



Analyses des régulations épigénétiques des éléments transposables chez *Drosophila melanogaster*

Jérémy Dufourt

► To cite this version:

Jérémy Dufourt. Analyses des régulations épigénétiques des éléments transposables chez *Drosophila melanogaster*. Biologie moléculaire. Université d'Auvergne - Clermont-Ferrand I, 2012. Français. NNT : 2012CLF1MM03 . tel-01167755

HAL Id: tel-01167755

<https://theses.hal.science/tel-01167755>

Submitted on 24 Jun 2015

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.

UNIVERSITE BLAISE PASCAL

UNIVERSITE D'Auvergne

Année 2012

N° d'ordre :

**ECOLE DOCTORALE
DES SCIENCES DE LA VIE ET DE LA SANTE**

N° d'ordre :

Thèse

Présentée à l'Université d'Auvergne
pour l'obtention du grade de Docteur d'Université

Spécialité : Biologie Moléculaire et Génétique

Dufourt JérémY

**Analyses des régulations épigénétiques des éléments transposables
chez *Drosophila melanogaster***

Soutenue publiquement le 04 mai 2012 devant le jury d'examen :

Président : M. Pierre Verrelle

**Rapporteurs : M. Frederic Bantignies
Mme Martine Simonelig**

**Membres : M. Eric Meyer
M. Stephane Ronsseray
Mme Chantal Vaury**

UMR CNRS 6293- Inserm 1103- Génétique Reproduction et Développement

UNIVERSITE BLAISE PASCAL

UNIVERSITE D'Auvergne

Année 2012

N° d'ordre :

**ECOLE DOCTORALE
DES SCIENCES DE LA VIE ET DE LA SANTE**

N° d'ordre :

Thèse

Présentée à l'Université d'Auvergne
pour l'obtention du grade de Docteur d'Université

Spécialité : Biologie Moléculaire et Génétique

Dufourt JérémY

**Analyses des régulations épigénétiques des éléments transposables
chez *Drosophila melanogaster***

Soutenue publiquement le 04 mai 2012 devant le jury d'examen :

Président : M. Pierre Verrelle

**Rapporteurs : M. Frederic Bantignies
Mme Martine Simonelig**

**Membres : M. Eric Meyer
M. Stephane Ronsseray
Mme Chantal Vaury**

UMR CNRS 6293- Inserm 1103- Génétique Reproduction et Développement

Résumé

Résumé

Les éléments transposables (ET) sont des séquences d'ADN trouvées chez tous les organismes vivants, et capables de se déplacer d'un site chromosomique à un autre. Ils sont une source de mutations et doivent donc être finement contrôlés par leurs hôtes. Afin de parer à leur mobilisation, les génomes ont mis en place des mécanismes de régulation (RNAi) impliquant des petits ARNs dont les siRNAs en sont la composante la plus connue. Récemment il a été mis en évidence chez la drosophile deux nouvelles classes de petits ARNs appelés piRNAs et endo-siRNAs. Les piRNAs contrôlent spécifiquement les ET dans les tissus reproducteurs qui comprennent des cellules germinales et des cellules somatiques : les cellules folliculaires. Les endo-siRNAs quant à eux, contrôlent ces ET dans les tissus somatiques. Il a été montré au laboratoire qu'Idefix, un retrotransposon à LTR, était régulé par un mécanisme post-transcriptionnel (PTGS). Celui-ci implique la voie des piRNAs et sa composante majeure, la protéine PIWI, dans les tissus reproducteurs de drosophiles. En revanche dans le reste des tissus de drosophiles cette régulation ne dépend pas de la protéine PIWI. Durant ma thèse je me suis intéressé à savoir si en plus de ce contrôle de type PTGS, il existait une régulation de type transcriptionnelle (TGS) appliquée sur les ET de drosophile dans les différents tissus aussi bien somatiques que germinaux. En étudiant les régulations que subissent divers transgènes composés d'un gène rapporteur et de divers fragments d'ET, j'ai montré que seule une régulation de type post-transcriptionnelle permettait de réguler les éléments transposables dans la lignée germinale femelle de drosophile. Cette régulation ayant une faiblesse précoce dans le développement des ovaires pouvant entraîner une mobilisation des éléments transposables dans certaines conditions sensibilisées. En revanche dans les tissus somatiques j'ai montré qu'une régulation transcriptionnelle s'ajoutait à la répression de type PTGS pour réprimer les ET. Cependant cette régulation transcriptionnelle présente une spécificité tissulaire puisqu'elle est observée dans les tissus somatiques de larves de drosophiles et absente dans les cellules somatiques folliculaires de l'ovaire. En conclusion divers systèmes de régulation mettent sous silence les éléments transposables en fonctions de la balance bénéfice/problèmes qu'ils apportent pour un tissu donné.

Abstract

Transposable elements (TEs) are DNA sequences found in all living organisms, and able to move from one chromosomal site to another. They are source of mutations and therefore must be finely controlled by their hosts. To counteract their mobilization, genomes have developed regulatory mechanisms (RNAi) involving small RNAs including the best-known siRNAs. Recently two novel classes of small RNAs called piRNAs and endo-siRNAs have been reported in *Drosophila*. The piRNAs specifically trigger TE repression in reproductive tissues, composed by germ cells and somatic follicular cells. The endo-siRNAs control those in somatic tissues. It has been shown by our group that Idefix, a LTR retrotransposon, is regulated by a posttranscriptional mechanism (PTGS). It implicates the piRNAs pathway and its major component, the PIWI protein, in reproductive tissues of *Drosophila*. By contrast, in the other *Drosophila* tissues, the regulation does not depend on the PIWI protein. During my PhD, I was interested to know if in addition to this PTGS, a transcriptional control (TGS) was necessary to control *Drosophila* TE in both the somatic and germinal tissues. By studying the regulations of sensor-transgenes carrying a reporter gene (GFP) and various fragments of ET acting as a target of the silencing pathways, I have shown that the post-transcriptional silencing is the only regulatory pathway targeting transposable elements in the *Drosophila* female germline. This regulation has a weakness early in the development of the ovaries that can lead to a mobilization of transposable elements under certain sensitized conditions. In somatic tissues I have shown that a transcriptional regulation is coupled to the PTGS. However, this transcriptional regulation has tissue specificity because it is only observed in somatic tissues of *Drosophila* outside of the ovaries, a PTGS with no TGS targeting TE in the somatic cells of the ovarian follicle.

Remerciements

C'est une tradition que de remercier au début d'un tel travail tous ceux qui, plus ou moins directement, ont contribué à le rendre possible. Ce n'est pas par habitude mais avec une réelle sincérité que je voudrais dire Merci à tous ceux qui à leur manière m'ont aidé à mener à bien cette thèse.

Ensuite comme j'aimerais rester fidèle à moi même, les fautes d'orthographe ne seront pas corrigées dans cette partie et sauf omission grave de ma part les personnes non remerciées n'ont pas contribué à cette thèse voir pour certains y ont été contreproductif.

Ce travail a été réalisé au laboratoire de Génétique Reproduction et Développement. Une unité de recherche qui donne à ses étudiants et chercheurs des moyens techniques formidables pour une recherche d'excellence et tout ça dans une soit disant petite ville. Pour cela je remercie le Professeur Bernard Dastugue, qui est l'un des instigateurs de cette dynamique de recherche clermontoise et Chantal Vaury qui à malgré les difficultés (nombreuses et encore je ne connais pas tout) réussis à créer cette unité de recherche qui est le GReD

Je remercie les rapporteurs de cette thèse Martine Simonelig et Frédéric Bantignies d'avoir accepté de juger ce travail. Merci également aux autres membres du jury qui ont accepté de juger ce travail Eric Meyer, Stéphane Ronsseray et Pierre Verrelle.

Merci Chantal Vaury pour m'avoir acceptée dans ton équipe et guidé pendant ces quatre ans, tout en me laissant une grande part de liberté. Comment dire en quelques mots tout ce que tu m'as apporté, merci pour ta passion communicative, pour avoir su m'écouter et me rassurer dans mes moments de doute, merci pour ta curiosité et ton esprit critique. Merci de m'avoir soutenus et remis à ma place afin que je ne m'égare pas. J'associe à ces remerciements Emilie Brasset car tu m'a apporté beaucoup sur le plan humain et intellectuel durant ces années et ce en ayant deux p'tits garçons pendant ce temps.

Merci à toute mon équipe et plus particulièrement, Silke Jensen toujours très critique mais fabuleusement gentille, Sophie Desset ingénieure de choc spécialiste de la récup dans les poubelles, sans toi le cahier des erreurs de labo serait vide. Nathalie Gueguen une vraie maman dans l'âme vraiment un grand merci à toi nath pour tout. Merci aux étudiants passés et présents : Rana, Colline, Mickael et Cynthia. Merci à mes petites étudiantes que j'ai eu en stage entre autres Marion et Dana, d'avoir réussi à supporter mon caractère de M....

Il est certain que ce travail n'aurait pu aboutir sans l'aide d'un grand nombre de personnes : je pense particulièrement à Eric merci pour ta bonne humeur, Agostina et bien sûr à la pièce maîtresse de ce laboratoire Fanfan, sans qui beaucoup de personnes au laboratoire se trouveraient démunies, Nadia pour ton efficacité et tout le personnel administratif : Frank et Marie –Diane.

Merci tout particulièrement à l'ensemble du groupe Droso/souris, pour ses réunions plus ou moins animées du jeudi, faites que ça perdure. Je pense bien sûr à Jean Louis, Muriel, Christophe, Vincent, Claire, Olivier, et le « petit nouveau » Philippe.

Merci à Jean Philippe pour avoir toujours des femelles vierges sous la main et pour ta bonne humeur et ta gaieté. Merci Cyril pour dire que tout va bien et que le temps est au beau fixe même sous le déluge.

Merci à tout les doctorants et post-doctorants présents et passés: Clothilde, Sylvain, Hervé (je mise sur le futur), Parvathy, François, Maxime, David...

Merci aux informaticiens, Yohan et Pierre. Pierre un deuxième merci pour tout ce que tu as fait sur le plan humain et scientifique durant ma fin de thèse.

Merci aux anciens Nicolas, tu resteras une de mes plus belles rencontres dans ce labo, je crois que ta bonne humeur manque cruellement au GReD. Merci à Laetitia et Caroline deux très belles rencontres humaines et scientifiques, continuez la science je vous embauche dès que je peux. Merci à Vanessa T et Sandrine pour votre soutien, vos délires et le temps de discussion sur la passerelle.

Merci à l'ensemble de ma famille pour tout simplement avoir été présents. Merci à Bruno mon « papa » adoptif, merci pour ton esprit critique et ton ouverture d'esprit. Je pense que tu m'as montré que quand on veut on peut.

Merci à mes deux frères Timothée et Thomas dont je suis si fier, merci pour cette complicité qui nous unit, merci pour les soirées détente sur le Xbox live et pour les délires de guerrier made in Tim. J'espère que vous réussirez à vous faire plaisir dans votre métier comme je le fais moi. Aussi bien toi Tim avec des enfants en difficultés que toi Tom dans la science (choisis la biologie c'est mieux que la chimie...).

Merci à ma maman pour tout cet amour et ce soutien qu'elle m'a apporté au cours de ces années. Merci de m'avoir toujours guidée et conseillée sans décider mais ne crois surtout pas que ta contribution s'arrête là, je continuerai à t'importuner encore longtemps. Encore un merci car tu n'en aura jamais trop, vu tout ce que tu as pu faire pour moi.

Merci à Tatiana pour tout, pour ta compréhension et ta patience sans limite depuis toutes ces années, et je sais qu'il en fallait. Merci pour ton soutien et ton amour quotidien car sans ça je n'aurais pas tenu durant toutes ces années. Merci de m'avoir toujours épaulé dans les moments de désespoir et calmé dans les moments d'euphories.

Une pensée pour mon chat PIWI qui nous a quitté le 9 mars 2012, sans toi je me sent un peu plus seul et surtout Aubergine ta copine de jeux qui se retrouve toute seule et miaule sans arrêt (dur d'écrire sa thèse avec des miaulements en musique de fond). Aubergine sans PIWI la voie des piRNAs est cassée alors je m'en vais vers de nouvelles aventures...

Index figures

Figure 1: Hétérochromatine et Euchromatine

Figure 2: Pourcentage d'éléments transposables dans divers génomes

Figure 3: Structure schématique des classes d'éléments transposables

Figure 4: Deux mécanismes de transposition des éléments transposables

Figure 5: Mécanisme de transposition des rétrotransposons à LTR

Figure 6: Mécanisme de transposition des rétrotransposons sans LTR

Figure 7: Mécanisme de transposition des transposons

Figure 8: Transposition répllicative et non répllicative

Figure 9 : Réarrangements intra-chromosomiques liés aux éléments transposables

Figure 10 : Réarrangements inter-chromosomiques liés aux éléments transposables

Figure 11 : Exemple d'apport bénéfique des éléments transposables: le système immunitaire acquis

Figure 12 : Co-suppression chez le pétunia

Figure 13: Exemples d'ARN interférence après injection d'ARN double brin dans différents modèles

Figure 14: Domaines protéiques conservés des protéines Dicer et Argonaute

Figure 15: Structure du complexe Dicer/ARNdb

Figure 16: Modèle pour le mécanisme de la voie des siRNAs

Figure 17: Identification des petits ARNs endogènes liés à Ago2

Figure 18: Divers type de génération de siRNAs endogènes (endo-siRNAs)

Figure 19: Protéines impliquées dans la voie des endo-siRNAs

Figure 20: Voie canonique des endo-siRNAs

Figure 21: Exemple de structure de pri-miRNA

Figure 22: Voie canonique des micro-RNAs

Figure 23: Production de miRNAs indépendante de Dicer

Figure 24: Théorie des cibles tampon des microRNAs

Figure 25: Caractéristiques des piRNAs de drosophile

Figure 26: Localisation des piRNAs sur le chromosome X de la drosophile

Figure 27: La voie canonique des piRNAs chez la drosophile

Figure 28: Biais de brin des piRNAs

Figure 29: piRNAs primaires et secondaires chez la drosophile

Figure 30: Le modèle *Ping-Pong* chez la souris

Figure 31: piRNAs primaires et secondaires chez la souris

Figure 32: Méthylations des arginines des protéines Aubergine et Argonaute 3

Figure 33: Divers protéines à domaine Tudor de drosophile et de souris

Figure 34: Méthylation des piRNAs

Figure 35: Le nuage chez la drosophile et la souris

Figure 36: Représentation des différents niveaux de compaction de l'ADN

Figure 37: Formation et organisation des histones

Figure 38: Principales modifications post-traductionnelles des histones

Figure 39: Profil d'expression des Histones Méthyl Transférase H3K9 dans l'ovaire de drosophile.

Figure 40: Les complexes de remodelage de la chromatine

Figure 41: Structure de la protéine hétérochromatique 1 (HP1) et du gène qui la code

Figure 42: Différentes isoformes de HP1 chez la drosophile

Figure 43: Distribution de HP1a

Figure 44: Distribution de la méthylation sur la lysine 9 de l'histone H3

Figure 45: Présence de HP1c sur des régions transcriptionnellement actives

Figure 46: Liste des protéines pouvant interagir avec HP1

Figure 47: Expression et rôle de Rhino chez la drosophile

Figure 48: Différents complexes impliquant les protéines du groupe Polycomb PcG

Figure 49: Conservation des principaux PcG chez les mammifères

Figure 50: Reconnaissance des Polycomb et Trithorax Response Element (PRE et TRE)

Figure 51: Les complexes du groupe Trithorax (TrxG)

Figure 52: Structure des méthyle-transférases de l'ADN (Dnmt)

Figure 53: Implication de la méthylation de l'ADN et des modifications de la chromatine dans le verrouillage de l'expression génique

Figure 54: Différents types de chromatine de drosophile

Figure 55: Enrichissement en diverses modifications d'histones des différents types de chromatine

Figure 56: Les chromatines rouges et jaunes

Figure 57: La chromatine noire

Figure 58: Organisation de la chromatine dans le noyau

Figure 59: Organisation des chromosomes à l'intérieur du noyau

Figure 60: Interactions physiques entre gènes, appelées phénomène de « gene kissing »

Figure 61: Les foyers Polycomb

Figure 62: Hi-C et différentes chromatines

Figure 63: RNAi et foyers Polycomb

Figure 64: Interaction entre la protéine hétérochromatique HP1 et PIWI

Figure 65: Méthylation ADN et piRNAs chez la souris

Figure 66: Réarrangement du génome des ciliés et PcG

Figure 67: Représentation de l'appareil reproducteur de la drosophile femelle

Figure 68: Représentation schématique d'une paire d'ovaire et d'un ovariole

Figure 69: Représentations schématiques d'un germarium et de follicules ovariens

Figure 70: Structure des allèles wIR6 , wIR6RevI et wIR6RevII et phénotype des yeux associé à ces allèles

Figure 71: Structure génomique d'*Idefix*

Figure 72 : Distribution génomique des séquences d'*Idefix*.

Figure 73: Expression d'*Idefix* en contexte mutant pour COM

Figure 74: Expression de la GFP à partir de transgènes portant un fragment d'*Idefix* en aval

Figure 75: Détection de petits ARNs antisens d'*Idefix*.

Figure 76: La régulation d'*Idefix* est dépendante de la protéine PIWI

Figure 77 : Distribution génomique des séquences d'*Idefix*.

Figure 78: La voie des piRNAs en cellules folliculaires de drosophile

Figure 79: Organisation du cluster de piRNAs *Flam*/COM

Figure 80: pUASp et pUAS_t dans les ovaires de drosophile

Figure 81: Structure des transgènes utilisés pour l'étude d'Idefix en ligné germinale

Figure 82: Expression d'Idefix en contexte mutant pour COM

Figure 83 : Trois classes d'éléments transposables régulés par les piRNAs

Figure 84: piRNA ciblant Idefix

Figure 85: Expression des transgènes pGgIdas et pGg^{DId} avec un activateur ubiquitaire Act-Gal4

Sommaire

INTRODUCTION 1

1) Les éléments transposables 2

a) Proportions dans les génomes 3

b) Classes d'éléments transposables et mécanismes de transposition 3

b).1 Les éléments de classe I 3

b).1.a Les éléments à LTR 4

b).1.b Les éléments sans LTR 6

b).2 Les éléments de classe II 8

c) Bénéfices et problèmes liés à l'insertion d'éléments transposables 11

c).1 Problèmes 11

c).2 Bénéfices 14

2) Régulation des éléments transposables 16

a) Les voies d'ARN interférence 16

a).1 La voie des petit ARNs interférents (siRNAs) 17

a).2 La voie des micros ARNs (miRNAs) 21

a).3 La voie des petits ARN se liant à PIWI (piRNAs) 26

b) Modifications de la chromatine 32

b).1 Les modifications covalentes d'histones 33

b).2 Les modifications non covalentes d'histones 35

c) Les protéines de compaction/décompaction de l'ADN 36

c).1 HP1 36

c).2 Les protéines des groupes Polycomb et Trithorax 40

d) <u>La méthylation de l'ADN</u>	44
e) <u>Organisation chromatinienne et nucléaire</u>	45
e).1 <u>Organisation chromatinienne</u>	45
e).2 <u>Les territoires chromosomiques</u>	47
e).3 <u>Interactions longues distances</u>	48
f) <u>RNAi, hétérochromatine et Organisation nucléaire</u>	49
3) <u>Le modèle d'étude</u>	51
a) <u>La drosophile</u>	51
b) <u>L'ovogénèse adulte de drosophile</u>	52
c) <u>Les éléments ZAM et Idefix</u>	54
c) .1 <u>historique</u>	54
c) .2 <u>Structure et expression de ZAM et Idefix</u>	55

<i>Résultats</i>	58
Publication 1	59
1) <u>Introduction</u>	59
2) <u>Analyse des éléments transposables euchromatiques</u>	60
3) <u>Conclusions</u>	61

Publication 2	64
----------------------	-----------

1) <u>Introduction</u>	<u>64</u>
------------------------	-----------

2) <u>Conclusions</u>	<u>66</u>
-----------------------	-----------

Publication 3	75
----------------------	-----------

1) <u>Introduction</u>	<u>75</u>
------------------------	-----------

2) <u>Conclusions</u>	<u>76</u>
-----------------------	-----------

<i>Conclusion</i>	77
-------------------	-----------

Références Bibliographiques

INTRODUCTION

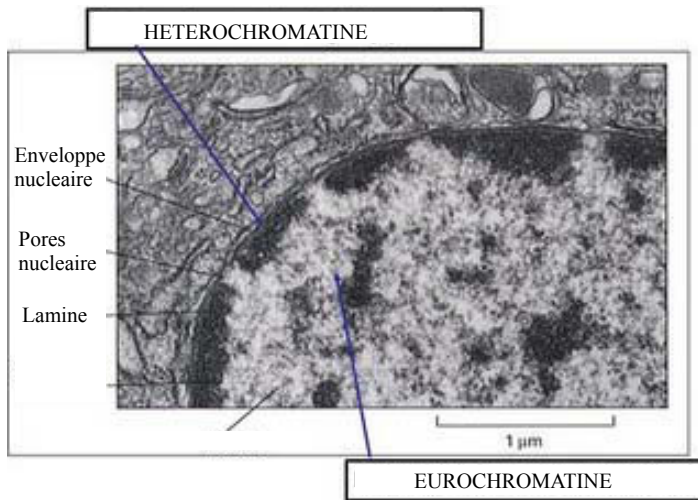


Figure 1: Hétérochromatine et Euchromatine

Image de microscopie électronique montrant des régions d'ADN denses en électrons (hétérochromatine) accrochées à la membrane nucléaire (Lamine) comparées aux régions beaucoup moins denses (euchromatine).[<http://genomedarkmatter.blogspot.com>]

INTRODUCTION

Le génome des eucaryotes qui ont été étudiés à ce jour est en grande partie constitué de séquences d'ADN dites non codantes. Ces analyses sont historiquement liées à l'étude de zones distinctes des chromosomes appelées hétérochromatine. En 1928 Heitz E (Heitz 1928) a observé des régions de chromosomes qui conservaient une structure condensée tout au long du cycle cellulaire mitotique (**Figure 1**) et qui furent ensuite montrées comme constituées majoritairement de séquences répétées. Cet ADN répétitif peut s'organiser en tandem, ou en séquences répétées dispersées et avoir une longueur variable.

On distingue principalement deux types de séquences répétées : a) les séquences hautement répétées en tandem comprenant : l'ADN satellite, minisatellite et microsatellite et b) les séquences moyennement répétées réparties sur l'ensemble du génome parmi lesquelles on distingue les éléments transposables.

1) Les éléments transposables

En caractérisant au début des années 1950 dans le génome du maïs une interaction entre deux *loci* génomiques nommés *Dissociator (Ds)* et *Activator (Ac)*, Barbara McClintock met en évidence l'existence de facteurs génomiques mobiles : les éléments transposables. En effet, elle démontre que *Ds* a la capacité de changer de position chromosomique en présence de *Ac* (McClintock 1956).

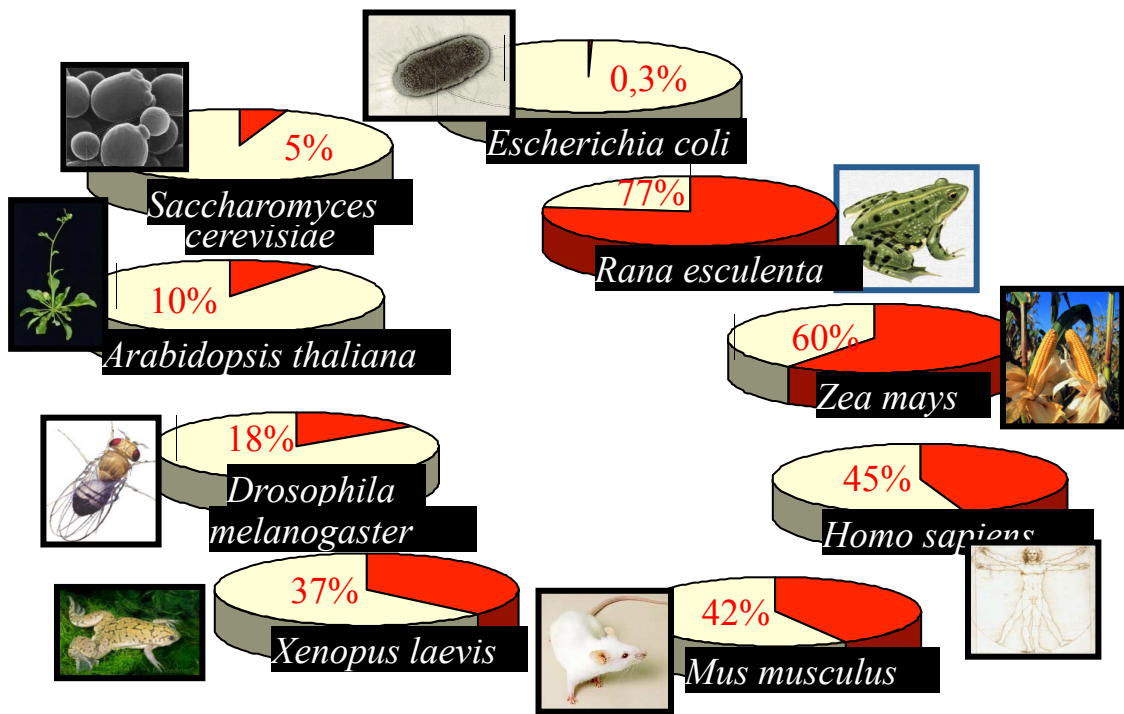


Figure 2: Pourcentage d'éléments transposables dans divers génomes

Les éléments transposables sont présents aussi bien dans les génomes procaryotes qu'eucaryotes mais en proportions très variables, indiquées en rouge. [Christian Biémont, Lyon]

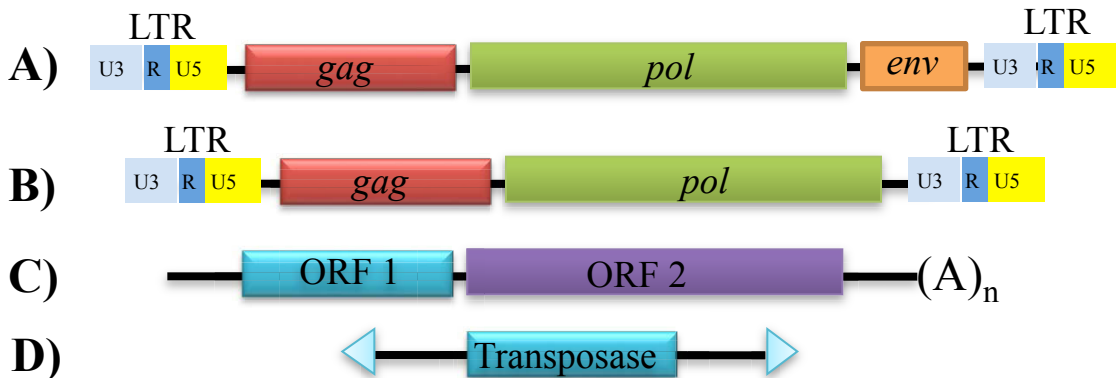


Figure 3: Structure schématique des classes d'éléments transposables

A) Rétrotransposon à LTR de type retrovirus, comprenant deux longues répétitions terminales directes (LTRs). Ces LTRs encadrent trois ORFs : *gag* (protéines de structure), *pol* (protéines de réplication) et *env* (protéines d'enveloppe)

B) Rétrotransposon à LTR, comprenant deux longues répétitions terminales directes. Ces LTRs encadrent deux ORFs : *gag* (protéines de structure) et *pol* (protéines de réplication).

C) Rétrotransposon sans LTR. Les rétrotransposons sans LTR possèdent une protéine de structure dans l'ORF1 (*Gag*) et une ORF2 pouvant coder diverses protéines nécessaires à la réplication. Ils sont terminés par une répétition poly-acétylée (A)_n.

A), B) et C) sont des éléments de classe I.

D) Transposons (Elément de classe II). Ils codent une transposase encadrée par deux répétitions inversées (triangles bleus).

a) Proportions dans les génomes

Les éléments transposables sont des séquences d'ADN capables de se déplacer d'un site chromosomique à un autre. Ce sont des composants naturels de presque tous les génomes séquencés aussi bien eucaryotes que procaryotes (**Figure 2**). Leur proportion dans les génomes est très variable, ils peuvent ne représenter qu'une faible proportion du génome, par exemple chez certaines bactéries (0,3% du génome de *Escherichia Coli*) (Mahillon and Chandler 1998) jusqu'à plus de 90% de certains génomes en particulier ceux de plantes. Ces éléments représentent par exemple respectivement 50% et 18% du génome de l'homme (Lander, Linton et al. 2001) et de *Drosophila melanogaster*.

b) Classes d'éléments transposables et mécanismes de transposition

On distingue principalement trois grandes familles d'éléments transposables en fonction de leur structure et de leur intermédiaire de transposition. Il existe ainsi les éléments à intermédiaire ARN (classe I), qui comprennent les rétrotransposons à LTR et les rétrotransposons, ainsi que les éléments à intermédiaire ADN (classe II) (**Figure 3**).

b).1 Les éléments de classe I

Les éléments transposables de classe I, qui ont une origine commune avec les virus à ARN (simple et double brin) (Ahlquist 2006) transposent *via* un intermédiaire ARN qui leur permet de "faxer" (Ronsseray, S métaphore pour étudiants M2 Génétique et physiologie) une copie d'eux même dans une autre région génomique. Ce système de transposition qui utilise un mode réplcatif dit "copier-coller" (**Figure 4**) permet d'augmenter grandement le nombre de copies de ces éléments dans les génomes. Ils représentent la grande majorité des éléments transposables. Les copies complètes codent une transcriptase inverse (RT) capable de synthétiser une molécule d'ADN

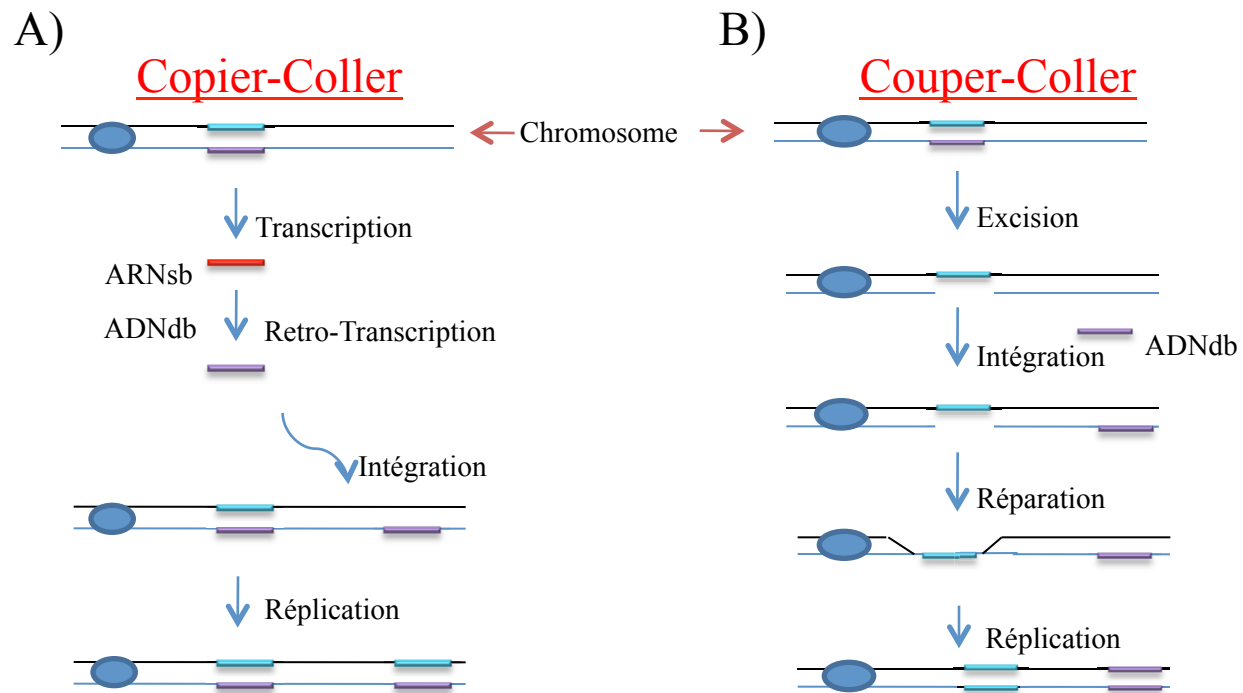


Figure 4: Deux mécanismes de transposition des éléments transposables

A) Mécanisme de transposition des éléments transposables de classe I: l'élément est transcrit en ARN simple brin (ARNsb) puis retro-transcrit en ADN double brin (ADNdb). L'ADNdb va s'insérer dans un nouveau locus génomique, sans excision de l'élément transposable de départ.

B) Mécanisme de transposition des éléments transposables de classe II: l'élément s'excise sous forme d'ADN double brin (ADNdb) et va se réinsérer dans un site génomique cible. La cassure laissée lors de son excision est réparée par recombinaison grâce à la chromatide sœur.

complémentaire (ADNc) à partir d'une matrice ARN simple brin et également une intégrase (IN) qui permet d'insérer le fragment d'ADNc *néo*-synthétisé dans un nouveau site chromosomique du génome de l'hôte. Bien que tous les chromosomes puissent être la cible d'une nouvelle insertion, la fréquence et la spécificité d'insertion varie dans le génome. Il en découle une distribution inégale le long des bras des chromosomes, reflet d'une préférence d'intégration pour certaines séquences consensus ou conformation structurale de l'ADN cible (Cost and Boeke 1998) (Jurka 1997; Jurka, Klonowski et al. 1998).

De nombreuses familles d'éléments de classe I ont été caractérisées chez la plupart des espèces eucaryotes mais aussi chez les procaryotes.

Selon leur structure génomique, deux groupes de retrotransposons ont été mis en évidence : ceux contenant des répétitions longues terminales (LTR) qui ont une structure de type provirale et ceux qui sont dépourvus de LTR qui ont une structure de type pseudogène.

b).1.a Les éléments à LTR

Les rétrotransposons à répétitions longues terminales (LTR) ont été détectés dans des génomes aussi variés que ceux de la levure (Ty) (Klein and Petes 1984), des insectes (éléments de type copia) (Biemont 1992), des plantes (Tnt1) (Grandbastien, Audeon et al. 1994) et de la souris (IAP) (Amariglio and Rechavi 1993).

