Morlando, M., Ballarino, M., Gromak, N., Pagano, F., Bozzoni, I., et Proudfoot, N. J. (2008). Primary microRNA transcripts are processed co-transcriptionally. *Nat Struct Mol Biol*, 15(9) : 902–909.
- Nakahara, K., Kim, K., Sciulli, C., Dowd, S. R., Minden, J. S., et Carthew, R. W. (2005). Targets of microRNA regulation in the *Drosophila* oocyte proteome. *Proc Natl Acad Sci USA*, 102(34) : 12023–12028.
- Nelson, P. T., Hatzigeorgiou, A. G., et Mourelatos, Z. (2004). miRNP :mRNA association in polyribosomes in a human neuronal cell line. *RNA*, 10(3) : 387–394.
- Newman, M. A., Thomson, J. M., et Hammond, S. M. (2008). Lin-28 interaction with the Let-7 precursor loop mediates regulated microRNA processing. *RNA*, 14(8) : 1539–1549.

- Nottrott, S., Simard, M. J., et Richter, J. D. (2006). Human let-7a miRNA blocks protein production on actively translating polyribosomes. *Nat Struct Mol Biol*, 13(12) : 1108–1114.
- Nykänen, A., Haley, B., et Zamore, P. D. (2001). ATP requirements and small interfering RNA structure in the RNA interference pathway. *Cell*, 107(3) : 309–321.
- Obernosterer, G., Leuschner, P. J., Alenius, M., et Martinez, J. (2006). Post-transcriptional regulation of microRNA expression. *RNA*, 12(7) : 1161–1167.
- O'Donnell, K. A., Wentzel, E. A., Zeller, K. I., Dang, C. V., et Mendell, J. T. (2005). c-Myc-regulated microRNAs modulate E2F1 expression. *Nature*, 435(7043) : 839–843.
- Okamura, K., Balla, S., Martin, R., Liu, N., et Lai, E. C. (2008a). Two distinct mechanisms generate endogenous siRNAs from bidirectional transcription in *Drosophila melanogaster*. *Nat Struct Mol Biol*, 15(6) : 581–590.
- Okamura, K., Chung, W. J., Ruby, J. G., Guo, H., Bartel, D. P., et Lai, E. C. (2008b). The *Drosophila* hairpin RNA pathway generates endogenous short interfering RNAs. *Nature*, 453(7196) : 803–806.
- Okamura, K., Hagen, J. W., Duan, H., Tyler, D. M., et Lai, E. C. (2007). The mirtron pathway generates microRNA-class regulatory RNAs in *Drosophila*. *Cell*, 130(1) : 89–100.
- Okamura, K., Ishizuka, A., Siomi, H., et Siomi, M. C. (2004). Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways. *Genes Dev*, 18(14) : 1655–1666.
- Olsen, P. H. et Ambros, V. (1999). The *lin-4* regulatory RNA controls developmental timing in *Caenorhabditis elegans* by blocking LIN-14 protein synthesis after the initiation of translation. *Dev Biol*, 216(2) : 671–680.
- Page, A. R., Kovacs, A., Deak, P., Torok, T., Kiss, I., Dario, P., Bastos, C., Batista, P., Gomes, R., Ohkura, H., Russell, S., et Glover, D. M. (2005). Spotted-dick, a zinc-finger protein of *Drosophila* required for expression of Orc4 and S phase. *EMBO J*, 24(24) : 4304–4315.
- Pak, J. et Fire, A. (2007). Distinct populations of primary and secondary effectors during RNAi in *C. elegans*. *Science*, 315(5809) : 241–244.
- Parker, J. S., Parizotto, E. A., Wang, M., Roe, S. M., et Barford, D. (2009). Enhancement of the seed-target recognition step in RNA silencing by a PIWI/MID domain protein. *Mol Cell*, 33(2) : 204–214.
