



**HAL**  
open science

# Etude de la biologie des clusters de piRNAs chez *Drosophila melanogaster* en utilisant comme modèle le locus flamenco

Nolwenn Mouniée

► **To cite this version:**

Nolwenn Mouniée. Etude de la biologie des clusters de piRNAs chez *Drosophila melanogaster* en utilisant comme modèle le locus flamenco. Génétique animale. Université Clermont Auvergne [2017-2020], 2019. Français. NNT : 2019CLFAC029 . tel-02409261

**HAL Id: tel-02409261**

**<https://theses.hal.science/tel-02409261>**

Submitted on 13 Dec 2019

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

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

ECOLE DOCTORALE DES SCIENCES DE LA VIE  
ET DE LA SANTE

N° d'ordre

**Thèse**

Présentée à l'Université Clermont Auvergne  
Pour l'obtention du grade de DOCTEUR

(Décret du 5 juillet 1984)

Spécialité : Génétique et Développement

Nolwenn MOUNIÉE

---

**Etude de la biologie des clusters de piRNAs chez  
*Drosophila melanogaster* en utilisant comme modèle le  
locus *flamenco***

---

Soutenue publiquement le mardi 16 juillet 2019 devant le jury d'examen :

Rapporteurs : CARRÉ Clément  
DUHARCOURT Sandra  
GILBERT Clément

Examineurs : ARNAUD Philippe  
GALLEGO Maria Eugenia

Directeur de Thèse : BRASSET Emilie

Équipe « Instabilités génétiques et contrôle par le génome de l'hôte »  
Laboratoire GReD (Génétique, Reproduction et Développement)  
CNRS (UMR6293) – INSERM (U1103)  
Université Clermont Auvergne



ECOLE DOCTORALE DES SCIENCES DE LA VIE  
ET DE LA SANTE

N° d'ordre

**Thèse**

Présentée à l'Université Clermont Auvergne  
Pour l'obtention du grade de DOCTEUR

(Décret du 5 juillet 1984)

Spécialité : Génétique et Développement

Nolwenn MOUNIÉE

---

**Etude de la biologie des clusters de piRNAs chez  
*Drosophila melanogaster* en utilisant comme modèle le  
locus *flamenco***

---

Soutenue publiquement le mardi 16 juillet 2019 devant le jury d'examen :

Rapporteurs : CARRÉ Clément  
DUHARCOURT Sandra  
GILBERT Clément

Examineurs : ARNAUD Philippe  
GALLEGO Maria Eugenia

Directeur de Thèse : BRASSET Emilie

Équipe « Instabilités génétiques et contrôle par le génome de l'hôte »  
Laboratoire GReD (Génétique, Reproduction et Développement)  
CNRS (UMR6293) – INSERM (U1103)  
Université Clermont Auvergne



## Résumé

Les éléments transposables (ETs) sont des séquences d'ADN mobiles retrouvées dans les génomes de toutes les espèces où ils ont été recherchés. Moteurs de l'évolution, ces éléments mobiles, présents en de nombreuses copies dans les génomes, ont joué un rôle majeur dans la dynamique des génomes en engendrant des mutations et des réarrangements chromosomiques. Néanmoins, étant des constituants majeurs des génomes, ils doivent être finement régulés dans le but de préserver l'intégrité génomique, et ainsi de conserver l'équilibre entre variabilité et stabilité des génomes. Afin de protéger l'information génétique de l'hôte transmise à la descendance, la régulation des ETs au niveau des gonades est effectuée par la voie des piRNAs, voie d'ARN interférent conservée chez les animaux. Bien qu'elle soit relativement bien décrite chez la drosophile et la souris, certaines étapes de cette voie restent encore incomprises. Durant ma thèse, j'ai exploré différents aspects de la biologie des clusters de piRNAs, en prenant comme modèle d'étude le locus *flamenco*. Le cluster de piRNAs *flamenco* est le producteur majeur de piRNAs dans les cellules folliculaires des ovaires de *Drosophila melanogaster*. Tout d'abord, j'ai analysé les fenêtres spatio-temporelles de l'expression du cluster de piRNAs *flamenco* tout au long du développement de la drosophile, de l'embryon à l'âge adulte. Ensuite, j'ai recherché, *in vivo*, la séquence des transcrits de *flamenco* qui serait suffisante pour induire l'adressage d'un transcrite chimérique à la voie de maturation des piRNAs. J'ai également exploré l'impact de certains facteurs sur la prise en charge de transcrits artificiels par la voie des piRNAs. Enfin, je me suis intéressée à la régulation génique que pourraient effectuer les piRNAs provenant de *flamenco* dans les ovaires de drosophile en recherchant, par des approches bioinformatiques et de biologie moléculaire, les gènes potentiellement reconnus, et par conséquent, régulés par les piRNAs de *flamenco*. L'ensemble de ces axes de recherche *in vivo* permettront d'avancer dans la compréhension de la biologie des clusters de piRNAs ainsi que sur les mécanismes moléculaires mis en jeu lors de la biogenèse des piRNAs chez la drosophile.



## Abstract

Transposable elements (TEs) are defined such as mobile DNA sequences found in genomes of all species where they were searched. As evolutionary drivers, these mobile elements, present in many copies in genomes, have played a major role in the genome dynamics by generating mutations and chromosomal rearrangements. Nevertheless, being major genome constituents, they must be finely regulated in order to preserve the genomic integrity, and thus, to maintain the balance between variability and stability of genomes. In order to protect the genetic information of the host transmitted to the offspring, the gonadal TE regulation is carried out by the piRNAs pathway, an interfering RNA pathway conserved in animals. Although this is relatively well described in *Drosophila* and in mouse, some steps of piRNA pathway are still misunderstood. During my thesis, I explored various aspects of piRNA cluster biology, using the *flamenco* locus as a model. This piRNA cluster is the main piRNA producer in the follicular cells of *Drosophila melanogaster* ovaries. First, I analyzed the spatio-temporal windows of *flamenco* piRNA cluster expression throughout the *Drosophila* development, from embryo to adulthood. Then, I searched, *in vivo*, the *flamenco* transcript sequence that would be sufficient to induce the addressing of a chimeric transcript to the piRNA processing pathway. I also explored the impact of some factors on the management of artificial transcripts by piRNAs. Finally, I was interested in the gene regulation that *flamenco*-derived piRNAs could make in *Drosophila* ovaries by searching, through bioinformatics and molecular biology approaches, the potentially recognized genes, and therefore, regulated by *flamenco* piRNAs. All of these *in vivo* research axes will advance in the understanding of the biology of piRNA clusters as well as the molecular mechanisms involved in the piRNA biogenesis in *Drosophila*.





## *Remerciements*

J'adresse tout d'abord tous mes remerciements à Clément CARRE, Sandra DUHARCOURT ainsi qu'à Clément GILBERT de l'honneur qu'ils m'ont fait en acceptant d'être rapporteurs de cette thèse et de se déplacer jusqu'à Clermont-Ferrand pour cette occasion. J'exprime ma gratitude à Maria GALLEGO ainsi que Philippe ARNAUD d'avoir gentiment accepté d'évaluer mes travaux de thèse en tant qu'examineurs.

Je remercie chaleureusement Chantal VAURY et Emilie BRASSET pour m'avoir accueillie, jeune Master 2 que j'étais, dans leur équipe "Instabilités génétiques et contrôle par le génome de l'hôte", au sein de laquelle, j'ai pu poursuivre mes travaux en thèse.

Je remercie tout particulièrement ma directrice de thèse, Emilie BRASSET, de m'avoir guidée durant ces quatre années de thèse, du début à la fin. Je tiens à te remercier pour ton admirable gentillesse, ta disponibilité de chaque moment, ton indéfectible soutien, mais surtout pour tes innombrables encouragements. J'espère avoir été à la hauteur de tes attentes.

Un grand merci à chacun des membres de cette équipe formidable ! Vous avez su répondre avec calme et patience aux questions quotidiennes dont je vous accablais. Merci à toi, Silke, pour ces moments "pause café" qui m'ont fait un bien fou, pour ton oreille attentive et tes nombreux conseils. Benjamin, Stéphanie et Marianne, la fine équipe, le trio infernal, sachez que, durant ma thèse, vous avez été ma bouffée d'oxygène ! Fous rires, expressions dérisoires et jeux de mots assurés !!! Vous l'aurez vite compris, vous allez tous cruellement me manquer... À parler de notre équipe, je m'aperçois que l'on ne fait pas juste parti de la même équipe, mais d'une mini-famille.

J'adresse aussi mes remerciements à Antoine BOIVIN, Laure TEYSSET, Carine MEGNIN et Hang NGUYEN pour leur participation à mes divers comités de suivi de thèse.

Je tiens à remercier également les drosophilistes du 5<sup>ème</sup> étage pour ces moments d'échanges scientifiques, mais aussi les non scientifiques !!! Le repiquage de mes 15 caisses passe beaucoup plus vite avec les blind-tests et les karaokés !

Je remercie bien également toutes les personnes du service technique et administratif du GReD.

Un grand merci à Graziella, Nathalie, Nadège, Fanfan et ma Nanny pour les repas du midi animés lors de ces trois premières années de thèse. Je tiens à te remercier Katarina pour ton sourire et ta bonne humeur à toute épreuve !



Je remercie les membres de l'équipe SAPIN qui m'ont diagnostiqué le syndrome du "sapin refoulé". Marilyne, Hélène, Corine, Loïc et Vincent pour votre bonne humeur, vos taquineries et votre bienveillance. Venir vous voir au 3ième est un réel bonheur !!!

Christelle GROSS, mon cher Docteur, je tiens à te remercier pour ces supers moments de joie, de folie et de complicité. Tu as été mon véritable pivot durant les 3 premières années de ma thèse. Ton absence au laboratoire durant cette quatrième année m'a beaucoup pesé je dois l'avouer. Je te souhaite avec Jérémy de belles destinations lors de tes prochains post-doctorats et surtout une bague et un petit garçon !!!

Manon, Marion, Katell, Lisa, Anissa, Blandine, Marine, je tiens à vous remercier beaucoup pour votre joie de vivre communicative, votre aide aux mots fléchés, nos discussions sur la terrasse et pour les soirées. Contentée d'être la championne de course en chaise en pièce à tri, qui va donc prendre mon titre l'année prochaine ? Les filles, vous êtes juste superbes !!! Je vous souhaite plein de bonnes choses pour la suite et j'espère garder contact avec vous !!!

Un grand merci également à ma famille pour son soutien sans faille et ses encouragements. Votre voyage en Auvergne pour l'occasion me fait plaisir au plus haut point !!!

Je remercie profondément mes parents qui ont toujours été là pour moi dans les bons comme dans les mauvais moments durant ma thèse, mais également au cours de ma vie. Votre amour, votre affection, vos encouragements et votre soutien quotidien m'ont été d'un grand réconfort. Même si vous n'étiez pas du domaine, vous avez essayé de vous investir, à votre mesure, dans la réussite de ce projet. Un grand merci à vous !

Enfin je tiens à remercier Cédric, mon mari pour ne pas avoir demandé le divorce durant ces derniers mois. Plus sérieusement, je te remercie du fond du cœur d'avoir toujours veillé sur moi et d'avoir pris soin de moi. Durant ma thèse, tu as su être présent pour moi dans les moments très difficiles et de douleurs qui sont advenus. Les gens nous comparent souvent à des inséparables, et je pense qu'ils ont bien raison. Sans toi, je n'aurais jamais eu autant de force pour me battre durant ces quatre années.

Un grand merci à tous ceux qui m'ont aidé de près ou de loin, à leur manière, durant mon travail de thèse.



# Sommaire

<b><i>Introduction</i></b> .....	<b>1</b>
<b>Chapitre 1 : Les Eléments Transposables</b> .....	<b>1</b>
<i>A. Barbara McClintock, la pionnière des éléments transposables</i> .....	<b>1</b>
<i>B. La Classification des éléments transposables</i> .....	<b>4</b>
<i>C. L'impact génomique des éléments transposables</i> .....	<b>7</b>
<b>Chapitre 2 : Les acteurs clés de la régulation des éléments transposables</b> ..	<b>13</b>
<i>A. La grande famille des protéines ARGONAUTES</i> .....	<b>14</b>
<i>B. Les endosiRNAs</i> .....	<b>15</b>
<i>C. Les piRNAs</i> .....	<b>17</b>
<b>Chapitre 3 : La biogenèse des piRNAs</b> .....	<b>21</b>
<i>A. Organismes modèles utilisés pour explorer la voie des piRNAs</i> .....	<b>21</b>
<i>B. La biogenèse des piRNAs primaires</i> .....	<b>23</b>
<i>C. L'amplification du pool de piRNAs dans les cellules germinales</i> .....	<b>32</b>
<i>D. La production de piRNAs en dehors des gonades</i> .....	<b>33</b>



Chapitre 4 : Les différentes fonctions biologiques des piRNAs.....	35
A. La régulation des ETs.....	36
B. La régulation des gènes.....	38
C. La réorganisation du génome.....	41
D. La défense virale.....	42
Chapitre 5 : Le modèle animal.....	43
A. La drosophile.....	43
B. Les gonades chez la drosophile.....	44
C. Modèle cellulaire, les cellules OSS.....	47
Chapitre 6 : Objectifs de l'étude.....	48
<b>Résultats.....</b>	<b>52</b>
Partie 1 : L'adressage des transcrits de <i>flamenco</i> à la voie des piRNAs....	53
Partie 2 : Expression d'un cluster de piRNAs au cours du développement de la drosophile.....	74
Partie 3 : Les piRNAs de <i>flamenco</i> , régulateurs géniques ?.....	92
<b>Discussion Générale.....</b>	<b>103</b>
<b>Références</b>	





# *Introduction*



# ***Introduction***

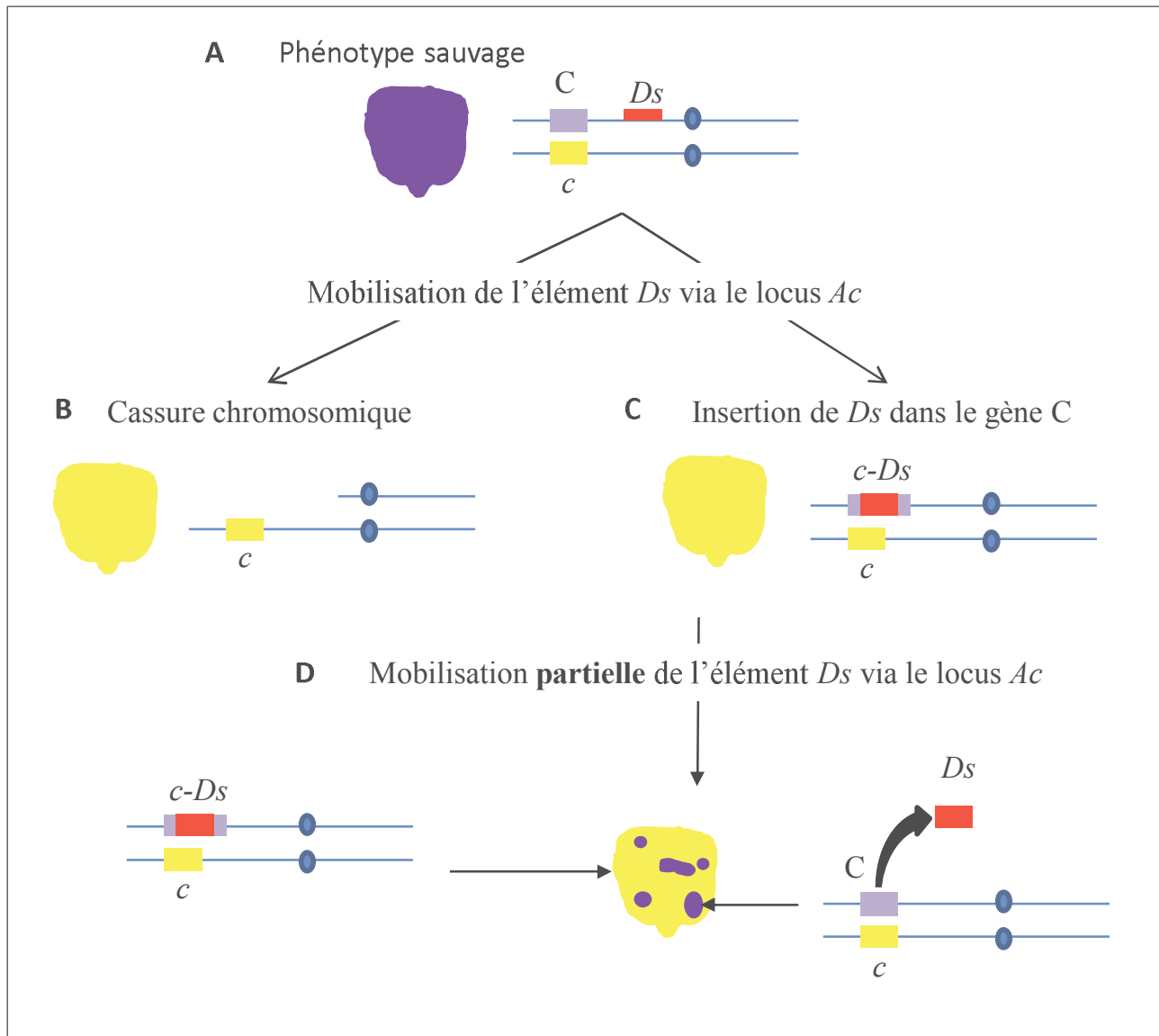
L'émergence du séquençage à haut débit ainsi que le développement de la bio-informatique ont permis de mettre en évidence que les génomes eucaryotes étaient majoritairement constitués de séquences d'ADN non codant puisque les transcrits issus de ces séquences n'étaient pas traduits en protéines. À titre d'exemple, la quasi-totalité du génome humain (soit environ 98%) est composée de séquences d'ADN non codant (Lander et al., 2001). Les Éléments Transposables (ETs), qui sont des séquences moyennement répétées et dispersées dans le génome, représentent plus de 70% des séquences d'ADN non codant chez l'Homme (De Koning et al., 2011). Présents chez tous les eucaryotes séquencés à ce jour, les ETs sont donc les constituants majeurs des génomes eucaryotes.

## **Chapitre 1 : Les Éléments Transposables**

Les ETs sont communément définis comme des séquences ADN ayant la capacité de se déplacer au sein d'un génome hôte *via* un mécanisme appelé la transposition (Haren et al., 1999 ; Kidwell et Lisch, 2001). En raison de leur insertion aléatoire, les ETs peuvent générer des mutations lorsqu'ils ne sont pas contrôlés par la cellule. Par ailleurs, ils sont maintenant reconnus comme étant importants pour la plasticité des génomes et leur évolution.

### ***A. Barbara McClintock, la pionnière des éléments transposables***

Les ETs ont été, pour la première fois, mis en évidence dans les années 50 par Barbara McClintock au sein d'une espèce végétale, le maïs. Cette cytogénéticienne américaine avait préalablement décrit, chez cette même espèce, les phénomènes de crossing-over, facilement observables en microscopie (Creighton et McClintock, 1931). Par la suite, elle étudia la transmission de la coloration des grains de maïs *via* des observations phénotypiques couplées à des analyses cytogénétiques. La mise en place de la coloration des grains de maïs implique notamment l'allèle *C* et le locus *Ds*. L'allèle *C* permet l'expression d'un pigment violet au niveau de la couche à aleurone alors que le locus *Ds* est un site fréquent de cassure du chromosome 9. En croisant des mutants de ces différents loci, des phénotypes inattendus ont été obtenus. En effet, quand l'allèle *C* est muté (allèle *c*), les grains de maïs sont jaunes. Cependant, elle observa des grains de maïs jaunes qui possédaient



**Figure 1 : Le système *Ds/Ac* chez le maïs décrit par Barbara McClintock.**

La coloration des grains de maïs dépend principalement du gène *C* (cadre violet) et de l'ET *Ds* (rectangle rouge). Les grains de maïs sont violets à l'état sauvage (A). Lors de la mobilisation de l'ET *Ds* aidée par l'ET *Ac*, la coloration des grains de maïs devient alors jaune. En effet, la mobilisation de l'ET *Ds* peut soit générer une cassure chromosomique avec perte du fragment portant le gène *C* (B), soit altérer le gène *C* en s'insérant dans ce dernier (C).

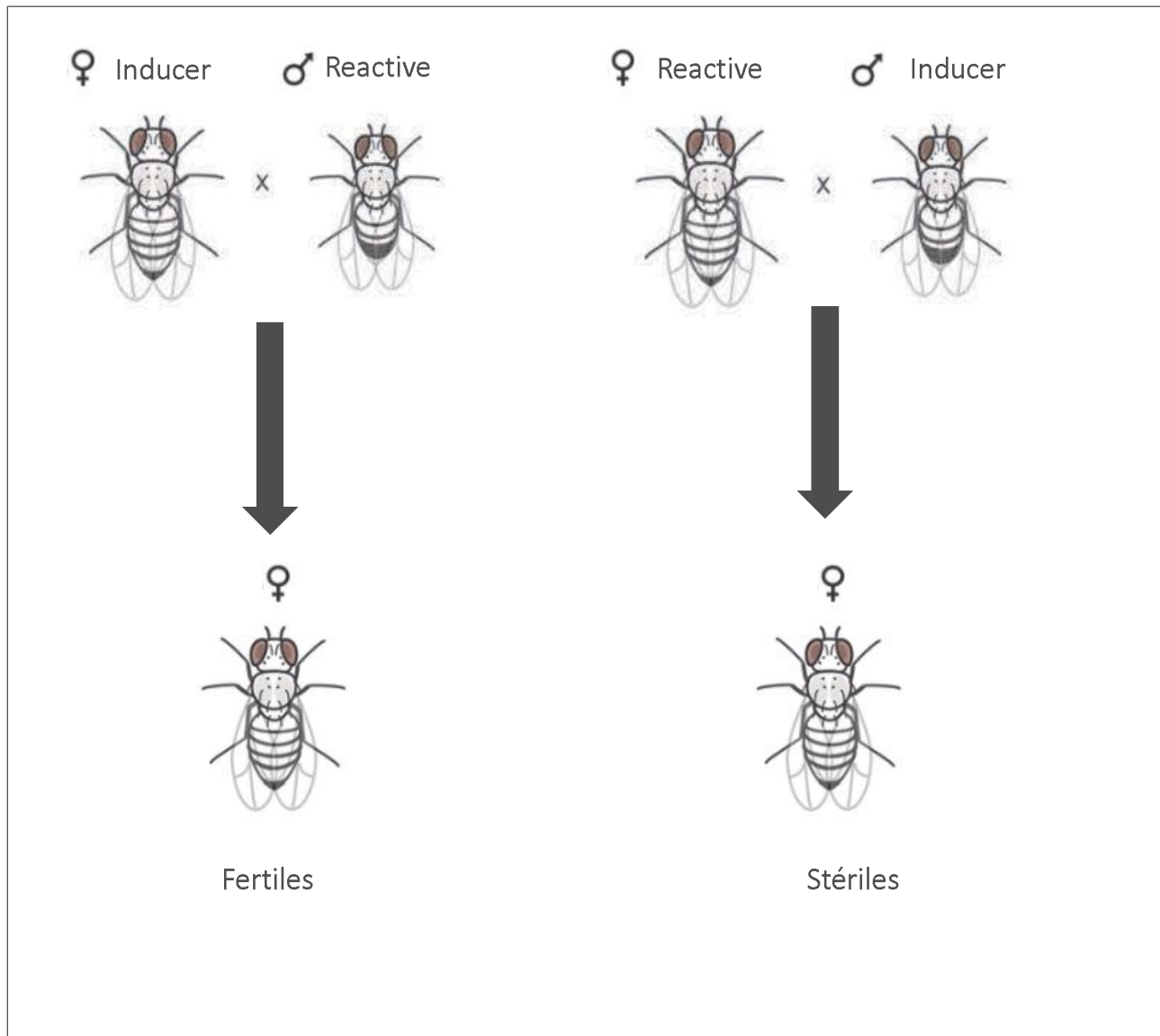
La présence de tâches violettes sur certains grains de maïs jaunes est due à une nouvelle mobilisation de l'élément *Ds* (D). La mutation du gène *C* est alors reversée permettant de reformer un gène *C* sauvage.

également des tâches violettes. La présence de ces tâches violettes était le reflet de mutations instables.

Après de nombreuses études génétiques et moléculaires, Barbara McClintock avait fait l'hypothèse que le locus *Ds* pouvait se déplacer de son emplacement chromosomique initial pour s'insérer dans l'allèle *C* (Figure 1). Cette insertion a pour effet de muter le gène *C*, et donc d'altérer son expression. La conséquence phénotypique de cette insertion est l'obtention d'un grain de maïs de couleur jaune. Cependant, des tâches violettes étaient observables sur les grains de maïs jaunes suggérant que, de façon clonale, cette mutation était réversible. Barbara McClintock comprit alors que la réapparition de la coloration violette dans certaines cellules de la couche à aleurone était due une nouvelle mobilisation du locus *Ds*, à un autre site génomique. En d'autres termes, l'excision du locus *Ds* dans certaines cellules de la couche à aleurone du grain de maïs a pour conséquence de reformer l'allèle *C* sauvage dans ces cellules. L'expression de l'allèle *C* était alors de nouveau rendue possible dans ces cellules, d'où la coloration violette de ces dernières, alors que les autres cellules qui composent la couche à aleurone restaient, quant à elles, jaunes puisqu'il n'y avait pas eu de nouvelle mobilisation du locus *Ds*.

En analysant de plus près ce phénomène, elle s'est aperçue que le locus *Ds* avait besoin d'un autre site génomique qu'elle appela le locus "activateur" ou encore "locus *Ac*". Elle décrit alors le système *Ds/Ac* ainsi que le mécanisme de transposition des séquences d'ADN mobiles chez le maïs. Par ailleurs, elle nomma ces séquences d'ADN mobiles les « gènes sauteurs » ou encore les « éléments de contrôle ».

Néanmoins, les confrères généticiens de Barbara McClintock ne reconnurent pas ses travaux puisqu'ils n'étaient pas cohérents avec la vision de la génétique de l'époque. Dans les années 1950, l'ADN était considéré comme une molécule stable dont l'information génétique transmise à la descendance était figée. Les découvertes de Barbara McClintock bouleversèrent le monde de la génétique puisqu'elle avait pu démontrer que les « gènes » n'étaient pas immobiles à un locus donné. Elle prouva également qu'une mutation pouvait être réversible, ce qui contredisait la définition d'une mutation établie dans les années 1950, à savoir qu'un gène était inactivé de façon définitive à partir du moment où il était muté. Néanmoins, l'ensemble des travaux de Barbara McClintock suggéraient que l'ADN possédait un certain degré de variabilité, mais aussi que l'information génétique pouvait être modulée et réorganisée *via* ces « éléments de contrôle ».



**Figure 2: La dysgénésie des hybrides.**

Lorsqu'une femelle possédant l'ET I (dite « Inducer ») est croisée avec un mâle n'ayant pas cet ET (nommé « Reactive »), la descendance est fertile. Lors du croisement réciproque d'une femelle Reactive avec un mâle Inducer, la descendance présente la caractéristique d'être stérile. La stérilité des femelles issues du croisement précédent est en grande partie due à l'absence de régulation de l'ET I.

## I. Les Génomes et les éléments transposables

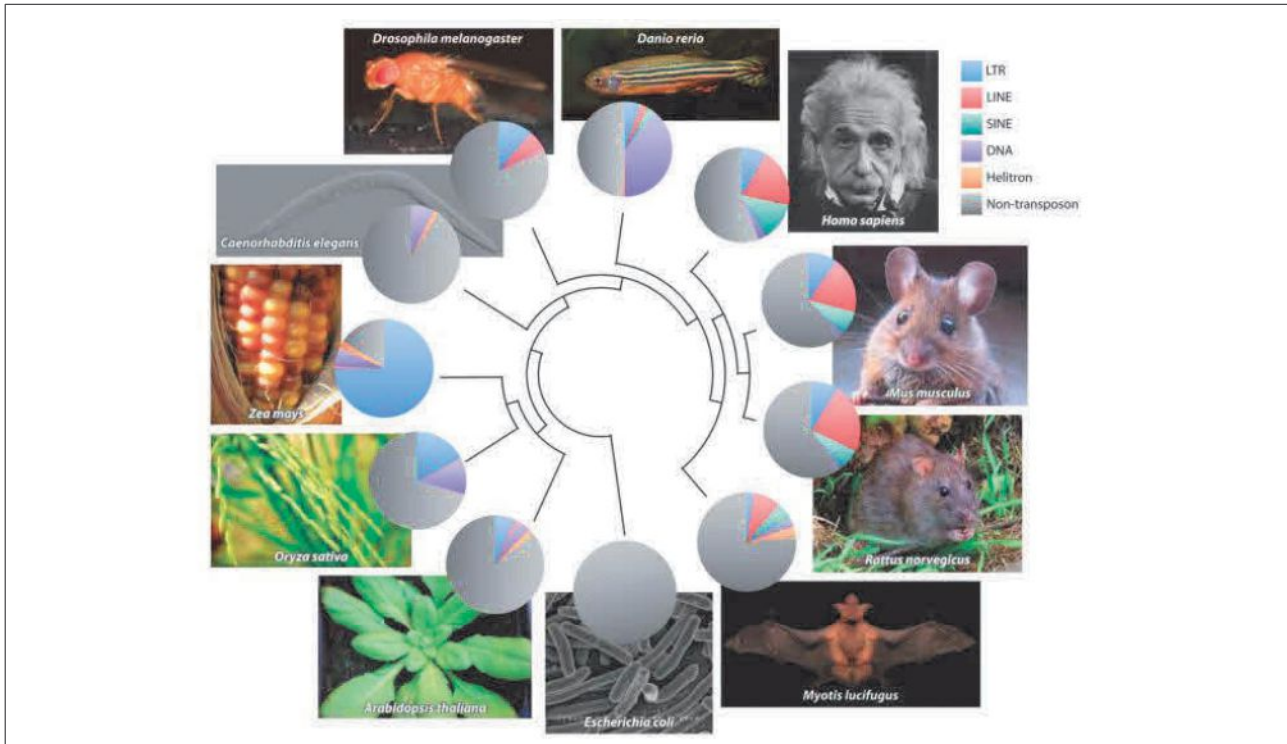
Il faudra attendre une dizaine d'années après les travaux de Barbara McClintock pour que d'autres séquences d'ADN mobiles soient également découvertes chez des procaryotes. Chez la bactérie *E. Coli*, les séquences IS furent identifiées comme étant responsables de la mutation de l'opéron *galactose* par James Alan Shapiro en 1969 (Shapiro et al., 1969). Puis en 1971, une autre séquence mobile, le transposon *Tn1*, a été identifiée comme conférant aux *Pseudomonas aeruginosa* une résistance à l'ampicilline (Datta et al., 1971). Dès lors, un certain nombre d'ETs furent découverts et caractérisés au sein des génomes procaryotes.

Quelques années plus tard (1976) à Clermont-Ferrand, le généticien Georges Picard observa l'apparition d'une stérilité chez les femelles issues d'un croisement génétique entre deux populations de drosophiles : les mouches dites «Inducer» (I) et les mouches dites «Reactive» (R) (Picard et al., 1976). De façon surprenante, les femelles issues du croisement d'une femelle R avec un mâle I étaient stériles, alors que les femelles issues du croisement réciproque sont, quant à elles, fertiles (Figure 2). Georges Picard trouva alors que l'apparition de la stérilité dans la descendance du croisement d'une femelle R avec un mâle I était due à la présence d'une séquence d'ADN mobile qu'il baptisa le facteur I. Il décrivit alors la dysgénésie des hybrides, c'est-à-dire l'incompatibilité génétique entre deux souches de drosophiles due à un polymorphisme génétique. En effet, le facteur I a été identifié comme étant un ET présent chez les mouches I, et à contrario, absent chez les mouches R. Un an après cette découverte clermontoise, un autre exemple de dysgénésie des hybrides chez la drosophile, le système P-M, a été décrit par Kidwell et ses collègues. Ce système est également basé sur un polymorphisme génétique entre deux souches de drosophiles lié à la présence/absence d'un autre ET, appelé l'élément P (Kidwell et al., 1977).

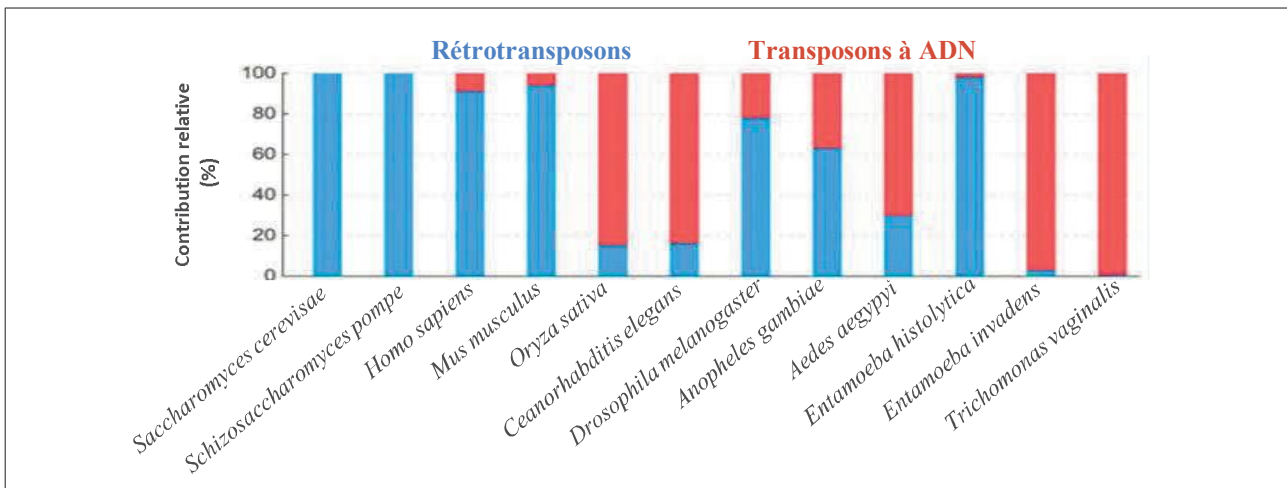
L'ensemble de ces découvertes a permis de réévaluer et d'appuyer les précédents résultats de Barbara McClintock. En revanche, ce n'est qu'en 1983 que ses travaux furent reconnus par ses pairs qui lui décernèrent le prix Nobel de physiologie et médecine.

Ultérieurement, différents projets de séquençage des génomes procaryotes et eucaryotes furent entrepris dans les années 90. Ces travaux ont notamment permis de mettre en évidence la présence universelle des ETs dans les génomes de toutes les espèces séquencées, à l'exception de la microsporidie *Enterocytozoon bieneusi* (Parisot et al., 2014).

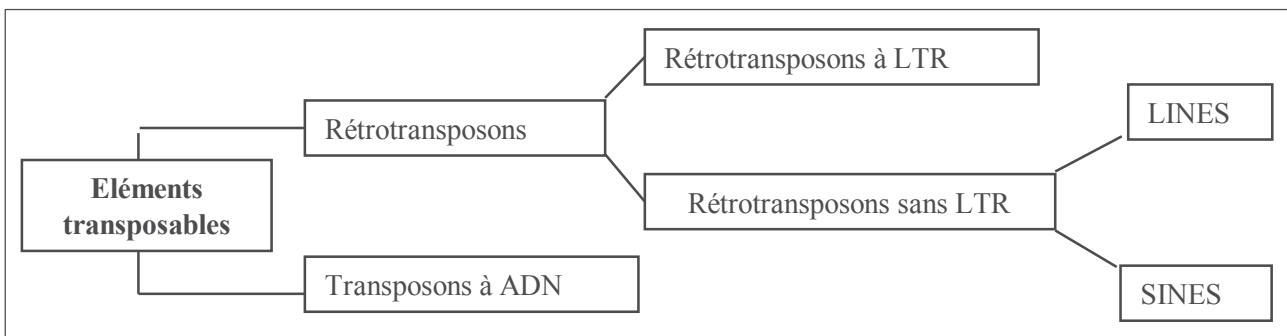




**Figure 3: Proportion génomique des ETs en fonction de l'espèce considérée.**  
 Figure adaptée de la figure 3 de l'article Huang et al., 2012



**Figure 4: Contribution relative des différentes classes d'ETs dans divers génomes eucaryotes.**  
 Figure adaptée de la figure 1 dans l'article Pray, 2008.



**Figure 5: Classification des ETs proposée par Finnegan en 1989.**

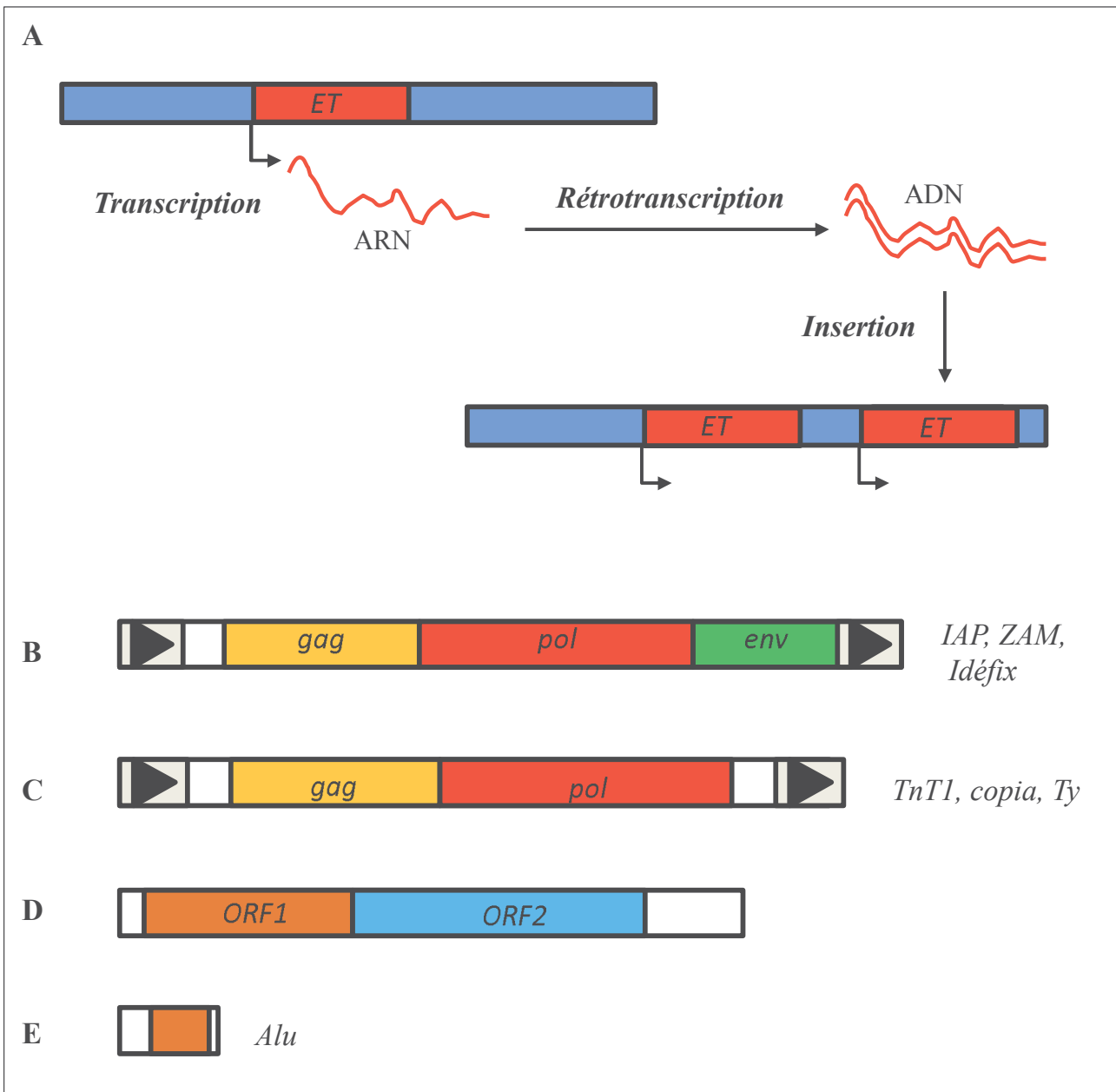
Ces données ont également révélé que les ETs étaient des constituants majeurs des génomes eucaryotes puisqu'ils constituent une fraction plus ou moins importante de ces derniers (Figure 3). Par exemple, les ETs représentent approximativement 12% du génome de *C. elegans* (*C. elegans* Sequencing Consortium, 1998), 15% du génome d'*Arabidopsis Thaliana* (De la Chaux et al., 2012), 15 à 20% du génome de *Drosophila* (Quesneville et al., 2005), 40% du génome murin, 65% du génome humain (Adrion et al., 2017 ; De Koning et al., 2011), plus de 77% du génome de *Pelophylax esculentus*, plus de 80% du génome du maïs (Schnable et al., 2009). Cependant, les proportions annoncées précédemment semblent être sous-estimées puisque l'émergence d'outils bio-informatiques de plus en plus performants permet l'identification et l'annotation de nouveaux ETs au sein de ces mêmes espèces.

La découverte des ETs au sein des différentes espèces étudiées nécessita de trouver une façon harmonieuse de les répertorier. En effet, les ETs partagent la capacité de se mouvoir dans les génomes, mais pour y parvenir, ils utilisent des mécanismes différents.

### ***B. La Classification des éléments transposables***

En 1989, David J. Finnegan proposa de les classer en fonction de leur mode de transposition (Finnegan, 1992 ; Finnegan, 1989). Grâce à ce paramètre, les ETs ont été scindés en deux grandes catégories : d'une part, les éléments mobiles de la classe I dont la transposition est effectuée selon un mécanisme de « copier-coller » *via* un intermédiaire ARN, et d'autre part, les ETs de la classe II se déplaçant sous forme de fragment ADN selon un mode "couper-coller" (Figures 4 et 5). Les ETs présents dans chacune de ces deux classes sont ensuite subdivisés en sous-groupes en fonction de leur structure.

Néanmoins, la découverte de nouveaux ETs tels que les MITES (ETs à répétitions inversées miniatures) présents dans les génomes des différents règnes du vivant excepté chez les protistes, a nécessité de réactualiser la classification des ETs initialement proposée par Finnegan. Récemment, Wicker et ses collègues ainsi que Kapitonov et Jurka ont proposé d'intégrer de nouveaux critères afin d'affiner la classification des ETs (Wicker et al., 2007 ; Kapitonov et Jurka, 2008). L'organisation des domaines fonctionnels des ETs, l'homologie de séquence avec des ETs fonctionnels ainsi que le site d'insertion génomique des ETs nouvellement identifiés ont été pris en compte pour l'optimisation de la classification de ces derniers. Grâce à ces nouveaux critères, de nombreux ETs non fonctionnels tels que les MITES ont pu être classés.



**Figure 6 : Mode de transposition des rétrotransposons et leurs structures.**

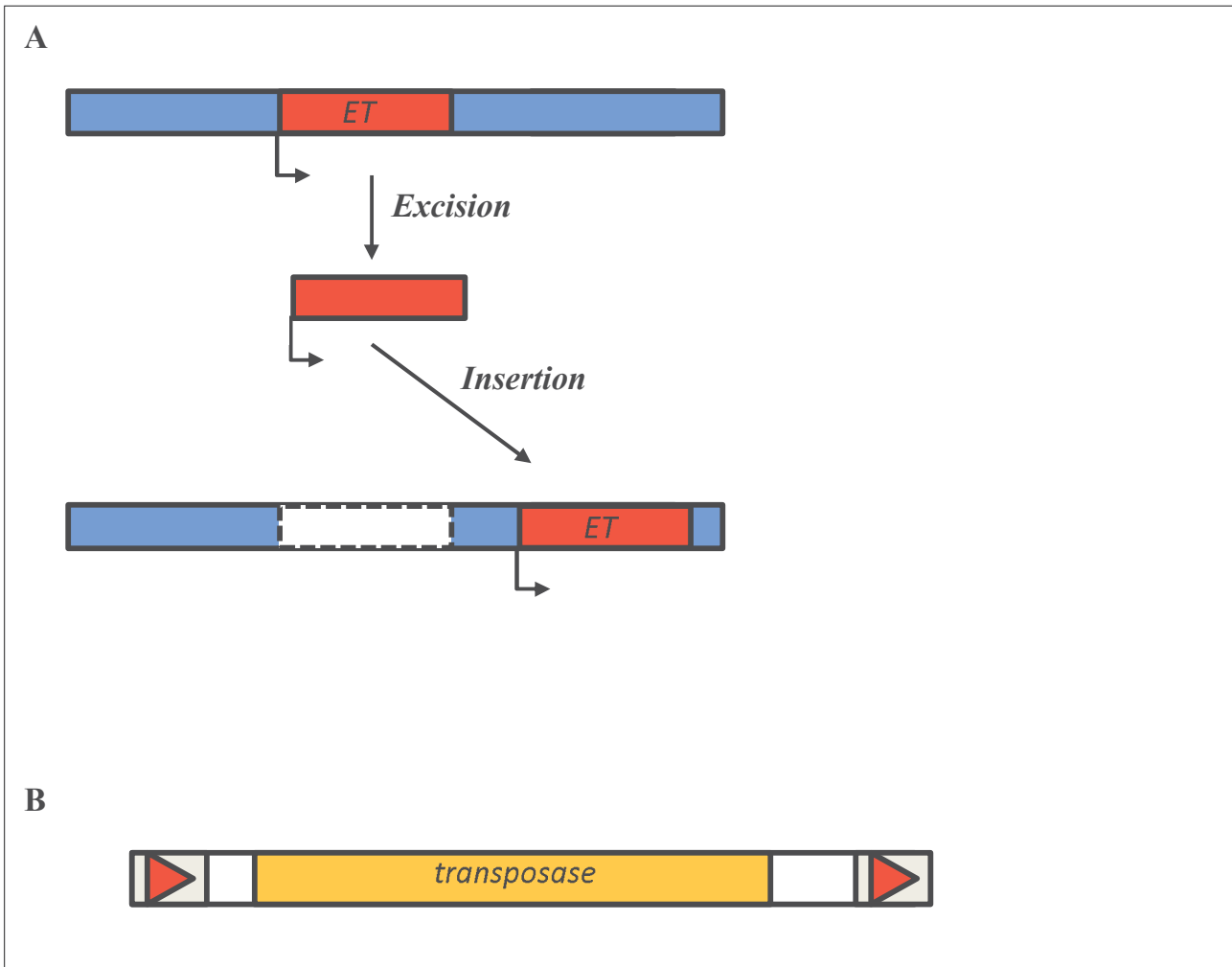
- A) Les rétrotransposons transposent sur un mode « copier-coller » grâce à un intermédiaire ARN.
- B) Structure des rétrotransposons à LTR composés du cadre de lecture *gag* (orange), du cadre de lecture *pol* (rouge) et du cadre de lecture *env* (vert). De longues répétitions terminales (LTR) (triangles noirs) encadrent ces séquences codantes.
- C) Structure des rétrotransposons à LTR composés du cadre de lecture *gag* (orange), du cadre de lecture *pol* (rouge). De longues répétitions terminales (LTR) (triangles noirs) encadrent ces séquences codantes.
- D) Structure des rétrotransposons non LTR de type LINES composés de deux cadres de lecture : ORF1 en marron et ORF2 en bleu.
- E) Structure des rétrotransposons non LTR de type SINES. Ces éléments ne sont pas fonctionnels.

Cependant, la classification des ETs présentée dans ce manuscrit est celle datant de 1989 puisque le projet d'amélioration de la classification des ETs est encore en cours. En effet, la classification de Wicker ne prend en compte que les ETs présents chez les eucaryotes. De plus, cette classification pourrait être rapidement modifiée puisque de nombreux ETs, entiers ou tronqués, sont encore actuellement découverts lors des analyses *in silico*.

### I. Les éléments de Classe I

Les éléments de classe I, ou rétrotransposons, se déplacent selon un mode de transposition dite « réplivative ». En effet, les rétrotransposons transposent *via* un intermédiaire ARN tout en laissant la copie initiale à sa position de départ (Figure 6A). L'intermédiaire ARN de l'ET est rétrotranscrit en ADN complémentaire (ADNc), puis il est intégré à un autre site génomique. Grâce à ce mode de transposition, les rétrotransposons ont pu augmenter le nombre de leurs copies génomiques, et ainsi, envahir les génomes. À titre d'exemple, ils représentent à eux seuls approximativement 40% des génomes humains et murins, soit la quasi-totalité des ETs présents dans ces deux espèces (Lander et al., 2001). Les rétrotransposons sont subdivisés en deux sous-groupes en fonction de la présence, ou non, de longues séquences terminales répétées particulières, appelées les «LTRs», mais également, en fonction de leurs mécanismes d'intégration au sein du génome hôte.

Les rétrotransposons, qui possèdent des LTRs à chacune de leurs extrémités, ont une structure très proche de celle des rétrovirus (Figure 6B et 6C). Ils sont notamment tous composés des gènes *pol* et *gag*. Le gène *gag* code des protéines de structures responsables de la formation de pseudo particules virales. Le gène *pol*, quant à lui, code une polyprotéine. Le clivage de la polyprotéine permet l'obtention de plusieurs protéines essentielles à la transposition des rétrotransposons telles que la protéase, l'intégrase, la ribonucléase H (Rnase H) et la reverse transcriptase. La reverse transcriptase permet de synthétiser l'ADNc en utilisant l'ARN messager du rétrotransposon comme matrice. Une fois la reverse transcription achevée, la RNase H dégrade l'ARN matriciel permettant de synthétiser, par la suite, un ADN double brin. Puis, l'ADN double brin est inséré dans le génome grâce à l'action de l'intégrase. Les rétrotransposons à LTRs les plus connus au sein des organismes modèles sont l'élément *TnT1* présent chez les plantes (Grandbastien et al., 1997), les éléments *Ty* retrouvés chez la levure (Williamson, 1983) et les éléments *copia* chez la drosophile.



**Figure 7 : Mode de transposition des transposons à ADN et leurs structures.**

- A) Les transposons transposent de façon directe à de nouveaux sites chromosomiques selon un processus « couper-coller ».
- B) Les transposons à ADN sont constitués d'une transposase (orange) encadrée par des séquences inversées et répétées (ITR) représentées par des triangles rouges.

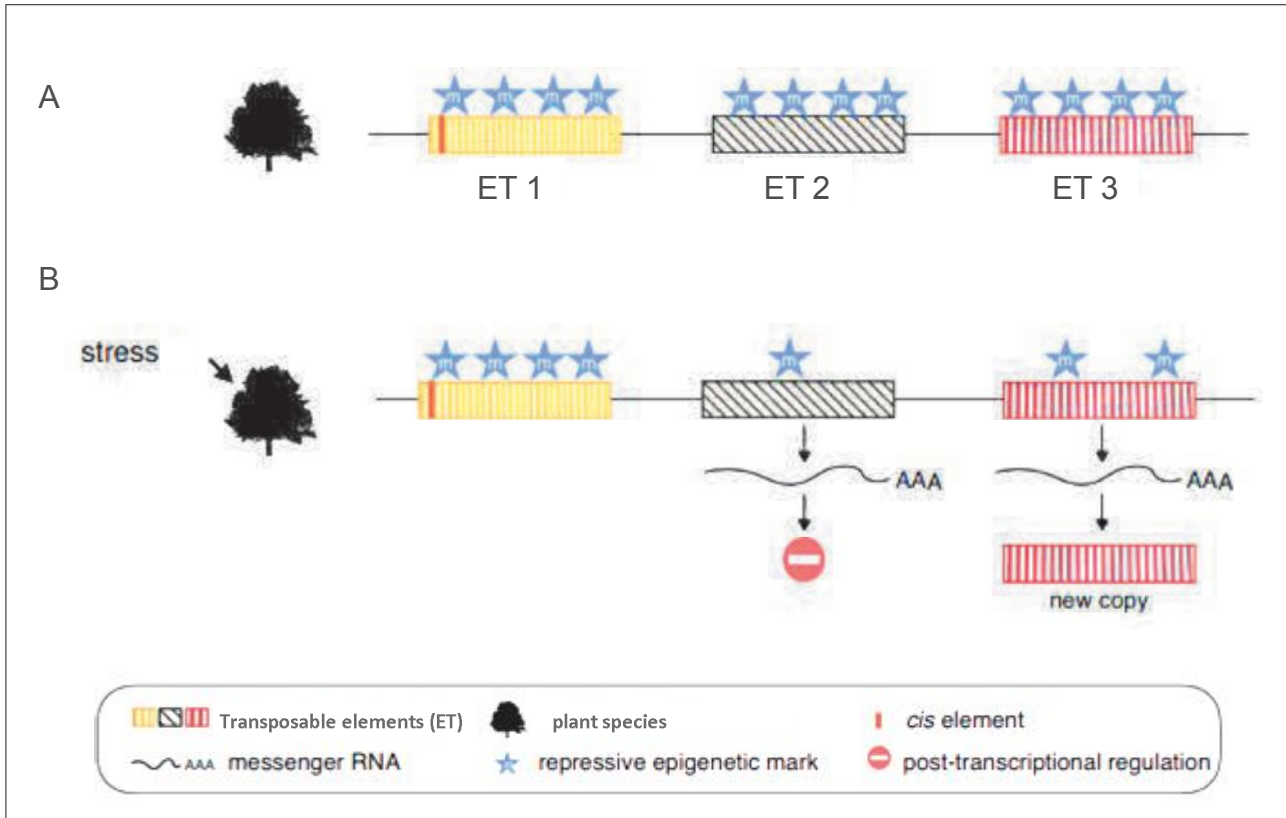
Certains rétrotransposons à LTRs possèdent un troisième cadre de lecture, le gène *env*, qui code une polyprotéine composée de deux sous-unités : une protéine transmembranaire et une glycoprotéine de surface impliquée dans la reconnaissance des récepteurs cellulaires mise en jeu lors d'une infection virale (Lerat et Capy et al., 1999 ; Leblanc et al., 1997 ; Pélisson et al., 1994). Les rétrotransposons à LTRs possédant le cadre de lecture *env* les plus connus au sein des organismes modèles sont les éléments *IAP* présents chez la souris ainsi que les éléments *Gypsy*, *ZAM* et *Idéfix* qui sont, quant à eux, retrouvés chez la drosophile (Reuss et al., 1996 ; Leblanc et al., 1997 ; Pélisson et al., 1994 ; Desset et al., 2003).

Les rétrotransposons n'ayant pas de séquences LTR, comme les LINEs et les SINEs, sont constitués à leur extrémité 3' de séquences riches en adénosine. Certains de ces ETs sont encore potentiellement fonctionnels comme chez l'Homme (Brouha et al., 2003 ; Deininger, 2011) et chez la drosophile. Les LINEs peuvent atteindre 4 à 7 kilobases et sont constitués des cadres de lecture codant une reverse transcriptase et une DNA endonucléase nécessaire à leur transposition (Figure 6D). Contrairement aux LINEs, les SINEs sont de taille plus modeste variant de 80 à 500 pb. Cette différence de taille est notamment expliquée par le fait que les SINEs ne possèdent pas le cadre de lecture codant la reverse transcriptase (Figure 6E). Par conséquent, la transposition des SINEs nécessite l'utilisation de la machinerie de rétrotransposition d'un autre ET. Le SINE le plus connu dans le génome humain est l'élément *Alu*.

## II. Les éléments de classe II

Les éléments de classe II sont appelés des transposons ou encore des ETs à ADN. Ils se déplacent selon une transposition dite « conservatrice » puisque la séquence ADN de l'ET est excisée, déplacée puis intégrée à un nouveau site génomique (Figure 7A). Ils représentent à eux seuls 3% du génome humain (Hellen et Brookfield, 2013). Néanmoins, l'augmentation du nombre de copies des éléments de classe II est rendue possible lorsque l'excision du transposon se produit suite à la réplication et que la néo-insertion de ce dernier est réalisée avant le passage de la fourche de réplication. C'est ainsi que, dans les années 1960, l'élément P a envahi le génome de *D. melanogaster* à partir d'une autre espèce, *D. willistoni*, dans un foyer d'Amérique centrale où ces deux espèces cohabitent (Daniels et al., 1990).

Les éléments de classe II sont composés à leurs extrémités de séquences terminales spécifiques appelées ITRs pour "internal terminal repeats" (Figure 7B). Les



**Figure 8 : Mobilisation des ETs à la suite d'un stress chez les végétaux.**

Exemple de trois ETs réprimés (ET 1, ET 2 et ET 3) par des marques épigénétiques telles que la méthylation de l'ADN dans des conditions normales chez un végétal (A). Lors d'un stress (biotique ou abiotique), les marques épigénétiques régulant l'ET 2 et l'ET 3 sont altérées (B). L'altération des marques épigénétiques permet alors leurs expressions. Les ETs peuvent donc se mobiliser afin de générer des mutations apportant un avantage sélectif face à cette nouvelle situation (cas de l'ET 3). Néanmoins, suite à leurs expressions, certains ETs sont régulés post-transcriptionnellement (cas de l'ET 2).

Figure adaptée de la figure 1 de l'article Lanciano et Mirouze, 2018.

séquences ITRs flanquent un cadre de lecture codant une enzyme, la transposase. Cette enzyme présente la particularité de reconnaître les séquences ITRs. La transcription de ces éléments permet l'obtention d'un ARN messenger qui est, par la suite, traduit au niveau du cytoplasme en protéine, la transposase. Cette enzyme est alors importée dans le noyau où elle reconnaîtra les séquences ITRs. Une fois les séquences ITRs reconnues par l'enzyme, l'ADN double brin est coupé à leur niveau. L'élément ainsi excisé va alors s'intégrer *via* cette même enzyme à un autre site chromosomique. Les éléments *Ds* et *Ac* du maïs décrit par Barbara McClintock appartiennent à cette catégorie d'ETs. Dans la littérature, d'autres transposons ont également été caractérisés tels que les éléments CACTA chez *Arabidopsis thaliana* et l'élément *bandit* chez *Ancylostoma caninum* (Laha et al., 2007).

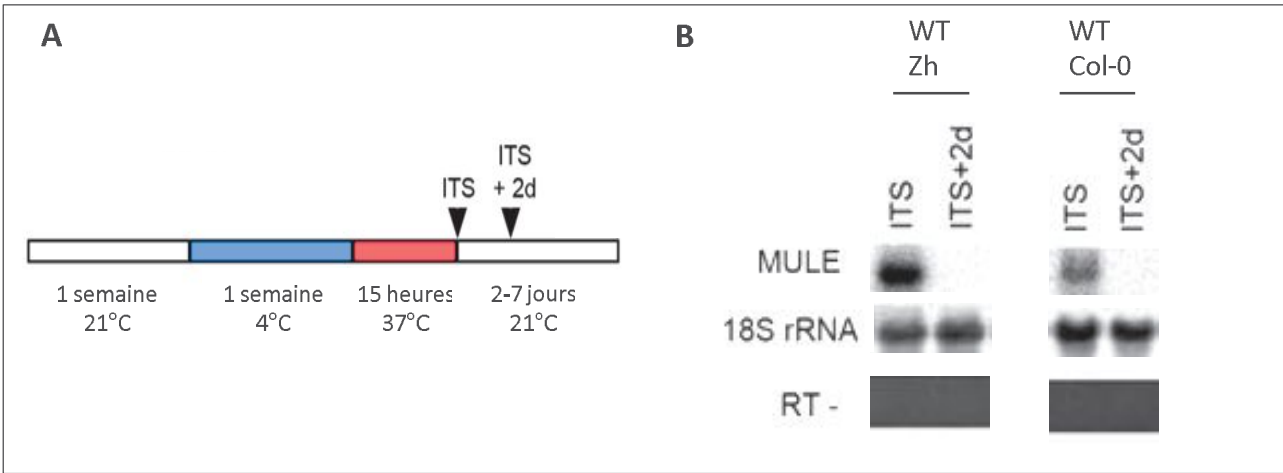
### ***C. L'impact génomique des éléments transposables***

Constituants majeurs des génomes eucaryotes, les ETs peuvent à la fois être perçus comme une source d'innovation génétique, mais aussi, comme une menace pour l'intégrité des génomes.