La structure des rétrotransposons à LTR a une organisation similaire à celle des rétrovirus (VIH, MMLV) (**Figure 3**). Ils présentent à chaque extrémité, des LTR identiques constitués des séquences U3, R, U5 retrouvées en orientation directe. Ils comprennent aussi en général deux cadres de lecture ouverts (ORF) codant pour diverses protéines, qui sont *gag* et *pol*. Les protéines de type Pol sont nécessaires à la réplication

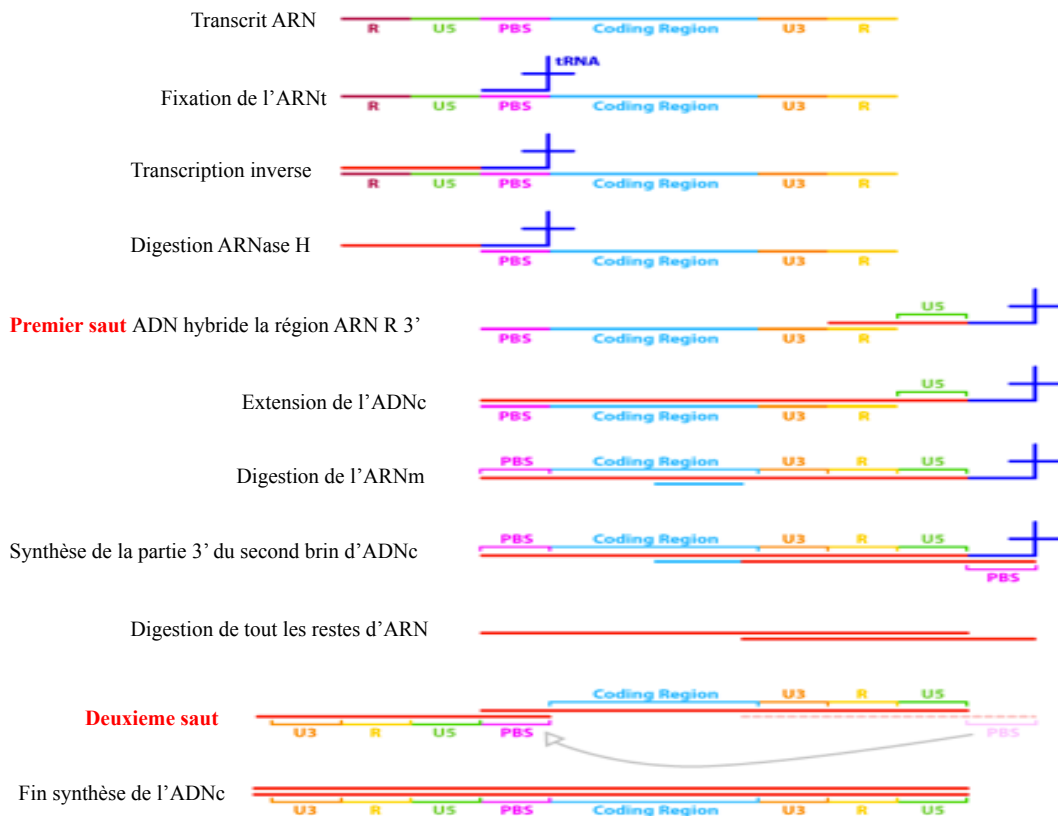


Figure 5: Mécanisme de transposition des rétrotransposons à LTR

L'amorçage du cycle de la transcription inverse est assuré par la fixation d'un ARNt dans la partie PBS (**Primer Binding Site**) du U5 du LTR 5'. La synthèse par la transcriptase inverse s'effectue jusqu'à l'extrémité 5' du LTR, ce qui produit une molécule d'ADN d'environ 200 pb. La transcriptase inverse, par son activité ARNase H, dégrade alors l'ARN complémentaire jusqu'à la partie R. La partie 3' de la copie d'ADN ainsi libérée s'hybride alors sur la partie complémentaire du LTR 3' **premier saut**. La synthèse du premier brin d'ADN est ainsi complète par extension de l'ADNc. Puis l'ARNm va être dégradé en partie sauf un fragment qui va servir d'amorce à la synthèse de la partie 3' du second brin d'ADNc. Puis **deuxième saut** hybridation entre les deux PBS. L'activité ADN polymérase de la transcriptase inverse permet la synthèse des deux brin d'ADN. On obtient une copie d'ADN linéaire à deux LTR. [<http://studentreader.com/interspersed-repeats/>]

(Intégrase, Transcriptase inverse, Protéase) et celles de type Gag codent pour les protéines de capsid. Chez certains rétrotransposons à LTR (Idefix, et Gypsy de *Drosophila melanogaster* par exemple) (Marlor, Parkhurst et al. 1986; Desset, Conte et al. 1999) il y a présence d'un troisième cadre de lecture *env*, codant pour les protéines d'enveloppe (**Figure 3**).

L'ARNm permettant la synthèse des protéines de type Gag, Pol et Env débute dans la partie R du LTR 5', ce LTR possédant généralement un promoteur de l'ARN polymérase II. La partie codant pour les protéines Env est épissée et traduite indépendamment. Les mécanismes produisant les protéines Gag et Pol sont de deux types. Cela dépend si les deux cadres de lecture sont en phase ou en décalage de phase. Dans les deux cas, les ORF *gag* et *pol* sont traduits pour produire une seule polyprotéine qui sera ensuite maturée. Dans le premier cas où les cadres de lecture sont en phase, le codon stop se trouvant en fin du *gag* sera effacé par un ARNt-glutamyl, permettant ainsi la traduction de l'ARNm correspondant à l'ORF *pol*. Dans le deuxième cas, le décalage sera annulé par un saut du ribosome, ce qui se produit dans environ 5 % des cas. Ces deux mécanismes permettent la régulation de l'expression de *pol*. En effet, peu de protéines de type Pol sont nécessaires pour la réplication, tandis que plus de protéines Gag sont nécessaires à la formation de particules virales.

L'amorçage du cycle de la transcription inverse est assuré par la fixation d'un ARNt dans la partie PBS (Primer Binding Site) située en aval du LTR 5' (**Figure 5**). La synthèse par la transcriptase inverse s'effectue jusqu'à l'extrémité 5' du LTR, ce qui produit une molécule d'ADN d'environ 200 pb. La transcriptase inverse, par son activité ARNase H, dégrade alors l'ARN complémentaire jusqu'à la partie R. La partie 3' de la copie d'ADN ainsi libérée effectue un premier saut et s'hybride alors sur la partie complémentaire du LTR 3' (**Figure 5**). La synthèse du premier brin d'ADN est ainsi complétée par extension

de l'ADNc. Puis l'ARNm va être dégradé en partie sauf un fragment qui va servir d'amorce à la synthèse de la partie 3' du second brin d'ADNc. Puis un deuxième saut est effectué et il y a hybridation entre les deux PBS. L'activité ADN polymérase de la transcriptase inverse permet la synthèse des deux brins d'ADNc. On obtient une copie d'ADNc linéaire à deux LTR.

Dans certains cas, une copie circulaire à deux LTR est obtenue mais elle ne semble pas pouvoir se réintégrer. L'intégrase coupe de façon décalée au site d'insertion et assure la ligation des extrémités coupées aux deux LTR, générant ainsi une duplication du site cible de part et d'autre des LTR.

b).1.b Les éléments sans LTR

Ces éléments diffèrent des rétrotransposons car ils ne possèdent pas de LTR. On en distingue deux sous catégories: les LINEs (Long INterspersed Elements) capables de se déplacer de façon autonome dans les génomes et les SINEs (Short INterspersed Elements) qui sont beaucoup plus courts et qui ne peuvent transposer seuls.

Les éléments de type LINE complets ont une longueur d'environ 6 kb. Cependant, la majorité des éléments détectés dans les génomes sont délétés en 5' et ne sont plus capables de transposer. Ils présentent deux ORF (**Figure 3**) qui ne sont pas dans le même cadre de lecture. Ils codent pour des protéines présentant de fortes homologies avec les transcriptases inverses et intégrases. Ils se terminent en 3' par une séquence riche en adénine qui correspond au site de polyadénylation des transcrits. Quelques rares éléments LINEs sont fonctionnels dont un des exemples le plus probant est le facteur I de la drosophile.

La transcription des LINEs produit un ARN correspondant à la longueur totale de l'élément. Il code pour deux protéines, traduites à partir du même ARNm. Dans le cycle de

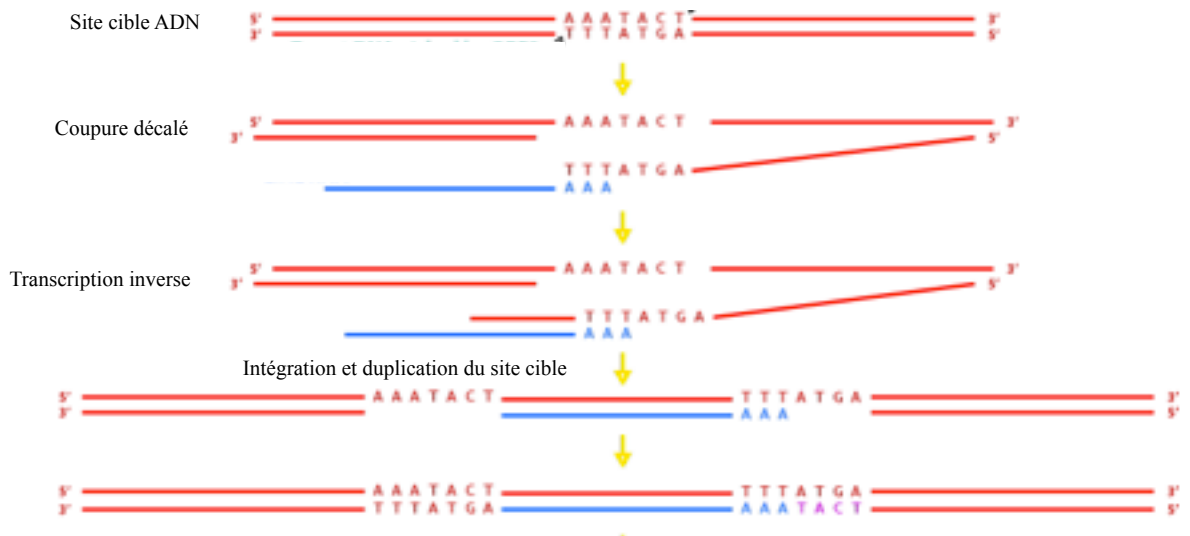


Figure 6: Mécanisme de transposition des rétrotransposons sans LTR

La transcription des LINEs produit un ARN correspondant à la longueur totale de l'élément (bleu). Dans le cycle de transposition, l'ARN est inversement transcrit simultanément à l'intégration de la copie ADN dans le génome. L'intégration se produit au niveau de coupures décalées et conduit à l'intégration d'une nouvelle copie et à la duplication du site cible. [<http://studentreader.com/interspersed-repeats/>]

transposition, l'ARN est inversement transcrit simultanément à l'intégration de la copie ADN dans le génome (**Figure 6**). L'intégration se produit au niveau de coupures décalées (Bucheton 1990). Ce mécanisme conduit à l'insertion d'un nouvel élément qui est très souvent délété en 5' car la transcriptase inverse se décroche de l'ARN en cours de synthèse. On peut les comparer aux pseudogènes qui peuvent être insérés dans le génome par des mécanismes analogues après épissage des introns. Ces éléments sont présents dans de nombreux génomes. Chez les plantes ils n'ont été détectés que dans quelques espèces dont le maïs (Bennetzen 1996), mais ils sont très représentés dans certains génomes, par exemple on peut en trouver jusqu'à 250 000 copies comme chez les mammifères.

Les éléments de type SINE ne possèdent pas dans leur structure de séquences codant pour une transcriptase inverse et sont donc incapables de transposer sans l'aide d'une transcriptase inverse fournie en trans. Ils ont été décrits dans le génome des mammifères où ils représentent une famille de séquences hautement répétées, environ 300 000 copies, ainsi que dans quelques plantes où ils sont très fortement représentés (jusqu'à 50 000 copies).

Les SINEs sont majoritairement transcrits par l'ARN polymérase III à partir d'un promoteur interne. Leur transcription inverse et leur réinsertion dans le génome doivent être assurées par des protéines se trouvant déjà dans le noyau. Ces enzymes pourraient être codées par d'autres éléments se trouvant ailleurs dans le génome. Les SINEs pourraient être générés accidentellement par l'intermédiaire de la machinerie répliquative des cellules à partir de précurseurs de type ARN qui seraient inversement transcrits et insérés dans le génome (Maestre, Tchenio et al. 1995).

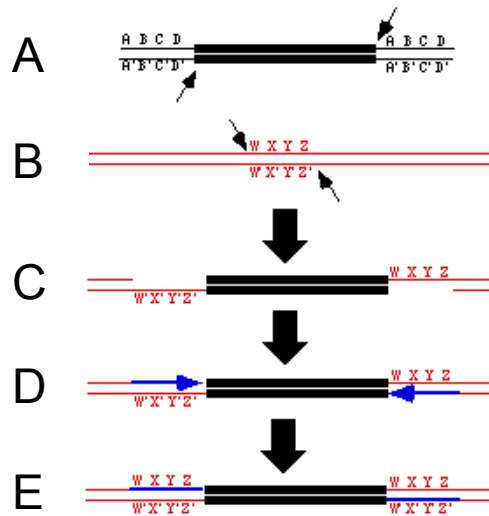


Figure 7: Mécanisme de transposition des transposons

A) La transposase de l'élément coupe l'extrémité 3' sur un brin d'ADN et l'extrémité 5' sur l'autre brin
B) La transposase effectue les mêmes coupures sur le site cible, **C)** l'extrémité 5' du transposon est liguée à l'extrémité 3' du site cible et inversement, **D)** Les nucléotides manquants sur le site cible dus à la coupure décalé sont synthétisés par l'ADN polymérase (flèches bleues), **E)** Le transposon est inséré dans son site cible et les séquences de part et d'autre de celui-ci sont dupliquées. [<http://www.sci.sdsu.edu/~smaloy/MicrobialGenetics/problems/transposons/transposition/>]

b).2 Les éléments de classe II

Les éléments de classe II se caractérisent par une transposition de type “couper-coller” (**Figure 4**) sans intermédiaire ARN. Ils codent pour une enzyme, la transposase, qui possède toutes les activités catalytiques nécessaires à la transposition de ces éléments. De plus, ils sont caractérisés par la présence d'ITR (Inverted Terminal Repeat) en orientation inverse, à chaque extrémité du gène codant la transposase (**Figure 3**). Ces ITR sont indispensables à la fixation de la transposase.

Chez les bactéries, on détecte les unités transposables les plus simples d'organisation, les séquences d'insertion (IS). Chez *E. coli*, environ 10 copies de chaque famille d'IS sont présentes dans le génome. Les IS sont des unités autonomes codant pour les protéines nécessaires à leur propre transposition. Chaque IS ne possède qu'un cadre de lecture ouvert codant pour la transposase. Les familles d'IS sont différentes en séquence mais présentent les mêmes caractéristiques dans leur organisation. Les IS sont de courtes séquences d'environ 1 000 pb présentant des répétitions inversées terminales (ITR) de 10 à 40 pb et, au point d'insertion, une duplication de la séquence cible. La séquence du site d'insertion varie pour une même IS, mais pas sa longueur. Cette séquence est dupliquée lors de l'intégration d'une IS, ce qui conduit à en trouver une copie à chaque extrémité de l'élément (**Figure 7**).

Des transposons plus complexes, appelés Tn, sont présents dans le génome bactérien. Ils ont la particularité d'être les vecteurs de séquences exogènes comme des gènes de résistance à des antibiotiques. Les Tn ont des structures composites, constituées de deux IS en répétition directe ou inverse, encadrant les séquences supplémentaires. Dans un Tn, l'IS permet sa propre transposition ou celle du complexe. Lorsque, dans un Tn, les deux IS sont différentes, chacune d'elles a la possibilité de mobiliser le Tn entier. Le choix

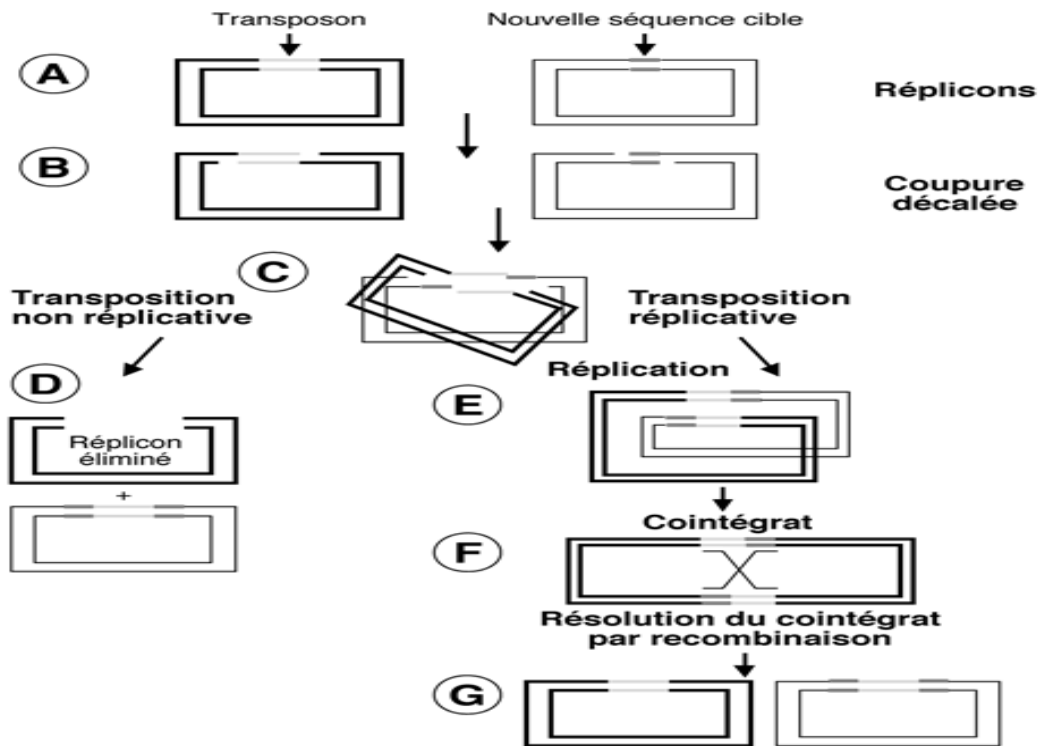


Figure 8: Transposition répliquative et non répliquative

A) Le transposon est intégré et le site de nouvelle intégration est sur un autre fragment d'ADN. **B)** La transposase produit une coupure décalée de part et d'autre du transposon et du site d'intégration. **C)** La transposase permet la formation d'une structure de type crossing-over, conduisant à l'échange de brins entre les deux fragment d'ADN. **D)** Dans la transposition non répliquative et conservative, les brins d'ADN du transposon donneur sont entièrement transférés au site accepteur. **E)** Lors de la transposition répliquative, la réplication conduit à la formation d'une seule molécule d'ADN circulaire, le co-intégrat, contenant un transposon à chaque extrémité des deux réplicons. **F)** La recombinaison homologue entre les transposons conduit à la résolution du cointégrat. **G)** Les deux fragments d'ADN alors obtenus contiennent chacun un transposon. [http://www.jle.com/fr/revues/bio_rech/vir/e-docs/00/03/F8/3B/article.phtml?fichier=images.htm]

de la transposition d'un Tn entier ou d'une IS seule se fait en fonction de la taille de la séquence située entre les deux IS : plus un Tn est court, plus il peut être mobilisé fréquemment en entier. D'autre part, selon la pression de sélection, le taux de transposition du complexe varie.

Deux modes de transposition des IS et donc des Tn ont été décrits. La transposition est soit répllicative et, dans ce cas, le transposon initial et sa copie transposée sont détectés dans le génome, soit non répllicative et conservative, et dans ce cas le transposon initial transpose et est détecté dans le génome hôte mais pas la copie originale (**Figure 8**).

Lors de la transposition répllicative, l'élément est copié. Cette copie de l'élément sera donc réinsérée. Les transposons de type Tn A transposent toujours de façon répllicative. Dans ce cas, la fusion entre un réplicon (molécule circulaire capable de répliquer) avec un transposon et un réplicon sans transposon aboutit à la formation d'un co-intégré. La transposase produit une coupure à un site précis de chaque côté du transposon ainsi qu'au site d'insertion. Les extrémités sont liées, ce qui produit une structure moléculaire de type crossing-over. La résolution par recombinaison de cette structure conduit à la formation de deux copies du transposon en orientation directe de part et d'autre de la jonction des deux réplicons dans le mécanisme réplcatif alors que dans le mécanisme non réplcatif seule une copie de l'élément transposable est retrouvée. La resolvase permet la résolution de ce complexe (**Figure 8**).

Chez les eucaryotes, des transposons proches des IS bactériens sont également présents et parfois très largement répandus selon les espèces. Leur organisation moléculaire ainsi que leurs modes de transposition sont très proches de ceux des transposons bactériens. Les mécanismes de transposition les plus étudiés sont ceux des éléments Ac/Ds chez le maïs, et P chez *Drosophila melanogaster*. Pour tous ces transposons, il existe dans un même génome des éléments complets et autonomes qui ont

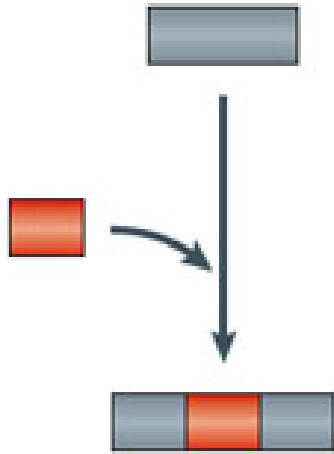
toutes les fonctions pour être mobiles, et des éléments non autonomes (défectifs) qui comportent des délétions plus ou moins importantes. Néanmoins, les éléments complets sont généralement capables de mobiliser en trans les éléments défectifs.

Chez le maïs, l'élément Ac est autonome tandis que les éléments Ds défectifs ne sont pas mobiles en l'absence d'Ac. Il existe environ 10 éléments par génome. L'élément Ac a une longueur de 4 563 pb. L'ARN de 3 500 pb transcrit à partir de cet élément comprend 5 exons. Il code pour une transposase de 807 aa. La transposition d'Ac n'est pas répliquative et le site donneur disparaît au cours de la transposition. Le site récepteur est fréquemment sur le même chromosome et très proche du site donneur.

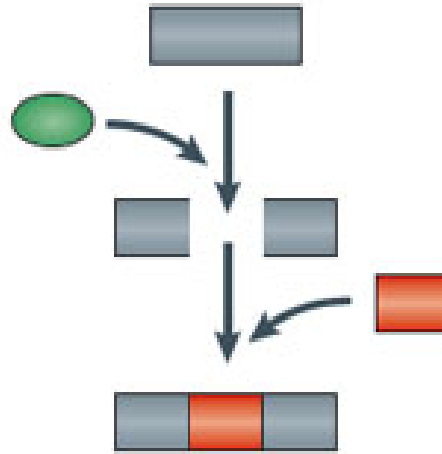
Chez *Drosophila melanogaster*, la transposition de l'élément P est à l'origine d'un phénomène de dysgénésie des hybrides. Le croisement des femelles de type M (Maternal) avec des mâles de cytotype P (Paternal) produit une descendance F1 stérile, ayant les gonades atrophiées. Le croisement réciproque produit des descendants fertiles. Cet élément a une longueur de 2,9 kb et possède, comme les IS, deux répétitions inversées terminales. Trente à cinquante copies sont présentes par souche de type P, les souches M en étant totalement dépourvues. L'élément P comporte 4 exons. Il est transcrit dans les cellules somatiques et germinales, mais ne transpose que dans la lignée germinale. La transposition de l'élément P est non répliquative et se produit par un mécanisme de couper-coller proche de celui des IS non réplcatifs.

Les éléments transposables, *via* leur capacité de transposition et leur caractère répétitif, conduisent à de multiples réarrangements chromosomiques comme par exemple des délétions de séquences ou des inversions, qui peuvent avoir des conséquences importantes sur l'expression des gènes adjacents. Ce sont les principales implications de ces modifications du génome que nous allons maintenant examiner.

a) Mutagenèse par insertion



b) Cassures double brin d'ADN



c) Nouveaux microsatellites



d) Conversion génique

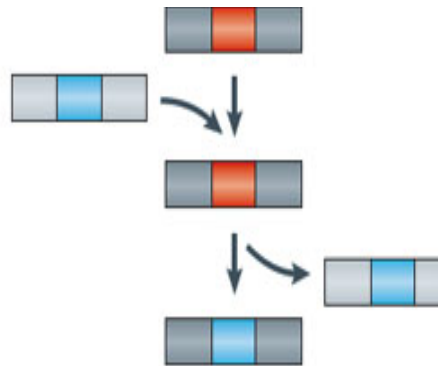


Figure 9 : Réarrangements intra-chromosomiques liés aux éléments transposables

a) La mutagenèse par insertion: un élément transposable (carré orange) peut s'insérer aléatoirement dans un site génomique (rectangle gris) et perturber sa régulation. **b)** Certaines protéines des ET ayant une activité endonucléase (rond vert) peuvent créer des cassures double brin dans le génome et donc conduire soit à une mauvaise réparation soit à une insertion d'un nouvel élément. **c)** La création de nouveaux microsatellites : En raison de leur abondance dans le génome, les éléments transposables peuvent générer des microsatellites à de nombreux loci dans le génome. La présence de régions répétées ectopiques en grand nombre peut dans certains cas générer des pathologies. **d)** La conversion génique: un élément transposable (carré bleu) contenu à un locus génomique (rectangle gris clair) peut se substituer à un autre élément (carré orange) situé à un autre locus génomique (gris foncé) sans qu'il n'y ait échange d'information. [Cordaux and Batzer 2009]

c) Bénéfices et problèmes liés à l'insertion d'éléments transposables

On ne connaît pas la fraction d'ADN répétitif bénéfique pour les organismes. Dans certains cas il peut y avoir un avantage au caractère répétitif (comme les séquences de l'ADN ribosomique (ADNr) répétées qui servent à produire rapidement des ARN ribosomiques (ARNr) lorsque cela est nécessaire. Mais la présence de ces séquences répétées en grand nombre ainsi que la capacité à se déplacer de certaines peut aussi conduire à divers problèmes. En 1988, ces éléments mobiles furent reconnus comme agents mutagènes capables de provoquer des pathologies chez l'homme lorsque, de façon indépendante, deux rétroéléments furent décrits comme responsables de plusieurs cas d'hémophilie suite à leur insertion dans le gène du Facteur VIII (Kazazian et al., 1988, Nature.). La présence d'ADN répétitif dans les génomes peut provoquer de nombreux réarrangements intra ou inter chromosomiques et ainsi provoquer diverses maladies.

c).1 Problèmes

La mutagenèse par insertion : en raison de l'impact phénotypique immédiat de ces insertions elles ont été les premières à être détectées. Cette mutagenèse est due à la capacité de mobilisation des éléments transposables qui peuvent altérer la fonction du génome, (et donc potentiellement influencer l'évolution du génome), en s'insérant dans les régions codantes des gènes (**Figure 9a**). Des exemples de troubles génétiques humains causés par une *néo*-insertion continuent à s'accumuler, et 65 cas ont été montrés pour causer des maladies héréditaires, comme l'hémophilie, la mucoviscidose, le syndrome de Apert, la neurofibromatose, la dystrophie musculaire de Duchenne, la β -thalassémie, l'hypercholestérolémie et des cancers du sein et du côlon. Par exemple chez l'homme,

l'insertion *de novo* d'un élément L1 est responsable de mutations dans le gène *myc* et est associée à un cancer du sein (Morse, Rotherg et al. 1988)

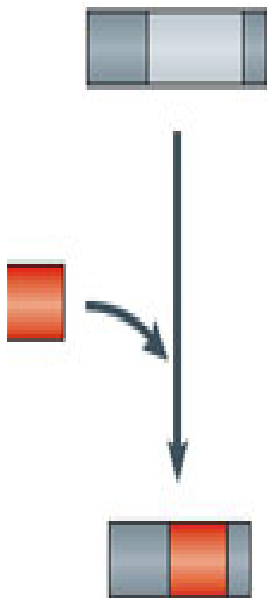
Globalement, il a été estimé qu'environ 0,3% des mutations du génome humain sont attribuables à des insertions *de novo* des éléments L1, Alu et SVA (Callinan and Batzer 2006).

Les cassures double brin d'ADN : Il a récemment été montré dans des lignées cellulaires de mammifères que le nombre de cassures d'ADN double-brin (DSB) générées par l'activité endonucléase des protéines de l'ORF2 des éléments L1, est beaucoup plus élevé que le nombre associé à l'insertion de L1 (Gasior, Wakeman et al. 2006) (**Figure 9b**). La mesure dans laquelle ces DSBs contribuent à l'instabilité génomique humaine reste inconnue car les niveaux d'expression de L1 dans ces conditions expérimentales étaient beaucoup plus élevés que ceux attendus dans des conditions cellulaires normales.

La créations de nouveaux microsatellites : En raison de leur abondance dans le génome les éléments transposables peuvent générer des microsatellites à de nombreux loci dans le génome (**Figure 9c**). Cela a été étudié, en particulier, pour l'éléments Alu (Arcot, Wang et al. 1995). Il existe au moins deux exemples de maladies génétiques qui sont causées par l'expansion de microsatellites issus de régions riches en éléments Alu (Justice, Den et al. 2001; Kurosaki, Ninokata et al. 2006).

La conversion génique : plusieurs études indiquent que certains éléments font de la conversion génique (**Figure 9d**) (Roy, Carroll et al. 2000), qui est un type de recombinaison qui se définit comme le transfert non réciproque d'informations entre des séquences homologues. L'élément Alu représente plus de 10% des séquences du génome humain, la conversion génique médiée par cet élément pourrait avoir un impact substantiel sur la diversité nucléotidique de l'ensemble de notre génome. Aussi, il pourrait nuire à l'utilisation des SNP situés dans les séquences Alu comme marqueurs génétiques car la

a) Suppression génomiques



b) Recombinaisons ectopiques

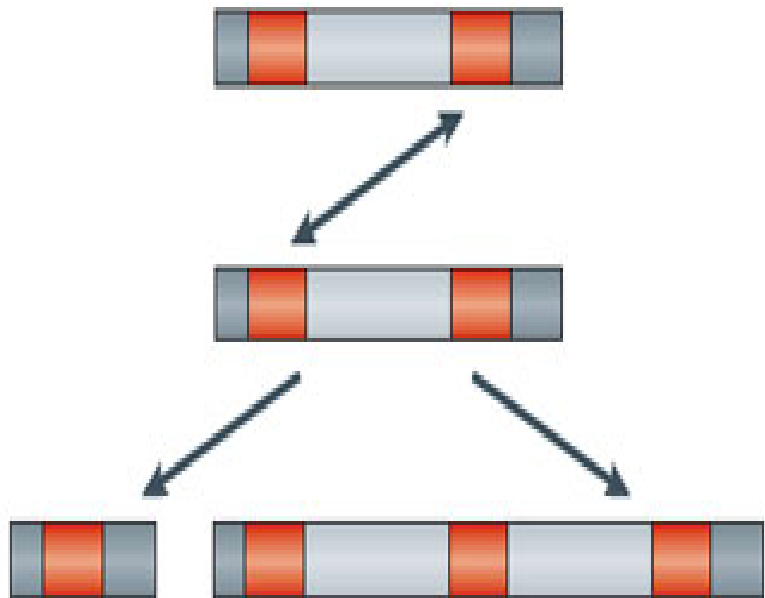


Figure 10 : Réarrangements inter-chromosomiques liés aux éléments transposables

En plus de générer une instabilité génomique locale intra-chromosomique, les rétrotransposons peuvent générer des réarrangements génomiques inter-chromosomiques, tels que des délétions, des duplications et des inversions. **a)** L'insertion des éléments sur de nouveaux sites génomiques entraîne parfois la suppression d'une séquence génomique adjacente et peut donc conduire à de la suppression génique. **b)** En raison de leur nombre de copies extrêmement élevé, les éléments transposables peuvent également créer des variations génomiques structurelles par recombinaison homologues non-allélique entre les éléments (recombinaisons ectopiques). Ce type de réarrangement ne résulte pas en gain ou en perte de séquences génomiques, il contribue à la variation génomique et peut être d'une importance fonctionnelle. [Cordaux and Batzer 2009]

conversion génique rendrait ces SNPs identiques par l'état plutôt que par filiation (Batzer and Deininger 2002) .

La suppression génomique : En plus de générer une instabilité génomique locale, les rétrotransposons peuvent générer des réarrangements génomiques, tels que des délétions, des duplications et des inversions. L'insertion d'éléments sur de nouveaux sites génomiques entraîne parfois la suppression d'une séquence génomique adjacente (**Figure 10a**). Ce phénomène a d'abord été observé à travers l'analyse de l'intégration de l'élément L1 dans des cultures de cellules humaines où environ 20% des insertions trouvées ont été associées à des réarrangements structuraux, y compris des suppressions qui variaient en taille de 1 pb à plus de 130 kb au niveau du site d'insertion (Gilbert, Lutz-Prigge et al. 2002). Au total, il a été estimé qu'au cours de l'évolution des primates, 45 000 suppressions par insertion pourraient avoir retiré plus de 30 Mb de séquence génomique (Cordaux 2008) .

Les recombinaisons ectopiques : En raison de leur nombre de copies extrêmement élevé, les éléments transposables peuvent également créer des variations génomiques structurelles par recombinaison homologue non-allélique entre les éléments (**Figure 10b**) . La contribution des éléments sur les inversions chromosomiques a été étudiée récemment par génomique comparative. Près de la moitié des inversions qui ont eu lieu dans le génome humain et celui du chimpanzé depuis leur divergence ont impliqué les éléments L1 et Alu et 20% de toutes ces inversions peuvent être clairement identifiées comme produits de recombinaison L1-L1 ou Alu-Alu (Lee, Han et al. 2008). Bien que ce type de réarrangement ne résulte pas en gain ou en perte de séquences génomiques, il contribue à la variation génomique et peut avoir une importance fonctionnelle (Lee, Han et al. 2008). Chez l'homme, deux maladies, celle de Charcot-Marie Tooth de type 1A et une neuropathie héréditaire (HNPP ou hereditary neuropathy with liability to pressure palsies),

sont dues à des crossing-over inégaux sur le chromosome 17 et il semble qu'un élément de la famille mariner soit à l'origine de cette recombinaison aberrante (Reiter, Murakami et al. 1996).