- Parker, J. S., Roe, S. M., et Barford, D. (2004). Crystal structure of a PIWI protein suggests mechanisms for siRNA recognition and slicer activity. *EMBO J*, 23(24) : 4727–4737.
- Parker, J. S., Roe, S. M., et Barford, D. (2005). Structural insights into mRNA recognition from a PIWI domain-siRNA guide complex. *Nature*, 434(7033) : 663–666.
- Parrish, S. et Fire, A. (2001). Distinct roles for RDE-1 and RDE-4 during RNA interference in *Caenorhabditis elegans*. *RNA*, 7(10) : 1397–1402.
- Parrish, S., Fleenor, J., Xu, S., Mello, C., et Fire, A. (2000). Functional anatomy of a dsRNA trigger : differential requirement for the two trigger strands in RNA interference. *Mol Cell*, 6(5) : 1077–1087.
- Péllisson, A., Sarot, E., Payen-Groschène, G., et Bucheton, A. (2007). A novel repeat-associated small interfering RNA-mediated silencing pathway downregulates complementary sense gypsy transcripts in somatic cells of the *Drosophila* ovary. *J Virol*, 81(4) : 1951–1960.
- Péllisson, A., Song, S. U., Prud'homme, N., Smith, P. A., Bucheton, A., et Corces, V. G. (1994). Gypsy transposition correlates with the production of a retroviral envelope-like protein under the tissue-specific control of the *Drosophila flamenco* gene. *EMBO J*, 13(18) : 4401–4411.
- Pillai, R. S., Bhattacharyya, S. N., Artus, C. G., Zoller, T., Cougot, N., Basyuk, E., Bertrand, E., et Filipowicz, W. (2005). Inhibition of translational initiation by Let-7 MicroRNA in human cells. *Science*, 309(5740) : 1573–1576.

- Piskounova, E., Viswanathan, S. R., Janas, M., LaPierre, R. J., Daley, G. Q., Sliz, P., et Gregory, R. I. (2008). Determinants of microRNA processing inhibition by the developmentally regulated RNA-binding protein Lin28. *J Biol Chem*, 283(31) : 21310–21314.
- Putnam, N. H., Srivastava, M., Hellsten, U., Dirks, B., Chapman, J., Salamov, A., Terry, A., Shapiro, H., Lindquist, E., Kapitonov, V. V., Jurka, J., Genikhovich, G., Grigoriev, I. V., Lucas, S. M., Steele, R. E., Finnerty, J. R., Technau, U., Martindale, M. Q., et Rokhsar, D. S. (2007). Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science*, 317(5834) : 86–94.
- Rand, T. A., Petersen, S., Du, F., et Wang, X. (2005). Argonaute2 cleaves the anti-guide strand of siRNA during RISC activation. *Cell*, 123(4) : 621–629.
- Rivas, F. V., Tolia, N. H., Song, J. J., Aragon, J. P., Liu, J., Hannon, G. J., et Joshua-Tor, L. (2005). Purified Argonaute2 and an siRNA form recombinant human RISC. *Nat Struct Mol Biol*, 12(4) : 340–349.
- Ro, S., Park, C., Young, D., Sanders, K. M., et Yan, W. (2007). Tissue-dependent paired expression of miRNAs. *Nucleic Acids Res*, 35(17) : 5944–5953.
- Rodriguez, A., Hilvo, M., Kytömäki, L., Fleming, R. E., Britton, R. S., Bacon, B. R., et Parkkila, S. (2007). Effects of iron loading on muscle : genome-wide mRNA expression profiling in the mouse. *BMC Genomics*, 8 : 379.
- Roger, A. J. et Hug, L. A. (2006). The origin and diversification of eukaryotes : problems with molecular phylogenetics and molecular clock estimation. *Philos Trans R Soc Lond B Biol Sci*, 361(1470) : 1039–1054.