#### ***I. Les éléments transposables, un atout génétique***

Au cours de l'évolution, une grande partie des ETs a été mutée perdant alors leur faculté à transposer. De plus, les ETs ayant conservé leur capacité de transposition sont généralement mis sous silence par la cellule par différents mécanismes de répression. En d'autres termes, les événements de transposition sont rares. Chez les bactéries et chez les levures, la fréquence moyenne de transposition avoisine celle des mutations spontanées, soit égale à  $10^{-6}$  événements par élément par génération. Chez la drosophile, la fréquence moyenne de transposition est estimée à  $10^{-4}$  événements par élément par génération (Charlesworth et Langley 1989). Néanmoins, dans certaines circonstances telles que l'induction d'un stress, cette fréquence de transposition peut être fortement augmentée. En réponse au stress, l'individu autorise alors la mobilisation des ETs afin de générer des mutations qui lui apportent un avantage sélectif face à cette nouvelle situation. L'hypothèse que les ETs soient des facteurs d'adaptation pouvant être responsables de l'évolution des génomes, et donc des événements de spéciation, avait préalablement été proposée par Barbara McClintock dans les années 80 (McClintock, 1984). De nombreux exemples de mobilisation d'ETs lors de stress biotiques et abiotiques ont été décrits chez les eucaryotes, et plus particulièrement chez les plantes (Figure 8). Ne pouvant fuir le danger, les plantes



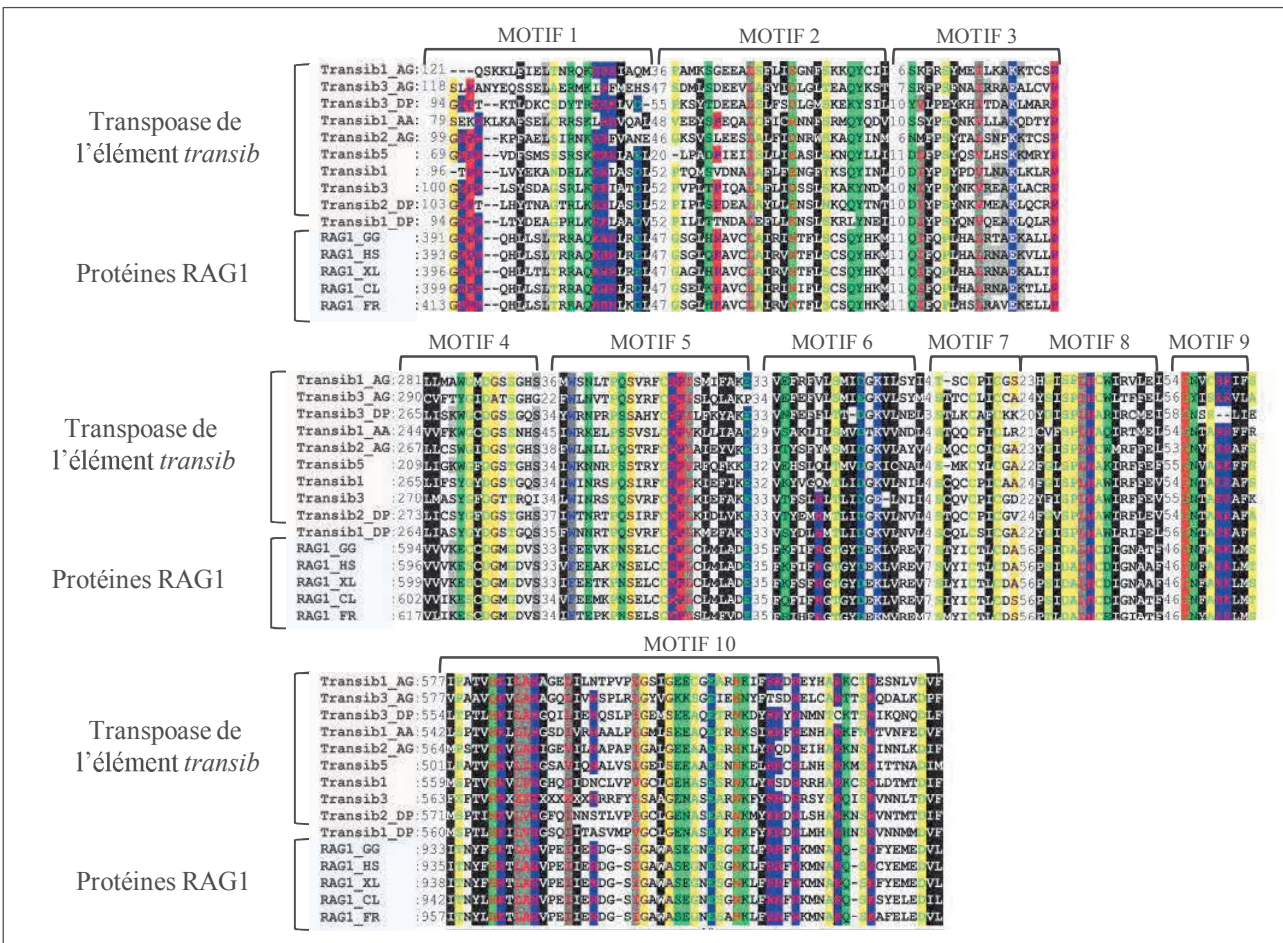


**Figure 9 : Le stress thermique induit l'activation transitoire de la transcription des ETs initialement réprimés chez *A. thaliana*.**

A) Schéma de l'expérience de stress thermique réalisée sur les écotypes Zurich (Zh) et Colombia (Col-0).

B) Northern Blot réalisé à partir des ARNs extraits des plantes Zh ou Col-0 à la suite du stress thermique (ITS) ou 2 jours après le stress thermique (ITS+2d). Détection des transcrits de l'ET MULE à la suite d'un stress thermique.

Figure adaptée des figures 2A et 3 de l'article Tittel-Elmer et al., 2010.



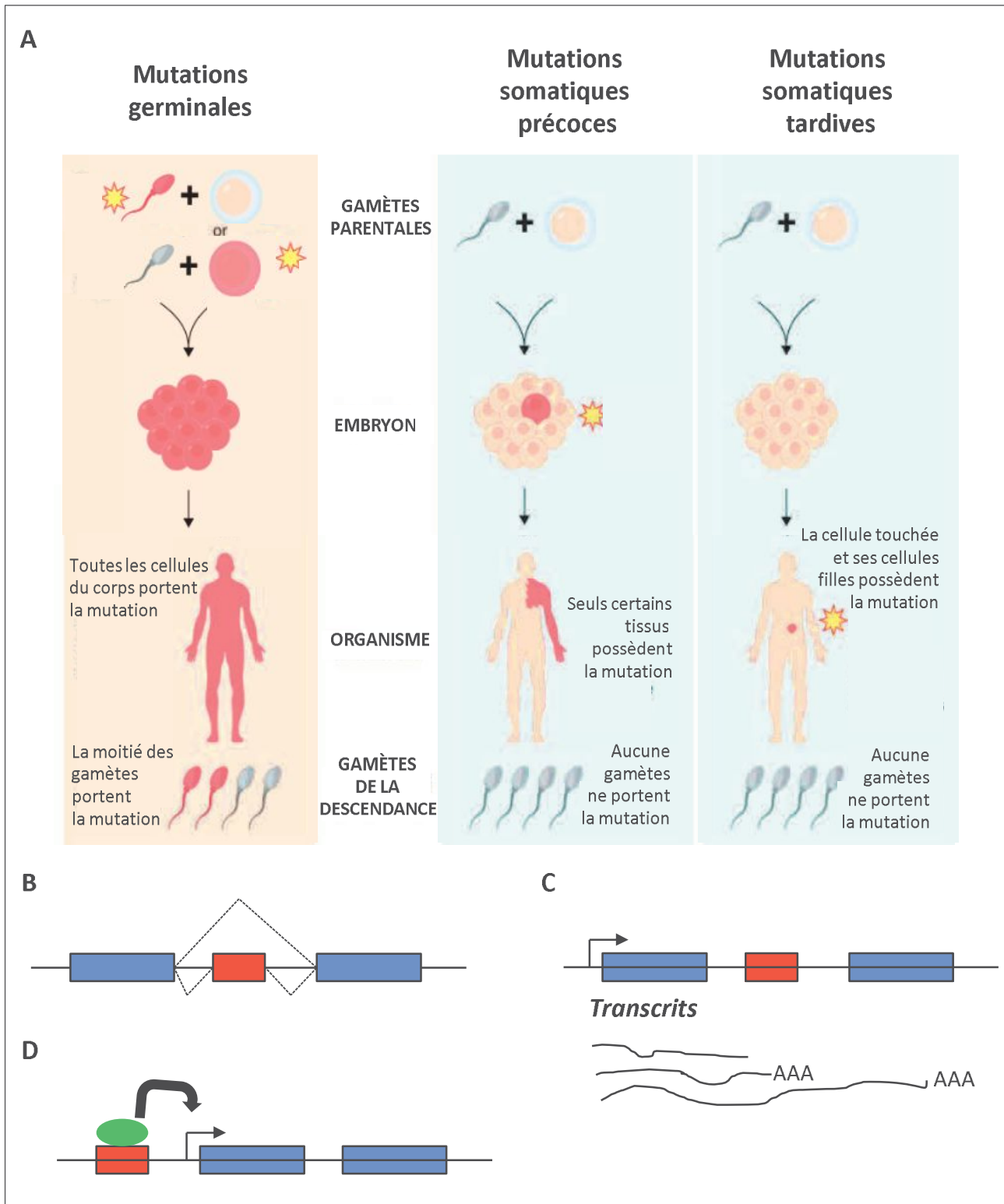
**Figure 10 : Alignement multiple de dix motifs conservés dans les protéines RAG1 et dans les transposases des ETs *Transib*.**

Figure adaptée de la figure 3 de l'article Kapitonov et Jurka, 2005.

autoriseraient la mobilisation des ETs afin de s'adapter rapidement à leur nouvel environnement. Les dérégulations d'ETs les plus connues sont celle de l'élément *Tnt1* chez le tabac dont l'expression peut être induite à la suite de stress biotique et abiotique (Beguiristain et al., 2001), celle de l'élément *mPing* chez le riz dont l'expression est activée en réponse au froid et à un stress salin (Yasuda et al., 2013), ou encore celles des éléments MULE et ONSEN chez *Arabidopsis thaliana* dont la transcription est activée lors d'un stress thermique (Figure 9) (Cavrak et al., 2014 ; Pecinka et al., 2010 ; Tittel-Elmer et al., 2010).

Par ailleurs, de nombreux gènes ont émergé chez les végétaux ainsi que chez les animaux grâce au processus de domestication des ETs (Volf, 2006). Ce processus moléculaire correspond à la création de néogènes à partir d'une séquence dérivant d'un élément transposable. De nombreux exemples de domestication d'ETs ont été décrits dans la littérature chez les eucaryotes, et plus particulièrement chez les mammifères. L'une des plus anciennes domestications d'ETs est celle de la transposase de l'élément *Transib* qui aurait eu lieu au sein du groupe monophylétique des Gnathostomes, il y a plus de 500 millions d'années (Figure 10) (Van Gent, Mizuuchi, et Gellert, 1996 ; Hiom et Gellert, 1998 ; Kapitonov et Jurka, 2005). La protéine résultante de cette domestication est la protéine RAG1. Cette enzyme joue un rôle primordial dans la recombinaison V(D)J durant le développement des cellules lymphocytaires. En effet, cette recombinase permet la création d'un répertoire infini d'immunoglobulines. La domestication de la transposase de l'élément *transib* est donc à la base du développement de notre système immunitaire. Un autre exemple de domestication fréquemment cité est la domestication du gène rétroviral *env* (issus du rétrovirus endogène HERV-W) à l'origine de l'apparition des mammifères placentaires. Les gènes dérivant de la séquence *env* sont les protéines syncytines. Ces protéines sont responsables notamment de la fusion cellulaire des cellules de la membrane extra-embryonnaire fœtale avec celles du tissu utérin maternel (Mi et al., 2000). Chez la souris, l'absence des protéines syncytines a pour conséquence d'induire un défaut placentaire léthal. L'altération de la formation du placenta chez ces souris est due à l'absence de fusion des cellules trophoblastiques, qui est une étape préalablement requise pour la formation de la membrane extra-embryonnaire fœtale (Dupressoir et al., 2009 ; Dupressoir et al., 2011). Ces expériences suggèrent donc que les syncytines sont des acteurs essentiels et nécessaires au développement du placenta.

Dans d'autres cas, les ETs concourent au maintien de la stabilité des génomes puisqu'ils participent au maintien structurel des chromosomes. Chez les eucaryotes, les télomères, régions d'ADN présentes à chaque extrémité des chromosomes, sont constitués de



**Figure 11: Répercussions engendrées par l'insertion d'un ET.**

A) Conséquences des mutations germinales, des mutations somatiques précoces ou des mutations somatiques tardives induites à la suite de la mobilisation d'un ET.

Figure adaptée de <https://kintalk.org/genetics-101/>.

L'insertion d'un ET dans un gène peut engendrer divers effets comme la modification de l'épissage de transcrits du gène (B), la perturbation de la terminaison de la transcription du gène (apparition de codons stop ou de queues polyA prématurés) (C) ou encore la modification de l'expression du gène (D).

séquences répétées en tandem. Ces extrémités sont maintenues intactes dans les cellules germinales grâce à une enzyme, la télomérase. Cependant, cette structure n'est pas universelle. Chez la drosophile, où aucune télomérase n'est codée par le génome, les extrémités télomériques sont composées de séquences alternées d'ETs : *HETA*, *TART* et *TAHRE* (Biessmann et al., 1992 ; Casacuberta et Pardue, 2005 ; Melnikova, Biessmann et Georgiev, 2005). Dans ce cas précis, la transcription des ETs et leur transposition permet d'assurer les fonctions télomériques préservant ainsi la stabilité du génome de la lignée germinale des drosophiles (Biessman et Madson 1992 ; Levis et al., 1993).

Malgré la découverte de nombreux avantages que confèrent les ETs aux génomes hôtes, ces composants majeurs des génomes restent une menace constante pour ces derniers.

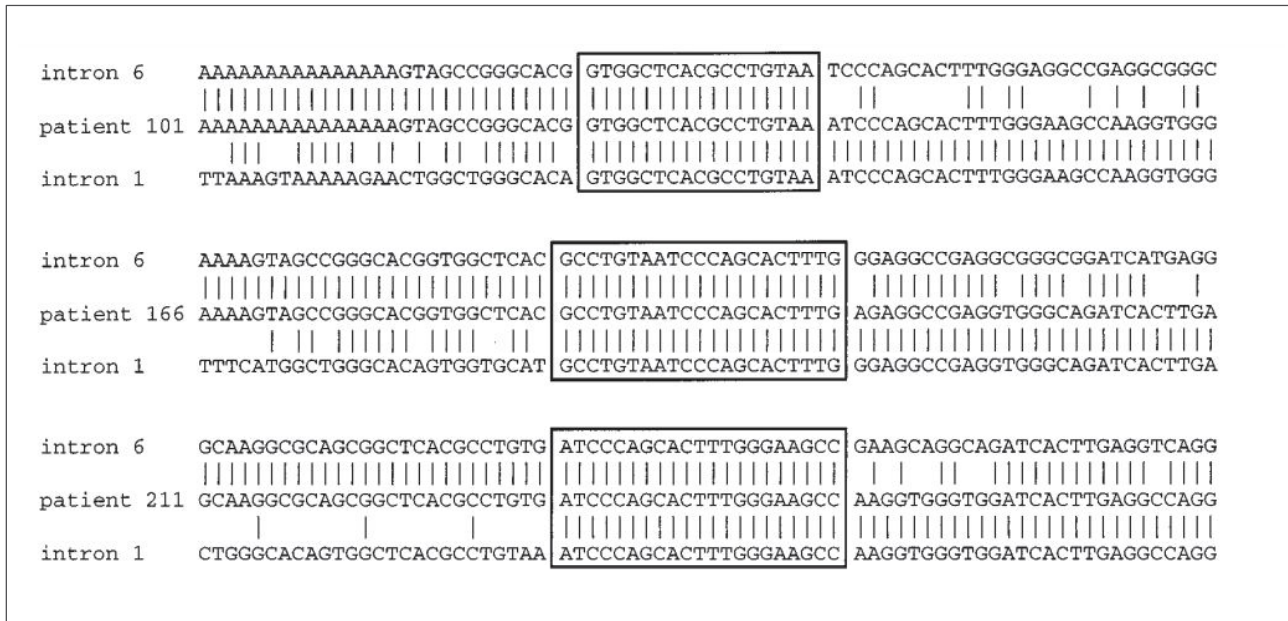
## *II. Les éléments transposables, un facteur d'instabilité génomique*

Les ETs sont des agents mutagènes présents en grand nombre dans les génomes hôtes. Bien qu'ils ne soient pas tous fonctionnels, certains d'entre eux ont gardé leur capacité à se déplacer dans les génomes. A titre d'exemple, le génome humain est composé, environ, d'une centaine d'éléments LINE1 potentiellement fonctionnels (Brouha et al., 2003). Malgré la présence d'un contrôle fin de la régulation des ETs, on estime qu'un individu sur cinquante posséderait une nouvelle insertion de l'élément LINE1 dans son génome. Chez la souris, les insertions des ETs sont responsables d'environ 10% des mutations *de novo* observées (Maksakova et al., 2006).

Les conséquences liées à la transposition aléatoire d'un ET seront plus ou moins délétères en fonction du site d'insertion, mais également en fonction de la cellule touchée (Figure 11). En effet, la transposition d'un ET au cours de la gamétogenèse n'aura pas le même impact que la transposition d'un ET au sein de cellules somatiques au cours du développement de l'individu.

### **a) Conséquences liées à la mobilisation d'un élément transposable au sein des cellules somatiques**

L'insertion aléatoire d'un ET au sein de cellules somatiques peut générer des mutations qui ne seront pas transmises à la descendance. Ces mutations peuvent être retrouvées, de façon clonale, au sein d'un petit groupe de cellules, mais aussi dans toutes les cellules composant un tissu/organe. La variation du nombre de cellules somatiques, dont le



**Figure 12 : Comparaison des séquences du gène *ALL1* chez des patients atteints de leucémie myéloïde aigüe.**

Séquence d'ADN de trois patients atteints de leucémie myéloïde aigüe ayant une duplication en tandem des exons 2 à 6 du gène *ALL1*. Les séquences ADN des patients sont comparées à celles de cellules germinales. Les lignes verticales indiquent les sites d'homologie de nucléotide. Les patients 101, 166 et 211 ont montré des signes de recombinaison entre les éléments *Alu* des introns 6 et 1. Les séquences encadrées indiquent des régions d'homologie chevauchantes, compatibles avec la formation d'hétéroduplexe lors de la recombinaison homologue.

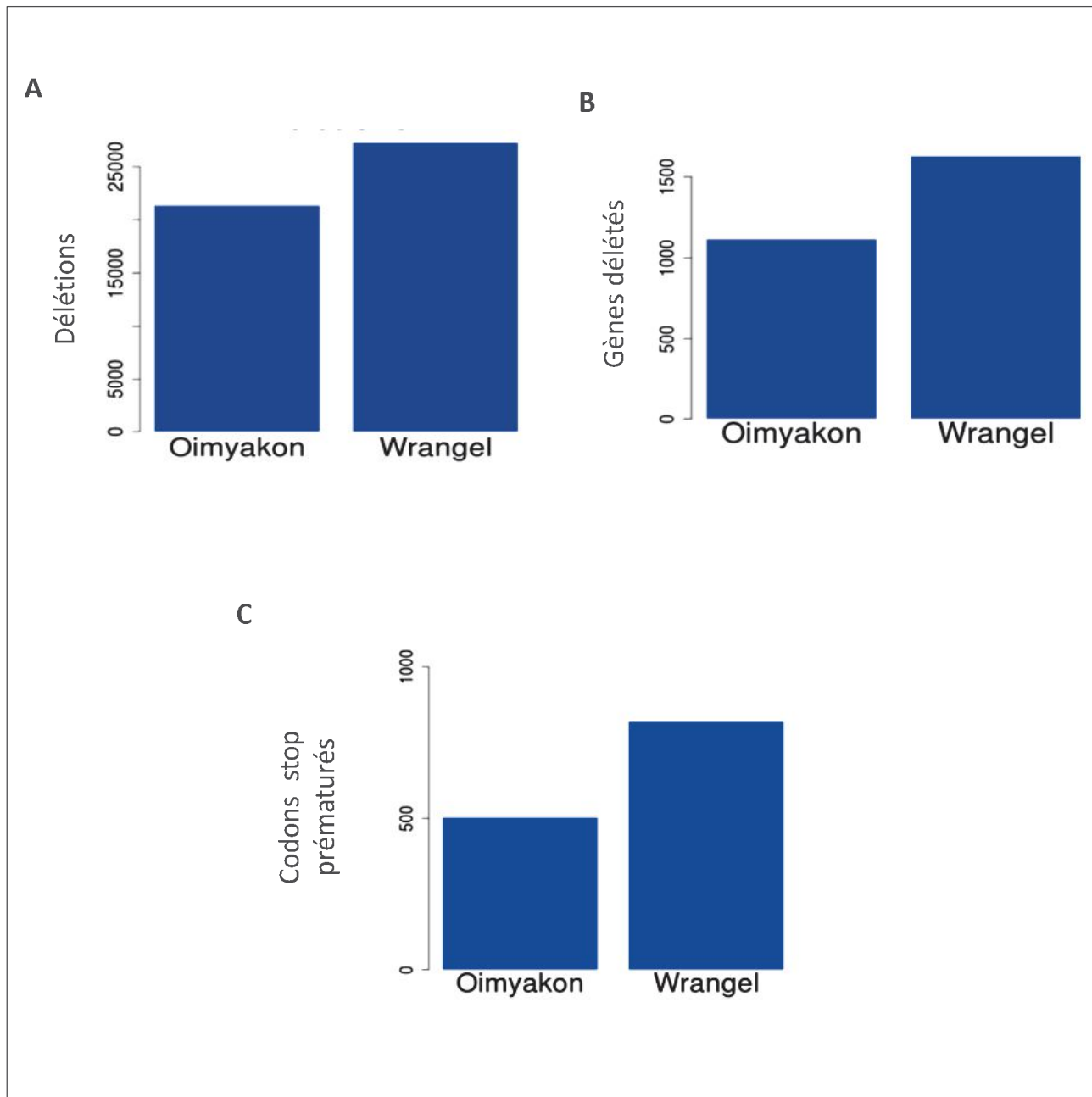
Figure adaptée de la figure 3 de l'article Strout et al., 1998.

génomique est affecté par l'insertion d'un ET, dépend du moment où la transposition a eu lieu. Les transpositions précoces d'ETs peuvent altérer le génome des cellules précurseurs pouvant aboutir à la création d'un lignage cellulaire aberrante, alors que la transposition tardive d'un ET au sein d'une cellule différenciée n'affectera que quelques cellules. Les insertions somatiques sont alors responsables de mosaïcisme puisque deux cellules ayant un génotype différent sont présentes au sein d'un même individu. La transposition d'ETs dans les cellules somatiques est plus fréquente que dans les cellules germinales (Hancks et Kazazian, 2012). Par ailleurs, les effets perturbateurs liés à la mobilisation d'ET ont été largement documentés dans les cellules somatiques.

L'altération de gènes induite par l'insertion somatique d'un ET a initialement été mise en évidence chez le maïs par Barbara McClintock. En effet, elle avait pu démontrer, grâce au système *Ds/Ac*, que les grains de maïs jaunes avec des tâches violettes traduisaient un mosaïcisme génétique lié à la transposition clonale de l'ET *Ds*.

Chez l'Homme, le premier cas d'insertion somatique d'un ET a été identifié chez un patient atteint d'un cancer colorectal en 1992 par Miki et ses collègues. L'origine des cancers colorectaux est le plus souvent associée à l'altération de l'expression du gène *APC*, de l'anglais *adenomatous polyposis coli*. Le gène *APC* est un gène dit "suppresseur de tumeur" puisqu'il code une protéine impliquée dans le processus prévenant l'inflammation pré-tumorale. L'ADN des cellules provenant de tumeurs colorectales a été comparé à celui des cellules environnantes. Miki et ses collègues se sont alors aperçus qu'un élément LINE-1 était inséré dans le gène *APC* des cellules tumorales (Miki et al., 1992). Cette insertion conduit à la création précoce d'un signal de polyadénylation altérant ainsi l'expression du gène *APC*. Dès lors, plus d'une centaine de maladies, dont certains cancers, ont été associées à l'insertion somatique d'un ET (Chenais, 2013 ; Hancks et Kazarian, 2016).

Présents en multicopies dans les génomes eucaryotes, les ETs peuvent également induire des réarrangements génomiques tels que des duplications, des délétions ou encore des inversions. En d'autres termes, lors de l'activation de la voie de réparation homologue, deux copies d'un même élément situées à deux localisations génomiques différentes pourront recombiner, et ainsi, générer des réarrangements chromosomiques plus ou moins importants. A titre d'exemple, différentes formes sévères de leucémies ont été associées à des réarrangements chromosomiques générés à la suite de recombinaisons somatiques entre les séquences *Alu* (Figure 12) (Strout et al., 1998 ; Caligiuri et al., 1998).



**Figure 13 : Analyses des mutations présentes dans le génome des mammoths de Oimyakon et celui des mammoths de l'île de Wrangel.**

Comparaison du nombre de délétions génomiques (A), de gènes délétés (B) et de codons stop prématurés (C) présents dans le génome des mammoths d'Oimyakon par rapport à celui des mammoths de l'île de Wrangel.

Adaptée de la figure 1 de Rogers et Slatkin, 2017.

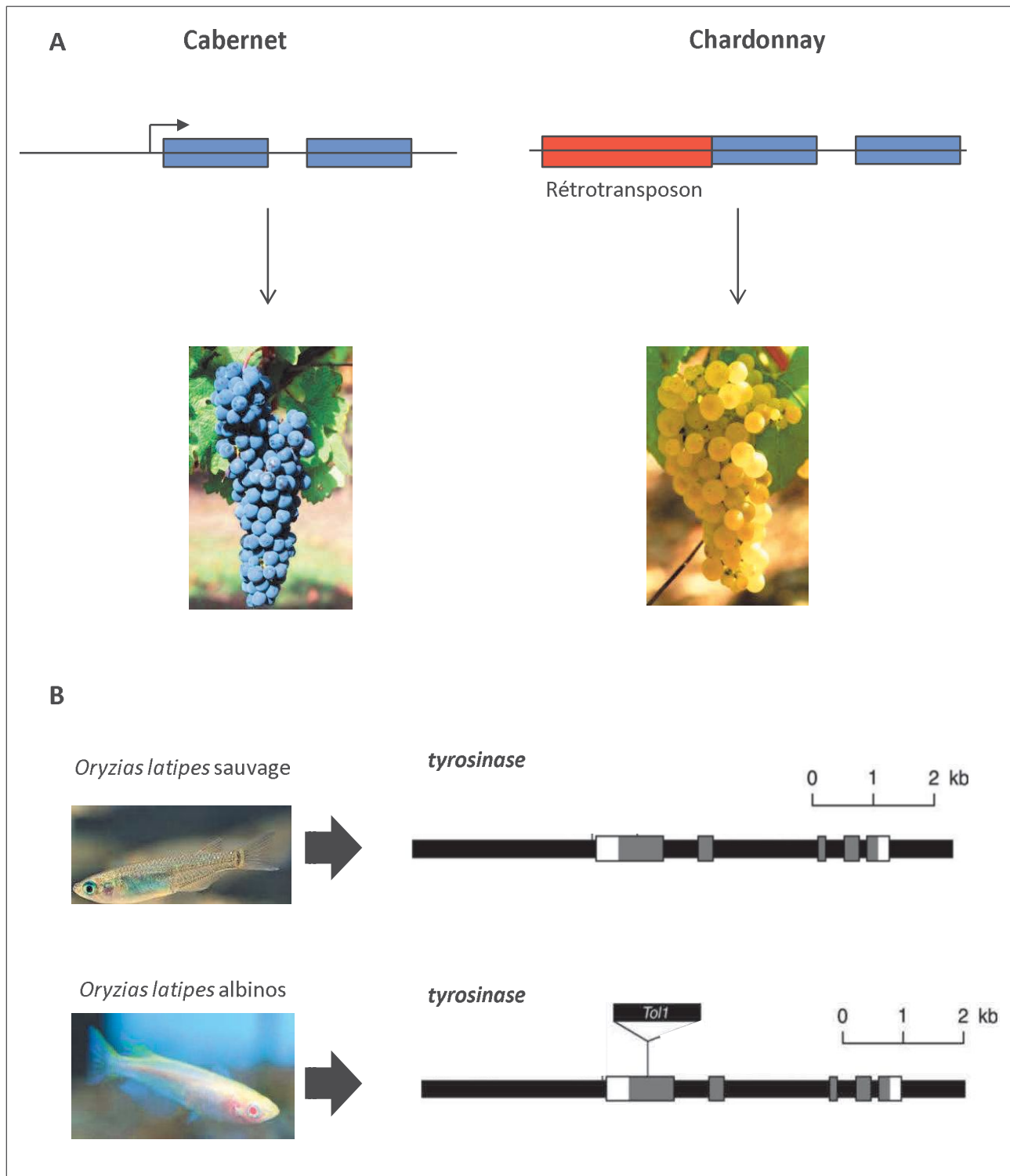
**b) Conséquences liées à la mobilisation d'un élément transposable au sein des cellules germinales**

Malgré une régulation fine des ETs au sein de cellules des tissus reproducteurs, certains ETs échappent aux différents mécanismes de contrôle mis en place par les cellules. Contrairement aux mutations somatiques, les mutations germinales induites notamment lors de la transposition d'un ET sont transmises à la future descendance. Ces mutations sont alors retrouvées dans toutes les cellules de la progéniture. La transposition germinale d'un ET peut avoir plusieurs conséquences, telles que l'apparition d'une stérilité, ou l'émergence de maladies allant jusqu'à une létalité précoce de la progéniture. Les mutations germinales liées à la transposition des ETs peuvent ainsi mettre en péril la survie d'une espèce.

L'un des meilleurs exemples de l'effet néfaste des ETs est celui du mammouth provenant de l'île de Wrangel. Récemment, le séquençage des génomes de deux spécimens de mammouths laineux a été réalisé (Palkopoulou et al. 2015). Le premier spécimen provient d'une population de nombreux mammouths localisés sur le continent sibérien, alors que l'autre spécimen provient, quant à lui, d'une population beaucoup plus petite de mammouths isolés sur l'île de Wrangel. À la suite du séquençage génomique de ces deux spécimens, une étude comparative a permis de révéler que le mammouth issu de l'île de Wrangel possédait de nombreuses mutations telles que de multiples délétions touchant dans la grande majorité des gènes, une accumulation de séquences appartenant aux ETs ainsi qu'une accumulation de codons-stops prématurés (Figure 13) (Rogers et Slatkin, 2017). L'ensemble de ces observations tend à penser que ces altérations génomiques ont été induites par une activité massive des ETs chez les mammouths isolés sur l'île de Wrangel. L'absence de régulation de ces ETs aurait eu pour conséquence d'altérer, de façon trop importante, le génome des mammouths de l'île pouvant potentiellement expliquer leur déclin sur l'île de Wrangel et serait donc responsable de leur extinction.

Chez la drosophile, la stérilité ou l'absence de développement de la descendance sont obtenues lors de la dysgénésie des hybrides engendrée par le ET *P* et l'ET *I* (Picard et al., 1976 ; Kidwell et al., 1977). C'est notamment sur ce critère que de nombreuses protéines ont été identifiées comme impliquées dans la régulation des ETs. En effet, lorsque ces dernières sont mutées, les mécanismes de régulation des ETs ne sont plus fonctionnels laissant alors la possibilité aux ETs de transposer. La dérégulation massive d'ETs est le plus souvent associée à la présence d'ovaires rudimentaires, ou quasi inexistantes, conduisant à une stérilité. Chez la





**Figure 14 : Phénotypes mutants engendrés à la suite de la mobilisation d'un ET.**

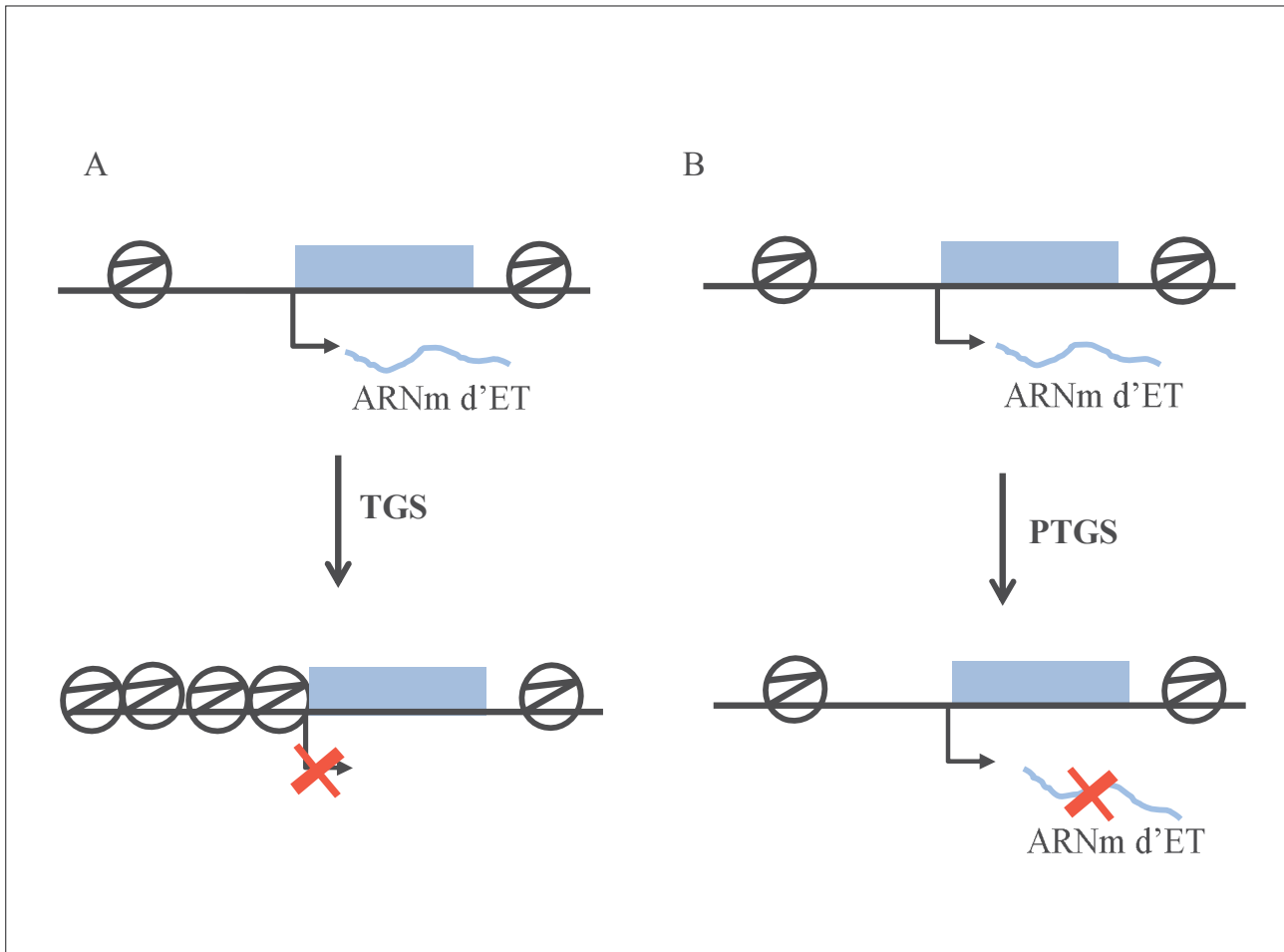
- A) La différence de coloration entre les grains de raisin Cabernet et des grains de raisin du Chardonnay est due à l'insertion d'un rétroélément dans le promoteur du gène codant un pigment noir. Le promoteur est alors altéré.
- B) Le phénotype albinos observé chez certains poissons *Oryzias latipes* est du à l'insertion de l'ET *Tol1* dans l'exon 1 du gène de la Tyrosinase, enzyme permettant la synthèse de mélanine.

drosophile, de nombreuses études ont pu confirmer qu'une augmentation du nombre d'insertions d'ETs était corrélée à une diminution du taux d'éclosion des œufs (Busseau et al., 1998 ; Khurana et al., 2011).

L'insertion d'un ET dans les régions régulatrices d'un gène ou de son promoteur peut entraîner une modification spatio-temporelle de son expression. C'est notamment le cas des grains de raisin des vignobles Chardonnay pour lesquels un rétroélément s'est inséré dans le promoteur d'un gène impliqué dans la synthèse d'un pigment noir au niveau des fruits (Figure 14A). La conséquence de cette insertion est une inactivation de ce gène, et par conséquent, l'obtention d'une coloration jaune des grains (Lisch 2013). À l'inverse, l'insertion d'un ET peut activer l'expression d'un gène normalement éteint dans un tissu. C'est notamment à l'issue de l'insertion d'un rétrotransposon de type *Copia* au niveau de la région régulatrice du gène *Ruby* que les oranges sanguines sont apparues (Butelli et al., 2012). En effet, *Ruby* est un facteur de transcription MYB qui permet la production d'un pigment rouge-orangé, l'anthocyanine. Dans le cas des oranges sanguines, le rétroélément contrôle l'expression du gène *Ruby* dans la chair du fruit.

Une perte de fonction peut également être engendrée à la suite de l'insertion germinale d'un ET au sein d'un exon ou d'un intron. Ces insertions peuvent être à l'origine de l'épissage alternatif du transcrit initial, de la création de sites de terminaisons précoces ou encore d'une perturbation du cadre de lecture du gène affecté (Cowley et Oakey, 2013 ; Elbarbary et al., 2016 ; Mita et Boeke, 2016). À titre d'exemple, l'insertion de séquence *LINE1* dans l'exon 14 du gène du facteur VIII est responsable de l'hémophilie A. Chez *Oryzias latipes*, le poisson *Madaka*, l'ET *Toll* a été identifié comme étant responsable de certains cas d'albinisme (Koga, 1995). L'albinisme est une maladie pour laquelle l'individu ne synthétise plus la mélanine, un pigment noir (Oetting et al. 2003). En étudiant la population des poissons Medaka, des individus albinos sont spontanément apparus. L'analyse génomique de ces individus a démontré que trois allèles mutants étaient présents dans la population, et que tous étaient liés à la présence d'un transposon touchant l'expression du gène codant la tyrosinase, l'enzyme qui permet la synthèse de la mélanine (Figure 14B) (Koga et al. 1995; Koga et al. 1996). L'altération de l'expression de la tyrosinase est due à l'insertion de l'élément *Toll* dans le premier exon de ce gène.

L'expression d'ETs peut également conduire à des réarrangements chromosomiques. En comparant le génome des tomates rondes (*Solanum lycopersicum*) avec celui des tomates



**Figure 15 : Représentation des différents niveaux de régulation des ETs.**

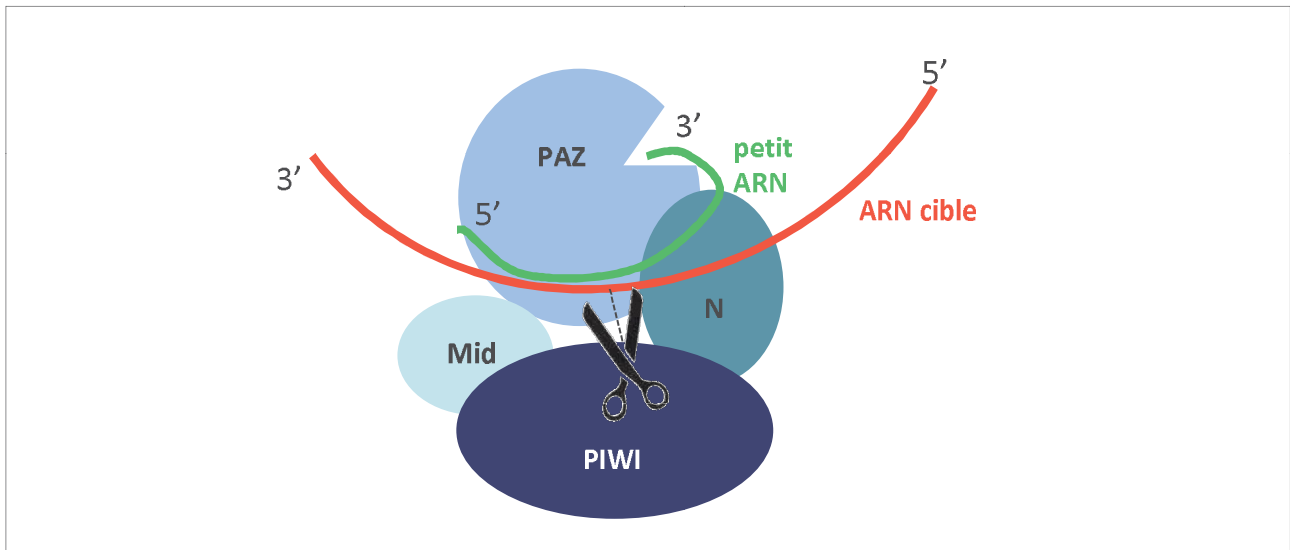
- A) La régulation transcriptionnelle (TGS) consiste à la mise en place d'une chromatine inactive au niveau du promoteur empêchant l'expression des ETs.
- B) La régulation post-transcriptionnelle (PTGS) consiste à dégrader les transcrits issus d'ETs actifs.

ovales, les scientifiques se sont aperçus que la différence de morphologie entre ces deux sortes de tomates était notamment due à la mobilisation de l'ET *Rider*. En effet, la mobilisation de cet ET a eu pour conséquence de générer la duplication puis la translocation du gène IQD12 (Xiao et al., 2008). La surexpression du gène IQD 12 au niveau des fruits des plants de tomates est responsable de l'allongement de ces derniers.

L'absence de régulation des ETs peut engendrer des effets néfastes sur un individu jusqu'à aboutir à l'extinction de l'espèce. En des millions d'années d'évolution, les ETs ont finement été contrôlés par les génomes des hôtes, atteignant ainsi l'équilibre entre variabilité des génomes et stabilité de ces derniers.

## Chapitre 2 : Les acteurs clés de la régulation des éléments transposables

Face au fort potentiel mutagène des ETs, les génomes hôtes ont développé divers mécanismes de défense afin de réguler finement l'activité des ETs, et ainsi, préserver la balance entre la stabilité et la variabilité des génomes. Chez les eucaryotes, les ETs sont contrôlés tant au niveau transcriptionnel qu'au niveau post-transcriptionnel (Figure 15). L'inhibition transcriptionnelle des ETs consiste à mettre en place une structure hétérochromatique au niveau de leurs promoteurs réprimant ainsi leur expression. La régulation post-transcriptionnelle consiste, quant à elle, à dégrader les transcrits des ETs actifs. Ces deux types de régulations utilisent le mécanisme d'ARN interférence dont les acteurs principaux sont des petits ARNs non codants et les protéines de la grande famille ARGONAUTE. Chez les animaux, les petits ARNs non codants sont répartis en trois grandes familles en fonction de leur taille : les miRNAs sont des petits ARNs non codants de 22 nucléotides pris en charge par la protéine AGO1, les siRNAs font quant à eux 21 nucléotides et sont pris en charge par la protéine AGO2, et enfin les piRNAs, dont la taille varie entre 23 et 29 nucléotides, sont chargés sur les différentes protéines appartenant à la sous-famille PIWI. Parmi les petits ARNs non codants, seuls les siRNAs (Chung et al., 2008 ; Czech et al., 2008 ; Kawamura et al., 2008) et les piRNAs (Brennecke et al., 2007 ; Aravin et al., 2006) ont été décrits comme étant impliqués dans la répression des ETs. Cependant, il a récemment été observé que l'altération de la voie de miRNAs au sein des cellules folliculaires de l'ovaire de drosophile avait pour conséquence de diminuer la production de piRNAs responsables de la mise sous silence d'ETs suggérant un lien possible entre la voie des miRNAs et celle des piRNAs (Mugat et al., 2015). La voie des endosiRNAs est active dans toutes les cellules de



**Figure 16 : Structure schématique d'une protéine ARGONAUTE.**

Les protéines ARGONAUTES présentent la caractéristique d'être constituées de quatre domaines protéiques : le domaine PAZ qui leur permet de s'associer aux petits ARNs, le domaine Mid, le domaine PIWI dont le rôle est de cliver l'ARN cible et le domaine N terminal.

Figure adaptée de Song et al., 2004.

l'organisme, alors que la voie des piRNAs est, quant à elle, exclusivement active au sein des cellules des tissus reproducteurs.


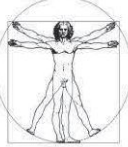

Néanmoins, un système additionnel de répression des ETs utilisant les protéines KRAB, protéines en doigt de zinc qui sont assimilées à des facteurs de transcription, a été mis en évidence chez les eucaryotes supérieurs (Yang, Wang et Macfarlan, 2017). En effet, des études récentes ont montré *via* des expériences d'immunoprécipitation de chromatine (ChIP), mais aussi, lors d'expériences de perte de fonction, que les protéines KRABs étaient importantes pour la reconnaissance et la mise en place de la régulation transcriptionnelle des ETs (Yang, Wang et Macfarlan, 2017).

### *A. La grande famille des protéines ARGONAUTES*

Lors du mécanisme d'ARN interférence, des petits ARNs de 20 à 30 nts sont chargés sur l'un des membres de la grande famille des protéines ARGONAUTES. Les complexes ainsi formés vont reconnaître par complémentarité de base (totale ou partielle) le transcrit du gène ou de l'ET à réprimer. Les complexes petit ARN/ARGONAUTE peuvent être soit importés au noyau pour initier une régulation transcriptionnelle, soit restés au niveau du cytoplasme pour réaliser une régulation post-transcriptionnelle.

Initialement, les protéines ARGONAUTES ont été décrites comme jouant un rôle important au cours du développement chez les végétaux (Bohmert et al., 1998), mais agissant aussi au cours de la division des cellules souches germinales de drosophile (Lin et Spradling, 1997). Les protéines ARGONAUTES sont caractérisées par la présence de quatre domaines : le domaine en N terminal, le domaine PAZ, le domaine Mid et le domaine PIWI (Figure 16). Les domaines PAZ et Mid sont les domaines qui permettent aux protéines de la famille Argonaute de fixer des petits ARNs (Song et al., 2004 ; Yan et al., 2003 ; Lingel et al., 2003). Le domaine PIWI, proche structurellement du domaine PAZ, confère une activité endonucléasique aux protéines ARGONAUTES. En d'autres termes, une fois que le transcrit à réprimer est reconnu par le petit ARN lié à la protéine ARGONAUTE, celle-ci coupe ce transcrit exerçant ainsi une régulation post-transcriptionnelle.

Présentes chez les bactéries, les archaebactéries et les eucaryotes, les protéines de la famille des protéines ARGONAUTES sont conservées à travers les espèces. Elles ont été subdivisées en 2 sous-familles : la sous-famille des protéines Argonautes dont l'expression est

ARGONAUTE						
Argonautes				PIWI		
	Ago1	Ago2	Ago3	Ago4	MIWI	MILI MIWI2
	Ago1	Ago2	Ago3	Ago4	HIWI HIWI3	HIWI2 HIWI4
	Ago1	Ago2		Ago3	Aub	Piwi
	Liaison aux miRNAs		Liaison aux endosRNAs		Liaison aux piRNAs	

**Tableau 1 : Les protéines de la grande famille ARGONAUTE chez l'homme, la souris et la drosophile.**

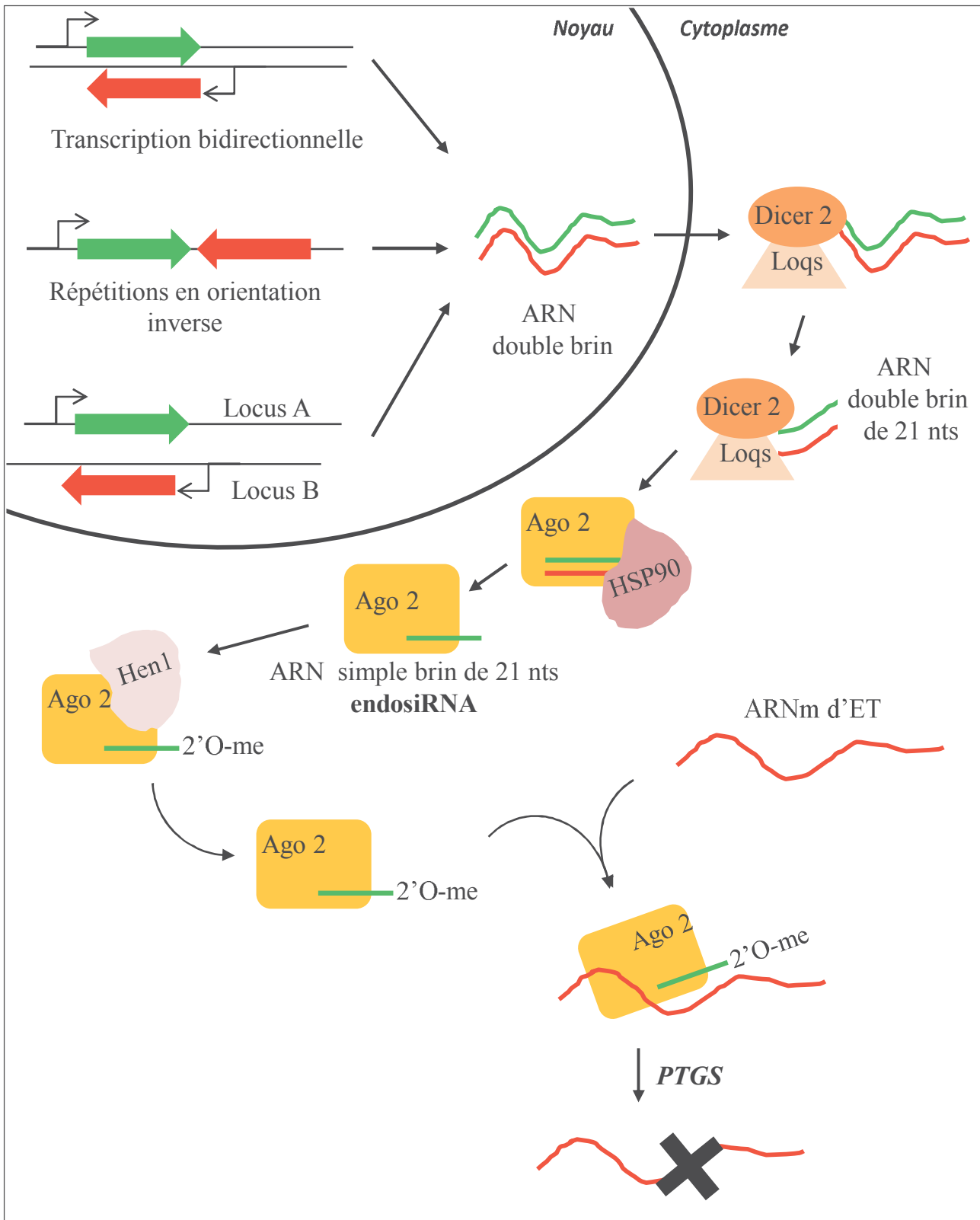
ubiquitaire (Peters et al., 2007 ; Hutvagner et Simard, 2008) et celle des protéines PIWI dont l'expression est restreinte aux cellules germinales, à l'exception des cellules folliculaires des ovaires de drosophile (Siomi et al., 2011). Une autre différence entre ces deux sous-familles est associée à la prise en charge des petits ARNs. En effet, les petits ARNs liés aux protéines Argonautes sont produits à partir de structure double brin alors que ceux liés aux protéines PIWI sont produits à partir de transcrits simples brins.

Le nombre de gènes codant les protéines de la famille ARGONAUTES sont différents en fonction des espèces (Tableau 1). A titre d'exemple, une seule protéine de la sous-famille Argonaute a été identifiée à ce jour chez *Schizosaccharomyces pombe*, alors que *C. elegans* possède 27 protéines de la sous-famille Argonaute. Chez la souris, sept gènes ARGONAUTES ont été identifiés : quatre d'entre eux codent des protéines de la sous-famille Argonautes (AGO1, AGO2, AGO3 et AGO4) et les trois autres codent des protéines de la sous-famille PIWI (MIWI, MILI et MIWI2). Les gènes codant les protéines AGO1, AGO2, AGO3 et AGO4 sont conservés chez l'Homme et un gène supplémentaire codant une quatrième protéine de la sous-famille PIWI a été répertorié (HIWI1, HIWI2, HIWI3 et HIWI4). La drosophile possède, quant à elle, deux gènes codant des protéines de la sous-famille argonaute (AGO1 et AGO2) et trois gènes codant des protéines de la sous-famille PIWI : les protéines Piwi, Aubergine (Aub) et AGO3. Ces trois protéines sont respectivement homologues aux protéines murines MIWI, MILI et MIWI2.

## ***B. Les endosiRNAs***

Les siRNAs/endosiRNAs sont des petits ARN interférents ayant la capacité de se lier aux protéines de la sous-famille Argonaute. Leur voie de biogenèse est relativement bien conservée chez les eucaryotes. Elle a initialement été décrite comme une voie de défense contre l'invasion de matériels génétiques extérieurs notamment lors des infections virales. En effet, de nombreuses études ont démontré que des petits ARNs étaient produits à partir des séquences étrangères puis chargés sur la protéine AGO2 dans le but d'exercer une régulation post-transcriptionnelle de ces derniers. En 2008, les petits ARNs provenant de têtes, d'embryons, d'ovaires de drosophile et de drosophiles entières ont été séquencés. L'alignement des petits ARN de 21 nucléotides (siRNAs) ainsi séquencés sur le génome de *Drosophila melanogaster* a révélé que ces derniers pouvaient également être codés par le génome lui-même (Czech et al., 2008 ; Ghildiyal et al., 2008 ; Okamura et Lai, 2008). En effet, ces siRNAs s'alignent majoritairement sur des régions riches en ETs, sur des loci



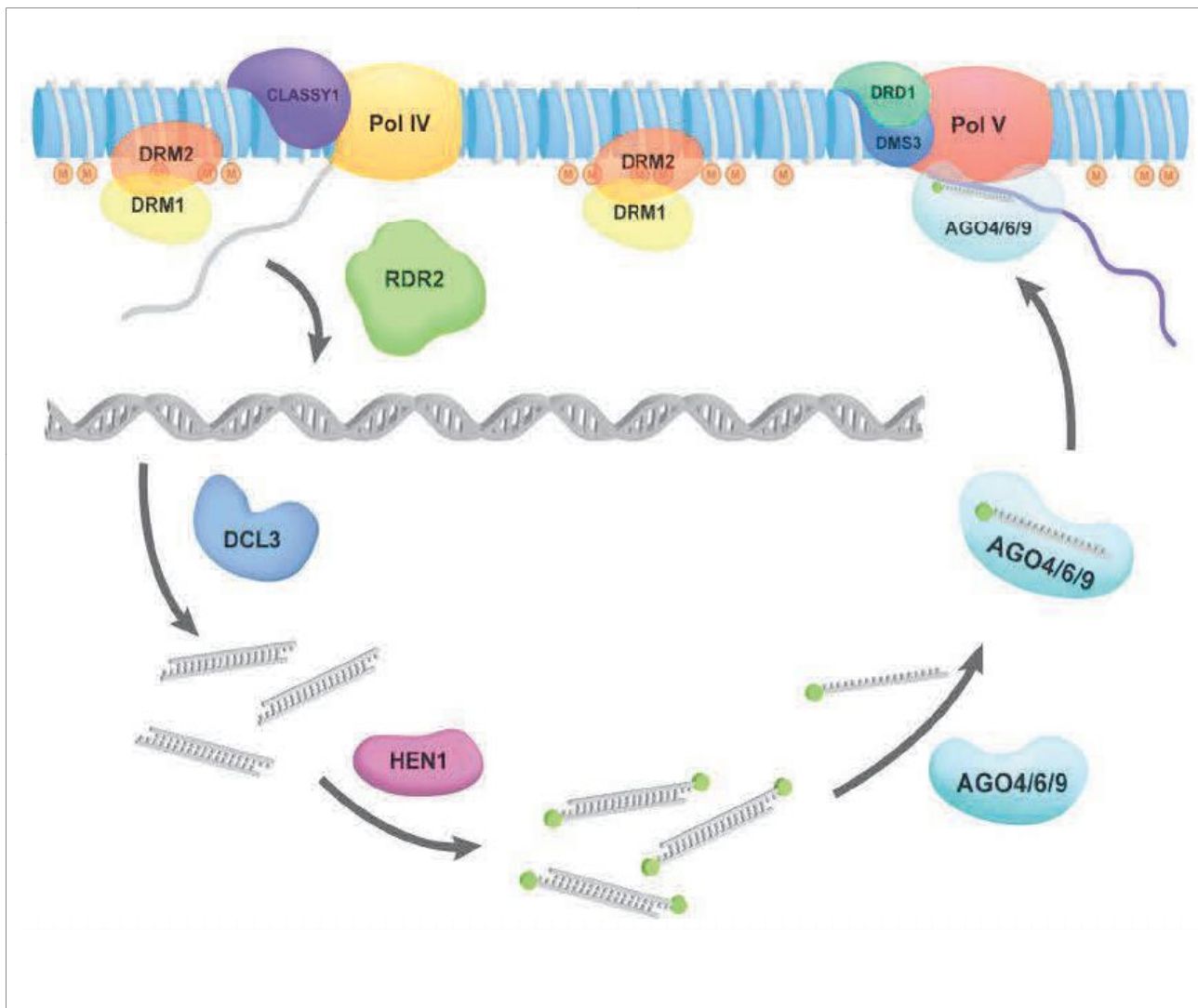


**Figure 17 : La voie des endosiRNAs chez la drosophile.**

La voie des endosiRNAs s'amorce avec la formation d'ARNs double brin reconnus au niveau du cytoplasme par la protéine DICER2 et son cofacteur Loquacious (Loqs). La protéine DICER2 clive l'ARN double brin en duplex de 21 nucléotides chargé sur la protéine AGO2 aidée de la protéine HSP90. Après la sélection de l'un des deux brins, un groupement 2'O-méthyle est ajouté sur le petit ARN simple brin de 21 nts, l'endosRNA. Par la suite, le complexe endosRNA/AGO2 reconnaît, par complémentarité de séquence, les transcrits des cibles à réprimer.

hétérochromatiques ainsi que sur des séquences intergéniques. Provenant de régions endogènes, ces petits ARNs chargés sur la protéine AGO2 ont été appelés les endosiRNAs. Les endosiRNAs ont été retrouvés orientés aussi bien en sens qu'en anti-sens par rapport aux séquences ADN dont ils sont homologues. Des expériences similaires chez la souris ont permis de corroborer ces résultats.