Ces séquences répétées, *via* divers mécanismes dont certains cités ci-dessus, peuvent provoquer de multiples maladies. En effet, un nombre sans cesse croissant de cancers a été associé à l'insertion *de novo* ou à des recombinaisons impliquant des séquences transposables (Belancio, Roy-Engel et al. 2010). Les données répertoriées dans la littérature indiquent que de nombreux éléments transposables présentent une forte activité dans les lignées tumorales telles que les myélomes et les neuroblastomes (Kuff and Fewell 1985) ou dans les lignées cellulaires traitées par des agents carcinogènes (Hsiao, Gattoni-Celli et al. 1986). Des particules rétrovirales de rétrovirus endogènes HTDV/HERV-K ont été détectées dans plus de 20 lignées dérivées de tumeurs des testicules, (Lower, Boller et al. 1993) indiquant qu'une activation de ces rétroéléments s'est mise en place dans ces lignées. On peut citer également les séquences IAP (Intracisternal A-particles) qui, chez la souris, voient leur expression et la production de particules accrues en cellules tumorales (Dupressoir and Heidmann 1997).

c).2 Bénéfices

Si le rôle mutagène des éléments transposables et leur implication dans la variabilité génétique sont très connus, leur influence bénéfique sur les génomes au cours de l'évolution a mis plus de temps à être acceptée. Or, on sait maintenant que ces éléments ont joué une part considérable dans la mise en place de certains systèmes de régulation des gènes. Leurs rôles, bénéfiques, comme protecteurs des télomères de la drosophile par les retrotransposons sans LTR Het-A et TART (Biessmann and Mason 1992; Biessmann, Valgeirsdottir et al. 1992), dans la formation des introns et des exons ainsi que dans la

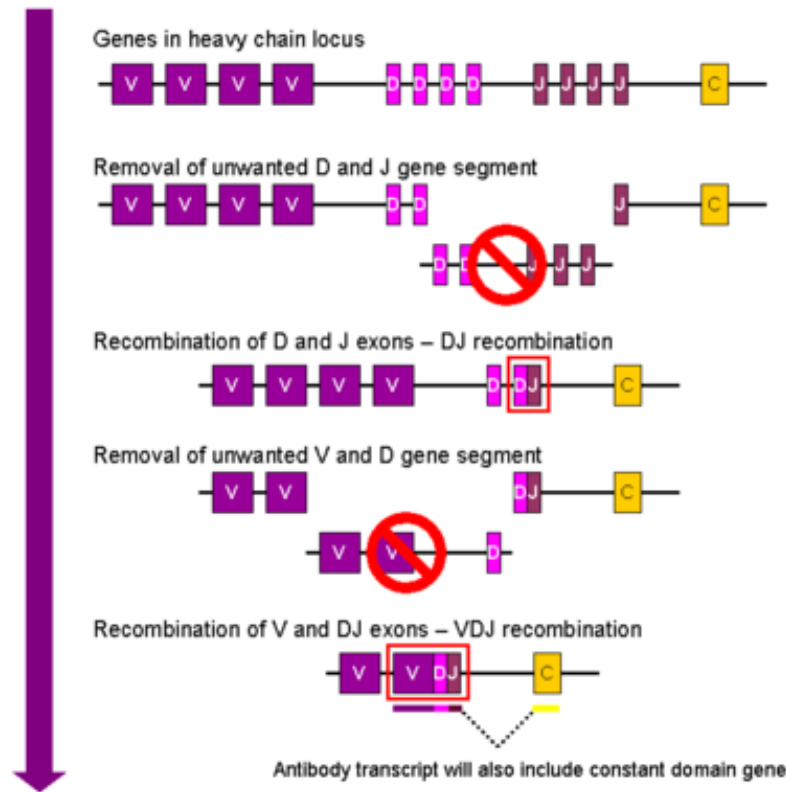


Figure 11 : Exemple d'apport bénéfique des éléments transposables: le système immunitaire acquis

L'apparition du système immunitaire acquis est un évènement fondamental qui permet aux organismes plus évolués de lutter contre les infections. Les immunoglobulines (Ig) sont formées de deux chaînes variables (la chaîne lourde et la chaîne légère) qui confèrent une variabilité importante au système immunitaire. Dans les cellules B en développement, il y a divers processus de recombinaison qui permettent la variabilité des anticorps. Premièrement, il y a recombinaison entre les régions variable D (Diversity) et J (Joining). Tout l'ADN situé entre les deux régions qui ont recombiné est supprimé. Cette recombinaison DJ est suivie par l'assemblage d'un des gènes de la région V (Variable) avec la région DJ nouvellement formée en amont, formant un gène VDJ réarrangé. Tous les autres gènes situés entre les segments V et D du gène nouvellement formé VDJ sont supprimés du génome de la cellule. Le transcrit primaire qui est généré contient la région VDJ et une région constante C (jaune). Cet ARN primaire est polyadénylé et épissé afin de supprimer la séquence entre le segment VDJ et le segment C. La traduction de cet ARNm conduit à la production de la chaîne lourde des Ig. [[http://en.wikipedia.org/wiki/V\(D\)J_recombination](http://en.wikipedia.org/wiki/V(D)J_recombination)]

mise en place de la variabilité du système immunitaire des vertébrés a clairement été établi. En effet le système immunitaire acquis doit être capable de reconnaître des millions d'antigènes différents. La diversité des anticorps produits est générée au niveau des lymphocytes au cours de leur développement, suite à des remaniements de l'ADN au niveau des gènes codant les immunoglobulines et les récepteurs des cellules T. Le processus conduisant à ces remaniements a été appelé Recombinaison V(D)J (**Figure 11**). Dans ce système, deux protéines, RAG1 et RAG2, sont essentielles pour la recombinaison. Elles possèdent, en effet, à la fois la propriété de reconnaître des séquences nucléotidiques spécifiques correspondant à des signaux de recombinaison placés au voisinage des segments V, D et J, et celle de couper l'ADN immédiatement à proximité de ces signaux. Ces signaux, composés d'héptamères et de nonamères très conservés et séparés les uns des autres par des séquences relativement homogènes de 12 ou 23 nucléotides, évoquent les répétitions terminales inversées de nombreux éléments transposables. L'élément à l'origine de *rag1* et *rag2* serait devenu immobile par la perte de ses ITR.

Les éléments transposables ont aussi été impliqués dans des phénomènes de spéciation. Trois mécanismes liés à la présence d'éléments transposables dans les génomes, et susceptibles de créer un isolement reproductif, ont été proposés: le réarrangement du programme génétique, l'incompatibilité mécanique des génomes et la dysgénésie des hybrides (Rose and Doolittle 1983).

On peut effectivement imaginer qu'une divergence de programmes de développement se produit entre individus ne possédant pas les mêmes types d'ADN mobiles. La formation de réarrangements chromosomiques pourrait entraîner un mauvais appariement des chromosomes homologues chez les hybrides et produire un début d'isolement post-reproductif. Dans ce cas les éléments transposables agiraient par leur rôle mutateur, sans être impliqués directement dans le mécanisme de spéciation.

F0



F1

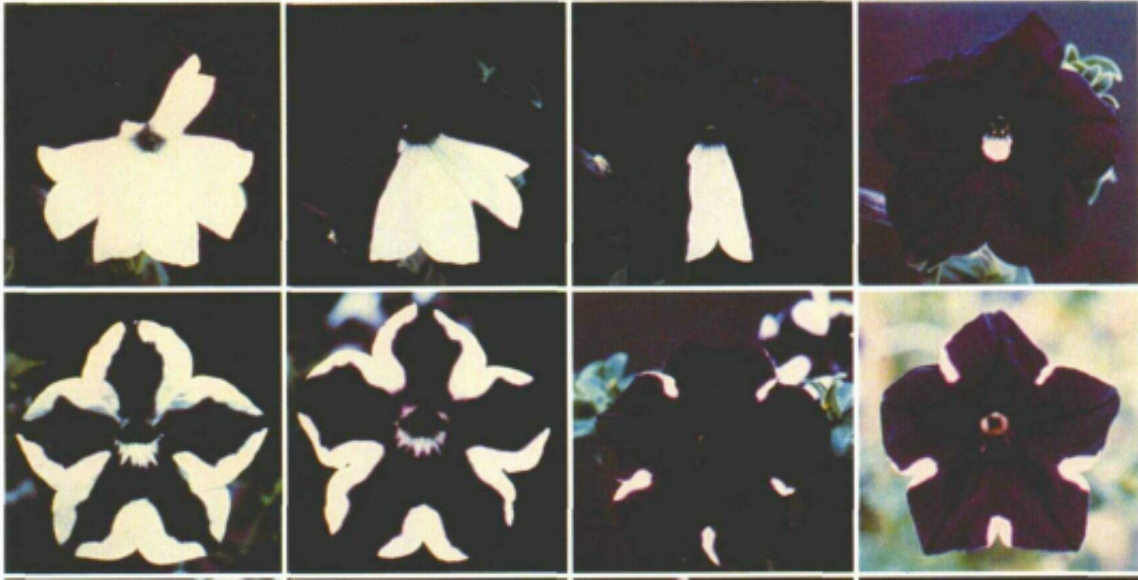


Figure 12 : Co-suppression chez le pétunia

En haut (F0) le pétunia parental à partir duquel des botanistes ont voulu surexprimer le gène codant pour la chalcone synthetase, qui est responsable de la couleur pourpre de la fleur. Pour cela ils ont introduit plusieurs copies du gène par transgènes. En dessous les différentes lignées issus de ce pétunia (F1) et qui surexprime ce gène. Au lieu d'avoir une augmentation de la coloration pourpre il apparaît des clones blancs. On parla alors de co-suppression: la présence de plusieurs copies de transgènes identiques entraîne une extinction de ceux ci par le génome de l'hôte. [Napoli *et al.*, 1990]

En revanche, dans le phénomène de dysgénésie des hybrides, mis en évidence avec les éléments P et I chez la drosophile, l'influence des éléments est plus directe. L'hypothèse la plus simple suppose que deux populations puissent être indépendamment envahies par des familles d'éléments transposables différents et que les hybrides résultant des croisements interpopulations soient stériles ou létaux. Dans ce cas les éléments seraient directement impliqués mais on ne connaît pour l'instant que des cas de dysgénésie partielle et restreints à des températures d'élevage particulières, ce qui limite la portée de cette hypothèse.

Tous ces mécanismes sont de remarquables illustrations du pouvoir de remodelage de la structure et du fonctionnement du génome par des éléments transposables et, donc, de leur impact sur l'évolution des organismes. Bien que la fréquence de transposition très faible de ces éléments (environ 10^{-6} par génération /élément) soit due au fait que beaucoup de copies sont dégénérées, le génome de l'hôte doit pouvoir finement contrôler la transposition des copies complètes afin de limiter leurs effets mutagènes.

2) Régulation des éléments transposables

a) Les voies d'ARN interférence

Voulant augmenter l'expression d'un gène de pétunia, la *chalcone synthétase*, en introduisant des copies supplémentaires par transgénèse dans le génome de la plante (Napoli, Lemieux et al. 1990), des botanistes constatèrent qu'au lieu d'obtenir la couleur pourpre attendue, certaines plantes présentaient des fleurs blanches : non seulement le transgène ne s'était pas exprimé, mais il avait de plus inhibé l'expression du gène endogène. On parla alors de co-suppression (**Figure 12**). Il a été démontré que cette co-

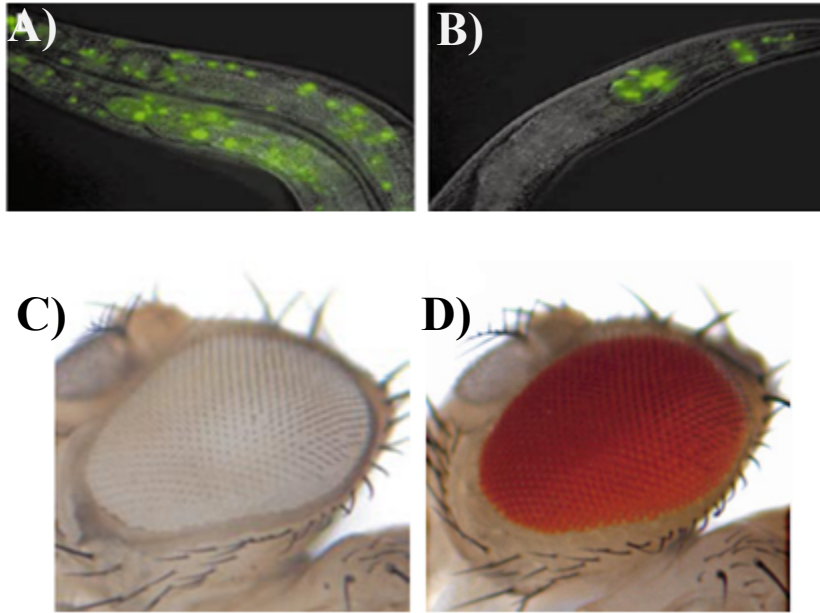


Figure 13: Exemples d'ARN interférence après injection d'ARN double brin dans différents modèles

De l'ARN double brin (ARNdb) peut être introduit expérimentalement dans un organisme pour réprimer des gènes par homologie de séquence. Chez le nématode, la répression peut être déclenchée par injection ou ingestion d'ARNdb. **A)** et **B)** Les nématodes *Caenorhabditis elegans* ont été modifiés pour exprimer la GFP dans les noyaux. A gauche **A)** ils ont été traités avec un ARNdb contrôle, tandis que ceux de droite **B)** ont été traités avec de l'ARNdb de la GFP, ce qui a pour effet de réprimer l'expression de la GFP. **C)** et **D)** Les drosophiles expriment le gène *white* responsable de la couleur rouge brique des yeux de la mouche. **C)** Les drosophiles expriment un ARN double brin homologue au gène *white* ce qui a pour résultat d'empêcher l'expression du gène endogène donnant des mouches aux yeux non pigmentés, comparées au type sauvage **D)**. [Hannon, 2002].

suppression se mettait en place au niveau transcriptionnel (TGS : *Transcriptional Gene Silencing*) par altération ou modification de l'ADN, par exemple par méthylation (Vaucheret, Beclin et al. 1998) ou hétérochromatinisation. Cette co-suppression pouvait également se mettre en place au niveau post-transcriptionnel (PTGS : *Post-Transcriptional Gene Silencing*) par dégradation rapide des transcrits homologues dans le cytoplasme. Ce dernier phénomène fut élucidé par Fire et Mello (Fire, Xu et al. 1998) (Prix Nobel de médecine en 2006) chez *Caenorhabditis elegans* où ils démontrèrent que la formation d'ARN double-brin (ARNdb) est responsable de l'induction du processus post-transcriptionnel, appelé ARN interférence (RNAi) (**Figure 13**).

Ce mécanisme a été retrouvé chez tous les eucaryotes étudiés. Il intervient dans de nombreux processus cellulaires tels que la régulation du développement, la croissance, la différenciation et l'apoptose, mais il joue aussi un rôle important dans la protection des organismes contre les virus et les éléments répétés (Plasterk 2002). Trois voies de régulation majeure par les petits ARNs sont maintenant connues qui sont respectivement par ordre de découverte : la voie des siRNAs, la voie des micros-RNAs, et celle des piRNAs.

a).1 La voie des petit ARNs interférents (siRNAs)

Suite à la découverte du mécanisme de RNAi, il a été mis en évidence une première voie de régulation : la voie des siRNAs (*small interfering RNA* / petit ARNs interférents).

Dans un premier temps, ce mécanisme a été mis en évidence chez les plantes pour son rôle constitutif de lutte contre les infections virales et fut appelé VIGS (*Virus Induced Gene Silencing*/ Mise sous silence de gènes induit par les virus). Conservé chez presque tous les eucaryotes étudiés, excepté la levure *Saccharomyces cerevisiae*, cette voie consiste en deux étapes distinctes : la phase d'initiation et la phase effectrice.

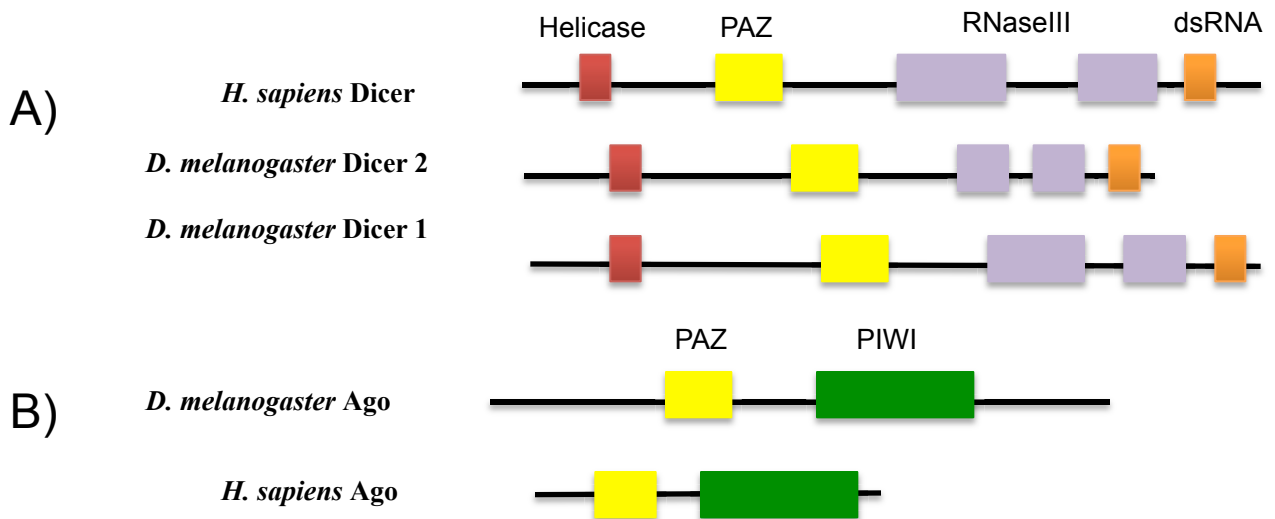


Figure 14: Domaines protéiques conservés des protéines Dicer et Argonaute

A) Les protéines Dicer (Dcr) de drosophile et humaine contiennent divers domaines protéiques importants et conservés: deux domaines Rnases III (violet), un domaine Helicase (rouge), un domaine PAZ (jaune) et un domaine de liaison aux ARN double brin (orange). La drosophile contient deux protéines Dicer, Dcr1 impliquée dans la voie des miRNAs et Dcr2 impliquée dans celle des siRNAs.

B) les protéines Argonaute contiennent deux domaines important: un domaine PAZ (jaune) et un domaine PIWI (vert). [Adapté de Meister and Tuschl 2004; Tsutsumi *et al.*, 2011]

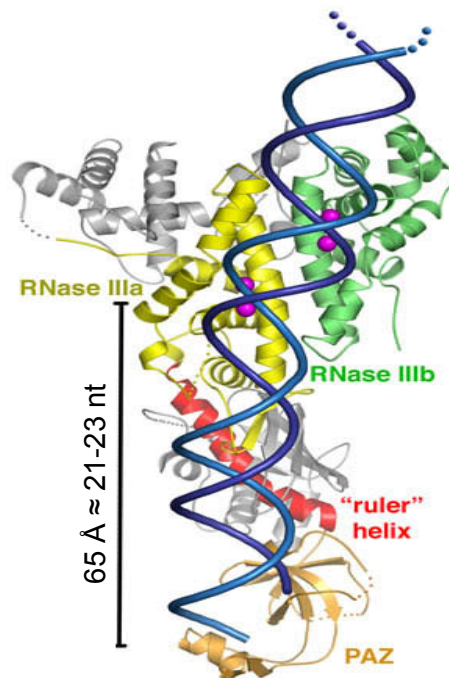


Figure 15: Structure du complexe Dicer/ARNdb

Sur cette représentation de la structure cristallographique de l'enzyme Dicer on peut remarquer qu'elle ressemble à une pince coupante avec comme poignée le domaine PAZ (orange) et le couperet de la lame les domaines RNase IIIa et IIIb (jaune et vert). Un connecteur plat mesurant 65 angströms sert de règle qui est utilisée pour mesurer des segments de 21 à 23 nucléotides de long. Un fragment d'ARN double brin (bleu foncé) est montré passant par l'enzyme Dicer. [Adapté de MacRae *et al.*, 2006]

La phase d'initiation est majoritairement effectuée par Dicer, une enzyme qui appartient à la classe des RNases III et qui a été identifiée comme étant responsable de la production de courts fragments d'ARNdb. L'activité nucléasique de Dicer ne s'exerce que sur des ARNdb. Cette protéine est localisée dans le cytoplasme des cellules et possède quatre domaines importants : un domaine ARN hélicase en N-terminal, un domaine PAZ (PIWI/Argonaute/Zwille), deux domaines RNase III et un motif de liaison aux ARNdb en C-terminal (**Figure 14**). Il n'a été caractérisé qu'une seule protéine Dicer chez les mammifères et chez *C. elegans* alors que chez *Drosophila melanogaster* on compte deux protéines de type Dicer : Dicer 1 (Dcr1) et Dicer 2 (Dcr2), cette dernière étant seulement impliquée dans la production des siRNAs.

La protéine Dicer repère les longs ARNdb (souvent issus d'une infection par un virus exogène) (**Figure 15**) puis, grâce à son activité nucléasique, les coupe en fragments d'une taille de 21 à 23 nucléotides. Appelés siRNAs, ces fragments ont la particularité de présenter à leur extrémité 3' deux nucléotides libres (non appariés) ainsi qu'un groupement phosphate à leur extrémité 5' (Tomari and Zamore 2005; Macrae, Zhou et al. 2006). Dicer n'agit pas sur une séquence spécifique, mais sur tout ARNdb et génère des fragments de taille constante.

Les petits ARNs doubles brins produits par Dicer sont ensuite reconnus par un complexe enzymatique dénommé RISC (RNA-Induced Silencing Complex) qui est composé de protéines Argonaute dotées d'activités hélicase ATP-dépendante et RNase comme par exemple la protéine Argonaute 2 (Ago2) chez la drosophile. Ces activités enzymatiques permettent le positionnement et l'alignement du siRNA sur l'ARNm cible qui lui est complémentaire, puis la coupure de cet ARNm. L'ARNm, ainsi fragmenté en ARNdb d'une vingtaine de nucléotides, ne peut plus donner lieu à un processus de traduction. Les fragments générés sont ultérieurement clivés par des RNases cellulaires.

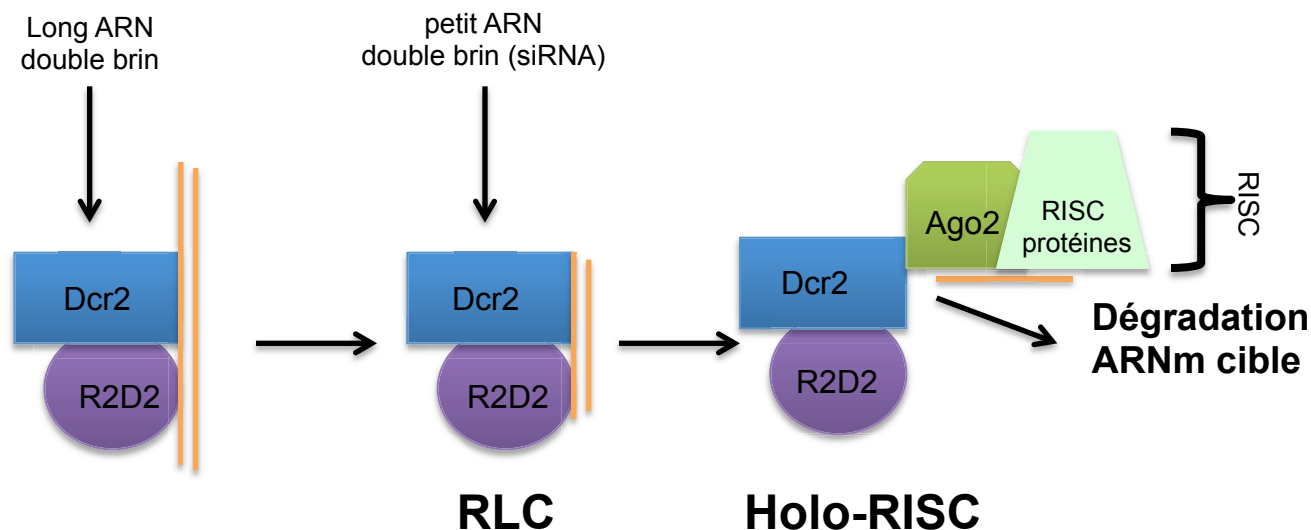


Figure 16: Modèle pour le mécanisme de la voie des siRNAs

De longs ARNdb, provenant soit d'ARNs synthétiques, soit issus de virus en répllication ou encore transcrits à partir de gènes nucléaires, sont reconnus et clivés par l'enzyme Dicer 2 (rectangle bleu) en présence de R2D2 (rond violet). Il en résulte la formation d'un complexe, contenant un petit ARN double brin (siRNA), appelé Risc Loading Complex (RLC) qui va charger le siRNA dans le complexe effecteur RISC (RNA-Induced Silencing Complex) et l'activer, ce qui va donner l'Holo-RISC (vert). Le complexe RISC qui contient entre autre la protéine Argonaute 2 (Ago2, vert foncé) va cibler un ARN endogène qui lui est homologue et ainsi réguler l'expression génique à de nombreux niveaux. De tels complexes agissent en favorisant soit la dégradation soit la répression traductionnelle de l'ARNm cible. Des complexes similaires ciblent également le remodelage de la chromatine. [Adapté de Pham *et al.*, 2004].

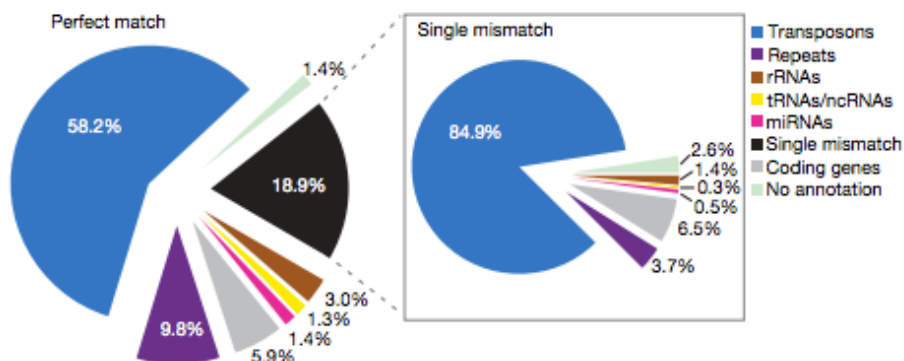


Figure 17: Identification des petits ARNs endogènes liés à Ago2

Sur le diagramme de gauche on peut observer la distribution des petits ARNs qui sont liés à la protéine Argonaute 2 (Ago2) en culture de cellules de drosophile et qui sont 100% identiques au génome. L'agrandissement à droite de la partie noire montre la distribution des petits ARNs ayant un nucléotide non identique au génome (single mismatch). On peut facilement observer que ces petits ARNs sont majoritairement identiques à des séquences d'éléments transposables (bleu) et donc potentiellement des régulateurs de ceux ci. Seul 8% correspondent à des gènes non répétés (gris et rose). [Kawamura *et al.*, 2008]

Les protéines de la famille argonaute (**Figure 14**) possèdent deux domaines importants et conservés : le domaine PAZ qui va fixer l'extrémité 3' du siRNA et PIWI qui va fixer l'extrémité 5' phosphate (Ma, Yuan et al. 2005; Parker, Roe et al. 2005). La torsion induite pas la fixation du siRNA dans le RISC va fragiliser l'appariement entre les deux brins du petit ARN et une hélicases ARN du complexe va séparer les deux brins et incorporer un seul brin. Ce complexe chargé et activé est appelé holo-RISC. Il va hybrider l'ARNm cible avec pour guide le petit ARN engagé dans le RISC (Kim, Lee et al. 2007). Chez la drosophile une protéine nommée R2D2, ayant deux domaines dsRBP, (Liu, Rand et al. 2003) agit comme partenaire de Dcr2 pour former un complexe RLC (RISC Loading Complex) afin d'effectuer le chargement du siRNA dans le RISC (**Figure 16**). R2D2 trouve les inégalités thermodynamiques, se lie à l'extrémité la plus stable de l'ARN duplex et oriente Ago2 sur le petit ARN double brin. Il a été montré que l'activité enzymatique endonucléolytique de la protéine Ago2 est responsable du retrait du brin passager du duplex siRNA (Miyoshi, Tsukumo et al. 2005; Diederichs and Haber 2007).

Les siRNAs ne sont pas uniquement un système de régulation antiviral exogène mais ils sont aussi codés dans le génome de l'hôte. Ces siRNAs endogènes ont été mis en évidence dans un premier temps chez les plantes comme les casiRNAs (Cis Acting RNA).

Récemment, des petits ARNs endogènes ayant de nombreuses similarités avec les siRNAs ont été mis en évidence chez la souris (Tam, Aravin et al. 2008; Watanabe, Totoki et al. 2008) et chez la drosophile (Chung, Okamura et al. 2008; Czech, Malone et al. 2008; Ghildiyal, Seitz et al. 2008; Kawamura, Saito et al. 2008; Okamura, Chung et al. 2008) et ont été appelés endo-siRNA ou esi-RNA (endogenous siRNAs ou petit ARNs endogènes) (**Figure 17**). Ces ARNs sont produits par des clusters génomiques (**Figure 18**) aussi bien dans les tissus reproducteurs que dans les tissus somatiques de drosophile (Czech, Malone

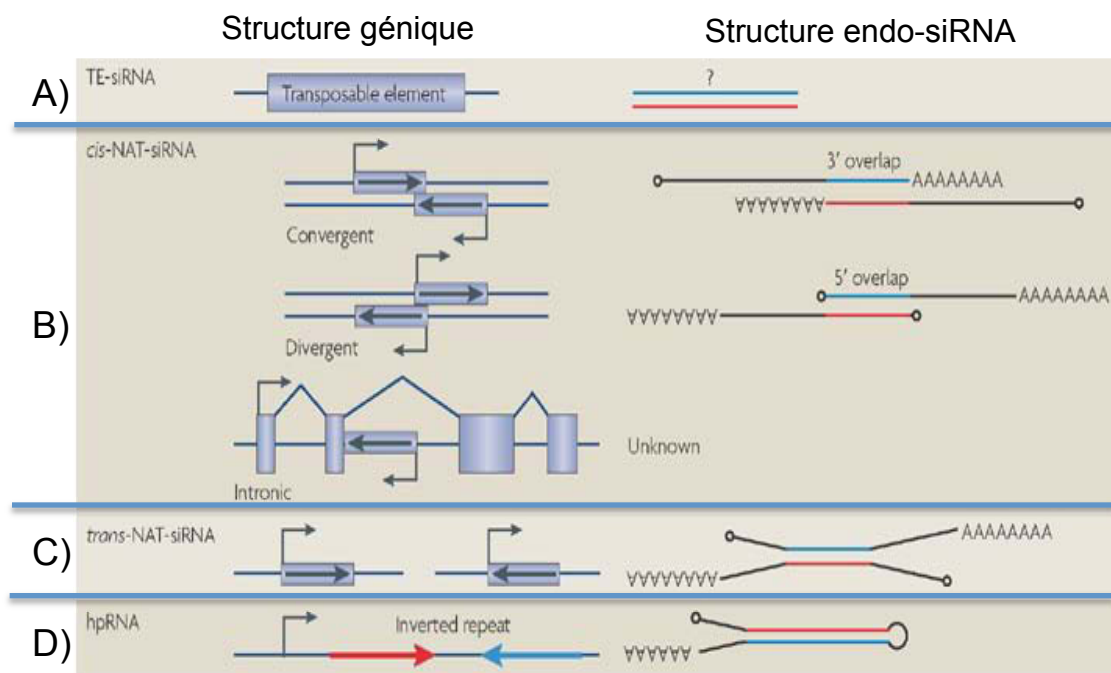


Figure 18: Divers type de génération de siRNAs endogènes (endo-siRNAs)

Sur ce schéma sont représenté différentes configuration génomiques qui peuvent entrainer la formation de petit ARNs double brin endogènes (endo-siRNAs). Sur la gauche on peut observer la configuration génique et sur la droite celle de l'ARN double brin pouvant être formé (en rouge et bleu). **A)** La transcription dans les deux sens d'un élément transposable, **B)** La transcription convergente ou divergente de deux gènes se chevauchant ou d'une transcription en sens inverse à partir de l'intron d'un gène, **C)** Deux gènes situés sur des régions génomiques éloignées mais ayant une partie commune en sens dans l'un et en antisens dans l'autre peuvent générer un appariement des ARNm. **D)** Un gène ou une région génique contenant deux fragments identiques en sens inverse dans la même unité de transcription peut entrainer un repliement de l'ARNm. [Katsutomo Okamura & Eric C. Lai 2008].

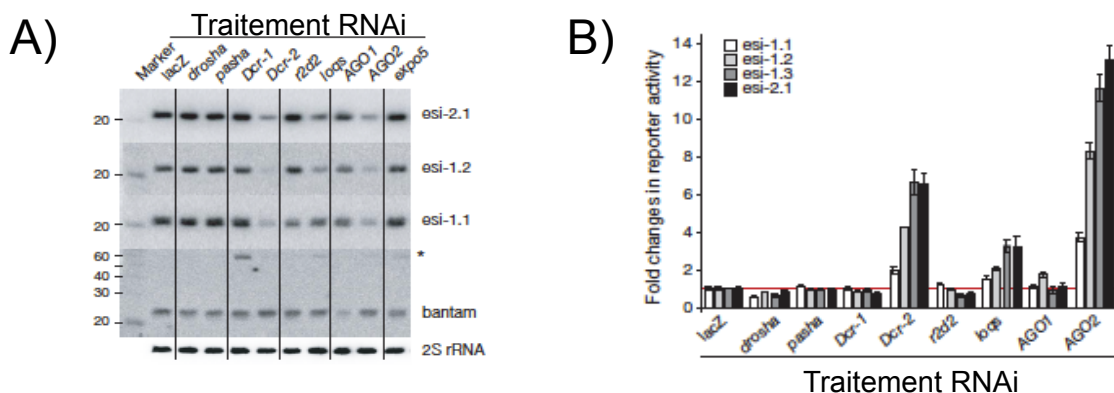


Figure 19: Protéines impliquées dans la voie des endo-siRNAs

A) Northern blots indiquant la quantité de siRNAs endogènes esi-1 and esi-2 (esi-2.1, esi-1.1 et esi-1.2) en cellules S2 transfectées avec des ARNs double brins contre divers gènes impliqués dans les voies d'ARN interférence. Les contrôles sont le pre-miRNA *bantam* et l'ARN ribosomique 2S. On peut observer que les gènes *dcr2*, *ago2* et *loquacious* (*loqs*) semblent impliqués dans la fabrication des endo-siRNAs. **B)** Essai luciférase avec 4 endo-siRNAs en cellules S2 traitées avec des ARNs doubles brin contre les différent gènes indiqués. Les gènes *dcr2*, *ago2* et *loquacious* sont impliqués dans la répression *via* les endo-siRNAs. [Czech *et al.*, 2008]

et al. 2008). Bien qu'ayant de fortes similarités structurales avec les siRNAs exogènes et que leur fabrication fasse intervenir les protéines Ago2 et Dcr2 chez la drosophile, il semble surprenant que R2D2 ne soit pas impliquée dans cette voie de maturation. En revanche la participation de la protéine Loquacious (Loqs), qui est un partenaire de Dcr1 dans la voie des miRNAs, est nécessaire pour la régulation *via* les endo-siRNAs (**Figure 19**). En effet des précurseurs ARNdb endogènes sont coupés par la protéine Dcr2 complexée avec Loqs et sont chargés avec la protéine Ago2 dans le RISC. Ce chargement ne semble pas faire intervenir la protéine R2D2 (**Figure 19**) (Chung, Okamura et al. 2008; Czech, Malone et al. 2008; Ghildiyal, Seitz et al. 2008; Kawamura, Saito et al. 2008; Okamura, Chung et al. 2008). Cependant une isoforme de Loqs (Loqs-PD) a été identifiée ce qui a permis de séparer la fonction des différentes isoformes de Loqs au cours de la biogenèse des miRNAs et des endo-siRNA (Hartig, Esslinger et al. 2009; Zhou, Czech et al. 2009). Le laboratoire de Siomi H&M a proposé qu'un complexe trimérique composé de Dcr-2, Loqs et R2D2 peut se former (Miyoshi, Miyoshi et al. 2010). La preuve expérimentale de ce complexe trimérique a été obtenue en surexprimant ces protéines, mais peut cependant ne pas refléter une situation physiologique. Très récemment, il a été montré que Loqs-PD et R2D2 sont en concurrence pour interagir avec Dcr2. Ces deux protéines ont une affinité pour le domaine hélicase de Dcr2. Bien que cette région soit assez grande pour accueillir plus d'une protéine, il a été montré que la diminution du niveau de R2D2 ou de Loqs-PD entraîne une activité accrue du silencing médié par l'autre voie. Dcr2 est donc le facteur limitant dans la régulation. En effet, la surexpression de Dcr-2 mais pas celle de R2D2 peut améliorer l'efficacité du RNAi (Dietzl, Chen et al. 2007). Un phénomène similaire a été observé concernant la concurrence entre les protéines Loqs-PA et Loqs-PB dans la voie des miRNAs.