- Ruby, J. G., Jan, C., Player, C., Axtell, M. J., Lee, W., Nusbaum, C., Ge, H., et Bartel, D. P. (2006). Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in *C. elegans*. *Cell*, 127(6) : 1193–1207.
- Ruby, J. G., Jan, C. H., et Bartel, D. P. (2007). Intronic microRNA precursors that bypass Drosha processing. *Nature*, 448(7149) : 83–86.
- Rybak, A., Fuchs, H., Smirnova, L., Brandt, C., Pohl, E. E., Nitsch, R., et Wulczyn, F. G. (2008). A feedback loop comprising lin-28 and let-7 controls pre-let-7 maturation during neural stem-cell commitment. *Nat Cell Biol*, 10(8) : 987–993.
- Saito, K., Ishizuka, A., Siomi, H., et Siomi, M. C. (2005). Processing of pre-microRNAs by the Dicer-1-Loquacious complex in *Drosophila* cells. *PLoS Biol*, 3(7) : e235.
- Saito, K., Nishida, K. M., Mori, T., Kawamura, Y., Miyoshi, K., Nagami, T., Siomi, H., et Siomi, M. C. (2006). Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the *Drosophila* genome. *Genes Dev*, 20(16) : 2214–2222.
- Saxena, S., Jonsson, Z. O., et Dutta, A. (2003). Small RNAs with imperfect match to endogenous mRNA repress translation. Implications for off-target activity of small inhibitory RNA in mammalian cells. *J Biol Chem*, 278(45) : 44312–44319.
- Schmitter, D., Filkowski, J., Sewer, A., Pillai, R. S., Oakeley, E. J., Zavolan, M., Svoboda, P., et Filipowicz, W. (2006). Effects of Dicer and Argonaute down-regulation on mRNA levels in human HEK293 cells. *Nucleic Acids Res*, 34(17) : 4801–4815.
- Schwarz, D. S., Hutvagner, G., Du, T., Xu, Z., Aronin, N., et Zamore, P. D. (2003). Asymmetry in the assembly of the RNAi enzyme complex. *Cell*, 115(2) : 199–208.
- Schwarz, D. S., Hutvagner, G., Haley, B., et Zamore, P. D. (2002). Evidence that siRNAs function as guides, not primers, in the *Drosophila* and human RNAi pathways. *Mol Cell*, 10(3) : 537–548.
- Schwarz, D. S., Tomari, Y., et Zamore, P. D. (2004). The RNA-induced silencing complex is a Mg²⁺-dependent endonuclease. *Curr Biol*, 14(9) : 787–791.

- Seipel, K., Yanze, N., et Schmid, V. (2004). The germ line and somatic stem cell gene *Cniwi* in the jellyfish *Podocoryne carnea*. *Int J Dev Biol*, 48(1) : 1–7.
- Seitz, H. (2009). Redefining microRNA targets. *Curr Biol*, 19(10) : 870–873.
- Seitz, H., Ghildiyal, M., et Zamore, P. D. (2008). Argonaute loading improves the 5' precision of both microRNAs and their miRNA* strands in flies. *Curr Biol*, 18(2) : 147–151.
- Seitz, H., Wee, L., Rentzsch, F., et Zamore, P. D. (en prep.). Coexistence of siRNAs and piRNAs over one billion years suggests an additional function for piRNAs. En préparation.
- Seitz, H. et Zamore, P. D. (2006). Rethinking the microprocessor. *Cell*, 125(5) : 827–829.
- Selbach, M., Schwanhäusser, B., Thierfelder, N., Fang, Z., Khanin, R., et Rajewsky, N. (2008). Widespread changes in protein synthesis induced by microRNAs. *Nature*, 455(7209) : 58–63.
- Sijen, T., Steiner, F. A., Thijssen, K. L., et Plasterk, R. H. (2007). Secondary siRNAs result from unprimed RNA synthesis and form a distinct class. *Science*, 315(5809) : 244–247.