Chez la souris et la drosophile, les endosiRNAs (petits ARNs de 21 nts) proviennent de la maturation de transcrits issus de séquences répétées en orientation inverse, ou de loci génomiques ayant une transcription bidirectionnelle (Ghildiyal et Zamore, 2009). La transcription de ces séquences particulières aboutit à la formation des précurseurs des endosiRNAs, les ARNs double brins (Figure 17). Une fois transcrits, les précurseurs sont exportés dans le cytoplasme où ils sont pris en charge par l'endonucléase DICER 2 **via** son domaine de liaison à l'ARN. Une fois fixée à l'ARN double brin, la protéine DICER 2, aidée de la protéine Loquacious ou de la protéine R2D2, mature l'ARN double brin en petits fragments d'ARN de 21 nts (Okamura et Lai, 2008). Les fragments d'ARN double brin générés sont ensuite chargés sur la protéine AGO2. Le chargement des fragments d'ARN double brin sur la protéine AGO2 requiert la protéine de choc thermique HSP90 (Miyoshi et al., 2010). La sélection du brin qui restera lié sur la protéine AGO2 est basée sur la stabilité thermodynamique des extrémités 5' de l'ARN de 21 nts (Liu et al., 2003 ; Tomari et al., 2004). Le brin lié à AGO2 est appelé endosiRNA, alors que l'autre brin est dégradé. Durant leur maturation, l'extrémité 5' des endosiRNAs est 5'monophosphatée alors que leur extrémité 3' est 2'O-méthylée par l'enzyme Hen1 (Okamura et al., 2004). L'endosiRNA chargé sur la protéine AGO2 forme alors le complexe RISC. Ce complexe permet la dégradation spécifique des ARNs messagers cibles tels que les transcrits d'ETs actifs. Cette régulation post-transcriptionnelle est basée sur la complémentarité nucléotidique entre l'endosiRNA et les transcrits cibles. La mutation ou la délétion d'un des acteurs de cette voie tels que DICER2 ou AGO2 engendre une augmentation de l'expression d'ETs dans tous les tissus de la drosophile (Ghildiyal et al., 2008 ; Kawamura et al., 2008 ; Chung et al., 2008 ; Okamura et Lai, 2008). Néanmoins, l'altération de la voie des endosiRNAs au sein de la lignée germinale de drosophile semble engendrer des effets moins drastiques que chez la souris. En effet, la voie murine des endosiRNAs joue un rôle important dans la mise sous silence des ETs dans la lignée germinale des femelles puisque son altération entraîne une stérilité des souris (Tam et



**Figure 18 : La voie de régulation des ETs chez les plantes, voie RdDM.**

La voie RdDM est une voie de méthylation ADN *via* de petits ARNs de 24 nucléotides.

La polymérase IV transcrit des ETs localisés dans des régions hétérochromatiques. Leurs transcrits sont ensuite pris en charge par la protéine RDR2 qui synthétise le brin d'ARN complémentaire. Les transcrits d'ETs double brin sont alors clivés en petits duplex de 24 nts par l'enzyme DCL3 (DICER like 3) puis méthylés au niveau de leurs extrémités 3' par HEN1. Les petits ARNs de 24 nts méthylés sont chargés sur la protéine AGO4 ou AGO6 ou AGO-9 afin de cibler les transcrits issus de loci génomiques particuliers transcrits par la polymérase V (Pol V). Ces loci sont alors mis sous silence.

Figure 2 de l'article Willmann et al., 2011

al., 2008 ; Watanabe et al., 2008 ). Chez la drosophile, l'altération de cette voie au sein de la lignée germinale entraîne une dérégulation d'un petit nombre d'ETs sans induction de stérilité.

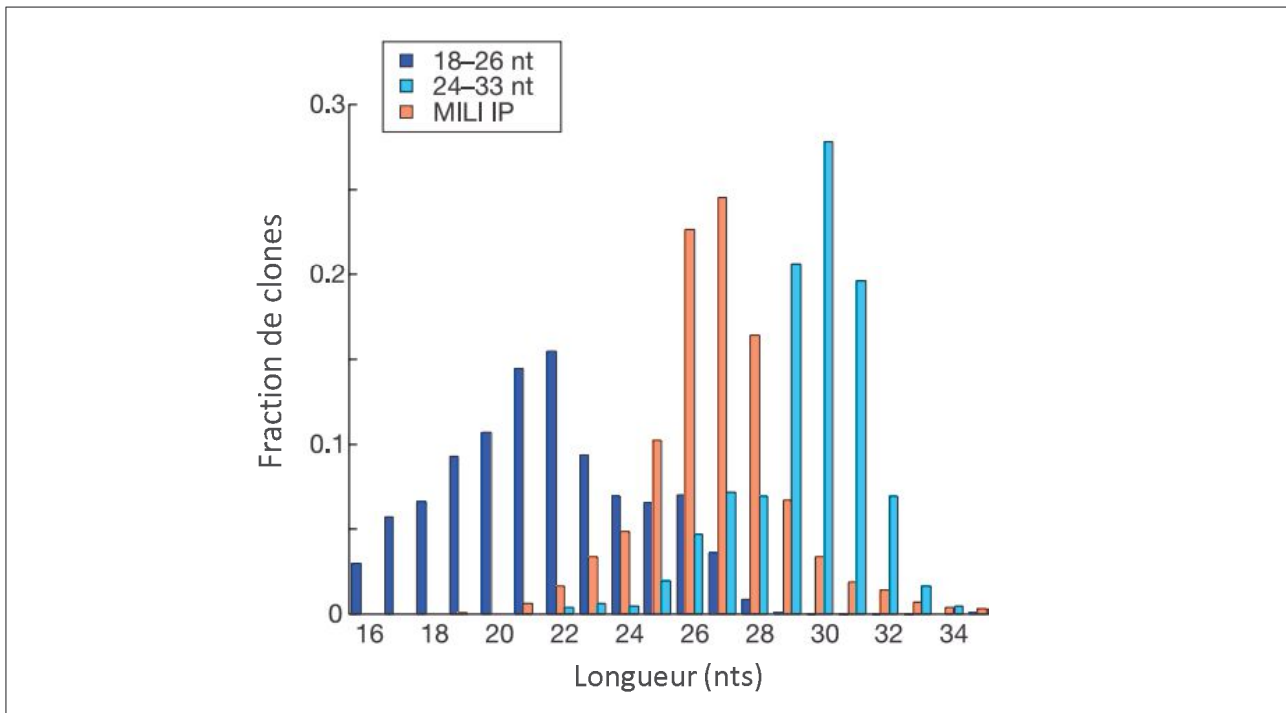
Chez les végétaux, les ETs sont transcrits par l'ARN polymérase IV (Figure 18). Les transcrits des ETs sont ensuite pris en charge par la protéine RDR2 qui permettra la production de transcrits complémentaires pour former un ARN double brin. Les ARNs double brin sont alors maturés en petits ARNs interférents de 24 nucléotides par la protéine Dicer-like3 (DCL3). Tout comme chez la souris et chez la drosophile, les petits ARNs interférents de 24 nucléotides vont être méthylés par l'enzyme Hen1 et pris en charge par une protéine de la sous-famille ARGONAUTE, la protéine AGO4. Le complexe endosRNA/AGO4 va, par complémentarité de base, reconnaître le transcrit naissant d'ETs transcrit par l'ARN polymérase V. Cette reconnaissance permet le recrutement notamment des protéines DRM1 et DRM2 qui sont impliquées dans la mise sous silence des ETs en méthylant l'ADN (Willmann et al., 2011).

### ***C. Les piRNAs***

Les piRNAs sont des petits ARNs non codants de 23 à 29 nucléotides qui ont la capacité de se lier aux protéines ARGONAUTES de la sous-famille PIWI. Contrairement aux endosRNAs, les piRNAs régulent l'expression des ETs exclusivement au sein des tissus gonadiques. Par conséquent, les piRNAs assurent la protection de l'information génétique transmise à la descendance (Aravin et al., 2007). Néanmoins, les piRNAs ont également pour fonction de réguler des gènes. C'est notamment grâce à cette fonction qu'ils ont été découverts.

#### ***I. La découverte des piRNAs***

Les premiers piRNAs ont été mis en évidence lors d'une étude visant à explorer la régulation post-transcriptionnelle du locus *Stellate* au sein des testicules de drosophile (Aravin et al., 2001). Le locus *Stellate* est une séquence répétée codant une protéine similaire à la  $\beta$ -sous-unité de la Caséine Kinase II. L'expression de cette protéine, au sein des testicules de drosophile, engendre la formation d'agrégats. Durant la spermatogenèse, les agrégats ainsi formés ont pour conséquence d'induire une stérilité. Lors d'une étude génétique visant à identifier les gènes impliqués dans l'inhibition de l'expression du gène *Stellate*, un site génomique composé d'une séquence similaire au locus *Stellate* fut identifié. Par ailleurs, ce



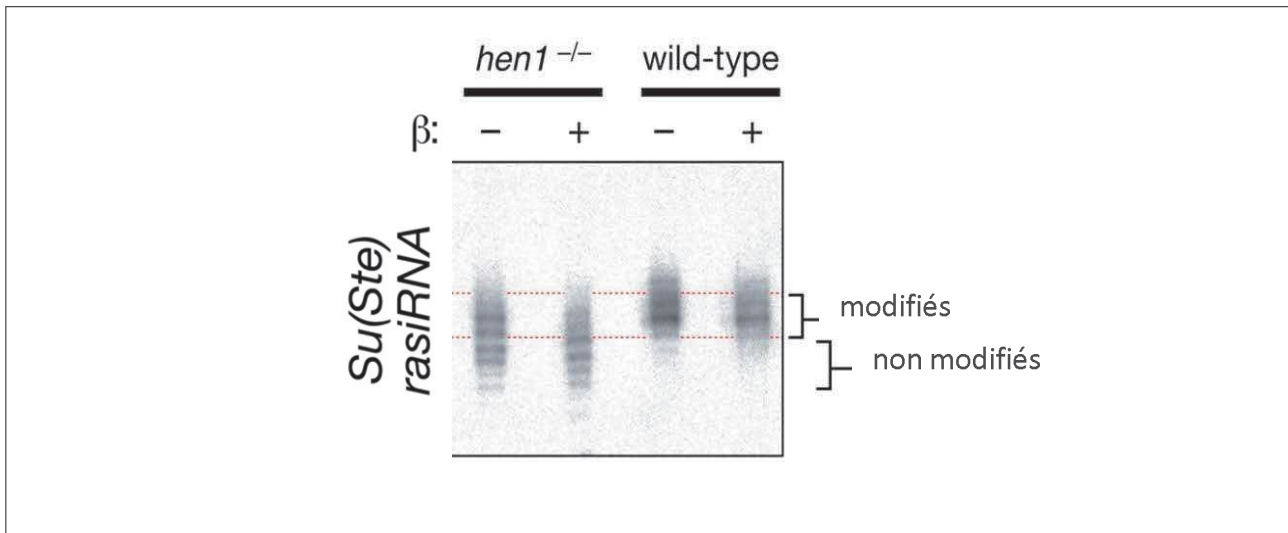
**Figure 19 : Distribution de tailles des petits ARNs provenant de testicules murins.**

Les ARNs totaux ont été extraits à partir des testicules de souris adultes. Les petits ARNs interagissant avec la protéine MILI, ainsi que les petits ARNs testiculaires de 18 à 26 nts et les petits ARNs testiculaires de 24 à 33 nts ont été clonés puis séquencés. La distribution de taille des petits ARNs clonés a été représentée en bleu foncé pour la librairie issue des petits ARNs testiculaires clonés de 18 à 26 nts, en bleu clair pour la librairie issue des petits ARNs testiculaires clonés de 24 à 33 nts et en orange pour la librairie issue des petits ARNs testiculaires immunoprécipités avec la protéine MILI.

Figure adaptée de la figure 1A présentée dans l'article Aravin et al., 2006.

site génomique localisé dans une région hétérochromatique du chromosome Y a été nommé *suppresseur de Stellate* puisqu'il est impliqué dans la répression de l'expression de ce gène. Les expériences de Northern Blot ont permis de révéler qu'au sein des ARNs totaux des testicules de drosophile, des petits ARNs de 25 à 27 nts étaient produits à partir du locus *Suppresseur de Stellate*, puisque chez des mâles ayant une délétion importante de ce locus, aucun petit RNA homologue à ce locus n'a pu être détecté. La présence de ces petits ARNs a été corrélée à la régulation du gène *Stellate*. À l'époque, ces petits ARNs ont été baptisés les rasiRNAs puisqu'ils permettaient la mise sous silence de séquences répétées dans le génome.

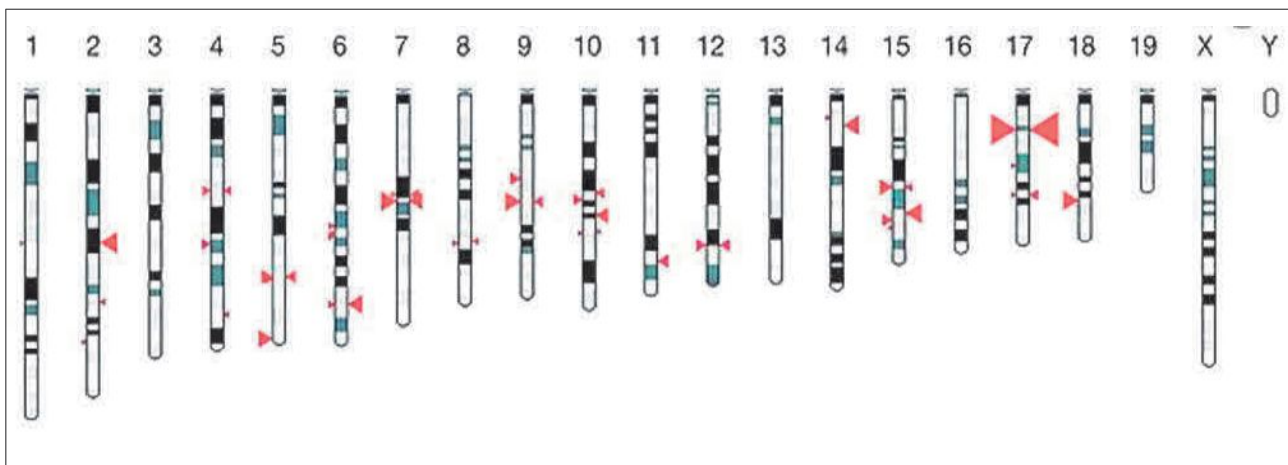
Cinq ans plus tard, chez la souris, une immunoprécipitation de la protéine MILI, une protéine appartenant à la sous-famille des protéines PIWI, a été réalisée à partir d'un lysat testiculaire de souris adulte (Vagin et al., 2006 ; Aravin et al., 2006 ; Girard et al., 2006 ; Grivna et al., 2006 ; Lau et al., 2006 ; Watanabe et al., 2006). Les petits ARNs liés à la protéine MILI ont été clonés puis séquencés. Les analyses bio-informatiques ont révélé que les petits ARNs chargés sur la protéine MILI étaient des petits ARNs 23 à 29 nts (Figure 19). La taille de ces petits ARNs était clairement distincte de celle des miRNAs et de celle des siRNAs suggérant qu'ils avaient identifié une nouvelle classe de petits ARNs (Aravin et al., 2006). Etant chargés sur la protéine MILI, mais également sur MIWI, une autre protéine de la sous-famille PIWI, ces petits ARNs de 23 à 29 nts, initialement baptisés les rasiRNAs, ont été renommés les piRNAs, de l'anglais Piwi-interacting RNAs. Malgré la distance évolutive qui sépare les animaux, les piRNAs sont retrouvés au sein des gonades de nombreuses espèces animales telles que la drosophile, la souris et l'Homme (Saito et al., 2006 ; Vagin et al., 2006 ; Brennecke et al., 2007 ; Ashe et al., 2012 ; Ha et al., 2014 ; Juliano et al., 2011 ; Li et al., 2013 ; Yang et al., 2013). Les analyses bio-informatiques comparant les piRNAs produits par différentes espèces animales ont révélé que la séquence des piRNAs ne semblait pas conservée entre ces espèces, contrairement à leur voie de biogenèse. En effet, la taille ainsi que la séquence/composition des piRNAs varient en fonction des espèces. Par exemple, chez la souris et la drosophile, la taille des piRNAs varie entre 23 et 29 nucléotides, alors que, chez *C.elegans* et *Tetrahymena*, les piRNAs ont, respectivement, une taille de 21 nucléotides et 27-30 nts. La caractérisation de ces petits ARNs a révélé que la majorité d'entre eux commençaient à leur extrémité 5' par une uridine monophosphatée. Tout comme les endosiRNAs, les piRNAs possèdent, à leur extrémité 3', un groupement 2'-O-méthyl ajouté par la méthyltransférase Hen1 lors de leur maturation (Horwich et al., 2007) (Figure 20).



**Figure 20 : L'extrémité 3' des piRNAs est méthylée par l'enzyme Hen 1.**

Northern Blot démontrant que les rasiRNAs (piRNAs) provenant du locus suppresseur du gène Stellate (Su(Ste)) sont méthylés dans leur région 3' puisqu'ils sont résistants à la réaction de  $\beta$ -élimination. La comparaison des résultats obtenus en condition sauvage par rapport à ceux obtenus en condition mutante pour l'enzyme Hen 1 indique que la méthylation de l'extrémité 3' des piRNAs de Su(Ste) est effectuée par l'enzyme Hen1.

Figure adaptée de la figure 2B présentée dans l'article Horwich et al., 2007.



**Figure 21 : Emplacement des clusters de piRNAs dans le génome de la souris.**

La taille des triangles est proportionnelle aux nombres de piRNAs immunoprécipités avec la protéine MILI. La position gauche ou droite des triangles rouges indique que les piRNAs produits sont orientés respectivement en anti-sens ou en sens par rapport à la transcription des loci. Figure adaptée de la figure 2A présente dans l'article Aravin et al., 2006.

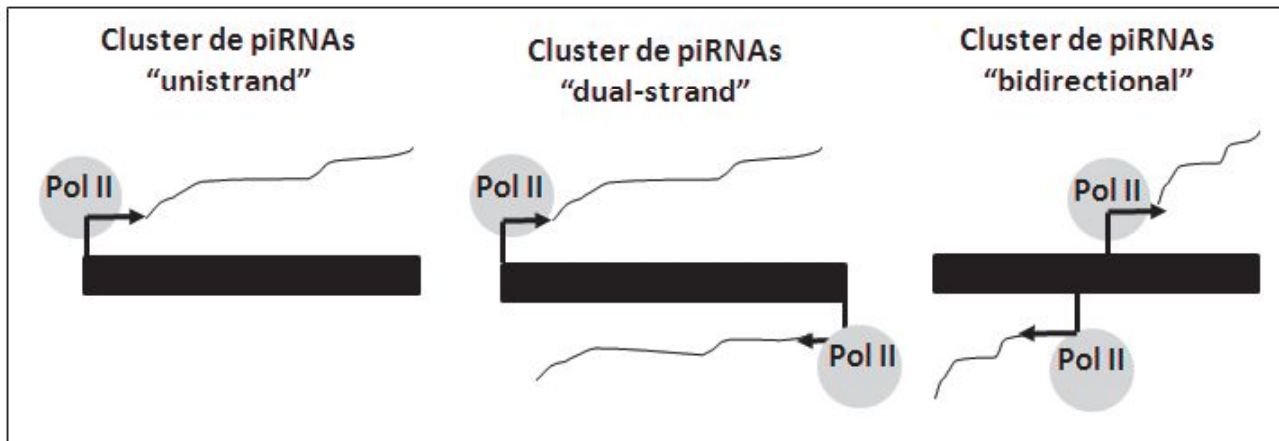
Cette modification leur permet notamment d'être protégés de l'ubiquitinylation et de la dégradation par l'exosome (Kirino et Mourelatos, 2007).

## II. L'origine génomique des piRNAs

Afin de mieux comprendre la biogénèse des piRNAs, l'origine génomique de ces derniers a été explorée. En 2006, quand les piRNAs dans les cellules germinales de souris mâles (Aravin et al., 2006 ; Girard et al., 2006) ont été découverts et cartographiés sur le génome, les auteurs de cette étude ont révélé qu'ils s'alignaient sur un seul des deux brins d'ADN d'une centaine de régions plus ou moins longues (Figure 21). Chez la drosophile, l'étude bio-informatique a été améliorée. En effet afin d'éviter que les piRNAs s'alignent à la fois sur les différentes copies génomiques des ETs qu'ils régulent et leurs sites producteurs, seuls les piRNAs ne s'alignant qu'une seule fois dans le génome ont été sélectionnés s'assurant ainsi de leur provenance. Les piRNAs ainsi sélectionnés ont été appelés les piRNAs uniques puisqu'ils ne sont retrouvés qu'à une seule position génomique. L'alignement des piRNAs uniques sur le génome de drosophile a permis d'identifier les régions productrices de ces derniers, les clusters de piRNAs (Brennecke et al., 2007). Ces clusters de piRNAs sont de taille variable, de quelques paires de bases à plusieurs centaines de kilobases. Chez les animaux, la plupart des clusters de piRNAs sont composés d'ETs. Autrement dit, les clusters de piRNAs répertorient les ETs pouvant être régulés par la voie des piRNAs au sein des cellules gonadiques. Néanmoins, certains loci producteurs de piRNAs sont dépourvus d'ETs et sont majoritairement localisés dans les 3' UTRs de gènes. Chez la drosophile, la région 3' UTR du gène *Traffic Jam* en est un exemple. Chez la souris, les piRNAs sont produits en deux vagues successives lors de la spermatogénèse. Les piRNAs pré-pachytènes proviennent de clusters de piRNAs majoritairement composés d'ETs protégeant ainsi les cellules germinales de potentielles transpositions ; alors que les piRNAs pachytènes sont issus de régions génomiques dépourvues d'ETs.

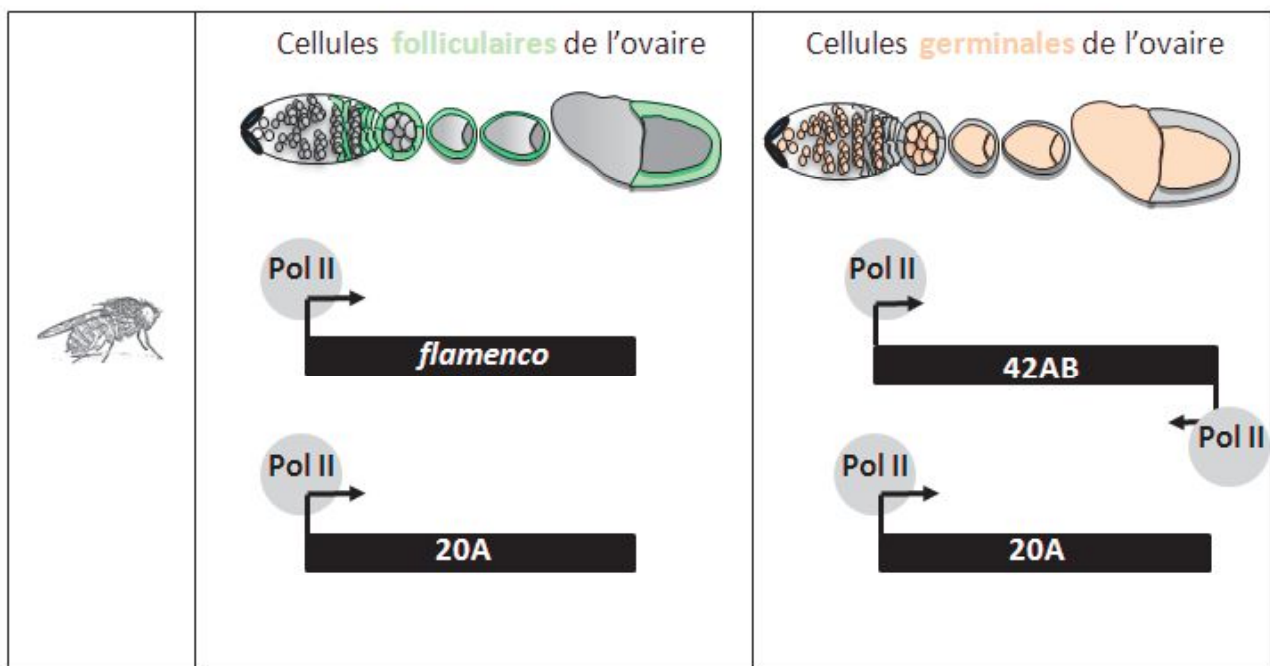
Les clusters de piRNAs ne représentent qu'une infime partie des génomes. A titre d'exemple, les clusters de piRNAs représentent près de 3,5% du génome de drosophile. Chez cette même espèce, 142 piRNAs clusters ont été identifiés et la majorité d'entre eux sont localisés dans des régions hétérochromatiques péri-centromériques ou télomériques (Brennecke et al., 2007). Parmi les 142 clusters de piRNAs, seuls sept d'entre eux ont été localisés dans des régions euchromatiques. À la suite de leurs découvertes chez la drosophile et la souris, les clusters de piRNAs ont été étudiés dans de nombreuses autres espèces





**Figure 22 : Les trois types de clusters de piRNAs identifiés.**

Les clusters de piRNAs sont classés en fonction de leurs propriétés transcriptionnelles. Les clusters de piRNAs « unistrand » regroupent l'ensemble des loci dont la transcription se fait à partir d'un seul brin d'ADN génomique. Les clusters de piRNAs « dual-strand » sont des loci transcrits à partir des deux brins d'ADN de façon convergente. Les clusters de piRNAs « bidirectional » sont assimilés à deux clusters de piRNAs « unistrand » transcrits de manière divergente.



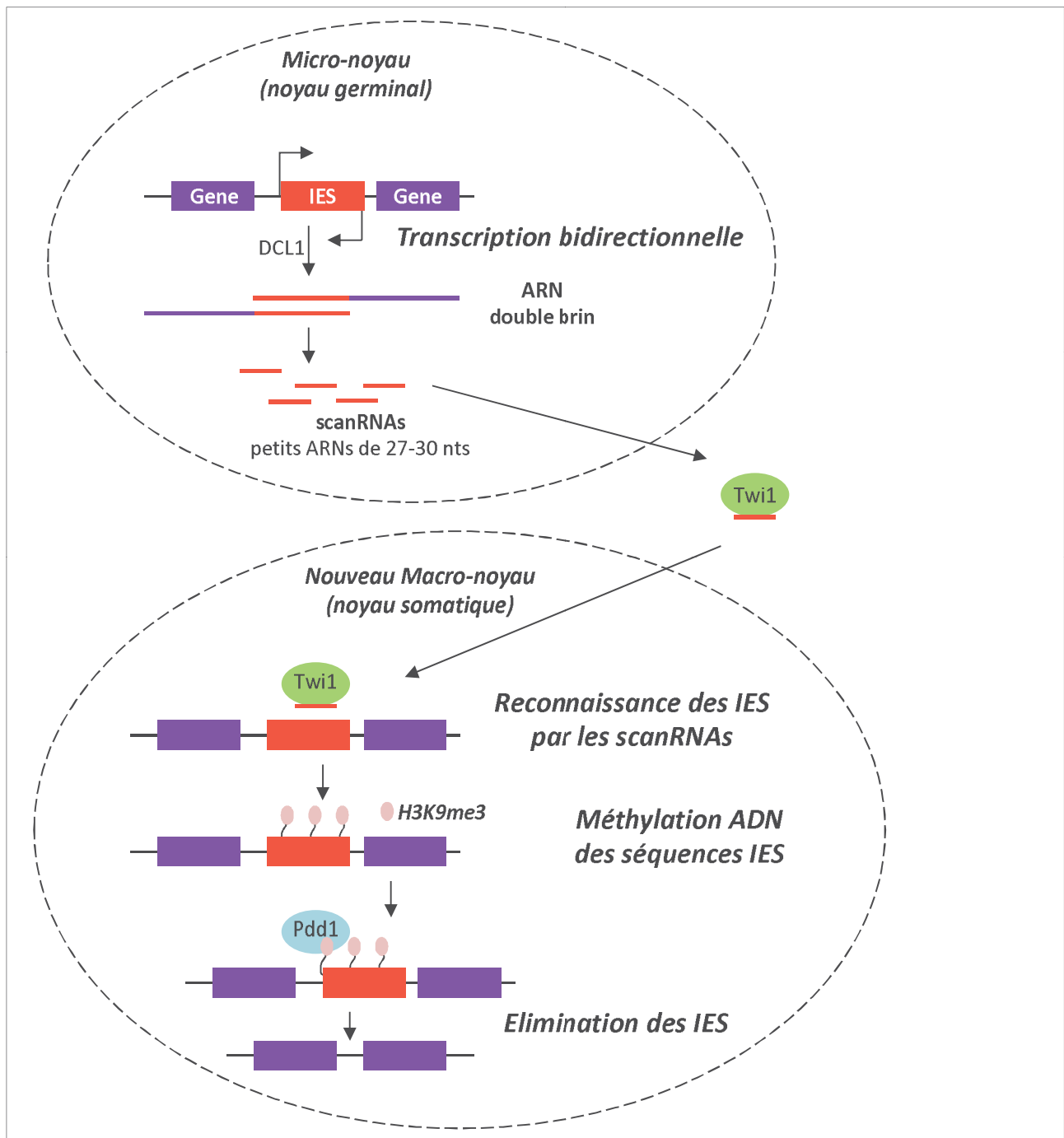
**Figure 23 : L'expression des clusters des piRNAs au sein des gonades femelles de drosophile.**

Au sein des cellules folliculaires des ovaires de drosophile (en vert), seuls les clusters de piRNAs « unistrand » sont exprimés. Dans les cellules germinales (ici en saumon), ce sont les clusters de piRNAs « dual-strand » qui sont majoritairement exprimés. Néanmoins, certains clusters de piRNAs « unistrand », tels que le locus 20A, s'expriment également dans ces cellules.

animales telles que *C. elegans* et le poisson-zèbre (Girard et al., 2006 ; Aravin et al., 2006 ; Ruby et al., 2006 ; Houwing et al., 2007 ; Grimson et al., 2008). Chez *C. elegans*, la majorité des petits ARNs 21U (piRNAs) proviennent de nombreux loci individuels regroupés en deux grandes régions localisées sur le chromosome IV. Ces larges régions génomiques ont été identifiées comme clusters de piRNAs (Ruby et al., 2006). Chez les mammifères, les clusters de piRNAs sont difficilement distinguables des unités de transcription de l'ARN polymérase II. De façon surprenante, la distribution des clusters de piRNAs chez les mammifères, tels que la souris, le rat et l'Homme, est relativement bien conservée alors que les piRNAs produits par ces mammifères ne semblent pas conservés (Girard et al., 2006 ; Lau et al., 2006).

Les clusters de piRNAs de la drosophile sont les plus étudiés et donc les mieux documentés dans la littérature. La classification des clusters de piRNAs chez la drosophile est basée sur leurs propriétés transcriptionnelles (Figure 22). Deux types de clusters de piRNAs ont ainsi pu être identifiés : les clusters de piRNAs "unistrand" c'est-à-dire transcrits à partir d'un seul des deux brins d'ADN, et les clusters de piRNAs "dual-strand" où les deux brins d'ADN sont transcrits de façon convergente (Brennecke et al., 2007 ; Malone et al., 2009 ; Senti et al., 2010 ; Yamanaka et al., 2014). Chez les mammifères, les loci producteurs de piRNAs sont majoritairement unistrands puisqu'ils sont transcrits unidirectionnellement par l'ARN polymérase II.

Chez la drosophile, les clusters de piRNAs "unistrand" sont majoritairement exprimés dans les cellules folliculaires de l'ovaire (Figure 23). Néanmoins, certains d'entre eux, tels que le cluster de piRNAs 20A, s'expriment également dans les cellules germinales de l'ovaire (Lau et al., 2009 ; Brennecke et al., 2007). Par ailleurs, les clusters de piRNAs "unistrand" murins, mais aussi ceux de *C. elegans*, sont exprimés au niveau des cellules germinales. Contrairement aux piRNA clusters « unistrand », les piRNAs cluster « dual-strand » sont uniquement retrouvés, à ce jour, chez les diptères, les lépidoptères ainsi que chez certains arthropodes (Klattenhoff et al., 2009 ; Mohn et al., 2014). Chez la drosophile, les clusters de piRNAs « dual-strand » sont exclusivement exprimés dans les cellules germinales de l'ovaire. Chez les mammifères, une troisième classe a été identifiée et nommée les clusters de piRNAs « bidirectionnel ». Ils ont été baptisés ainsi car ils sont transcrits *via* un promoteur central d'où émane une transcription divergente. C'est notamment le cas des clusters de piRNAs pachytènes murins (Aravin et al., 2006 ; Girard et al., 2006).



**Figure 24 : Voie des scanRNAs chez *Tetrahymena*.**

La transcription bidirectionnelle des séquences dérivant des séquences d'ETs (IES) dans le micronoyau (MIC, noyau germinal) engendre la formation d'ARNs double brin. Les longs transcrits double brin sont alors clivés par l'enzyme DCL1 (Dicer-like 1 protein) en petits ARNs de 27 à 30 nts, les «scanRNAs». Ces petits ARNs sont ensuite chargés sur les protéines Twi1. Une fois acheminés au nouveau Macronoyau (MAC, noyau somatique), les complexes scanRNA/Twi1 reconnaissent, par complémentarité de base, les séquences IES engendrant la mise en place de marques histones H3K9m3 à leur niveau. Les marques histones H3K9m3 sont ensuite reconnues par la protéine Pdd1 *via* son domaine chromoshadow. La liaison de la protéine Pdd1 sur les régions IES induit le recrutement de la machinerie impliquée dans l'excision des séquences IES. Après élimination des séquences IES, l'ADN du nouveau MAC est uniquement constitué de séquences codantes.

## Chapitre 3 : La biogenèse des piRNAs

Les piRNAs sont une large population de petits ARNs de 23 à 29 nts enrichis dans les tissus gonadiques. De nombreuses recherches ont été menées chez les métazoaires afin de comprendre, en détail, comment ces petits ARNs de 23 à 29 nts, dérivant des clusters de piRNAs, sont produits au sein des cellules germinales.

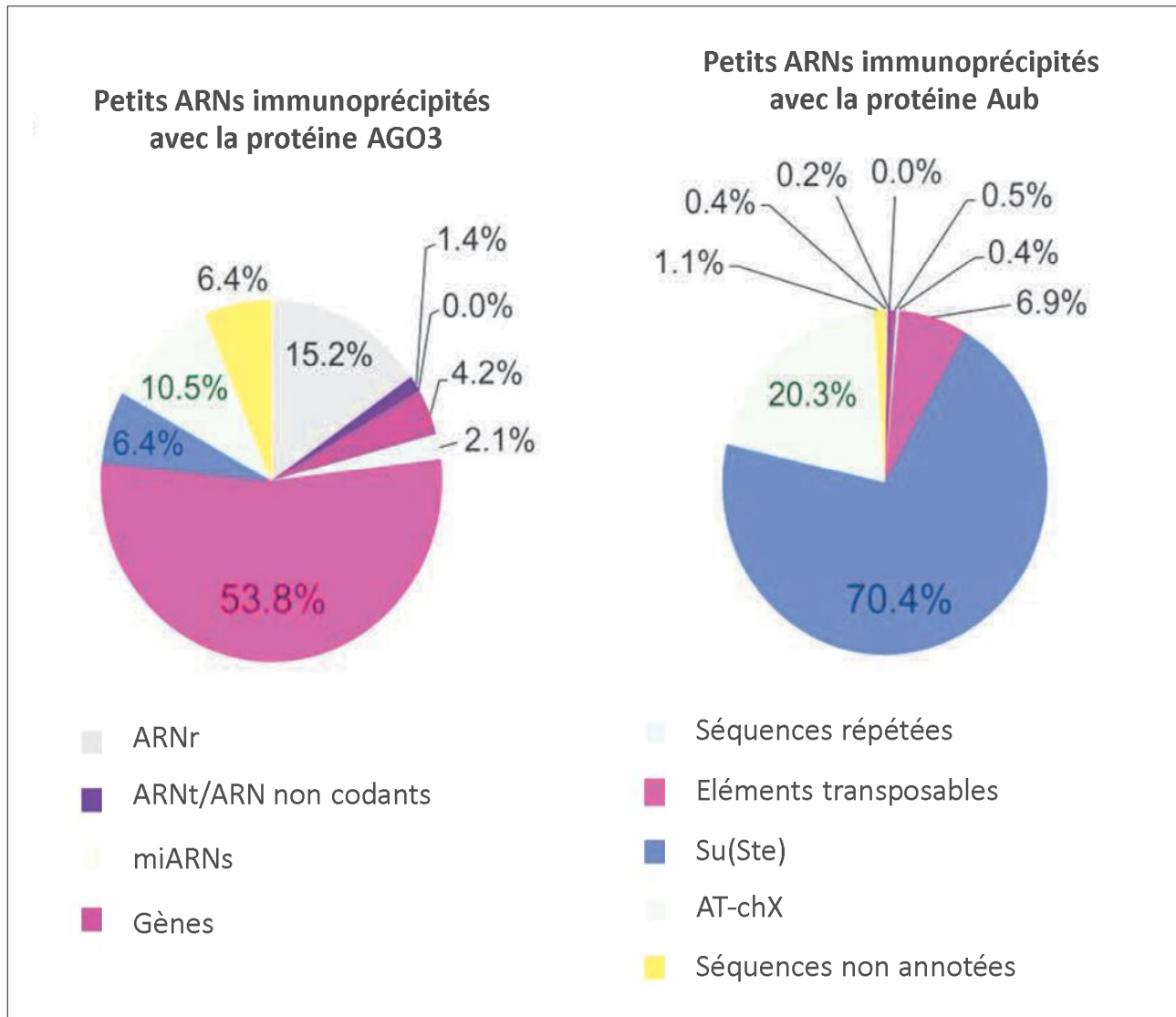
La biogenèse des piRNAs est relativement bien conservée chez les métazoaires à l'exception de certaines espèces telles que la plupart des nématodes (Sarkies et al., 2015 ; Grimson et al., 2008). La biogenèse des piRNAs peut être le résultat de deux voies distinctes qui ont été relativement bien décrites chez la drosophile, mon modèle d'étude. D'une part, les piRNAs peuvent être produits par la voie de biogenèse dite "primaire". La production de piRNAs est alors réalisée par la maturation des longs transcrits simple brin issus de la transcription des clusters de piRNAs. D'autre part, les piRNAs peuvent être produits par la voie d'amplification des piRNAs appelée la voie secondaire ou encore le cycle ping-pong. Contrairement à la biogenèse des piRNAs primaires, le cycle ping-pong est exclusivement germinale. Les piRNAs qui en sont issus dérivent des transcrits provenant des clusters de piRNAs, mais aussi, des transcrits d'ETs actifs. La maturation de ces transcrits dépend de deux protéines de la famille PIWI qui ont une expression exclusivement germinale, Aubergine (Aub) et AGO3.

Contrairement à de nombreuses espèces eucaryotes, les petits ARNs de 27-30 nts produits par *Tétrahymena* émanent de la maturation d'un ARN double brin (Mochizuki et al., 2002; Mochizuki et Gorovsky, 2005). Les piRNAs de *Tétrahymena* sont alors appelés des scanRNAs et dérivent des séquences internes éliminées qui regroupent des régions répétées du génome, des ETs, ainsi que des régions enrichies en A-T (Figure 24).

### *A. Organismes modèles utilisés pour explorer la voie des piRNAs*

Les études visant à explorer la voie des piRNAs ont été menées dans les deux modèles animaux où ils avaient été initialement identifiés, à savoir la drosophile et la souris. De façon intéressante, la voie des piRNAs a été explorée au sein des gonades femelles chez la drosophile, alors que chez la souris, cette voie a été examinée au sein des gonades mâles.

La voie des piRNAs chez la drosophile a été décrite comme étant active aussi bien dans les ovaires que dans les testicules de drosophile. Une des raisons qui permet d'expliquer



**Figure 25 : Profils des petits ARN associés à la protéine AGO3 et à la protéine Aub dans les testicules de drosophile.**

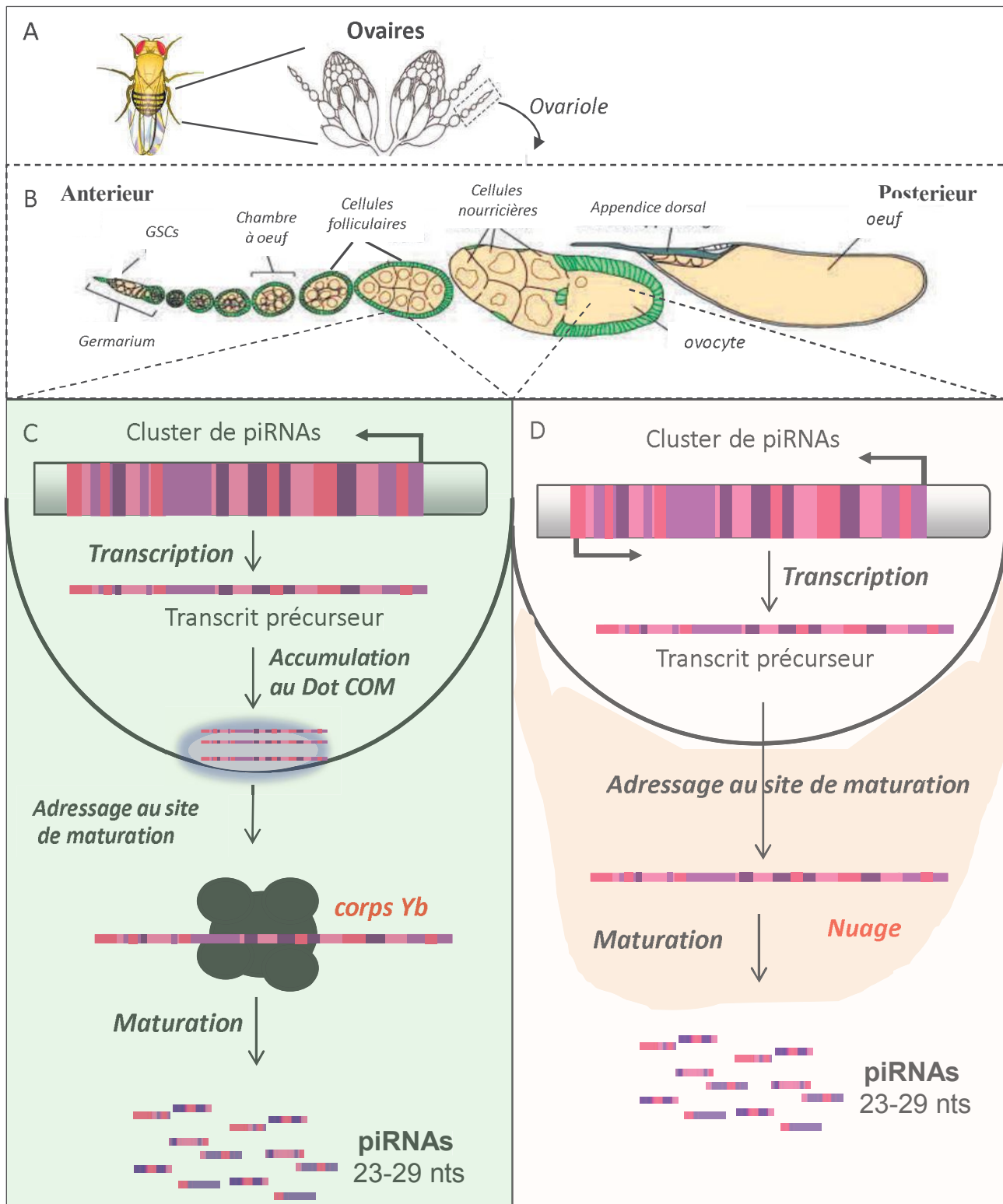
Parmi les piRNAs immunoprécipités avec la protéine AGO3, les piRNAs dérivant des séquences d'éléments transposables sont les plus abondants. La classe la plus abondante de piRNAs associés à Aub est celle dérivée des transcrits du locus *Suppressor of Stellate* [Su (Ste)]. Les piRNAs dérivant des séquences d'éléments transposables ne représentent que 6.9% des piRNAs immunoprécipités avec la protéine Aubergine (Aub).

Figure adaptée de la figure 2C de l'article Nagao et al., 2010.

que l'étude de la voie des piRNAs a été menée préférentiellement sur les ovaires de drosophile plutôt que sur les testicules de drosophile est liée à la production des piRNAs dans ces deux tissus. En effet, la voie des piRNAs a été décrite comme permettant la régulation fine des ETs au sein des tissus gonadiques chez les animaux. Contrairement aux piRNAs produits dans les ovaires de drosophiles, les piRNAs produits au niveau des gonades mâles dérivent principalement de loci dépourvus de séquence d'ETs tels que le locus *Suppresseur de Stellate* et d'un autre locus situé sur le chromosome X (Nishida et al., 2007). En effet, seuls 10% des piRNAs chargés sur la protéine Aub et 54% des piRNAs chargés sur la protéine Ago3 dérivent d'ETs (Nagao et al., 2010) (Figure 25). L'absence de l'une de ces deux protéines n'induit pas ou peu de dérégulation des ETs dans les testicules de drosophile, alors que l'inverse est observable dans les ovaires de drosophile (Brennecke et al., 2007 ; Nagao et al., 2010).

À l'inverse de la drosophile, la voie des piRNAs chez la souris a été préférentiellement décrite au sein des testicules. Dans cette espèce, seule l'altération de la voie des piRNAs chez le mâle entraîne une dérégulation des ETs mais également une stérilité (Reuter et al., 2011). Chez les souris femelles, aucune stérilité n'est observée notamment en absence de l'une des protéines PIWI murines (Carmell et al., 2007, Deng et Lin, 2002, Kuramochi-Miyagawa et al., 2004). Malgré le fait que les protéines PIWI soient exprimées dans les ovocytes murins, seule une infime quantité de piRNAs a été détectée dans ces cellules (Aravin et al., 2008 ; Ding et al., 2013 ; Lim et al., 2013). Cependant, de façon intéressante, les siRNAs sont, quant à eux, relativement nombreux au sein des ovocytes murins (Tam et al., 2008 ; Watanabe et al., 2008). En effet, les siRNAs sont retrouvés en de plus grande quantité que les piRNAs dans les ovocytes murins, alors que cette balance au sein des ovocytes est inversée chez les autres mammifères. Cette différence de production de petits ARNs, propre au groupe murin, peut être expliquée par le fait que la souris a perdu, au cours de son évolution, l'une des quatre protéines PIWI présentes chez les mammifères. En parallèle, il a été démontré qu'une isoforme de DICER a été créée suite à l'insertion d'un rétrotransposon (Flerm et al., 2013). Cette isoforme est exprimée spécifiquement dans les ovocytes murins et pourrait être le reflet d'un avantage sélectif pour amplifier la production d'endosRNAs afin de réguler les ETs.

Ces différentes données permettent ainsi d'expliquer pourquoi chez la souris l'étude de la voie des piRNAs a été réalisée sur des gonades mâles, alors que chez la drosophile, elle a été menée sur des gonades femelles. La voie de biogenèse des piRNAs décrite par la suite est principalement focalisée sur la voie des piRNAs au sein de mon modèle d'étude, la



**Figure 26 : La voie de biogenèse primaire des piRNAs présente dans les ovaires de drosophile.**

A) Représentation schématique du tissu reproducteur de la femelle drosophile constitué d'une paire d'ovaires renfermant chacun environ 16 ovarioles. B) Représentation schématique d'une ovariole. La partie la plus antérieure et la plus précoce de l'ovariole est appelée le germarium. Il contient les cellules souches germinales et folliculaires. Le germarium est suivi d'une succession de follicules à différents stades de maturation jusqu'à la formation d'un œuf. Chaque follicule est composé de 16 cellules germinales (saumon) entourées de cellules folliculaires (vert). C et D) Voie simplifiée de la biogenèse des piRNAs primaires active dans les cellules folliculaires (C) et dans les cellules germinales (D).

drosophile. Au sein du tissu reproducteur femelle de la drosophile, les ovaires, la voie de biogenèse des piRNAs a été décrite comme étant active dans les cellules germinales, mais également dans les cellules somatiques qui les entourent, les cellules folliculaires (Figure 26). Néanmoins, la voie des piRNAs des cellules folliculaires et celle des cellules germinales de drosophile sont quelque peu différentes en termes d'acteurs protéiques, mais également en termes de piRNAs produits.

### ***B. La biogenèse des piRNAs primaires***

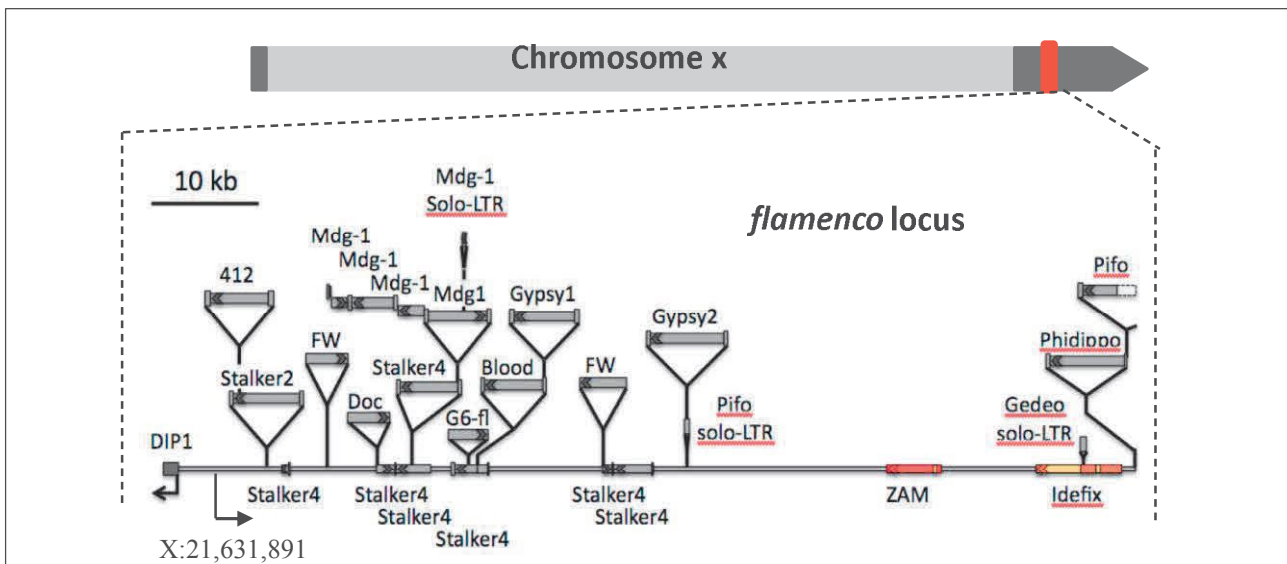
La voie de biogenèse des piRNAs primaires chez la drosophile est active aussi bien dans les cellules germinales que dans les cellules folliculaires. La production des piRNAs par la voie de biogenèse primaire peut être déclinée en trois grandes étapes (Figure 26). L'initiation de la voie de biogenèse des piRNAs primaire débute par la transcription des clusters de piRNAs. Dans un second temps, les longs transcrits simple brin issus de la transcription des clusters de piRNAs sont reconnus par la cellule pour être acheminés à leur site de maturation. Une fois adressés à leur site de maturation, les longs transcrits simple brin sont coupés en petits ARNs de 23 à 29 nts alors chargés sur l'un des membres de la sous-famille PIWI.

Chez *C. elegans*, la biogenèse des piRNAs primaires est quelque peu différente de celle rencontrée chez la drosophile. Contrairement aux transcrits précurseurs de piRNAs de drosophile qui sont de longs ARNs simple brin, les transcrits précurseurs de piRNAs chez *C. elegans* sont de petits transcrits de 25 à 26 nts. Dans un second temps, une première coupure a lieu précisément entre le deuxième et le troisième nucléotide de ces transcrits générant ainsi l'extrémité 5' des futurs endosiRNAs. Par la suite, les petits ARNs chargés sur une protéine homologue à la protéine Piwi de drosophile, la protéine PRG-1. Enfin, la taille des petits ARNs chargés sur la protéine PRG-1 est ajustée à 21 nts (Ruby et al., 2006 ; Cecere et al., 2012).

#### ***I. Les transcriptions des piRNAs clusters chez *Drosophila melanogaster****

La transcription des clusters de piRNAs « unistrand » et « dual-strand » ainsi que les protéines impliquées dans leur transcription respective sont différentes.

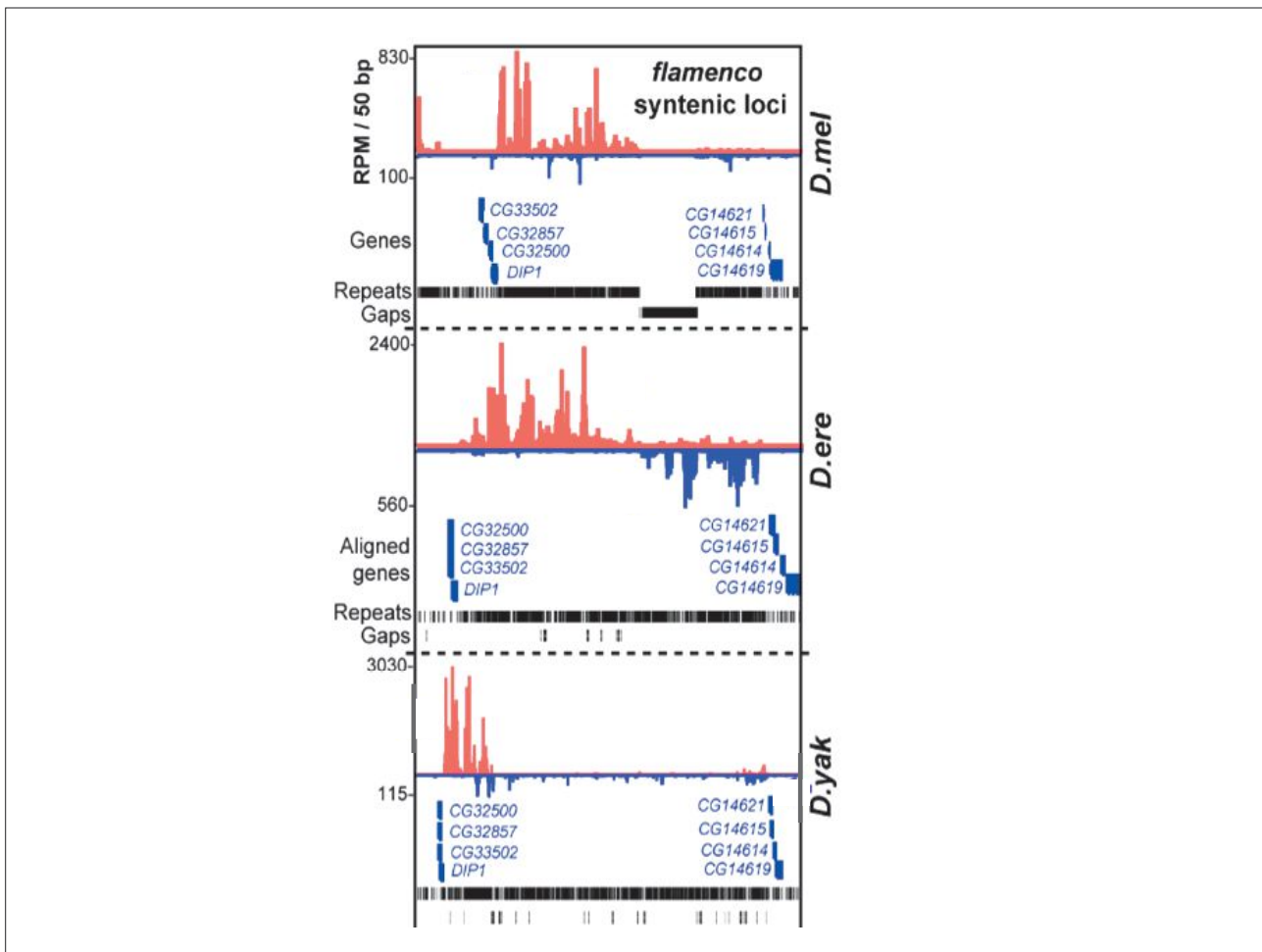




**Figure 27 : Structure du cluster de piRNAs *flamenco* chez *D. melanogaster*.**

Le locus *flamenco* est localisé dans la région péri-centromérique du chromosome X. Il est situé plus précisément à 1,7 Kb du gène DIP1. Il est composé d'un enchevêtrement d'éléments transposables. Seuls les éléments transposables récents ont été représentés.

Figure adaptée de la figure 1 de l'article Zanni et al., 2013.



**Figure 28 : Capture d'écran du « genome browser » des loci orthologues au locus *flamenco* chez *D. melanogaster*, *D. erecta* et *D. yakuba*.**

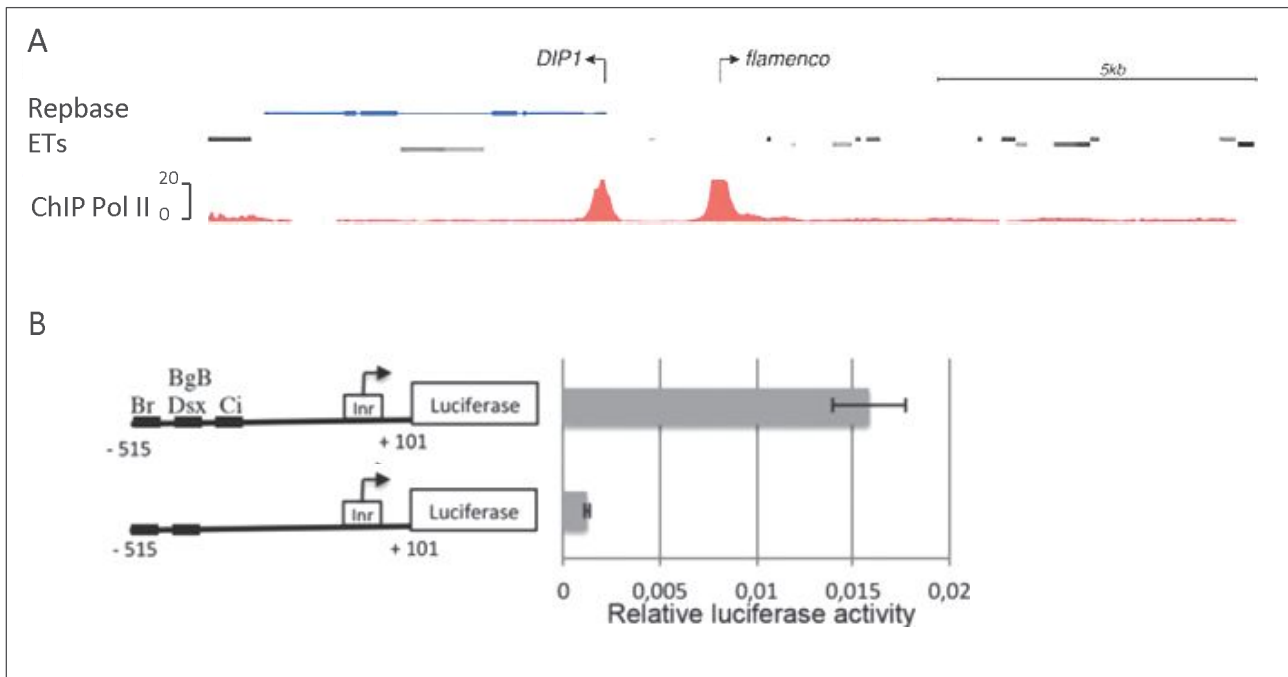
Pour chaque espèce, les piRNAs sens sont représentés par des pics rouges et les piRNAs orientés en anti-sens par rapport au sens de transcription sont, quant à eux, représentés en bleu. Figure adaptée de la figure 1H de l'article Chirn et al., 2015.

### a) Les clusters de piRNAs « unistrand »

Les clusters de piRNAs «unistrand» sont les seuls clusters de piRNAs à s'exprimer au sein des cellules somatiques des ovaires de drosophile, les cellules folliculaires. Néanmoins, certains clusters de piRNAs «unistrand», tels que le cluster de piRNAs *20A*, s'expriment également au sein des cellules germinales.