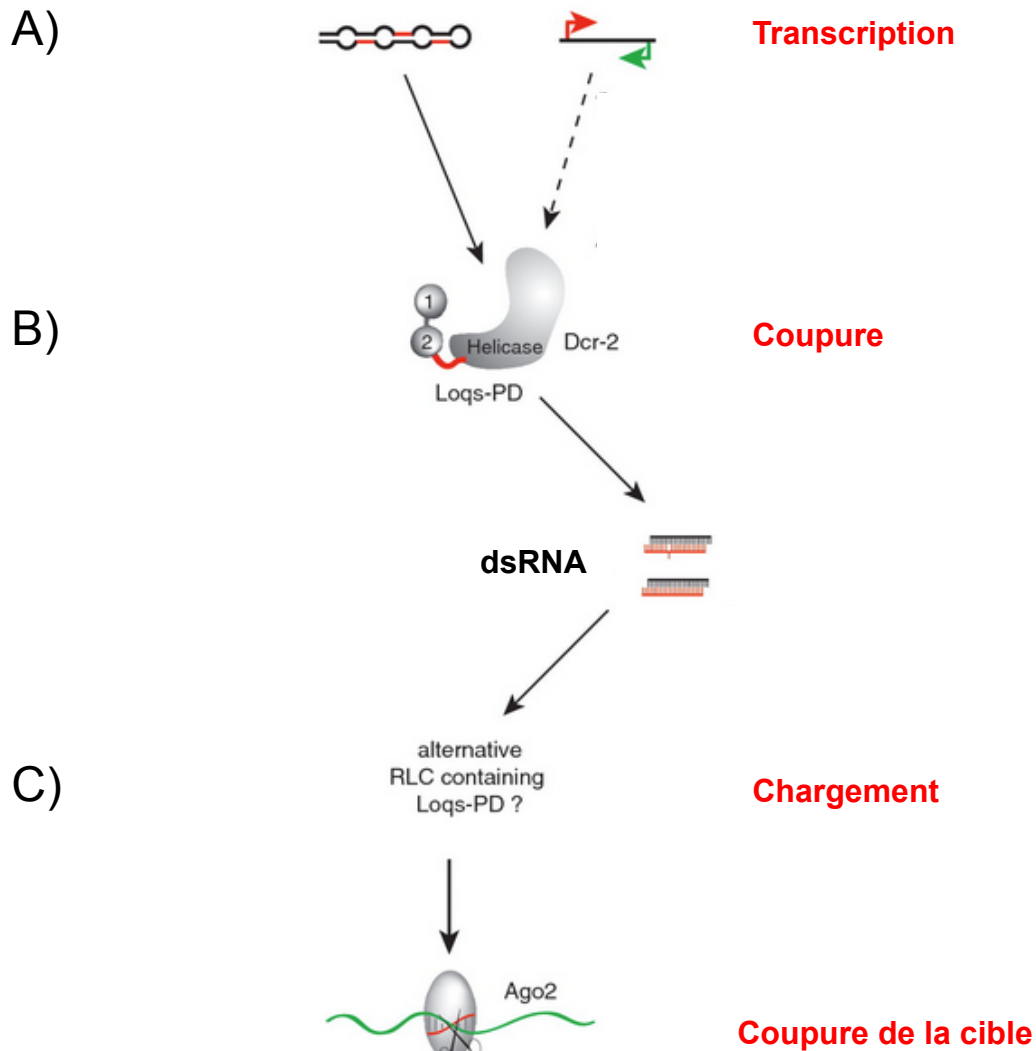


Figure 20: Voie canonique des endo-siRNAs

A) La transcription de loci génomiques structurés (cf Figure 18) va produire des transcrits double brins. **B)** Ces transcrits vont être pris en charge par la protéine Loquacious (Loqs-PD) qui contient deux domaines de liaison à l'ARN double-brin en N-terminal (numérotés 1 et 2) qui va s'associer via son domaine C-terminal avec le domaine hélicase N-terminal de Dcr-2. Ce complexe de type RLC va couper les grands ARNs double brin en petits ARNdb qui vont ensuite être chargés dans un complexe RISC (**C**) qui contient la protéine Ago 2. Le complexe RISC chargé va cibler et détruire sa cible. [Adapté de Hartig et Forstemann 2010]

La voie des endo-siRNAs a été impliquée dans la régulation des éléments transposables dans les tissus somatiques de drosophile (**Figure 20**) (Czech, Malone et al. 2008; Ghildiyal, Seitz et al. 2008; Kawamura, Saito et al. 2008; Okamura, Chung et al. 2008) ainsi que plus récemment dans la longévité et la résistance au stress chez *D. melanogaster* grâce à leur rôle dans la régulation de l'homéostasie métabolique (Lim, Oh et al. 2011).

a).2 La voie des micros ARNs (miRNAs)

Les micros-RNAs (miRNA) ont été découverts en 1993 par Victor Ambros, Rosalind Lee et Rhonda Feinbaum lors d'une étude du gène *lin-14* chez *C. elegans* (Lee, Feinbaum et al. 1993). Ils ont constaté que l'abondance des protéines LIN-14 est régulée par un ARN codé par le gène *lin-4*. Rétrospectivement, *lin-4* était le premier miRNA identifié. Ce n'est qu'en 2000 qu'un deuxième miRNA a été caractérisé: *let-7* (Reinhart, Slack et al. 2000), qui s'est révélé être conservé dans de nombreuses espèces, (Pasquinelli, Reinhart et al. 2000) indiquant l'existence d'un phénomène plus large.

Les miRNAs sont des petits ARNs d'environ 22 nucléotides qui sont temporellement et spatialement exprimés de manière spécifique pour réguler le développement et divers processus physiologiques chez les eucaryotes supérieurs (Flynt and Lai 2008; Herranz and Cohen 2010). Les miRNAs sont générés à partir d'ARN endogènes en forme de "tige boucle" (Kim, Behlke et al. 2005). Leur fonction est de servir de guide dans la régulation traductionnelle des gènes par appariement généralement dans la région 3' non traduite (UTR) des ARNm cibles. La liaison d'un miRNA à l'ARNm cible conduit généralement à une modulation de la traduction bien que d'autres fonctions puissent leur être attribuées comme la fermeture de la chromatine (Kim, Saetrom et al. 2008).

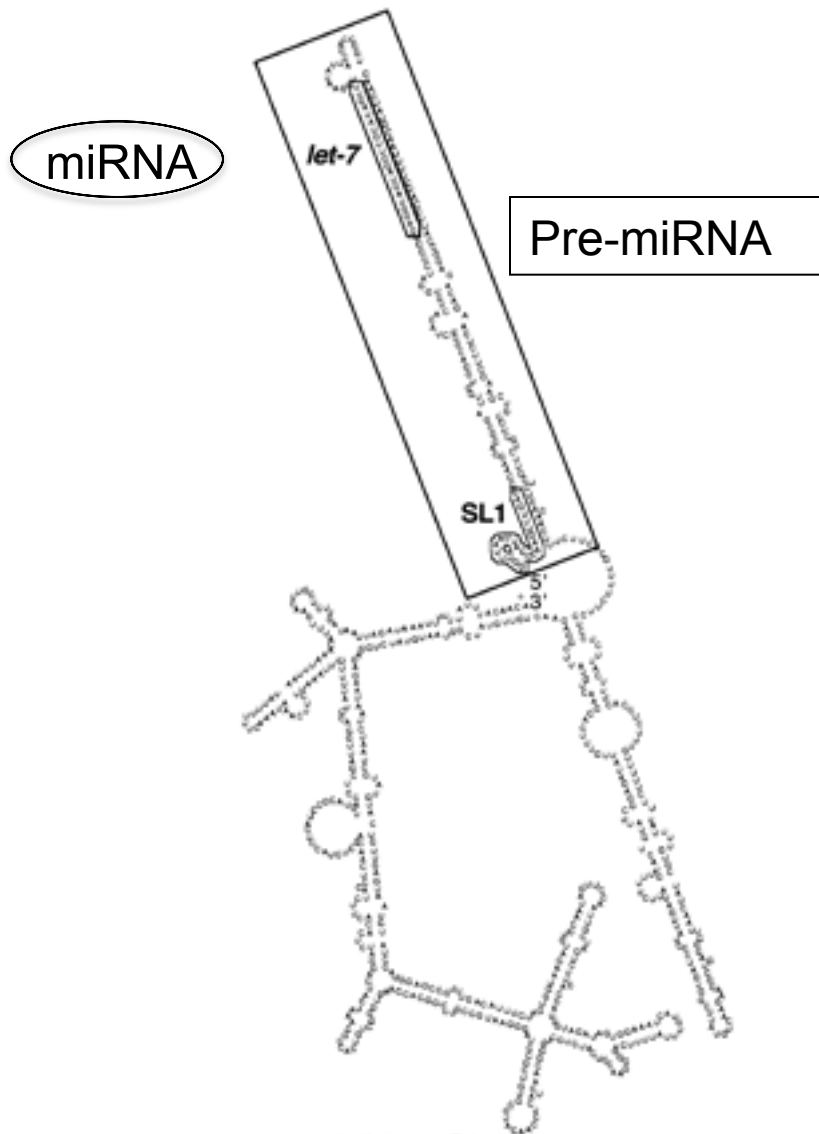


Figure 21: Exemple de structure de pri-miRNA

Structure secondaire du pri-miRNA Let7 de *C-elegans* ayant une taille de 728 nucléotides. En encadré le *pre-mir Let7* et entouré en noir le miRNA *Let7*. On peut remarquer que la structure secondaire est en forme de tige boucle avec des mésappariements à plusieurs endroits.[Adapté de Lee *et al.*, 2003]

Les miRNAs sont phylogénétiquement conservés et à l'heure actuelle on compte par exemple 154 miRNAs chez *C. elegans*, 152 miRNAs chez *D. melanogaster*, 337 miRNAs chez *D. rerio*, et 695 miRNAs chez l'homme. 55% des miRNAs de *C. elegans* ont des homologues chez l'homme, ce qui leur confère un rôle important à travers l'évolution (Ibanez-Ventoso, Vora et al. 2008; Friedman, Farh et al. 2009). Les miRNAs sont le plus souvent codés dans les introns des gènes même si on peut en trouver dans les exons ou dans des régions intergéniques. La transcription des miRNAs est médiée par l'ARN polymérase II (Pol II) (Lee, Kim et al. 2004), bien qu'un groupe mineur de miRNAs qui sont associés aux éléments transposables de type SINE Alu puisse être transcrit par l'ARN polymérase III (Pol III) (Borchert, Lanier et al. 2006).

Classiquement la maturation des miRNAs commence dans le noyau. Les transcrits primaires (Pri-miRNA) (**Figure 21**) qui sont générés par la Pol II peuvent faire plusieurs kilobases (kb) et contiennent une ou des régions en forme de tige boucle. La première étape de la maturation des miRNAs est un clivage au niveau de la tige boucle, ce qui va libérer une petite tige boucle que l'on appelle un pré-miRNA (Lee, Jeon et al. 2002). Cette réaction, qui a lieu dans le noyau, est réalisée par une enzyme de la classe des RNase III : la protéine Drosha (Lee, Ahn et al. 2003) et son cofacteur DGCR8 chez l'homme et Pasha chez *D. melanogaster* et *C. elegans* (une protéine liant les ARN double brin). Ces protéines vont former un complexe connu sous le nom de complexe microprocesseur, d'une taille d'environ 500 kDa chez *D. melanogaster* et 650 kDa chez l'homme (Gregory, Yan et al. 2004).

Un pri-miRNA classique est constitué d'une ou plusieurs tiges boucles d'environ 33 paires de bases, suivies d'un fragment d'ARN simple brin. Pasha interagit avec le pri-miRNA à la jonction entre l'ARN simple brin et la tige boucle, et Drosha coupe environ 11 pb après la jonction entre l'ARN simple brin et l'ARN double brin de la tige boucle,

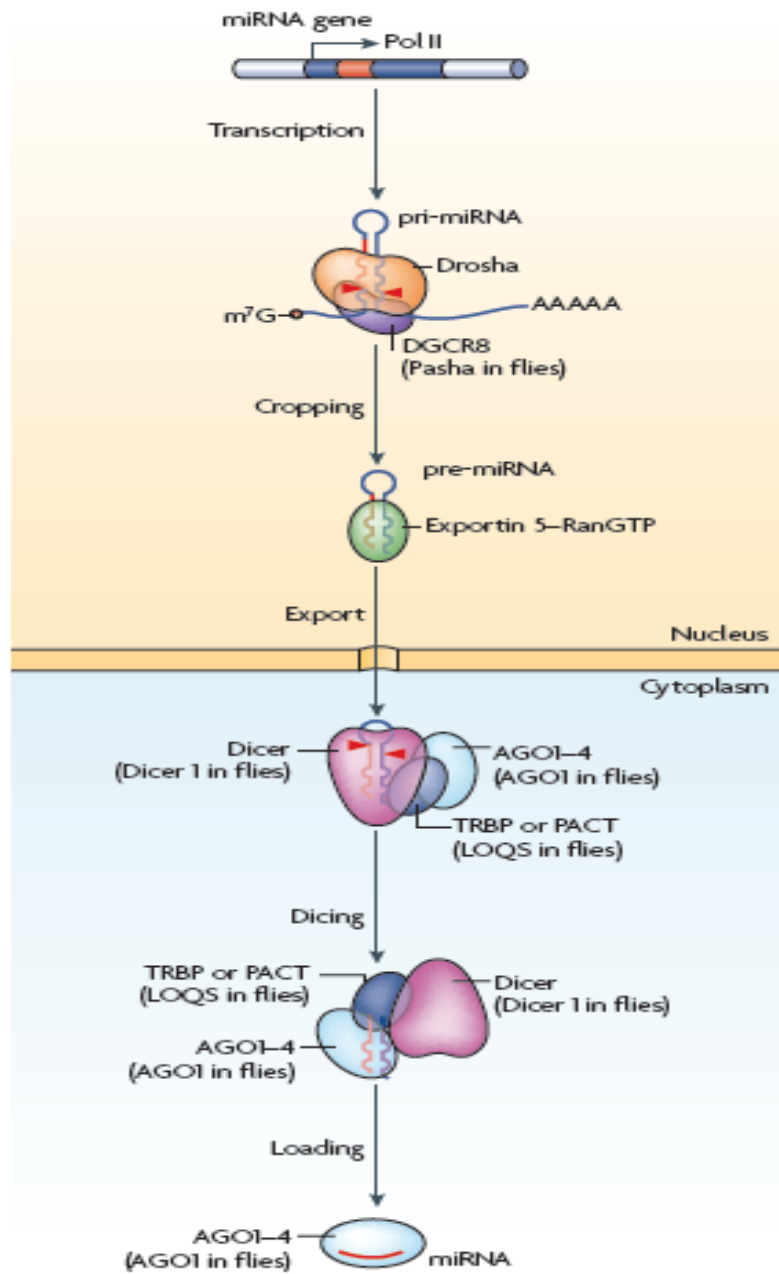


Figure 22: Voie canonique des micro-RNAs

Après transcription par l'ARN polymérase II, le pri-miRNA est transformé en pre-miRNA par un complexe enzymatique qui inclut l'enzyme nucléaire de type RNase III, Drosha et DGRC8 (Pasha). Le pre-miRNA est transporté vers le cytoplasme par l'Exportin-5 et une protéine GTPase Ran. Une fois dans le cytoplasme, le pre-miRNA est transformé en un duplex de ~ 22 nucléotides par une autre enzyme RNase III, Dicer, en association avec TRBP (Loquacious). Le miRNA est ensuite chargé dans le RISC (RNA Induced Silencing Complex) qui contient des protéines Argonautes. Les miRNAs lient majoritairement la région 3' non traduite de leur ARNm cible afin de les déstabiliser ou de bloquer leur traduction par exemple.[Kim *et al.*, 2009]

généralisant ainsi un pre-miRNA (**Figure 22**) (Han, Lee et al. 2006). Des études ont montré que la maturation des pri-miRNAs pourrait être un processus co-transcriptionnel (se déroulant en même temps que la transcription sur le site génomique qui code le miRNA) (Kim and Kim 2007) (Morlando, Ballarino et al. 2008).

Il existe une voie de fabrication de pre-miRNAs indépendante de Drosha qui intervient dans le cas de petites tiges boucles (Mirtrons) (Ruby, Jan et al. 2007), d'ARN de transfert (tRNA) (Babiarz, Ruby et al. 2008) ou des petits ARNs nucléolaires (snoRNA) (Saraiya and Wang 2008).

Les pre-miRNAs ont une taille d'environ 55-70 nt et sont constitués d'une extrémité 3' sortante de deux nucléotides générée par la coupure *via* Drosha. Ils sont exportés vers le cytoplasme *via* une protéine l'Exportine 5 (Exp5) (**Figure 22**) (Kim 2004). L'Exportine 5 est un membre de la famille des transporteurs nucléaires qui, comme tous les autres membres de cette famille, nécessite l'association avec le cofacteur Ran lié au GTP. Après export nucléaire l'Exp5 relâche le pre-miRNA dans le cytoplasme grâce à l'hydrolyse du GTP (Bohnsack, Czaplinski et al. 2004; Lund, Guttinger et al. 2004).

Suite à l'exportation cytoplasmique, les pré-miRNAs sont clivés à proximité de la tige boucle (à environ deux tours d'hélice) par Dicer (**Figure 22**), libérant un miRNA duplexe d'environ 22 nt (Hutvagner, McLachlan et al. 2001; Ketting, Fischer et al. 2001; Tsutsumi, Kawamata et al. 2011). Chez les organismes comme par exemple, *D. melanogaster*, où deux protéines Dicer sont connues, une seule intervient dans la voie des miRNAs: Dcr1. Diverses protéines sont associées avec Dicer chez *D. melanogaster* Dcr1 est associé à Loquacious (Loqs ou R3D1), cette protéine contient trois domaines de liaison à l'ARN double brin (dsRBD/double strand RNA Binding Domain) lui permettant de fixer le pre-miRNA (Forstemann, Tomari et al. 2005; Jiang, Ye et al. 2005). Chez l'homme Dicer interagit avec deux protéines TRBP (TAR RNA-binding protein) (Haase, Jaskiewicz

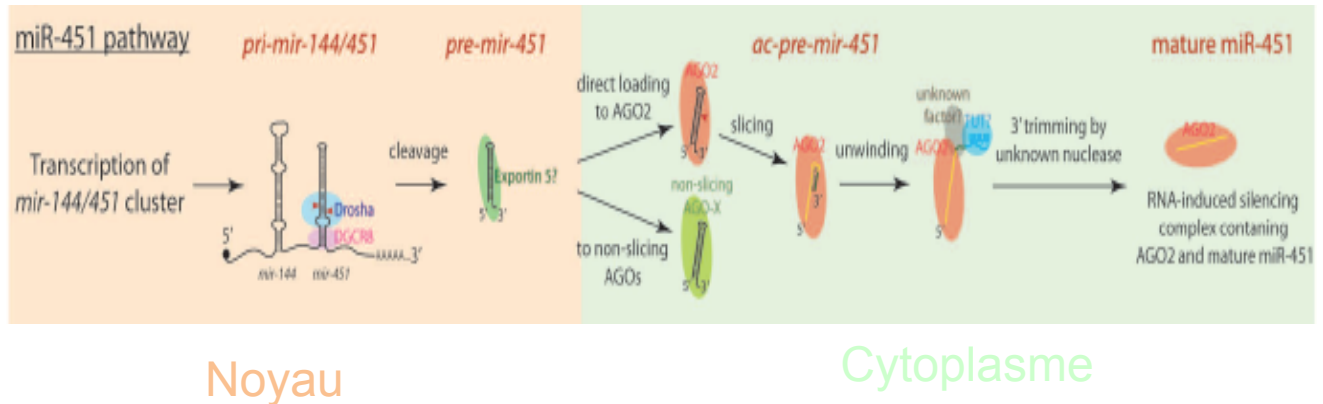


Figure 23: Production de miRNAs indépendante de Dicer

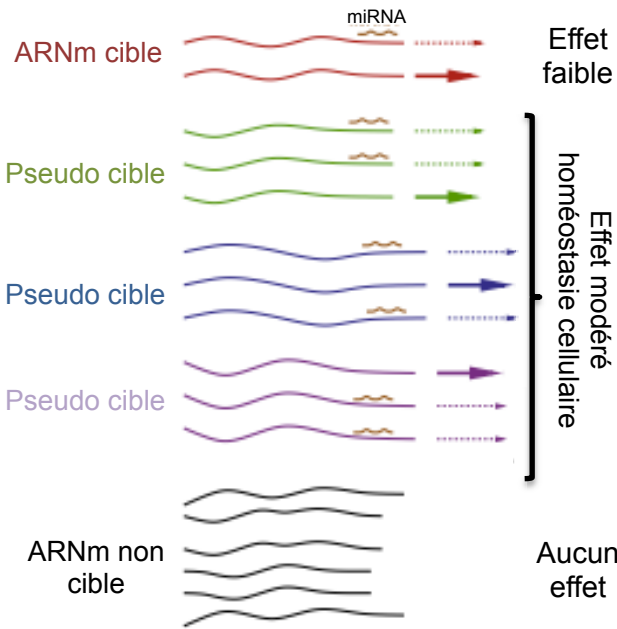
Le miRNA mir-451 se trouve sur un opéron dont le transcrit primaire contient aussi mir-144. Le pri-miRNA 451 est maturé en pré-mir-451 par Drosha/DGCR8 (Pasha) qui va générer une épingle à cheveux de 42nt qui va être exportée du noyau par l'Exportin-5. Le pre-mir-451 a une structure très particulière: la tige boucle ne fait que 17 nt et l'appariement est de 100% contrairement aux miRNAs classiques. A la sortie du noyau le pre-mir-451 ne peut être correctement chargé dans Dicer qui requière au moins une tige boucle supérieure à 19 nt. Le pre-mir-451 est donc directement chargé dans les protéines Ago. Chez les mammifères le chargement dans Ago2 va générer un (ac)-pre-mir-451(Ago cleaved pre-mir-451) de 30 nt. Ceci est possible grâce à l'activité nucléase de Ago2 mais aussi car l'appariement est de 100% sur la tige boucle de pre-mir-451. Il va y avoir ensuite action d'une autre nucléase (pour l'instant inconnue) qui va permettre de maturer le (ac)-pre-mir-451 en mir-451. [Adapté de Yang et Lai 2010]

et al. 2005) et PACT (Protein ACTivator of PKR) (Lee, Hur et al. 2006). Des études ont indiqué que TRBP serait un régulateur négatif de la protéine kinase dépendante des ARN double brin PKR, tandis que PACT serait un activateur de PKR. A l'heure actuelle, on ignore s'il existe une connexion fonctionnelle entre la voie des miRNAs et la signalisation cellulaire par PKR. Il existe certain pre-miRNA qui sont directement chargé dans le complexe RISC sans passer par la protéine Dicer présentant une nouvelle voie de régulation indépendante de cette dernière (**Figure 23**).

Après le clivage par Dicer, le miRNA duplex est chargé dans une protéine Argonaute (Ago1 chez la Drosophile) de façon à générer le complexe effecteur RISC. Un seul brin du duplex va être chargé dans la protéine Argonaute (le brin guide), tandis que l'autre brin (le brin passager) est majoritairement dégradé ou peut être chargé dans Ago2 chez la drosophile (Czech, Zhou et al. 2009). Les études sur les siRNAs indiquent que la stabilité thermodynamique des deux extrémités du duplex détermine quel brin va être conservé (Khvorova, Reynolds et al. 2003; Schwarz, Hutvagner et al. 2003). De nombreuses études sur les miRNAs ont suggéré que les mêmes règles pouvaient s'appliquer à la plupart des miRNAs, mais pas à tous car pour certains miRNAs on retrouve la même quantité de chaque brin chargé dans le RISC (Khvorova, Reynolds et al. 2003).

Dicer, TRBP et les protéines Argonautes contribuent à l'assemblage du RLC chez l'homme, alors que chez la drosophile le RLC comprend Dcr1, Loqs et ago1. La présence de plusieurs protéines Ago différentes dans une cellule entraine un choix multiple de protéines Ago pour les petits ARNs lors de l'assemblage du RISC. Chez *D. melanogaster*, le facteur important qui détermine ce choix est la structure du précurseur ARN (Forstemann, Horwich et al. 2007). Les miRNAs duplex qui présentent souvent des hétéroduplexes sont préférentiellement chargés dans Ago1, alors que les siRNAs souvent

Cellule avec un transcriptome de type A



Cellule avec un transcriptome de type B

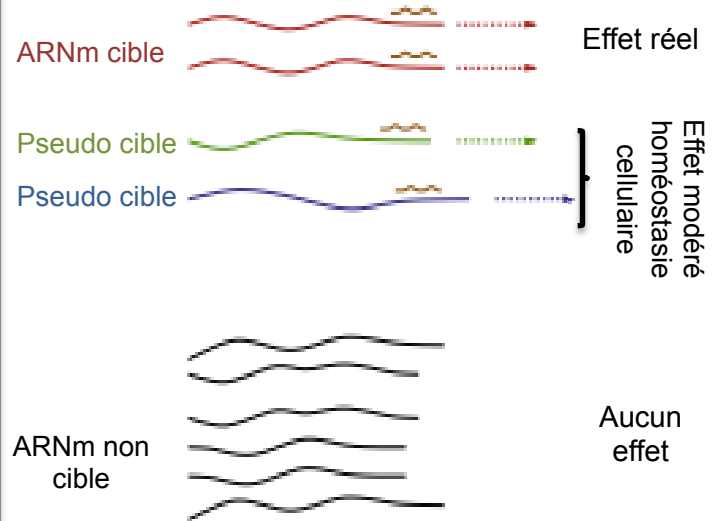


Figure 24: Théorie des cibles tampon des microRNAs

Selon les premières observations sur les miRNAs, il existerait plusieurs centaines de cibles pour chaque miRNA. Mais une hypothèse intéressante tente de démontrer qu'un miRNA n'aurait que quelques véritables cibles (rouge) et que les autres seraient des cibles tampons (bleu, vert, violet). Pour démontrer cela, on prend l'exemple de deux types cellulaires avec un transcriptome différent en quantité de pseudo cibles mais avec la même quantité de véritable ARNm cibles. Dans les cellules de type A la production d'un miRNA n'entraîne pas d'effet phénotypique visible car il y a beaucoup de pseudo cibles (dont la modulation de la traduction peut être régulé par l'homéostasie cellulaire, beaucoup de gène peuvent avoir des variations d'expression importantes sans entrainer d'effet sur la viabilité cellulaire) ce qui entraîne un faible effet sur les véritables cibles. Dans les cellules de type B, il y a peu de pseudo cibles et leur répression est tamponnée par la cellule. Dans ce cas, on peut voir apparaître les véritables cibles qui doivent être finement régulées en dose. (les flèches représentent la traduction: plus le trait est épais plus il y a de traduction et moins il y a d'effet de répression *via* les miRNAs). [Adapté de Seitz 2009]

sous forme de duplexe sans mésappariements sont incorporés dans Ago2 (Okamura, Liu et al. 2009). Ce mécanisme est conservé chez *C. elegans*, où les protéines Ago-like ALG-1 s'associent avec des petits ARNs précurseurs ayant des mésappariements tandis que RDE-1 s'associe avec des précurseur ARNdb parfaits, (Steiner, Hoogstrate et al. 2007). Cette modeste quantité de complémentarité a permis aux miRNAs animaux d'accumuler des centaines de cibles car ils reconnaissent majoritairement leur cible grâce à 7 nucléotides situés en 5' (nucléotide 2 à 8) que l'on nomme la séquence « seed » (Bartel 2009; Friedman, Farh et al. 2009). Cette théorie des cibles abondantes est remise en cause. En effet, en émettant l'hypothèse que beaucoup de cibles des miRNAs seraient des cibles tampons régulées par l'homéostasie cellulaire et non des ARNm réellement régulés par ces miRNAs, il existerait que peu de véritables cibles pour un miRNA donné (Seitz 2009) (**Figure 24**). Le manque de complémentarité entre le miRNA et sa cible empêche un clivage de la cible et entraîne majoritairement des phénomènes de déadenylation et de modulation de la traduction. Une complémentarité parfaite entraîne une dégradation de la cible ce qui est le cas des siRNAs.

Les défauts de régulation des miRNAs sont impliqués dans certaines maladies comme les cancers (Esteller 2011). Les miRNAs peuvent être utilisés comme marqueur thérapeutique par exemple certains microRNAs sont spécifiques des muscles, en particulier du coeur. La survenue d'un infarctus du myocarde entraîne une fuite de certains miRNAs dans la circulation et leur dosage pourrait être un marqueur biologique d'infarctus (Wang, Zhu et al. 2010; Zhong, Wang et al. 2010) ou comme marqueur de cancer (Lu, Getz et al. 2005). La voie des miRNAs ne semble pas avoir de rôle direct à l'heure actuelle dans la régulation des éléments transposables bien que le « crosstalk » entre la voie des endo-siRNAs et celle des miRNAs pourrait faire entrevoir une possible intervention de ceux ci

dans la régulation des éléments transposables. Mais récemment, une voie majeure de régulation des éléments transposable a été mise en évidence, celle des piRNAs.

a).3 La voie des petits ARN se liant à PIWI (piRNAs)

Découverte chez la souris et le rat mais majoritairement étudiée dans les ovaires de drosophile, la voie des piRNAs est composée de très nombreux acteurs qui permettent de maintenir l'intégrité des organes reproducteurs femelle ou male en fonction des organismes. Cette voie de petits ARNs mise en évidence très récemment semble conservée dans de très nombreux organismes eucaryotes tels que l'homme (Qiao, Zeeman et al. 2002), la souris (Kuramochi-Miyagawa, Kimura et al. 2001), la drosophile (Lin and Spradling 1997), le poisson zèbre (Tan, Lee et al. 2002), les ciliés (Mochizuki, Fine et al. 2002) etc.... Il semblerait que la voie des piRNAs soit le régulateur majeur des éléments transposables dans la lignée germinale.

La première protéine découverte comme acteur principal de cette voie de régulation est la protéine PIWI (P element Induced WImpi testes) qui a été mise en évidence dans une recherche de mutants affectant la division asymétrique des cellules souches germinales de drosophile (Lin and Spradling 1997). L'analyse fine du rôle de cette protéine a montré qu'elle était un acteur clef de la voie d'ARN interférence chez la drosophile (Pal-Bhadra, Bhadra et al. 2002), et que sa mutation induisait une perte de contrôle des éléments transposables tels que gypsy (Sarot, Payen-Groschene et al. 2004) ou Idefix (Desset, Buchon et al. 2008) chez la drosophile, ou encore L1 (Carmell, Girard et al. 2007) chez la souris. Le rôle majeur de cette protéine dans cette voie de régulation a donné son nom à la voie: piRNA (PIWI interacting RNA ou ARN se liant à la protéine PIWI)

Chez la souris, il a été identifié dans un premier temps deux orthologues de PIWI, appelés Miwi et Mili (Deng and Lin 2002; Kuramochi-Miyagawa, Kimura et al. 2004),

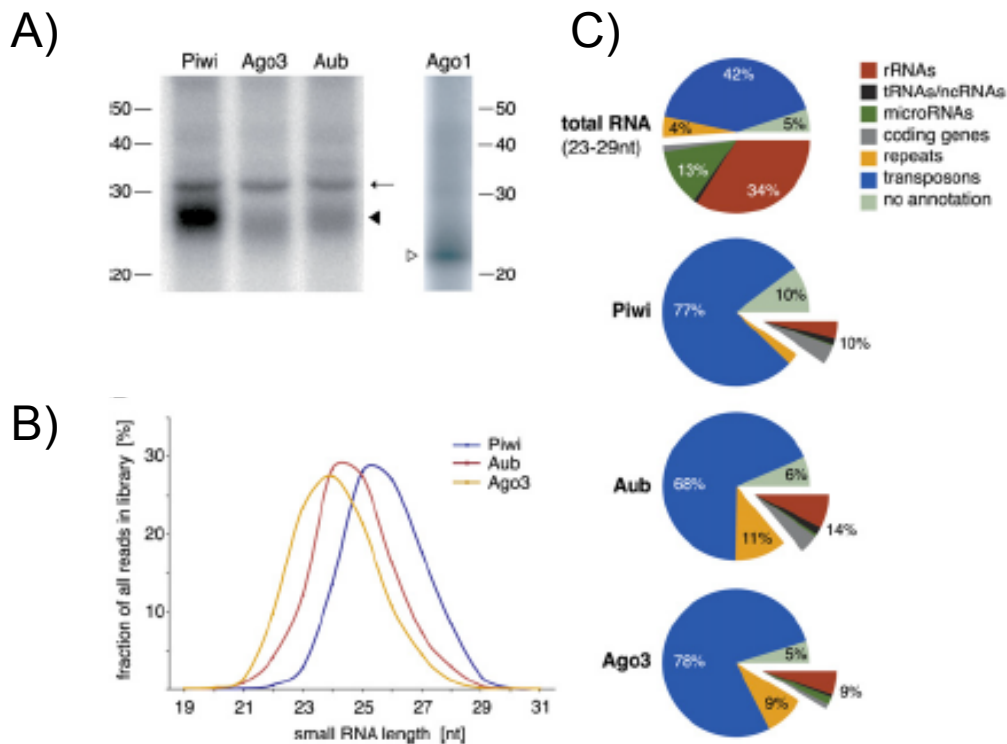


Figure 25: Caractéristiques des piRNAs de drosophile

A) Sur ce northern blot on peut observer que la taille des petits ARNs se liant à PIWI, Aubergine, et Ago3 est comprise entre 24 et 30 nucléotides (pointe de flèche noire) contrairement aux miRNAs liés à Ago1 (pointe de flèche blanche). **B)** Distribution de ces petits ARNs en fonction de leur taille et de la protéine avec laquelle ils ont été précipités. **C)** Homologie de séquences de ces piRNAs avec le génome, la majorité d'entre eux correspondent à des séquences d'éléments transposables (bleu). [Brennecke *et al.*, 2007]



Figure 26: Localisation des piRNAs sur le chromosome X de la drosophile

Schéma du chromosome X de drosophile avec représentation de la zone pérécitromérique (gris foncé) où est détectée une grande densité de piRNAs (barres vertes).

ayant un rôle dans le bon déroulement de la spermatogenèse mais pas dans la maintenance des cellules souches germinales. Ce n'est qu'en 2007 (Carmell, Girard et al. 2007) qu'une troisième protéine, Miwi 2, a été mise en évidence comme ayant un rôle dans le maintien de l'intégrité des cellules souches germinales. Il semblerait donc que les protéines de la famille PIWI soient conservées chez la souris pour leur rôle dans le maintien de l'intégrité de la lignée germinale.