- Sledz, C. A., Holko, M., de Veer, M. J., Silverman, R. H., et Williams, B. R. (2003). Activation of the interferon system by short-interfering RNAs. *Nat Cell Biol*, 5(9) : 834–839.
- Song, J. J., Liu, J., Tolia, N. H., Schneidman, J., Smith, S. K., Martienssen, R. A., Hannon, G. J., et Joshua-Tor, L. (2003). The crystal structure of the Argonaute2 PAZ domain reveals an RNA binding motif in RNAi effector complexes. *Nat Struct Biol*, 10(12) : 1026–1032.
- Song, J. J., Smith, S. K., Hannon, G. J., et Joshua-Tor, L. (2004). Crystal structure of Argonaute and its implications for RISC Slicer activity. *Science*, 305(5689) : 1434–1437.
- Stark, A., Brennecke, J., Russell, R. B., et Cohen, S. M. (2003). Identification of *Drosophila* MicroRNA Targets. *PLoS Biol*, 1(3) : E60.
- Steiner, F. A., Hoogstrate, S. W., Okihara, K. L., Thijssen, K. L., Ketting, R. F., Plasterk, R. H., et Sijen, T. (2007). Structural features of small RNA precursors determine Argonaute loading in *Caenorhabditis elegans*. *Nat Struct Mol Biol*, 14(10) : 927–933.
- Steiner, F. A., Okihara, K. L., Hoogstrate, S. W., Sijen, T., et Ketting, R. F. (2009). RDE-1 slicer activity is required only for passenger-strand cleavage during RNAi in *Caenorhabditis elegans*. *Nat Struct Mol Biol*, 16(2) : 207–211.
- Sullivan, C. S. et Ganem, D. (2005). A virus-encoded inhibitor that blocks RNA interference in mammalian cells. *J Virol*, 79(12) : 7371–7379.
- Tabara, H., Sarkissian, M., Kelly, W. G., Fleenor, J., Grishok, A., Timmons, L., Fire, A., et Mello, C. C. (1999). The *rde-1* gene, RNA interference, and transposon silencing in *C. elegans*. *Cell*, 99(2) : 123–132.
- Tabara, H., Yigit, E., Siomi, H., et Mello, C. C. (2002). The dsRNA binding protein RDE-4 interacts with RDE-1, DCR-1, and a DExH-box helicase to direct RNAi in *C. elegans*. *Cell*, 109(7) : 861–871.
- Taira, K. (2006). Retraction : Induction of DNA methylation and gene silencing by short interfering RNAs in human cells. *Nature*, 441(7097) : 1176.
- Tam, O. H., Aravin, A. A., Stein, P., Girard, A., Murchison, E. P., Cheloufi, S., Hodges, E., Anger, M., Sachidanandam, R., Schultz, R. M., et Hannon, G. J. (2008). Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. *Nature*, 453(7194) : 534–538.
- Tchurikov, N. A. et Kretova, O. V. (2007). Suffix-specific RNAi leads to silencing of F element in *Drosophila melanogaster*. *PLoS One*, 2(5) : e476.
- Tijsterman, M., Ketting, R. F., Okihara, K. L., Sijen, T., et Plasterk, R. H. (2002). RNA helicase MUT-14-dependent gene silencing triggered in *C. elegans* by short antisense RNAs. *Science*, 295(5555) : 694–697.

- Tomari, Y., Du, T., et Zamore, P. D. (2007). Sorting of *Drosophila* small silencing RNAs. *Cell*, 130(2) : 299–308.
- Tomari, Y., Matranga, C., Haley, B., Martinez, N., et Zamore, P. D. (2004). A protein sensor for siRNA asymmetry. *Science*, 306(5700) : 1377–1380.
- Vagin, V. V., Sigova, A., Li, C., Seitz, H., Gvozdev, V., et Zamore, P. D. (2006). A distinct small RNA pathway silences selfish genetic elements in the germline. *Science*, 313(5785) : 320–324.