Le séquençage des piRNAs provenant des cellules OSS (cellules dérivées des cellules souches somatiques ovariennes) a permis de mettre en évidence que près de 80% des piRNAs uniques produits par ces cellules proviennent d'un locus de plus de 180 Kb situé au niveau de la région péricentromérique du chromosome X, le locus *flamenco*. Ce même locus avait initialement été décrit génétiquement comme impliqué dans la régulation de trois ETs : *gypsy*, *ZAM* et *Idéfix* (Prud'homme et al., 1995 ; Desset et al., 1999). En d'autres termes, le cluster de piRNAs *flamenco* est le producteur majeur de piRNAs au sein des cellules folliculaires (Malone et al., 2009). Le cluster de piRNAs *flamenco* est composé à 95% d'ETs orientés du centromère au télomère (Zanni et al., 2013). Cinquante-deux ETs ont été identifiés dans le cluster de piRNAs *flamenco* dont trois d'entre eux, *mdg1*, l'élément *F* et l'élément *Stalker 4*, sont présents en multicopies (Figure 27). Les piRNAs produits par *flamenco* sont orientés en anti-sens par rapport aux séquences d'ETs qui le composent, suggérant que la transcription de *flamenco* est dirigée du télomère au centromère. Néanmoins, aucun promoteur n'avait à l'époque été identifié.

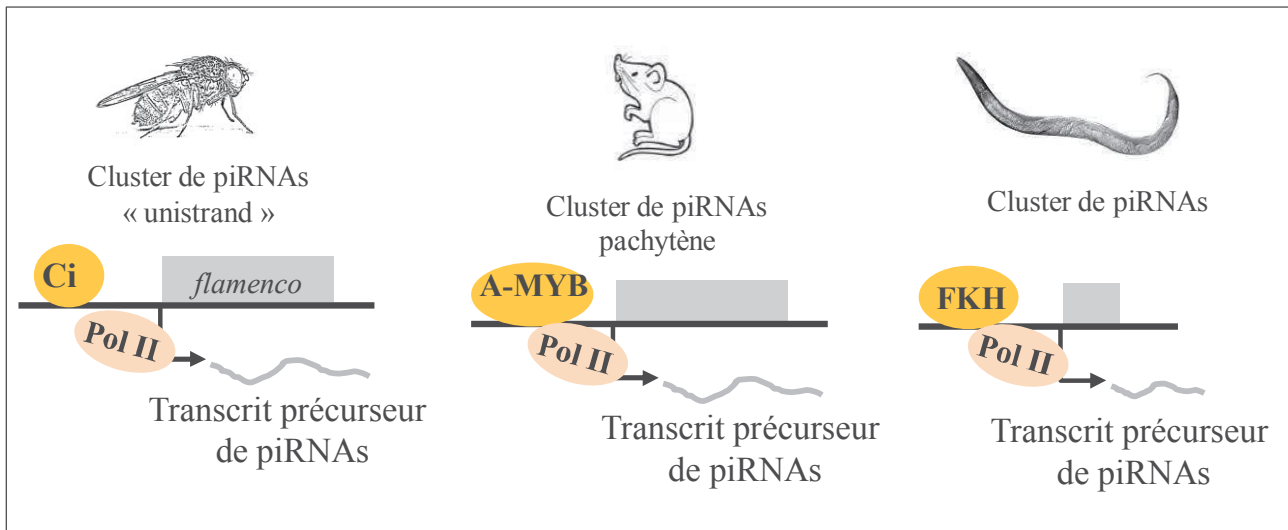
Au cours de l'évolution des drosophiles, le locus *flamenco* serait apparu il y a environ 10 à 12 millions d'années selon des études comparatives menées par trois équipes indépendantes (Malone et al., 2009 ; Goriaux et al., 2014 ; Chirn et al., 2015). Grâce aux alignements multiples, des loci orthologues au cluster de piRNAs *flamenco* ont été identifiés notamment chez *D. yakuba*, *D. simulans* et *D. erecta*. Ces loci sont également localisés dans la région péricentromérique du chromosome X, proche du gène *dip1* (Figure 28). Ils sont enrichis en séquences d'ETs dont l'orientation est majoritairement en anti-sens par rapport à leur transcription. Les ETs qui les composent appartiennent majoritairement à la famille *gypsy*. Pour vérifier que ces locus étaient des clusters de piRNAs fonctionnels, les petits ARNs produits dans les ovaires de *D. erecta* ont été séquencés, puis bio-informatiquement analysés (Chirn et al., 2015). Tout comme chez *D. melanogaster*, le locus *flamenco*-like de chez *D. erecta* produit de nombreux piRNAs à partir du même brin génomique. Les propriétés transcriptionnelles des clusters de piRNAs « unistrand » ont été explorées en prenant comme



**Figure 29 : Propriétés transcriptionnelles du cluster de piRNAs *flamenco* chez *D. melanogaster*.**

A) Profil du recrutement de la polymérase II (Pol II, en rouge) le long de la région environnant le cluster de piRNAs *flamenco* obtenu par ChIP-seq Pol II effectué sur les cellules OSS. Les séquences d'ETs sont représentés par des barres grises au dessus des résultats du ChIP-seq Pol II. Figure adaptée de la figure S6 de l'article Sienski et al., 2012.

B) Mesure de l'activité luciférase après des transfections transitoires de plasmides possédant ou non le site de liaison du facteur de transcription *Cubitus interruptus* (Ci). Figure adaptée de la figure 3C de l'article Goriaux et al., 2014.



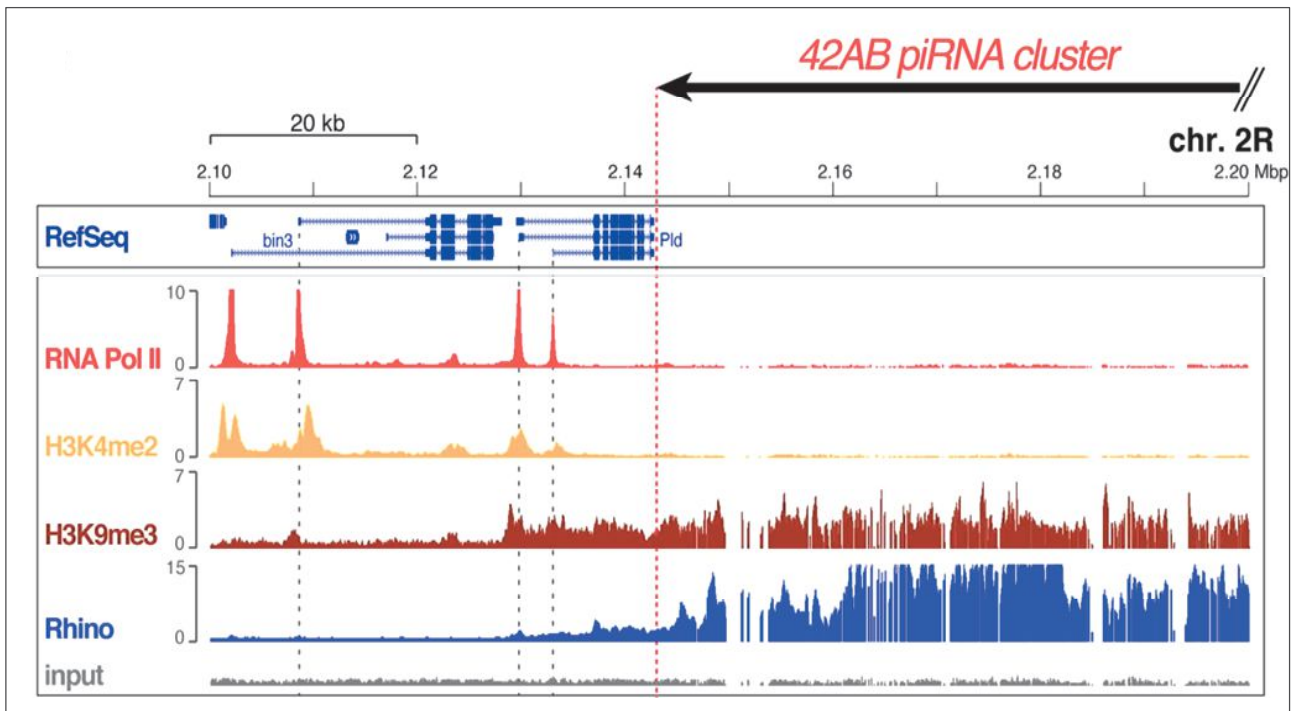
**Figure 30 : Activation de la transcription des clusters de piRNAs « unistrand » chez la drosophile, chez la souris et chez *C. elegans*.**

Trois facteurs de transcription non spécifiques aux clusters de piRNAs « unistrand » ont été identifiés. Les facteurs de transcription *Cubitus interruptus* (Ci), A-MYB et Forkhead (FKH) sont requis pour l'activation de la transcription des clusters de piRNAs « unistrand » respectivement chez la drosophile, chez la souris et chez *C. elegans*.

modèle le cluster de piRNAs *flamenco*. Les analyses comparatives des orthologues de *flamenco* couplées à des analyses transcriptionnelles réalisées dans notre équipe ont permis d'identifier le site majeur d'initiation de transcription de *flamenco* situé à la position X :21 631 891 (Release R6), soit à 1733 bases en amont du gène *dip1*. Des analyses bio-informatiques ont permis de découvrir un promoteur de type INR DPE dépendant de l'ARN polymérase II. Ces résultats ont été confirmés par des expériences de CHIP-Seq révélant une accumulation spécifique de l'ARN polymérase II (Sienski et al., 2012) au site d'initiation de la transcription que notre équipe a identifié (Figure 29A). L'étude des marques chromatinienne au niveau de *flamenco* a révélé que ce locus était pauvre en marques histones répressives H3K9me3, contrairement au cluster de piRNAs *20A* qui, quant à lui, semble être enrichi en marques H3K9me3 dans son extrémité 5' (Mohn et al., 2014). La transcription dépendante de l'ARN polymérase II des clusters de piRNAs « unistrand » est un paramètre conservé chez les animaux tels que chez *C. elegans* et chez les mammifères (Li et al., 2013 ; Ruby et al., 2006).

L'activation de la transcription du cluster de piRNAs *flamenco* est sous le contrôle du facteur de transcription *Cubitus Interruptus* (Ci) (Figure 29B) (Goriaux et al., 2014). De façon surprenante, chez les animaux, les facteurs de transcription identifiés comme étant impliqués dans l'activation transcriptionnelle des clusters de piRNAs ne sont pas uniquement dédiés à la transcription des clusters (Figure 30). En effet, l'activation de la transcription des clusters de piRNAs « unistrand » est placée sous le contrôle du facteur de transcription A-MYB chez la souris et du facteur de transcription Forkhead chez *C. elegans* (Li et al., 2013 ; Ruby et al., 2006). Les clusters de piRNAs de *C. elegans* ont la particularité de posséder un motif CTGTTTCA situé 40 nucléotides en amont de leurs promoteurs et encadré par des séquences riches en AT. En outre, les séquences enrichies en AT sont reconnues spécifiquement par le facteur de transcription Forkhead qui est responsable de l'expression des clusters de piRNAs (Cecere et al., 2012).

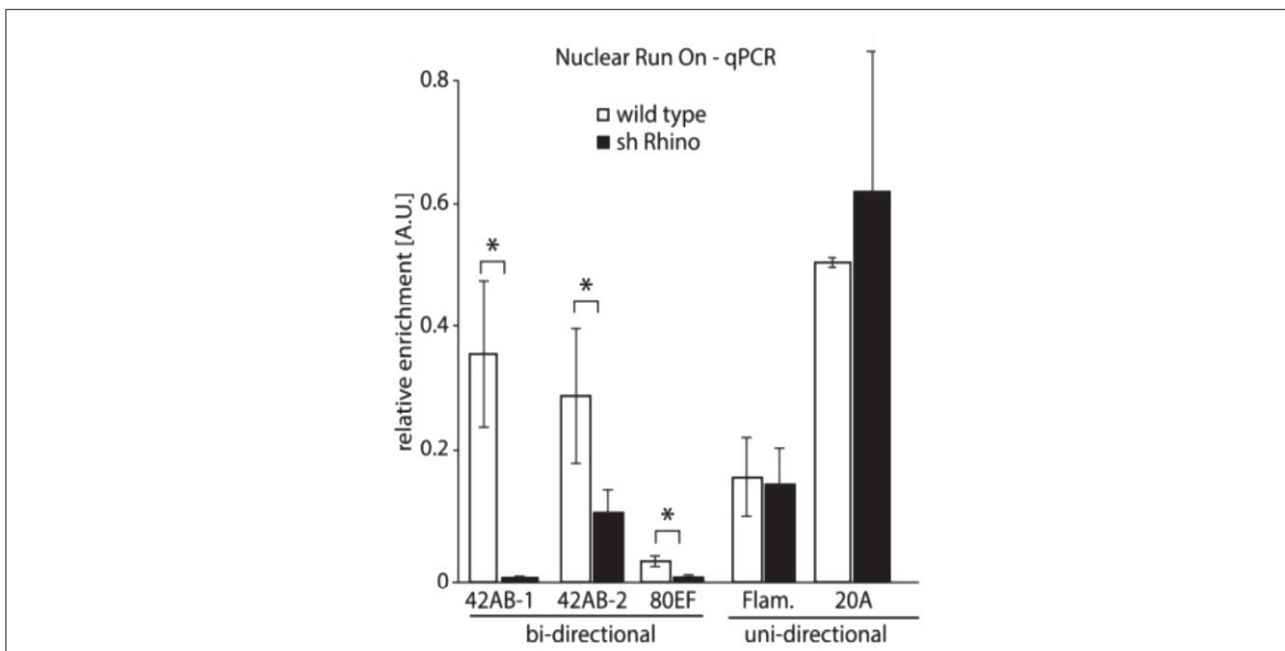
Les transcrits précurseurs des clusters de piRNAs folliculaires tels que le piRNA cluster *flamenco* sont alternativement épissés, donnant lieu à de multiples transcrits qui seront ensuite maturés en piRNAs (Goriaux et al., 2014). Par ailleurs, les transcrits issus des clusters de piRNAs « unistrand » sont également coiffés à leurs extrémités 5' (Mohn et al., 2014). Tout comme les transcrits des clusters de piRNAs « unistrand » chez la drosophile (Chen et al., 2016), les transcrits de clusters de piRNAs « unistrand » des mammifères sont polyadénylés.



**Figure 31 : Caractéristiques transcriptionnelles du cluster de piRNAs « dual-strand » majoritaire dans les cellules germinales des ovaires de drosophile, le locus *42AB*.**

L'enrichissement de la polymérase II (Pol II) sur le locus *42AB* est représenté en rouge, celui des marques histones H3K4me2 en jaune, celui des marques histones H3K9me3 en marron et celui de la protéine Rhino en bleu.

Figure adaptée de la figure 1D de l'article Mohn et al., 2014.



**Figure 32 : La protéine Rhino est requise pour l'expression des clusters de piRNAs « dual-strand ».**

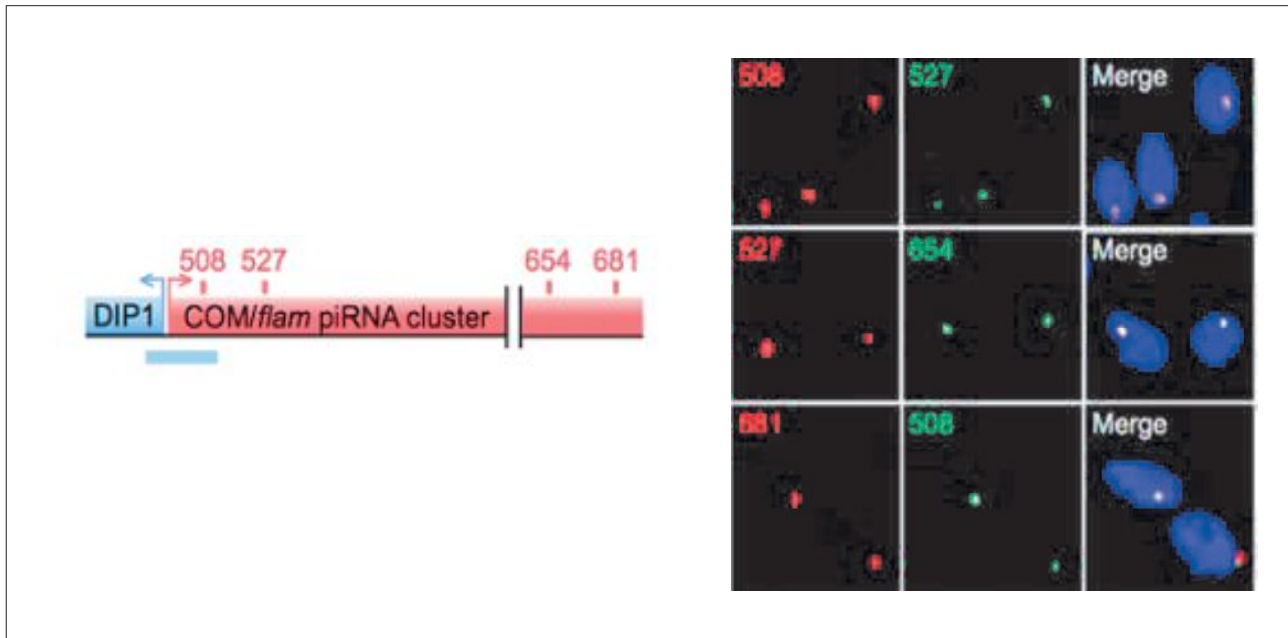
A la suite d'un « Run On », les transcrits naissant issus des clusters « dual-strand » et des clusters « unistrand », en condition sauvage et en absence de la protéine Rhino, ont été quantifiés par qRT-PCR.

Figure adaptée de la figure 6C de l'article Le Thomas et al., 2014.

## **b) Les clusters de piRNAs « dual-strand »**

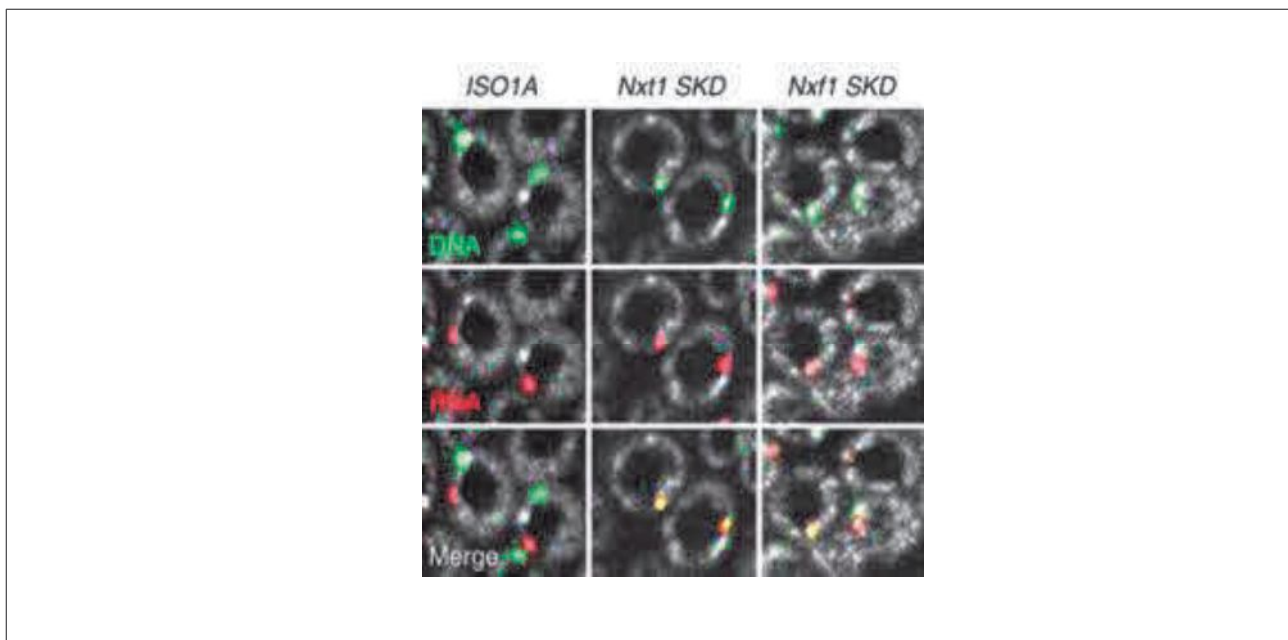
Contrairement aux clusters de piRNAs « unistrand », les clusters de piRNAs « dual-strand », sont exclusivement exprimés dans les cellules germinales des ovaires de drosophile. Les producteurs majeurs de piRNAs au sein des cellules germinales d'ovaires de drosophile sont les clusters de piRNAs *42AB*, *80F* et *38C1* (Brennecke et al., 2007). Le piRNA cluster *42AB* est un piRNA cluster qui s'étend approximativement sur 240 Kb. Contrairement à *flamenco*, les ETs qui le composent sont orientés de façon aléatoire. Les piRNAs produits à partir de ces sites génomiques sont donc orientés en anti-sens, mais également en sens par rapport aux séquences d'ETs présentes dans ces derniers. Etant le producteur majeur de piRNAs des cellules germinales, le piRNA cluster *42AB* a été étudié en détails par de nombreuses équipes.

Contrairement aux clusters de piRNAs « unistrand », les clusters de piRNAs "dual strand" ne sont pas exprimés grâce à la voie canonique de transcription. L'étude des propriétés transcriptionnelles des piRNA clusters « dual-strand », *via* des immunoprécipitations de chromatine contre l'ARN polymérase II (Figure 31) ainsi que des expériences de GRO-seq a révélé que des initiations de transcription étaient présentes tout du long de ces clusters « dual-strand », suggérant qu'ils ne possèdent pas de promoteurs propres uniques (Chen et al., 2016 ; Mohn et al., 2014, Le Thomas et al., 2014). La transcription des clusters de piRNAs « dual-strand » dépend du complexe RDC composé par les protéines Rhino, Deadlock et Cutoff dont la présence sur les clusters de piRNAs a notamment été mise en évidence lors des expériences de ChIP (Mohn et al., 2014 ; Pane et al., 2011 ; Klattenhoff et al., 2009). La protéine Rhino est une protéine hétérochromatique paralogue à la protéine HP1 permettant la mise en place d'un environnement hétérochromatique. La protéine Rhino reconnaît les marques répressives H3K9me3 *via* son chromo-domaine. En d'autres termes, l'expression des clusters de piRNAs « dual-strand » chez la drosophile requiert préalablement un dépôt des marques H3K9me3 le long de ces derniers (Figure 32) (Rangan et al., 2011). En effet, la déplétion de l'une des enzymes responsables de la mise en place des marques H3K9me3, Eggless, a pour conséquence d'induire une diminution de l'expression des clusters de piRNAs corrélée à un défaut de production des piRNAs (Le Thomas et al., 2014). Les marques H3K9me3 sur les clusters de piRNAs « dual-strand » sont déposées durant l'embryogenèse *via* les complexes piRNA/Piwi hérités maternellement (Akkouche et al., 2017). Il a été montré qu'une déplétion de la protéine Piwi durant l'embryogenèse engendre une perte, ou du moins, une diminution



**Figure 33 : Les transcrits de *flamenco* s'accumulent dans un même foyer nucléaire, le Dot COM.**

Double RNA FISH réalisé sur les ovaires de drosophiles *ISO1A*. Les sondes testées sont dirigées contre différentes régions de *flamenco* et sont représentées en rouge et en vert. Leurs positions respectives sont placées sur le locus *flamenco* schématisé à gauche des résultats des expériences de Double RNA FISH. Figure adaptée de la figure 1A et 1D de l'article Dennis et al., 2013.



**Figure 34 : Le transit nucléaire des transcrits de *flamenco* dépend des protéines Nxf1 et Nxt1.**

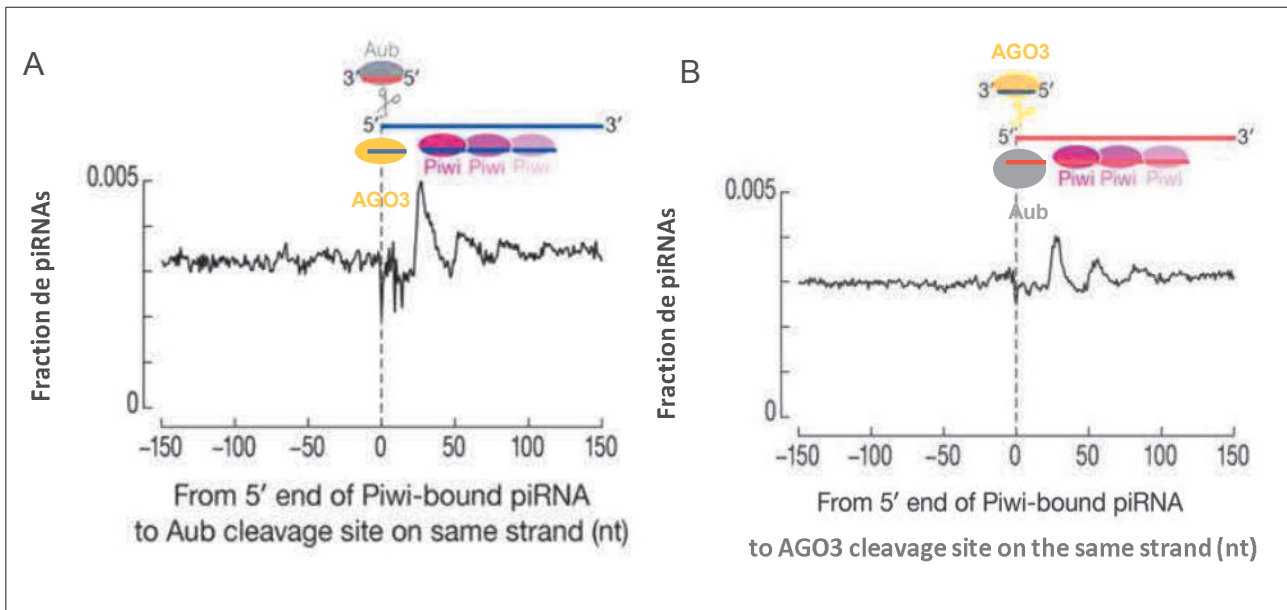
Expériences de double ADN/ARN FISH réalisées sur les ovaires de drosophiles en contexte sauvage (*ISO1A*) et en absence de la protéine Nxt1 ou de la protéine Nxf1 au sein des cellules folliculaires. Les sondes ciblant l'ADN de *flamenco* sont en vert et les sondes ciblant les transcrits de *flamenco* sont en rouge. Figure adaptée de la figure 2G de l'article Dennis et al., 2016.

des marques H3K9me3 sur l'ensemble des clusters de piRNAs « dual-strand » chez les adultes. En absence des marques histones H3K9me3, une diminution de l'expression des clusters de piRNAs est observée chez ces individus induisant une dépression d'ETs (Akkouche et al., 2017). De même, en condition mutante *rhino*, il a été observé par qRT-PCR que la quantité de transcrits des clusters de piRNAs « dual-strand » est diminuée impliquant une chute de la production de piRNAs. A titre d'exemple, la production de piRNAs à partir du cluster « dual-strand » *42AB* chute d'environ 97% dans des conditions mutantes pour Rhino. Une fois recrutée sur les clusters de piRNAs « dual-strand », la protéine Rhino interagit également avec la protéine Deadlock *via* son chromoshadow domaine. La protéine Deadlock recrute à son tour de façon transitoire la protéine Cutoff dont le rôle est d'inhiber l'épissage des transcrits naissants, mais également d'inhiber les signaux d'arrêt de la transcription de l'ARN polymérase II afin que cette dernière transcrive les clusters de piRNAs de façon continue (Pane et al., 2011; Mohn et al., 2014 ; Zhang et al., 2014 ; Chen et al., 2016). De plus, la protéine Cutoff en se liant aux transcrits naissants des clusters empêche la mise en place de la coiffe en 5' ce qui a pour conséquence d'abolir leur traduction. La protéine UAP56 est également retrouvée lié aux transcrits des clusters de piRNAs germinaux inhibant ainsi leur épissage. Les transcrits issus des clusters de piRNAs « dual-strand » ne sont pas polyadénylés. Récemment, l'équipe de Zamore a décrit que la protéine Maelstrom permet d'inhiber la transcription canonique des ETs composant les clusters de piRNAs afin d'éviter qu'elle ne puisse entrer en compétition avec la transcription du cluster de piRNAs lui-même (Chang et al., 2019). En effet, en absence de Maelstrom, les auteurs ont observé une augmentation de la transcription canonique à partir des promoteurs d'ETs composant les clusters entraînant une diminution de la production des piRNAs.

## II. Adressage des transcrits précurseurs de piRNAs à la voie des piRNAs

Avant d'être exportés au cytoplasme, les transcrits précurseurs des piRNAs des cellules folliculaires (« unistrand ») sont accumulés en un focus nucléaire, le Dot COM (Figure 33) (Dennis et al., 2013). Le transit nucléaire des transcrits de *flamenco* de son site de transcription au Dot COM dépend du complexe de l'exon-jonction puisqu'en absence d'un des constituants de ce complexe, il a été montré par des expériences de RNA/DNA FISH que les transcrits de *flamenco* restent localisés à leur site de transcription (Figure 34). Les protéines du complexe de l'exon-jonction sont connues pour être associées aux transcrits en cours d'épissage. Une fois acheminés et accumulés au Dot COM, les transcrits de *flamenco* sont





### Figure 35: Le mécanisme de Phasing chez la drosophile.

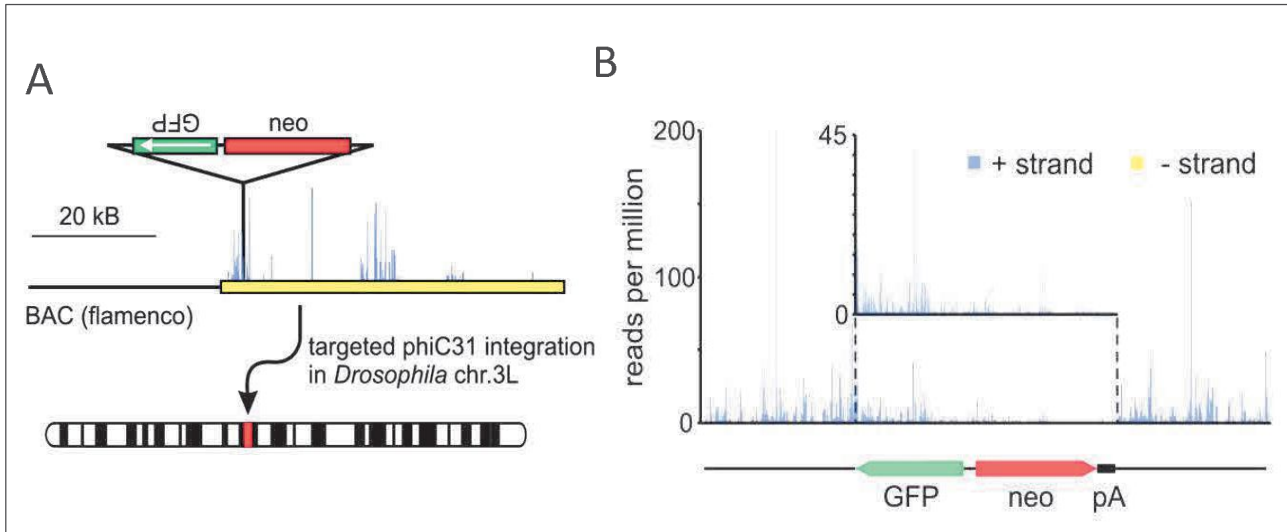
Analyse de la distance entre les extrémité 5' des piRNAs liés à Piwi et issus d'un transcrit et les sites de clivages générés par les protéines Aub (A) ou AGO3 (B) sur ce même transcrit. Les séquences ARNs en rouge sont en sens inverse des séquences ARNs représentées en bleu. Suite au cycle ping-pong réalisé par les protéines Aub et AGO3, le reste du transcrit est maturé de façon phasique par la protéine Zucchini et la protéine Piwi. Figure adaptée de la figure 3 de l'article Han et al., 2015.

exportés au cytoplasme *via* les exportines Nxt1 et Nxf1 et la protéine UAP56 (Dennis et al., 2016). Il est intéressant de noter qu'en absence de l'export cytoplasmique des transcrits de *flamenco*, l'assemblage des corps Yb, structure cytoplasmique multiprotéique où a lieu la maturation en piRNA, est altéré. Au sein des cellules germinales, le transit des transcrits de clusters de piRNAs à la membrane nucléaire est effectué par l'hélicase UAP56 (Zhang et al., 2012 ; Zhang et al., 2014). Après avoir traversés la membrane nucléaire, les transcrits précurseurs de piRNAs germinaux sont pris en charge par la protéine Vasa.

Une fois exportés dans le cytoplasme, les transcrits précurseurs de piRNAs doivent, dans un premier temps, être reconnus par la cellule comme étant des transcrits précurseurs de piRNAs pour être, dans un second temps, acheminés vers leur site de maturation. Néanmoins, l'analyse des transcrits précurseurs de piRNAs, notamment par des analyses bio-informatiques n'a pas permis d'identifier de structure particulière ou de motif particulier qui les distingueraient des ARNs messagers. Chez la drosophile, les transcrits de piRNAs clusters germinaux (« dual-strand ») sont des transcrits atypiques qui pourraient être facilement différenciés des ARNs messagers. En revanche, les transcrits de clusters folliculaires (« unistrand ») possèdent une coiffe en 5', sont épissés et polyadénylés tout comme les transcrits provenant de gènes. L'une des questions les plus énigmatiques concernant la biogenèse des piRNAs, qui reste à ce jour encore en suspens, est de savoir comment des transcrits provenant des clusters de piRNAs, et plus particulièrement provenant des piRNAs clusters « unistrand », sont sélectionnés pour être maturés en piRNAs.

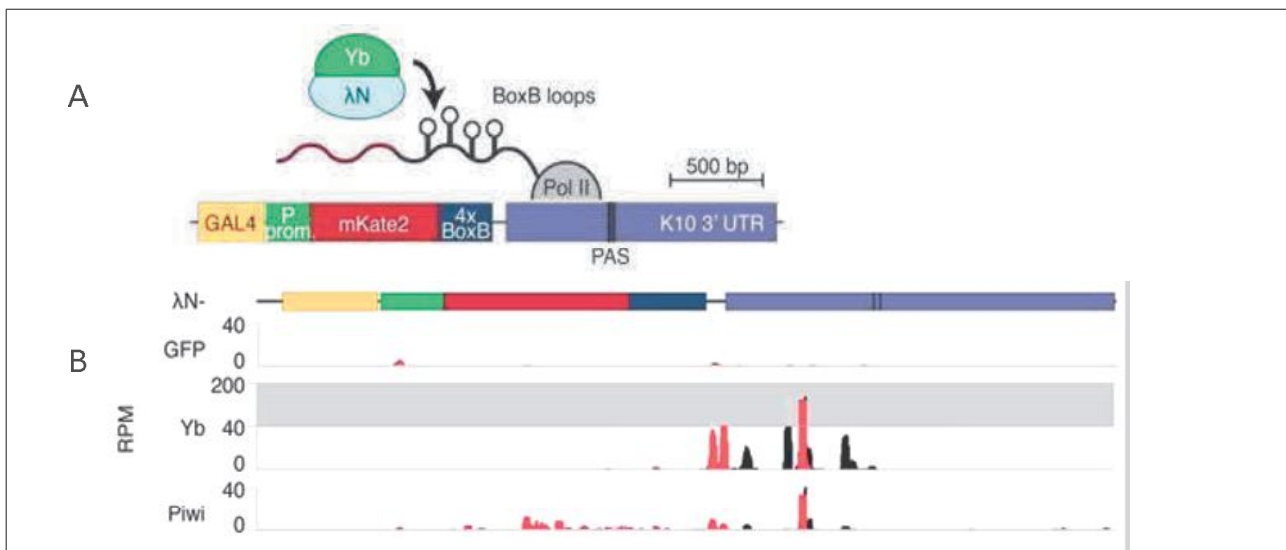
Au sein des cellules germinales de l'embryon, les complexes piRNA/PIWI hérités maternellement ont pour rôle d'initier la biogenèse des piRNAs (Le Thomas et al., 2014). En effet, la reconnaissance des transcrits naissant de clusters de piRNAs « dual-strand » par les complexes piRNA/PIWI hérités maternellement conduit à la mise en place de structures hétérochromatiques au niveau des clusters de piRNAs germinaux (structures essentielles à leur expression). Par ailleurs, la reconnaissance des transcrits de clusters de piRNAs « dual-strand », mais également des transcrits d'ETs actifs, par les complexes piRNA/PIWI hérités permet d'initier le cycle ping-pong. Il a alors été proposé que les protéines Aub génèrent la première coupure des transcrits formant ainsi le premier piRNA. Cette coupure initie alors la maturation du reste des transcrits par les protéines Piwi et Zucchini *via* le mécanisme de Phasing (Figure 35) (Han et al., 2015 ; Mohn et al., 2015).

Au sein des cellules folliculaires, la protéine Aub n'est pas exprimée suggérant que la reconnaissance des transcrits issus des clusters de piRNAs « unistrand » utilise un mécanisme



**Figure 36 : Production de piRNAs artificiels à partir d'un cluster de piRNAs ectopique chez *D. melanogaster*.**

A) Représentation schématique de l'insertion de la cassette GFP anti-sens/néomycine (Neo) dans le locus *flamenco* (en jaune) contenu dans un BAC. Par la suite, le BAC a été inséré dans le génome de la mouche dans un site spécifique localisé dans le chromosome 3. B) Alignement des piRNAs séquencés le long des différentes régions constituant le BAC. Les piRNAs sens sont représentés par des pics bleus et ceux étant en anti-sens par rapport au sens de transcription sont en jaune. Les piRNAs alignés sur les régions de *flamenco* flanquant la cassette GFP/Neo proviennent du *flamenco* endogène mais également du *flamenco* ectopique. Figure adaptée des figures 2A et 2C de l'article Muerdter et al., 2012.



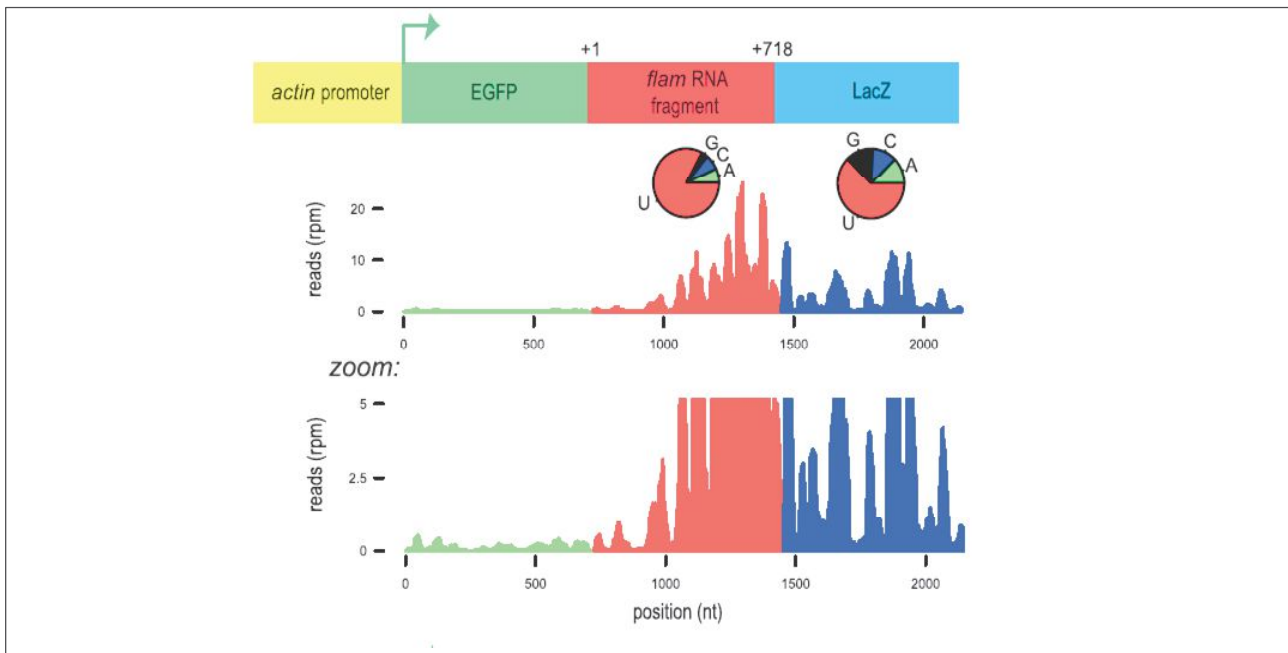
**Figure 37 : Le recrutement artificiel des protéines Piwi et Yb sur un transcrit engendre sa maturation en piRNAs dans les cellules folliculaires des ovaires de drosophile.**

A) Schéma de la construction utilisée pour étudier l'effet du recrutement de composants de la voie piRNA sur l'ARN. B) Alignement des piRNAs séquencés le long du transcrit artificiel lors du recrutement artificiel des protéines GFP, Yb ou Piwi sur ce dernier au sein des cellules folliculaires. Le profil d'alignement en noir représente l'alignement de tous les piRNAs et celui en rouge correspond à l'alignement de piRNAs provenant uniquement du transcrit artificiel. Figure adaptée des figures 1A et 1C de l'article Rogers et al., 2017.

différent pour la reconnaissance et la maturation du précurseur en piRNAs. Afin de mieux comprendre comment les transcrits provenant de piRNAs cluster « unistrand » sont sélectionnés par la cellule pour être maturés en piRNAs, de nombreuses études ont utilisé le cluster de piRNA *flamenco* comme modèle chez la drosophile. En 2012, des lignées transgéniques de drosophile ont été créées en intégrant, au site génomique *61E2*, un BAC contenant une partie importante (plus de 60 Kb) du locus *flamenco*, dans laquelle une séquence rapportrice codant la GFP avait été insérée préalablement (Muerdter et al., 2012). Cette étude avait pour but d'explorer l'impact potentiel de l'environnement génomique d'un cluster de piRNAs « unistrand » sur sa capacité à produire des piRNAs. En d'autres termes, savoir si la reconnaissance des transcrits de *flamenco* était définie par son environnement génomique. Le suivi de la production de piRNAs provenant du *flamenco* ectopique a été rendu possible grâce à la séquence rapportrice GFP. Lors du séquençage des petits ARNs des ovaires des mouches transgéniques, des piRNAs provenant de la séquence rapportrice GFP ont été identifiés (Figure 36). Ces résultats suggèrent que la sélection des transcrits précurseurs de clusters de piRNAs « unistrand » pour être maturés en piRNAs est indépendante de l'environnement génomique de ces derniers.

Récemment, des expériences de liaison artificielle de différentes protéines de la voie des piRNAs sur des transcrits ont été menées *in vivo* par deux équipes indépendantes. Les lignées transgéniques développées dans ces études possèdent i) un transgène codant une protéine impliquée dans la biogenèse des piRNAs fusionnée au N-peptide, mais aussi, ii) un transgène permettant l'expression d'un transcrit artificiel rapporteur possédant des séquences *BoxB* (Figure 37). Le N-peptide présente la caractéristique de pouvoir reconnaître et de se lier aux boucles des boîtes *BoxB*. C'est-à-dire que les auteurs de ces deux articles ont pu explorer, de façon artificielle grâce à ce modèle, la capacité de chacune des protéines de la voie de biogenèse à adresser les transcrits rapporteurs à la voie des piRNAs. L'analyse bio-informatique des petits ARNs a permis d'identifier que le recrutement artificiel des protéines Piwi, Zucchini, Yb, Armitage ou encore Shutdown sur un transcrit était suffisant pour initier la production de piRNAs à partir des transcrits artificiels (Rogers et al., 2017 ; Pandey et al., 2017).

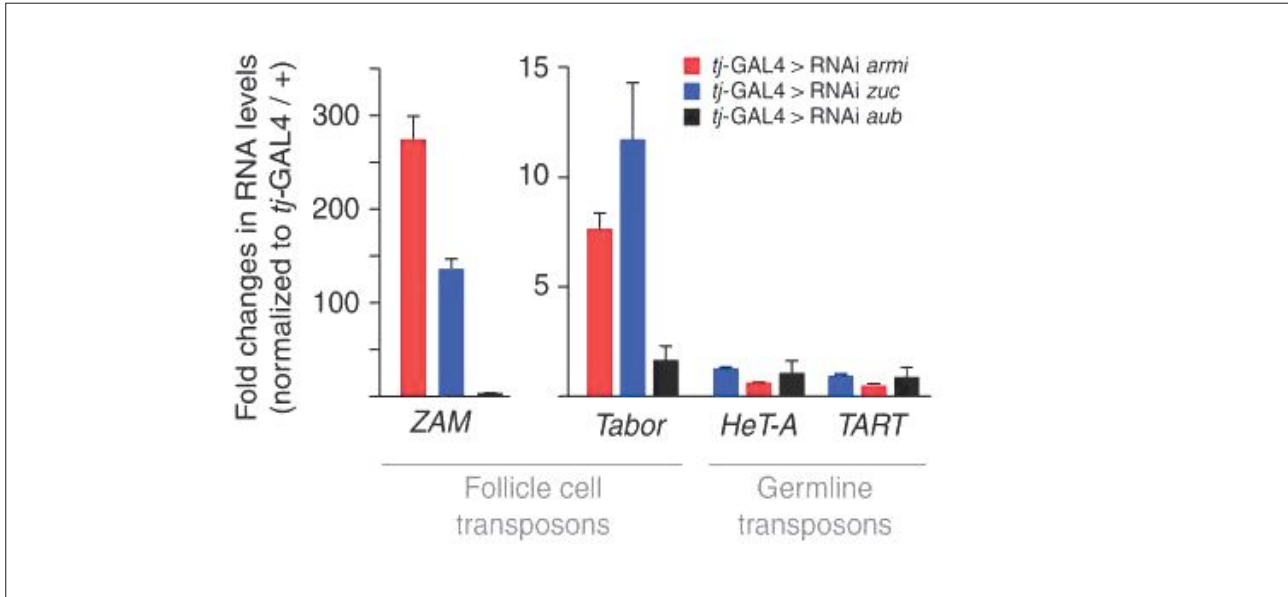
Afin d'analyser les interactions entre certaines protéines de la voie des piRNAs avec les transcrits issus de clusters de piRNAs, des expériences de CLiP (Cross-Linking and Immuno-Precipitation) ont été réalisées sur les cellules OSS (cellules dérivant des cellules souches folliculaires) (Ishizu et al., 2015). Les expériences de Yb-CLIP ont permis de révéler



**Figure 38 : La région +1 +718 du locus *flamenco* permet la maturation de la séquence placée en aval en piRNAs.**

Après des transfections transitoires dans les cellules OSS, les piRNAs liés à la protéine Piwi ont été séquencés puis analysés. L'alignement des piRNAs sur les séquences composant la construction est représenté ci-dessus. Des piRNAs sont produits à partir de la séquence LacZ (en bleu) présente en aval du fragment de *flamenco* +1 +718 (en rouge) mais aucun piRNA provenant de la séquence placée en amont (en vert) n'est observé.

Figure adaptée de la figure 4A de l'article Homolka et al., 2015.



**Figure 39 : Zucchini et Armitage sont impliquées dans la régulation des ETs dans les cellules folliculaires.**

Résultats de qRT PCR mesurant le niveau d'expression des ETs *ZAM*, *Tabor*, *HeT-A* et *TART* dans les ovaires de drosophiles exprimant un ARN interférent ciblant les transgènes *armitage* (*armi*), *zucchini* (*zuc*) ou *aubergine* (*aub*) dans les cellules folliculaires.

Figure adaptée de la figure 3B de l'article Olivieri et al., 2010.

que la protéine Yb était recrutée tout au long des transcrits de *flamenco*, mais aucune région spécifique des transcrits de *flamenco* n'a pu être identifiée comme étant responsable de sa prise en charge par la voie des piRNAs. Cependant, en 2015, des expériences de transfections transitoires réalisées sur les cellules OSS dans notre équipe, et en collaboration avec l'équipe de Ramesh Pillai, ont permis l'identification d'une séquence présente en 5' du locus *flamenco* qui est suffisante pour induire la maturation d'un gène rapporteur en piRNAs. En effet, les 718 premiers nucléotides du transcrit de *flamenco* placés en amont du gène rapporteur de la luciférase permet la maturation du transcrit de la luciférase en piRNAs (Figure 38) (Homolka et al., 2015).

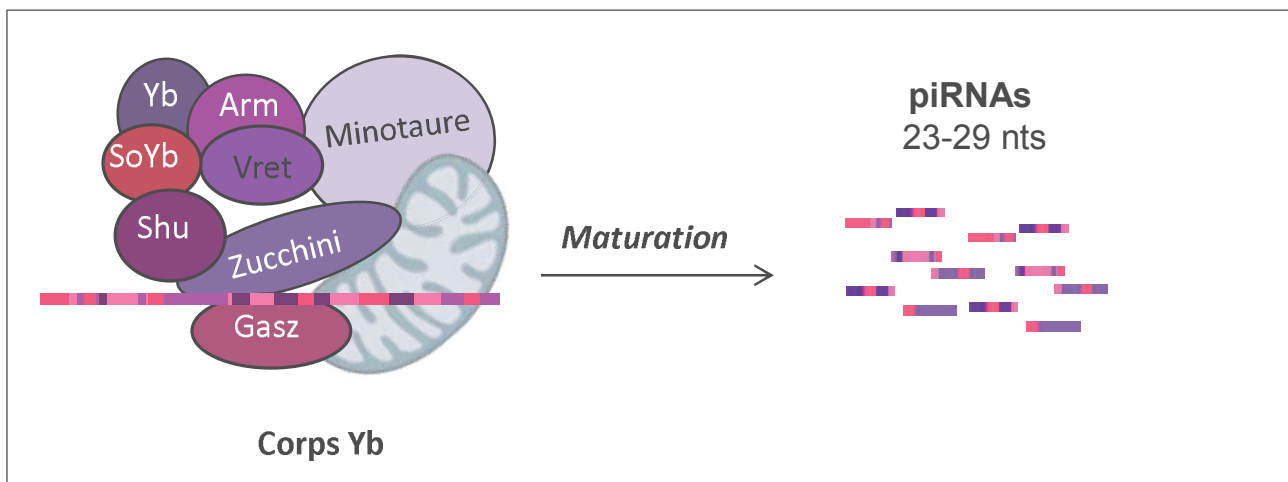
Néanmoins, *in vivo*, la sélection des transcrits de clusters de piRNAs « unistrand » tels que *flamenco* reste encore à élucider à ce jour.

### III. *La maturation des transcrits issus des clusters de piRNAs*

Une fois sélectionnés, les transcrits de clusters de piRNAs sont acheminés à leur site de maturation. La maturation des transcrits provenant de piRNAs cluster « unistrand » (exprimés dans les cellules folliculaires) a lieu au sein du **corps Yb**, une structure cytoplasmique multiprotéique, alors que celle des transcrits provenant de piRNAs cluster "dual-strand" (exprimés dans les cellules germinales) a lieu dans le **nuage**.

#### a) **Maturation au sein du corps Yb**

L'identification des différents composants des corps Yb a été rendue possible *via* divers cribles génétiques réalisés au sein des cellules OSS, mais également *in vivo* par ARN interférence (Figure 39) (Olivieri et al., 2010 ; Saito et al. 2010). Dans certains contextes mutants, des ovaires rudimentaires, voire parfois absents, ont été observés. L'apparition de ces phénotypes est le reflet d'une altération de la production des piRNAs et est corrélée à l'expression (ou la perte de répression) d'un transgène senseur de la répression des ETs. C'est ainsi que les protéines Yb, Armitage (Armi), Zucchini, Vreteno, Shutdown, et Sister of YB ont été identifiées (Saito et al., 2010 ; Handler et al., 2013). Hiérarchiquement, la formation des corps Yb commence avec l'accumulation de la protéine Yb. En effet, en son absence, les autres protéines composant ce corps multiprotéique ne s'accumulent plus en foyer et leur localisation semble diffuse dans le cytoplasme. En absence de la protéine Armi, seule la localisation de la protéine Yb n'est pas affectée suggérant que le recrutement de la protéine Armi intervient après l'accumulation de la protéine Yb. La localisation de la protéine Vreteno,



**Figure 40 : Représentation schématique du corps Yb.**

Le corps Yb est une structure multiprotéique responsable de la maturation des transcrits de clusters de piRNAs folliculaires en piRNAs. Le corps Yb est notamment composé des protéines Yb, Armitage (Arm), Sister of Yb (So YB), Vreteno (Vret), Shutdown (Shu), Zucchini, Minotaure et de GasZ.

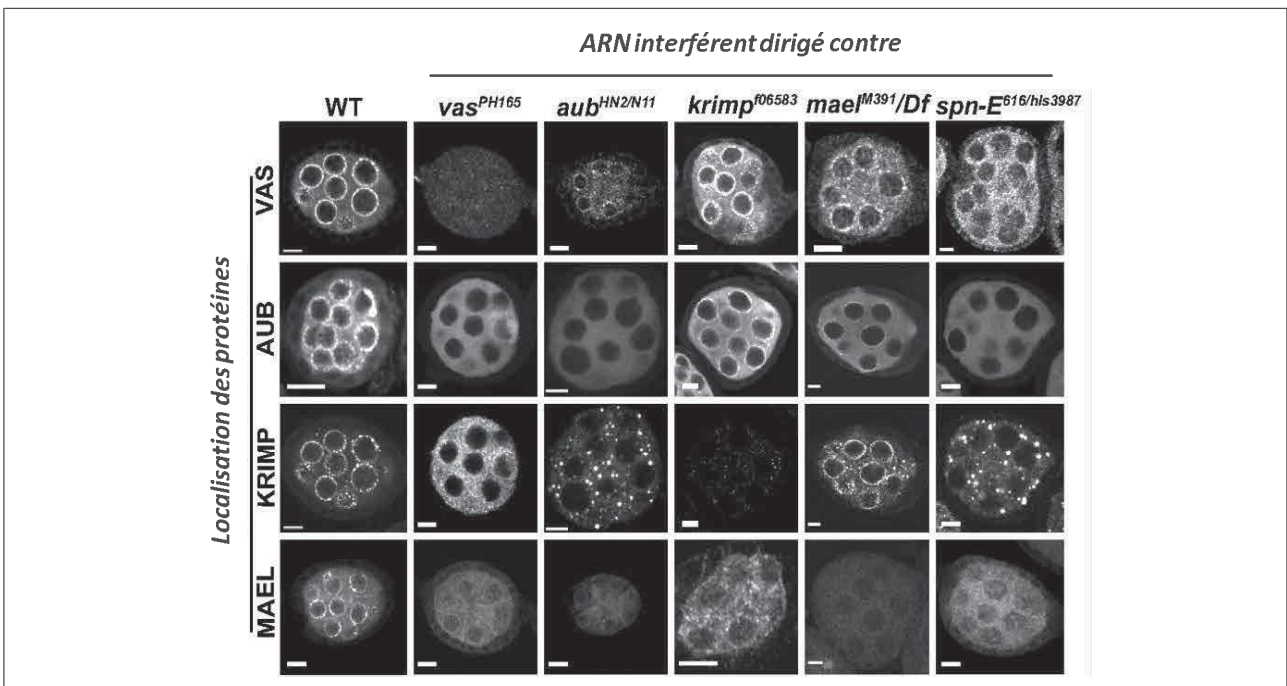
quant à elle, requiert les protéines Yb et Armitage. Les protéines Yb, Armitage, Vreteno sont quant à elles requises pour que la protéine Shutdown soit localisée au niveau des corps Yb (Olivieri et al., 2012 ; Preall et al., 2012). Chacune de ces protéines est composée d'un domaine protéique relié au métabolisme des ARNs. La protéine “female sterile (1) Yb” ou Yb, ainsi que la protéine Armitage sont des hélicases. Les protéines Vreteno et Shutdown ont quant à elles des domaines TUDORs (Handler et al., 2011 ; Zamparini et al., 2011). La protéine Zucchini est une endonucléase localisée à la membrane mitochondriale responsable de la génération des extrémités 5' des piRNAs (Figure 40). La protéine Zucchini possède deux cofacteurs également localisés au niveau de la membrane extérieure de la mitochondrie, les protéines GasZ et Minotaure (Vagin et al., 2013). Le rôle de la protéine GasZ serait de recruter les protéines Armitage et Piwi à la membrane mitochondriale, les rapprochant ainsi de la protéine Zucchini (Handler et al., 2013). L'altération de l'expression des protéines Yb, Armitage et Zucchini a notamment pour conséquence de perturber la localisation de la protéine Piwi qui devient alors cytoplasmique, et de diminuer la production de piRNAs. L'ensemble de ces résultats est corrélé à une dérégulation d'ETs suggérant que ces protéines jouent un rôle critique dans la biogenèse des piRNAs. Chez la souris, les protéines homologues aux protéines Vasa, (Mvh), Armitage (MOV10L1), Vreteno, Shutdown et Zucchini (MITOPLD) sont également des facteurs cruciaux lors de la biogenèse des piRNAs primaires puisqu'en absence d'une de ces protéines, une stérilité chez les mâles est observée (Zheng et al., 2010 ; Huang et al., 2011). La dernière étape de la biogenèse des piRNAs est le raccourcissement de l'extrémité 3' des piRNAs suivi d'une 2'-O-méthylation sur cette même extrémité par la méthyltransférase Hen1. En effet, les piRNAs produits par Zucchini présentent une taille intermédiaire allant de 25 à 50 nucléotides. Il a récemment été montré que l'exonucléase Nibbler participerait à la formation des extrémités 3' des piRNAs (Hayashi et al., 2016). Une fois produits, les piRNAs primaires sont chargés sur la protéine Piwi en requérant l'aide de la protéine chaperonne HSP90 (Izumi et al., 2013).

#### **b) Maturation au sein du nuage**

Dans les cellules germinales des ovaires de drosophile, la maturation des transcrits de piRNAs clusters se produit dans le nuage, une structure électroniquement dense qui entoure le noyau des cellules germinales.

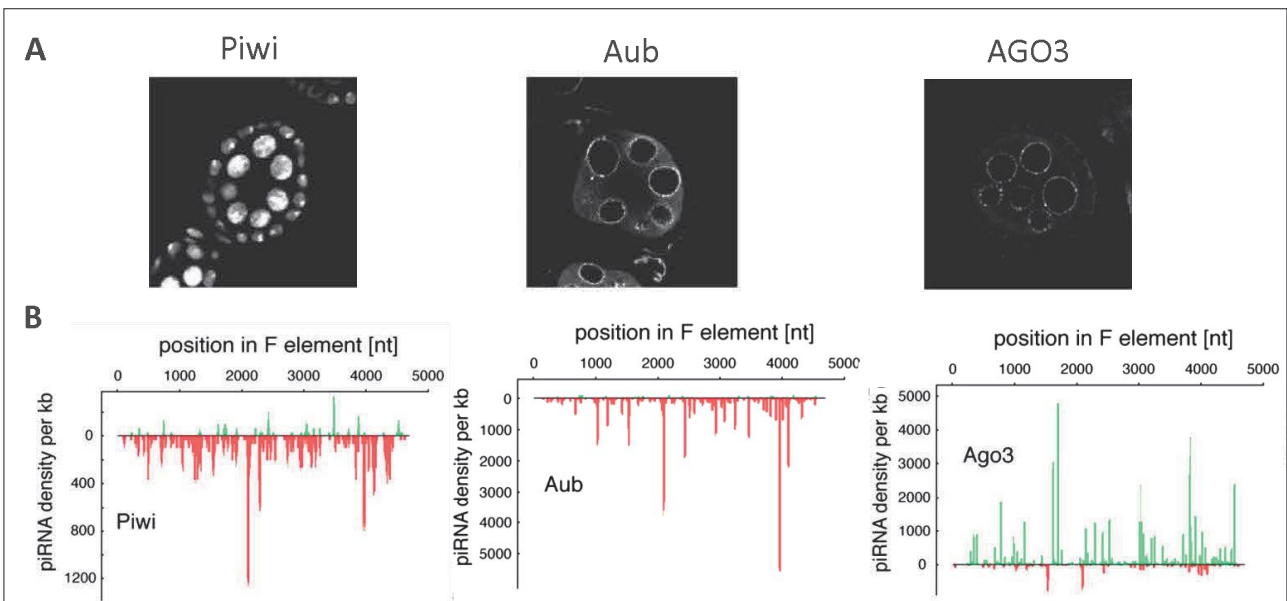
L'hélicase Vasa prend en charge les transcrits provenant des clusters de piRNAs germinaux dès leur entrée dans le cytoplasme pour les acheminer au nuage (Zhang et al.,





**Figure 41 : Assemblage hiérarchique des protéines du nuage.**

L'immunomarquage des protéines Vasa, Aubergine, Krimper et Maelstrom lors de l'expression germinale d'ARNs interférents ciblant les transcrits *vasa*, *aubergine*, *krimper*, *maelstrom* et *spindle-E* a permis d'établir l'ordre de recrutement de ces protéines au nuage. La localisation des composants du nuage dans les cellules germinales dépend de la présence de la protéine Spindle-E (*spn-E*) localisée à la périphérie nucléaire. Les autres protéines constituant le nuage sont alors successivement recrutées selon l'ordre suivant : Vasa>Aubergine>Krimper>Maelstrom. Figure adaptée de la figure A4 de l'article Lim et Kai, 2007.