En 2006, deux laboratoires mettaient en évidence que les protéines Miwi et Mili pouvaient se lier à de petits ARNs ayant des caractéristiques très particulières, les piRNAs (Aravin, Gaidatzis et al. 2006; Girard, Sachidanandam et al. 2006). Leur longueur de 24 à 30 nts est supérieure à celle des autres petits ARNs connus (**Figure 25**), ils semblent provenir de quelques loci ponctuels du génome d'une taille allant de 1 à plus de 100 kb (**Figure 26**). Enfin, ils sont majoritairement complémentaires des séquences des ARNm des éléments transposables.

Les caractéristiques de ces petits ARNs semblent être proches des petits ARNs appelés rasi-RNA (repeat associated small interfering RNA), qui sont des ARNs régulateurs pouvant hybrider les séquences répétées et qui sont majoritairement étudiés chez les plantes.

Deux autres protéines majeures de la famille argonaute ont été mises en évidence comme jouant un rôle dans la voie des piRNAs chez la drosophile : ce sont les protéines Aubergine (Aub), qui est entre autre responsable de l'élaboration des axes embryonnaires (Dorso-ventral et Antéro-postérieur) en régulant les morphogènes Gurken et Oskar, et la protéine Argonaute 3 (AGO3).

Les piRNAs sont formés en absence de protéine Dicer (Vagin, Sigova et al. 2006) . En 2007, un article publié par le de Greg Hannon décrit pour la première fois le mécanisme d'action des piRNAs sur le contrôle des éléments transposables dans les ovaires de

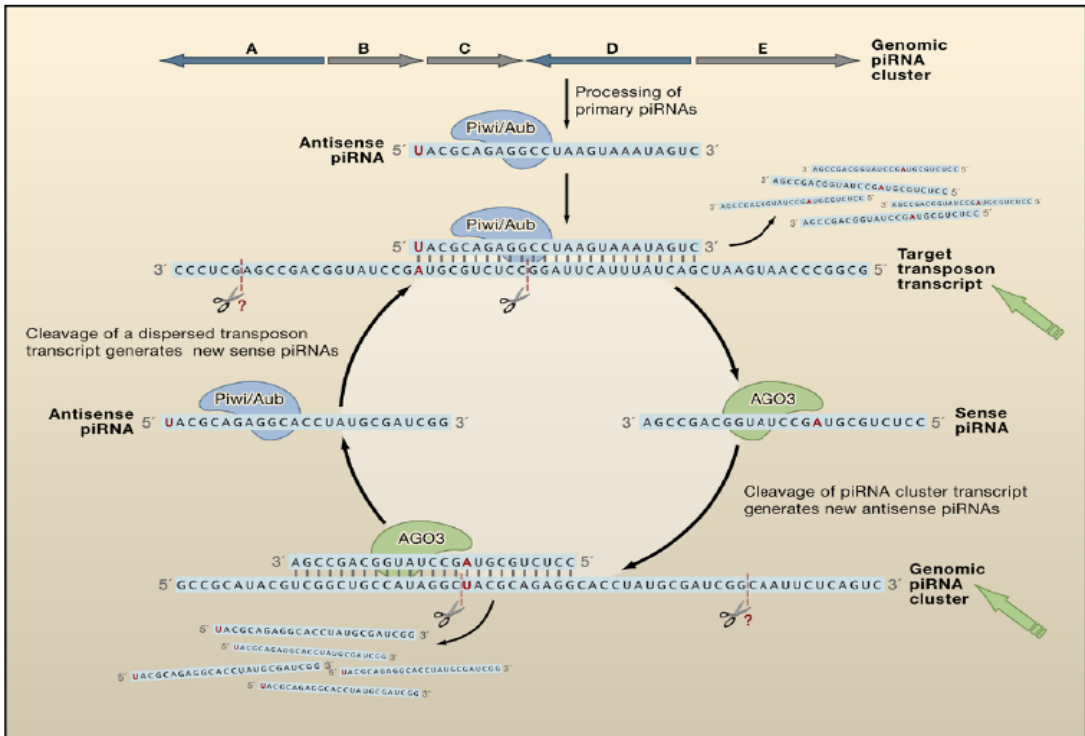


Figure 27: La voie canonique des piRNAs chez la drosophile

Le cycle commence avec le clivage des piRNAs primaires qui sont dérivés de copies de transposons défectueux situées dans les régions hétérochromatiques (marquées A à E en haut de la figure), par les protéines PIWI. Les piRNAs primaires sont antisens aux éléments transposables actifs et sont liés aux protéines PIWI ou Aubergine (Aub). Ensemble, ces complexes PIWI/Aub-piRNA primaire ciblent puis clivent les transcrits qui leur sont complémentaires, générant de petits ARNs sens qui se fixent à la protéine Ago3. Ces petits ARNs sont aussi appelés piRNAs. Ce second complexe Ago3-piRNA dirige un second événement de clivage d'un autre transcrit issu du cluster hétérochromatique, ce qui crée de nouveaux piRNAs antisens capables de se fixer à PIWI ou Aubergine, installant un cycle d'amplification auto-renforcé. [Brennecke *et al.*, 2007].

drosophile, « le modèle *Ping-Pong* ». Les séquences des piRNAs sont homologues à celles de loci péricentromériques (cluster de piRNAs) composés de séquences d'éléments transposables défectives (**Figure 27**). Ces clusters de piRNAs vont être transcrits en un ou plusieurs précurseurs ARN simple brin. Ces précurseurs simple brin sont majoritairement des petits ARNs antisens à l'ARN messager des éléments transposables. Ce mécanisme est dû à une insertion préférentielle des éléments transposables défectifs en orientation inverse du sens de transcription des clusters de piRNAs (**Figure 28**). Il a récemment aussi été montré que la protéine Qin, qui est une E3 ubiquitine ligase à domaine Tudor, permettrait de maintenir un biais en faveur des piRNAs antisens (Zhang, Xu et al. 2011).

Les grands transcrits précurseurs vont ensuite être coupés en piRNAs dit primaires. Ces piRNAs sont pris en charge par les protéines PIWI et Aubergine qui ciblent les ARNm qui sont complémentaires aux piRNA créant ainsi de petits ARNs sens dit piRNAs secondaires. Les piRNAs secondaires sont alors pris en charge par la protéine AGO3. Ce nouveau complexe nucléoprotéique ARN/Ago3 cible à son tour les transcrits issus des cluster de piRNAs créant ainsi une boucle d'amplification appelée modèle *Ping-Pong* (Brennecke, Aravin et al. 2007) (**Figure 29 A**). Les piRNAs primaires ont en majorité une base uracile en 5' ceci qui est dû à l'affinité de la protéine PIWI pour les petits ARNs débutant par un nucléotide uracile (Kawaoka, Izumi et al. 2011). L'hybridation du piRNA sur son ARNm cible ainsi que la coupure dix nucléotides après la première base génère des piRNAs secondaires avec une base adénine en dixième position, ce qui est le reflet d'une signature de *Ping-Pong* (**Figure 29 A,B**). On peut donc trouver des paires de piRNAs chevauchant sur 10 bases (paires de *Ping-Pong*). Ce mécanisme de régulation est conservé chez les mammifères. En effet, il a été montré chez la souris que MIWI2 peut s'associer aux petits ARNs issus des transcrits péricentromériques et cibler les ARNm des éléments transposables, où une boucle *Ping-Pong* peut se mettre en place grâce à la protéine MILI

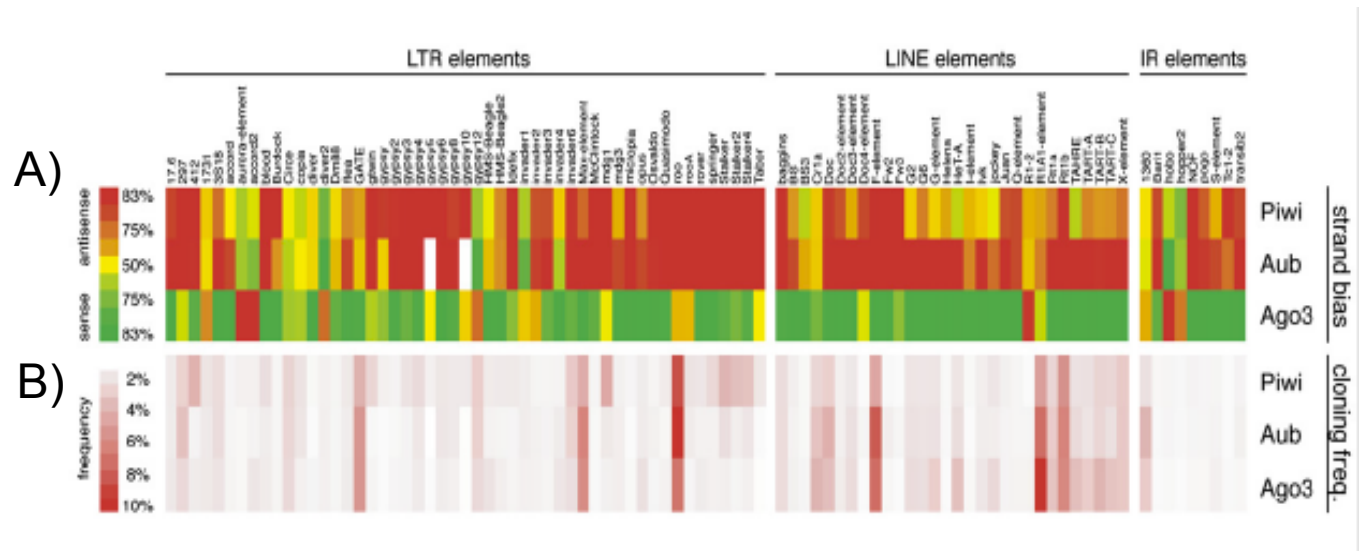


Figure 28: Biases de brin des piRNAs

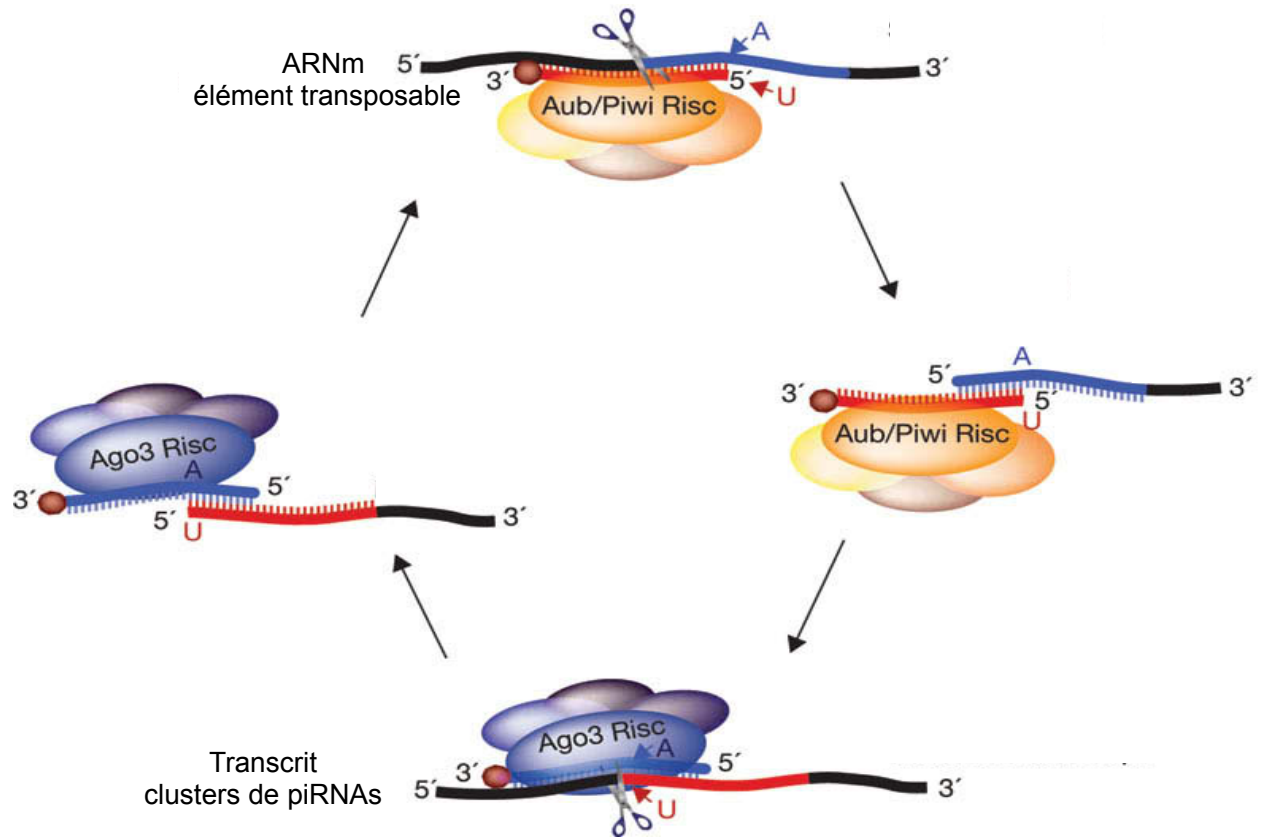
Sur le panneau du haut **(A)** est indiqué le biais de sens des piRNAs liés avec les trois protéines Argonautes: PIWI, Aubergine et Ago3. Le sens est rapporté au transcrit des éléments transposables canoniques indiqués au dessus qui sont classé en trois catégories: les retrotransposons à LTR, les Long Interspersed Nuclear Elements (LINE) et les transposons (IR elements) . L'intensité de couleur indique le degré de biais, le vert indique une majorité de piRNAs sens, le rouge de piRNAs antisens et le jaune l'absence de différence entre sens et antisens. PIWI et Aubergine sont majoritairement liées avec des piRNAs antisens tandis que Ago3 est liée à des piRNAs sens. Sur le panneau du bas **(B)** est représenté la fréquence de clonage de ces piRNAs précipités avec les différentes protéines. Une majorité de piRNA antisens est retrouvée. (l'intensité de rouge indique la fréquence de clonage). [Brennecke *et al.*, 2007]

(**Figure 30**). De façon surprenante surprenant les piRNAs primaires chez la souris ont une base adénine en position 10 (**Figure 31**) alors que les piRNAs secondaires ont une base uracile en 5' ce qui est l'inverse de ce qui est observé chez la drosophile (Aravin, Sachidanandam et al. 2008).

En plus des protéines Argonaute de la famille PIWI, de nombreux autres facteurs ont été mis en évidence majoritairement chez la drosophile dans la voie des piRNAs:

Les protéines à domaine Tudor : L'une des familles les plus importantes (au moins quantitativement) de protéines ayant un rôle dans la voie des piRNAs sont les protéines de la famille Tudor. Cette famille a reçu son nom lors d'un crible génétique chez *Drosophila melanogaster* visant à rechercher des gènes exprimés maternellement dont la mutation se traduit par une létalité ou une stérilité dans la descendance (Boswell and Mahowald 1985). La structure du domaine Tudor a été décrite plus tardivement en 1997 et a été majoritairement étudiée chez l'homme (Callebaut and Mornon 1997; Ponting 1997). Les protéines à domaine Tudor ont la capacité de lier les résidus arginine ou lysine diméthylés de certaines protéines. Une étude récente a montré que les protéines PIWI de drosophile, souris et xenope contiennent des résidus arginines diméthylés symétriquement (sDMAs), et que le facteur médiateur de cette modification post-traductionnelle est la protéine arginine méthyltransférase 5 (PRMT5) (Kirino, Kim et al. 2009). Les sDMAs sont l'une des diverses modifications par groupement méthyl trouvées sur les arginines spécifiques de certaines protéines et sont connues pour modifier la capacité d'une protéine à effectuer ses activités biologiques (Bedford and Clarke 2009). Il n'est donc pas surprenant que les protéines à domaine Tudor agissent dans la régulation des éléments transposables *via* la voie des piRNAs grâce à leurs interactions avec les sDMAs des protéines de la famille PIWI (**Figure 32**) (Siomi, Mannen et al. 2010).

A)



B)

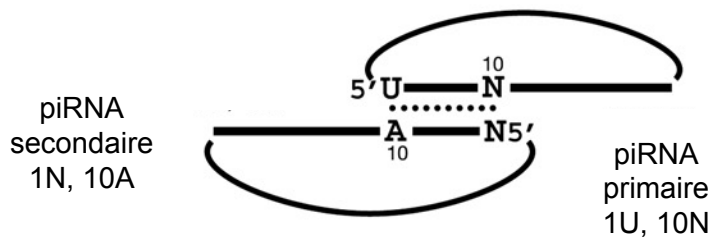


Figure 29: piRNAs primaires et secondaires chez la drosophile

Il existe deux types de piRNAs: les primaires qui ont majoritairement une base uracyle en 5' (5'U) et qui sont associés avec les protéines PIWI et Aubergine et les secondaires qui contiennent une base adénine en dixième position et sont associés à la protéine Argonaute 3 (Ago3) en majorité. **A)** La première catégorie (rouge) contient en majorité des piRNAs qui dérivent des transcrits des clusters de piRNAs et qui sont pris en charge par PIWI grâce à son affinité pour les 5'U. Le complexe PIWI/piRNA primaire va alors cibler les ARNm issus des éléments transposables et les couper dix nucléotides en amont de la base 5' du piRNA primaire, ce qui va générer des petits ARNs ayant une base Adénine en dixième position (bleu). Ces petits ARNs vont être pris en charge par la protéine Ago3. **B)** Représentation schématique de l'appariement entre les piRNAs primaires et secondaires. [Adapté de Seila et Sharp 2008]

Le domaine Tudor a été trouvé dans les protéines de nombreux organismes eucaryotes, comme les champignons, les plantes et les animaux. Les protéines PIWI semblent avoir conservé évolutivement leurs résidus arginine diméthylés. Chez la drosophile, de nombreuses protéines à domaine Tudor ont été mises en évidence dans la régulation des éléments transposables *via* la voie des piRNAs comme Krimper (Krimp) (Lim and Kai 2007), Vreteno (Vret) (Handler, Olivieri et al. 2011; Zamparini, Davis et al. 2011), Tejas (Patil and Kai 2010), Kumo (Anand and Kai 2011), Qin (Zhang, Xu et al. 2011) etc. De nombreuses protéines à domaine Tudor sont aussi impliquées dans la voie des piRNAs chez la souris (Vagin, Wohlschlegel et al. 2009) (**Figure 33**).

Les nucléases : Il a dans un premier temps été montré que la perte de fonction des gènes *zucchini* (*zuc*) et *squash* (*squ*), qui codent pour des nucléases putatives, altère la production de piRNAs et occasionne une levée de répression des éléments transposables chez la drosophile (Pane, Wehr et al. 2007). Ces protéines pourraient être une alternative aux protéines Dicer. La protéine Squash possède un domaine RNase tandis que Zucchini est une protéine de la famille des phospholipases D ayant une activité nucléase putative. Des études récentes chez la souris ont montré que MITOPLD, l'orthologue de Zuc, est une protéine de la famille des phospholipases D associée aux mitochondries (Watanabe, Chuma et al. 2011) (Huang, Gao et al. 2011).

Les méthyl-transférases : Les piRNAs possèdent un groupement 2'-O-méthyl à leur extrémité 3'. Une enzyme Pimet (dmHen1) (Horwich, Li et al. 2007) assure cette méthylation après interaction avec les protéines Argonautes (**Figure 34**). L'absence de cette protéine entraîne une modification de l'abondance et de la taille des piRNAs mais ne semble avoir qu'un rôle mineur sur la régulation des éléments transposables. Cette modification a d'abord été trouvée sur les miRNAs et siRNAs des plantes (Yang, Ebright

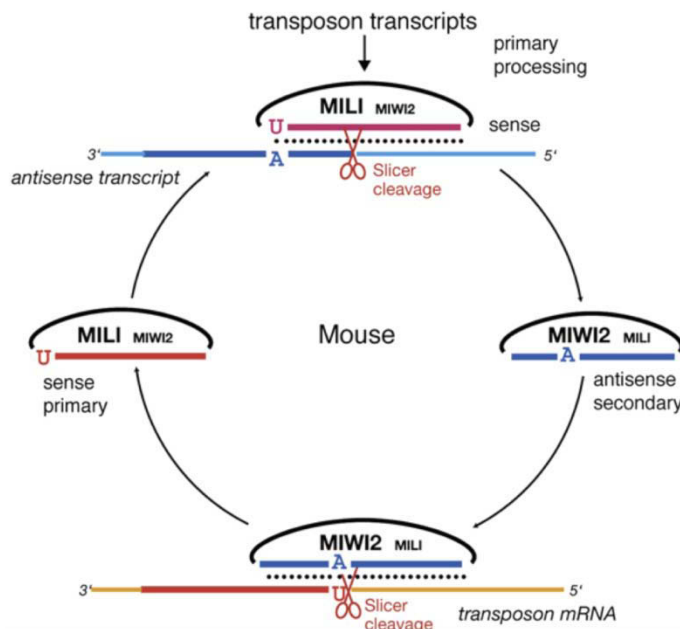


Figure 30: Le modèle *Ping-Pong* chez la souris

Chez la souris il existe un modèle d'amplification des piRNAs comme chez la drosophile. les piRNAs issus des clusters (bleu) sont majoritairement liés à la protéine MIWI2 et les piRNAs issus de la dégradation des ARNm (jaune) des éléments transposables sont liés à la protéine MILI (rouge). [Aravin *et al.*, 2008]

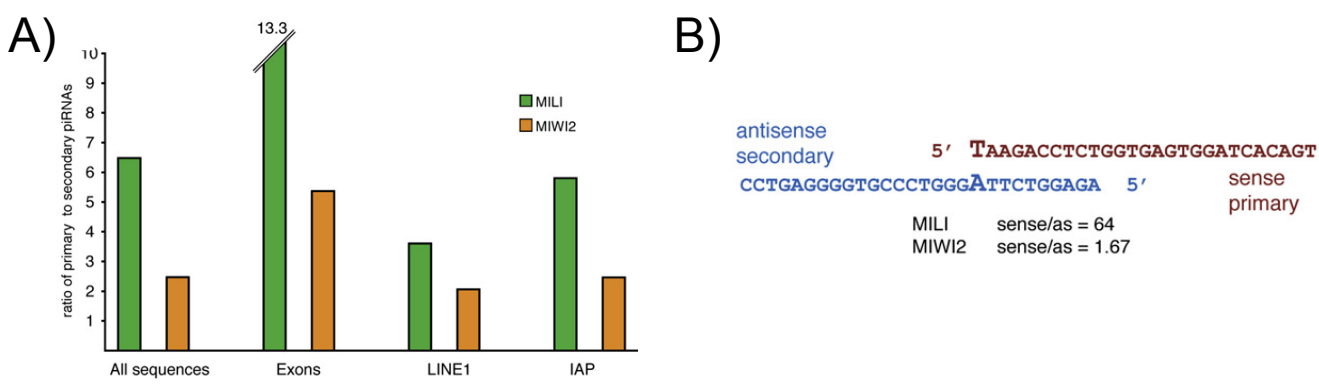


Figure 31: piRNAs primaires et secondaires chez la souris

Comme chez la drosophile, il existe deux types de piRNAs chez la souris. Toutefois il existe une différence majeure: chez la souris les piRNAs primaires dérivent majoritairement des ARNm des éléments transposables et les secondaires proviennent des clusters de piRNAs. **A)** Sur ce diagramme est représenté le ratio de piRNAs primaires/piRNAs secondaires et on peut observer que les piRNAs primaires sont majoritairement liés à la protéines MILI (vert) et peu à la protéine MIWI2 (orange). **B)** Représentation schématique de l'appariement entre les piRNAs primaires (rouge), qui sont majoritairement sens, et les piRNAs secondaires (bleu), qui sont majoritairement antisens. [Aravin *et al.*, 2008]

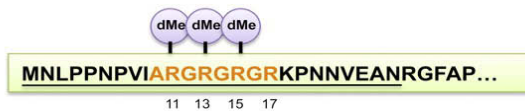
et al. 2006) et est conservée chez les piRNAs de mammifères (Kirino and Mourelatos 2007; Kirino and Mourelatos 2007; Ohara, Sakaguchi et al. 2007)

Les hélicases ARN : Trois hélicases ARN majeures ont été mises en évidence dans la régulation des éléments transposables et la production des piRNAs. Ce sont les protéines Armitage, qui est l'homologue de l'hélicase ARN d'*Arabidopsis thaliana* SDE3 (Cook, Koppetsch et al. 2004), Spindle-E et Vasa (Liang, Diehl-Jones et al. 1994; Aravin, Klenov et al. 2004; Tomari, Du et al. 2004; Vagin, Klenov et al. 2004). De façon intéressante Vasa contient des sDMAs et pourrait donc interagir avec les protéines à domaines Tudor.

En plus de ces différentes familles de protéines, il a été mis en évidence que d'autres facteurs étaient impliqués dans la régulation des éléments transposables *via* les piRNAs. Il s'agit entre autres de la protéine Maelstrom (Mael) (Findley, Tamanaha et al. 2003) et du gène *cutoff* (*cuff*) (Chen, Pane et al. 2007) qui semble être l'homologue du gène *rail* de levure qui est impliqué dans la maturation des ARN ribosomiques (Xue, Bai et al. 2000; Kim, Krogan et al. 2004). Par exemple la perte de *cuff* conduit à une dérepression des éléments transposables similaire à la mutation d'*aubergine* (Chen, Pane et al. 2007).

De manière très intéressante la grande majorité de ces protéines, en plus d'être conservées chez les mammifères, sont localisées dans une structure électronique dense qui entoure le noyau des cellules germinales et qui est appelée « nuage » (**Figure 35**). Les protéines Argonaute : Aubergine et Ago3 (Brennecke, Aravin et al. 2007) sont aussi présentes dans cet organite périnucléaire qui a été caractérisé pour la première fois dans les cellules nourricières de l'ovaire de drosophile (Mahowald 1971; Hay, Ackerman et al. 1988; Hay, Jan et al. 1988). Cette structure servirait à remodeler les complexes ribonucléoprotéiques issus du noyau avant leur transport dans le cytoplasme. Mais le nuage

Aub N-terminal



AGO3 N-terminal



Figure 32: Méthylations des arginines des protéines Aubergine et Argonaute 3

Représentation schématique de la séquence N-terminale de deux protéines de la famille PIWI de drosophile: Aubergine (Aub) et Argonaute 3 (Ago3). Les acides aminés représentés en orange sont les arginines qui peuvent être symétriquement diméthylé (sDMAs) (rond violet) et donc être potentiellement reconnues par les protéines à domaine Tudor. [Siomi et al., 2010]

Protéine Tudor drosophile

Proteine Tudor	Fonction	Reference
Tudor	Germline development, piRNA biogenesis	Schupbach and Wieschaus 1986; Nishida et al. 2009
Spindle E	piRNA biogenesis, histon methylation	Pal-Bhadra et al. 2004; Lim and Kai 2007
Tudor-SN	RNAi pathway, implicated in RNA editing and degradation	Caudy et al. 2003; Scadden 2005
CG7082	Unknown	
CG8589	Unknown	
CG8920	Unknown	
CG13472	Unknown	
CG14303	Unknown	
Female sterile (1) Yb	Male germline stem cell maintenance	Szarmay et al. 2009
Krimper	piRNA biogenesis	Lim and Kai 2007
MBD-R2	Unknown	
Ovarian tumor	RNA localization nurse cell chromosome dispersion	Goodrich et al. 2004
Polycomblike	Chromatin modification	Lonie et al. 1994
Survival motor neuron	snRNP assembly	Miguel-Aliaga et al. 2000
Yu	Formation of long-term memory, protein kinase A binding	Lu et al. 2007
CG4771	Unknown	
CG9684	Unknown	
CG9925	Unknown	
CG15042	Unknown	
CG15930	Unknown	
CG17454	Unknown	
CG30390	Unknown	
CG31755	Unknown	

Protéine Tudor Souris

Proteine Tudor	Fonction	Reference
TDRD1/MTR-1	Spermatogenesis, piRNA biogenesis	Chuma et al. 2006; Reuter et al. 2009
TDRD2/TDRKH	Implicated in spermatogenesis	Chen et al. 2009
TDRD3	Unknown	
TDRD4/RNF17	Implicated in spermatogenesis	Pan et al. 2005
TDRD5	Implicated in spermatogenesis	Smith et al. 2004
TDRD6	Spermatogenesis, miRNA expression	Hosokawa et al. 2007; Vasileva et al. 2009
TDRD7/TRAP	Implicated in spermatogenesis	Hosokawa et al. 2007
TDRD8/STK31	Implicated in spermatogenesis	Chen et al. 2009
TDRD9	Spermatogenesis, piRNA biogenesis	Shoji et al. 2009
TDRD12	Unknown	
AKAP1/AKAP121	Mitochondrial metabolism	Livigni et al. 2006
ARID4A	Implicated in chromatin remodeling	Wu et al. 2008
ARID4B	Implicated in chromatin remodeling	Wu et al. 2008
KDM4A/JMJD2A	Histone demethylation, self-renewal in embryonic stem cells	Loh et al. 2007
KDM4B/JMJD2B	Histone demethylation	Fodor et al. 2006
KDM4C/JMJD2C	Histone demethylation, self-renewal in embryonic stem cells	Loh et al. 2007
LBR	Implicated in morphological granulocyte maturation	Cohen et al. 2008
MTF2	Unknown	
SETDB1/ESET	Histone methylation, maintenance of embryonic stem cell state	Bilodeau et al. 2009
SMN1	snRNP assembly	Gabanella et al. 2005
SMNDC1	SMN-related protein	Talbot et al. 1998
SND1/P100	Coactivators for signal transducer, implicated in RNA editing and degradation, fly Tudor-SN ortholog	Paukku et al. 2003; Scadden. 2005
PHF19	Unknown	
PHF20	Unknown	
PHF20L1	Unknown	
ZGPAT	Unknown	

Figure 33: Divers protéines à domaine Tudor de drosophile et de souris

Tableau contenant les différentes protéines à domaine Tudor annotées chez la drosophile et la Souris avec leurs fonctions, ainsi que les gènes codant de putatives protéines à domaine Tudor. Chez la drosophile CG4771 et CG30390 ont récemment été caractérisés et sont appelés respectivement *vreteno* et *sister of Yb*. Les protéines codées par ces gènes ont un rôle dans la fabrication des piRNAs et la régulation des éléments transposables.[Siomi et al., 2010]

a aussi été largement impliqué dans la régulation des éléments transposables car par exemple la mutation de *krimp* qui est nécessaire à son assemblage, engendre une altération de la production de piRNAs et de la régulation des éléments transposables (Lim and Kai 2007). L'interaction des protéines à domaine Tudor avec les protéines de la famille PIWI permettrait la localisation de ces dernières dans le nuage et pourrait aussi modifier leur activité. Il a été reporté dans la littérature que la voie des piRNAs pouvait être impliquée dans le remodelage de la chromatine.

b) Modifications de la chromatine

Il existe deux grands types de chromatine : une où l'ADN est particulièrement accessible et qui est pauvre en éléments répétés mais riche en gènes, que l'on appelle euchromatine et l'autre qui inversement contient très peu de gènes mais porte de nombreuses régions riches en éléments répétés, l'hétérochromatine. Cette dernière est la forme compactée de la chromatine et correspond à des régions peu ou pas transcrites. Il en existe deux formes: l'hétérochromatine constitutive et l'hétérochromatine facultative. La première correspond à des portions de chromosomes condensés à l'interphase dont les plus grandes parties se trouvent à proximité des centromères et des télomères et apparemment inactives dans toutes les cellules. La deuxième correspond à des portions d'hétérochromatine plus ou moins condensée et pouvant se situer sur n'importe quelle région chromosomique.

L'organisation de l'ADN génomique en une structure plus ou moins compacte qu'est la chromatine (**Figure 36**) est notamment assurée par les protéines histones. Ces dernières s'organisent en un cylindre d'octamères constitué de deux hétérodimères des histones H2A et H2B et de deux hétérodimères des histones H3 et H4 (Luger, Rechsteiner et al. 1997) (**Figure 37 A**) autour duquel s'enroulent 146 pb d'ADN génomique. C'est

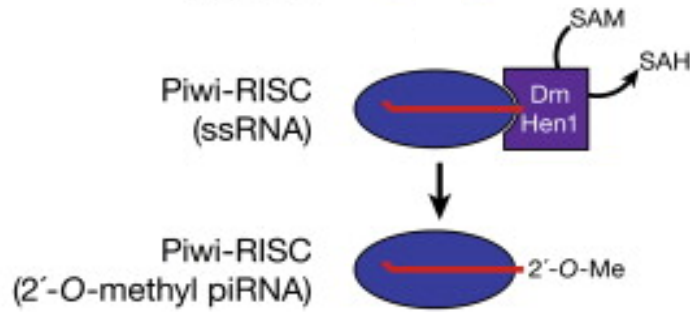


Figure 34: Méthylation des piRNAs

Les piRNAs (rouge) sont pris en charge par les protéines de la famille PIWI dans un complexe de type RISC (Piwi-RISC; bleu) où ils vont être maturés. La méthyl-transférase Hen1(violet) est associée à ce complexe et va méthyliser le piRNA en 3', générant une extrémité 2'- O – méthyl. [Horwich *et al.*, 2007]

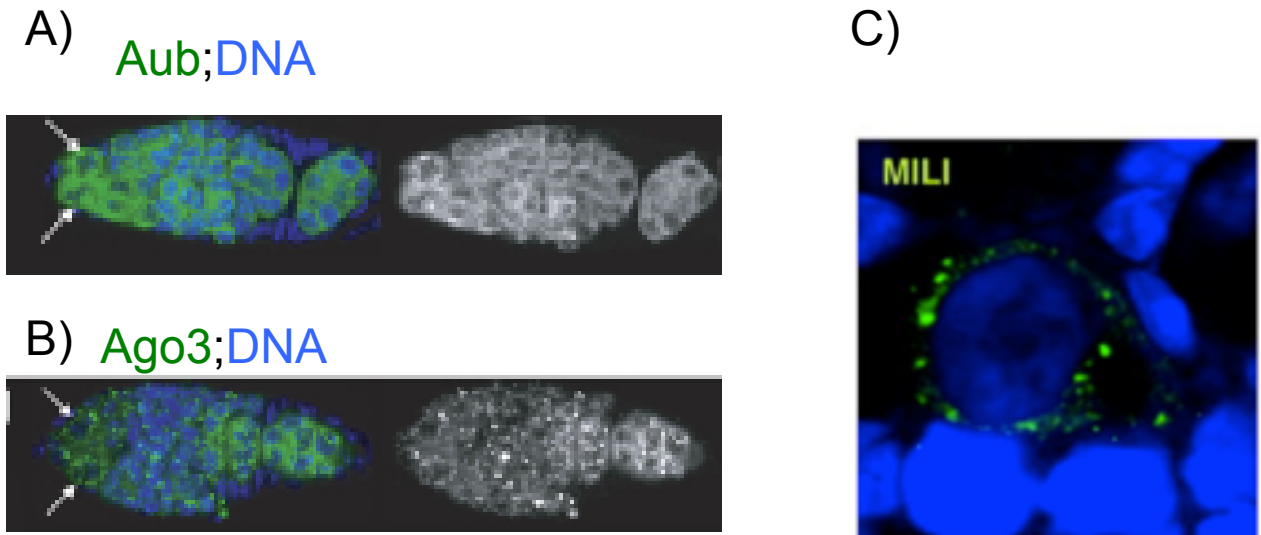


Figure 35: Le nuage chez la drosophile et la souris

Le nuage est une structure dense aux électrons qui entoure le noyau des cellules de la lignée germinale et qui contient de nombreuses protéines impliquées dans la voie des piRNAs. Chez la drosophile les protéines Aubergine (A) (vert ou blanc) et Ago3 (B) (vert ou blanc) sont présentes dans cette structure. Les flèches blanches représentent les cellules souches germinales C) Chez la souris, on peut observer la présence de MILI (vert). L'ADN est marqué en bleu. [Brennecke *et al.*, 2007; Vagin *et al.*, 2009]

l'unité structurale de la chromatine : le nucléosome. L'agencement des nucléosomes le long de l'ADN s'organise sur environ 200 pb et sur les quelques 50 pb séparant deux nucléosomes est fixé l'histone H1 (**Figure 37 B**). Les queues amino-terminales des histones H3 et H4 dépassent du corps globulaire de l'octamère d'histone et, par leur charge positive, assurent l'interaction avec l'ADN, chargé négativement (Fischle, Kiermer et al. 2001).