- van Rij, R. P., Saleh, M. C., Berry, B., Foo, C., Houk, A., Antoniewski, C., et Andino, R. (2006). The RNA silencing endonuclease Argonaute 2 mediates specific antiviral immunity in *Drosophila melanogaster*. *Genes Dev*, 20(21) : 2985–2995.
- Vastenhouw, N. L., Fischer, S. E., Robert, V. J., Thijssen, K. L., Fraser, A. G., Kamath, R. S., Ahringer, J., et Plasterk, R. H. (2003). A genome-wide screen identifies 27 genes involved in transposon silencing in *C. elegans*. *Curr Biol*, 13(15) : 1311–1316.
- Vasudevan, S., Tong, Y., et Steitz, J. A. (2007). Switching from repression to activation : microRNAs can up-regulate translation. *Science*, 318(5858) : 1931–1934.
- Vinther, J., Hedegaard, M. M., Gardner, P. P., Andersen, J. S., et Arctander, P. (2006). Identification of miRNA targets with stable isotope labeling by amino acids in cell culture. *Nucleic Acids Res*, 34(16) : e107.
- Viswanathan, S. R., Daley, G. Q., et Gregory, R. I. (2008). Selective blockade of microRNA processing by Lin28. *Science*, 320(5872) : 97–100.
- Wakiyama, M., Takimoto, K., Ohara, O., et Yokoyama, S. (2007). Let-7 microRNA-mediated mRNA deadenylation and translational repression in a mammalian cell-free system. *Genes Dev*, 21(15) : 1857–1862.
- Wang, X. H., Aliyari, R., Li, W. X., Li, H. W., Kim, K., Carthew, R., Atkinson, P., et Ding, S. W. (2006). RNA interference directs innate immunity against viruses in adult *Drosophila*. *Science*, 312(5772) : 452–454.
- Wang, Y., Juranek, S., Li, H., Sheng, G., Tuschl, T., et Patel, D. J. (2008a). Structure of an Argonaute silencing complex with a seed-containing guide DNA and target RNA duplex. *Nature*, 456(7224) : 921–926.
- Wang, Y., Sheng, G., Juranek, S., Tuschl, T., et Patel, D. J. (2008b). Structure of the guide-strand-containing Argonaute silencing complex. *Nature*, 456(7219) : 209–213.
- Watanabe, T., Takeda, A., Tsukiyama, T., Mise, K., Okuno, T., Sasaki, H., Minami, N., et Imai, H. (2006). Identification and characterization of two novel classes of small RNAs in the mouse germline : retrotransposon-derived siRNAs in oocytes and germline small RNAs in testes. *Genes Dev*, 20(13) : 1732–1743.
- Watanabe, T., Totoki, Y., Toyoda, A., Kaneda, M., Kuramochi-Miyagawa, S., Obata, Y., Chiba, H., Kohara, Y., Kono, T., Nakano, T., Surani, M. A., Sakaki, Y., et Sasaki, H. (2008). Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature*, 453(7194) : 539–543.
- Wu, L., Fan, J., et Belasco, J. G. (2006). MicroRNAs direct rapid deadenylation of mRNA. *Proc Natl Acad Sci USA*, 103(11) : 4034–4039.
- Wulczyn, F. G., Smirnova, L., Rybak, A., Brandt, C., Kwidzinski, E., Ninnemann, O., Strehle, M., Seiler, A., Schumacher, S., et Nitsch, R. (2007). Post-transcriptional regulation of the *let-7* microRNA during neural cell specification. *FASEB J*, 21(2) : 415–426.
- Yan, K. S., Yan, S., Farooq, A., Han, A., Zeng, L., et Zhou, M. M. (2003). Structure and conserved RNA binding of the PAZ domain. *Nature*, 426(6965) : 468–474.

- Yang, D., Lu, H., et Erickson, J. W. (2000). Evidence that processed small dsRNAs may mediate sequence-specific mRNA degradation during RNAi in *Drosophila* embryos. *Curr Biol*, 10(19) : 1191—1200.