**Figure 42 : Caractérisation des protéines PIWI au sein des ovaires de drosophiles.**

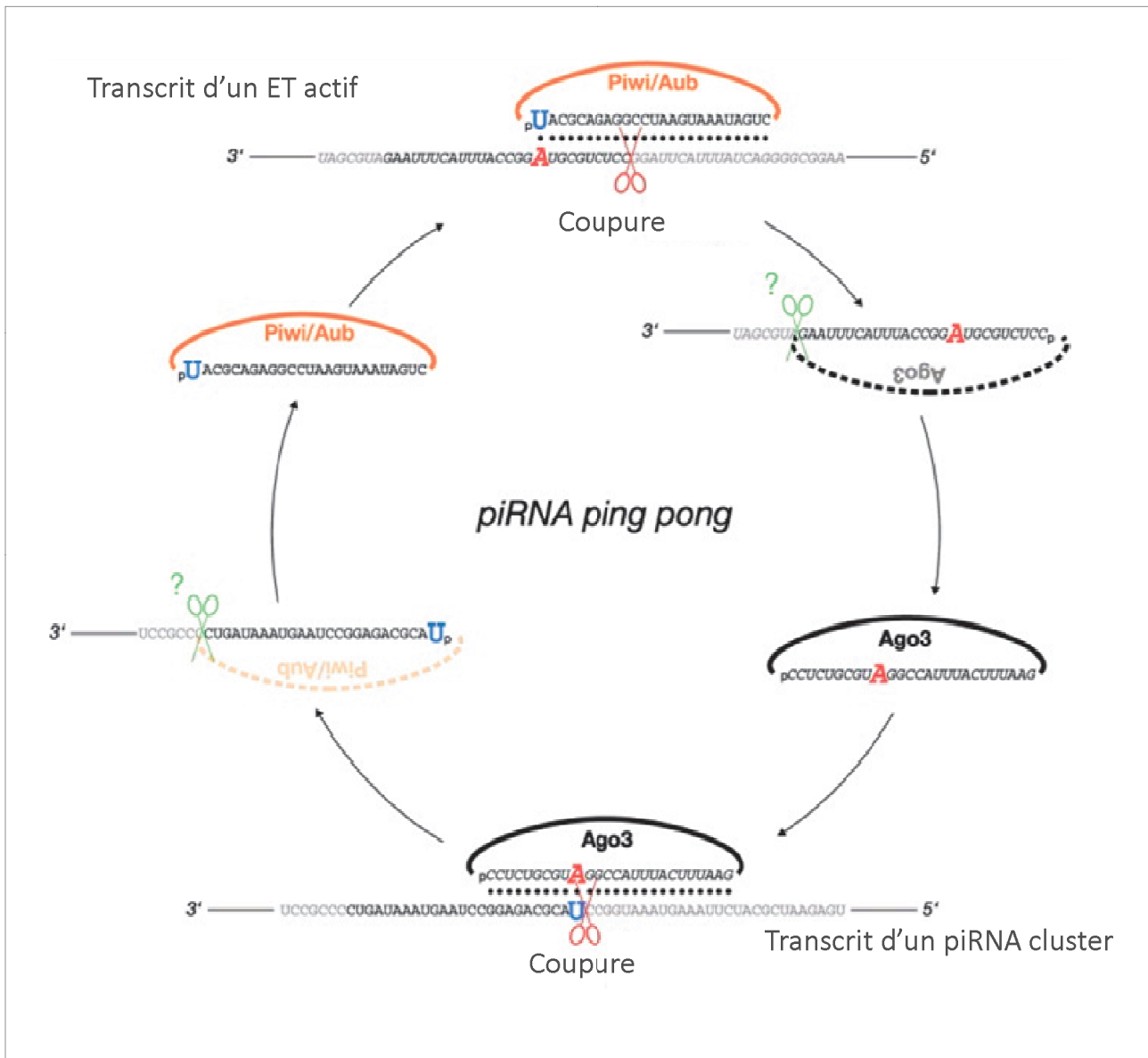
A) Localisation des protéines de la sous famille PIWI dans les ovaires de drosophile. Les protéines Aub et AGO3 s'accumulent à la périphérie nucléaire, dans le nuage dans les cellules germinales alors que la protéine Piwi est localisée dans les noyaux des cellules folliculaires ainsi que dans les noyaux des cellules germinales. Figure adaptée de la figure 2E de Malone et al., 2009. B) Les protéines Piwi et Aub prennent majoritairement en charge les piRNAs anti-sens par rapport au sens de transcription de l'élément transposable F (en rouge). A l'inverse, les piRNAs sens par rapport au sens de transcription de l'ET F (vert) sont préférentiellement pris en charge par la protéine AGO3. Figure adaptée de la figure 5B de Brennecke et al., 2007.

2012). Le recrutement de Vasa au nuage nécessite préalablement l'accumulation de la protéine Spindle-E (Lim et Kai, 2007). Les autres composants du nuage s'accumulent alors au niveau du nuage selon un ordre hiérarchique (Figure 41) (Lim et Kai, 2007 ; Harris et Macdonald, 2001 ; Findley et al., 2003). La plupart des protéines composant le corps Yb, à l'exception de la protéine Yb elle-même, sont retrouvées accumulées dans le **nuage** (Malone et al., 2009). En effet, les protéines Zucchini, Armitage, Sister of Yb, Vreteno, Shutdown, Gasz et Hen1 sont aussi des composants du nuage. La maturation des transcrits précurseurs de piRNAs au sein du nuage partage les mêmes étapes que celle des transcrits précurseurs de piRNAs au sein du corps Yb à savoir la génération de l'extrémité 5' des piRNAs par la protéine Zucchini et le raccourcissement des extrémités 3' des piRNAs par une exonucléase. Une fois produits, les piRNAs primaires sont chargés sur la protéine Piwi ou sur la protéine Aubergine.

### *C. L'amplification du pool de piRNAs dans les cellules germinales*

Ce mécanisme, spécifique aux cellules germinales, permet l'amplification cytoplasmique du pool de piRNAs primaires. Les protéines impliquées dans ce processus sont deux protéines de la sous-famille PIWI, la protéine Aubergine et la protéine AGO3. Ces deux protéines sont exprimées uniquement dans les cellules germinales (Figure 42A). Elles ont en commun d'être localisées au niveau du nuage, de pouvoir se lier aux ARNs, mais également, de pouvoir les couper *via* leur activité endonucléasique. Les séquençages des piRNAs liés aux différentes protéines PIWI ont révélé que la majorité des piRNAs chargés sur les protéines Piwi et Aubergine commençaient par une uridine, alors que ceux liés à la protéine AGO3 présentaient préférentiellement une adénine en position 10. De plus, il est important de noter que la protéine Aubergine lie majoritairement des piRNAs anti-sens (par rapport à la séquence d'ETs) et que la protéine AGO3 se lie, quant à elle, préférentiellement à des piRNAs sens (par rapport à la séquence d'ETs) (Figure 42B).

Au niveau du nuage, les piRNAs primaires peuvent être pris en charge, soit par la protéine Piwi et exercer une répression transcriptionnelle, soit par la protéine Aubergine afin de réaliser une régulation post-transcriptionnelle des ETs. Le complexe piRNA primaire/Aubergine reconnaît, par complémentarité de séquence, l'ARNm cytoplasmique d'un ET actif (sens) (Figure 43). Une fois le complexe fixé au transcrit cible, la protéine Aubergine génère, *via* son activité d'endonucléase, une coupure entre le dixième nucléotide et le onzième nucléotide de la région ciblée par le piRNA primaire (Gunawardane et al., 2007). Le premier nucléotide (U) du piRNA primaire (anti-sens) correspond au dixième nucléotide (A) du



**Figure 43 : Représentation schématique du cycle ping-pong.**

Le complexe piRNA primaire/Aub reconnaît, par complémentarité de séquence, le transcrit d'un ET actif. La protéine Aub, *via* son activité endonucléasique, génère une coupure entre le dixième nucléotide et le onzième nucléotide de la région ciblée par le piRNA primaire (anti-sens). Le piRNA secondaire (sens) contenant une adénosine (A) en 10<sup>ème</sup> position est alors pris en charge par la protéine AGO3. Le complexe AGO3/piRNA secondaire va à son tour reconnaître, par complémentarité de base, le transcrit provenant d'un cluster de piRNAs. Grâce à son activité endonucléasique, AGO3 va couper également entre le dixième nucléotide et le onzième nucléotide de la région ciblée par le piRNA secondaire, générant ainsi l'extrémité 5' du piRNA pris en charge par la protéine Aub.

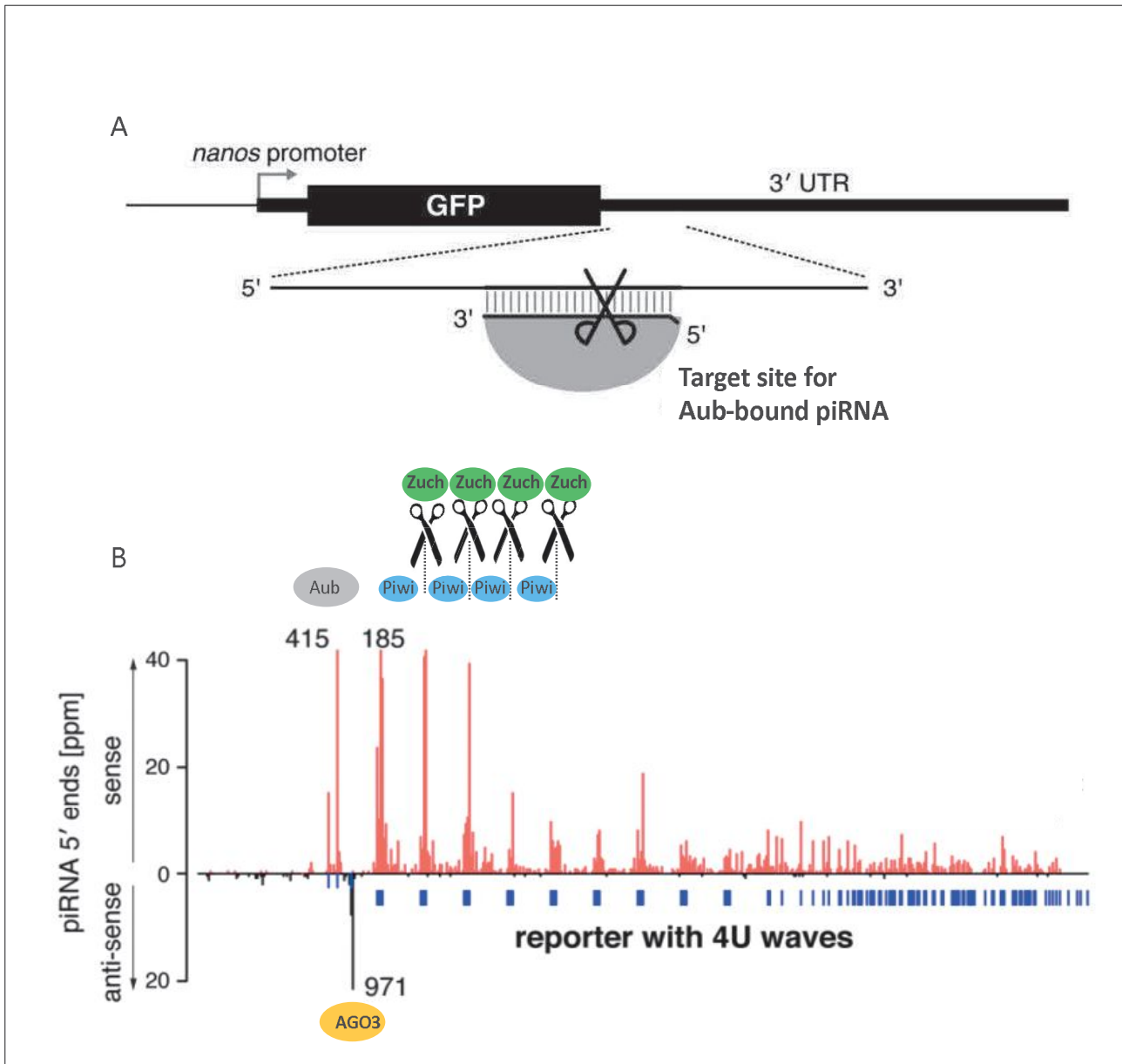
Figure adaptée de la figure 7 de Brennecke et al 2007.

piRNA secondaire (sens). Le piRNA secondaire sens est ainsi pris en charge par la troisième protéine de la famille PIWI, AGO3. En d'autres termes, les piRNAs primaires chargés sur la protéine Aubergine et les piRNAs secondaires chargés sur la protéine AGO3 partagent une homologie de séquence de 10 nucléotides. Le piRNA chargé sur la protéine AGO3 va également subir un raccourcissement de son extrémité 3' ainsi qu'une 2'O-méthylation par la méthyltransférase Hen1. Le complexe AGO3/piRNA va à son tour couper le transcrite ciblé par le piRNA, un transcrite provenant d'un cluster de piRNAs. Grâce à son activité endonucléasique, AGO3 va couper également entre le dixième nucléotide et le onzième nucléotide de la région ciblée par le piRNA secondaire générant ainsi l'extrémité 5' du piRNA pris en charge à son tour par la protéine Aubergine. Ce mécanisme appelé plus communément la boucle d'amplification ping-pong permet d'augmenter le réservoir de piRNAs produits dans la lignée germinale pour une répression efficace des ETs dans la lignée germinale de l'individu et dans la descendance. En effet, ces piRNAs germinaux sont transmis maternellement à l'embryon précoce et sont indispensables à la réactivation des clusters de piRNAs dans la génération suivante. Chez la drosophile, le cycle ping-pong est préférentiellement réalisé *via* une interaction hétérotopique entre Ago3 et Aub.

Chez la souris, la production des piRNAs secondaires peut être effectuée par les protéines MILI et MIWI2, les protéines homologues à Aubergine et AGO3. Cependant, la coupure des transcrits est réalisée grâce à l'endonucléase MILI. Contrairement à la boucle ping-pong chez la drosophile, le cycle ping-pong murin utilise majoritairement une interaction homotypique MILI/MILI (De Fazio et al., 2011).

Le ping-pong permet également la production des piRNAs dits "tertiaires" ou "trail piRNAs" (Han et al., 2015 ; Mohn et al., 2015). Ils sont générés à partir d'un transcrite préalablement coupé en 5' par Aub ou AGO 3 lors du cycle ping-pong. En effet, la coupure endonucléasique du transcrite ciblé par les protéines Aub ou AGO3 induit la maturation du reste du transcrite de manière phasique par les protéines Piwi et les protéines Zucchini. Pour cela, les protéines Piwi se fixent les unes derrière les autres sur le transcrite clivé (Figure 44). La protéine Zucchini coupe de part et d'autre des protéines Piwi produisant ainsi, de façon cyclique, des piRNAs 5'U. Cette nouvelle source de production de piRNAs est appelée le "Phasing" (Han et al., 2015 ; Mohn et al., 2015).

#### ***D. La production de piRNAs en dehors des gonades***



**Figure 44 : Mécanisme de Phasing**

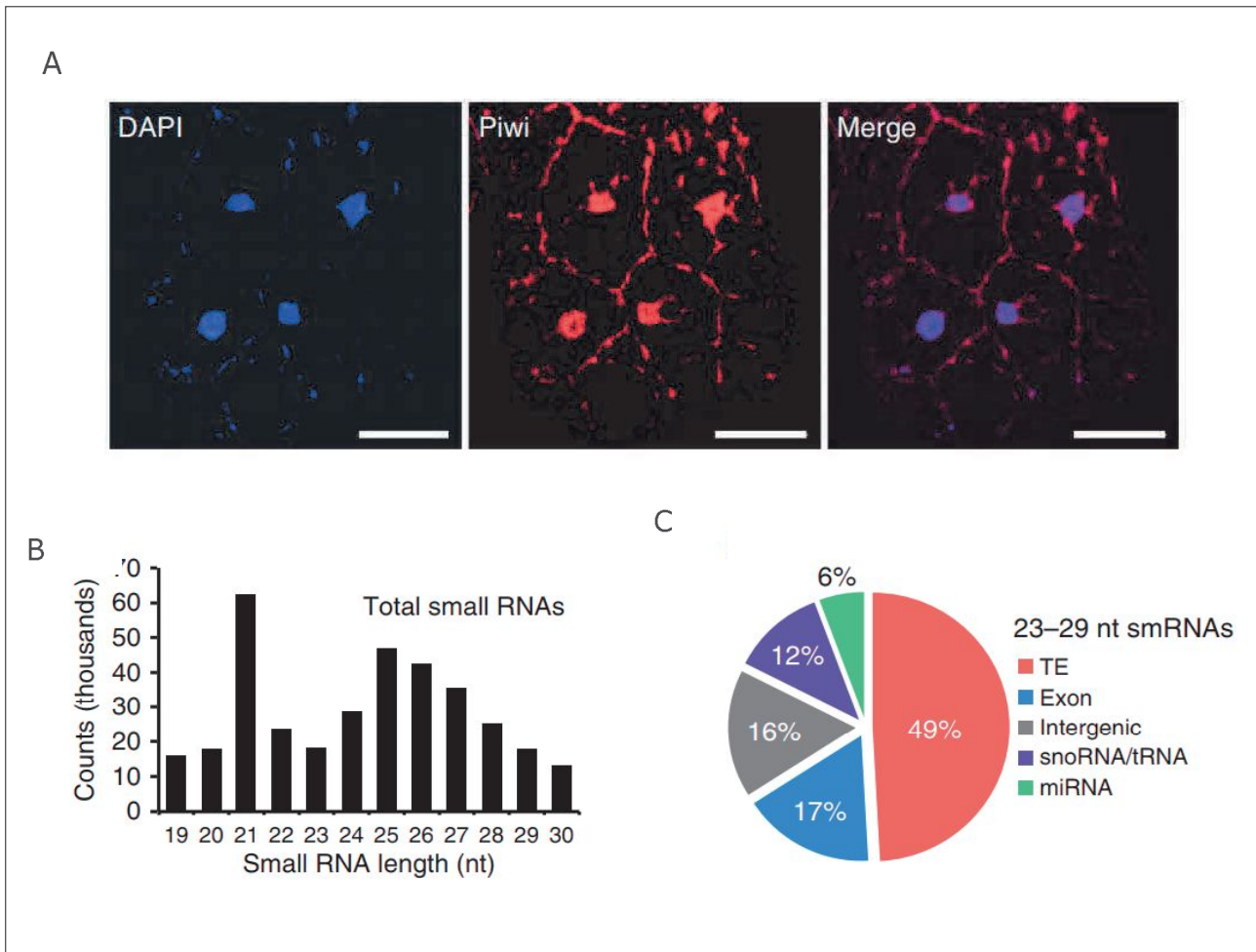
- A) Transgène senseur de la biogenèse des piRNAs possédant un site ciblé par les complexes piRNA/Aub.
- B) Profil des extrémités 5' des piRNAs s'alignant sur le transcrit rapporteur (Rouge : piRNAs provenant du transcrit rapporteur, noir : piRNAs anti-sens au transcrit rapporteur). Le rapporteur possède un site ciblé par les complexes piRNAs/Aub (en gris). Ce site est placé en amont d'une séquence composée de 4 uridines tous les 26 nucléotides (en bleu). Suite au cycle ping-pong réalisé par les protéines Aub (en gris) et AGO3 (en orange), le reste du transcrit est maturé de façon phasique par la protéine Zucchini (en vert avec des ciseaux) et la protéine Piwi (en bleu clair).

Figure adaptée des figures 3A et 3C Mohn et al., 2015.

La voie des piRNAs a initialement été découverte et décrite comme permettant de réguler spécifiquement la régulation des ETs au sein des tissus gonadiques des animaux. De plus, il a été montré que les acteurs clés de la voie des piRNAs, notamment les protéines PIWI et les piRNAs, sont essentielles pour un déroulement optimal de la spermatogenèse et l'ovogenèse chez de nombreuses espèces animales. Cependant, les protéines PIWI et les piRNAs ont également été découverts au sein de tissus non gonadiques chez certaines espèces animales suggérant alors que la voie des piRNAs pourrait être de la même façon active dans d'autres tissus pour réguler les ETs et les gènes.

Les protéines PIWI ont été retrouvées chez de nombreux organismes allant des *Porifera* (éponges) à l'Homme. Leur expression a été rapportée initialement au sein des cellules gonadiques. Cependant, plusieurs études ont démontré que les protéines PIWI étaient également exprimées dans divers types cellulaires allant de cellules souches pluripotentes à des cellules différenciées. À titre d'exemple, chez l'Homme, les protéines PIWI ont été retrouvées exprimées dans les cellules germinales, mais aussi dans les cellules souches hématopoïétiques et dans un certain nombre de cancers (Sharma et al., 2001 ; Lee et al., 2011). Chez *Planaria*, les protéines PIWI sont également exprimées au sein des cellules souches totipotentes appelées les néoblastes (Reddien et al., 2005). Chez la drosophile, les protéines PIWI ont été retrouvées dans les gonades, le cerveau, les glandes salivaires ainsi que dans le corps gras (Perrat et al., 2013 ; Brower-Toland et al., 2007 ; Jones et al., 2016). Par ailleurs, la mutation des protéines Aub et AGO3, décrites comme des protéines germinales, conduit à une dépression des ETs dans le cerveau suggérant que la répression des ETs au sein des cellules non gonadiques peut également être conduite par la voie des piRNAs (Perrat et al., 2013). De plus, l'absence de la protéine Piwi au niveau du corps gras chez la drosophile induit une dépression d'ETs tels que l'élément *gypsy*, une réduction du stockage du glycogène et des triglycérides, ainsi qu'une durée de vie plus courte (Jones et al., 2016). En d'autres termes, l'ensemble de ces données suggère que la voie des piRNAs semble être également fonctionnelle dans certains tissus non gonadiques, et par conséquent, que des piRNAs pourraient être produits en dehors des gonades chez ces animaux.

Chez la drosophile, il a été rapporté que des piRNAs seraient produits au niveau du corps gras (Figure 45) mais également au niveau de leurs têtes (Jones et al., 2016 ; Yan et al., 2011 ; Mirkovic-Hösle et Förtemann, 2014). Néanmoins, une étude récente a pu démontrer que les piRNAs identifiés dans les bibliothèques des têtes de drosophiles provenaient des petits



**Figure 45 : Présence de la protéine Piwi et de piRNAs dans le corps gras de drosophile.**

- A) La protéine Piwi (en rouge) est localisée dans les noyaux de cellules du corps gras. La membrane de ces cellules est auto-fluorescente responsable d'un marquage aspécifique (en rouge). Les noyaux sont marqués au DAPI en bleu. Les barres d'échelle représentent 20  $\mu$ m.
- B) Distribution de tailles des petits ARNs retrouvés dans les cellules du corps gras de la drosophile.
- C) Origine génomique des petits ARNs de 23 à 29 nts.

Figure adaptée des figures 1 B, 1D et 1E de l'article Jones et al., 2016.

ARNs issus de gonades suite à une contamination lors des extractions des ARNs (Van Den Beek et al., 2018).

En 2011, Lee et ses collègues ont mis en évidence la présence de piRNAs au niveau de l'hippocampe des souris. Leur présence a également été détectée dans d'autres tissus chez la souris tels que le cerveau, le cœur, le foie, les poumons, la rate, les reins et les testicules (Lee et al., 2011). Cependant, leur production semble être plus importante dans l'hippocampe, dans le cœur et dans le foie. Bien qu'ils aient une taille de 24 à 32 nts et qu'ils soient liés à la protéine MIWI, la plupart de ces piRNAs (soit 40,3%) ne commencent pas par une uridine, signature de la production des piRNAs primaires. C'est-à-dire que les piRNAs détectés dans ces tissus ne possèdent pas l'ensemble des caractéristiques propres des piRNAs.

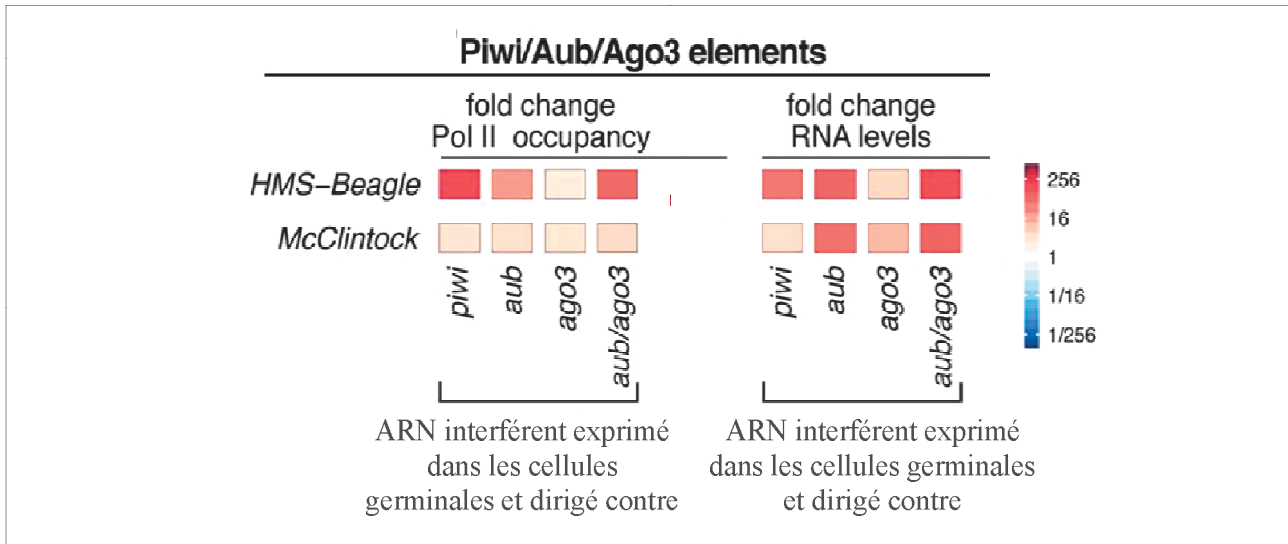
Chez l'Homme, des petits ARNs de 23-29 nts ont été identifiés comme des piRNAs dans des échantillons de plasma, de lait maternel, d'urine, de salives et de tumeurs cancéreuses (Freedman et al., 2016 ; Rubio et al., 2018). Néanmoins, une seconde analyse de ces résultats a pu démontrer qu'une seule fraction des petits ARNs identifiés correspondrait aux piRNAs (Tosar, Rovira et Cayota, 2018).

Pour lever les différentes ambiguïtés sur la production des piRNAs en dehors des gonades, il faut s'assurer que les petits ARNs de 23 à 32 nts partagent plusieurs caractéristiques attribuées aux piRNAs telles qu'un biais d'uridine à leur extrémité 5', un groupement 2'O-méthyle au niveau de leur extrémité 3' ainsi qu'avoir la capacité de se lier aux protéines PIWI.

## **Chapitre 4 : Les différentes fonctions biologiques des piRNAs**

Initialement, la fonction biologique des piRNAs était restreinte à la régulation des ETs, au niveau du tissu reproducteur, dans le but de protéger l'information génétique transmise à la descendance. Sachant que les piRNAs dérivants de séquence d'ETs ont la capacité de réguler finement l'activité de ces derniers, les piRNAs dérivants de régions génomiques dépourvues d'ETs pourraient alors avoir la capacité de réguler l'expression de gènes. Depuis quelques années, de nombreuses études ont été menées afin d'explorer les divers rôles biologiques des piRNAs. Les différentes fonctions attribuées aux piRNAs ont été retrouvées tant au niveau des tissus gonadiques qu'au niveau de tissus somatiques non gonadiques.





**Figure 46 : Mise en évidence du type de régulation permettant la mise sous silence d'un ET donné.**

La comparaison du recrutement de la polymérase II (Pol II) ainsi que la quantité de transcrits d'un ET entre une condition sauvage et des conditions mutantes pour les protéines Piwi, Aub et AGO3 sont présentées. L'ET *HMS-Beagle* serait préférentiellement réprimé via une régulation transcriptionnelle puisqu'une augmentation du recrutement de la Pol II sur cet ET est observée en absence de la protéine Piwi. L'ET *McClintock* est, quant à lui, préférentiellement régulé post transcriptionnellement. Aucun enrichissement de l'ARN polymérase II n'est observé sur cet ET en condition mutante pour l'une de ces deux protéines. Néanmoins, la quantité des transcrits de cet ET est, quant à elle, augmentée en condition mutante pour l'une de ces deux protéines. Figure adaptée de la figure 5C de Senti et al., 2015.

## *A. La régulation des ETs*

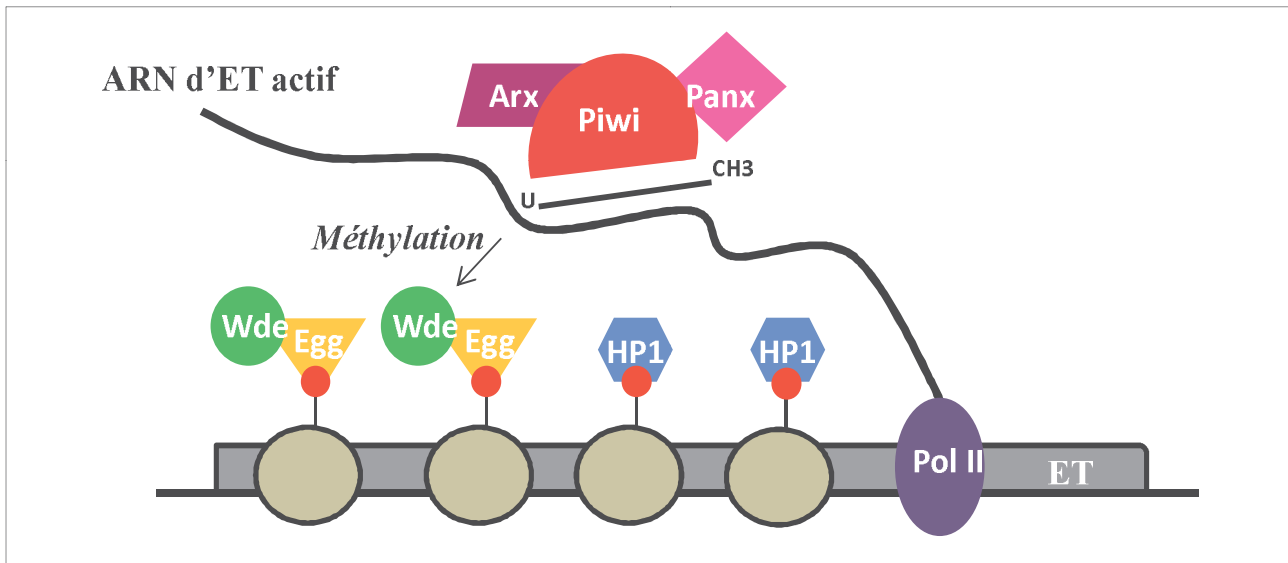
Chez les animaux, la voie des piRNAs permet de verrouiller l'activité des ETs soit en induisant une inhibition transcriptionnelle de l'ET ciblé, soit en coupant l'ARN messager de ce dernier. Néanmoins, chaque ET peut être préférentiellement contrôlé par l'une ou l'autre de ces régulations. En effet, Senti et ses collaborateurs ont classé en 2015 les ETs en fonction de leur régulation putative. Pour cela, ils ont comparé le recrutement de l'ARN polymérase II ainsi que la quantité des transcrits d'ETs entre une condition sauvage et une condition mutante soit de la protéine aubergine, soit de la protéine AGO3. Si en condition mutante pour Aubergine ou AGO3, l'ARN polymérase II est enrichie au niveau du site de transcription de l'ET étudié par rapport à un contexte sauvage, alors la régulation de ce dernier est transcriptionnelle (Figure 46). Si, au contraire, aucun enrichissement de l'ARN polymérase II n'est observé en condition mutante par rapport à la condition sauvage, mais que la quantité d'ARN messagers est, quant à elle, augmentée, alors ces ETs sont préférentiellement régulés post-transcriptionnellement.

### *I. Régulation transcriptionnelle*

La régulation transcriptionnelle des ETs chez la drosophile et chez la souris est initiée respectivement par les protéines nucléaires Piwi et MIWI2. Ce mode de régulation permet, de façon indirecte, de réduire la quantité de transcrits d'un élément donné en modifiant notamment la configuration chromatinienne au niveau du promoteur des ETs.

La répression transcriptionnelle englobe le remplacement des marques histones permissives H3K4me2 par des marques histones répressives H3K9me3, mais comprend aussi la méthylation de l'ADN de dinucléotides CpG au sein des cellules germinales de mammifères (Aravin et al., 2008 ; Kuramochi-Miyagawa et al., 2008). La régulation transcriptionnelle peut être découpée en trois grandes étapes.

La première étape de la régulation transcriptionnelle des ETs chez la drosophile consiste à reconnaître le transcrit naissant d'un ET actif (Figure 47). Cette reconnaissance est basée sur la complémentarité de séquence entre le piRNA chargé sur la seule protéine PIWI nucléaire, la protéine Piwi, et le transcrit naissant de l'ET à réprimer. La mutation du signal de localisation nucléaire de cette protéine Piwi engendre la disparition des marques hétérochromatiques nécessaires à la répression transcriptionnelle des ETs (Klenov et al.,



**Figure 47 : Représentation schématique de la répression transcriptionnelle d'un ET par la voie des piRNAs.**

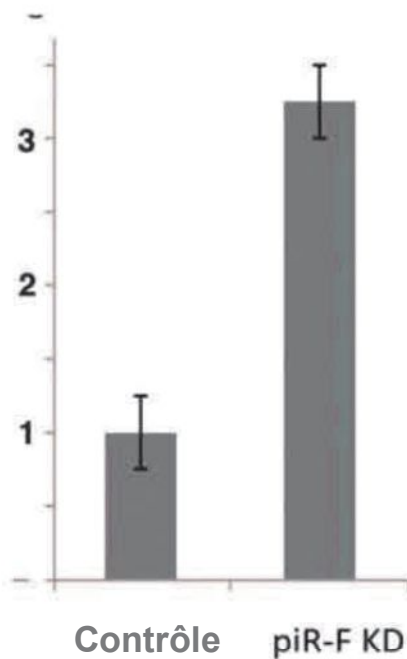
La première étape de la régulation transcriptionnelle des ETs chez la drosophile est la reconnaissance des transcrits d'ETs actifs via le complexe piRNA/Piwi aidé des protéines Astérix (Arx) et Panoramix (Panx). La seconde étape est le recrutement de la méthyltransférase Eggless et de son cofacteur Windel afin de déposer des marques histones H3K9me3 le long de la séquence de l'ET à réprimer. La protéine HP1 reconnaissant les marques histones H3K9me3 induit par la suite la mise en place d'une structure hétérochromatique verrouillant ainsi la mise sous silence de l'ET.

2014 ; Klenov et al., 2011 ; Sienski et al., 2012). En revanche, l'activité endonucléasique de Piwi n'est pas requise lors de la répression transcriptionnelle (Saito et al., 2009 ; Sienski et al., 2012). Les cofacteurs Astérix et Panoramix assisteront le complexe piRNA/Piwi lors de son appariement avec le transcrit naissant de l'ET à réprimer. La protéine Astérix stabilise l'interaction entre le complexe piRNA/Piwi alors que la protéine Panoramix induit, quant à elle, la répression transcriptionnelle de l'élément ciblé par le complexe piRNA/Piwi. L'ordre de recrutement des différentes protéines impliquées dans la mise en place d'une structure hétérochromatique au niveau du promoteur de l'ET ciblé reste, à ce jour, encore inconnu. Néanmoins, la méthyltransférase Eggless, aidée de son cofacteur Windei, dépose des marques histones H3K9me3 au niveau du promoteur de l'ET à réprimer (Handler et al., 2013). Le promoteur de l'ET ciblé est ainsi rendu inaccessible à la machinerie transcriptionnelle. Une fois mise en place, la répression de l'expression des ETs est maintenue notamment *via* la protéine HP1 qui se lie aux marques histones H3K9me3 (Figure 47). Toutefois, le dépôt des marques histones H3K9me3 au niveau des transposons est faiblement affectée en absence d'une autre protéine également impliquée dans la répression transcriptionnelle des ETs, à savoir la protéine Maelstrom. Cependant, au sein des mutants Maelstrom, le nombre de transcrits naissant d'ETs actifs est augmenté. La dérégulation observée chez les mutants *Maelstrom* est comparable à celle obtenue chez les mutants *Piwi* (Sienski et al., 2012). En d'autres termes, la protéine Maelstrom est une protéine importante pour la répression transcriptionnelle des ETs, mais dont le mode d'action reste encore inconnu à ce jour.

## II. Régulation post-transcriptionnelle

La régulation post-transcriptionnelle des ETs est, quant à elle, initiée par les protéines PIWI impliquées dans le cycle ping-pong, à savoir les protéines Aubergine et AGO3 chez la drosophile, les protéines SIWI et BmAgo3 chez *Bombyx mori* et les protéines MILI et MIWI chez la souris (Brennecke et al., 2007 ; Kawaoka et al., 2009 ; Aravin et al., 2007 ; Reuter et al., 2011 ; Di Giacomo et al., 2013 ; Gunawardane et al., 2007 ; De Fazio et al., 2011). Comme décrit précédemment, le cycle ping-pong consiste à amplifier le pool de piRNAs au sein des cellules germinales, mais permet aussi de diminuer directement la quantité de transcrits d'ETs actifs. Les transcrits coupés par la protéine Aub sont par la suite soit dégradés, soit maturés en piRNAs par le cycle ping-pong et/ou par le processus de Phasing.

### Change in CREB2 mRNA levels



**Figure 48 : Régulation transcriptionnelle du gène CREB2 par le piRNA piR-F chez *Aplysia*.**

Mesure du niveau d'expression du gène CREB2 dans les ganglions pleuraux contenant des neurones sensoriels. Les ganglions ont préalablement été incubés huit heures soit dans une solution contrôle (contrôle), soit dans une solution contenant des 2'-O-méthyl oligoribonucléotides ciblant spécifiquement les piRNAs piR-F (piR-F KD). La diminution de la quantité de piRNAs piR-F (piR-F KD) augmente significativement le niveau d'expression du gène CREB2.

Figure adaptée de la figure 6C Rajasethupathy et al., 2012.

## ***B. La régulation des gènes***

Tout comme les ETs, l'expression des gènes peut également être contrôlée transcriptionnellement et post-transcriptionnellement par la voie des piRNAs. Les piRNAs mis en jeu dans ces régulations sont le plus souvent des piRNAs ciblant des séquences d'ETs présentes dans les régions 3'UTR de certains transcrits de gènes, mais également des piRNAs dont les séquences sont dépourvues d'ETs.

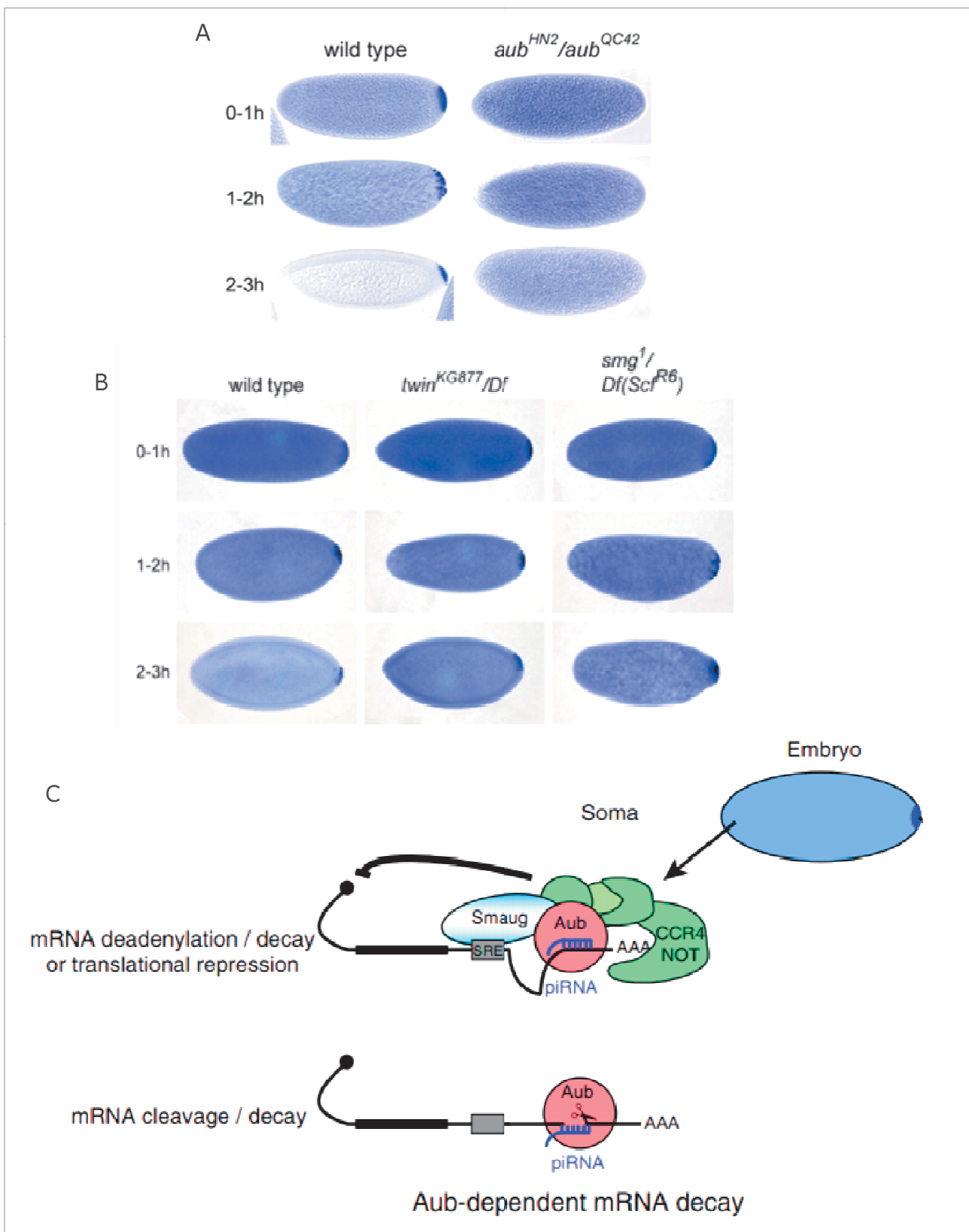
### ***I. Régulation transcriptionnelle***

Dans un certain nombre d'études, les complexes piRNAs/PIWI ont été décrits comme permettant la mise sous silence de gènes en induisant la mise en place d'une structure hétérochromatique au niveau des promoteurs de ces derniers, pouvant aller jusqu'à la méthylation ADN des régions promotrices chez certains animaux. En d'autres termes, la voie des piRNAs permettrait de réguler transcriptionnellement l'expression de certains gènes notamment au cours du développement d'un organisme.

En 2012, une étude portant sur la formation de la mémoire chez la limace de mer (*Aplysia*) a révélé que les piRNAs permettaient de mettre sous silence l'un des répresseurs transcriptionnels majeurs impliqués dans l'inhibition du processus de mémorisation, le gène CREB2 (Figure 48). En effet, chez *Aplysia*, des piRNAs ont été trouvés au niveau du système nerveux central. Ces piRNAs neuraux partagent les mêmes caractéristiques que ceux retrouvés au niveau des cellules germinales. Il a été montré qu'un piRNA, en particulier le piRNA piR-F, était produit en quantité importante en présence de sérotonine, un neurotransmetteur impliqué lors des processus de mémoire. Ce piRNA chargé sur l'unique protéine PIWI d'*Aplysia*, la protéine AcaPiwi, reconnaît, par complémentarité de séquence, le transcrit naissant du gène *creb2*, et recruterait une ADN méthyltransférase pour méthyler les îlots CpG proches de son promoteur (Rajaseethupathy et al., 2012). L'inhibition de l'expression de ce gène par le complexe acaPiwi/piR-F conduit à l'augmentation de la plasticité synaptique essentielle à l'acquisition de la mémoire à long terme.

### ***II. Régulation post-transcriptionnelle***

Lors des diverses études menées pour explorer les différentes fonctions des piRNAs, il a été démontré que les piRNAs chargés sur les protéines PIWI pouvaient également réguler post-transcriptionnellement des gènes cibles en diminuant le taux de leurs transcrits de façon



**Figure 49 : Régulation post-transcriptionnelle des transcrits *nanos* dans l'embryon de drosophile.**

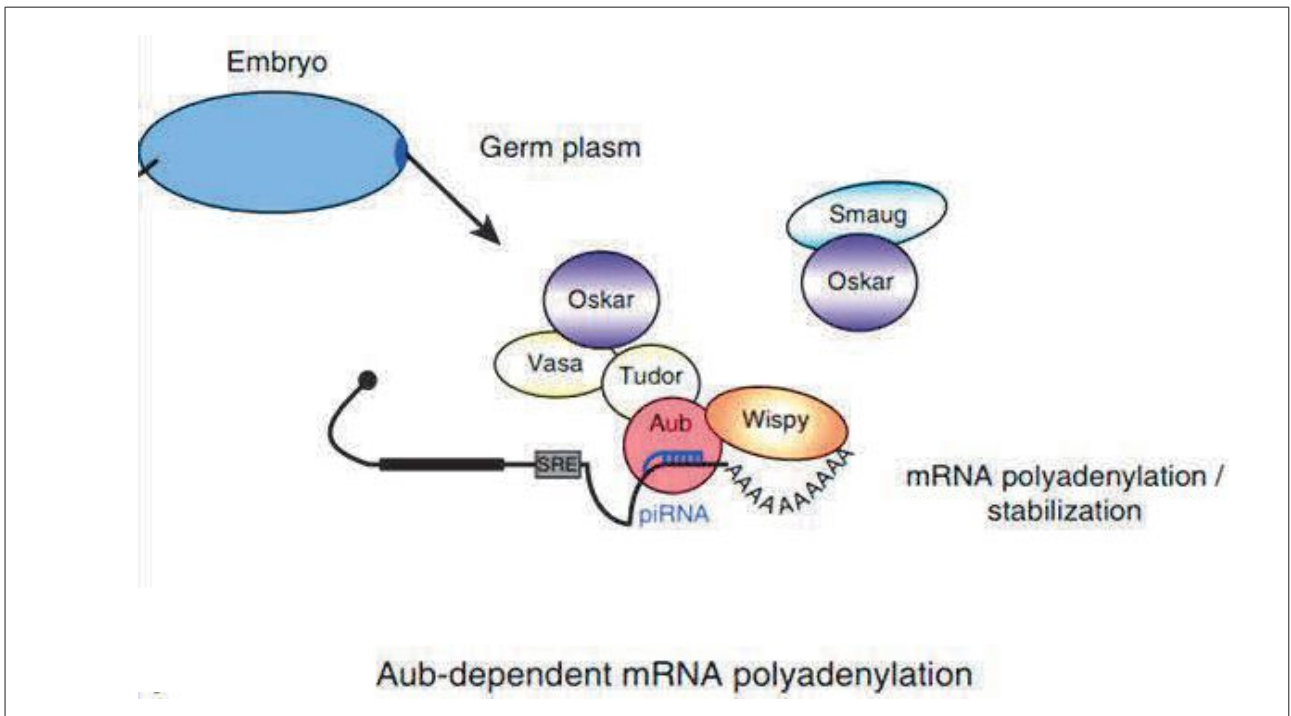
Hybridations in situ des transcrits de *nanos* en condition mutante pour Aub (A), Twin (un composant du complexe CCR-4 NOT) ou Smaug (B). C) Lors de la transition zygotique, les transcrits de *nanos* présents au niveau des parties antérieure et médiane de l'embryon sont dégradés après avoir été déadénylés *via* le complexe CCR4-NOT couplé à la protéine Smaug, ou dégradés à la suite de la coupure endonucléasique générée par la protéine Aub. Figure adaptée de 1C de l'article Rouget et al., 2010 ; de la figure 2C de l'article Zaessinger et al., 2006 ; de la figure 5D de l'article Dufourt et al., 2017.

directe. Ces régulations post-transcriptionnelles ont été mises en évidence, par exemple, au cours de différentes étapes de la gamétogenèse (notamment lors de la différenciation des spermatides ronds et lors de l'élongation des spermatides) chez la souris, mais également, durant l'embryogenèse chez la drosophile.

L'un des exemples les plus détaillés de régulation génique effectuée par les piRNAs chez la drosophile est celui de la régulation post transcriptionnelle des ARN messagers maternels du gène *nanos* au cours de l'embryogenèse précoce. Au cours de l'ovogenèse, des transcrits maternels sont produits en grande quantité dans les cellules nourricières, puis sont transférés à l'ovocyte afin d'assurer les premières étapes de l'embryogenèse. À un stade plus avancé de l'embryogenèse, les transcrits maternels sont massivement dégradés lorsque le zygote commence à exprimer son propre génome. Ce phénomène a été appelé la transition maternozygotique. Avant cette dernière, les transcrits du gène *nanos* sont localisés dans l'embryon précoce avec une accumulation préférentielle au pôle postérieur de l'embryon, dans les futures cellules germinales (Figure 49A). Lors de la transition zygotique, les transcrits de *nanos* présents au niveau de la partie antérieure et médiane de l'embryon sont dégradés après avoir été déadénylés *via* le complexe CCR4-NOT couplé à la protéine Smaug (Figure 49B) (Zaessinger et al., 2006). En 2010, Rouget et ses collègues ont montré que la protéine Aub était requise pour que l'ARN messenger de *nanos* soit dégradé dans la partie antérieure et médiane de l'embryon lors de la transition maternozygotique. En effet, en absence d'Aub, les transcrits du gène *nanos* ne sont pas dégradés et restent localisés dans la partie somatique de l'embryon après la transition maternozygotique (Rouget et al., 2010). Les analyses bio-informatiques ont révélé que parmi les piRNAs chargés sur la protéine Aub, certains sont complémentaires à la région 3' UTR du transcrit du gène *nanos*. C'est-à-dire que le complexe piRNA/Aub reconnaît le transcrit du gène *nanos*. Une fois que le complexe est lié à l'ARN messenger du gène *nanos*, une régulation post-transcriptionnelle est mise en place. Le recrutement du complexe piRNA/Aub sur les transcrits du gène *nanos* induit le recrutement du complexe CCR4-NOT et de la protéine Smaug. Ces protéines vont déadényler les transcrits qui seront par la suite dégradés. Il a également été proposé que la dégradation des transcrits *nanos* pourrait également être due uniquement à l'activité endonucléasique d'Aub qui couperait alors l'ARN messenger *nanos* à la suite de sa reconnaissance.

Cependant, la régulation post-transcriptionnellement du gène *nanos* lors de la transition maternozygotique est spécifique à la partie antérieure et médiane de l'embryon. En effet, après la transition maternozygotique, l'accumulation des transcrits du gène *nanos* au





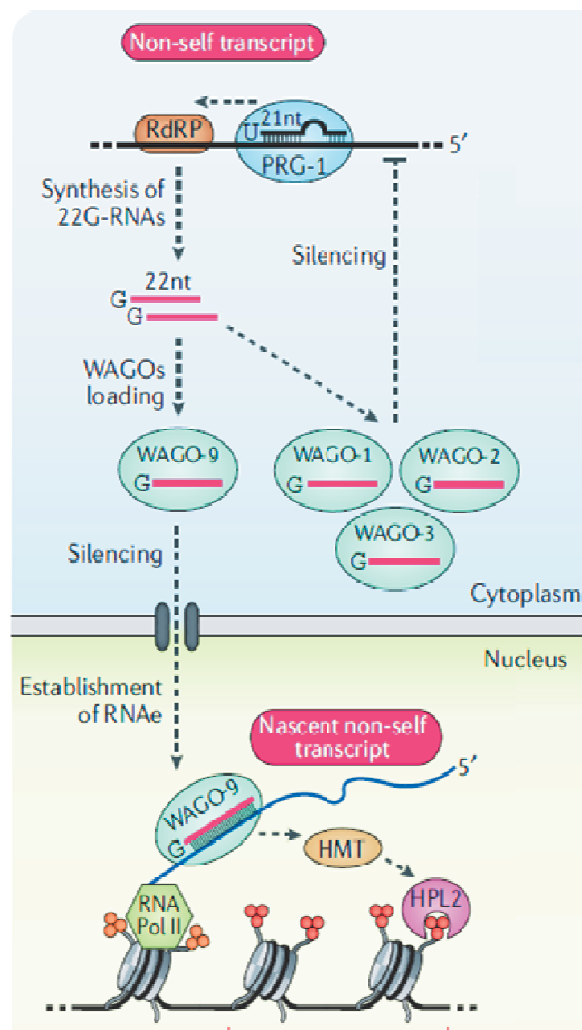
**Figure 50 : Les complexes piRNA/Aub jouent un rôle dans la stabilisation des transcrits maternels au pôle postérieur des embryons de drosophile.**

La reconnaissance des transcrits maternels *nanos* par le complexe piRNA/Aub induit le recrutement de la protéine germinale Wispy, une polyA polymérase. La polymérase Wispy génère alors une élévation de la queue polyA permettant ainsi de stabiliser les transcrits de *nanos* maternels au sein des cellules germinales embryonnaires.

Figure adaptée de la figure 5D de l'article Dufourt et al., 2017

sein des futures cellules germinales localisées au pôle postérieur de l'embryon est encore observable suggérant que le complexe piRNA/Aub ait un rôle différent de celui qu'il exerce ailleurs dans l'embryon. Au sein des futures cellules germinales, le complexe piRNA/Aub protégerait donc les transcrits du gène *nanos* (Figure 49 C). Lors de la mutation des séquences ciblées de l'ARN messenger de *nanos* par les piRNAs, les piRNAs ne peuvent plus reconnaître et protéger les transcrits maternels du gène *nanos* induisant alors la perte de leur accumulation au sein des futures cellules germinales au pôle postérieur de l'embryon. En effet, la reconnaissance des transcrits maternels *nanos* par le complexe piRNA/Aub induit le recrutement de la protéine germinale Wispy, une polyA polymérase (Figure 50). La polymérase Wispy génère alors une élongation de la queue polyA permettant ainsi de stabiliser les transcrits de *nanos* maternels au sein des cellules germinales embryonnaires (Dufourt et al., 2017). Par ailleurs, le complexe CCR4-NOT est absent au pôle postérieur. En d'autres termes, ces données incitent à penser que les piRNAs liés aux protéines Aub jouent un rôle dans la reconnaissance et la stabilisation des transcrits maternels au pôle postérieur. Par conséquent, les piRNAs liés aux protéines Aub ont un rôle dans la spécification des cellules germinales ainsi que dans leurs développements. Récemment, des expériences de iCLIP réalisées chez l'embryon précoce de drosophile ont permis de démontrer que plusieurs centaines de transcrits maternels étant reconnus par le complexe piRNA/Aub (Barckmann et al., 2015). Dès lors, d'autres transcrits tels que le transcrit du gène *dude*, jouant un rôle dans le métabolisme des adénosines monophosphates cycliques, ont été identifiés comme étant régulés par les complexes piRNA/Aub lors de la transition zygotique.

Chez *Bombyx mori*, la différenciation mâle (WZ) femelle (WW) repose sur la production de piRNAs spécifiques chez les femelles. Tout comme pour les locus *Stellate* et *Supressor of Stellate* de la drosophile, chez *B. mori* le locus *Fem* localisé sur le chromosome W est responsable de la production de piRNAs dont les séquences sont homologues aux transcrits du gène *Masculinizer*. Le gène *Masculinizer* est localisé, quant à lui, sur le chromosome Z. Les piRNAs produits à partir du locus du chromosome W chez les femelles sont liés à la protéine Siwi, une protéine PIWI. La reconnaissance des transcrits par complémentarité de séquence avec le piRNA chargé sur la protéine Siwi induit une coupure de l'ARN messenger du gène *Masculinizer* (Kiuchi et al., 2014). Lors de l'étude des piRNAs produits au sein des embryons femelles de *Bombyx mori*, une signature ping-pong a été retrouvée entre des piRNAs qui dérivait de l'ARN messenger du gène *Masculinizer*, préférentiellement liés aux protéines bmAgo3, et les piRNAs produits par le locus *Fem*,



**Figure 51 : Reconnaissance des transcrits du « non-soi » via les piRNAs chez *C. Elegans*.**

Les transcrits des séquences du « non-soi » sont reconnus par les complexes piRNA/PRG-1. Une fois que le piRNA interagit avec sa cible, la protéine PRG-1 recrute une ARN polymérase ARN dépendante (RdRP) pour synthétiser de petits ARN interférents secondaires (siARNs, 22G-ARN) en utilisant la cible comme matrice de transcription.

Les ARNs 22G chargés sur WAGO, protéine Argonaute, sont importés au noyau et reconnaissent, par complémentarité de séquence avec le siARN, le transcrit naissant d'un locus alors identifié comme du « non-soi ». La reconnaissance du transcrit par le complexe induit le recrutement d'une histone méthyltransférase (HMT) et de la protéine homologue à la protéine HP1, HPL2, afin de mettre en place des structures hétérochromatiques pour mettre sous silence les régions du « non-soi ».

Figure adaptée de la figure 5B de l'article Ozata et al., 2019.

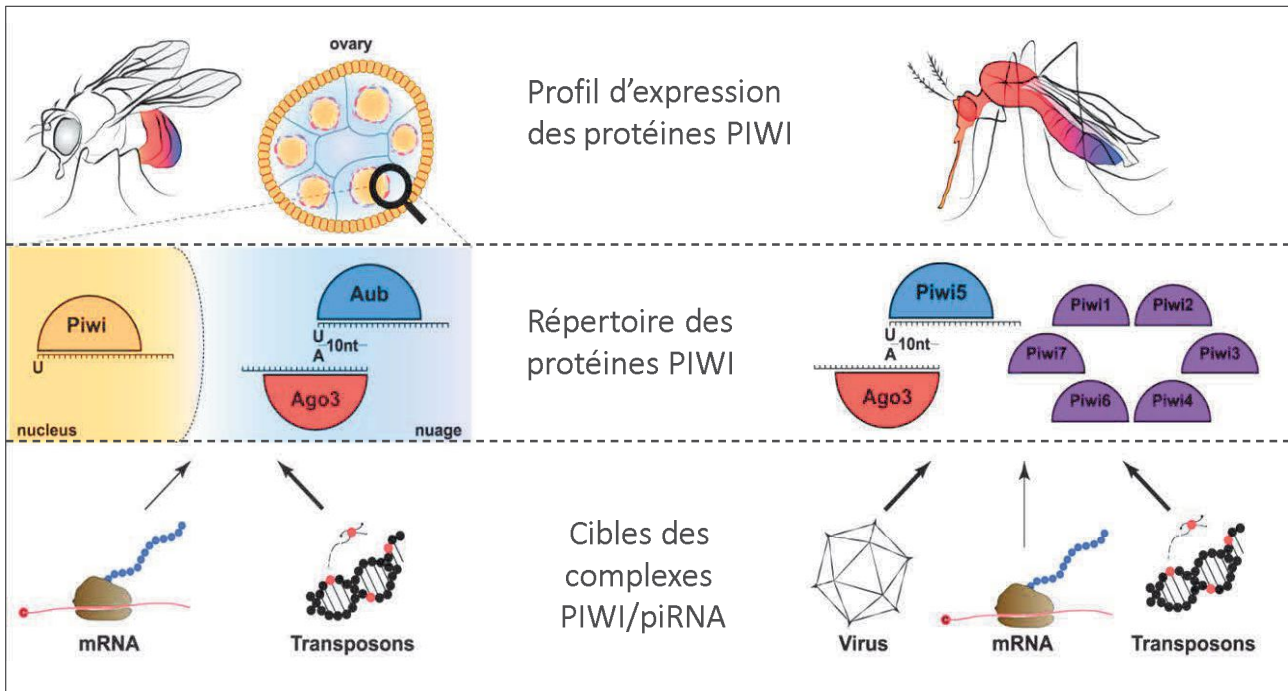
chargées sur la protéine Siwi. Ces données suggèrent que la coupure du transcrite du gène *Masculinizer* par le complexe piRNA *Fem/Siwi* induit sa maturation en piRNAs et permet ainsi d'initier le cycle ping-pong chez les embryons femelles. En absence de la production de piRNAs à partir du locus situé sur le chromosome W des femelles, les embryons se masculinisent évoquant que la détermination du sexe chez ces animaux dépend des piRNAs.