On peut distinguer quatre grands mécanismes de modifications induisant une compaction ou un relâchement de la structure chromatinienne (Lee and Young 2000) (Narlikar, Fan et al. 2002): les modifications covalentes et non covalentes (complexes de remodelage) d'histones, les protéines de compaction/décompaction de la chromatine et la méthylation de l'ADN.

b).1 Les modifications covalentes d'histones

Il existe deux modifications covalentes majeures d'histones: l'acétylation et la méthylation

L'acétylation des histones consiste en un transfert réversible d'un groupement acétyle sur les résidus lysines des queues amino-terminales des histones. La neutralisation par le groupement acétyle des charges positives des résidus lysines conduit à une déstabilisation de la structure nucléosomal par perte d'affinité des histones pour la double hélice d'ADN. Dès lors, il s'ensuit un relâchement de la structure chromatinienne permettant un accès plus facile à l'ADN pour les facteurs de transcription. Le transfert du groupement acétyle est assuré par des protéines nucléaires possédant une activité histone acétyle-transférase (HAT) (**Figure 38**). Si l'acétylation des histones est médiée par des protéines à activité HAT, le processus inverse, la désacétylation, peut être catalysé par des protéines possédant une activité histone désacétylase (HDAC) (Hassig, Tong et al. 1998),

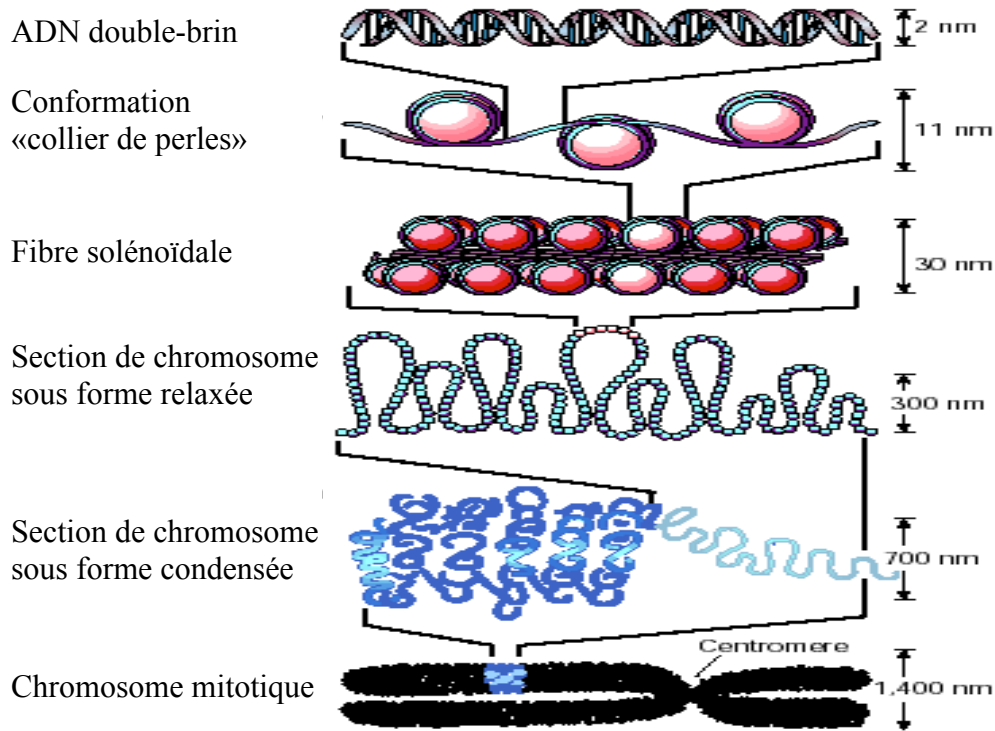


Figure 36: Représentation des différents niveaux de compaction de l'ADN

La chromatine est un complexe nucléo-protéique qui permet l'agencement de l'ADN dans le noyau et la régulation de la transcription. Ce complexe s'organise en différents niveaux de compaction, de l'état le plus relâché qu'est l'ADN double brin, au plus compacté, le chromosome mitotique. [Felsenfeld et Groudine, 2003]

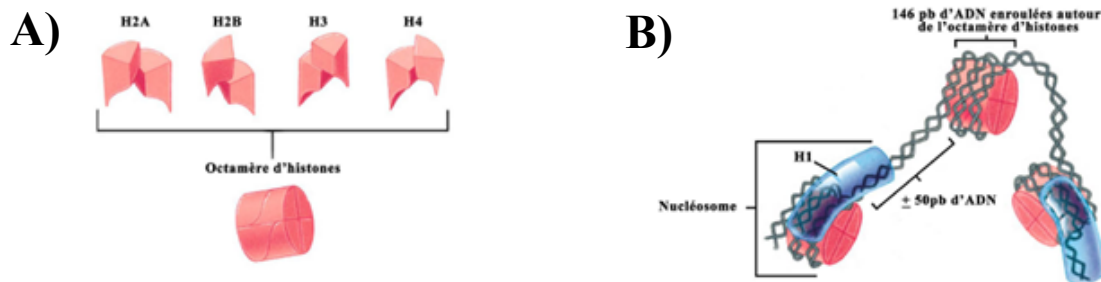


Figure 37: Formation et organisation des histones

A) L'octamère d'histone est composé de quatre protéines histones différentes (H2A, H2B, H3 et H4), chacune présente en deux exemplaires.

B) Le nucléosome se définit par l'enroulement de 146 pb d'ADN autour de l'octamère d'histones. Chaque nucléosome est séparé de l'autre par environ 50 pb d'ADN sur lesquelles se fixe l'histone H1.

[Principles of Biochemistry, Horton, 2006]

ce qui aura pour effet une compaction de la chromatine impliquant aussi la méthylation des histones.

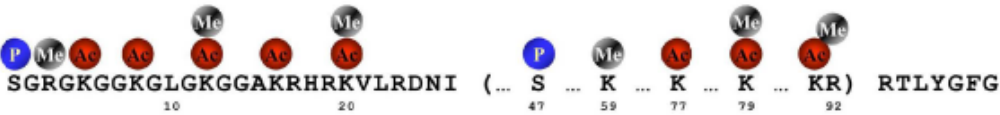
La méthylation des histones touche les résidus arginines et lysines des queues amino-terminales d'histones (**Figure 38**). Cette modification post-traductionnelle est médiée par les protéines histone méthyl-transférases (HMT) (Zhang and Reinberg 2001) qui assurent l'addition de un, deux, voire trois (à ce jour uniquement sur résidu lysine) groupement méthyle sur leurs résidus cibles. L'histone H3 apparaît comme l'histone la plus méthylée avec l'histone H4. La méthylation des histones peut aussi bien conduire à une activation ou à une répression de la transcription qui va dépendre de la nature du résidu modifié et de la combinaison de plusieurs modifications affectant les histones. Cette combinaison de modifications responsables de variations d'expression des gènes a été appelée le code histone (Jenuwein and Allis 2001).

A titre d'exemple du code histone, la méthylation de la lysine 4 de l'histone H3 (H3K4) est corrélée à une activation transcriptionnelle : les lysines, dès lors qu'elles sont méthylées peuvent servir de point d'ancrage à la protéine Cdh1p qui *via* son chromodomaine (CD), peut à son tour recruter une protéine ayant une activité HAT (Pray-Grant, Daniel et al. 2005) et donc permettre une activation du gène voisin. Il en est de même pour la triméthylation de l'histone H3 sur sa lysine 36 qui est une marque de l'élongation de la transcription (Lee and Shilatifard 2007) et qui peut être lié par le CD de la protéines MRG15 (Sun, Hong et al. 2008). A l'opposé, la méthylation des lysines 9 et 27 de l'histone H3 (H3K9, H3K27) est associée majoritairement à une répression transcriptionnelle *via* le recrutement de HP1 (Heterochromatic protein 1) ou des protéines de la famille des Polycomb qui jouent un rôle direct dans la condensation chromatinienne (Daniel, Pray-Grant et al. 2005). Chez la drosophile, on connaît deux protéines majeures qui permettent de méthyler l'histone H3 sur la lysine 9, il s'agit de Su(var)3-9 et Dsetdb1

HISTONE H3



HISTONE H4



HISTONE H2A



HISTONE H2B

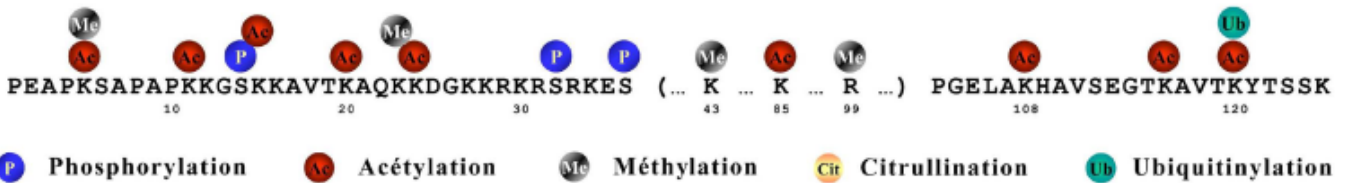


Figure 38: Principales modifications post-traductionnelles des histones

Représentation de la composition en acides aminés des queues N-terminales des différentes protéines histones et des modifications post-traductionnelles qu'elles peuvent subir: phosphorylation (bleu), acétylation (rouge), méthylation (gris), citrullination (jaune) et ubiquitinylation (vert). [Abcam et Upstate]

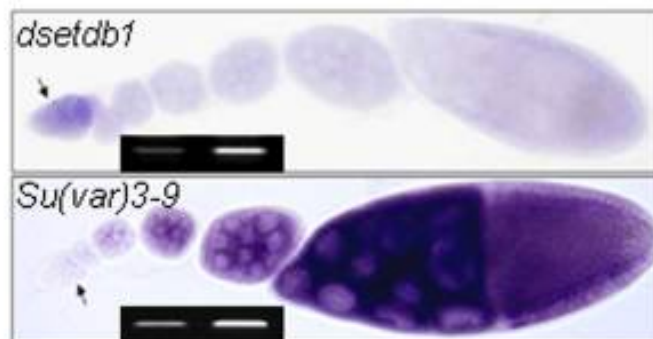


Figure 39: Profil d'expression des Histones Méthyl Transferase H3K9 dans l'ovaire de drosophile.

Chez la drosophile, deux gènes codent pour des protéines qui sont capables de méthylater les histones sur la lysine 9 de l'histone H3 : *dsetdb1* et *su(var)3-9*. Comme on peut voir sur cette hybridation *in situ* d'ovaire adulte de drosophile, ces deux gènes ont un profil d'expression temporellement et spatialement différent. *dsetdb1* est majoritairement exprimé précocement dans un compartiment appelé germarium (flèche noire) et *su(var)3-9* semble s'exprimer à la sortie du germarium. [Yoon J *et al.*, 2010]

(aussi connue sous le nom de Eggless). Leur profil d'expression dans les ovaires de drosophile semble être complémentaire (**Figure 39**).

Jusqu'à récemment, la méthylation des histones était considérée comme un processus irréversible, notamment lié au fait que le temps de demi-vie des histones et des lysines méthylées est identique. Très récemment, des enzymes capables de déméthyliser tant les résidus arginines que lysines ont été découvertes. Il s'agit notamment pour les résidus arginines et lysines des déméthylases PADI4 (Cuthbert, Daujat et al. 2004) (Wysocka, Allis et al. 2006) et LSD1 (Forneris, Binda et al. 2005).

D'autres modifications sont connues telles que la phosphorylation (répression et activation), la sumoylation (répression), l'ubiquitinylation ou la citrullination (activation) (**Figure 38**).

b).2 Les modifications non covalentes d'histones

L'organisation chromatinienne en nucléosomes peut être remodelée *via* des complexes multiprotéiques dont l'activité est dépendante de l'ATP. Ces complexes de remodelage de la chromatine (**Figure 40**) hydrolysent l'ATP *via* leur sous-unité catalytique à activité ATPase, et, au moyen de l'énergie ainsi libérée, provoquent un déplacement nucléosomal suite à la modification physique de la structure chromatinienne par torsion de l'ADN. Le déplacement nucléosomal permet ou restreint l'accès de l'ADN aux complexes de transcription (Kingston and Narlikar 1999). Les complexes de remodelage de la chromatine sont classés en trois familles selon l'homologie de séquence de leur sous-unité catalytique (Langst and Becker 2004). Il s'agit des familles SWI/SNF, ISW1 et Mi2-NURD.

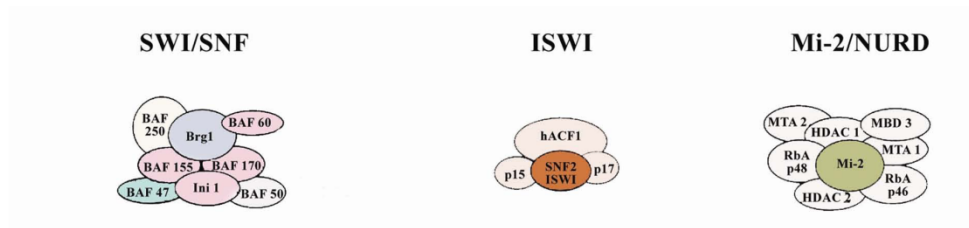


Figure 40: Les complexes de remodelage de la chromatine

Représentation des trois familles de complexes de remodelage de la chromatine. Les sous-unités catalytiques contenant l'activité ATPase sont représentées en bleu (SWI/SNF), orange (ISWI) et vert (NURD). [Narlikar *et al.*, 2002]

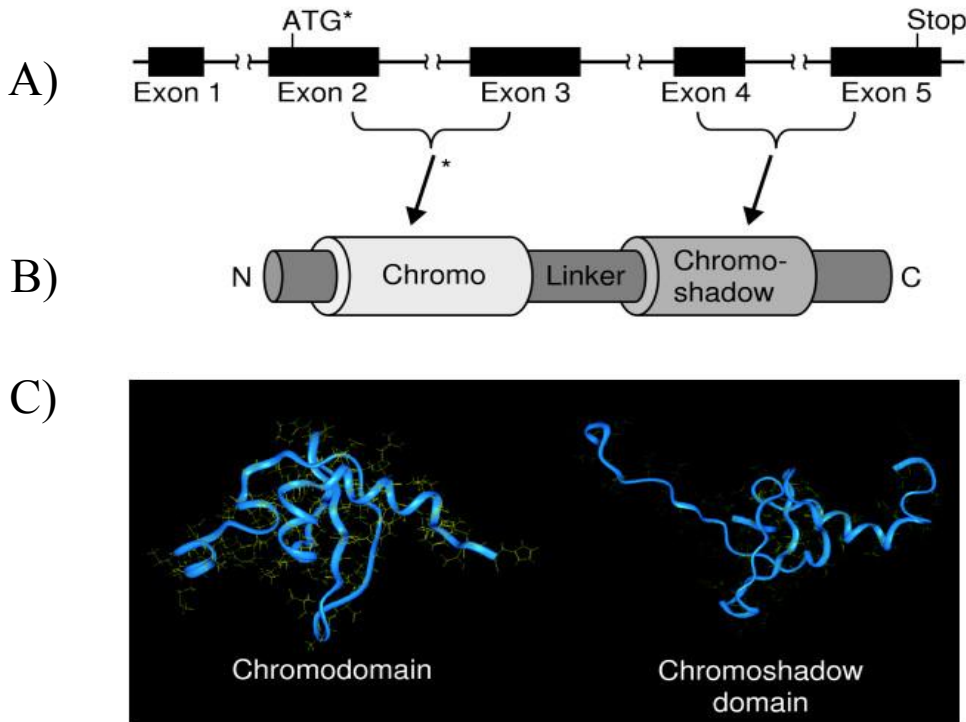


Figure 41: Structure de la protéine hétérochromatique 1 (HP1) et du gène qui la code

A) La structure génomique des gènes codant les diverses protéines HP1 est conservée. Chaque gène est composé de cinq exons séparés par quatre introns. Le début de traduction (ATG) et la fin (Codon stop) sont indiqués dans les exons deux et cinq respectivement. B) Représentation linéaire de la structure des protéines HP1 contenant un ChromoDomaine (CD) codé par une partie des exons deux et trois, une régions interdomaine Hinge et un ChromoShadowDomaine (CSD) codé par une partie des exons quatre et cinq. N: Région Amino-Terminale, C: Région Carboxy-Terminale. C) Représentation tridimensionnelle du CD et CSD de HP1. [Lomberk *et al.*, 2006]

c) Les protéines de compaction/décompaction de l'ADN

De nombreuses protéines permettent, en se fixant aux modifications covalentes d'histones, de compacter ou décompacter la chromatine. La protéine hétérochromatique 1 (HP1) et les protéines de la famille des Polycomb (PcG) et Trithorax sont des acteurs majeurs de ce remodelage.

c).1 HP1

HP1 est une petite protéine non histone identifiée comme une composante de l'hétérochromatine qui agit comme un modificateur de la variegation par effet de position (PEV) chez la drosophile (Eissenberg, James et al. 1990). HP1 est phylogénétiquement conservée et est retrouvée chez presque tous les eucaryotes à l'exception de la levure (Singh, Miller et al. 1991). Il existe plusieurs isoformes de HP1. En général, les protéines HP1 se composent de deux domaines hautement conservés: un chromodomaine (CD) N-terminal, qui est requis pour l'interaction avec la chromatine (Paro and Hogness 1991), et un chromoshadow domaine (CSD) C-terminal, qui permet l'interaction avec d'autres protéines à CSD (Aasland and Stewart 1995). Ces deux domaines sont séparés par une région charnière Hinge (HIN), de longueur variable, qui a été impliquée dans diverses interactions (**Figure 41**), y compris avec l'histone H1, la liaison à l'ADN et l'ARN, ainsi qu'à la chromatine (Zhao, Heyduk et al. 2000; Nielsen, Oulad-Abdelghani et al. 2001; Nielsen, Schneider et al. 2001; Smothers and Henikoff 2001; Meehan, Kao et al. 2003). Le CD de HP1 reconnaît spécifiquement et se lie à l'histone H3 méthylée sur la lysine 9 (H3K9me) (Bannister, Zegerman et al. 2001; Lachner, O'Carroll et al. 2001). Cette interaction est importante pour le recrutement de HP1 sur les régions hétérochromatiques. Le CSD est requis pour l'interaction de HP1 avec de nombreuses autres protéines à CSD

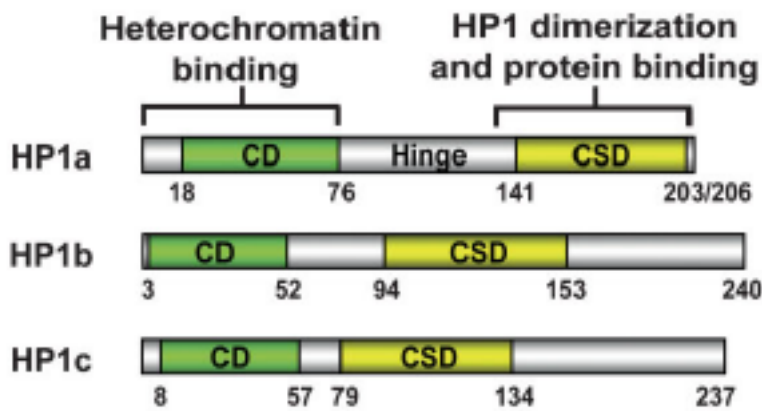


Figure 42: Différentes isoformes de HP1 chez la drosophile

Représentation schématique des trois isoformes de HP1 décrites chez la drosophile: HP1a, HP1b et HP1c. Toutes les isoformes contiennent un ChromoDomaine (CD; vert) qui permet de lier l'histone H3 méthylée sur la lysine 9 (marque de l'hétérochromatine) et un ChromoShadowDomaine (CSD; jaune) qui est une région de dimérisation de HP1, ou de liaison à divers autres facteurs protéiques. Les chiffres indiqués sont les positions des acides aminés des bornes des différents domaines. [Piacentini and Pimpinelli 2010]

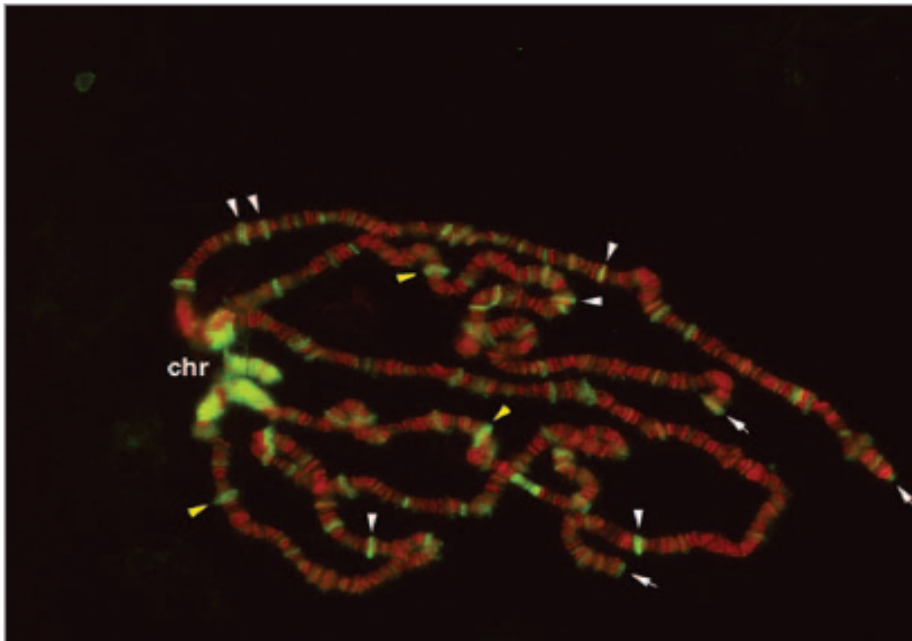


Figure 43: Distribution de HP1a

Immunomarquage sur chromosomes polytènes de glandes salivaires de drosophile avec un anticorps reconnaissant la protéine HP1a (vert). Cette protéine est située en majorité dans une région hétérochromatique regroupant les centromères des différents chromosomes, appelée chromocentre (chr). On peut aussi observer la présence de cette protéine sur les bras des chromosomes dans des régions euchromatiques (pointes de flèches jaunes et blanches) et sur les télomères (flèches blanches). [Piacentini et Pimpinelli 2010]

qui possèdent donc un motif conservé PxVxL (Brasher, Smith et al. 2000; Cowieson, Partridge et al. 2000). Le CSD de HP1 a été impliqué dans diverses interactions protéine-protéine avec des facteurs qui comprennent entre autres des modificateurs de la chromatine, des régulateurs de la transcription, des facteurs liés au cycle cellulaire, des protéines nécessaires à la réplication et des protéines de réparation de l'ADN. Il a été suggéré que le CSD de HP1 peut être responsable du ciblage continu de HP1 sur l'hétérochromatine en l'absence de méthylation de la lysine 9 de l'histone H3 (Maison, Bailly et al. 2002).

La plupart des eucaryotes présentent plusieurs isoformes de la protéine HP1. Chez la drosophile et les mammifères, trois principales protéines HP1 sont connues: HP1a, HP1b et HP1c (**Figure 42**), et HP1 α , HP1 β et HP1 γ respectivement (Kwon and Workman 2008). Toutes les isoformes possèdent la même organisation en différents domaines, mais peuvent différer dans leurs fonctions, leurs profils d'expression et leurs localisations. Fait intéressant, la distribution de HP1a sur les chromosomes polytènes de glandes salivaires de drosophile ne se limite pas aux régions hétérochromatiques mais elle est également observée sur les régions euchromatiques des bras des chromosomes (**Figure 43**) (Cowell, Aucott et al. 2002; Fanti, Berloco et al. 2003). Sachant que toutes les isoformes de HP1 contiennent un CD avec une affinité pour la méthylation H3K9 qui est majoritairement présent dans l'hétérochromatine (**Figure 44**), il est surprenant de trouver les différentes HP1 localisées dans différentes régions euchromatiques du génome.

Le rôle de HP1 dans la compaction de la chromatine a été largement étudié, mais en ce qui concerne son rôle activateur il n'existe que des preuves émergentes chez la drosophile et chez les mammifères qui suggèrent que HP1 joue un rôle dans la régulation positive des gènes dans l'euchromatine. HP1 γ et H3K9me3 ont été identifiées comme

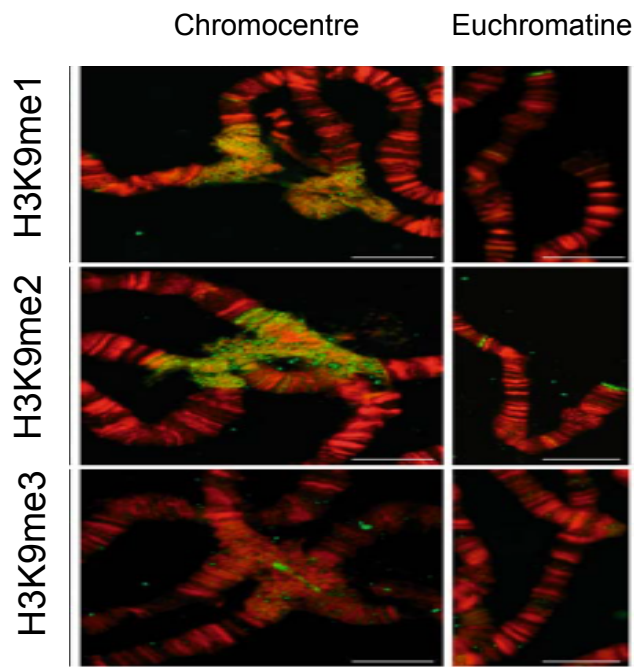
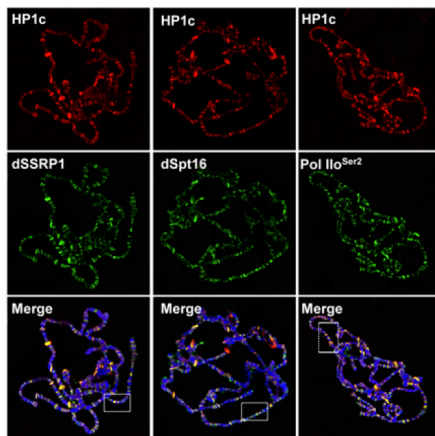


Figure 44: Distribution de la méthylation sur la lysine 9 de l'histone H3

Immunomarquage sur chromosomes polytenes de glandes salivaires de drosophile avec un anticorps reconnaissant différents niveaux de méthylation de la lysine 9 de l'histone H3 (H3K9) en vert. L'ADN est marqué en rouge. La colonne de gauche représente un agrandissement sur le chromocentre et celle de droite sur les bras des chromosomes. Il y a présence de H3K9me, H3K9me2, H3K9me3 sur le chromocentre et absence sur les bras des chromosomes, indiquant que ces modifications de la chromatine sont spécifiques de l'hétérochromatine. [Ebert *et al.*, 2006]

A)



B)

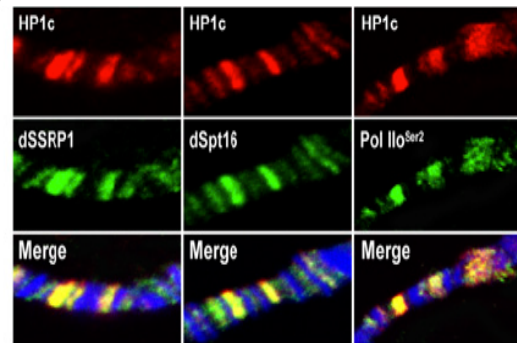


Figure 45: Présence de HP1c sur des régions transcriptionnellement actives

A) Immunolocalisation des protéines HP1c (rouge) avec les protéines du complexe activateur de la transcription FACT (dSSRP1 et dSpt16) (vert) et la forme active de l'ARN polymérase 2 (Pol Ilo^{Ser2}) (vert). HP1c colocalise sur les bras des chromosomes polytenes de glandes salivaires de drosophile avec ces différentes protéines, signe de la présence de HP1c sur des gènes transcriptionnellement actifs.

B) Agrandissement (carré blanc en A) des régions où HP1c colocalise avec dSSRP1, dSpt16 et Pol Ilo^{Ser2}. [Kwon *et al.*, 2010]

jouant un rôle inattendu dans l'activation transcriptionnelle chez les mammifères grâce à leur association avec l'ARN polymérase II (Pol II) (Vakoc, Mandat et al. 2005). Par ailleurs, une étude récente a également lié l'élongation de la transcription par l'ARN Pol II et HP1c *via* le complexe FACT (FACilitated Chromatin Transcription) (Kwon, Florens et al. 2010) qui est composé d'un hétérodimère des protéines SSRP1 (Structure specific recognition protein 1) et Spt16 (**Figure 45**) (Schlesinger and Formosa 2000; Cao, Bendall et al. 2003). Il semblerait que l'isoforme HP1c ait un rôle plus important dans l'activation de la transcription, mais la disparition de HP1a de sites euchromatiques par un traitement *in vivo* à la RNase chez la drosophile indique également son association avec des transcrits naissants (Piacentini, Fanti et al. 2003). Plus récemment, il a été démontré que HP1a s'associe à la forme active de la Pol II (phosphorylée sur la serine 2) ainsi qu'aux transcrits naissants par son chromodomaine (Piacentini, Fanti et al. 2009). Un criblage du génome entier en utilisant la technique de RIPchip (immunoprécipitation d'ARN sur puce) en culture de cellules de drosophile a permis d'identifier plus de 100 cibles actives de HP1. Bien que ces cibles aient été identifiées en cellules en culture, il semblerait que la plupart des sites liant HP1 le long des chromosomes polytènes de larves soient identiques.

Un rôle positif de HP1 dans l'expression génique a également été suggéré par d'autres analyses montrant que HP1 colocalise et coprécipite avec différentes classes de hnRNPs (heterogenous nuclear Ribo-Nucleo-Proteins). Il s'agit notamment du hnRNP de type A/B HRB87F (Haynes, Johnson et al. 1991). De plus, des expériences de DamID (DNA adenine methyltransferase identification) avec HP1 ont montré la présence de la protéine sur les bras des chromosomes et sur toute la longueur des gènes excepté la partie 5' UTR. Ces mêmes gènes étaient cibles de modifications d'histones de type

Protein	Organism	HP1 variant	Domain
Chromatin modifying proteins			
Clr4	<i>S. Pombe</i>	Swi6	Nd
Su(var)3-9	<i>Drosophila</i>	HP1	CSD
Suv39h1	Mouse, Human	mHP1β, hHP1 β	Nd
Su(var)3-7	<i>Drosophila</i>	HP1	CSD
G9a	Human	hHP1α, hHP1γ	Nd
SETDB1	Human	hHP1	Nd
dSETDB1	<i>Drosophila</i>	HP1a	Nd
Epel	<i>S. Pombe</i>	Swi6	Nd
Dnmt3a/Dnmt3b	Mouse	mHP1α	Nd
ATRX/HP1-BP38	Mouse	mHP1α, mHP1β	CSD
Pim-1	Human	hHP1γ	CSD
CKII	<i>Drosophila</i>	HP1	Nd
dAF10	<i>Drosophila</i>	HP1	CSD
MITR, HDAC4/5	Mouse	mHP1α	Hinge
Transcription-associated proteins			
RNA polymerase II	Human	hHP1γ	Nd
RNA polymerase II	<i>Drosophila</i>	HP1a, HP1c	CD
FACT	<i>Drosophila</i>	HP1a, HP1b, HP1c	CSD
WOC	<i>Drosophila</i>	HP1c	Nd
ROW	<i>Drosophila</i>	HP1c	Nd
Polycomb	Human	hHP1α, hHP1γ	CSD
KAP1/TIFβ	Mouse	mHP1α, mHP1β, mHP1γ	CSD
KAP1/TIFβ	Human	hHP1α, hHP1γ	CSD
TRF1/PIN2	Mouse	mHP1β	Nd
TAFII130	Human	hHP1α, hHP1γ	CSD
TIF1α	Mouse	mHP1α, mHP1β, mHP1γ	CSD
mSNF2β	Mouse	mHP1α	CSD
Rb	Human	hHP1α, hHP1γ	Nd
E2F-6	Human	hHP1γ	Nd
BRG1	Human	hHP1α	CSD
POF	<i>Drosophila</i>	HP1	Nd
DNA replication and repair			
CAF-1 p150	Mouse	mHP1α, mHP1β	CSD
CAF-1 p150	Human	hHP1α	CSD
CAF-1 p180	<i>Drosophila</i>	HP1a	Nd
Ku70	Human	hHP1α hHP1γ	CSD
BRCA-1	Human	hHP1α	Nd
ORC1-6	<i>Drosophila</i>	HP1	CD,CSD
HOAP	<i>Drosophila</i>	HP1	Nd
SuUR	<i>Drosophila</i>	HP1	CSD + Hinge
RNA-associated proteins			
DDP1, Hrb87F, PEP	<i>Drosophila</i>	HP1a	Nd
PIWI	<i>Drosophila</i>	HP1a	CSD

Figure 46: Liste des protéines pouvant interagir avec HP1

Les protéines HP1 peuvent interagir avec de nombreuses protéines majoritairement *via* leurs ChromoShadowDomaine (CSD). Sur ce tableau sont présentés de gauche à droite les protéines qui interagissent avec HP1, l'organisme où l'interaction a été trouvée, l'isoforme de HP1 qui est impliquée et le domaine de HP1 nécessaire pour l'interaction. [Piacentini et Pimpinelli 2010]

H3K4me2 et du variant d'histone de type H3.3, qui est lié à une activation transcriptionnelle (de Wit, Greil et al. 2007).

Comment peut-on expliquer les différentes fonctions de la protéine HP1 ? On sait qu'elle est capable de lier H3K9me et d'interagir directement ou indirectement avec plusieurs protéines ayant de grandes variétés de fonctions. Ces interacteurs (**Figure 46**) sont impliqués dans des processus cellulaires tels que la régulation transcriptionnelle, les modifications de la chromatine, la réplication, la réparation de l'ADN, et la mise en place de l'architecture nucléaire (Kwon and Workman 2008). Fait intéressant, ces interactions peuvent être spécifiques d'une isoforme de HP1 ou être possibles avec les trois isoformes.