- Yang, N. et Kazazian, Jr., H. H. (2006). L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells. *Nat Struct Mol Biol*, 13(9) : 763–771.
- Yekta, S., Shih, I.-h., et Bartel, D. P. (2004). MicroRNA-directed cleavage of *HOXB8* mRNA. *Science*, 304(5670) : 594—596.
- Yi, R., Qin, Y., Macara, I. G., et Cullen, B. R. (2003). Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev*, 17(24) : 3011–3016.
- Yigit, E., Batista, P. J., Bei, Y., Pang, K. M., Chen, C. C., Tolia, N. H., Joshua-Tor, L., Mitani, S., Simard, M. J., et Mello, C. C. (2006). Analysis of the *C. elegans* Argonaute family reveals that distinct Argonautes act sequentially during RNAi. *Cell*, 127(4) : 747—757.
- Yuan, Y. R., Pei, Y., Ma, J. B., Kuryavyyi, V., Zhadina, M., Meister, G., Chen, H. Y., Dauter, Z., Tuschl, T., et Patel, D. J. (2005). Crystal structure of *A. aeolicus* Argonaute, a site-specific DNA-guided endoribonuclease, provides insights into RISC-mediated mRNA cleavage. *Mol Cell*, 19(3) : 405–419.
- Zamore, P. D., Tuschl, T., Sharp, P. A., et Bartel, D. P. (2000). RNAi : double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell*, 101(1) : 25—33.
- Zhang, H., Kolb, F. A., Jaskiewicz, L., Westhof, E., et Filipowicz, W. (2004). Single processing center models for human Dicer and bacterial RNase III. *Cell*, 118(1) : 57–68.
- Zhang, L., Ding, L., Cheung, T. H., Dong, M. Q., Chen, J., Sewell, A. K., Liu, X., Yates, 3rd, J. R., et Han, M. (2007). Systematic identification of *C. elegans* miRISC proteins, miRNAs, and mRNA targets by their interactions with GW182 proteins AIN-1 and AIN-2. *Mol Cell*, 28(4) : 598–613.
- Zhao, Y., Samal, E., et Srivastava, D. (2005). Serum response factor regulates a muscle-specific microRNA that targets *Hand2* during cardiogenesis. *Nature*, 436(7048) : 214–220.

Index

- A**
Archæ 14
- B**
biologie des systèmes 130
brin guide, brin passager 11
- C**
cellules ES 131
coiffe 22
- D**
découverte des piRNA 36
Deutérostomiens 128
divergence des Animaux 105
divergence des Eucaryotes 30
domaine DUF1785 14, 106
domaine Paz 14, 106
domaine Piwi 14, 106
dsRBD 12, 19, 20
dysgénie des hybrides 102
- E**
Eubactéries 14
exaptation 125
- F**
flamenco 86
- G**
graine d'un microARN 16, 25, 52
GW182 23
- H**
hétérochromatine 17, 86
Hen1 13, 106
HP1 86
Hydre 104, 125
- I**
impact physiologique des microARN .. 19, 119
initiation de la traduction 22
IRES 22
- L**
Lépidoptère 86, 125
Loquacious 19, 43
- M**
Méduse 123
méiose 125
maturation de RISC 57
Microprocessor 20, 53, 131
miRBase 46
modèle du ping-pong 70, 86, 102, 107
- N**
nomenclature des protéines Argonaute .. 14, 16
- P**
peptide naissant 24
Protostomiens 128
- Q**
queue poly(A) 22
- R**
R3D1 voir Loquacious
régénération 124
réponse interféron 11, 30, 43
rasiRNA 30
RdRP 13, 43
reproduction asexuée 122
RLC 12
- S**
shRNA 11
SNP 129
- T**
triade catalytique 14, 70
triage des petits ARN 21, 57