La régulation post-transcriptionnelle des gènes par les piRNAs semble être une fonction conservée au niveau gonadique chez les animaux. En effet, des piRNAs, n'ayant pas de complémentarité avec les ETs, ont été retrouvés en quantité non négligeable au sein des gonades de nombreuses espèces telles que le cochon, le rat, le poulet, le marmoset et l'Homme (Gebert et al., 2015 ; Hirano et al., 2014 ; Lau et al., 2006 ; Li et al., 2013 ; Williams et al., 2015). Récemment, deux études indépendantes suggèrent que la régulation post-transcriptionnelle des gènes par des piRNAs serait une fonction ancestrale des piRNAs (Lewis et al., 2018 ; Praher et al., 2017).

### ***C. La réorganisation du génome***

Chez certaines espèces, les piRNAs chargés sur les protéines PIWI participent à la protection des génomes en discriminant le « soi » du « non soi » comme c'est le cas pour *C. elegans* (Figure 51) (Izumi et Tomar, 2014) et *Tétrahymena* (Kataoka et Mochizuki, 2011). En effet, chez *Tétrahymena*, les séquences reconnues comme faisant partie du "non-soi", telles que les ETs ou encore des régions inconnues riches en A-T, sont éliminées lors de la reproduction sexuée (Figure 24). Ce réarrangement génomique leur permet ainsi de se protéger des ETs en les éliminant, contrairement aux autres animaux qui les répriment *via* la voie des piRNAs.

Durant la reproduction sexuée, deux micronucléi (MIC) haploïdes parentaux (noyaux germinaux) fusionnent pour former le nouveau noyau zygotique. Les macronucléi (MAC) parentaux (noyaux polyploïdes somatiques) sont quant à eux perdus. Le noyau zygotique se divise alors pour former un nouveau noyau germinale diploïde, MIC, et un nouveau noyau somatique polyploïde, MAC. Le nouveau MAC va, par la suite, subir une élimination des séquences apparentées aux ETs, les séquences IES (internal eliminated sequences) afin de ne garder que les gènes qui seront transcrits (Figure 24). Pour cela, des scanRNAs sont produits par la protéine Dcl1 au niveau du noyau germinale (MIC) et sont chargés sur la protéine Twi1 (Mochizuki et al., 2002). L'absence d'une de ces deux protéines, Dcl1 et Twi1, entraîne un



**Figure 52 : Comparaison des voies des piRNAs actives chez *Drosophila melanogaster* et chez *Aedes aegypti*.**

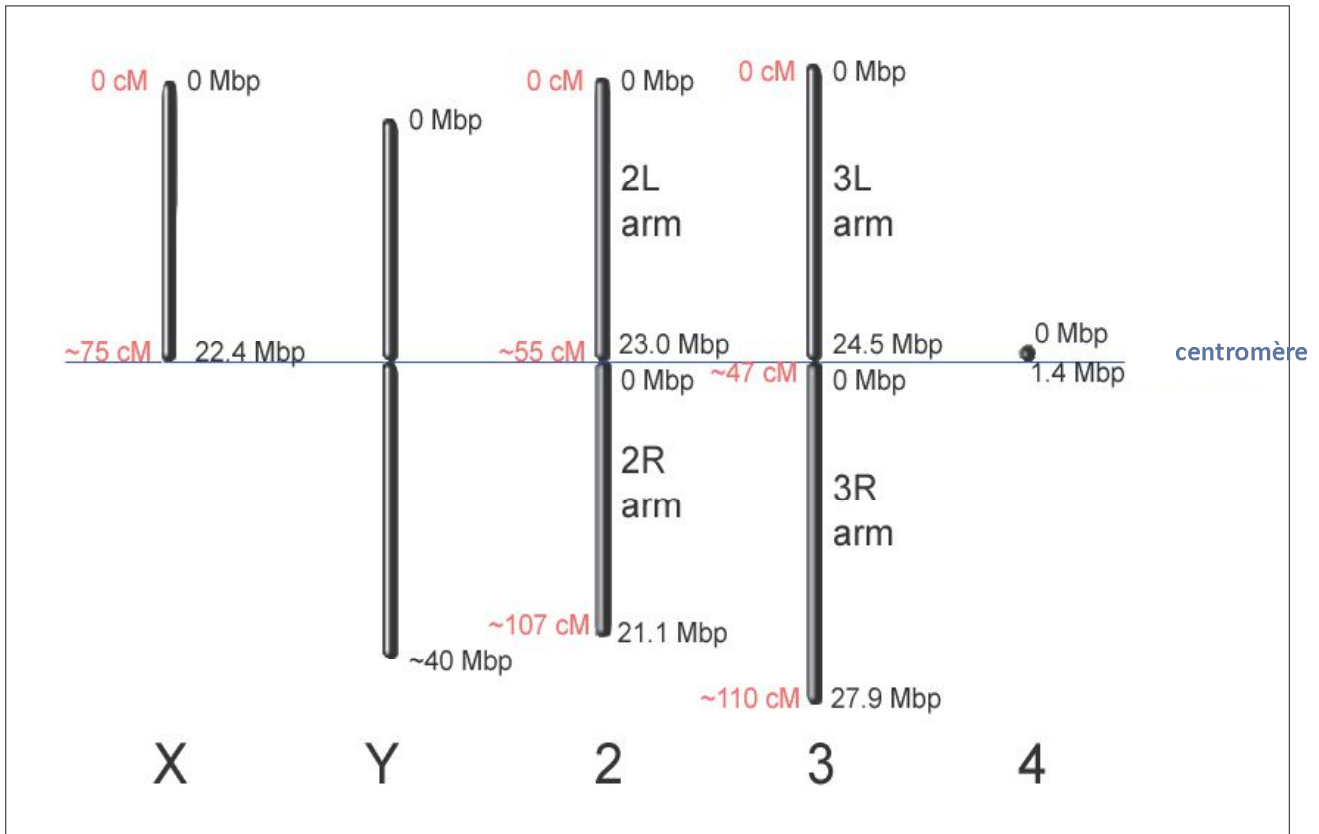
Chez la drosophile, les protéines PIWI sont presque exclusivement exprimées dans les tissus gonadiques. La protéine Piwi est une protéine nucléaire exprimée dans les cellules germinales et les cellules somatiques de l'ovaire, alors que l'expression des protéines Aub et Ago3 est restreinte aux cellules germinales. Les piRNAs de la drosophile sont principalement dérivés de séquences d'ETs et, dans une moindre mesure, d'ARNm.

Chez *Aedes aegypti*, les protéines PIWI sont au nombre de huit : Piwi 1-7 et Ago3. Certaines de ces protéines sont exprimées en dehors des gonades. Les piRNAs d'*Aedes aegypti* sont produits à partir des séquences d'ETs, des ARNm mais aussi à partir d'ARNs viraux.

blocage de l'élimination des séquences IES au niveau du nouveau MAC suggérant que ces deux protéines ont, respectivement, un rôle crucial pour la production des scanRNAs et la prise en charge de ces derniers. Le complexe Twil/scanRNA est ensuite acheminé au MAC où il reconnaît, par complémentarité de séquence, les transcrits naissant des IES. Par la suite, des marques histones répressives H3K9 seraient déposées par la protéine *ezl1* sur les régions ciblées par le complexe Twil/scanRNA. La protéine HP1 est ensuite recrutée afin de réprimer l'expression des séquences IES dans un premier temps. Cette étape préliminaire à l'élimination des séquences IES du génome du MAC est relativement similaire à celle mise en place pour la mise sous silence des ETs chez la drosophile. Dans un second temps, l'ensemble de ces marques hétérochromatiques est reconnu par un mécanisme permettant l'élimination des séquences IES du génome du MAC encore inconnu à ce jour.

#### ***D. La défense virale***

La défense virale au sein des tissus somatiques a initialement été décrite comme étant assurée par les voies de siRNAs. Néanmoins, chez certaines espèces animales telles que les moustiques, la défense virale requiert la voie des piRNAs (Lewis et al., 2018 ; Schnettler et al., 2013 ; Morazzani, et al., 2012 ; Miesen, Girardi et van Rij, 2015). Pour cela, les transcrits de chaque brin des transcrits viraux sont pris en charge par les protéines Piwi5 et Ago3 (Figure 52). Ces deux protéines vont former un complexe hétérotypique initiant ainsi le cycle ping-pong. Les transcrits viraux sont peu à peu maturés en piRNAs évitant ainsi la production de nouveaux virus. Néanmoins, chez la drosophile, il a été montré que la défense virale ne faisait pas intervenir la voie des piRNAs (Petit et al., 2016). La différence de fonction des complexes piRNA/PIWI entre la drosophile et le moustique peut être expliquée d'un point de vue évolutif. En effet, chez la drosophile seules trois protéines PIWI sont codées par le génome alors que chez le moustique, notamment *Aedes aegypti*, les protéines PIWI sont en nombre de sept. Il a ainsi été proposé qu'au cours de l'évolution des espèces, les protéines PIWI chez le moustique auraient acquis, en plus de leur fonction régulatrice des ETs, des fonctions supplémentaires antivirales, alors que les protéines PIWI chez la drosophile auraient gardé un rôle restreint à la régulation des ETs. Alternativement, il peut-être aussi proposé que la voie des piRNAs a perdu une fonction ancestrale de lutte antivirale.



**Figure 53 : Représentation schématique des chromosomes de la drosophile.**

Le génome de la drosophile est réparti sur 4 chromosomes dont trois autosomes (chromosome 2, 3 et 4).

Données issues du NCBI et Calvalo, 2002.

## Chapitre 5 : Le modèle animal

### A. *La drosophile*

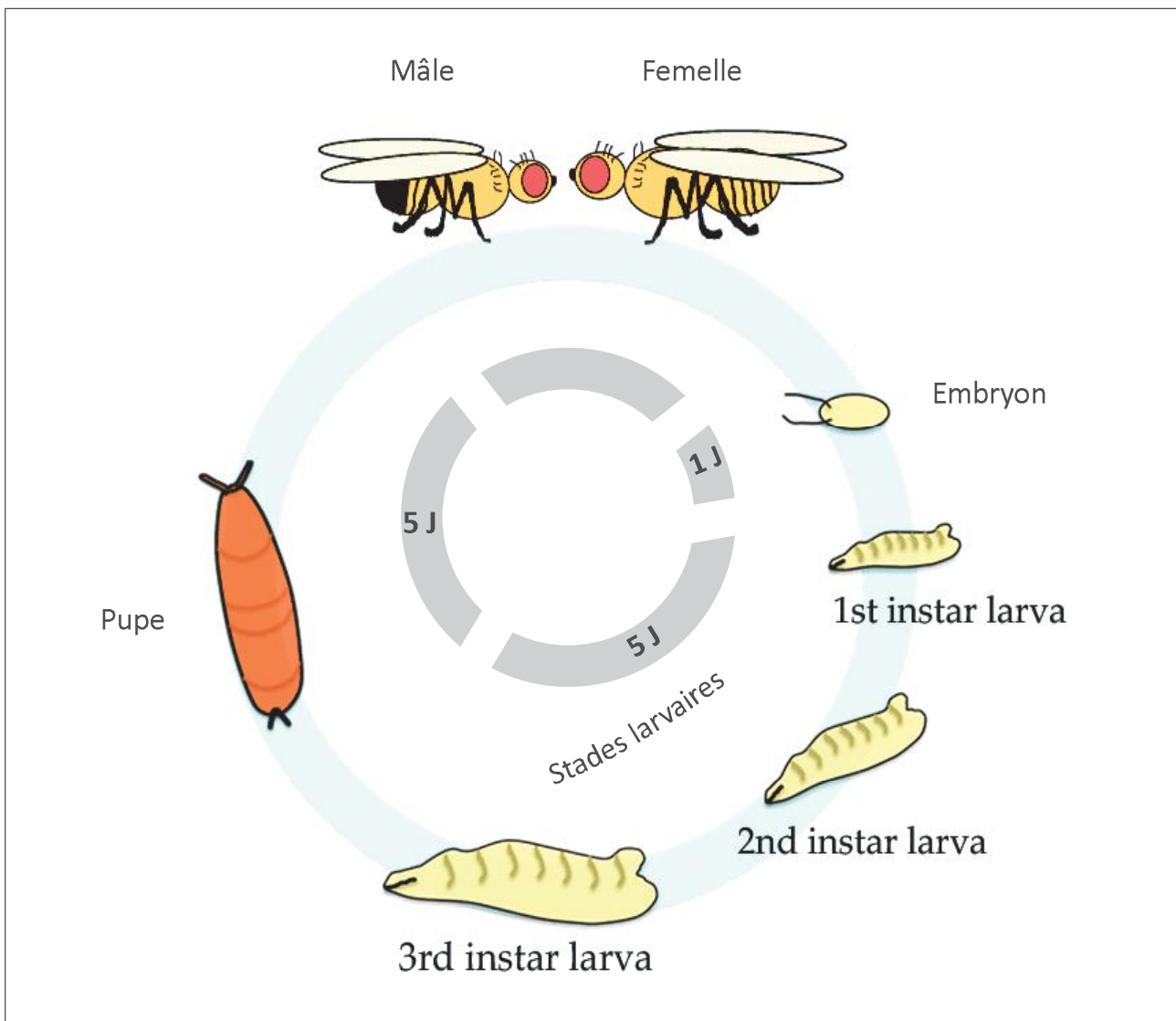
La drosophile, ou communément appelée mouche du vinaigre, est un insecte appartenant à l'ordre des Diptères. Cet insecte ne faisant que quelques millimètres est la star des modèles animaux notamment dans les domaines de la génétique et du développement. Plus de **213 960** articles scientifiques concernant la drosophile étaient publiés sur PubMed Central le 11 mai 2019.

Dans les années 1920, Thomas Hunt Morgan, embryologiste et généticien, (1866-1945) a notamment utilisé les drosophiles pour étudier la transmission des caractères parentaux à la génération suivante. Au cours des générations, une mouche aux yeux blancs est fortuitement apparue parmi la population de mouches sauvages dont les yeux sont rouges. Il suivit alors la descendance de cet individu qui fut le premier mutant génétique étudié dans l'histoire de la génétique. En 1933, il reçut un prix Nobel de Physiologie ou Médecine pour ces travaux démontrant que les chromosomes étaient le support de l'hérédité. Par ailleurs, Thomas Hunt Morgan et Hermann Müller ont réalisé les premières cartes génétiques de la drosophile en étudiant différents mutants. De nos jours, une importante collection de drosophiles mutantes est disponible et elle s'accroît chaque année.

Le génome de la drosophile est un génome de petite taille. Il est constitué de 180 millions de paires de bases, soit une vingtaine de fois plus petit que celui de l'Homme. Il est réparti en quatre chromosomes (Figure 53) : le premier chromosome est un chromosome sexuel (X ou Y) et les 3 autres (le chromosome 2, le chromosome 3 et le chromosome 4) sont des autosomes (Bosco et al., 2007). Le projet de séquençage du génome de la drosophile a débuté en 1991. Le séquençage et l'annotation du génome de la drosophile ont été terminés neuf années plus tard (Adams et al., 2000 ; Celniker et al., 2002). En 2000, près de 98% de la partie codante du génome de la drosophile a été séquencée, soit plus de 120 Mb contenant environ 13 600 gènes.

Malgré une distance évolutive importante avec l'Homme, de nombreux gènes, de nombreuses voies de régulations dont la voie des piRNAs et de nombreux réseaux moléculaires sont conservés entre l'Homme et la drosophile. À titre d'exemple, près de 75% des gènes impliqués dans des pathologies humaines, telles que le cancer, ont des gènes





**Figure 54 : Le cycle de vie de la drosophile.**



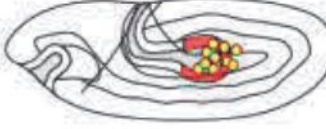





L'embryon pondu sur le milieu éclot 24 heures après en larve. La larve va croître durant 5 jours avant de s'encapsuler dans une pupa. Le stade pupal dure également 5 jours. Durant cette période, la larve va se métamorphoser en mouche adulte qui émerge à l'issue du stade pupal. Au bout de quelques heures, les femelles vont s'accoupler avec les mâles et stockeront le sperme de ces derniers dans la spermathèque pour l'utiliser ultérieurement.

homologues chez la drosophile (Bier, 2005). Les nombreux outils génétiques développés dans ce modèle ont permis l'acquisition rapide de connaissances sur de nombreuses voies conservées chez les animaux. Un des avantages de cet organisme est que les mâles n'effectuent pas de recombinaison ce qui facilite les analyses génétiques.

L'intérêt de ce modèle ne se limite pas à la génétique. En effet, les conditions d'élevage de cet animal sont faciles et relativement peu coûteuses, que ce soit en termes de nourriture, de conditionnement ou encore de stockage. La drosophile possède également une capacité reproductive impressionnante par rapport aux autres modèles. En effet, son cycle de génération n'est que d'une dizaine de jours à 20°C contre 40 jours pour la souris (Figure 54). La drosophile peut générer 25 générations en un an. Cependant, la durée de ce cycle de génération est dépendante de la température. En effet, le cycle peut durer 50 jours à 12°C, alors qu'à 30°C, il ne dure que 8 jours. De plus, une femelle drosophile peut à elle seule pondre entre 25 et 35 œufs par jour. Autrement dit, un nombre important de descendants peut donc être obtenu rapidement. L'œuf pondu sur le milieu éclot 24 heures après en larve. La larve va croître durant 5 jours avant de s'encapsuler dans une pupa. Le stade pupal dure également 5 jours. Durant cette période, la larve va se métamorphoser en mouche adulte qui émerge à l'issue du stade pupal. Au bout de quelques heures, les femelles vont s'accoupler avec les mâles et stockeront le sperme de ces derniers dans la spermathèque pour l'utiliser ultérieurement. L'espérance de vie d'une drosophile est en moyenne de un à deux mois.

### ***B. Les gonades chez la drosophile***

L'ovocyte est fécondé lors de son passage dans la spermathèque. La fécondation va induire des mitoses incomplètes. En effet, le noyau zygotique subit une succession de divisions nucléaires sans cytotélerèse. La fréquence de ces divisions est estimée à une division nucléaire environ toutes les huit minutes. Les noyaux ainsi créés vont se répartir dans le cytoplasme de l'embryon formant ainsi un syncytium. On parle alors d'embryon syncytial. Lors de la 10<sup>ième</sup> division, les noyaux vont migrer vers la périphérie. C'est le stade de blastoderme syncytial durant lequel débute la cellularisation. Les premières cellules à être cellularisées sont les cellules germinales primordiales (PGCs) localisées dans la partie postérieure de l'embryon au stade 5 (Figure 55). La formation des PGCs dépend notamment de l'accumulation de protéines postérieures telles qu'Oskar, Vasa, Tudor et Aub.

Stage		Step	Genes
Antérieur ↔ Postérieur			
Stage 5 -2.5 h AEL		Formation and transcriptional silencing	<i>osk, vas, aub, tud, mtr RNA, gcl, nos, pum, pgc</i>
Stage 8 -3.5 h AEL		Adhesion to midgut	
Stage 9 -4.5 h AEL		Transgut migration	<i>tre-1, srp, hkb</i>
Stage 10 -5h10m AEL		Movement on midgut	<i>Wun and wun2</i>
Stage 11 -7h AEL		Migration to mesoderm	<i>hmgcr, fpps, quemao, GTT1, tin, zfh1, bryn, slam</i>
Stage 12 -9h AEL		Association with gonadal mesoderm	
Stage 13 -10.5 h AEL		Germ line-soma alignment	<i>abdA, abdB, cli, tin, six4/5</i>
Stage 14 -11.5 h AEL		Gonad coalescence	<i>toi, E-cadherin</i>

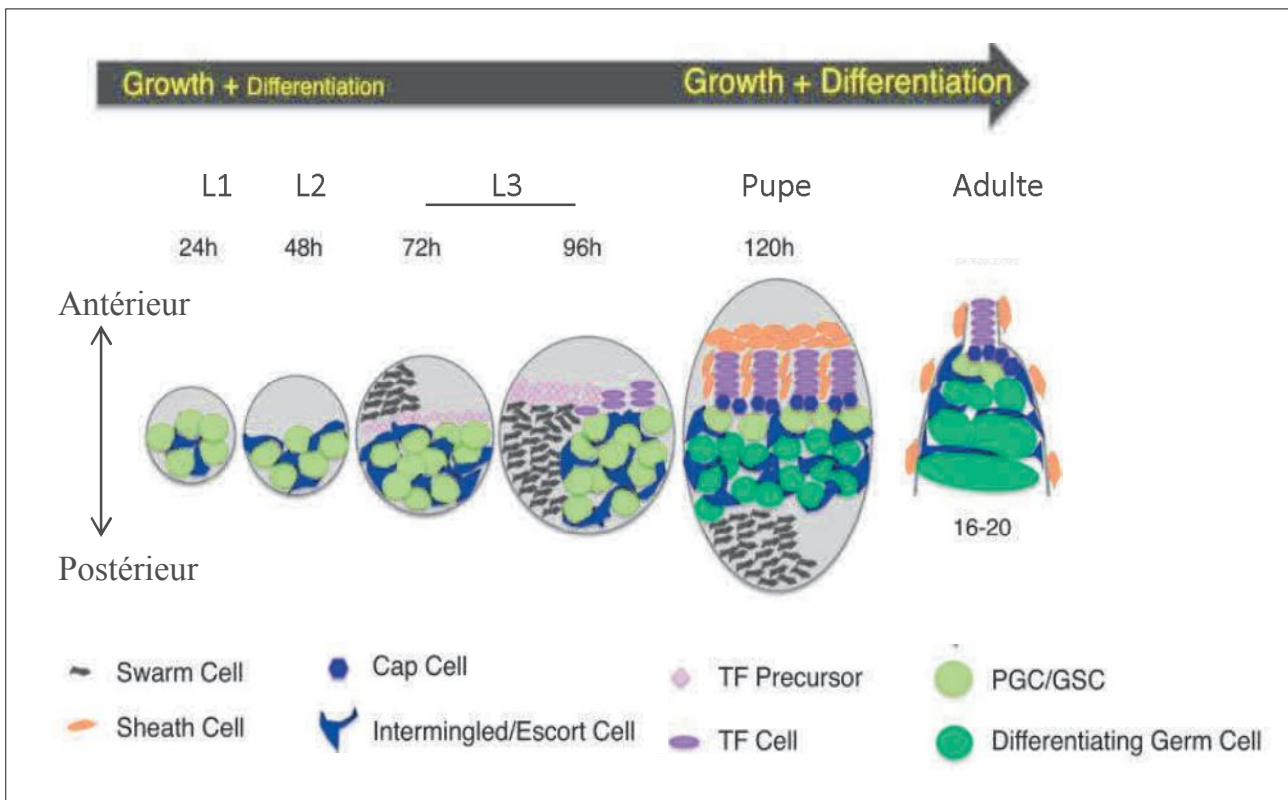
**Figure 55 : Développement des gonades au cours de l'embryogenèse de la drosophile.**

Synthèse des différentes étapes du développement des gonades embryonnaires. Les cellules germinales sont représentées en jaune et l'intestin est représenté en rouge. Le mésoderme est représenté uniquement au stade 10 en vert. Les précurseurs des cellules somatiques des gonades sont représentées en vert à partir du stade 11. Les gènes exprimés spécifiquement aux différents stades sont listés à droite de la figure.

Figure adaptée de la figure 2 de Santos et Lehmann, 2004.

Adjacentes aux cellules donnant dans la partie postérieure de l'intestin moyen, les cellules PGCs vont migrer de manière passive le long de la face dorsale durant la gastrulation. Lors de l'invagination de l'intestin moyen postérieur au stade 7, les PGCs sont internalisées passivement à l'intérieur de l'embryon. Puis au stade 9, les PGCs migrent activement à travers l'épithélium de l'intestin moyen postérieur en direction du mésoderme adjacent (Figure 55). Au stade 12, les PGCs migrent bilatéralement pour s'associer successivement aux trois groupes de précurseurs gonadiques somatiques (qui sont répartis dans les parasegments 10, 11 et 12) en commençant par les pools de précurseurs les plus postérieurs. Un pool additionnel de précurseurs gonadiques somatiques, indépendant des trois autres, est retrouvé dans le parasegment 13 chez les futures femelles et mâles. Néanmoins, au cours du développement, les précurseurs somatiques du parasegment 13 sont éliminés par apoptose chez les futures femelles alors que chez les futurs mâles, ces cellules vont migrer vers l'avant pour être intégrées à la gonade mâle. Au stade 13, c'est à dire une fois que les précurseurs gonadiques somatiques les plus antérieurs ont été intégrés à la future gonade, l'ensemble des cellules fusionnent pour former des gonades embryonnaires allongées de chaque côté de l'embryon en cours de développement. Les précurseurs gonadiques somatiques vont alors recouvrir les PGCs *via* des extensions membranaires : c'est la coalescence des gonades (Jaglarz et Howard, 1995). Vers le stade 15, les cellules gonadiques se compactent pour former des gonades sphériques. L'éclosion de la larve est réalisée au stade 17.

Tout au long des stades larvaires, les PGCs et les précurseurs des cellules somatiques composant les gonades mâles et femelles vont proliférer et se différencier aboutissant à la formation respective des futurs testicules et des futurs ovaires (Gilboa et Lehmann, 2006). Néanmoins, la taille des gonades mâles est plus importante que celle des gonades femelles lors des stades larvaires. Chez les larves femelles, les PGCs vont subir quatre mitoses asynchrones successives passant ainsi d'environ 25 PGCs par ovaire à plus de 150 PGCs au stade pupal. Lors du premier (L1) et du second (L2) stade larvaire, les PGCs, localisées dans la partie postérieure des gonades femelles, sont entourées par un groupe de cellules somatiques et interstitielles, les "Escort Cells" ou "Intermingled Cells" (Figure 56). Lors du stade L3, les PGCs vont s'aligner au centre du futur ovaire juste en dessous des filaments terminaux pour initier la formation des ovarioles. Les PGCs sont alors entourées des cellules somatiques provenant du lignage cellulaire des "Escort Cells", mais aussi des "Cap Cells" (cellules différenciées des "Escort Cells"). Ces différentes cellules somatiques ainsi que les cellules des filaments terminaux vont constituer les différentes niches (à raison de 2 à 3 PGCs



**Figure 56 : Développement des gonades au cours des stades larvaires, du stade pupal et au stade adulte de la drosophile.**

Au cours des stades larvaires, les cellules souches germinales (en vert clair) se multiplient et s'entourent des « Cap cells » et des « Escort cells », deux types cellulaires ayant une origine somatique (bleu). Cet ensemble de cellules reste localisé au niveau postérieur de l'ovaire jusqu'au stade pupal. Pendant ce temps, dans la partie antérieure de la gonade se mettent en place les filaments terminaux.

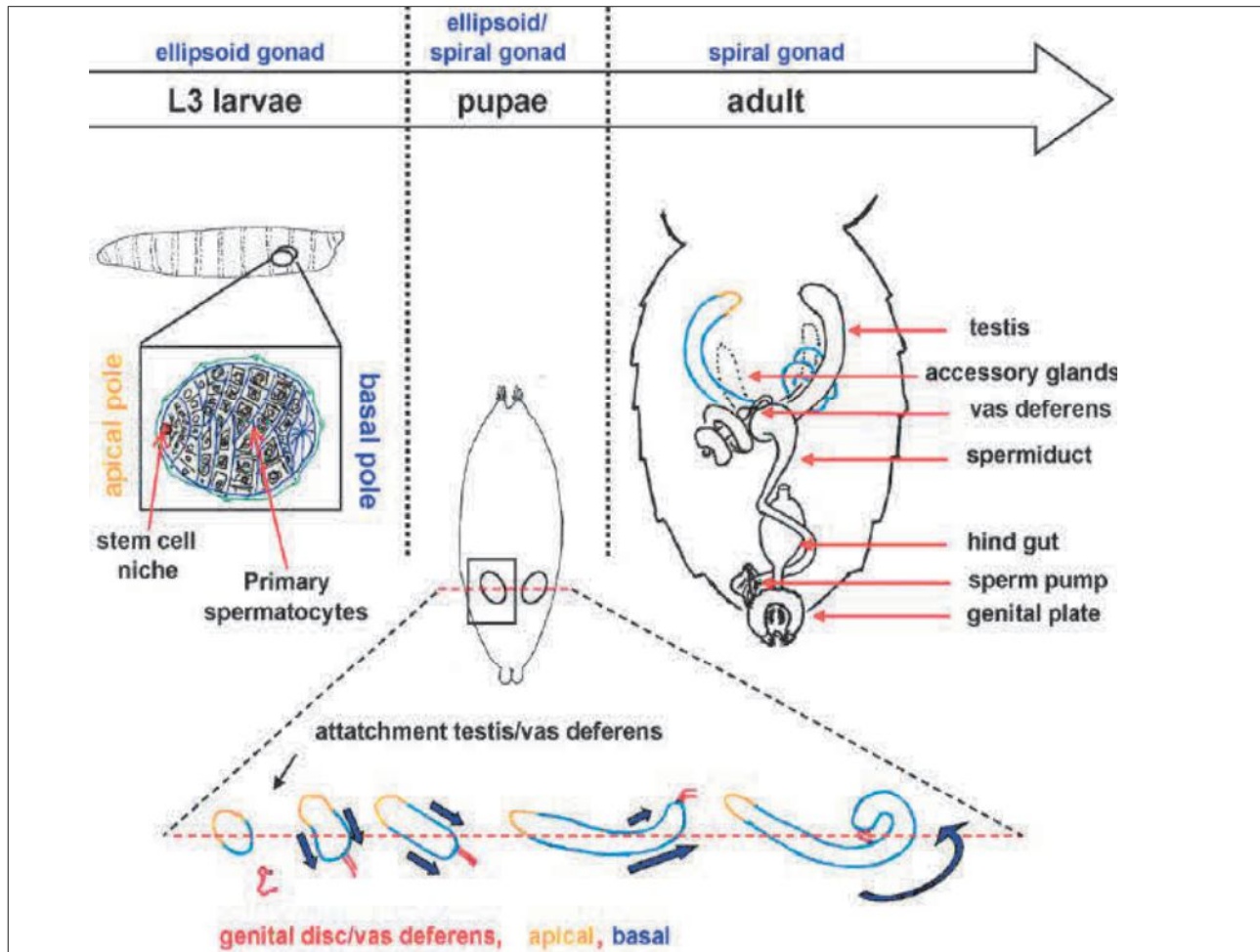
Durant le stade pupal, les cellules des filaments terminaux vont contacter les « cap cells » et ainsi faire migrer l'ensemble des cellules localisées en postérieur vers l'antérieur de la gonade formant ainsi l'ébauche du germarium. Durant cette migration, les cellules germinales les plus loin des « cap cells » vont commencer à entrer en mitose incomplète (vert foncé) amorçant ainsi l'ovogenèse. En fin de métamorphose, les gonades femelles sont contactées et attachées aux oviductes et les différentes niches sont individuellement séparées pour former les 16 à 20 ovarioles de chaque ovaire.

Figure adaptée de la figure 1 de l'article Gilboa, 2015

par niche). Les cellules germinales femelles des larves L3 restent bloquées au stade d'ovogonies (avant la méiose). Lors de la transition du stade L3 vers les stades pupaux, les cellules de la niche ainsi que les futures cellules germinales souches vont migrer au niveau de la partie antérieure du germarium du futur ovaire. La structure des gonades est par la suite modifiée lors des stades pupaux. Entre 36 et 42h après l'entrée en pupe, les gonades femelles sont contactées et attachées aux oviductes. En fin de métamorphose, les différentes niches sont individuellement séparées pour former les 16 à 20 ovarioles de chaque ovaire (Figure 26A et 26B).

Chez le mâle, la spermatogenèse commence dès le stade larvaire L1 après que la niche ait été mise en place dans la partie apicale du futur testicule. Les premières spermatides matures sont observées suite à la méiose II en début de stade pupal. En d'autres termes, la spermatogenèse est fonctionnelle chez le mâle dès le stade pupal. Trente heures après l'entrée en pupe, les futurs testicules vont fondamentalement changer de forme en passant d'une gonade ovoïde à une structure allongée et en spirale (Figure 57). En parallèle, les vésicules séminales ainsi que le canal déférent (produites à partir du disque germinal) vont fusionner avec les testicules.

Chez la drosophile adulte, le tissu reproducteur femelle est constitué d'une paire d'ovaires renfermant chacun environ seize ovarioles (Figure 58). Chaque ovariole est organisée en une succession de follicules à des stades de maturation progressifs aboutissant à la formation d'un œuf selon l'axe antéro-postérieur. La partie la plus antérieure et la plus précoce de l'ovariole, appelée le germarium, contient les cellules souches germinales et les cellules souches somatiques folliculaires. Les deux cellules souches germinales sont localisées dans une niche à l'apex du germarium. Elles vont ensuite se diviser de façon asynchrone, aboutissant à la formation d'une cellule fille souche germinale qui restera localisée dans la niche, et d'une cellule localisée plus postérieurement (le cystoblaste), qui va se détacher de la niche pour subir quatre divisions mitotiques successives incomplètes (absence de cytotérièse). À la fin des différentes divisions, un cyste de seize cellules reliées entre elles par des ponts cytoplasmiques est obtenu. Parmi ces seize cellules, seules deux d'entre elles vont posséder quatre ponts cytoplasmiques contre trois pour les autres. Ces cellules vont entrer en prophase I et ainsi devenir des pro-ovocytes. Néanmoins, seule une de ces deux cellules est sélectionnée pour donner l'ovocyte et reste bloquée en prophase I. Le second pro-ovocyte, quant à lui, sort de la méiose pour subir une endoréplication de son ADN et devient alors une cellule nourricière. Le cyste est alors composé de quinze cellules nourricières et d'un ovocyte.



**Figure 57 : Développement des testicules chez la drosophile.**

Au stade larvaire L3, les gonades sont ellipsoïdes. Le pôle apical (orange) contient la niche de cellules souches germinales et somatique alors que le pôle basal (bleu) sont localisés les spermatocytes primaires. Au stade pupal, les testicules subissent une élongation (flèches bleues). Le canal déférent dérivant du disque génital va alors contacté et se fixé à la base du testicule qui s'enroulera par la suite. Au stade adulte, les testicules en spirales ainsi que tous les composants génitaux internes et externes dérivés du disque génital sont mis en place.

Figure adaptée de la figure 3 de l'article Coutelis et al., 2008.

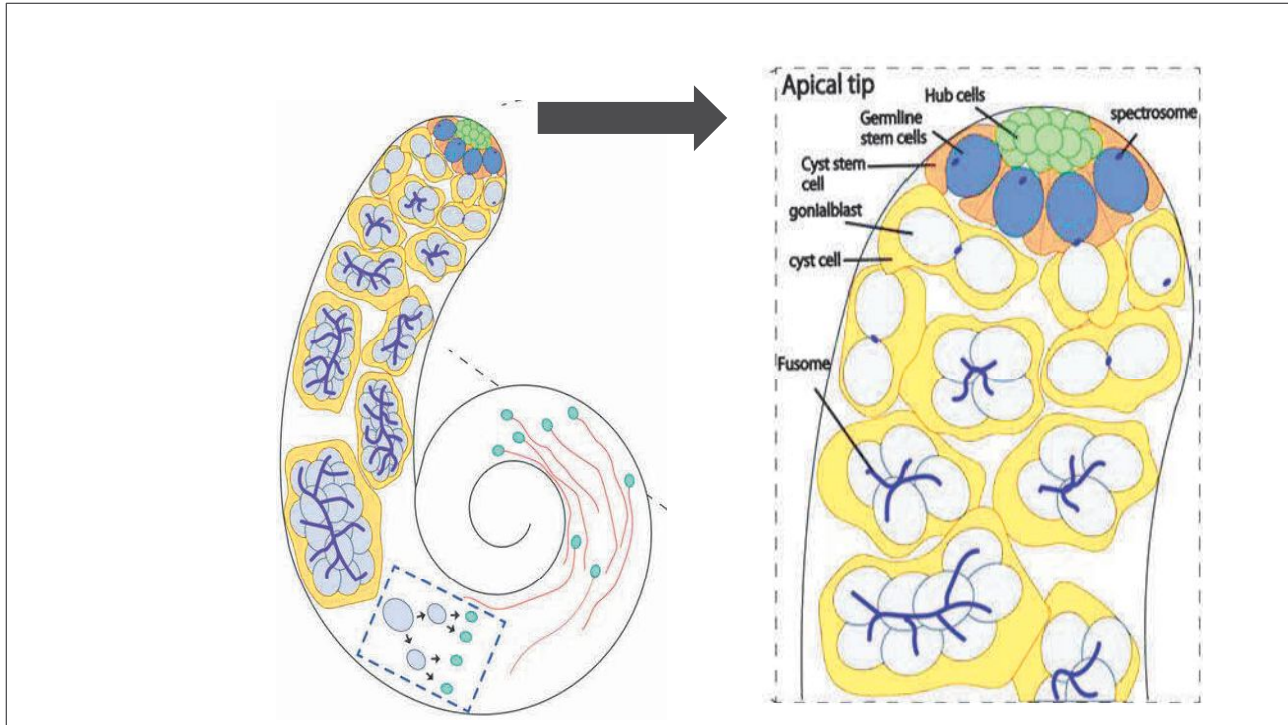
Les cellules nourricières synthétisent un nombre important d'ARNs messagers et de protéines qui seront déversés dans l'ovocyte durant les stades tardifs de l'ovogenèse. Dans la partie postérieure du germarium, les 16 cellules germinales sont encapsulées par une monocouche de cellules somatiques, les cellules folliculaires issues des cellules souches somatiques, pour former un follicule ou chambre d'œuf (Spradling, 1993). Les cellules folliculaires sont reliées les unes aux autres par des jonctions serrées. À la sortie du germarium, le follicule néoformé poursuit sa maturation lors de sa progression le long de l'ovariole jusqu'à la formation d'un œuf. L'ovogenèse le long des ovarioles dure une dizaine de jours.

Le tissu reproducteur des drosophiles mâles adultes est constitué, quant à lui, de deux testicules en forme de spirale. L'apex des testicules renferme les cellules germinales souches qui, comme pour les cellules germinales souches ovariennes, se divisent de façon asymétrique. Contrairement au germarium, l'apex des testicules de drosophile contient 7 à 8 cellules germinales souches (Figure 58). Lors de leur division asymétrique, une des cellules filles restera une cellule germinale souche, alors que l'autre cellule fille se différenciera par la suite. Cette cellule fille est appelée le "gonioblaste" et est entourée de deux cellules somatiques issues des cellules somatiques souches. Chaque gonioblaste subit quatre divisions incomplètes (absence de cytokinèse) aboutissant à la formation d'un cyste composé de 16 spermatogonies connectées entre elles par des ponts intercytoplasmiques, les "rings canals". Les spermatogonies vont alors commencer à se différencier en spermatocytes. Au cours de leur transit testiculaire, les spermatocytes primaires vont effectuer une méiose I et II aboutissant à la formation de 64 spermatides rondes renfermées dans un seul cyste. Durant la différenciation des spermatocytes en spermatides, des changements de taille et de forme seront observés. Les spermatides vont subir une élongation lors de leur transit testiculaire. Après la compaction des noyaux, les spermatides sont individualisées puis acheminées à la sortie du testicule pour être stockées dans la vésicule séminale. De façon surprenante, aucune division n'est observée pour les deux cellules somatiques qui recouvrent les cellules germinales dès leur différenciation.

### ***C. Modèle cellulaire, les cellules OSS***

Afin d'étudier les différentes propriétés des cellules souches germinales et des cellules souches folliculaires, Niki et ses collègues ont développé deux modèles cellulaires en 2006. Ils ont en effet mis en culture des cellules issues d'ovaires de drosophiles mutantes pour le





**Figure 58 : Représentation schématique du tissu reproducteur mâle adulte chez la drosophile.**

A l'apex, le testicule est composé des cellules du hub (vert) suivies des cellules germinales souches (bleu). Lors de la division asymétrique de la cellule germinale souche, la cellule fille la plus postérieure à la niche quitte cette dernière. Lors de son détachement de la niche, elle s'entoure alors de deux cellules du cyste d'origine somatiques (jaunes) formant un « gonioblaste ». Après avoir subi quatre divisions incomplètes, le cyste est composé de 16 spermatogonies. Durant leur transit testiculaires, les spermatogonies se différencient en spermatocytes. Les spermatocytes réalisent ensuite la méiose I et II aboutissant à la formation de 64 spermatides ronds. (encadré bleu). Les spermatides vont subir une élongation permettant à la suite l'individualisation des spermatides allongées en spermatozoïdes à la fin du testicule.

Figure adaptée de la figure 1B de l'article Gleason et al., 2018.

facteur de différenciation *bam*. Cette mutation empêche la différenciation des cellules germinales et donne un phénotype de sac de billes au germarium (Bag of Marbles = BAM).

La première lignée cellulaire qu'ils ont obtenue dérive uniquement des cellules souches folliculaires puisque l'ensemble de ces cellules expriment le marqueur folliculaire *FasIII*. Elles ont été nommées les cellules OSS, ou OSC, pour Ovarian Somatic Sheet Cells. Les cellules OSS sont des cellules semi-adhérentes qui sont cultivées dans un milieu adapté à la culture de cellules de drosophile (appelé le milieu M3 enrichi d'un broyat de drosophile). La température de culture des cellules OSS est de 25°C. L'utilisation de cette lignée lors de nombreux criblages a notamment permis l'identification des différents acteurs de la voie des piRNAs primaires, la seule voie de biogenèse des piRNAs présente dans les cellules folliculaires de l'ovaire de drosophile.

La seconde lignée cellulaire qu'ils ont obtenue contient un mélange de cellules souches germinales et de cellules souches folliculaires. En effet, la mise en culture de cellules germinales seules n'est pas rendue possible suggérant que les cellules souches germinales ont besoin d'un soutien somatique.

## Chapitre 6 : Objectifs de l'étude

Les ETs sont des séquences ADN mobiles retrouvées dans les génomes de toutes les espèces où ils ont été recherchés. Moteurs de l'évolution, ces éléments mobiles présents en de nombreuses copies dans les génomes ont joué un rôle majeur dans la dynamique des génomes en engendrant des mutations et des réarrangements chromosomiques. Néanmoins, étant des constituants majeurs des génomes, ils doivent être finement régulés dans le but de préserver l'intégrité des génomes et ainsi conserver l'équilibre entre variabilité et stabilité des génomes. Afin de protéger l'information génétique de l'hôte transmise à la descendance, la régulation des ETs au niveau de gonades est effectuée par la voie des piRNAs. Cette voie régule l'expression des ETs de façon transcriptionnelle et post-transcriptionnelle en utilisant le mécanisme d'ARN interférent dont les acteurs principaux sont des petits ARNs non codants de 23 à 29 nucléotides et les protéines de la grande famille ARGONAUTE. De nombreuses recherches ont été menées chez les animaux afin de comprendre, en détail, comment les petits ARNs de 23 à 29 nts étaient produits au sein des cellules germinales. Le développement des techniques de séquençage et du domaine de la bio-informatique a permis une rapide évolution des connaissances sur cette voie notamment chez la drosophile et la souris.



Afin de mieux comprendre la biogenèse des piRNAs, leur origine a été explorée en les séquençant et les cartographiant sur les génomes. Chez la drosophile, mon modèle d'étude, 142 loci producteurs de piRNAs ont ainsi été identifiés. Ces loci majoritairement hétérochromatiques sont composés d'un enchevêtrement d'ETs. Dans cette espèce, ces loci, appelés clusters de piRNAs, ont été classés en deux catégories en fonction de leurs propriétés transcriptionnelles. D'une part, les clusters de piRNAs "unistrand" sont transcrits par l'ARN polymérase II à partir d'un seul des deux brins d'ADN dans les cellules folliculaires des ovaires de drosophile. Et d'autre part, les clusters de piRNAs "dual-strand" sont transcrits de façon convergente par l'ARN polymérase II à partir des deux brins d'ADN exclusivement dans les cellules germinales des ovaires de cette espèce. L'ensemble des études réalisées sur les gonades femelles des drosophiles adultes ont permis d'identifier les caractéristiques transcriptionnelles de ces deux types de clusters de piRNAs. Cependant, les conditions nécessaires à leur activation et donc à la mise en place de la voie des piRNAs au cours du développement restent encore inconnues à ce jour. De nombreuses questions se posent donc : **À quel moment du développement la voie des piRNAs est-elle mise en place pour protéger l'information génétique transmise à la descendance ? À quel moment a lieu l'activation de l'expression des clusters de piRNAs ? Quels sont les acteurs impliqués dans l'initiation de l'expression des clusters de piRNAs ? L'expression des clusters de piRNAs est-elle concomitante à l'expression des acteurs protéiques de la voie des piRNAs ? La voie des piRNAs est-elle mise en place suite à la différenciation de sexe ? La voie des piRNAs est-elle active en dehors des gonades au cours du développement de la drosophile ?** Pour répondre à ces différentes questions, j'ai analysé en détail, avec l'aide du Docteur Benjamin Bertin, le profil d'expression d'un cluster de piRNAs au cours du développement de la drosophile en prenant comme modèle le cluster de piRNAs *flamenco*, producteur majeur de piRNAs au sein des cellules folliculaires des ovaires de drosophile.

L'expression des clusters de piRNAs permet l'obtention de longs transcrits simple brin adressés au site de maturation pour être coupés en petits ARNs de 23 à 29 nucléotides. Les transcrits issus des clusters "unistrand" sont accumulés en un foyer nucléaire, appelé le Dot COM, avant d'être exportés au cytoplasme alors que les transcrits issus des clusters de piRNAs "dual-strand" sont, quant à eux, directement acheminés à la membrane nucléaire pour être exportés au cytoplasme. Une fois dans le cytoplasme, les transcrits des clusters de piRNAs sont reconnus et acheminés à leur site de maturation. L'une des questions les plus énigmatiques dans ce domaine est de savoir **comment sont adressés les transcrits**



**provenant de clusters de piRNAs à leur site de maturation.** Dans la littérature, il a été démontré que les piRNAs hérités maternellement ont un rôle clé pour la reconnaissance et la maturation des transcrits des clusters de piRNAs "dual-strand". Néanmoins, les mécanismes d'adressage des transcrits issus de clusters de piRNAs "unistrand" restent encore inconnus à ce jour. Contrairement aux transcrits des clusters de piRNAs "dual-strand", les transcrits des clusters de piRNAs "unistrand" partagent de nombreuses caractéristiques avec les transcrits de gènes. Ils sont coiffés à leur extrémité 5' et possèdent une queue poly(A) à leur extrémité 3'. De plus, les transcrits provenant du cluster de piRNAs "unistrand" *flamenco*, locus producteur majeur de piRNA des cellules folliculaires, subissent des épissages alternatifs. En d'autres termes, les clusters de piRNAs "unistrand" tels que *flamenco* se comporteraient comme des gènes. **Alors comment la cellule distingue-t-elle les transcrits émanant de clusters de piRNAs "unistrand" devant être maturés en piRNAs des transcrits provenant de gènes devant être traduits ? Quels sont les mécanismes permettant la reconnaissance des transcrits provenant de clusters de piRNAs "unistrand" pour être maturés en piRNAs ?** Pour expliquer ce phénomène, mon hypothèse est la suivante : il y aurait, au sein de *flamenco*, une séquence ou une structure ARN qui serait reconnue par l'une des protéines de la voie des piRNAs permettant ainsi la maturation des transcrits de *flamenco* en piRNAs. Une étude menée par notre équipe et en collaboration avec Ramesh Pilai a montré qu'une production de piRNAs artificiels provenant des séquences placées en aval de la région transcrite de *flamenco* (+1 à +718 nts par rapport à son site d'initiation de la transcription) était observée lors de transfections transitoires au sein des cellules OSS. Lorsque cette région de *flamenco* est placée entre deux gènes rapporteurs, elle permet la production de piRNAs provenant de la séquence placée en aval de *flamenco*, mais pas de la séquence placée en amont. Ceci suggère que cette région de *flamenco* fonctionne en position interne et de façon unidirectionnelle. **Cette région de *flamenco* serait-elle la clé de la sélection et de la maturation des transcrits de *flamenco* en piRNAs ? Quelle est la particularité de cette région ? Est-elle également responsable de l'adressage des transcrits de *flamenco* *in vivo* ? Quels sont les facteurs impliqués dans la reconnaissance de cette région de *flamenco* ? Cette sélection est-elle effectuée dès la transcription des clusters de piRNAs ou une fois que ces derniers sont épissés ? La sélection a-t-elle lieu au niveau du noyau ou du cytoplasme ? Les éléments nécessaires à la sélection et à la maturation d'un transcrit issu du cluster de piRNAs *flamenco* sont-ils cellule-spécifiques ?** Pour le savoir, j'ai généré divers transgènes possédant différentes parties de la région de *flamenco* en amont du gène rapporteur *tomato*, codant une protéine fluorescente. L'ensemble des transgènes ont été placés sous le contrôle du



promoteur endogène de *flamenco*. Les transgènes ont ensuite été injectés en mouches pour permettre leur insertion en une seule copie à un site d'insertion précis (sur la bande chromosomique 53B2 du chromosome 2). Des approches de séquençage, de biologie moléculaire, d'imagerie confocale et de bio-informatique ont été menées sur les différentes lignées transgéniques ainsi créées afin de répondre à ces diverses questions.

Les piRNAs issus de la maturation des transcrits des clusters de piRNAs sont par la suite pris en charge par les protéines PIWI. Les complexes piRNA/PIWI vont reconnaître, par complémentarité de séquence, les transcrits des ETs actifs à réprimer. Néanmoins, de plus en plus d'études ont démontré que les complexes piRNA/PIWI avaient également la capacité de réguler des gènes chez certains animaux. Contrairement aux ETs, la reconnaissance des transcrits de gènes par les complexes piRNA/PIWI est basée sur un appariement incomplet. Sachant que les cellules folliculaires jouent un rôle important lors de la formation de la gonade et de sa physiologie chez les drosophiles femelles, **est-ce que les piRNAs produits à partir de clusters de piRNAs "unistrand" présentent la capacité de réguler certains gènes impliqués dans l'un de ces deux processus ? Cette régulation est-elle effectuée lors des fenêtres temporelles précises ? Quels sont les critères de reconnaissance des transcrits de gènes devant être régulés ? Les acteurs impliqués dans la régulation génique menée par les piRNAs sont-ils différents de ceux impliqués dans la régulation de l'expression des ETs ? Quels sont les mécanismes mis en jeu lors de la régulation des gènes par les piRNAs ?** Afin de trouver des réponses à ces questions, une étude très préliminaire consistant à identifier les gènes potentiellement régulés par les piRNAs provenant du cluster de piRNAs *flamenco* a été initiée durant ma thèse.

Bien que la voie des piRNAs soit relativement bien décrite chez la drosophile, la biogenèse des piRNAs n'est pas encore totalement comprise. L'ensemble de ces axes de recherches *in vivo* permettront ainsi d'avancer dans la compréhension de la biologie des clusters de piRNAs "unistrand", mais aussi des mécanismes moléculaires mis en jeu lors de la biogenèse des piRNAs en utilisant la drosophile comme modèle d'étude.





*Discussion*  
*générale*



## ***Discussion générale***

### **Défense de l'information génétique de la lignée germinale au cours du développement de la drosophile**

Chez la drosophile, l'information génétique qui sera transmise à la descendance est renfermée dans les cellules germinales primordiales, premières cellules à se différencier dans l'embryon. Ces cellules sont les premières cellules à s'individualiser au pôle postérieur de l'embryon dès le stade 4, c'est-à-dire juste avant la transition maternozygotique qui a lieu au stade 5. Les cellules germinales primordiales vont par la suite migrer dans l'embryon. À partir du stade 12, ces cellules vont migrer dans le mésoderme, puis vont être attirées vers un sous-ensemble de cellules mésodermiques qui deviendront les futures cellules somatiques des gonades. Au stade 13, les cellules germinales primordiales sont entourées par les précurseurs des cellules somatiques de gonades. Tout comme la drosophile, les cellules germinales primordiales de nombreux animaux, tels que les mammifères, vont migrer dans les embryons, puis elles vont rentrer en contact avec les futures cellules somatiques des gonades pour former les futures gonades.

L'information génétique durant les stades précoces du développement de la drosophile sera protégée de la mobilisation massive d'ETs grâce aux complexes piRNA/Aub et piRNA/Piwi déposés maternellement dans l'embryon (Akkouche et al., 2017; Le Thomas et al., 2014). En effet, lors de l'initiation de l'expression du génome zygotique au stade 5, les clusters de piRNAs, tout comme les ETs, vont être exprimés. Les complexes piRNA/PIWI déposés maternellement permettraient alors la reconnaissance des transcrits de clusters de piRNAs « dual-strand », mais également des transcrits d'ETs actifs afin d'initier le cycle ping-pong. La protection de l'information génétique de la lignée germinale est donc assurée en continu grâce à la transmission des piRNAs de la mère à l'embryon. Il a été proposé que les protéines Aub associées à des piRNAs complémentaires généreraient la première coupure des transcrits d'ETs formant ainsi les premiers piRNAs zygotiques. Cette coupure initie alors la maturation du reste des transcrits par les protéines Piwi et Zucchini *via* le mécanisme de Phasing. En d'autres termes, l'activation de la voie des piRNAs germinale aurait lieu très précocement dans les cellules germinales primordiales. Cependant, l'activation de la voie des



piRNAs au sein des cellules somatiques de soutien protégeant la lignée germinale au sein des gonades est un domaine qui n'a pas été encore exploré.

L'activation de la voie des piRNAs dans les cellules somatiques des gonades résulte de i) l'expression de clusters de piRNAs "unistrand", dont *flamenco* est le principal, mais également de ii) l'expression des protéines de la voie des piRNAs. Durant ma thèse, j'ai donc étudié, avec l'aide de Benjamin BERTIN, le profil d'expression d'un cluster de piRNAs "unistrand" au cours du développement de la drosophile en prenant comme modèle le cluster de piRNAs *flamenco*. Le cluster de piRNAs *flamenco* est localisé dans la région péricentromérique du chromosome X et s'étend sur plus de 500 kb selon la release R6. Bien que localisé dans une région hétérochromatique, l'étude des marques chromatiniennes au niveau de ce cluster de piRNAs a révélé qu'il n'était pas enrichi en marques histones répressives H3K9me3 (Mohn et al., 2014). Le promoteur de *flamenco* est un promoteur de type INR DPE dépendant de l'ARN polymérase II (Goriaux et al., 2014; Sienski et al., 2012). L'activation de l'expression de ce cluster dépend du facteur de transcription Cubitus Interruptus (Ci) (Goriaux et al., 2014). Dans les ovaires de drosophiles, il a été montré que le cluster de piRNAs *flamenco* est exprimé exclusivement dans les cellules folliculaires (Dennis et al., 2016, 2013). Néanmoins, l'initiation de l'expression du cluster de piRNAs *flamenco* dans les cellules somatiques des ovaires de drosophiles, et par conséquent, l'initiation de la voie des piRNAs dans ces mêmes cellules au cours du développement était encore inconnue à ce jour.

#### ***A. Le piRNA cluster flamenco est exprimé très précocement du développement de la drosophile***

Grâce au transgène senseur de l'expression de *flamenco* que j'ai développé, nous avons démontré que le cluster de piRNAs *flamenco* s'exprimait précocement dans les cellules somatiques des gonades embryonnaires femelles, mais aussi des gonades embryonnaires mâles. En effet, le cluster de piRNAs *flamenco* commence à être exprimé dans les précurseurs des cellules somatiques des gonades dès qu'elles interagissent avec les cellules germinales primordiales au stade embryonnaire 13. Après la coalescence des cellules gonadiques au stade 14, l'expression du cluster *flamenco* semble augmentée dans les cellules somatiques des gonades des deux sexes, ce qui suggère que *flamenco* s'exprime dès que les cellules somatiques acquièrent l'identité de cellules folliculaires précurseurs. En outre, nous montrons également par des expériences d'hybridation *in situ* ARN (FISH) que les transcrits endogènes de *flamenco* sont déjà accumulés en un foyer nucléaire suggérant que le transit nucléaire des



transcrits de *flamenco* est également fonctionnel dès le début de la gonadogenèse. Dans les cellules folliculaires des ovaires de drosophiles adultes, ce site d'accumulation est appelé Dot COM et fait face au site de maturation cytoplasmique, le corps Yb (Dennis et al., 2013). Cette accumulation précoce des transcrits de *flamenco* dans les précurseurs des cellules somatiques des gonades suggère que la voie des piRNAs est active dans les précurseurs de cellules somatiques folliculaires très précocement au cours de l'embryogenèse.

Durant les stades larvaires et le stade adulte des drosophiles mâles et femelles, nous avons démontré que le cluster de piRNAs *flamenco* était uniquement exprimé dans les cellules somatiques des gonades en contact avec les cellules germinales. Ces données suggèrent, une fois de plus, que le contact entre les cellules de la lignée germinale et les cellules somatiques des gonades pourrait être un pré requis à l'expression du cluster de piRNAs *flamenco*, et par conséquent, à l'activation de la voie des piRNAs dans ces cellules somatiques des gonades. La communication entre les cellules gonadiques somatiques et la lignée germinale serait-elle la clé de l'activation de la voie des piRNAs dans les cellules somatiques des gonades en contact avec les cellules germinales ?

Au cours de l'embryogenèse, les cellules germinales primordiales migrent tout d'abord à travers l'endoderme postérieur, se déplacent ensuite vers la surface dorsale de l'endoderme, pour finalement migrer dans le mésoderme adjacent. Une fois localisées dans le mésoderme, cellules germinales primordiales sont attirées vers un sous-ensemble de cellules mésodermiques (les précurseurs des cellules somatiques de la gonade) pour former la gonade embryonnaire. L'attraction des cellules germinales primordiales est induite par l'expression de l'Hydroxyméthylglutaryl-CoA réductase (HMGCR) dans les précurseurs des cellules somatiques des gonades (Van Doren et al., 1998). Cette enzyme est initialement exprimée dans le mésoderme, puis son expression est restreinte à des blocs de cellules mésodermiques localisés dans les parasegments 10 à 12. Ces blocs de cellules correspondent aux précurseurs des cellules somatiques des gonades. Vers le stade 13, ces cellules vont se joindre pour former une bande continue de cellules finement attachées aux cellules germinales primordiales. Dans les embryons mutants *hmgcr*, la plupart des cellules germinales primordiales localisées dans le mésoderme se dispersent dans l'embryon au lieu de s'attacher aux précurseurs des cellules somatiques des gonades. De plus, l'expression ectopique du gène *hmgcr* dans les cellules de l'épiderme et des cellules du système nerveux est suffisante pour attirer les cellules germinales primordiales vers des tissus vers lesquels elles ne migreraient normalement pas (Van Doren et al., 1998). Il serait alors intéressant de savoir si, dans des embryons mutants *hmgcr*,





l'expression de *flamenco* est activée ou non dans les précurseurs des cellules somatiques des gonades en absence de contact avec les cellules germinales primordiales. Des expériences complémentaires pourront être réalisées afin de savoir si l'expression de *flamenco* est initiée dans les cellules qui attirent artificiellement les cellules germinales en exprimant ectopiquement le gène *hmgcr* dans les cellules de l'épiderme et des cellules du système nerveux par exemple.