Chaque isoforme de HP1 peut subir des modifications post-traductionnelles, telles que des acétylations, des phosphorylations, des méthylations..., et plusieurs sites peuvent avoir différents types de modifications (Lomberk, Bensi et al. 2006; LeRoy, Weston et al. 2009). Ces modifications peuvent se trouver à la fois dans le CD et le CSD, ce qui suggère qu'elles peuvent jouer un rôle dans la modulation des interactions ou des fonctions de HP1. Par exemple, l'hyperphosphorylation de HP1a est associée à la formation de l'hétérochromatine et affecte sa localisation chromosomique. La mutation au niveau de ces sites de phosphorylation empêche la mise sous silence. Néanmoins une forme hypophosphorylée de HP1 peut s'associer avec des sous-unités du complexe ORC et est plus étroitement associée avec des chromosomes en interphase (Shareef, King et al. 2001). Ces constatations laissent penser que la phosphorylation peut contrôler HP1 en modifiant ses interactions.

En 2001 il a été mis en évidence une quatrième protéine de type HP1 chez la drosophile : la protéine Rhino (HP1d) (Volpe, Horowitz et al. 2001; Vermaak, Henikoff

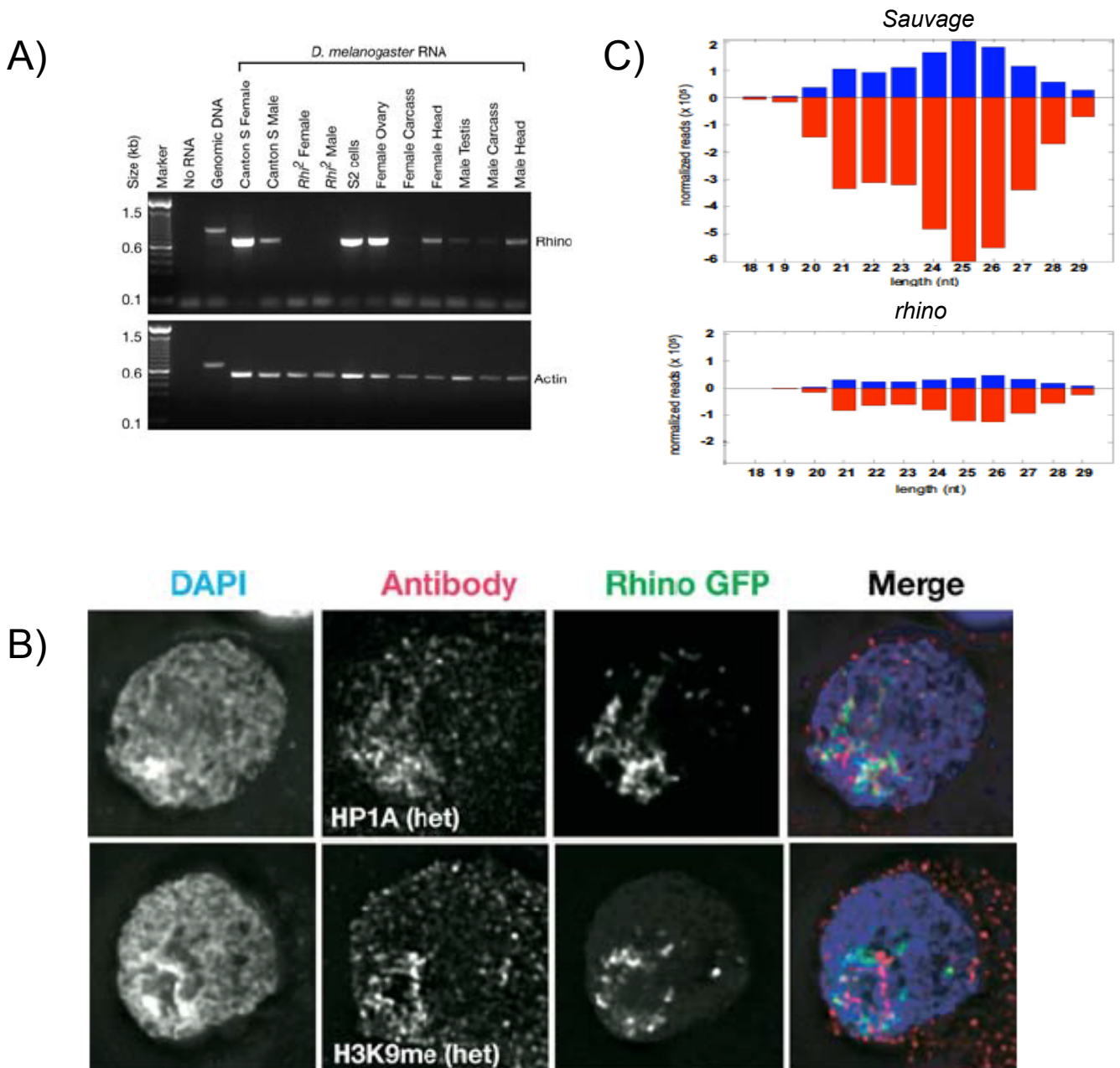


Figure 47: Expression et rôle de Rhino chez la drosophile

A) rt-PCR sur l'ARNm du gène *rhino* sur différent tissus de drosophile. L'expression de ce gène semble restreinte à certains tissus comme les ovaires (l'actine sert de contrôle de charge). **B)** Colocalisation entre HP1a ou H3K9me (rouge) avec la protéine Rhino (vert). Cette dernière est localisée dans les régions denses en ADN (bleu foncé) mais ne semble pas localisée sur les mêmes régions que HP1a et H3K9me. [Vermaak *et al.*, 2005]. **C)** Séquençage des piRNAs de drosophile sauvage (panneau du haut) et mutante pour *rhino* (panneau du bas). L'absence de *rhino* entraîne une baisse de la quantité de piRNAs sens (bleu) et antisens (rouge). [Klattenhoff *et al.*, 2009]

et al. 2005). Cette protéine majoritairement localisée dans l'hétérochromatine permettrait directement ou indirectement l'activation de la transcription des clusters de piRNAs (Klattenhoff, Xi et al. 2009) (**Figure 47**) et est impliquée dans la mise sous silence des éléments transposables. Cette protéine est exprimée spécifiquement dans les ovaires de drosophile. Une cinquième isoforme HP1e a aussi été trouvée et est spécifique des testicules chez la drosophile (Vermaak, Henikoff et al. 2005).

c).2 Les protéines des groupes Polycomb et Trithorax

Les protéines Polycomb et Trithorax ont été identifiées pour la première fois chez la drosophile en 1949 et la découverte de leur implication dans le développement a valu le prix Nobel à Lewis en 1995. Ces protéines semblent être conservées chez de nombreux eucaryotes (Li, Tatsuke et al. 2011). Les groupes de protéines Polycomb (PcG) et Trithorax (TrxG) contribuent respectivement au maintien de la répression ou de l'activation transcriptionnelle de gènes régulés au cours du développement pendant la division cellulaire. Ces protéines sont nécessaires pour initier des modifications épigénétiques et leur expression est essentielle pour le maintien de ces modifications au cours du développement. Récemment, il a été montré que les protéines PcG étaient impliquées dans de nombreux autres processus cellulaires en particulier chez les mammifères comme l'inactivation du chromosome X, l'empreinte parentale, la plasticité des cellules souches embryonnaires, le développement tumoral, la régulation des rétrovirus et les voies d'ARN interférence (Pasini, Bracken et al. 2008; Zhao, Sun et al. 2008; Martinez, Schuettengruber et al. 2009; Golding, Zhang et al. 2010; Mills 2010; Kim, Kim et al. 2011; O'Loghlen, Munoz-Cabello et al. 2012). Je m'intéresserais plus particulièrement aux PcG car ce sont les complexes répresseurs.

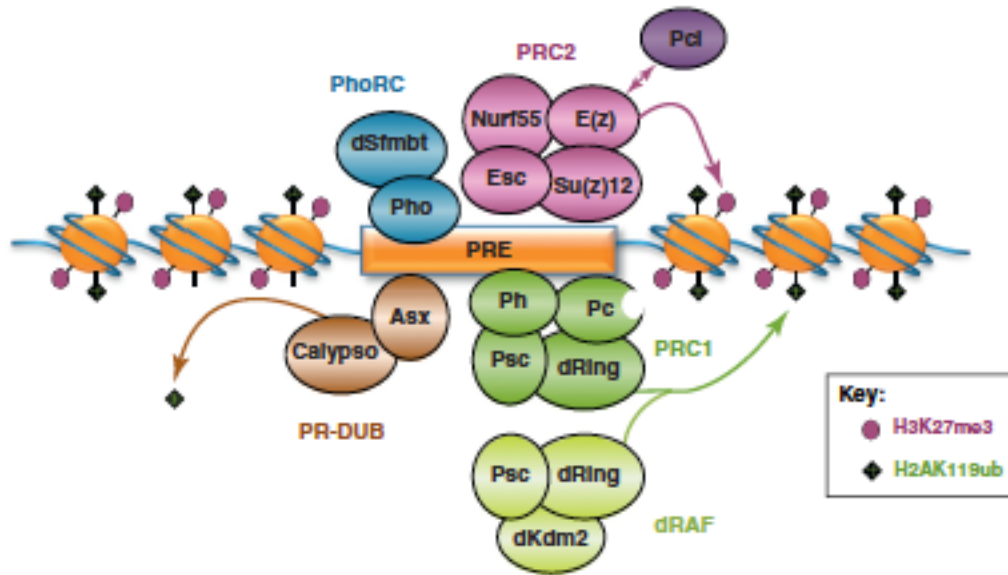


Figure 48: Différents complexes impliquant les protéines du groupe Polycomb PcG

Chez la drosophile, cinq complexes ont été décrits (PRC1, PRC2, dRAF, PhoRC, PR-DUB) mais deux complexes PcG principaux PRC2 et PRC1 ont été majoritairement étudiés. Le complexe répressif polycomb 2 (PRC2; rose) regroupe plusieurs protéines telles que enhancer of Zeste (E(Z)), extra sex combs (ESC), SU(Z)12 mais aussi une protéine liant les histones NURF-55. E(Z) est une HMT qui méthyle les lysines 27 et 9 de l'histone H3 (H3K27me³; rond rose). Le complexe répressif polycomb 1 (PRC1; vert foncé,) grâce à la fixation du CD de la protéine Pc sur H3K27me³, vient se lier à la chromatine. Après sa stabilisation, le PRC1 qui comprend les protéines telles que Polycomb (Pc), Polyhomeotic (Ph), Extra-Sex-Comb (ESC) et dRING, une ubiquitine ligase qui ajouterait un groupement ubiquitine sur la lysine 119 de l'histone H2A (H2AK119ub; losange noir), verrouillerait ses cibles en bloquant l'accès de la chromatine. Les autres complexes ont été moins étudiés car mis en évidence plus récemment, mais leur rôle semble être en partie défini. Le complexe répresseur Pleihomeotic (Pho) (PhoRC; bleu) contient deux protéines Pho et dSfmbt. Son rôle pourrait être de recruter le PRC2 sur l'ADN. Le complexe dRAF (d-Ring-Associated-Factors; vert clair), composé des protéines dRing, Psc, et dKdm2, jouerait un rôle dans la déposition de H2AK119ub. Le complexe PR-DUB (Polycomb repressive deubiquitinase complex; marron), composé des protéines Asx et Calypso, jouerait un rôle dans l'effacement de H2AK119ub. Tous ces complexes viennent se fixer sur des régions d'ADN appelées PRE (Polycomb Response Element). [Bantignies et Cavalli 2011]

L'un des premiers gènes à avoir été cloné est *polycomb* (*pc*). La protéine Pc possède un chromodomaine identique à 58% au domaine de la protéine HP1 de drosophile (Paro and Hogness 1991). Ces protéines forment et agissent au sein de complexes multiprotéiques. A l'heure actuelle, chez la drosophile, cinq complexes ont été décrits (PRC1, PRC2, dRAF, PhoRC, PR-DUB) (**Figure 48**) mais deux complexes PcG principaux PRC2 et PRC1 ont été majoritairement étudiés. Le complexe répressif polycomb 2 (PRC2) regroupe plusieurs protéines telles que Enhancer of Zeste (E(Z)), Extra Sex Combs (ESC), SU(Z)12 mais aussi une protéine liant les histones, NURF-55. E(Z) est une HMT qui méthyle la lysines 27 de l'histone H3 (H3K27me). Le code histone ainsi mis en place est alors reconnu par le complexe répressif polycomb 1 PRC1 grâce à la fixation du CD de la protéine Pc sur H3K27me³ (Cao, Wang et al. 2002; Czermin, Melfi et al. 2002; Kuzmichev, Nishioka et al. 2002; Muller, Hart et al. 2002). Après sa stabilisation, le PRC1, (**Figure 48**) qui comprend les protéines Polycomb (PC), Polyhomeotic (PH), Extra-Sex-Comb (ESC) et dRING, une ubiquitine ligase qui ajouterait un groupement ubiquitine sur la lysine 119 de l'histone H2A (H2AK119ub), verrouillerait ses cibles en bloquant l'accès de la chromatine aux complexes de remodelage tel que SWI/SNF (Francis, Saurin et al. 2001; Francis, Kingston et al. 2004).

Les autres complexes ont été moins étudiés, car mis en évidence plus récemment, mais leur rôle semble être en partie défini. Le complexe répresseur Pleihomeotic (Pho) (PhoRC) contient deux protéines Pho et dSfmbt et son rôle pourrait être de recruter le PRC2 sur l'ADN. Le complexe dRAF (d-Ring-Associated-Factors), composé des protéines dRing, Psc, et dKdm2, jouerait un rôle dans la déposition de H2AK119ub. Le complexe PR-DUB (Polycomb Repressive DeUBiquitinase complex), composé des protéines Asx et Calypso, jouerait un rôle dans l'effacement de H2AK119ub (Scheuermann, de Ayala Alonso et al. 2010; Beisel and Paro 2011). La plupart de ces complexes mis en évidence

drosophile

Mammifères

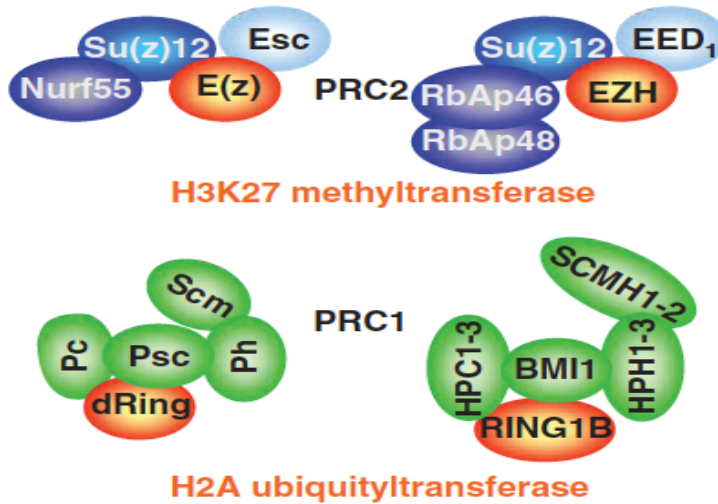


Figure 49: Conservation des principaux PcG chez les mammifères

Chez les mammifères, il a été montré que les complexes PRC1 et PRC2 sont conservés. Les activités enzymatiques le sont aussi: E(z) chez la drosophile méthyle H3K27 et son orthologue EZH a la même activité. De même pour dRing et RING1B qui peuvent ubiquityniler H2AK119. [Schuettengruber et Cavalli 2009]

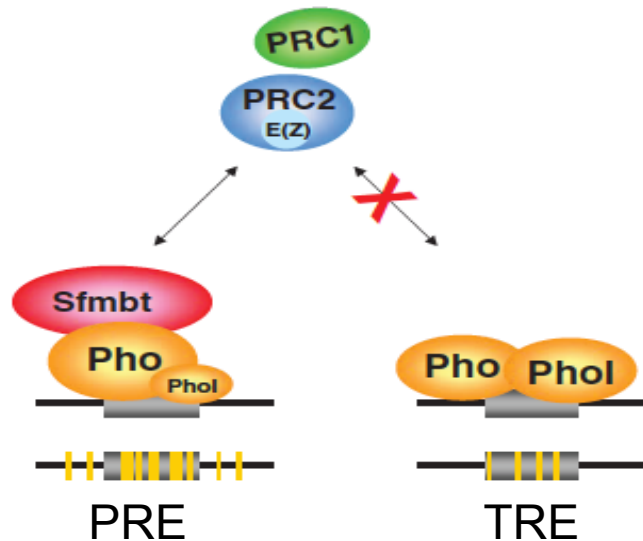


Figure 50: Reconnaissance des Polycomb et Trithorax Response Element (PRE et TRE)

Malgré la controverse, il ne semble pas exister de séquences ADN consensus permettant de prédire des PRE et TRE. Il semblerait que la protéine Pho serait présente sur une majorité de PRE et serait donc cruciale pour le recrutement des protéines PcG (PRC1;vert et PRC2;bleu). Inversement la présence de Phol serait un indicateur du maintien de l'activité transcriptionnelle et donc peu présente sur les PRE. La présence d'une activité transcriptionnelle peut être due aux protéines des complexes TrxG et donc Phol pourrait être un indicateur des TRE. [Schuettengruber et Cavalli 2009]

chez la drosophile semblent être conservés chez les mammifères (**Figure 49**) chez lesquels il semblerait que le recrutement des PcG soit étroitement lié à des ARNs non codants.

Suite à des analyses menées afin d'identifier tous les sites cibles des protéines Polycomb, il a tout d'abord été suggéré que ces protéines intervenaient dans la régulation de plus de 200 gènes (Schwartz, Kahn et al. 2006; Tolhuis, de Wit et al. 2006). Il semblerait en fait que plus de 1000 gènes soient régulés par les PcG chez la drosophile soit 7% des gènes (Bantignies and Cavalli 2011; Enderle, Beisel et al. 2011) et approximativement autant qu'en cellules souches embryonnaires (ESCs) (Brookes, de Santiago et al. 2012) (Bantignies and Cavalli 2011; Enderle, Beisel et al. 2011). La plupart des cibles identifiées sont des gènes codant pour des facteurs de transcription ou intervenant dans différentes voies de signalisation. Ces résultats étayent l'idée que ces protéines Polycomb sont hautement impliquées dans la coordination du développement en formant de larges complexes répressifs.

Ces complexes protéiques sont connus pour se fixer sur des séquences régulatrices appelées PRE/TRE pour Polycomb and Trithorax Response Element (Chan, Rastelli et al. 1994). La taille de ces séquences peut varier d'une centaine à plusieurs milliers de paires de bases. Leur position par rapport au gène régulé est aussi très variable : proches ou chevauchant le promoteur pour certains PRE/TRE, ou bien situées jusqu'à des dizaines de kilobases de celui-ci, en amont ou en aval. Une des caractéristiques de ces séquences est qu'elles ne possèdent pas ou peu de ressemblances entre elles.

De manière intéressante, la localisation de protéines se liant à certains motifs de l'ADN semble être un bon indicateur de la localisation de PRE/TRE (Schuettengruber and Cavalli 2009; Schuettengruber, Ganapathi et al. 2009). Il s'agit des protéines GAGA factor (GAF), Pipsqueak (PSQ), Pleiohomeotic (PHO), Pleiohomeotic-like (PHOL), Dorsal Switch Protein 1 (DSP1), Zeste, Grainyhead (GH) et SP1/KLF. Il semblerait que la

Complexes TrxG
modificateurs
d'histones

Complexes TrxG
de remodelage
de la chromatine
ATP dépendant

Complex	Comment	Mammals	<i>D. melanogaster</i>	<i>C. elegans</i>	Plants
Histone-modifying complexes					
COMPASS	Global gene activation; mediates bulk trimethylation of K4 on Histone H3 (H3K4me3)	SET1A,B CXXC1 WDR82 ASH2L, DPY30, HCF1, RBBP5, WDR5	SET1 CXXC1 WDR82 ASH2, DPY30, HCF1, RBBP5, WDR5	SET-2	
COMPASS-like	HOX-family gene regulation, tumour suppressor activity and HAT activity for H4K16	MLL1-2 MOF Menin ASH2L, DPY30, HCF1, RBBP5, WDR5	TRX Menin ASH2, DPY30, HCF1, RBBP5, WDR5		ATX1-5 ASH2R, RBL, WDR5A
	Activates nuclear receptors in mammals and ectosome-induced promoters in <i>D. melanogaster</i> ; demethylase activity for H3K27me3 plays a part in dosage compensation and the attenuation of RAS signalling in <i>C. elegans</i>	MLL3,4 UTX NCOA6 PA1 PTIP ASH2L, DPY30, RBBP5, WDR5	TRR UTX NCOA6 PA1 PTIP ASH2, DPY30, RBBP5, WDR5	SET-18 UTX-1 ASH-2, DPY-30, WDR-5	
TAC1	Counteracts PcG silencing, but also has a more global role in gene activation; has HAT activity for H3K27 in <i>D. melanogaster</i>		TRX CBP SBF1		
ASH1	Counteracts PcG silencing and has HMT activity for H3K36 and HAT activity for H3K27 in <i>D. melanogaster</i>	ASH1L	ASH1 CBP		
ATP-dependent chromatin-remodelling complexes					
SWI/SNF	Binds acetylated histones via its bromodomain, regulates cell cycle, cell signalling, proliferation and chromosome segregation	BRM, BRG1 Actin, BAF45A, BAF45B, BAF45C, BAF45D, BAF47, BAF53A, BAF53B, BAF60A, BAF60B, BAF60C, BAF155, BAF170, BAF180, BAF200, BAF250A, BAF250B, BAF57, BRD7, BRD9	BRD Actin, BAF60, BAF180, BAP55, BAP111, Delao, BAP170, OSA, MOR, SAYP, SNR1		
ISWI	Binds unmodified histones via its SANT domain and reads H3K4me3 via its PHD finger; regulates transcription, cell fate determination and differentiation	BPTF RBAP46, RBAP48, SNF2L	NURF301 ISWL, NURF38, RBAP46, RBAP48	NURF ISW-1	CHR11.177
CHD1, CHD2	Bind H3K4me3 via their chromodomains; have roles in development and ES cell maintenance; H3.3 deposition in <i>D. melanogaster</i>	CHD1.2	CHD1.2		
CHD3, CHD4	HDAC activity; regulate transcription, replication and DNA repair; regulate cell fate determination in <i>C. elegans</i> and counteract PcG silencing in plants	CHD3,4 HDAC1,2 RBAP46, RBAP48, MBD2, MBD3, MTA1, MTA2, MTA3, p66	MI2 RPD3 MBD2, MBD3, MTA, p55, p66, p68	HDA-1 LET-418 MEP-1	PKL
CHD6, CHD7, CHD8	Bind dimethylated H3K4 (H3K4me2) and H3K4me3 via their chromodomains; counteract PcG silencing and regulate transcriptional elongation in <i>D. melanogaster</i>	CHD8 ASH2L, RBBP5, WDR5	CHD6 CHD7 CHD8 KIS-L		

Figure 51: Les complexes du groupe Trithorax (TrxG)

Les protéines TrxG peuvent être divisées en deux classes majoritaires en fonction de leur activité moléculaire. La première classe (bleu) comprend des facteurs à domaine SET qui peuvent modifier post-traductionnellement les queues des histones. La seconde classe (orange) contient des facteurs de remodelage de la chromatine ATP-dépendant. La plupart de ces protéines TrxG exercent leur fonction dans le cadre de grands complexes multi protéiques qui coopèrent pour réguler l'expression des gènes. Les protéines TrxG sont évolutivement conservées (représenté sur ce schéma les complexes conservés chez les mammifères, les plantes, la drosophile et *C.elegans*. [Schuettengruber *et al.*, 2011]

protéine PHO soit présente sur une majorité de PRE et serait donc cruciale pour le recrutement des protéines PcG. Inversement, la présence de PHOL ou Zeste serait un indicateur du maintien de l'activité transcriptionnelle et ces protéines sont donc peu présentes sur les PRE (**Figure 50**). La présence d'une activité transcriptionnelle peut être due aux protéines des complexes TrxG.

Les protéines TrxG, comme les protéines PcG qui régulent de nombreux gènes en plus des gènes homéotiques, sont aussi impliquées dans de nombreux processus cellulaires (Schuettengruber, Martinez et al. 2011) comme la méthylation des histones et le remodelage de la chromatine, la tumorigénèse, le renouvellement des cellules souches embryonnaires (Ang, Tsai et al. 2011), la différenciation cellulaire et la prolifération (Bagchi, Papazoglu et al. 2007). Les protéines TrxG ont également été impliquées dans l'inactivation du chromosome X (Pullirsch, Hartel et al. 2010), l'apoptose (Tyagi and Herr 2009), la régulation du cycle cellulaire (Liu, Cheng et al. 2007), ainsi que dans la réponse au stress, au cours de laquelle ils activent la sénescence et les points de contrôle des dommages à l'ADN (Greer, Maures et al. 2010; Siebold, Banerjee et al. 2010).

Cependant, on ne sait pas si ces complexes régulent ces processus cellulaires essentiellement grâce à leur fonction antagoniste aux PcG, ou par une activation de l'expression des gènes. Les protéines TrxG sont évolutivement conservées et peuvent être divisées en trois classes en fonction de leur activité moléculaire. Une première classe comprend des facteurs à domaine SET qui peuvent méthyler les queues des histones. La seconde classe contient des facteurs de remodelage de la chromatine ATP-dépendant (**Figure 51**). La troisième classe comprend des protéines TrxG qui peuvent se lier directement à des séquences d'ADN spécifiques et est composée de modificateurs des histones et de protéines de remodelage chromatinien, ainsi que des protéines qui ne sont pas classées dans les deux premières catégories. La plupart de ces protéines TrxG exercent

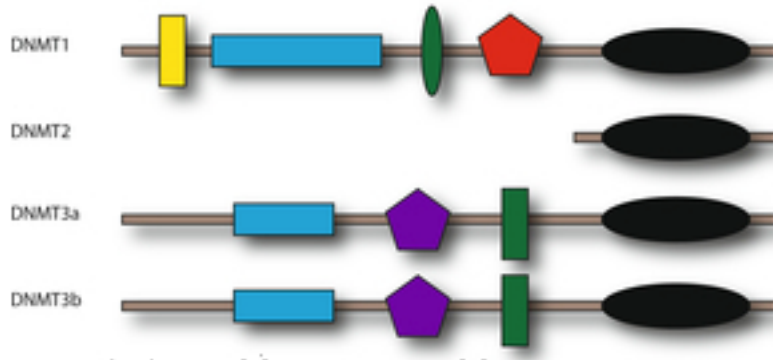
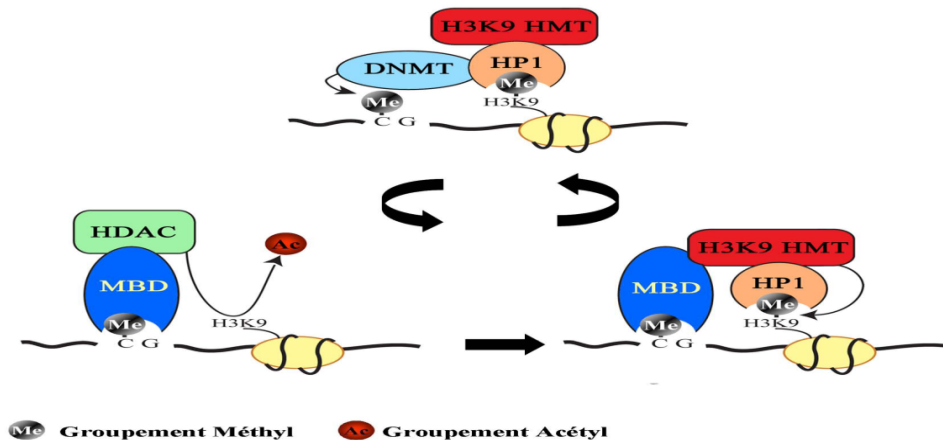


Figure 52: Structure des méthyle-transférases de l'ADN (Dnmt)

Représentation graphique des domaines protéiques des quatre DNMTs mammifères. Les domaines sont représentés par les couleurs suivantes: rectangle jaune domaine de liaison DMAP1, rectangle bleu région avec des domaines de faible complexité, ellipse verte domaine à doigt de zinc de type CXXC, pentagone rouge domaine d'homologie BAH, ellipse noire domaine catalytique DNMT, pentagone pourpre domaine PWWP, rectangle vert domaine à doigt de zinc de type PHD. [http://www.springerimages.com/Images/Chemistry/1-10.1007_s00214-009-0681-2-0]



Me Groupement Méthyl Ac Groupement Acétyl

Figure 53: Implication de la méthylation de l'ADN et des modifications de la chromatine dans le verrouillage de l'expression génique

Les DNMT sont recrutées par HP1 qui est un répresseur qui lie les histones H3 méthylées sur la lysine 9. Les DNMT méthylent alors l'ADN. L'ADN méthylé permet la liaison de protéines à domaines MBD qui peuvent recruter des histones déacétylases (HDAC) et des méthylases d'histone (HMT), qui permettent de verrouiller l'expression génique grâce à la formation d'une structure chromatiniennne compacte. [Fuks et al., 2005]

leur fonction dans le cadre de grands complexes multi protéiques qui coopèrent pour réguler l'expression des gènes (Schuettengruber, Martinez et al. 2011).

d) La méthylation de l'ADN

En plus des modifications covalentes ou non des histones, l'ADN est lui-même sujet à une modification covalente : la méthylation. Cette modification de l'ADN est associée à un verrouillage réversible de l'expression des gènes et cible presque exclusivement les résidus cytosines (Fuks 2003; Fuks 2005). La méthylation de l'ADN est impliquée dans de nombreux processus biologiques tels que l'inactivation du chromosome X ou l'empreinte parentale.

Chez les mammifères, quatre enzymes méthylant activement les cytosines ont été identifiées. Il s'agit des méthyl-transférases de l'ADN Dnmt1, 3a, 3b et 3l (**Figure 52**). La méthylation des cytosines réprime la transcription génique notamment *via* le recrutement de protéines à domaine MBD (Methyl Binding Domain) qui reconnaissent de façon sélective les cytosines méthylées (**Figure 53**). Ces protéines MBD sont elles-mêmes capables de recruter une activité HDAC assurant ainsi la désacétylation des histones, et, ainsi, la condensation de la chromatine (Rountree, Bachman et al. 2001). Chez la drosophile, seule une Dnmt '*like*' est connue (Dnmt2) mais son mécanisme d'action reste à élucider. Divers rôle lui ont été attribués comme la méthylation des ARNs de transferts (tRNAs) (Schaefer, Pollex et al. 2010) ou encore la méthylation des éléments transposables et le maintien de l'intégrité des télomères (Phalke, Nickel et al. 2009).

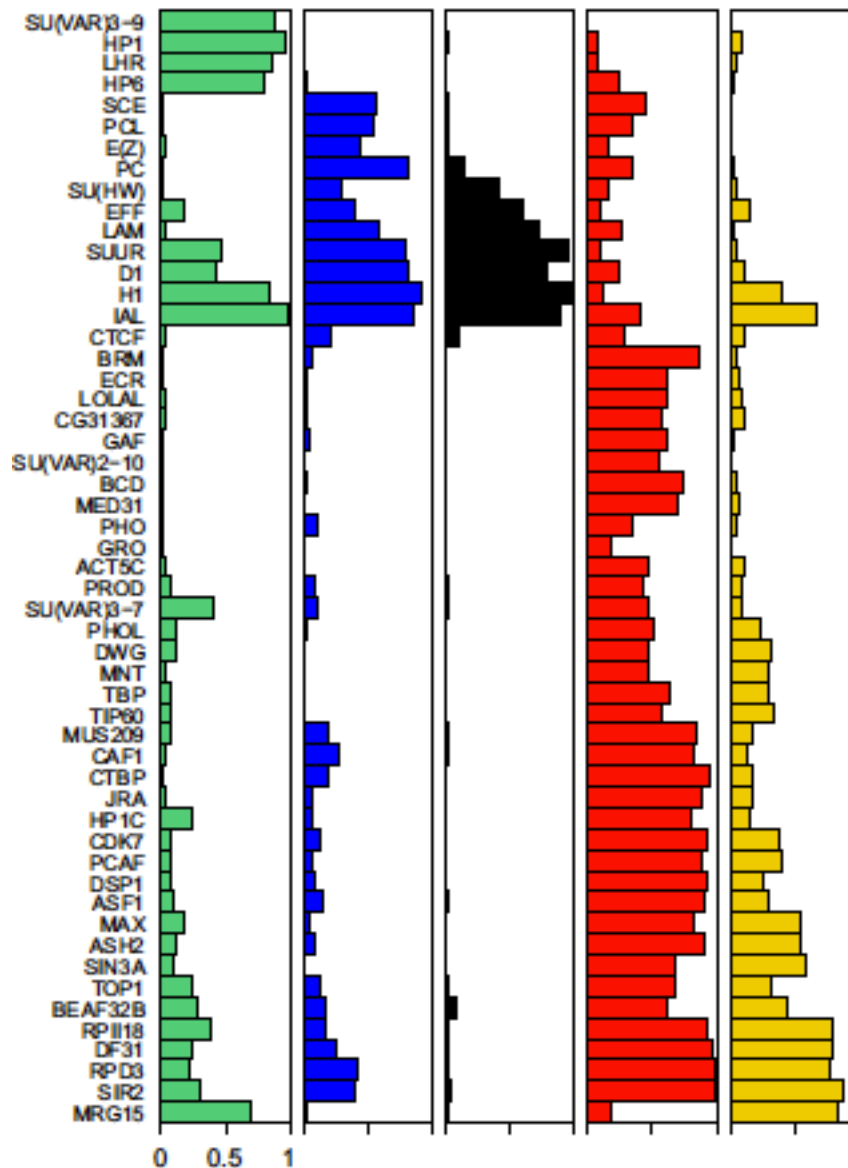


Figure 54: Différents types de chromatine de drosophile

Il a été mis en évidence chez la drosophile cinq types de chromatines distinctes grâce à une cartographie fine par DamID (DNA adenine methyltransferase identification) de la localisation de 53 protéines chromatinienne. Ces cinq types de chromatine ont été classifiés selon un code couleur: les chromatines de type fermé (bleu, noire et verte) et celle de type ouverte (jaune et rouge). La hauteur des barres indique l'enrichissement en diverse protéines. [Filion *et al.*, 2010]

e) Organisation chromatinienne et nucléaire

e).1 Organisation chromatinienne

La structure chromatinienne est constituée de plusieurs niveaux de compaction, et les chromosomes subissent plusieurs niveaux d'organisation. Le premier niveau d'organisation a été mis en évidence par Heitz dès 1928 entre les régions euchromatiques et hétérochromatiques de l'ADN.

De manière très intéressante, il n'existe pas de colocalisation entre les protéines de type HP1 et les protéines PcG (de Wit, Greil et al. 2007) indiquant que les zones de chromatine compacte (hétérochromatine) peuvent être séparées en plusieurs sous classes d'hétérochromatine. Récemment, il a en effet été mis en évidence chez la drosophile cinq types de chromatine distincte (Filion, van Bemmelen et al. 2010) grâce à une cartographie fine par DNA adenine methyltransferase identification (DamID) de la localisation de 53 protéines chromatinienne. Ces cinq types de chromatine ont été classifiés selon un code couleur : les chromatines de type fermé (bleu, noire et verte) et celle de type ouverte (jaune et rouge) (**Figure 54**).

La chromatine verte correspond aux régions hétérochromatiques qui sont principalement recouvertes par les protéines HP1 et Su(var)3-9 et se trouve localisée majoritairement sur le chromosome 4 et les régions centromériques. Cette chromatine est largement enrichie en H3K9me2 (**Figure 55 A**).