### ***B. flamenco, cible du facteur de transcription Traffic Jam***

L'un des potentiels acteurs impliqués dans l'expression de *flamenco* pourrait être le facteur de transcription de type Maf codé par le gène *traffic jam (tj)*. Ce facteur de transcription est connu pour être impliqué dans la migration de cellules somatiques dites "border cells" dans ovaires de drosophiles (Gunawan et al., 2013), mais aussi dans la morphogenèse des gonades mâles et femelles de drosophiles (Li et al., 2003). De plus, le profil d'expression du gène *tj* correspond parfaitement au profil d'expression de *flamenco*. Tout comme *flamenco*, le gène *tj* s'exprime dès le stade embryonnaire 13 dans les cellules somatiques des futures gonades mâles et femelles. Son expression est restreinte par la suite aux cellules somatiques directement en contact avec les cellules souches germinales et les cellules germinales différenciées au cours de la gonadogenèse et de la gamétogénèse (Li et al., 2003). De façon intéressante, l'expression des facteurs de transcriptions de type Maf dans les gonades des mammifères est également restreinte aux cellules somatiques en contact avec les cellules de la lignée germinale (DeFalco and Capel, 2009). L'analyse bio-informatique des résultats de ChIP Seq réalisés sur cellules OSS et utilisant un anticorps dirigé contre le facteur de transcription Tj a révélé des sites de fixation de Tj dans la région promotrice de *flamenco*. Ces informations suggéreraient donc que l'expression de *flamenco* pourrait être directement induite par ce facteur de transcription.

Ces résultats sont appuyés par des expériences *in vivo* d'inactivation de l'expression de Tj spécifiquement dans les cellules somatiques folliculaires. En effet, l'expression d'une construction RNAi ciblant les transcrits du gène *tj* dans les cellules somatiques des ovaires de drosophiles a eu pour conséquence d'induire la formation d'ovaires rudimentaires similaires à ceux obtenus lors de la mutation du cluster de piRNAs *flamenco*. La présence d'ovaires rudimentaires dans un mutant *tj* pourrait peut-être être la conséquence d'une expression altérée du cluster de piRNAs *flamenco* entraînant ainsi la dérégulation de certains ETs. En effet, nous avons montré que l'altération de l'expression du gène *tj* était responsable de la



dérépression d'un des éléments transposables connus pour être régulés par le cluster de piRNAs *flamenco*, l'élément transposable ZAM. Des expériences d'expression ectopique du gène *tj* pourraient également être conduites pour savoir si l'expression du facteur de transcription Tj est suffisante pour induire l'expression du cluster de piRNAs *flamenco* dans des cellules connues pour ne pas exprimer *flamenco* telles que les cellules germinales. De façon intrigante, le facteur de transcription Tj est connu pour activer l'expression de la protéine Piwi (Saito et al., 2009). Par ailleurs, des expériences d'hybridation *in situ* ont révélé que d'autres gènes codant des protéines de la voie des piRNAs telles que la protéine Armitage sont également exprimés dans les gonades d'embryons au stade 13. Nait alors l'idée que Tj pourrait être le Facteur de Transcription "chef d'orchestre" qui initierait l'activation de la voie des piRNAs dans les cellules somatiques adjacentes des cellules germinales.

### ***C. La voie de signalisation Hedgehog pourrait-elle induire l'activation de l'expression du gène *tj* et de *flamenco* dans les cellules adjacentes aux cellules germinales ?***

Sachant que l'expression du gène *tj* et du cluster de piRNAs *flamenco* commencent lors de la mise en contact des cellules germinales avec les précurseurs des cellules somatiques des gonades et que leur expression est plus importante après la coalescence des gonades chez l'embryon, il serait également envisageable que l'activation de leur expression soit due à un facteur de transcription commun produit à la suite d'une communication entre les cellules des gonades. Des études ont montré que l'enzyme HMGCR, requise pour l'attraction des cellules germinales primordiales aux précurseurs des cellules somatiques des gonades, était impliquée dans la voie de signalisation Hedgehog (Deshpande et al., 2001). Ils ont également démontré que la migration de la lignée germinale vers les précurseurs des cellules somatiques des gonades et la coalescence dans les gonades embryonnaires étaient extrêmement sensibles à l'activité de la voie de signalisation Hedgehog. De façon intéressante, il a été montré que l'activateur transcriptionnel direct des gènes cibles de la voie Hedgehog est le facteur de transcription Ci. Or, le facteur de transcription Ci a précédemment été décrit comme permettant l'expression du cluster de piRNAs *flamenco* dans les cellules en culture dérivées des cellules souches folliculaires (Goriaux et al., 2014). De plus, nos analyses *in silico* ont confirmé la présence d'un site de liaison de Ci autour du promoteur de *flamenco*. Enfin, une étude récente a montré que Tj serait activé par le facteur de transcription Ci en réponse à la voie de signalisation Hedgehog (Lai et al., 2017). Il serait donc crucial de savoir si



l'expression du *flamenco*, mais aussi celle du gène *tj*, dépendent du facteur de transcription Ci, et par conséquent, de savoir si elles dépendent de la voie de signalisation Hedgehog.

#### ***D. Régulation de l'expression du cluster de piRNAs flamenco et du gène tj par une autre voie de signalisation ?***

La voie de signalisation Notch est une voie conservée chez les animaux qui est impliquée dans de nombreux processus développementaux basés sur la communication intercellulaire. L'activation de la voie Notch repose sur deux éléments : le récepteur transmembranaire Notch et son ligand Delta (Bray, 2006; Kopan and Ilagan, 2009). Au sein des ovaires de drosophiles, il a été montré que les cellules germinales produisaient le ligand Delta et que le récepteur Notch était présent sur les membranes de toutes les cellules somatiques composant l'ovaire (Xu and Gridley, 2012). Pour toutes les cellules somatiques des ovaires en contact avec les cellules germinales, les ligands Delta sont fixés sur leurs récepteurs Notch engendrant par la suite des clivages protéolytiques des récepteurs. A la suite de ces clivages, la partie cytoplasmique du récepteur de Notch est libérée de la membrane et est importée au noyau. Le domaine intracellulaire de Notch interagit avec un complexe de trois protéines ayant la capacité de se lier à l'ADN (CBF-1 / Su (H) / LAG-1) ainsi qu'avec le coactivateur transcriptionnel Mastermind pour activer l'expression de gènes cibles. En d'autres termes, les cellules somatiques des ovaires ayant un contact avec les cellules germinales présentent une activation de la voie Notch importante par rapport aux cellules somatiques interfolliculaires. De façon intéressante, le patron d'activation de la voie Notch dans les ovaires de drosophiles (Assa-Kunik et al., 2007; Xu and Gridley, 2012) correspond au profil d'expression de *flamenco* et du gène *tj* suggérant que la voie Notch pourrait être impliquée dans la mise en place de l'activation de l'expression de *flamenco* et peut-être de la voie des piRNAs dans les cellules somatiques folliculaires. De plus, le récepteur Notch est exprimé dans les précurseurs des cellules somatiques des gonades à partir du stade 12, c'est-à-dire juste avant que ces cellules attirent les cellules germinales primordiales (Kitadate and Kobayashi, 2010). Pour savoir si l'activation de la voie Notch au sein des cellules somatiques des gonades de drosophiles est responsable de l'expression de *flamenco*, le profil d'expression de *flamenco* au cours du développement sera étudié en utilisant des lignées exprimant dans les cellules folliculaires une construction RNAi ciblant les transcrits du gène *Notch* ou en utilisant des lignées de drosophiles mutantes pour ce gène. Des expériences d'activation constitutive de la voie Notch au sein des cellules interfolliculaires pourrait être réalisées dans le but de savoir



si la voie Notch est suffisante pour induire l'expression du cluster de piRNAs *flamenco* dans des cellules connues pour ne pas exprimer le locus *flamenco* telles que les cellules "stalks".

La description du profil d'expression de *flamenco* et ainsi que les analyses portant sur la régulation de son expression ont ouvert de nouvelles pistes pour les futures études visant à comprendre l'établissement de la voie piRNAs dans les gonades de drosophiles.

### **Le devenir des transcrits des clusters de piRNAs, un choix cellulaire ?**

Chez les animaux, la transcription des clusters de piRNAs dépend de l'ARN polymérase II tout comme les gènes. Contrairement aux transcrits de gènes, les transcrits issus des clusters de piRNAs sont des longs ARNs non codants sélectionnés pour être maturés en petits ARNs de 23 à 29 nts, les piRNAs. L'une des questions les plus énigmatiques concernant la biogenèse des piRNAs, qui reste à ce jour encore en suspens, est de savoir comment des transcrits provenant des clusters de piRNAs sont sélectionnés pour être maturés en piRNAs.

La reconnaissance des transcrits provenant des clusters de piRNAs "dual-strand" par les cellules germinales chez la drosophile serait liée au fait qu'ils soient transcrits de façon non canonique. En effet, ils ne sont ni épissés ni coiffés et la majorité d'entre eux ne seraient pas polyadénylés (Brennecke et al., 2007; Chen et al., 2016; Malone et al., 2009; Mohn et al., 2014; Sienski et al., 2012). De plus, il a été montré que la maturation des transcrits artificiels en piRNAs dans les cellules germinales d'ovaires de drosophiles pouvait être induite par la liaison d'un des composants du nuage sur ces transcrits (Rogers et al., 2017). Il a été proposé que la reconnaissance des transcrits de clusters de piRNAs "dual-strand" serait réalisée dès leur transcription par la protéine Cutoff et UAP56. Une fois reconnus, ces transcrits sont acheminés à la membrane nucléaire par un transit nucléaire impliquant les protéines Bootlegger, Nxf3 et Nxt1 (Elmaghraby et al., 2019). Ce transit nucléaire est différent de celui des transcrits de gènes. L'exportine Crm1 permet ensuite leur transfert au cytoplasme où ils seront pris en charge par la protéine Vasa dont le rôle serait de les amener au nuage pour y être maturés en piRNAs.

Contrairement aux transcrits des piRNAs clusters "dual-strand" présents uniquement chez la drosophile, les transcrits des clusters de piRNAs "unistrand" présents chez de nombreux animaux tels que les mammifères et chez la drosophile présentent les





caractéristiques d'être coiffés et polyadénylés tout comme les ARNs messagers. Notre équipe a démontré que les transcrits de *flamenco* sont alternativement épissés et accumulés en un foyer nucléaire appelé le Dot COM (Dennis et al., 2016, 2013; Goriaux et al., 2014). Le transit nucléaire de *flamenco* requière les exportines Nxt1 et Nfx1, protéines également impliquées dans l'export des ARNs messagers. Alors comment les cellules folliculaires distinguent-elles les transcrits de *flamenco* devant être maturés en piRNAs des transcrits provenant de gènes devant être traduits ? Et quels sont les mécanismes permettant la reconnaissance des transcrits provenant de piRNA cluster "unistrand" pour être maturés en piRNAs ? Chez les eucaryotes, la distinction entre ces deux types d'ARNs semble complexe et fait actuellement l'objet de nombreuses recherches.

### ***A. Identification in vivo de la séquence responsable de la maturation des transcrits de flamenco***

La première stratégie développée par certaines équipes pour comprendre comment les transcrits du cluster de piRNAs *flamenco* sont sélectionnés consistait à analyser l'interaction des différentes protéines de la voie des piRNAs sur les transcrits de *flamenco* via des expériences de CLIP (Cross-Linking and Immuno-Precipitation) (Ishizu et al., 2015). Les expériences de Yb-CLIP réalisées sur les cellules OSS en culture ont notamment permis de révéler que la protéine Yb était recrutée tout au long des transcrits de *flamenco*, mais aucune région spécifique des transcrits de *flamenco* n'a pu être identifiée comme étant responsable de sa prise en charge par la voie des piRNAs.

La seconde stratégie développée par notre équipe consistait à identifier la séquence responsable de la maturation des transcrits de *flamenco* en piRNAs. Des expériences de transfections transitoires au sein des cellules OSS en culture réalisées par notre équipe en collaboration avec Ramesh Pilai ont démontré que les 718 premiers nucléotides transcrits de *flamenco* permettaient la maturation de la séquence rapportrice, placée en aval, en piRNAs (Homolka et al., 2015). Néanmoins, les approches de séquençage, de biologie moléculaire, d'imagerie confocale et de bio-informatique m'ont permis de démontrer que, *in vivo*, ni le premier exon, ni le premier intron du locus *flamenco*, et par déduction, les 718 premiers nucléotides de *flamenco* ne suffiraient à initier l'adressage des transcrits chimériques à la voie des piRNAs, ou du moins, elles ne permettraient qu'un très faible adressage des transcrits à leur site de maturation. Une des difficultés de cette analyse réside dans le fait que la séquence responsable de l'adressage des transcrits pourrait peut-être être présente dans les premiers



nucléotides des transcrits de *flamenco*, mais que cette séquence n'est pas la capacité d'induire une forte expression de *flamenco*. La faible expression des transgènes possédant la région unique transcrite de *flamenco* pourrait être expliquée par l'absence de séquences d'ETs. Or, les ETs ont la particularité de posséder des sites de liaison pour certains facteurs de transcription. En effet, l'ajout d'une séquence de *flamenco* contenant des ETs en amont de la séquence rapportrice *tomato* a permis d'améliorer l'expression des transgènes suggérant que l'expression de *flamenco* dépendrait des facteurs de transcription présents dans les séquences d'ETs composant ce locus. La présence d'une séquence issue du cluster de piRNAs *flamenco* dans les transcrits chimériques n'a cependant pas empêché leur traduction. Les transcrits des transgènes étant préférentiellement traduits en protéines, j'ai donc recherché les facteurs requis pour leur adressage à la voie de maturation des piRNAs. Les paramètres pouvant influencer l'adressage des transcrits chimériques à la voie de maturation des piRNAs testés dans cette étude sont la stabilité des transcrits, l'environnement génomique ainsi que la longueur du fragment de *flamenco* placé en amont du gène rapporteur. Bien que l'augmentation de la longueur du fragment de *flamenco* placé en amont du gène rapporteur ait permis d'améliorer la prise en charge des transcrits artificiels par la voie de maturation en piRNAs, la quantité de piRNAs homologues à la séquence *tomato* détectés dans les ovaires des lignées transgéniques reste relativement faible. La faible détection de piRNAs homologues à la séquence *tomato* produits dans les cellules folliculaires pourrait cependant être liée à la trop grande quantité de piRNAs produits dans la lignée germinale et recueillis dans les échantillons séquencés issus d'ovaires entiers.

L'utilisation de transgènes pour identifier la séquence responsable de l'adressage des transcrits de *flamenco* dans les cellules folliculaires *in vivo* n'est probablement pas l'outil optimal. L'utilisation de la technologie de CRISPER-Cas 9 est actuellement l'approche qui semble la plus adaptée pour identifier la ou les régions nécessaires à la reconnaissance des transcrits de *flamenco* (Adli, 2018; Pickar-Oliver and Gersbach, 2019). Les futures expériences de CRISPER-Cas9 nous permettront de supprimer précisément différentes régions du locus endogène *flamenco*, et par conséquent, d'identifier la ou les séquences requises pour que les transcrits de *flamenco* soient adressés à leur site de maturation dans les cellules folliculaires. Une analyse détaillée de ces séquences permettra à long terme d'identifier les facteurs nécessaires pour la sélection et la maturation des transcrits de *flamenco* en piRNAs, et par conséquent, de progresser dans la compréhension de la biogenèse des piRNAs.



## ***B. Amélioration de la résolution des résultats d'une étude in vivo***

Lors de la comparaison de l'alignement des piRNAs provenant des cellules OSS en culture et des ovaires sur les 6334 premières bases de *flamenco*, peu de piRNAs produits à partir de cette région ont été détectés dans les ovaires entiers par rapport aux cellules OSS. Le séquençage des petits ARNs liés à la protéine Piwi permettrait de s'affranchir des piRNAs germinaux produits lors du cycle ping-pong, et ainsi, d'améliorer la détection des piRNAs produits dans les cellules folliculaires. Le séquençage des petits ARNs liés à la protéine Piwi à partir d'ovaires entiers a permis la détection de piRNAs artificiels produits à partir des transcrits d'un transgène exprimé spécifiquement dans les cellules folliculaires dans trois études récentes (Muerdter et al., 2012; Pandey et al., 2017; Rogers et al., 2017). En d'autres termes, la détection des éventuels piRNAs homologues de la séquence *tomato* dans nos lignées transgéniques pourrait être améliorée en séquençant uniquement les petits ARNs liés à la protéine Piwi. Néanmoins, la protéine Piwi est une protéine de la voie des piRNAs exprimée dans les deux types cellulaires qui composent l'ovaire : les cellules somatiques de l'ovaire et les cellules germinales. Pour aller plus loin dans l'optimisation de la résolution de ces expériences *in vivo*, des expériences de FACS pourraient être envisagées afin de trier les cellules folliculaires (de la Cruz and Edgar, 2008; Tuchin et al., 2011). Afin de s'affranchir de leur force d'expression, les transgènes développés dans cette étude pourraient être couplés à un transgène UASp GFP piloté par le driver Traffic jam (driver spécifique des cellules folliculaires) pour que le tri des cellules folliculaires soit réalisé sur la fluorescence de la protéine GFP.

## ***C. Distinction entre sélection et maturation des transcrits de cluster de piRNAs***

La biogenèse de piRNAs résulte de nombreuses étapes successives : la transcription du cluster de piRNAs, le transit nucléaire de ses transcrits, l'export cytoplasmique des transcrits, l'adressage des transcrits à la voie de maturation des piRNAs et enfin leur maturation en piRNAs. Récemment, deux équipes ont développé *in vivo* un système artificiel permettant de séparer l'étape de sélection des transcrits devant être maturés en piRNAs de l'étape de maturation de ces derniers. Ce système consiste à recruter artificiellement une protéine de la voie des piRNAs sur n'importe quel transcrit afin de savoir si son recrutement sur un transcrit est suffisant à initier la maturation de ce dernier en piRNAs. Ils ont ainsi démontré que n'importe quelle séquence qui était artificiellement adressée à la voie des piRNAs était maturée en piRNAs. En d'autres termes, l'adressage artificiel d'un transcrit au site de



maturation suffit à le maturer en petits ARNs de 23 à 29 nts. Cependant, la reconnaissance des transcrits devant être maturés en piRNAs reste encore inconnue.

Néanmoins, avant leurs exports cytoplasmiques, les transcrits de clusters de piRNAs "unistrand" tels que *flamenco*, sont accumulés en un foyer nucléaire au sein des cellules folliculaires des ovaires de drosophiles (Dennis et al., 2016, 2013). Ce site nucléaire d'accumulation des transcrits de clusters de piRNAs "unistrand", le Dot COM, fait directement face au corps Yb, structure multiprotéique responsable de la maturation des transcrits en piRNAs. En d'autres termes, le flux de transcrits de clusters de piRNAs "unistrand" exportés est concentré sur une faible zone de la membrane nucléaire. Les transcrits exportés au cytoplasme sont alors proches du corps Yb permettant alors leurs prises en charge par la voie de maturation des piRNAs. Par ailleurs, la mutation des protéines impliquées dans la formation du corps Yb telles que Armitage et Zucchini n'affecte pas l'accumulation des transcrits de *flamenco* dans le Dot COM (Dennis et al., 2013). Les exportines Nxt1 et Nxf1 ainsi que les protéines du complexe exon-jonction et la protéine UAP56 sont requises pour le transit nucléaire des transcrits du locus *flamenco*. L'absence de l'expression de *flamenco* dans les cellules folliculaires d'ovaires des lignées BG ainsi que le blocage de l'export des transcrits de *flamenco* altèrent la formation du corps Yb (Dennis et al., 2016). L'accumulation des transcrits serait alors une étape intermédiaire nécessaire entre la sélection des transcrits de *flamenco* et leur maturation en piRNAs. Sachant que les transcrits des transgènes développés dans cette étude sont majoritairement traduits en protéines, des expériences de double RNA FISH nous permettraient de savoir si ces transcrits colocalisent ou non avec les transcrits endogènes de *flamenco*. L'absence de leur accumulation dans le Dot COM serait ainsi la preuve que les transcrits des transgènes ne sont pas reconnus comme des transcrits provenant de clusters de piRNAs, d'où leur faible maturation en piRNAs. Néanmoins les critères de sélection permettant l'accumulation des transcrits au Dot COM restent encore à découvrir.

#### ***D. Paramètres permettant la distinction entre les ARNs messagers et les ARNs des clusters de piRNAs "unistrand"***

L'analyse des critères permettant à la cellule de distinguer les ARNs messagers des longs ARNs non codants, tels que les ARNs précurseurs provenant de la transcription des clusters "unistrand", chez les eucaryotes est un axe de recherche relativement complexe. Les ARNs messagers sont en effet constitués de régions codantes bornées par deux régions transcrites

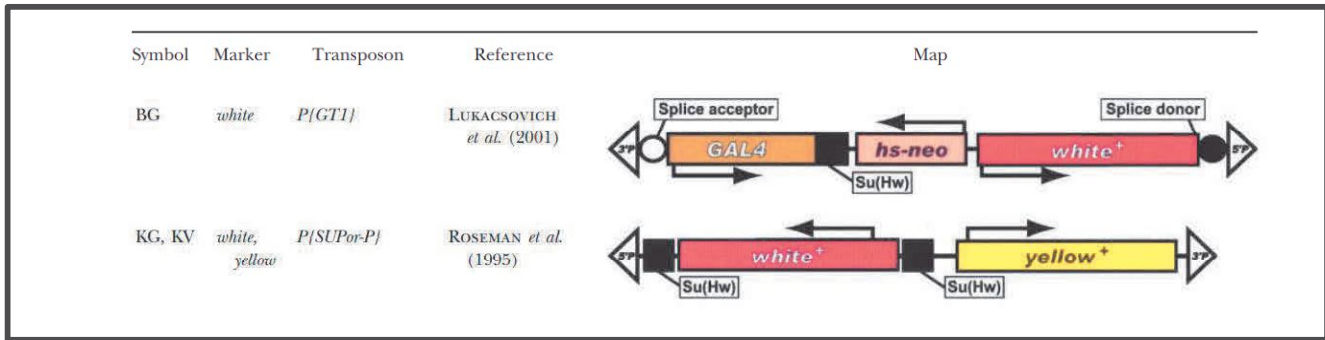




mais non traduites, les régions UTRs. Sachant que les transgènes sont composés d'une région codante (la séquence *tomato*) suivie de la région 3'UTR du gène K'10 de drosophiles, il se pourrait alors que les 2086 premières bases du transcrit de *flamenco* additionnées en aval de la séquence *tomato* soient considérées comme une région 5'UTR par les cellules folliculaires. Chez les eucaryotes, les longs ARNs non codants ont été définis comme des transcrits de la polymérase II dont la taille des unités de transcription est supérieure à 5kb (Sun et Kraus, 2014). C'est notamment le cas des transcrits des clusters de piRNAs tels que *flamenco* chez la drosophile. Bien que la résolution des résultats de cette étude ne soit pas optimale, l'insertion d'une plus grande séquence de *flamenco* (~ 6kb) en amont de la séquence *tomato* a permis d'améliorer l'adressage des transcrits à la voie de maturation des piRNAs. Ces données suggèrent que le choix de l'adressage des transcrits dépendrait alors de la longueur totale du transcrit.

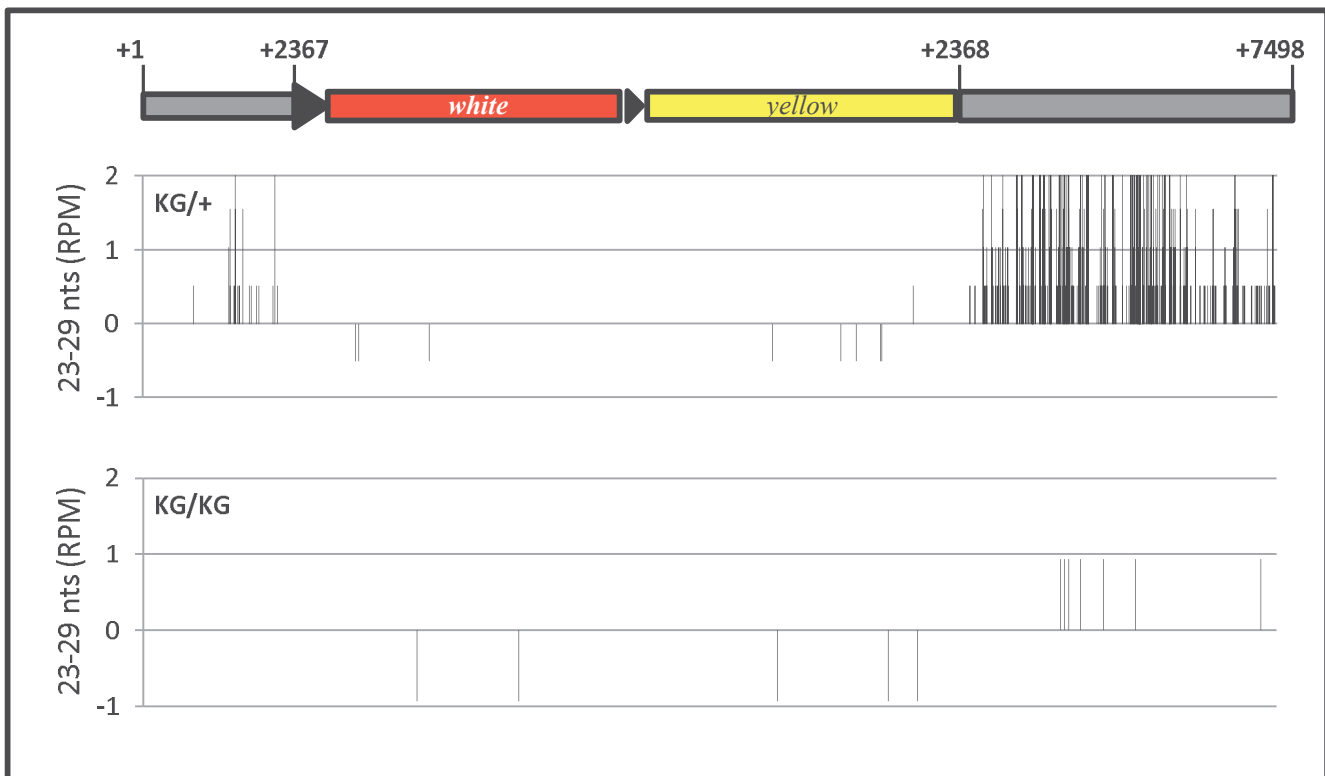
Chez les eucaryotes, la région transcrite non traduite (UTRs) localisée dans la partie 3' d'un transcrit possède notamment un signal de polyadénylation (polyA) plus ou moins fort dont le rôle est essentiel pour sa stabilité, son transit intracellulaire et donc sa traduction (Shatkin and Manley, 2000). Les données obtenues lors de cette étude laissent penser que la stabilité conférée aux transcrits *via* différents signaux polyA pourrait potentiellement avoir un impact sur la quantité de transcrits adressés à la voie des piRNAs. La génération de transcrits chimériques dépourvus de signal de polyA permettrait ainsi d'explorer l'impact de la polyadénylation des transcrits de clusters de piRNAs "unistrand" tels que *flamenco* sur leur sélection. Bien que les transcrits des clusters de piRNAs "unistrand" soient décrits comme polyadénylés, l'influence et l'importance de la polyadénylation des transcrits précurseurs de *flamenco* sur leur sélection restent encore inconnues (Chen et al., 2016). En effet, tout comme pour les transcrits de clusters de piRNAs "unistrand", le rôle régulateur des signaux de polyA sur les longs transcrits non codants des eucaryotes est encore inconnu à ce jour.

L'adressage préférentiel des transcrits chimériques à la machinerie traductionnelle pourrait être lié au fait que les transgènes possèdent une séquence codante, la séquence *tomato* débutant par un codon d'initiation de la traduction. Or, par définition, les longs ARNs non codants se distinguent des ARNs messagers de par le fait qu'ils ne possèdent pas de séquence codante. De façon intéressante, la lignée KG mutante pour *flamenco* est le résultat de l'insertion d'un transgène à la fin de l'exon 2 du locus *flamenco* (Mével-Ninio et al., 2007). Le transgène inséré est composé de la séquence du gène *white* (orientée en anti-sens par rapport au sens de transcription de *flamenco*) suivie de la séquence du gène *yellow* (par rapport au



**Figure 59 : Structure des transgènes présents dans les lignée BG et KG.**

Le transgène BG est composé du gène GAL4 (en marron) fusionné à la séquence « *suppressor of Hairy-wing* » (Su(Hw)) (carré noir), du gène de résistance à la néomycine (en rose) et du gène *white* (en rouge). Le transgène KG est quant à lui composé du gène *white* (en rouge) encadré par deux séquences « *suppressor of Hairy-wing* » (Su(Hw)) (carrés noirs) et du gène *yellow* (en jaune). Les pieds de P sont représentés par des triangles blancs. Figure adaptée du tableau 1 de l'article Bellen et al., 2004.



**Figure 60 : Diminution drastique de la production de piRNAs provenant de *flamenco* à la suite de l'insertion du transgène KG.**

Alignement des petits ARNs de 23 à 29 nts uniques issus des ovaires de drosophile KG/+ et des ovaires de drosophile KG/KG sur l'allèle *KG00476*. Les alignements ont été normalisés en reads per millions (RPM). Seules les positions 5' des piRNAs ont été représentées.

Les séquences *flamenco* (+1 à +2367 et de +2368 à +7498 par rapport au promoteur de *flamenco*) sont représentées en gris, celle du gène *white* en rouge et celle du gène *yellow* en jaune. Les 2 séquences « *suppressor of Hairy-wing* » sont représentées par des triangles noirs.

Le transgène KG s'est inséré à la position X:21,634,258 (+2367).

sens de transcription de *flamenco*) (Figure 59). Chaque séquence codante possède son propre promoteur mimant ainsi l'insertion d'ETs au sein de *flamenco* (Bellen et al., 2004). En rédigeant ce manuscrit, je me suis donc demandé si l'insertion du transgène KG affectait uniquement la maturation des transcrits de *flamenco* en aval de son site d'insertion, ou si elle affectait de façon globale la maturation des transcrits de *flamenco*. Ces analyses m'ont permis de montrer que l'insertion du transgène KG dans le cluster de piRNAs *flamenco* entraîne une diminution drastique de la production des piRNAs provenant de *flamenco* en amont et en aval de son site d'insertion (Figure 60). Pourtant, l'insertion de la séquence rapportrice *gfp* orientée en anti-sens dans un locus ectopique de *flamenco* n'a pas altéré la capacité de ce cluster de piRNAs ectopique à produire des piRNAs (Muerdter et al., 2012). Ces données suggèrent donc que l'introduction d'une séquence codante au sein d'un cluster de piRNAs altérerait la prise en charge de ces transcrits par la voie de maturation des piRNAs. En d'autres termes, la présence de la séquence codante *tomato* pourrait être responsable du faible adressage des transcrits chimériques à la voie de maturation des piRNAs. La séquence *tomato* a donc été placée en orientation inverse par rapport au sens de transcription des transgènes afin de tester cette hypothèse.

Une meilleure connaissance des critères permettant à un transcrit d'être sélectionné pour être par la suite maturé en piRNAs offrirait la possibilité de détourner la voie des piRNAs pour réguler transcriptionnellement et post-transcriptionnellement des gènes. Cependant, bien que de nombreuses expériences aient permis de générer des piRNAs artificiels, l'exploration de la fonctionnalité de ces piRNAs n'a encore jamais été testée. Je suis actuellement en train de développer un système permettant de savoir si les piRNAs dérivant de la séquence rapportrice *tomato* sont fonctionnels, c'est-à-dire, de savoir si les piRNAs produits artificiellement à partir d'une séquence rapportrice *tomato* anti-sens auraient la capacité de reconnaître, et par conséquent, de réguler les transcrits dont ils seraient complémentaires.

Si les piRNAs artificiellement produits sont effectivement fonctionnels, ils seraient alors des outils relativement intéressants pour réguler spécifiquement l'expression d'un gène de façon transcriptionnelle et post-transcriptionnelle.



## Les piRNAs provenant de *flamenco* sont-ils aussi des régulateurs géniques ?

Les régions UTR des transcrits de gènes sont des régions clés impliquées dans la régulation de l'expression des gènes, de la stabilité des transcrits, de leur localisation cellulaire. Cependant, de nombreux transcrits de gènes possèdent des fragments d'ETs dans leur région 5'UTR et 3'UTR. Se pose alors la question de savoir si le rôle régulateur des régions UTR ne serait pas lié à la présence de ces fragments d'ETs ? En d'autres termes, la régulation de ces transcrits serait-elle liée à la voie des piRNAs ? En effet, les piRNAs produits dans les cellules des gonades des animaux ont pour but de reconnaître, par complémentarité de séquence, les transcrits des ETs actifs. Il se pourrait donc que les piRNAs produits à partir des clusters de piRNAs permettent donc également de reconnaître les régions UTR des transcrits contenant des fragments d'ETs pour les réguler.

L'un des exemples les plus détaillés de régulation génique effectuée par les piRNAs chez la drosophile est celui de la régulation post-transcriptionnelle des ARN messagers maternels du gène *nanos* au cours de l'embryogenèse précoce. La régulation post-transcriptionnelle des transcrits *nanos* est induite à la suite de la reconnaissance des fragments des ETs *412* et *roo* par les piRNAs liés sur la protéine Aub (Rouget et al., 2010; Zaessinger et al., 2006). L'appariement entre les piRNAs et les fragments des ETs induit soit une dégradation des transcrits dans les parties antérieure et médiane de l'embryon *via* le recrutement du complexe CCR4-NOT et de la protéine Smaug, soit une stabilisation des transcrits *nanos* au pôle postérieur de l'embryon en recrutant la polyA polymérase Wispy (Dufourt et al., 2017; Zaessinger et al., 2006). Les piRNAs émanant des clusters de piRNAs "dual-strand" auraient donc la capacité de réguler à la fois les ETs mais aussi de réguler les gènes. De plus, les transcrits des transgènes développés dans notre équipe pour suivre la répression des ETs dans les cellules folliculaires sont constitués d'une séquence codante la GFP fusionnée à un fragment d'ETs tels que l'élément *Idéfix* ou *ZAM*, deux ETs régulés par le cluster de piRNAs *flamenco* (Duc et al., 2019; Dufourt et al., 2014). Notre équipe a démontré que la présence de ces fragments d'ET était suffisant pour induire une répression du transgène suggérant alors que les clusters de piRNAs somatiques peuvent également être impliqués dans la régulation génique des cellules folliculaires.

L'ensemble de ces données laisseraient donc penser que les piRNAs produits à partir de cluster de piRNAs pourraient également permettre la régulation de gènes au cours de la gonadogenèse et de la gamétogenèse. En effet, des phénotypes d'ovaires rudimentaires, voire



parfois inexistantes, sont induits lors de la mutation du locus *flamenco*, le cluster de piRNAs producteur majeur de piRNAs dans les cellules folliculaires (Mével-Ninio et al., 2007). En d'autres termes, l'altération de la production de piRNAs provenant du locus *flamenco* entraîne, à elle seule, des défauts d'ovaires (Brennecke et al., 2007; Mével-Ninio et al., 2007). Bien que ces phénotypes mutants puissent être le reflet d'une augmentation massive de l'expression des ETs, il serait également envisageable que l'absence de production de piRNAs à partir du locus *flamenco* engendre une dérégulation d'un ou plusieurs gènes ayant pour conséquence l'apparition de phénotype d'ovaires atrophiés. Afin de tester cette hypothèse, j'ai essayé d'identifier les gènes qui pourraient potentiellement être reconnus, et donc régulés, par les piRNAs produits à partir du cluster de piRNAs *flamenco* utilisant des approches bio-informatique et de biologie moléculaire.

Les résultats préliminaires ont permis de découvrir que le gène *Zydeco* serait une cible potentielle des piRNAs produits par *flamenco* puisque la diminution de piRNAs provenant de *flamenco* induit une diminution du niveau d'expression de ce gène. Ces résultats suggèrent que les piRNAs de *flamenco* stabiliseraient alors les transcrits du gène *zydeco* dans les ovaires de drosophiles. Néanmoins, la recherche d'informations sur ce gène ne mentionne aucune fonction liée à l'ovogenèse. En effet, le gène *Zydeco* est décrit comme étant un échangeur membranaire de cations sodium/calcium dépendant du potassium impliqué dans la dépolarisation des membranes des cellules gliales (Freeman, 2015). Alors pourquoi les piRNAs de *flamenco* stabiliseraient-ils les transcrits de ce gène dans les ovaires de drosophiles ? La protéine *Zydeco* est-elle exprimée à la membrane des cellules folliculaires des ovaires de drosophiles ? Son expression est-elle restreinte aux cellules préalablement décrites comme exprimant *flamenco* ? Le gène *zydeco* est-il différentiellement exprimé au cours de l'ovogenèse ? La protéine *Zydeco* participe-t-elle à la physiologie de l'ovaire ? Une analyse détaillée de ce gène au sein des ovaires de drosophiles nous permettra de répondre à ces diverses questions.

Dans la littérature, il a été montré que la reconnaissance des gènes par les piRNAs est souvent basée sur un appariement incomplet. En effet, un appariement de quatorze bases entre les piRNAs et leurs transcrits cibles serait suffisant chez l'embryon de drosophiles pour que ces derniers exercent leur rôle de régulateur génique (Barckmann et al., 2015; Rouget et al., 2010). Les futures recherches *in silico* visant à identifier les gènes potentiellement régulés par le cluster de piRNAs *flamenco* seront donc optimisées en autorisant un nombre plus important de mésappariements, mais aussi en analysant l'ensemble des piRNAs produits par *flamenco* et





non pas que les piRNAs uniques, afin de ne pas sous-estimer le nombre de transcrits potentiellement régulés par le cluster de piRNAs *flamenco*.

L'identification de gènes pouvant être régulés par les piRNAs issus de la maturation de transcrits de clusters de piRNAs engendre un certain nombre de questions : la régulation génique effectuée par ces piRNAs est-elle restreinte à des fenêtres spatio-temporelles au cours du développement ? Quels sont les critères de reconnaissance des transcrits de gènes devant être régulés ? Les acteurs impliqués dans la régulation génique menée par les piRNAs sont-ils différents de ceux impliqués dans la régulation de l'expression des ETs ? Quels sont les mécanismes mis en jeu lors de la régulation des gènes par les piRNAs ? Ce sont autant de questions auxquelles les futures recherches permettront de répondre.



# *Références*

## Références

- Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F., George, R.A., Lewis, S.E., Richards, S., Ashburner, M., Henderson, S.N., Sutton, G.G., Wortman, J.R., Yandell, M.D., Zhang, Q., Chen, L.X., Brandon, R.C., Rogers, Y.-H.C., Blazej, R.G., Champe, M., Pfeiffer, B.D., Wan, K.H., Doyle, C., Baxter, E.G., Helt, G., Nelson, C.R., Gabor, G.L., Miklos, Abril, J.F., Agbayani, A., An, H.-J., Andrews-Pfannkoch, C., Baldwin, D., Ballew, R.M., Basu, A., Baxendale, J., Bayraktaroglu, L., Beasley, E.M., Beeson, K.Y., Benos, P. V, Berman, B.P., Bhandari, D., Bolshakov, S., Borkova, D., Botchan, M.R., Bouck, J., Brokstein, P., Brottier, P., Burtis, K.C., Busam, D.A., Butler, H., Cadieu, E., Center, A., Chandra, I., Cherry, J.M., Cawley, S., Dahlke, C., Davenport, L.B., Davies, P., Pablos, B. de, Delcher, A., Deng, Z., Mays, A.D., Dew, I., Dietz, S.M., Dodson, K., Doup, L.E., Downes, M., Dugan-Rocha, S., Dunkov, B.C., Dunn, P., Durbin, K.J., Evangelista, C.C., Ferraz, C., Ferriera, S., Fleischmann, W., Fosler, C., Gabrielian, A.E., Garg, N.S., Gelbart, W.M., Glasser, K., Glodek, A., Gong, F., Gorrell, J.H., Gu, Z., Guan, P., Harris, M., Harris, N.L., Harvey, D., Heiman, T.J., Hernandez, J.R., Houck, J., Hostin, D., Houston, K.A., Howland, T.J., Wei, M.-H., Ibegwam, C., Jalali, M., Kalush, F., Karpen, G.H., Ke, Z., Kennison, J.A., Ketchum, K.A., Kimmel, B.E., Kodira, C.D., Kraft, C., Kravitz, S., Kulp, D., Lai, Z., Lasko, P., Lei, Y., Levitsky, A.A., Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Mattei, B., McIntosh, T.C., McLeod, M.P., McPherson, D., Merkulov, G., Milshina, N. V, Mobarry, C., Morris, J., Moshrefi, A., Mount, S.M., Moy, M., Murphy, B., Murphy, L., Muzny, D.M., Nelson, D.L., Nelson, D.R., Nelson, K.A., Nixon, K., Nusskern, D.R., Pacleb, J.M., Palazzolo, M., Pittman, G.S., Pan, S., Pollard, J., Puri, V., Reese, M.G., Reinert, K., Remington, K., Saunders, R.D.C., Scheeler, F., Shen, H., Shue, B.C., Sidén-Kiamos, I., Simpson, M., Skupski, M.P., Smith, T., Spier, E., Spradling, A.C., Stapleton, M., Strong, R., Sun, E., Svirskas, R., Tector, C., Turner, R., Venter, E., Wang, A.H., Wang, X., Wang, Z.-Y., Wassarman, D.A., Weinstock, G.M., Weissenbach, J., Williams, S.M., Woodage, T., Worley, K.C., Wu, D., Yang, S., Yao, Q.A., Ye, J., Yeh, R.-F., Zaveri, J.S., Zhan, M., Zhang, G., Zhao, Q., Zheng, L., Zheng, X.H., Zhong, F.N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H.O., Gibbs, R.A., Myers, E.W., Rubin, G.M., Venter, J.C., 2000. The Genome Sequence of *Drosophila melanogaster*. *Science* (80-. ). 287, 2185 LP – 2195. doi:10.1126/science.287.5461.2185
- Adli, M., 2018. The CRISPR tool kit for genome editing and beyond. *Nat. Commun.* 9, 1911. doi:10.1038/s41467-018-04252-2
- Adrion, J.R., Song, M.J., Schrider, D.R., Hahn, M.W., Schaack, S., 2017. Genome-Wide Estimates of Transposable Element Insertion and Deletion Rates in *Drosophila Melanogaster*. *Genome Biol. Evol.* 9, 1329–1340. doi:10.1093/gbe/evx050
- Akkouche, A., Mugat, B., Barckmann, B., Varela-Chavez, C., Li, B., Raffel, R., Pélisson, A., Chambeyron, S., 2017. Piwi Is Required during *Drosophila* Embryogenesis to License Dual-Strand piRNA Clusters for Transposon Repression in Adult Ovaries. *Mol. Cell* 66, 411-419.e4. doi:10.1016/j.molcel.2017.03.017

- Araújo, S.J., 2015. The hedgehog signalling pathway in cell migration and guidance: What we have learned from *Drosophila melanogaster*. *Cancers (Basel)*. 7, 2012–2022. doi:10.3390/cancers7040873
- 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., Tuschl, T., 2006. A novel class of small RNAs bind to MILI protein in mouse testes. *Nature* 442, 203–207. doi:10.1038/nature04916
- Aravin, A.A., Hannon, G.J., Brennecke, J., 2007a. The Piwi-piRNA Pathway Provides an Adaptive Defense in the Transposon Arms Race. *Science (80- )*. 318, 761–764. doi:10.1126/science.1146484
- Aravin, A.A., Naumova, N.M., Tulin, A. V., Vagin, V. V., Rozovsky, Y.M., Gvozdev, V.A., 2001. Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the *Drosophila melanogaster* germline. *Curr. Biol.* 11, 1017–27.
- Aravin, A.A., Sachidanandam, R., Bourc'his, D., Schaefer, C., Pezic, D., Toth, K.F., Bestor, T., 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. doi:10.1016/j.molcel.2008.09.003
- Aravin, A.A., Sachidanandam, R., Girard, A., Fejes-Toth, K., Hannon, G.J., 2007b. Developmentally Regulated piRNA Clusters Implicate MILI in Transposon Control. *Science (80- )*. 316, 744–747. doi:10.1126/science.1142612
- Ashe, A., Sapetschnig, A., Weick, E.-M., Mitchell, J., Bagijn, M.P., Cording, A.C., Doebley, A.-L., Goldstein, L.D., Lehrbach, N.J., Le Pen, J., Pintacuda, G., Sakaguchi, A., Sarkies, P., Ahmed, S., Miska, E.A., 2012. piRNAs Can Trigger a Multigenerational Epigenetic Memory in the Germline of *C. elegans*. *Cell* 150, 88–99. doi:10.1016/j.cell.2012.06.018
- Assa-Kunik, E., Torres, I.L., Schejter, E.D., Johnston, D.S., Shilo, B.-Z., 2007. *Drosophila* follicle cells are patterned by multiple levels of Notch signaling and antagonism between the Notch and JAK/STAT pathways. *Development* 134, 1161 LP – 1169. doi:10.1242/dev.02800
- Barckmann, B., El-Barouk, M., Pélisson, A., Mugat, B., Li, B., Franckhauser, C., Lavier, A.S.F., Mirouze, M., Fablet, M., Chambeyron, S., 2018. The somatic piRNA pathway controls germline transposition over generations. *Nucleic Acids Res.* 46, 9524–9536. doi:10.1093/nar/gky761
- Barckmann, B., Pierson, S., Dufourt, J., Papin, C., Armenise, C., Port, F., Grentzinger, T., Chambeyron, S., Baronian, G., Desvignes, J.-P., Curk, T., Simonelig, M., 2015. Aubergine iCLIP Reveals piRNA-Dependent Decay of mRNAs Involved in Germ Cell Development in the Early Embryo. *Cell Rep.* 12, 1205–1216. doi:10.1016/j.celrep.2015.07.030
- Beguiristain, T., Grandbastien, M.A., Puigdomènech, P., Casacuberta, J.M., 2001. Three Tnt1 subfamilies show different stress-associated patterns of expression in tobacco. Consequences for retrotransposon control and

- evolution in plants. *Plant Physiol.* 127, 212–21. doi:10.1104/PP.127.1.212
- Bellen, H.J., Levis, R.W., Liao, G., He, Y., Carlson, J.W., Tsang, G., Evans-Holm, M., Hiesinger, P.R., Schulze, K.L., Rubin, G.M., Hoskins, R.A., Spradling, A.C., 2004. The BDGP Gene Disruption Project. *Genetics* 167, 761–781. doi:10.1534/genetics.104.026427
- Bier, E., 2005. *Drosophila*, the golden bug, emerges as a tool for human genetics. *Nat. Rev. Genet.* doi:10.1038/nrg1503
- Biessmann, H., Mason, J.M., 1992. Genetics and molecular biology of telomeres. *Adv. Genet.* 30, 185–249.
- Biessmann, H., Valgeirsdottir, K., Lofsky, A., Chin, C., Ginther, B., Levis, R.W., Pardue, M.L., 1992. HeT-A, a transposable element specifically involved in "healing" broken chromosome ends in *Drosophila melanogaster*. *Mol. Cell Biol.* 12, 3910–8. doi:10.1128/mcb.12.9.3910
- Bohmert, K., Camus, I., Bellini, C., Bouchez, D., Caboche, M., Benning, C., 1998. AGO1 defines a novel locus of *Arabidopsis* controlling leaf development. *EMBO J.* 17, 170–180. doi:10.1093/emboj/17.1.170
- Bosco, G., Campbell, P., Leiva-Neto, J.T., Markow, T.A., 2007. Analysis of *Drosophila* species genome size and satellite DNA content reveals significant differences among strains as well as between species. *Genetics* 177, 1277–1290. doi:10.1534/genetics.107.075069
- Brasset, E., Taddei, A.R., Arnaud, F., Faye, B., Fausto, A.M., Mazzini, M., Giorgi, F., Vaury, C., 2006. Viral particles of the endogenous retrovirus ZAM from *Drosophila melanogaster* use a pre-existing endosome/exosome pathway for transfer to the oocyte. *Retrovirology* 3, 1–9. doi:10.1186/1742-4690-3-25
- Bray, S.J., 2006. Notch signalling: a simple pathway becomes complex. *Nat. Rev. Mol. Cell Biol.* 7, 678–689. doi:10.1038/nrm2009
- Brennecke, J., Aravin, A.A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., Hannon, G.J., 2007. Discrete Small RNA-Generating Loci as Master Regulators of Transposon Activity in *Drosophila*. *Cell* 128, 1089–1103. doi:10.1016/j.cell.2007.01.043
- Brennecke, J., Malone, C.D., Aravin, A.A., Sachidanandam, R., Stark, A., Hannon, G.J., 2008. An epigenetic role for maternally inherited piRNAs in transposon silencing. *Science* (80-. ). 322, 1387–1392. doi:10.1126/science.1165171
- Brouha, B., Badge, R.M., Farley, A.H., Moran, J. V., Lutz-Prigge, S., Kazazian, H.H., Schustak, J., 2003. Hot L1s account for the bulk of retrotransposition in the human population. *Proc. Natl. Acad. Sci.* 100, 5280–5285. doi:10.1073/pnas.0831042100
- Brouha, Brook, Schustak, J., Badge, R.M., Lutz-Prigge, S., Farley, A.H., Moran, J. V., Kazazian, H.H., Jr., 2003. Hot L1s account for the bulk of retrotransposition in the human population. *Proc. Natl. Acad. Sci. U. S. A.* 100, 5280–5. doi:10.1073/pnas.0831042100

- Brower-Toland, B., Findley, S.D., Jiang, L., Liu, L., Yin, H., Dus, M., Zhou, P., Elgin, S.C.R., Lin, H., 2007. *Drosophila* PIWI associates with chromatin and interacts directly with HP1a. *Genes & Dev.* 21, 2300–2311. doi:10.1101/gad.1564307
- Busseau, I., Malinsky, S., Balakireva, M., Chaboissier, M.C., Teninges, D., Bucheton, A., 1998. A genetically marked I element in *Drosophila melanogaster* can be mobilized when ORF2 is provided in trans. *Genetics* 148, 267–275.
- Butelli, E., Licciardello, C., Zhang, Y., Liu, J., Mackay, S., Bailey, P., Reforgiato-Recupero, G., Martin, C., 2012. Retrotransposons Control Fruit-Specific, Cold-Dependent Accumulation of Anthocyanins in Blood Oranges. *Plant Cell* 24, 1242–1255. doi:10.1105/tpc.111.095232
- Caligiuri, M.A., Strout, M.P., Lawrence, D., Arthur, D.C., Baer, M.R., Yu, F., Knuutila, S., Mrózek, K., Oberkircher, A.R., Marcucci, G., de la Chapelle, A., Elonen, E., Block, A.W., Rao, P.N., Herzig, G.P., Powell, B.L., Ruutu, T., Schiffer, C.A., Bloomfield, C.D., 1998. Rearrangement of ALL1 (MLL) in acute myeloid leukemia with normal cytogenetics. *Cancer Res.* 58, 55–9.
- Carmell, M.A., Girard, A., van de Kant, H.J.G., Bourc'his, D., Bestor, T.H., de Rooij, D.G., Hannon, G.J., 2007. MIWI2 Is Essential for Spermatogenesis and Repression of Transposons in the Mouse Male Germline. *Dev. Cell* 12, 503–514. doi:10.1016/j.devcel.2007.03.001
- Casacuberta, E., Pardue, M.-L., 2005. <i>HeT-A</i> and <i>TART</i>, two <i>Drosophila</i> retrotransposons with a bona fide role in chromosome structure for more than 60 million years. *Cytogenet. Genome Res.* 110, 152–159. doi:10.1159/000084947
- Cavrak, V. V., Lettner, N., Jamge, S., Kosarewicz, A., Bayer, L.M., Mittelsten Scheid, O., 2014. How a Retrotransposon Exploits the Plant's Heat Stress Response for Its Activation. *PLoS Genet.* 10, e1004115. doi:10.1371/journal.pgen.1004115
- Cecere, G., Zheng, G.X.Y., Mansisidor, A.R., Klymko, K.E., Grishok, A., 2012. Article Promoters Recognized by Forkhead Proteins Exist for Individual 21U-RNAs. *Mol. Cell* 47, 734–745. doi:10.1016/j.molcel.2012.06.021
- Celniker, S., Wheeler, D., Kronmiller, B., Carlson, J., Halpern, A., Patel, S., Adams, M., Champe, M., Dugan, S., Frise, E., Hodgson, A., George, R., Hoskins, R., Laverty, T., Muzny, D., Nelson, C., Pacleb, J., Park, S., Pfeiffer, B., Richards, S., Sodergren, E., Svirskas, R., Tabor, P., Wan, K., Stapleton, M., Sutton, G., Venter, C., Weinstock, G., Scherer, S., Myers, E., Gibbs, R., Rubin, G., 2002. Finishing a whole-genome shotgun: Release 3 of the *Drosophila melanogaster* euchromatic genome sequence. *Genome Biol.* 3, research0079.1-0079.14. doi:10.1186/gb-2002-3-12-research0079
- Chang, T.H., Mattei, E., Gainetdinov, I., Colpan, C., Weng, Z., Zamore, P.D., 2019. Maelstrom Represses Canonical Polymerase II Transcription within Bi-directional piRNA Clusters in *Drosophila melanogaster*. *Mol. Cell* 73, 291-303.e6. doi:10.1016/j.molcel.2018.10.038



- Charlesworth, B., Lapid, A., Canada, D., 2019. The distribution of transposable elements within and between chromosomes in a population of *Drosophila melanogaster*. I. Element frequencies and distribution. *Genet. Res., Camb* 60, 103–114. doi:10.1017/S0016672300030792
- Chen, Y.C.A., Stuwe, E., Luo, Y., Ninova, M., Le Thomas, A., Rozhavskaia, E., Li, S., Vempati, S., Laver, J.D., Patel, D.J., Smibert, C.A., Lipshitz, H.D., Fejes Toth, K., Aravin, A.A., 2016. Cutoff Suppresses RNA Polymerase II Termination to Ensure Expression of piRNA Precursors. *Mol. Cell* 63, 97–109. doi:10.1016/j.molcel.2016.05.010
- Chénais, B., 2013. Transposable elements and human cancer: A causal relationship? *Biochim. Biophys. Acta - Rev. Cancer* 1835, 28–35. doi:10.1016/j.bbcan.2012.09.001
- Chirn, G., Rahman, R., Sytnikova, Y.A., Matts, J.A., Zeng, M., Gerlach, D., Yu, M., Berger, B., Naramura, M., Kile, B.T., Lau, N.C., 2015. Conserved piRNA Expression from a Distinct Set of piRNA Cluster Loci in Eutherian Mammals. *PLOS Genet.* 11, e1005652. doi:10.1371/journal.pgen.1005652
- Chung, W.-J., Okamura, K., Martin, R., Lai, E.C., 2008. Endogenous RNA Interference Provides a Somatic Defense against *Drosophila* Transposons. *Curr. Biol.* 18, 795–802. doi:10.1016/j.cub.2008.05.006
- Cowley, M., Oakey, R.J., 2013. Transposable Elements Re-Wire and Fine-Tune the Transcriptome. *PLoS Genet.* 9, e1003234. doi:10.1371/journal.pgen.1003234
- Creighton, H.B., McClintock, B., 1931. A Correlation of Cytological and Genetical Crossing-Over in *Zea Mays*. *Proc. Natl. Acad. Sci. U. S. A.* 17, 492–7. doi:10.1073/pnas.17.8.492
- 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., Brennecke, J., 2008. An endogenous small interfering RNA pathway in *Drosophila*. *Nature* 453, 798–802. doi:10.1038/nature07007
- Daniels, S.B., Peterson, K.R., Strausbaugh, L.D., Kidwell, M.G., Chovnick, A., 1990. Evidence for horizontal transmission of the P transposable element between *Drosophila* species. *Genetics* 124, 339–55.
- Dansereau, D.A., Lasko, P., 2008. RanBPM regulates cell shape, arrangement, and capacity of the female germline stem cell niche in *Drosophila melanogaster*. *J. Cell Biol.* 182, 963–977. doi:10.1083/jcb.200711046
- Datta, N., Hedges, R.W., Shaw, E.J., Sykes, R., Richmond, M.H., 1971. Properties of an R Factor from *J. Bacteriol.* 108, 1244–1249.
- De Fazio, S., Bartonicek, N., Di Giacomo, M., Abreu-Goodger, C., Sankar, A., Funaya, C., Antony, C., Moreira, P.N., Enright, A.J., O'Carroll, D., 2011. The endonuclease activity of Mili fuels piRNA amplification that silences LINE1 elements. *Nature* 480, 259–263. doi:10.1038/nature10547
- De La Chaux, N., Tsuchimatsu, T., Shimizu, K.K., Wagner, A., 2012. The predominantly selfing plant

- Arabidopsis thaliana* experienced a recent reduction in transposable element abundance compared to its outcrossing relative *Arabidopsis lyrata*. *Mob. DNA* 3, 1–18. doi:10.1186/1759-8753-3-2
- de la Cruz, A.F.A., Edgar, B.A., 2008. Flow Cytometric Analysis of *Drosophila* Cells BT - *Drosophila: Methods and Protocols*, in: Dahmann, C. (Ed.), . Humana Press, Totowa, NJ, pp. 373–389. doi:10.1007/978-1-59745-583-1\_24
- DeFalco, T., Capel, B., 2009. Gonad Morphogenesis in Vertebrates: Divergent Means to a Convergent End. *Annu. Rev. Cell Dev. Biol.* 25, 457–482. doi:10.1146/annurev.cellbio.042308.13350.GONAD
- Deininger, P., 2011. Alu elements: Know the SINEs. *Genome Biol.* doi:10.1186/gb-2011-12-12-236
- Deng, W., Lin, H., 2002. miwi, a murine homolog of piwi, encodes a cytoplasmic protein essential for spermatogenesis. *Dev. Cell* 2, 819–30.
- Dennis, C., Brasset, E., Sarkar, A., Vaury, C., 2016. Export of piRNA precursors by EJC triggers assembly of cytoplasmic Yb-body in *Drosophila*. *Nat. Commun.* 7, 1–12. doi:10.1038/ncomms13739
- Dennis, C., Zanni, V., Brasset, E., Eymery, A., Zhang, L., Mteirek, R., Jensen, S., Rong, Y.S., Vaury, C., 2013. “Dot COM”, a Nuclear Transit Center for the Primary piRNA Pathway in *Drosophila*. *PLoS One* 8. doi:10.1371/journal.pone.0072752
- Deshpande, G., Swanhart, L., Chiang, P., Schedl, P., 2001. Hedgehog signaling in germ cell migration. *Cell* 106, 759–769. doi:10.1016/S0092-8674(01)00488-3
- Desset, S., Conte, C., Dimitri, P., Calco, V., Dastugue, B., Vaury, C., 1999. Mobilization of two retroelements, ZAM and Idefix, in a novel unstable line of *Drosophila melanogaster*. *Mol. Biol. Evol.* 16, 54–66. doi:10.1093/oxfordjournals.molbev.a026038
- Desset, S., Conte, C., Dimitri, P., Dastugue, B., Vaury, C., 1995. Mobilization of Two Retroelements , ZAM and Idefix , in a Novel Unstable Line of *Drosophila melanogaster*. *Mol. Biol.* 54–66.
- Desset, S., Meignin, C., Dastugue, B., Vaury, C., 2003. COM, a heterochromatic locus governing the control of independent endogenous retroviruses from *Drosophila melanogaster*. *Genetics* 164, 501–509.
- Di Giacomo, M., Comazzetto, S., Saini, H., De Fazio, S., Carrieri, C., Morgan, M., Vasiliauskaite, L., Benes, V., Enright, A.J., O’Carroll, D., 2013. Multiple Epigenetic Mechanisms and the piRNA Pathway Enforce LINE1 Silencing during Adult Spermatogenesis. *Mol. Cell* 50, 601–608. doi:10.1016/j.molcel.2013.04.026
- Ding, X., Guan, H., Li, H., 2013. Characterization of a piRNA binding protein Miwi in mouse oocytes. *Theriogenology* 79, 610-615.e1. doi:10.1016/j.theriogenology.2012.11.013
- Duc, C., Yoth, M., Mounié, N., Jensen, S., Vaury, C., Brasset, E., 2019. Trapping a somatic endogenous retrovirus into a germline piRNA cluster immunizes the germline against further invasion Keywords.

- Dufourt, J., Bontonou, G., Chartier, A., Jahan, C., Meunier, A.-C., Pierson, S., Harrison, P.F., Papin, C., Beilharz, T.H., Simonelig, M., 2017. piRNAs and Aubergine cooperate with Wispy poly(A) polymerase to stabilize mRNAs in the germ plasm. *Nat. Commun.* 8, 1305. doi:10.1038/s41467-017-01431-5
- Dufourt, J., Dennis, C., Boivin, A., Gueguen, N., Théron, E., Goriaux, C., Pouchin, P., Ronsseray, S., Brassat, E., Vaury, C., 2014. Spatio-temporal requirements for transposable element piRNA-mediated silencing during *Drosophila* oogenesis. *Nucleic Acids Res.* 42, 2512–2524. doi:10.1093/nar/gkt1184
- Dupressoir, A., Vernochet, C., Bawa, O., Harper, F., Pierron, G., Opolon, P., Heidmann, T., 2009. Syncytin-A knockout mice demonstrate the critical role in placentation of a fusogenic, endogenous retrovirus-derived, envelope gene. *Proc. Natl. Acad. Sci.* 106, 12127–12132. doi:10.1073/pnas.0902925106
- Dupressoir, A., Vernochet, C., Harper, F., Guegan, J., Dessen, P., Pierron, G., Heidmann, T., 2011. A pair of co-opted retroviral envelope syncytin genes is required for formation of the two-layered murine placental syncytiotrophoblast. *Proc. Natl. Acad. Sci.* 108, E1164–E1173. doi:10.1073/pnas.1112304108
- Elbarbary, R.A., Lucas, B.A., Maquat, L.E., 2016. Retrotransposons as regulators of gene expression. *Science* (80-. ). 351, aac7247–aac7247. doi:10.1126/science.aac7247
- Elmaghraby, M.F., Andersen, P.R., Pühringer, F., Meixner, K., Lendl, T., Tirian, L., Brennecke, J., 2019. A heterochromatin-specific RNA export pathway facilitates piRNA production. doi:10.1101/596171
- Findley, S.D., Tamanaha, M., Clegg, N.J., 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, 859–71.
- Finnegan, D.J., 1992. Transposable elements. *Curr. Opin. Genet. Dev.* 2, 861–7.
- Finnegan, D.J., 1989. Eukaryotic transposable elements and genome evolution. *Trends Genet.* 5, 103–7.
- Fisher B, Weizmann R, Frise E, Hammonds A, Tomancak P, Beaton A, Berman B, Quan E, Shu S, Lewis S, Rubin G, Barale C, Laguertas E, Quinn J, Ghosh A, Hartenstein V, Ashburner M, Celniker S. BDGP insitu homepage.2012. <http://flybase.org/reports/FBrf0219073.html>.
- Flemr, M., Malik, R., Franke, V., Nejeplinska, J., Sedlacek, R., Vlahovicek, K., Svoboda, P., 2013. A Retrotransposon-Driven Dicer Isoform Directs Endogenous Small Interfering RNA Production in Mouse Oocytes. *Cell* 155, 807–816. doi:10.1016/j.cell.2013.10.001
- Freedman, J.E., Gerstein, M., Mick, E., Rozowsky, J., Levy, D., Kitchen, R., Das, S., Shah, R., Danielson, K., Beaulieu, L., Navarro, F.C.P., Wang, Y., Galeev, T.R., Holman, A., Kwong, R.Y., Murthy, V., Tanriverdi, S.E., Koupenova, M., Mikhalev, E., Tanriverdi, K., 2016. Diverse human extracellular RNAs are widely detected in human plasma. *Nat. Commun.* 7, 11106. doi:10.1038/ncomms11106
- Freeman, M.R., 2015. *Drosophila* Central Nervous System Glia. *Cold Spring Harb. Perspect. Biol.* 7, a020552.

doi:10.1101/cshperspect.a020552

- Gebert, D., Ketting, R.F., Zischler, H., Rosenkranz, D., 2015. piRNAs from Pig Testis Provide Evidence for a Conserved Role of the Piwi Pathway in Post-Transcriptional Gene Regulation in Mammals. *PLoS One* 10, e0124860.
- Ghildiyal, M., Seitz, H., Horwich, M.D., Li, C., Du, T., Lee, S., Xu, J., Kittler, E.L.W., Zapp, M.L., Weng, Z., Zamore, P.D., 2008. Endogenous siRNAs Derived from Transposons and mRNAs in *Drosophila* Somatic Cells. *Science* (80-. ). 320, 1077–1081. doi:10.1126/science.1157396
- Ghildiyal, M., Zamore, P.D., 2009. Small silencing RNAs: an expanding universe. *Nat. Rev. Genet.* 10, 94–108. doi:10.1038/nrg2504
- Gilboa, L., Lehmann, R., 2006. Soma–germline interactions coordinate homeostasis and growth in the *Drosophila* gonad. *Nature* 443, 97–100. doi:10.1038/nature05068
- Girard, A., Sachidanandam, R., Hannon, G.J., Carmell, M.A., 2006. A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* 442, 199–202. doi:10.1038/nature04917
- Goh, S., Siong, W., Falciatori, I., Tam, O.H., Burgess, R., Meikar, O., Kotaja, N., Hammell, M., Hannon, G.J., 2015. piRNA-directed cleavage of meiotic transcripts regulates spermatogenesis. *Genes Dev.* 29, 1032–1044. doi:10.1101/gad.260455.115.5
- Goriaux, C., Renaud, Y., Brasset, E., Desset, S., Vaury, C., 2014. Transcriptional properties and splicing of the flamenco piRNA cluster. *EMBO Rep.* 15, 411–418. doi:10.1002/embr.201337898
- Gou, L.-T., Dai, P., Yang, J.-H., Xue, Y., Hu, Y.-P., Zhou, Y., Kang, J.-Y., Wang, X., Li, H., Hua, M.-M., Zhao, S., Hu, S.-D., Wu, L.-G., Shi, H.-J., Li, Y., Fu, X.-D., Qu, L.-H., Wang, E.-D., Liu, M.-F., 2014. Pachytene piRNAs instruct massive mRNA elimination during late spermiogenesis. *Cell Res.* 24, 680.
- Grandbastien, M.A., Lucas, H., Morel, J.B., Mhiri, C., Vernhettes, S., Casacuberta, J.M., 1997. The expression of the tobacco Tnt1 retrotransposon is linked to plant defense responses. *Genetica* 100, 241–52.
- Grimson, A., Srivastava, M., Fahey, B., Woodcroft, B.J., Chiang, H.R., King, N., Degnan, B.M., Rokhsar, D.S., Bartel, D.P., 2008. Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. *Nature* 455, 1193–1197. doi:10.1038/nature07415
- Grivna, S.T., Beyret, E., Wang, Z., Lin, H., 2006. A novel class of small RNAs in mouse spermatogenic cells. *Genes Dev.* 20, 1709–1714. doi:10.1101/gad.1434406
- Gunawan, F., Arandjelovic, M., Godt, D., 2013. The Maf factor Traffic jam both enables and inhibits collective cell migration in *Drosophila*; oogenesis. *Development* 140, 2808 LP – 2817. doi:10.1242/dev.089896

- Gunawardane, L.S., Saito, K., Nishida, K.M., Miyoshi, K., Kawamura, Y., Nagami, T., Siomi, H., Siomi, M.C., 2007. A Slicer-Mediated Mechanism for Repeat-Associated siRNA 5' End Formation in *Drosophila*. *Science* (80- ). 315, 1587–1590. doi:10.1126/science.1140494
- Ha, H., Song, J., Wang, S., Kapusta, A., Feschotte, C., Chen, K.C., Xing, J., 2014. A comprehensive analysis of piRNAs from adult human testis and their relationship with genes and mobile elements. *BMC Genomics* 15, 1–16. doi:10.1186/1471-2164-15-545
- Han, B.W., Wang, W., Li, C., Weng, Z., Zamore, P.D., 2015. Noncoding RNA. piRNA-guided transposon cleavage initiates Zucchini-dependent, phased piRNA production. *Science* 348, 817–821. doi:10.1126/science.aaa1264
- Hancks, D.C., Kazazian, H.H., 2016. Roles for retrotransposon insertions in human disease. *Mob. DNA* 7, 9. doi:10.1186/s13100-016-0065-9
- Hancks, D.C., Kazazian, H.H., Jr., 2012. Active human retrotransposons: variation and disease. *Curr. Opin. Genet. Dev.* 22, 191–203. doi:10.1016/j.gde.2012.02.006
- Handler, D., Meixner, K., Pizka, M., Lauss, K., Schmied, C., Gruber, F.S., Brennecke, J., 2013. The genetic makeup of the *Drosophila* piRNA pathway. *Mol. Cell* 50, 762–777. doi:10.1016/j.molcel.2013.04.031
- Handler, D., Olivieri, D., Novatchkova, M., Gruber, F.S., Meixner, K., Mechtler, K., Stark, A., Sachidanandam, R., Brennecke, J., 2011. A systematic analysis of *Drosophila* TUDOR domain-containing proteins identifies Vreteno and the Tdrd12 family as essential primary piRNA pathway factors. *EMBO J.* 30, 3977–3993. doi:10.1038/emboj.2011.308
- Haren, L., Ton-Hoang, B., Chandler, M., 1999. Integrating DNA: Transposases and Retroviral Integrases. *Annu. Rev. Microbiol.* 53, 245–281. doi:10.1146/annurev.micro.53.1.245
- Harris, A.N., Macdonald, P.M., 2001. Aubergine encodes a *Drosophila* polar granule component required for pole cell formation and related to eIF2C. *Development* 128, 2823–32.
- Hayashi, R., Schnabl, J., Handler, D., Mohn, F., Ameres, S.L., Brennecke, J., 2016. Genetic and mechanistic diversity of piRNA 3'-end formation. *Nature* 539, 588–592. doi:10.1038/nature20162
- Heiman, M., Kulicke, R., Fenster, R.J., Greengard, P., Heintz, N., 2014. Cell type-specific mRNA purification by translating ribosome affinity purification (TRAP). *Nat. Protoc.* 9, 1282–1291. doi:10.1038/nprot.2014.085
- Hellen, E.H.B., Brookfield, J.F.Y., 2013. The diversity of class II transposable elements in mammalian genomes has arisen from ancestral phylogenetic splits during ancient waves of proliferation through the genome. *Mol. Biol. Evol.* 30, 100–8. doi:10.1093/molbev/mss206
- Hiom, K., Melek, M., Gellert, M., 1998. DNA transposition by the RAG1 and RAG2 proteins: a possible source

of oncogenic translocations. *Cell* 94, 463–70.

- Hirano, T., Iwasaki, Y.W., Lin, Z.Y.-C., Imamura, M., Seki, N.M., Sasaki, E., Saito, K., Okano, H., Siomi, M.C., Siomi, H., 2014. Small RNA profiling and characterization of piRNA clusters in the adult testes of the common marmoset, a model primate. *RNA* 20, 1223–1237. doi:10.1261/rna.045310.114
- Homolka, D., Pandey, R.R., Goriaux, C., Brassat, E., Vaury, C., Sachidanandam, R., Fauvarque, M.O., Pillai, R.S., 2015. PIWI Slicing and RNA Elements in Precursors Instruct Directional Primary piRNA Biogenesis. *Cell Rep.* 12, 418–428. doi:10.1016/j.celrep.2015.06.030
- Horwich, M.D., Li, C., Matranga, C., Vagin, V., Farley, G., Wang, P., Zamore, P.D., 2007. The *Drosophila* RNA Methyltransferase, DmHen1, Modifies Germline piRNAs and Single-Stranded siRNAs in RISC. *Curr. Biol.* 17, 1265–1272. doi:10.1016/j.cub.2007.06.030
- 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., Plasterk, R.H.A., Hannon, G.J., Draper, B.W., Ketting, R.F., 2007. A Role for Piwi and piRNAs in Germ Cell Maintenance and Transposon Silencing in Zebrafish. *Cell* 129, 69–82. doi:10.1016/j.cell.2007.03.026
- Huang, H., Gao, Q., Peng, X., Choi, S.-Y., Sarma, K., Ren, H., Morris, A.J., Frohman, M.A., 2011. piRNA-Associated Germline Nuage Formation and Spermatogenesis Require MitoPLD Profusogenic Mitochondrial-Surface Lipid Signaling. *Dev. Cell* 20, 376–387. doi:10.1016/j.devcel.2011.01.004
- Hutvagner, G., Simard, M.J., 2008. Argonaute proteins: key players in RNA silencing. *Nat. Rev. Mol. Cell Biol.* 9, 22–32. doi:10.1038/nrm2321
- Ishizu, H., Iwasaki, Y.W., Hirakata, S., Ozaki, H., Iwasaki, W., Siomi, H., Siomi, M.C., 2015. Somatic Primary piRNA Biogenesis Driven by cis-Acting RNA Elements and trans-Acting Yb. *Cell Rep.* 12, 429–440. doi:10.1016/j.celrep.2015.06.035
- Izumi, N., Kawaoka, S., Yasuhara, S., Suzuki, Y., Sugano, S., Katsuma, S., Tomari, Y., 2013. Hsp90 facilitates accurate loading of precursor piRNAs into PIWI proteins. *RNA* 19, 896–901. doi:10.1261/rna.037200.112
- Jaglarz, M.K., Howard, K.R., 1995. The active migration of *Drosophila* primordial germ cells. *Development* 121, 3495–503.
- Jones, B.C., Wood, J.G., Chang, C., Tam, A.D., Franklin, M.J., Siegel, E.R., Helfand, S.L., 2016. A somatic piRNA pathway in the *Drosophila* fat body ensures metabolic homeostasis and normal lifespan. *Nat. Commun.* 7, 13856. doi:10.1038/ncomms13856
- Juliano, C., Wang, J., Lin, H., 2011. Uniting Germline and Stem Cells: The Function of Piwi Proteins and the piRNA Pathway in Diverse Organisms. *Annu. Rev. Genet.* 45, 447–469. doi:10.1146/annurev-genet-110410-132541

- Junion, G., Bataillé, L., Jagla, T., Da Ponte, J.P., Tapin, R., Jagla, K., 2007. Genome-wide view of cell fate specification: Ladybird acts at multiple levels during diversification of muscle and heart precursors. *Genes Dev.* 21, 3163–3180. doi:10.1101/gad.437307
- Kapitonov, V. V., Jurka, J., 2008. A universal classification of eukaryotic transposable elements implemented in Repbase. *Nat. Rev. Genet.* 9, 411–412. doi:10.1038/nrg2165-c1
- Kapitonov, V. V., Jurka, J., 2005. RAG1 Core and V(D)J Recombination Signal Sequences Were Derived from Transib Transposons. *PLoS Biol.* 3, e181. doi:10.1371/journal.pbio.0030181
- Kataoka, K., Mochizuki, K., 2011. Programmed DNA elimination in Tetrahymena: a small RNA-mediated genome surveillance mechanism. *Adv. Exp. Med. Biol.* 722, 156–173. doi:10.1007/978-1-4614-0332-6\_10
- Kawamura, Y., Saito, K., Kin, T., Ono, Y., Asai, K., Sunohara, T., Okada, T.N., Siomi, M.C., Siomi, H., 2008. *Drosophila* endogenous small RNAs bind to Argonaute 2 in somatic cells. *Nature* 453, 793–797. doi:10.1038/nature06938
- Kawaoka, S., Hayashi, N., Suzuki, Y., Abe, H., Sugano, S., Tomari, Y., Shimada, T., Katsuma, S., 2009. The *Bombyx* ovary-derived cell line endogenously expresses PIWI/PIWI-interacting RNA complexes. *RNA* 15, 1258–1264. doi:10.1261/rna.1452209
- Khurana, J.S., Wang, J., Xu, J., Koppetsch, B.S., Thomson, T.C., Nowosielska, A., Li, C., Zamore, P.D., Weng, Z., Theurkauf, W.E., 2011. Adaptation to P element transposon invasion in *Drosophila melanogaster*. *Cell* 147, 1551–63. doi:10.1016/j.cell.2011.11.042
- Kidwell, M.G., Kidwell, J.F., Sved, J.A., 1977. Hybrid Dysgenesis in *DROSOPHILA MELANOGASTER*: A Syndrome of Aberrant Traits Including Mutation, Sterility and Male Recombination. *Genetics* 86, 813–33.
- Kidwell, M.G., Lisch, D.R., 2001. PERSPECTIVE: TRANSPOSABLE ELEMENTS, PARASITIC DNA, AND GENOME EVOLUTION. *Evolution (N. Y.)* 55, 1–24. doi:10.1111/j.0014-3820.2001.tb01268.x
- Kim, K.W., Tang, N.H., Andrusiak, M.G., Wu, Z., Chisholm, A.D., Jin, Y., 2018. A Neuronal piRNA Pathway Inhibits Axon Regeneration in *C. elegans*. *Neuron* 97, 511-519.e6. doi:10.1016/j.neuron.2018.01.014
- Kirino, Y., Mourelatos, Z., 2007. Mouse Piwi-interacting RNAs are 2'-O-methylated at their 3' termini. *Nat. Struct. Mol. Biol.* 14, 347–348. doi:10.1038/nsmb1218
- Kitadate, Y., Kobayashi, S., 2010. Notch and Egfr signaling act antagonistically to regulate germ-line stem cell niche formation in *Drosophila* male embryonic gonads. *Proc. Natl. Acad. Sci. U. S. A.* 107, 14241–14246. doi:10.1073/pnas.1003462107
- Kiuchi, T., Koga, H., Kawamoto, M., Shoji, K., Sakai, H., Arai, Y., Ishihara, G., Kawaoka, S., Sugano, S., Shimada, T., Suzuki, Y., Suzuki, M.G., Katsuma, S., 2014a. A single female-specific piRNA is the primary determiner of sex in the silkworm. *Nature* 509, 633–636. doi:10.1038/nature13315

- Kiuchi, T., Koga, H., Kawamoto, M., Shoji, K., Sakai, H., Arai, Y., Ishihara, G., Kawaoka, S., Sugano, S., Shimada, T., Suzuki, Y., Suzuki, M.G., Katsuma, S., 2014b. A single female-specific piRNA is the primary determiner of sex in the silkworm. *Nature* 509, 633.
- Klattenhoff, C., Xi, H., Li, C., Lee, S., Xu, J., Khurana, J.S., Zhang, F., Schultz, N., Koppetsch, B.S., Nowosielska, A., Seitz, H., Zamore, P.D., Weng, Z., Theurkauf, W.E., 2009. The *Drosophila* HP1 Homolog Rhino Is Required for Transposon Silencing and piRNA Production by Dual-Strand Clusters. *Cell* 138, 1137–1149. doi:10.1016/j.cell.2009.07.014
- Klenov, M.S., Lavrov, S.A., Korbut, A.P., Stolyarenko, A.D., Yakushev, E.Y., Reuter, M., Pillai, R.S., Gvozdev, V.A., 2014. Impact of nuclear Piwi elimination on chromatin state in *Drosophila melanogaster* ovaries. *Nucleic Acids Res.* 42, 6208–6218. doi:10.1093/nar/gku268
- Klenov, M.S., Sokolova, O.A., Yakushev, E.Y., Stolyarenko, A.D., Mikhaleva, E.A., Lavrov, S.A., Gvozdev, V.A., 2011. Separation of stem cell maintenance and transposon silencing functions of Piwi protein. *Proc. Natl. Acad. Sci. U. S. A.* 108, 18760–5. doi:10.1073/pnas.1106676108
- Koga, A., Inagaki, H., Bessho, Y., Hori, H., 1995. Insertion of a novel transposable element in the tyrosinase gene is responsible for an albino mutation in the medaka fish, *Oryzias latipes*. *MGG Mol. Gen. Genet.* 249, 400–405. doi:10.1007/BF00287101
- Koga, A., Suzuki, M., Inagaki, H., Bessho, Y., Hori, H., 1996. Transposable element in fish [11]. *Nature* 383, 30. doi:10.1038/383030a0
- Koning, A.P.J. de, Gu, W., Castoe, T.A., Batzer, M.A., Pollock, D.D., 2011. Repetitive Elements May Comprise Over Two-Thirds of the Human Genome. *PLoS Genet.* 7, e1002384. doi:10.1371/JOURNAL.PGEN.1002384
- Kopan, R., Ilagan, M.X.G., 2009. The canonical Notch signaling pathway: unfolding the activation mechanism. *Cell* 137, 216–233. doi:10.1016/j.cell.2009.03.045
- Kuramochi-Miyagawa, S., 2008. DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes.pdf. *Genes ...* 1, 908–917. doi:10.1101/gad.1640708.9
- Kuramochi-Miyagawa, S., Kimura, T., Ijiri, T.W., Isobe, T., Asada, N., Fujita, Y., Ikawa, M., Iwai, N., Okabe, M., Deng, W., Lin, H., Matsuda, Y., Nakano, T., 2004. Mili, a mammalian member of piwi family gene, is essential for spermatogenesis. *Development* 131, 839–849. doi:10.1242/dev.00973
- Laha, T., Loukas, A., Wattanasatitarpa, S., Somprakhon, J., Kewgrai, N., Sithithaworn, P., Kaewkes, S., Mitreva, M., Brindley, P.J., 2007. The bandit, a New DNA Transposon from a Hookworm—Possible Horizontal Genetic Transfer between Host and Parasite. *PLoS Negl. Trop. Dis.* 1, e35. doi:10.1371/journal.pntd.0000035
- Lai, C.-M., Lin, K.-Y., Kao, S.-H., Chen, Y.-N., Huang, F., Hsu, H.-J., 2017. Hedgehog signaling establishes



- precursors for germline stem cell niches by regulating cell adhesion. *J. Cell Biol.* 216, 1439 LP – 1453. doi:10.1083/jcb.201610063
- Lau, N.C., Robine, N., Martin, R., Chung, W., Niki, Y., Berezikov, E., Lai, E.C., 2009. Abundant primary piRNAs, endo-siRNAs, and microRNAs in a *Drosophila* ovary cell line. *Genome Res.* 1776–1785. doi:10.1101/gr.094896.109.In
- Lau, N.C., Seto, A.G., Kim, J., Kuramochi-Miyagawa, S., Nakano, T., Bartel, D.P., Kingston, R.E., 2006. Characterization of the piRNA Complex from Rat Testes. *Science* (80- ). 313, 363–367. doi:10.1126/science.1130164
- Le Thomas, A., Stuwe, E., Li, S., Du, J., Marinov, G., Rozhkov, N., Chen, Y.-C.A., Luo, Y., Sachidanandam, R., Toth, K.F., Patel, D., Aravin, A.A., 2014. Transgenerationally inherited piRNAs trigger piRNA biogenesis by changing the chromatin of piRNA clusters and inducing precursor processing. *Genes Dev.* 28, 1667–1680. doi:10.1101/gad.245514.114
- Leblanc, P., Dasset, S., Dastugue, B., Vaury, C., 1997. Invertebrate retroviruses: ZAM a new candidate in *D. melanogaster*. *EMBO J.* 16, 7521–7531. doi:10.1093/emboj/16.24.7521
- Lee, E.J., Banerjee, S., Zhou, H., Jammalamadaka, A., Arcila, M., Manjunath, B.S., Kosik, K.S., 2011. Identification of piRNAs in the central nervous system. *RNA* 17, 1090–1099. doi:10.1261/rna.2565011
- Lerat, E., Capy, P., 1999. Retrotransposons and retroviruses: analysis of the envelope gene. *Mol. Biol. Evol.* 16, 1198–1207. doi:10.1093/oxfordjournals.molbev.a026210
- Levis, R.W., Ganesan, R., Houtchens, K., Tolar, L.A., Sheen, F.M., 1993. Transposons in place of telomeric repeats at a *Drosophila* telomere. *Cell* 75, 1083–93.
- Lewis, S.H., Quarles, K.A., Yang, Y., Tanguy, M., Frézal, L., Smith, S.A., Sharma, P.P., Cordaux, R., Gilbert, C., Giraud, I., Collins, D.H., Zamore, P.D., Miska, E.A., Sarkies, P., Jiggins, F.M., 2018. Pan-arthropod analysis reveals somatic piRNAs as an ancestral defence against transposable elements. *Nat. Ecol. Evol.* 2, 174–181. doi:10.1038/s41559-017-0403-4
- Li, M.A., Alls, J.D., Avancini, R.M., Koo, K., Godt, D., 2003. The large Maf factor traffic jam controls gonad morphogenesis in *Drosophila*. *Nat. Cell Biol.* 5, 994–1000. doi:10.1038/ncb1058
- Li, X.Z., Roy, C.K., Dong, X., Bolcun-Filas, E., Wang, J., Han, B.W., Xu, J., Moore, M.J., Schimenti, J.C., Weng, Z., Zamore, P.D., 2013. An Ancient Transcription Factor Initiates the Burst of piRNA Production during Early Meiosis in Mouse Testes. *Mol. Cell* 50, 67–81. doi:10.1016/j.molcel.2013.02.016
- Lim, A.K., Kai, T., 2007. Unique germ-line organelle, nuage, functions to repress selfish genetic elements in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci.* 104, 6714–6719. doi:10.1073/pnas.0701920104
- Lim, A.K., Lorthongpanich, C., Chew, T.G., Tan, C.W.G., Shue, Y.T., Balu, S., Gounko, N., Kuramochi-

- Miyagawa, S., Matzuk, M.M., Chuma, S., Messerschmidt, D.M., Solter, D., Knowles, B.B., 2013. The nuage mediates retrotransposon silencing in mouse primordial ovarian follicles. *Development* 140, 3819–3825. doi:10.1242/dev.099184
- Lin, H., Spradling, A.C., 1997. A novel group of pumilio mutations affects the asymmetric division of germline stem cells in the *Drosophila* ovary. *Development* 124, 2463–76.
- Lingel, A., Simon, B., Izaurralde, E., Sattler, M., 2003. Structure and nucleic-acid binding of the *Drosophila* Argonaute 2 PAZ domain. *Nature* 426, 465–469. doi:10.1038/nature02123
- Lisch, D., 2013. How important are transposons for plant evolution? *Nat. Rev. Genet.* 14, 49–61. doi:10.1038/nrg3374
- Liu, Q., Rand, T.A., Kalidas, S., Du, F., Kim, H.-E., Smith, D.P., Wang, X., 2003. R2D2, a Bridge Between the Initiation and Effector Steps of the *Drosophila* RNAi Pathway. *Science* (80-. ). 301, 1921–1925. doi:10.1126/science.1088710
- Maksakova, I.A., Romanish, M.T., Gagnier, L., Dunn, C.A., Van De Lagemaat, L.N., Mager, D.L., 2006. Retroviral elements and their hosts: Insertional mutagenesis in the mouse germ line. *PLoS Genet.* 2, 1–10. doi:10.1371/journal.pgen.0020002
- Malone, C.D., Brennecke, J., Dus, M., Stark, A., McCombie, W.R., Sachidanandam, R., Hannon, G.J., 2009. Specialized piRNA Pathways Act in Germline and Somatic Tissues of the *Drosophila* Ovary. *Cell* 137, 522–535. doi:10.1016/j.cell.2009.03.040
- Melnikova, L., Biessmann, H., Georgiev, P., 2005. The Ku protein complex is involved in length regulation of *Drosophila* telomeres. *Genetics* 170, 221–235. doi:10.1534/genetics.104.034538
- Melom, J.E., Littleton, J.T., 2013. Mutation of a NCKX eliminates glial microdomain calcium oscillations and enhances seizure susceptibility. *J. Neurosci.* 33, 1169–1178. doi:10.1523/JNEUROSCI.3920-12.2013
- Mével-Ninio, M., Pelisson, A., Kinder, J., Campos, A.R., Bucheton, A., 2007. The flamenco locus controls the gypsy and ZAM retroviruses and is required for *Drosophila* oogenesis. *Genetics* 175, 1615–1624. doi:10.1534/genetics.106.068106
- Mi, S., Lee, X., Li, X., Veldman, G.M., Finnerty, H., Racie, L., LaVallie, E., Tang, X.-Y., Edouard, P., Howes, S., Keith, J.C., McCoy, J.M., 2000. Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. *Nature* 403, 785–789. doi:10.1038/35001608
- Miesen, P., Girardi, E., van Rij, R.P., 2015. Distinct sets of PIWI proteins produce arbovirus and transposon-derived piRNAs in *Aedes aegypti* mosquito cells. *Nucleic Acids Res.* 43, 6545–6556. doi:10.1093/nar/gkv590
- Miki, Y., Nishisho, I., Horii, A., Miyoshi, Y., Utsunomiya, J., Kinzler, K.W., Vogelstein, B., Nakamura, Y.,

1992. Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a colon cancer. *Cancer Res.* 52, 643–5.
- Mirkovic-Hösle, M., Förstemann, K., 2014. Transposon Defense by Endo-siRNAs, piRNAs and Somatic piRNAs in *Drosophila*: Contributions of Loqs-PD and R2D2. *PLoS One* 9, e84994. doi:10.1371/journal.pone.0084994
- Mita, P., Boeke, J.D., 2016. How retrotransposons shape genome regulation. *Curr. Opin. Genet. Dev.* 37, 90–100. doi:10.1016/j.gde.2016.01.001
- Miyoshi, T., Takeuchi, A., Siomi, H., Siomi, M.C., 2010. A direct role for Hsp90 in pre-RISC formation in *Drosophila*. *Nat. Struct. Mol. Biol.* 17, 1024–1026. doi:10.1038/nsmb.1875
- Mochizuki, K., Fine, N.A., Fujisawa, T., Gorovsky, M.A., 2002. Analysis of a piwi-related gene implicates small RNAs in genome rearrangement in *Tetrahymena*. *Cell* 110, 689–699. doi:10.1016/S0092-8674(02)00909-1
- Mochizuki, K., Gorovsky, M.A., 2005. A Dicer-like protein in *Tetrahymena* has distinct functions in genome rearrangement, chromosome segregation, and meiotic prophase. *Genes Dev.* 19, 77–89. doi:10.1101/gad.1265105
- Mohn, F., Handler, D., Brennecke, J., 2015. piRNA-guided slicing specifies transcripts for Zucchini-dependent, phased piRNA biogenesis. *Science* 348, 812–817. doi:10.1126/science.aaa1039
- Mohn, F., Sienski, G., Handler, D., Brennecke, J., 2014. The Rhino-Deadlock-Cutoff Complex Licenses Noncanonical Transcription of Dual-Strand piRNA Clusters in *Drosophila*. *Cell* 157, 1364–1379. doi:10.1016/j.cell.2014.04.031
- Morazzani, E.M., Wiley, M.R., Murreddu, M.G., Adelman, Z.N., Myles, K.M., 2012. Production of Virus-Derived Ping-Pong-Dependent piRNA-like Small RNAs in the Mosquito Soma. *PLoS Pathog.* 8, e1002470. doi:10.1371/journal.ppat.1002470
- Moteki, S., Price, D., 2002. Functional coupling of capping and transcription of mRNA. *Mol. Cell* 10, 599–609. doi:10.1016/S1097-2765(02)00660-3
- Muerdter, F., Olovnikov, I., Molaro, A., Rozhkov, N. V., Czech, B., Gordon, A., Hannon, G.J., Aravin, A.A., 2012. Production of artificial piRNAs in flies and mice. *RNA* 18, 42–52. doi:10.1261/rna.029769.111
- Mugat, B., Akkouche, A., Serrano, V., Armenise, C., Li, B., Brun, C., Fulga, T.A., Van Vactor, D., Pélisson, A., Chambeyron, S., 2015. MicroRNA-Dependent Transcriptional Silencing of Transposable Elements in *Drosophila* Follicle Cells. *PLOS Genet.* 11, e1005194. doi:10.1371/journal.pgen.1005194
- Nagao, A., Mituyama, T., Huang, H., Chen, D., Siomi, M.C., Siomi, H., 2010. Biogenesis pathways of piRNAs loaded onto AGO3 in the *Drosophila* testis. *RNA* 16, 2503–2515. doi:10.1261/rna.2270710

- Niki, Y., Yamaguchi, T., Mahowald, A.P., 2006. Establishment of stable cell lines of *Drosophila* germ-line stem cells. *Proc. Natl. Acad. Sci.* 103, 16325–16330. doi:10.1073/pnas.0607435103
- Nishida, K.M., Saito, K., Mori, T., Kawamura, Y., Nagami-Okada, T., Inagaki, S., Siomi, H., Siomi, M.C., 2007. Gene silencing mechanisms mediated by Aubergine piRNA complexes in *Drosophila* male gonad. *RNA* 13, 1911–1922. doi:10.1261/rna.744307
- Oetting, W.S., Fryer, J.P., Shriram, S., King, R.A., 2003. Oculocutaneous albinism type 1: The last 100 years. *Pigment Cell Res.* 16, 307–311. doi:10.1034/j.1600-0749.2003.00045.x
- Okamura, K., Ishizuka, A., Siomi, H., Siomi, M.C., 2004. Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways. *Genes Dev.* 18, 1655–1666. doi:10.1101/gad.1210204
- Okamura, K., Lai, E.C., 2008. Endogenous small interfering RNAs in animals. *Nat. Rev. Mol. Cell Biol.* 9, 673–678. doi:10.1038/nrm2479
- Olivieri, D., Senti, K.-A., Subramanian, S., Sachidanandam, R., Brennecke, J., 2012. The cochaperone shutdown defines a group of biogenesis factors essential for all piRNA populations in *Drosophila*. *Mol. Cell* 47, 954–969. doi:10.1016/j.molcel.2012.07.021
- Olivieri, D., Sykora, M.M., Sachidanandam, R., Mechtler, K., Brennecke, J., 2010. An in vivo RNAi assay identifies major genetic and cellular requirements for primary piRNA biogenesis in *Drosophila*. *EMBO J.* 29, 3301–3317. doi:10.1038/emboj.2010.212
- Ozata, D.M., Gainetdinov, I., Zoch, A., O'Carroll, D., Zamore, P.D., 2019. PIWI-interacting RNAs: small RNAs with big functions. *Nat. Rev. Genet.* 20, 89–108. doi:10.1038/s41576-018-0073-3
- Palkopoulou, E., Mallick, S., Skoglund, P., Enk, J., Rohland, N., Li, H., Omrak, A., Vartanyan, S., Poinar, H., Götherström, A., Reich, D., Dalén, L., 2015. Complete Genomes Reveal Signatures of Demographic and Genetic Declines in the Woolly Mammoth. *Curr. Biol.* 25, 1395–1400. doi:10.1016/j.cub.2015.04.007
- Pandey, R.R., Homolka, D., Chen, K.-M., Sachidanandam, R., Fauvarque, M.-O., Pillai, R.S., 2017. Recruitment of Armitage and Yb to a transcript triggers its phased processing into primary piRNAs in *Drosophila* ovaries. *PLOS Genet.* 13, e1006956.
- Pane, A., Jiang, P., Zhao, D.Y., Singh, M., Schüpbach, T., 2011. The Cutoff protein regulates piRNA cluster expression and piRNA production in the *Drosophila* germline. *EMBO J.* 30, 4601–4615. doi:10.1038/emboj.2011.334
- Parisot, N., Pelin, A., Gasc, C., Polonais, V., Belkorchia, A., Panek, J., El Alaoui, H., Biron, D.G., Brasset, E., Vaury, C., Peyret, P., Corradi, N., Peyretailade, É., Lerat, E., 2014. Microsporidian genomes harbor a diverse array of transposable elements that demonstrate an ancestry of horizontal exchange with metazoans. *Genome Biol. Evol.* 6, 2289–300. doi:10.1093/gbe/evu178

- Park, D., Hadžić, T., Yin, P., Rusch, J., Abruzzi, K., Rosbash, M., Skeath, J.B., Panda, S., Sweedler, J. V., Taghert, P.H., 2011. Molecular organization of *Drosophila* neuroendocrine cells by Dimmed. *Curr. Biol.* 21, 1515–1524. doi:10.1016/j.cub.2011.08.015
- Pecinka, A., Dinh, H.Q., Baubec, T., Rosa, M., Lettner, N., Scheid, O.M., 2010. Epigenetic Regulation of Repetitive Elements Is Attenuated by Prolonged Heat Stress in *Arabidopsis*. *Plant Cell* 22, 3118–3129. doi:10.1105/tpc.110.078493
- Péligon, A., Song, S.U., Prud'homme, N., Smith, P.A., Bucheton, A., Corces, V.G., 1994a. 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, 4401–11.
- Péligon, A., Song, S.U., Prud'homme, N., Smith, P.A., Bucheton, A., Corces, V.G., 1994b. 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, 4401–11.
- Perrat, P.N., DasGupta, S., Wang, J., Theurkauf, W., Weng, Z., Rosbash, M., Waddell, S., 2013. Transposition-driven genomic heterogeneity in the *Drosophila* brain. *Science* 340, 91–95. doi:10.1126/science.1231965
- Peters, L., Meister, G., 2007. Argonaute Proteins: Mediators of RNA Silencing. *Mol. Cell* 26, 611–623. doi:10.1016/J.MOLCEL.2007.05.001
- Petit, M., Mongelli, V., Frangeul, L., Blanc, H., Jiggins, F., Saleh, M.-C., 2016. piRNA pathway is not required for antiviral defense in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. U. S. A.* 113, E4218–E4227. doi:10.1073/pnas.1607952113
- Picard, G., 1976. Non-mendelian female sterility in *Drosophila melanogaster*: hereditary transmission of I factor. *Genetics* 83, 107–23.
- Pickar-Oliver, A., Gersbach, C.A., 2019. The next generation of CRISPR–Cas technologies and applications. *Nat. Rev. Mol. Cell Biol.* doi:10.1038/s41580-019-0131-5
- Praher, D., Zimmermann, B., Genikhovich, G., Columbus-Shenkar, Y., Modepalli, V., Aharoni, R., Moran, Y., Technau, U., 2017. Characterization of the piRNA pathway during development of the sea anemone *Nematostella vectensis*. *RNA Biol.* 14, 1727–1741. doi:10.1080/15476286.2017.1349048
- Preall, J.B., Czech, B., Guzzardo, P.M., Muerdter, F., Hannon, G.J., 2012. shutdown is a component of the *Drosophila* piRNA biogenesis machinery. *RNA* 18, 1446–57. doi:10.1261/rna.034405.112
- Prud'homme, N., Gans, M., Masson, M., Terzian, C., Bucheton, A., 1995. Flamenco, a gene controlling the gypsy retrovirus of *Drosophila melanogaster*. *Genetics* 139, 697–711.
- Quesneville, H., Bergman, C.M., Andrieu, O., Autard, D., Nouaud, D., Ashburner, M., Anxolabehere, D., 2005. Combined evidence annotation of transposable elements in genome sequences. *PLoS Comput. Biol.* 1,

0166–0175. doi:10.1371/journal.pcbi.0010022

- Rajasethupathy, P., Antonov, I., Sheridan, R., Frey, S., Sander, C., Tuschl, T., Kandel, E.R., 2012. A role for neuronal piRNAs in the epigenetic control of memory-related synaptic plasticity. *Cell* 149, 693–707. doi:10.1016/j.cell.2012.02.057
- Rangan, P., Malone, C.D., Navarro, C., Newbold, S.P., Hayes, P.S., Sachidanandam, R., Hannon, G.J., Lehmann, R., 2011. piRNA Production Requires Heterochromatin Formation in *Drosophila*. *Curr. Biol.* 21, 1373–1379. doi:10.1016/j.cub.2011.06.057
- Reddien, P.W., Oviedo, N.J., Jennings, J.R., Jenkin, J.C., Sánchez Alvarado, A., 2005. SMEDWI-2 Is a PIWI-Like Protein That Regulates Planarian Stem Cells. *Science* (80-. ). 310, 1327–1330. doi:10.1126/science.1116110
- Reuss, F.U., Frankel, W.N., Moriwaki, K., Shiroishi, T., Coffin, J.M., 1996. Genetics of Intracisternal-A-Particle-Related Envelope-Encoding Proviral Elements in Mice, *JOURNAL OF VIROLOGY*.
- Reuter, M., Berninger, P., Chuma, S., Shah, H., Hosokawa, M., Funaya, C., Antony, C., Sachidanandam, R., Pillai, R.S., 2011. Miwi catalysis is required for piRNA amplification-independent LINE1 transposon silencing. *Nature* 480, 264–267. doi:10.1038/nature10672
- Rogers, A.K., Situ, K., Perkins, E.M., Toth, K.F., 2017. Zucchini-dependent piRNA processing is triggered by recruitment to the cytoplasmic processing machinery. *Genes Dev.* 31, 1858–1869. doi:10.1101/gad.303214.117
- Rogers, R.L., Slatkin, M., 2017. Excess of genomic defects in a woolly mammoth on Wrangel island. *PLOS Genet.* 13, e1006601. doi:10.1371/journal.pgen.1006601
- Rouget, C., Papin, C., Boueux, A., Meunier, A.-C., Franco, B., Robine, N., Lai, E.C., Pelisson, A., Simonelig, M., 2010. Maternal mRNA deadenylation and decay by the piRNA pathway in the early *Drosophila* embryo. *Nature* 467, 1128–1132. doi:10.1038/nature09465
- Rubio, M., Bustamante, M., Hernandez-Ferrer, C., Fernandez-Orth, D., Pantano, L., Sarria, Y., Piqué-Borras, M., Vellve, K., Agramunt, S., Carreras, R., Estivill, X., Gonzalez, J.R., Mayor, A., 2018. Circulating miRNAs, isomiRs and small RNA clusters in human plasma and breast milk. *PLoS One* 13, e0193527. doi:10.1371/journal.pone.0193527
- Ruby, J.G., Jan, C., Player, C., Axtell, M.J., Lee, W., Nusbaum, C., Ge, H., Bartel, D.P., 2006. Large-Scale Sequencing Reveals 21U-RNAs and Additional MicroRNAs and Endogenous siRNAs in *C. elegans*. *Cell* 127, 1193–1207. doi:10.1016/j.cell.2006.10.040
- Saito, K., Inagaki, S., Mituyama, T., Kawamura, Y., Ono, Y., Sakota, E., Kotani, H., Asai, K., Siomi, H., Siomi, M.C., 2009a. A regulatory circuit for piwi by the large Maf gene traffic jam in *Drosophila*. *Nature* 461, 1296–1299. doi:10.1038/nature08501

- Saito, K., Ishizu, H., Komai, M., Kotani, H., Kawamura, Y., Nishida, K.M., Siomi, H., Siomi, M.C., 2010. Roles for the Yb body components Armitage and Yb in primary piRNA biogenesis in *Drosophila*. Supplemental Material: Supplemental Materials and Methods. *Genes Dev.* 19, 2493–2498. doi:10.1101/gad.1989510.GENES
- Saito, K., Nishida, K.M., Mori, T., Kawamura, Y., Miyoshi, K., Nagami, T., Siomi, H., 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. doi:10.1101/gad.1454806
- Sarkies, P., Selkirk, M.E., Jones, J.T., Blok, V., Boothby, T., Goldstein, B., Hanelt, B., Ardila-Garcia, A., Fast, N.M., Schiffer, P.M., Kraus, C., Taylor, M.J., Koutsovoulos, G., Blaxter, M.L., Miska, E.A., 2015. Ancient and Novel Small RNA Pathways Compensate for the Loss of piRNAs in Multiple Independent Nematode Lineages. *PLOS Biol.* 13, e1002061. doi:10.1371/journal.pbio.1002061
- Schnable, P.S., Ware, D., Fulton, R.S., Stein, J.C., Wei, F., Pasternak, S., Liang, C., Zhang, J., Fulton, L., Graves, T.A., Minx, P., Reily, A.D., Courtney, L., Kruchowski, S.S., Tomlinson, C., Strong, C., Delehaunty, K., Fronick, C., Courtney, B., Rock, S.M., Belter, E., Du, F., Kim, K., Abbott, R.M., Cotton, M., Levy, A., Marchetto, P., Ochoa, K., Jackson, S.M., Gillam, B., Chen, W., Yan, L., Higginbotham, J., Cardenas, M., Waligorski, J., Applebaum, E., Phelps, L., Falcone, J., Kanchi, K., Thane, T., Scimone, A., Thane, N., Henke, J., Wang, T., Ruppert, J., Shah, N., Rotter, K., Hodges, J., Ingenthron, E., Cordes, M., Kohlberg, S., Sgro, J., Delgado, B., Mead, K., Chinwalla, A., Leonard, S., Crouse, K., Collura, K., Kudrna, D., Currie, J., He, R., Angelova, A., Rajasekar, S., Mueller, T., Lomeli, R., Scara, G., Ko, A., Delaney, K., Wissotski, M., Lopez, G., Campos, D., Braidotti, M., Ashley, E., Golser, W., Kim, H., Lee, S., Lin, J., Dujmic, Z., Kim, W., Talag, J., Zuccolo, A., Fan, C., Sebastian, A., Kramer, M., Spiegel, L., Nascimento, L., Zutavern, T., Miller, B., Ambroise, C., Muller, S., Spooner, W., Narechania, A., Ren, L., Wei, S., Kumari, S., Faga, B., Levy, M.J., McMahan, L., Van Buren, P., Vaughn, M.W., Ying, K., Yeh, C.-T., Emrich, S.J., Jia, Y., Kalyanaraman, A., Hsia, A.-P., Barbazuk, W.B., Baucom, R.S., Brutnell, T.P., Carpita, N.C., Chaparro, C., Chia, J.-M., Deragon, J.-M., Estill, J.C., Fu, Y., Jeddelloh, J.A., Han, Y., Lee, H., Li, P., Lisch, D.R., Liu, S., Liu, Z., Nagel, D.H., McCann, M.C., SanMiguel, P., Myers, A.M., Nettleton, D., Nguyen, J., Penning, B.W., Ponnala, L., Schneider, K.L., Schwartz, D.C., Sharma, A., Soderlund, C., Springer, N.M., Sun, Q., Wang, H., Waterman, M., Westerman, R., Wolfgruber, T.K., Yang, L., Yu, Y., Zhang, L., Zhou, S., Zhu, Q., Bennetzen, J.L., Dawe, R.K., Jiang, J., Jiang, N., Presting, G.G., Wessler, S.R., Aluru, S., Martienssen, R.A., Clifton, S.W., McCombie, W.R., Wing, R.A., Wilson, R.K., 2009. The B73 Maize Genome: Complexity, Diversity, and Dynamics. *Science* (80- ). 326, 1112 LP – 1115. doi:10.1126/science.1178534
- Schnettler, E., Donald, C.L., Human, S., Watson, M., Siu, R.W.C., McFarlane, M., Fazakerley, J.K., Kohl, A., Frangkoudis, R., 2013. Knockdown of piRNA pathway proteins results in enhanced Semliki Forest virus production in mosquito cells. *J. Gen. Virol.* 94, 1680–9. doi:10.1099/vir.0.053850-0
- Senti, K.-A., Brennecke, J., 2010. The piRNA Pathway : Guardian of the Genome – A Fly ’ s Perspective. *Trends Genet.* 26, 499–509. doi:10.1016/j.tig.2010.08.007.

- Senti, K.-A., Jurczak, D., Sachidanandam, R., Brennecke, J., 2015. piRNA-guided slicing of transposon transcripts enforces their transcriptional silencing via specifying the nuclear piRNA repertoire. *Genes Dev.* 29, 1747–62. doi:10.1101/gad.267252.115
- Shapiro, J., Machattie, L., Eron, L., Ihler, G., Ippen, K., Beckwith, J., 1969. Isolation of pure lac operon DNA. *Nature* 224, 768–74.
- Sharma, A.K., Nelson, M.C., Brandt, J.E., Wessman, M., Mahmud, N., Weller, K.P., Hoffman, R., 2001. Human CD34(+) stem cells express the hiwi gene, a human homologue of the Drosophila gene piwi. *Blood* 97, 426–34.
- Shatkin, A.J., Manley, J.L., 2000. The ends of the affair: Capping and polyadenylation. *Nat. Struct. Biol.* 7, 838–842. doi:10.1038/79583
- Sienski, G., Dönertas, D., Brennecke, J., 2012. Transcriptional silencing of transposons by Piwi and maelstrom and its impact on chromatin state and gene expression. *Cell* 151, 964–980. doi:10.1016/j.cell.2012.10.040
- Siomi, M.C., Sato, K., Pezic, D., Aravin, A.A., 2011. PIWI-interacting small RNAs: the vanguard of genome defence. *Nat. Rev. Mol. Cell Biol.* 12, 246–258. doi:10.1038/nrm3089
- Song, J.J., Smith, S.K., Hannon, G.J., Joshua-Tor, L., 2004. Crystal structure of argonaute and its implications for RISC slicer activity. *Science* (80-. ). 305, 1434–1437. doi:10.1126/science.1102514
- Spradling, A.C. (1993). Developmental genetics of oogenesis. *Bate, Martinez Arias, 1993* : 1--70.
- Strout, M.P., Marcucci, G., Bloomfield, C.D., Caligiuri, M.A., 2002. The partial tandem duplication of ALL1 (MLL) is consistently generated by Alu-mediated homologous recombination in acute myeloid leukemia. *Proc. Natl. Acad. Sci.* 95, 2390–2395. doi:10.1073/pnas.95.5.2390
- Tam, O.H., Aravin, A.A., Stein, P., Girard, A., Murchison, E.P., Cheloufi, S., Hodges, E., Anger, M., Sachidanandam, R., Schultz, R.M., Hannon, G.J., 2008. Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. *Nature* 453, 534–538. doi:10.1038/nature06904
- Théron, E., Dennis, C., Brassat, E., Vaury, C., 2014. Distinct features of the piRNA pathway in somatic and germ cells: From piRNA cluster transcription to piRNA processing and amplification. *Mob. DNA* 5, 1–11. doi:10.1186/s13100-014-0028-y
- Tittel-Elmer, M., Bucher, E., Broger, L., Mathieu, O., Paszkowski, J., Vaillant, I., 2010. Stress-Induced Activation of Heterochromatic Transcription. *PLoS Genet.* 6, e1001175. doi:10.1371/journal.pgen.1001175
- Tomari, Y., Matranga, C., Haley, B., Martinez, N., Zamore, P.D., 2004. A Protein Sensor for siRNA Asymmetry. *Science* (80-. ). 306, 1377–1380. doi:10.1126/science.1102755



- Tosar, J.P., Rovira, C., Cayota, A., 2018. Non-coding RNA fragments account for the majority of annotated piRNAs expressed in somatic non-gonadal tissues. *Commun. Biol.* 1, 2. doi:10.1038/s42003-017-0001-7
- Tuchin, V. V., Tárnok, A., Zharov, V.P., 2011. In vivo flow cytometry: a horizon of opportunities. *Cytometry. A* 79, 737–745. doi:10.1002/cyto.a.21143
- Vagin, V. V., Sigova, A., Li, C., Seitz, H., Gvozdev, V., Zamore, P.D., 2006. A Distinct Small RNA Pathway Silences Selfish Genetic Elements in the Germline. *Science* (80-. ). 313, 320–324. doi:10.1126/science.1129333
- Vagin, V. V., Yu, Y., Jankowska, A., Luo, Y., Wasik, K.A., Malone, C.D., Harrison, E., Rosebrock, A., Wakimoto, B.T., Fagegaltier, D., Muerdter, F., Hannon, G.J., 2013. Minotaur is critical for primary piRNA biogenesis. *RNA* 19, 1064–77. doi:10.1261/rna.039669.113
- van den Beek, M., da Silva, B., Pouch, J., Ali Chaouche, M.E.A., Carré, C., Antoniewski, C., 2018. Dual-layer transposon repression in heads of *Drosophila melanogaster*. *RNA* 24, 1749–1760. doi:10.1261/rna.067173.118
- Van Doren, M., Broihier, H.T., Moore, L.A., Lehmann, R., 1998. HMG-CoA reductase guides migrating primordial germ cells. *Nature* 396, 466–469. doi:10.1038/24871
- van Gent, D.C., Mizuuchi, K., Gellert, M., 1996. Similarities between initiation of V(D)J recombination and retroviral integration. *Science* 271, 1592–4.
- Volff, J.-N., 2006. Turning junk into gold: domestication of transposable elements and the creation of new genes in eukaryotes. *BioEssays* 28, 913–922. doi:10.1002/bies.20452
- Watanabe, T., Cheng, E.C., Zhong, M., Lin, H., 2015. Retrotransposons and pseudogenes regulate mRNAs and lncRNAs via the piRNA pathway in the germline. *Genome Res.* 25, 368–380. doi:10.1101/gr.180802.114
- Watanabe, T., Takeda, A., Tsukiyama, T., Mise, K., Okuno, T., Sasaki, H., Minami, N., 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, 1732–1743. doi:10.1101/gad.1425706
- 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., Sasaki, H., 2008. Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature* 453, 539–543. doi:10.1038/nature06908
- Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J.L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., Paux, E., SanMiguel, P., Schulman, A.H., 2007. A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* 8, 973–982. doi:10.1038/nrg2165
- Williams, Z., Morozov, P., Juranek, S., Tuschl, T., Williams, Z., Morozov, P., Mihailovic, A., Lin, C., Puvvula,

- P.K., Juranek, S., 2015. Resource Discovery and Characterization of piRNAs in the Human Fetal Ovary. *CellReports* 13, 854–863. doi:10.1016/j.celrep.2015.09.030
- Williamson, V.M., 1983. Transposable Elements in Yeast. *Int. Rev. Cytol.* 83, 1–25. doi:10.1016/S0074-7696(08)61684-8
- Willmann, M.R., Endres, M.W., Cook, R.T., Gregory, B.D., 2011. The Functions of RNA-Dependent RNA Polymerases in Arabidopsis. *Arab. B.* 9, e0146. doi:10.1199/tab.0146
- Wu, W.-S., Brown, J.S., Chen, T.-T., Chu, Y.-H., Huang, W.-C., Tu, S., Lee, H.-C., 2019. piRTarBase: a database of piRNA targeting sites and their roles in gene regulation. *Nucleic Acids Res.* 47, D181–D187. doi:10.1093/nar/gky956
- Xiao, H., Jiang, N., Schaffner, E., Stockinger, E.J., van der Knaap, E., 2008. A Retrotransposon-Mediated Gene Duplication Underlies Morphological Variation of Tomato Fruit. *Science (80- )*. 319, 1527–1530. doi:10.1126/science.1153040
- Xu, J., Gridley, T., 2012. Notch Signaling during Oogenesis in *Drosophila melanogaster*. *Genet. Res. Int.* 2012, 648207. doi:10.1155/2012/648207
- Yamanaka, S., Siomi, M.C., Siomi, H., 2014. piRNA clusters and open chromatin structure. *Mob. DNA* 5, 22. doi:10.1186/1759-8753-5-22
- Yan, K.S., Yan, S., Farooq, A., Han, A., Zeng, L., Zhou, M.-M., 2003. Structure and conserved RNA binding of the PAZ domain. *Nature* 426, 469–474. doi:10.1038/nature02129
- Yan, Z., Hu, H.Y., Jiang, X., Maierhofer, V., Neb, E., He, L., Hu, Y., Hu, H., Li, N., Chen, W., Khaitovich, P., 2011. Widespread expression of piRNA-like molecules in somatic tissues. *Nucleic Acids Res.* 39, 6596–6607. doi:10.1093/nar/gkr298
- Yang, P., Wang, Y., Macfarlan, T.S., 2017. The Role of KRAB-ZFPs in Transposable Element Repression and Mammalian Evolution. *Trends Genet.* 33, 871–881. doi:10.1016/j.tig.2017.08.006
- Yang, Q., Hua, J., Wang, L., Xu, B., Zhang, H., Ye, N., Zhang, Z., Yu, D., Cooke, H.J., Zhang, Y., Shi, Q., 2013. MicroRNA and piRNA Profiles in Normal Human Testis Detected by Next Generation Sequencing. *PLoS One* 8, e66809. doi:10.1371/journal.pone.0066809
- Yasuda, K., Ito, M., Sugita, T., Tsukiyama, T., Saito, H., Naito, K., Teraishi, M., Tanisaka, T., Okumoto, Y., 2013. Utilization of transposable element mPing as a novel genetic tool for modification of the stress response in rice. *Mol. Breed.* 32, 505–516. doi:10.1007/s11032-013-9885-1
- Zaessinger, S., Busseau, I., Simonelig, M., 2006. Oskar allows nanos mRNA translation in *Drosophila* embryos by preventing its deadenylation by Smaug/CCR4. *Development* 133, 4573 LP – 4583. doi:10.1242/dev.02649

- Zamparini, A.L., Davis, M.Y., Malone, C.D., Vieira, E., Zavadil, J., Sachidanandam, R., Hannon, G.J., Lehmann, R., 2011. Vreteno, a gonad-specific protein, is essential for germline development and primary piRNA biogenesis in *Drosophila*. *Development* 138, 4039–4050. doi:10.1242/dev.069187
- Zanni, V., Eymery, A., Coiffet, M., Zytnicki, M., Luyten, I., Quesneville, H., Vaury, C., Jensen, S., 2013. Distribution, evolution, and diversity of retrotransposons at the flamenco locus reflect the regulatory properties of piRNA clusters. *Proc. Natl. Acad. Sci. U. S. A.* 110, 19842–19847. doi:10.1073/pnas.1313677110
- Zhang, F., Wang, J., Xu, J., Zhang, Z., Koppetsch, B.S., Schultz, N., Vreven, T., Meignin, C., Davis, I., Zamore, P.D., Weng, Z., Theurkauf, W.E., 2012. UAP56 couples piRNA clusters to the perinuclear transposon silencing machinery. *Cell* 151, 871–884. doi:10.1016/j.cell.2012.09.040
- Zhang, P., Kang, J.-Y., Gou, L.-T., Wang, J., Xue, Y., Skogerboe, G., Dai, P., Huang, D.-W., Chen, R., Fu, X.-D., Liu, M.-F., He, S., 2015. MIWI and piRNA-mediated cleavage of messenger RNAs in mouse testes. *Cell Res.* 25, 193.
- Zhang, Z., Wang, J., Schultz, N., Zhang, F., Parhad, S.S., Tu, S., Vreven, T., Zamore, P.D., Weng, Z., Theurkauf, W.E., 2014. The HP1 Homolog Rhino Anchors a Nuclear Complex that Suppresses piRNA Precursor Splicing. *Cell* 157, 1353–1363. doi:10.1016/j.cell.2014.04.030
- Zheng, K., Xiol, J., Reuter, M., Eckardt, S., Leu, N.A., McLaughlin, K.J., Stark, A., Sachidanandam, R., Pillai, R.S., Wang, P.J., 2010. Mouse MOV10L1 associates with Piwi proteins and is an essential component of the Piwi-interacting RNA (piRNA) pathway. *Proc. Natl. Acad. Sci.* 107, 11841–11846. doi:10.1073/pnas.1003953107