La chromatine bleue correspond aux régions couvertes par les protéines PcG (Pc, E(Z), PCL...) et recouvre les gènes connus comme étant régulés par ces protéines comme les gènes homéotiques. Cette chromatine est largement enrichie en H3K27me3 (**Figure 55 B**)

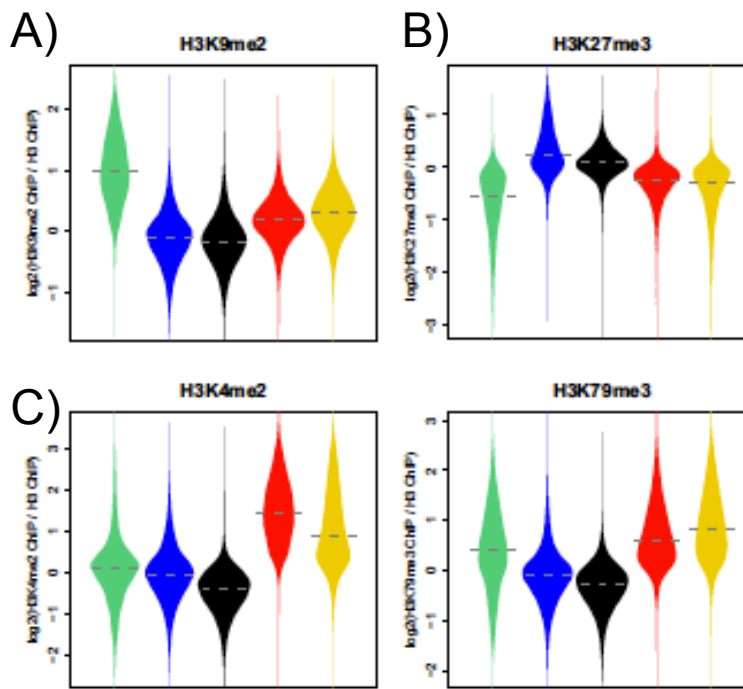


Figure 55: Enrichissement en diverses modifications d'histones des différents types de chromatine

A) La chromatine verte est enrichie en H3K9me2, **B)** La chromatine bleue est enrichie en H3K27me3. Ces deux marques sont des marques de fermeture de la chromatine ce qui correspond aux protéines retrouvées dans ces deux types de chromatine. **C)** Les chromatines jaune et rouge, qui sont de type euchromatique, sont enrichies en modifications d'histone de type ouverte comme H3K4me2 et H3K79me3. En revanche ces marques sont faiblement présentes sur les chromatines de type hétérochromatique (verte, bleue et noire). [Filion *et al.*, 2010]

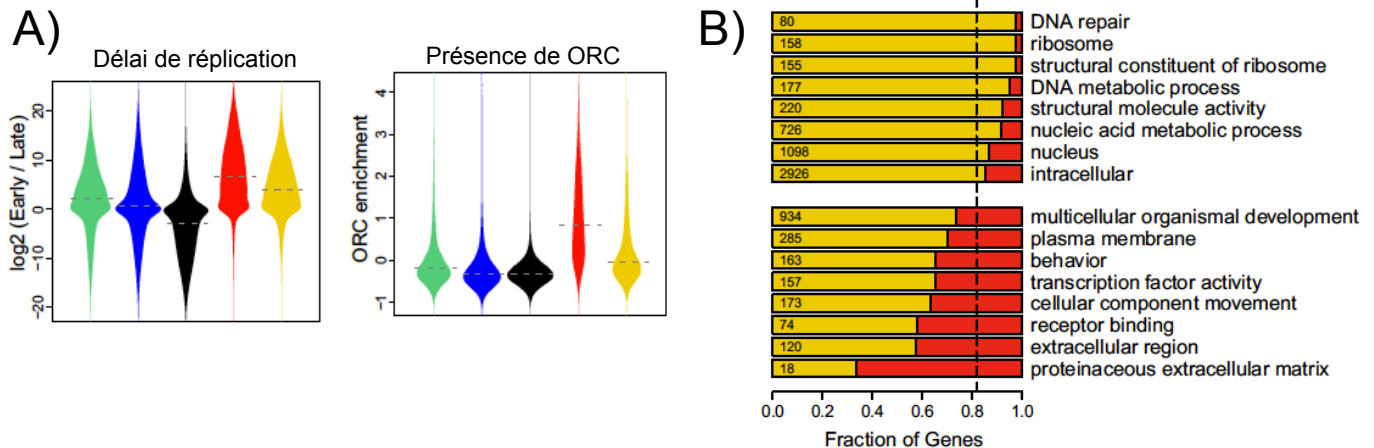


Figure 56: Les chromatines rouges et jaunes

Ces deux type de chromatine correspondent à de la chromatine transcriptionnellement active (euchromatine). **A)** Elles présentent une réplication plus précoce signe des régions euchromatiques, mais la jaune est un peu plus tardive que la rouge ce qui est confirmé par la forte présence de la protéine ORC (Origin Recognition Complex). **B)** Ces deux types d'euchromatine ne contiennent pas les mêmes type de gènes. La chromatine jaune est préférentiellement retrouvée sur les gènes exprimés ubiquitairement au cours du développement alors que les gènes avec un profil d'expression plus restreint semblent être couverts par de la chromatine rouge. Ces derniers gènes nécessitent de nombreux facteurs régulateurs spécifiques ce qui peut expliquer la présence de plus de complexes sur la chromatine rouge. [Filion *et al.*, 2010]

Inversement à ces deux premiers types d'hétérochromatine, il a été mis en évidence deux types d'euchromatine. Ces deux types, définis par les couleurs rouge et jaune, sont riches en gènes transcrits et en protéines activatrices de la transcription comme par exemple ASH2 qui appartient à la famille des TrxG et qui est une sous unité du complexe HMT H3K4 (Nagy, Griesenbeck et al. 2002). On retrouve des marques de chromatine ouverte telle que H3K79me3 et H3K4me2 (**Figure 55 C**) et peu de marques de fermeture (**Figure 55 A, B**). Mais ces deux type de chromatine, bien qu'apparemment euchromatiques et actifs en transcription, ne sont pas complètement marqués par les mêmes complexes ce qui permet de les différencier. La chromatine rouge est marquée par plus de complexes activateurs que la chromatine jaune et semble aussi être répliquée plus précocement ce qui coïncide avec la présence majoritaire de la protéine ORC (Origin Recognition Complex) (MacAlpine, Gordan et al. 2010) comme si c'était de la « super euchromatine » (**Figure 56 A**). Ces deux types de chromatine semblent être présents sur des familles de gènes distincts. La chromatine jaune est préférentiellement retrouvée sur les gènes exprimés ubiquitairement au cours du développement alors que les gènes avec un profil d'expression plus restreint semblent être couverts par de la chromatine rouge (**Figure 56 B**). Ces derniers gènes nécessitent de nombreux facteurs régulateurs spécifiques ce qui peut expliquer la présence de plus de complexes sur la chromatine rouge.

La chromatine noire comprend 48% des régions analysées et est donc le type le plus abondant dans le génome de la drosophile. Elle est enrichie en diverses protéines telles que l'histone H1, SUUR (Suppressor of Under Replication), D1 (connue pour se lier aux régions satellites), IAL (IplI-aurora-like kinase) et la Lamine (**Figure 57 A**). La particularité de cette chromatine est qu'elle n'en a aucune si ce n'est qu'elle semble être pauvre en gènes transcrits et avoir une capacité de mise sous silence plus forte que les

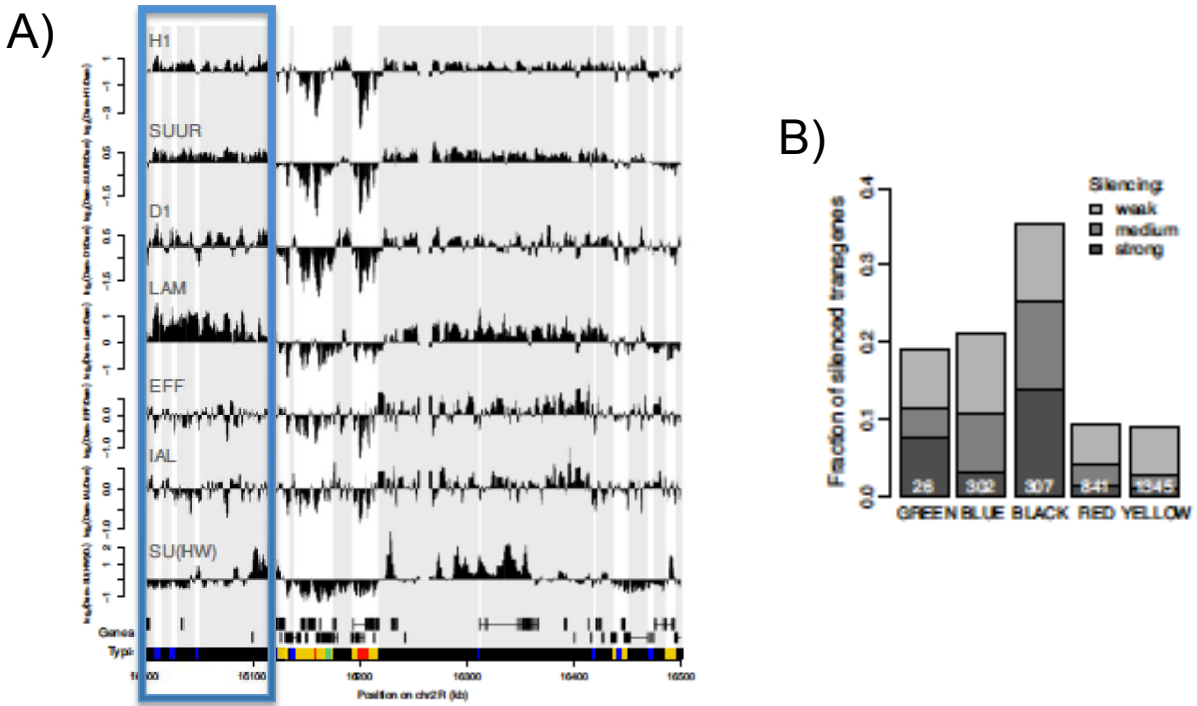


Figure 57: La chromatine noire

A) La chromatine noire est enrichie en diverses protéines telles que l'histone H1, SUUR (Suppressor of Under Replication), D1 (connue pour se lier aux régions satellite), IAL (IplI-aurora-like kinase) et la Lamine (rectangle bleu). **B)** Cette chromatine a la capacité de mettre sous silence beaucoup plus fortement des transgènes que les autres types de chromatine, même celles de type hétérochromatique comme la chromatine bleue et la chromatine verte. [Filion *et al.*, 2010]

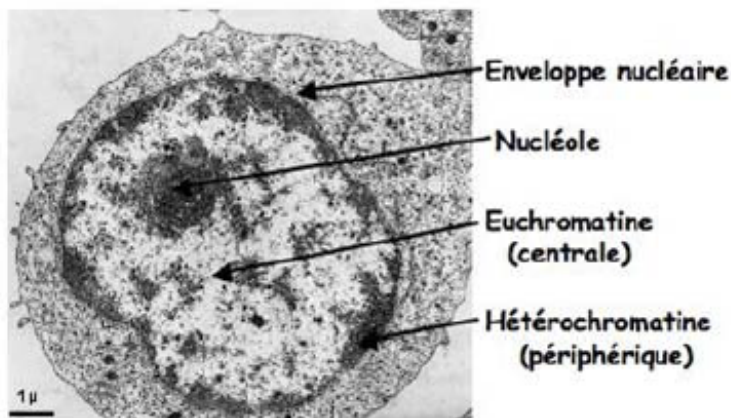


Figure 58: Organisation de la chromatine dans le noyau

La chromatine est présente sous deux aspects à l'intérieur du noyau. L'euchromatine, faiblement condensée et riche en gènes, est située à l'intérieur du noyau tandis que l'hétérochromatine, très condensée et pauvre en gènes, est située en périphérie, proche de l'enveloppe nucléaire. [www.facbio.com]

autres types de chromatine fermé (**Figure 57 B**). En revanche en cellule de drosophile, les gènes qui sont recouverts par cette chromatine peuvent s'exprimer dans certains tissus (Chintapalli, Wang et al. 2007) (**Figure 57 C**) ce qui pourrait être le reflet d'une chromatine transitoire entre deux états chromatinien distincts et régulés développementalement comme la chromatine bleue.

Ces différents types de chromatine, qui sont un premier niveau d'organisation nucléaire, ne sont pas situés dans le noyau de façon aléatoire. En effet, l'hétérochromatine est en grande majorité localisée en périphérie nucléaire tandis que l'euchromatine se situe préférentiellement à l'intérieur du noyau (**Figure 58**).

e).2 Les territoires chromosomiques

Les techniques d'Hybridation In Situ à Fluorescence, ou FISH, appliquées aux chromosomes entiers ont permis d'observer un deuxième niveau d'organisation du génome au sein du noyau. En effet, les chromosomes se répartissent dans des territoires chromosomiques propres à chacun (**Figure 59 A**) (Cremer and Cremer 2001). Ces différents territoires peuvent varier selon l'espèce et les tissus concernés. Deux modèles différents tentent de définir les limites de ces territoires. Le premier, imaginé par l'équipe Cremer, suppose l'existence d'un milieu inter-chromatinien sans chromatine séparant les différents territoires et contenant divers facteurs protéiques comme ceux nécessaires à la transcription (**Figure 59 B**). A l'inverse, selon l'équipe Pombo, les différents territoires sont séparés par une zone contenant un entremêlât de chromatine provenant des différents chromosomes voisins (**Figure 59 C**) (Branco and Pombo 2006).

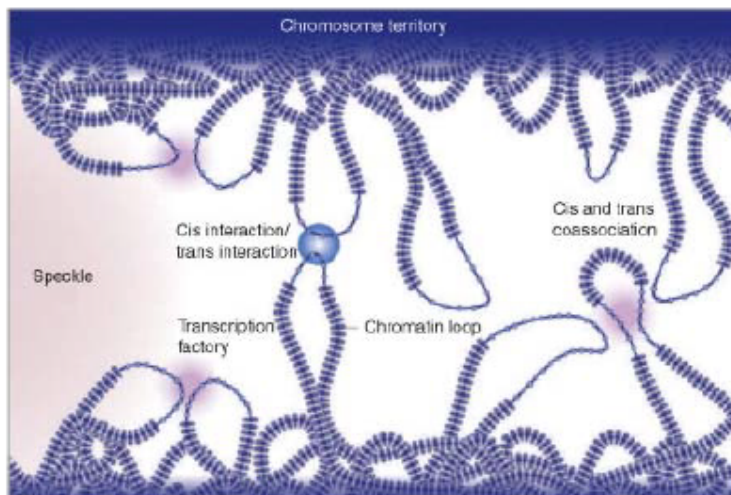
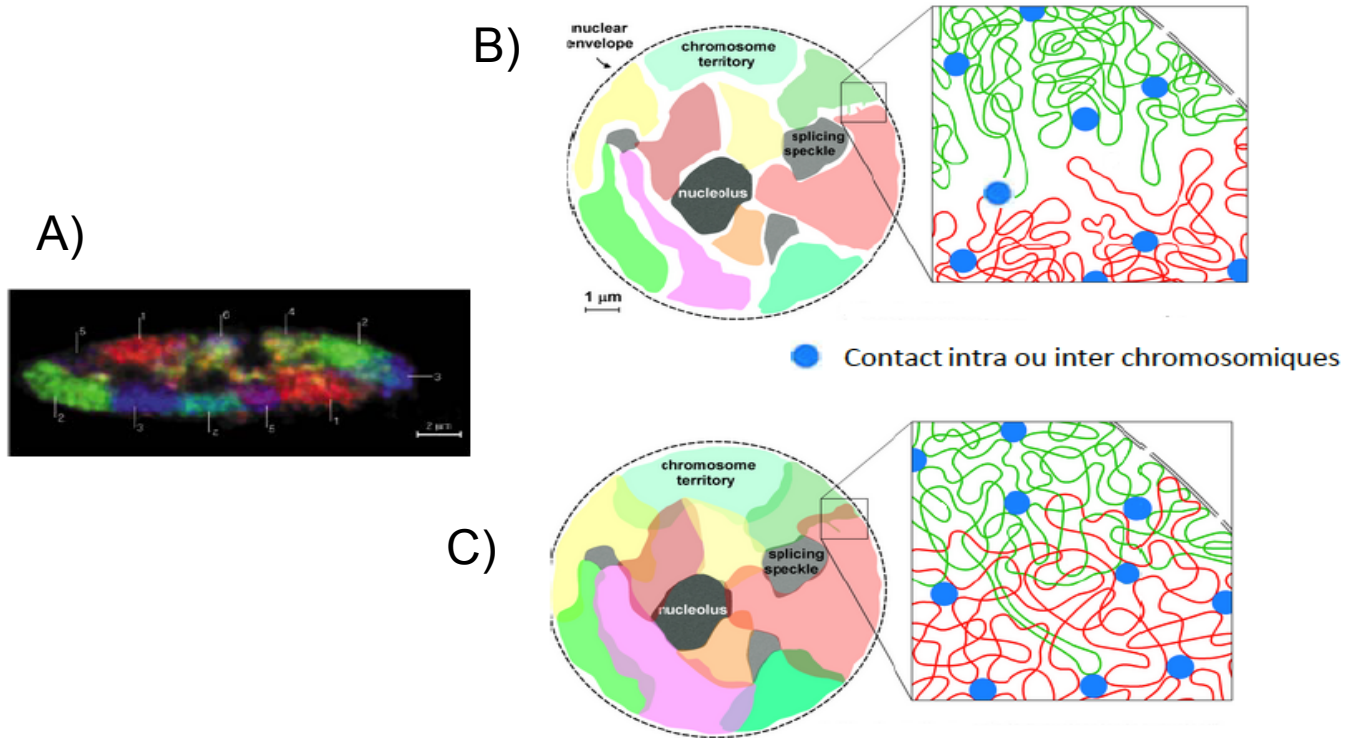


Figure 60: Interactions physiques entre gènes, appelées phénomène de « gene kissing »

Des interactions longues distances sont mises en place entre gènes situés sur le même chromosome ou sur des chromosomes différents, afin d'assurer le fonctionnement de certains mécanismes, comme la régulation transcriptionnelle. [Cremer et Cremer 2010]

e).3 Interactions longues distances

Bien que répartis en territoires chromosomiques distincts, des interactions entre chromosomes demeurent cependant nécessaires notamment lors des événements de réparation ou de réplication de l'ADN, ou encore pour la régulation de l'expression génique (**Figure 60**). C'est ainsi que de nombreuses boucles d'interactions entre des gènes co-régulés ont été mises en évidence. C'est le phénomène de « gène kissing » (Cremer and Cremer). Un des exemples le plus connus concerne les gènes soumis à la régulation par des protéines du type Polycomb. Il a été montré que ces gènes se regroupent par des interactions à longues distances au sein de foyers nucléaires contenant les protéines de type Polycomb (**Figure 61**). Si ce regroupement est aboli des mutations affectant le développement harmonieux de l'individu sont observées (Bantignies, Roure et al. 2011)

Récemment, grâce à la technique de Capture de la Conformation des Chromosomes (3C), qui permet la détection de régions génomiques qui sont à proximité *in vivo* (Dekker, Rippe et al. 2002; Sexton, Bantignies et al. 2009), et l'intégration de cette technique avec des méthodes génomiques, la topologie de la chromatine a pu être étudiée à une résolution et un débit inaccessible par microscopie. Par exemple, en 3C, il a été identifié de nombreuses boucles de chromatine qui rapprochaient des gènes à proximité d'éléments de régulation distants de plusieurs kilobases (Lanzuolo, Roure et al. 2007). La technique de Capture de la Conformation des Chromosome sur puce (4C) a permis de mettre en évidence des réseaux qui regroupent des gènes transcriptionnellement actifs (Schoenfelder, Sexton et al. 2010; Noordermeer, de Wit et al. 2011), des gènes soumis à empreinte (Zhao, Tavoosidana et al. 2006) ou des gènes réprimés par les protéines PcG (Bantignies, Roure et al. 2011).

Durant ces trois dernières années, des équipes ont couplé la technique de 3C avec les techniques de séquençage à haut débit (Hi-C) et ont pu donner une cartographie fine

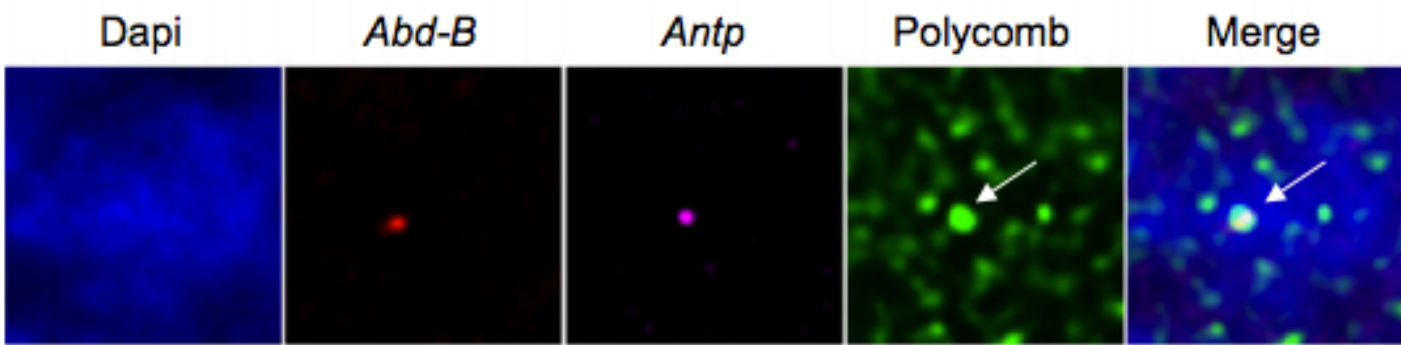


Figure 61: Les foyers Polycomb

Il a été montré que les gènes soumis à la régulation par des protéines du type Polycomb se regroupent par des interactions longues distances au sein de foyers nucléaires contenant les protéines de type Polycomb. Sur cet immuno-FISH les loci *Abd-B* (rouge) et *Antp* (violet) sont regroupés (flèche) au sein de foyers contenant la protéine Polycomb (vert). L'ADN est marqué au DAPI (bleu). [Bantignies *et al.*, 2011]

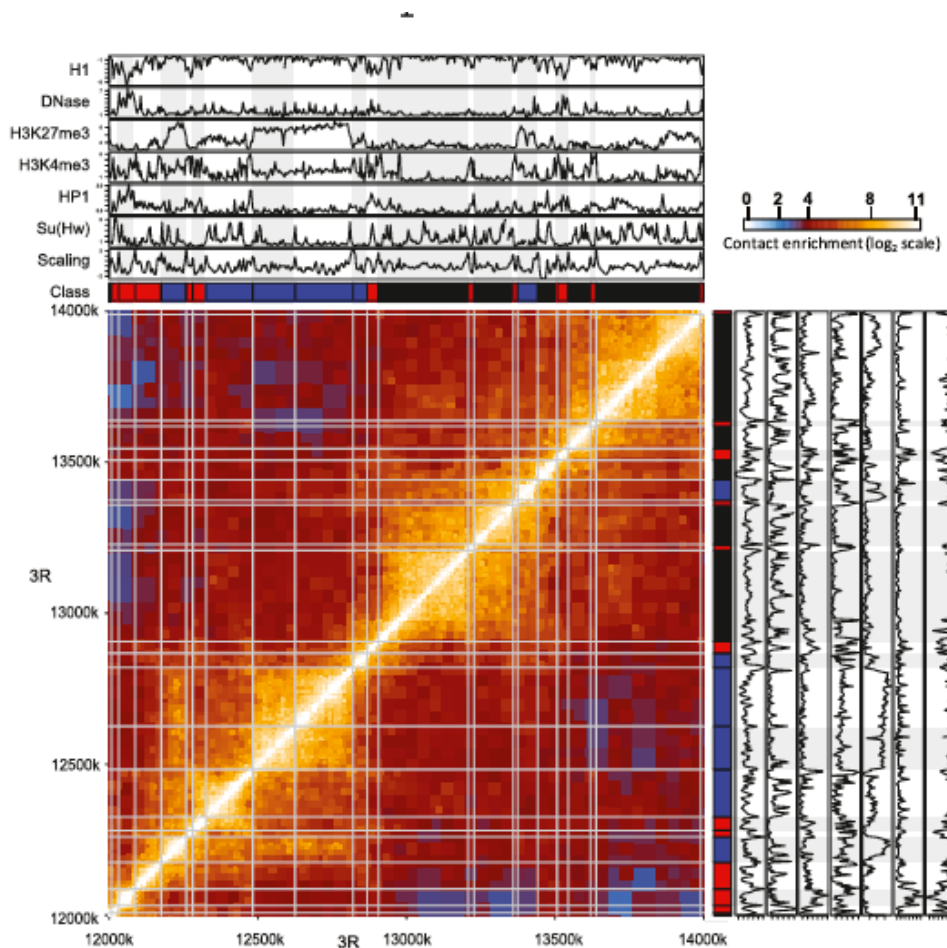


Figure 62: Hi-C et différentes chromatines

Cartographie des contacts d'une région de 2Mb du bras droit du chromosome 3 de drosophile par Hi-C. Il y a plus d'interaction entre deux types de chromatine identiques qu'entre des types différents. L'intensité de rouge représente la force d'interaction entre deux régions (blanc: interaction extrêmement forte, bleue: peu ou pas d'interactions). [Sexton *et al.*, 2012]

des interactions entre les chromosomes chez l'homme (Lieberman-Aiden, van Berkum et al. 2009), la levure (Duan, Andronescu et al. 2010; Tanizawa, Iwasaki et al. 2010), et la drosophile, où, de manière intéressante il semblerait que les différents types de chromatine (en colorama) décrits plus tôt semblent interagir préférentiellement avec de la chromatine de même type (**Figure 62**) (Sexton, Yaffe et al. 2012).

f) RNAi, hétérochromatine et Organisation nucléaire

(Le triangle des Bermudes)

Les régulations épigénétiques sont donc majoritairement réalisées par trois voies différentes : les voies d'ARN interférence permettant de faire du PTGS, la fermeture de la chromatine qui peut faire du TGS et l'organisation nucléaire qui peut localiser une région génomique plutôt dans une région du noyau transcriptionnellement active ou inactive. Ces trois voies, bien que faisant intervenir des acteurs différents, entrent en jeu de manière simultanée pour réguler finement les gènes et les éléments transposables.

L'organisation nucléaire étant un domaine d'étude dont les outils ont été développés récemment, il existe peu de données sur la régulation des éléments transposables par cette voie. En revanche, on sait que le déplacement d'un gène d'une région du noyau transcriptionnellement active vers une région inactive permet d'hétérochromatiser le gène et inversement. Il a aussi été montré chez la drosophile que les interactions longues distances entre les gènes régulés *via* Polycomb semblent être perturbées chez les individus mutés pour des protéines des voies de RNAi, notamment lorsque le gène *piwi* est muté (Grimaud, Bantignies et al. 2006). En accord avec ces dérégulations, des protéines des voies de RNAi ont été retrouvées dans les foyers Polycomb (**Figure 63**) Ces expériences suggèrent donc que certaines protéines du

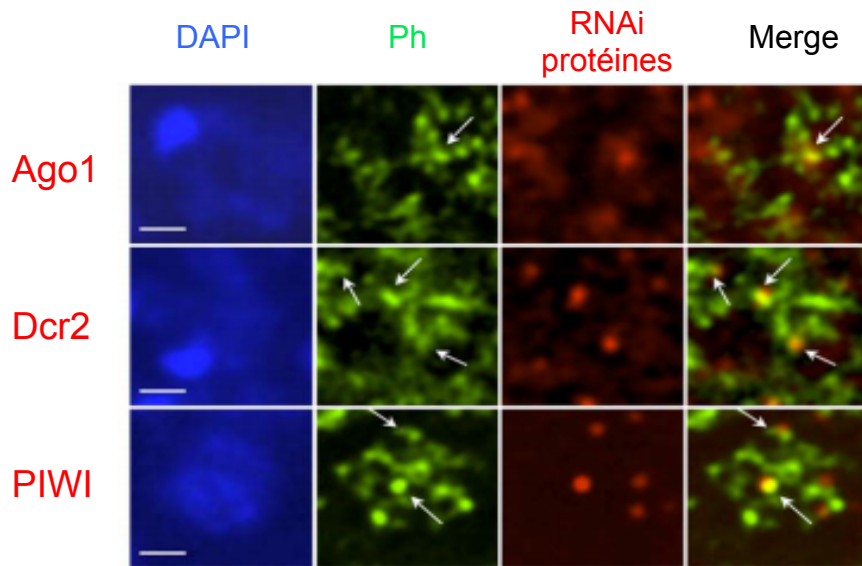


Figure 63: RNAi et foyers Polycomb

Co-immunomarquage entre les foyers de la protéine du PRC1 Polyhomeotic (PH; vert) et des protéines des voies de RNAi (rouge) dans les disques imaginaux de drosophile. PH colocalise (flèches) avec Ago1 (miRNAs), Dcr2 (siRNAs) et PIWI (piRNAs). Des protéines des voies de RNAi sont donc retrouvées dans les foyers Polycomb. Ces expériences suggèrent que certaines protéines du mécanisme RNAi seraient importantes pour l'établissement d'une régulation d'expression dépendante de l'organisation du génome au sein du volume nucléaire. [Grimmaud *et al.*, 2006]

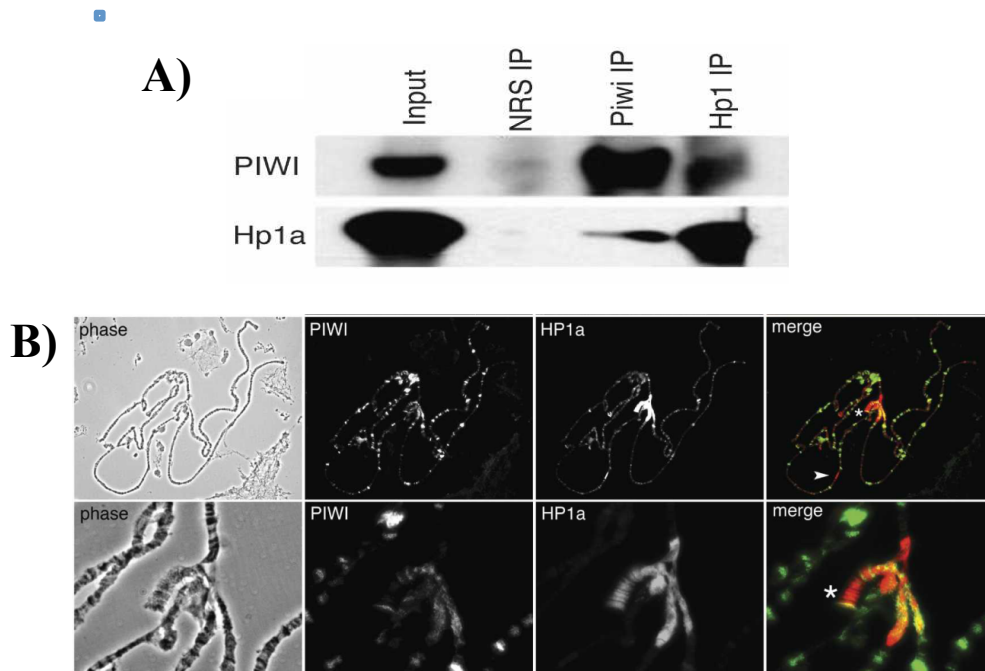


Figure 64: Interaction entre la protéine hétérochromatique HP1 et PIWI

A) Des expériences d'immunoprécipitation avec un anticorps anti-PIWI (PIWI IP) montrent une interaction entre PIWI et HP1 et réciproquement: en immunoprécipitant avec un anticorps anti-HP1 (Hp1 IP) on précipite aussi PIWI.

B) Localisation de PIWI et de HP1 par immunomarquage sur chromosome polytène de drosophile. On peut remarquer que PIWI et HP1 colocalisent et ce particulièrement au niveau du chromocentre (merge). [Brent Brower-Toland *et al.*, 2007]

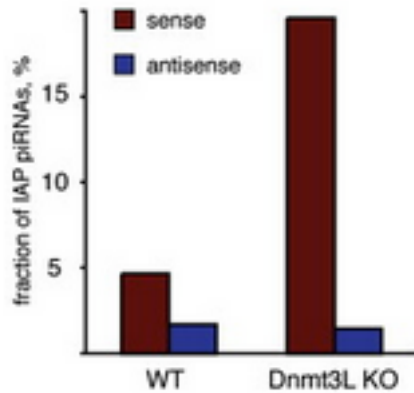
mécanisme RNAi, telle PIWI, seraient importantes pour l'établissement d'une régulation d'expression dépendante de l'organisation du génome au sein du volume nucléaire.

Parallèlement à cela on connaît beaucoup plus de régulations croisées entre les protéines impliquées dans le PTGS et le TGS dans la régulation des éléments transposables.

L'une des premières études majeure qui a montré que la voie d'ARN interférence pouvait être couplée à la mise en place d'un TGS sur les éléments transposables grâce à des modifications de l'état chromatinien a été publiée en 2002. Elle montre (Volpe, Kidner et al. 2002) que certains loci riches en éléments transposables chez *Saccharomyces pombe* sont ciblés par des siRNAs qui permettent le recrutement de plusieurs protéines dont SWI6 (homologue de la protéine hétérochromatique HP1 chez l'homme) et Ctr4 (protéine connue pour méthyliser l'histone H3 sur la lysine 9). Pour la première fois, il était démontré que la production de petits ARNs permettait de recruter des protéines modifiant la structure chromatinienne enveloppant les éléments transposables, de fermer la chromatine et donc d'empêcher la transcription. Chez la drosophile, il a été démontré que la mutation des gènes *piwi*, *aubergine* et *spindle-E* (hélicase) entraînait une baisse de la méthylation H3K9 et une délocalisation de HP1 (Pal-Bhadra, Leibovitch et al. 2004). Une interaction directe entre PIWI et HP1 a également été rapportée (Brower-Toland, Findley et al. 2007) (**Figure 64**). Cependant il a été montré que la protéine PIWI entraînait une perte d'hétérochromatine et l'absence de HP1 dans des régions subtélomériques (Yin and Lin 2007). Il a aussi été montré que les protéines Rhino et HP1 peuvent interagir avec la protéine Cuff (Pane, Jiang et al. 2011).

La méthylation de l'ADN et la voie des piRNAs peut aussi être liée à la régulation des éléments transposables chez les mammifères. La méthylation de leur séquence ADN

A)



B)

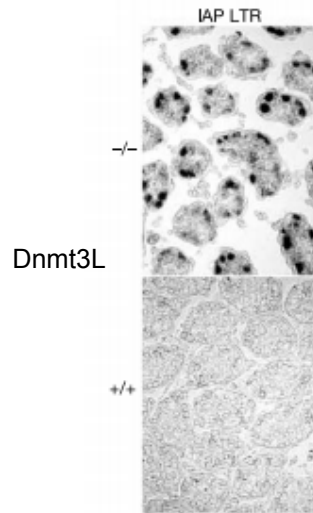
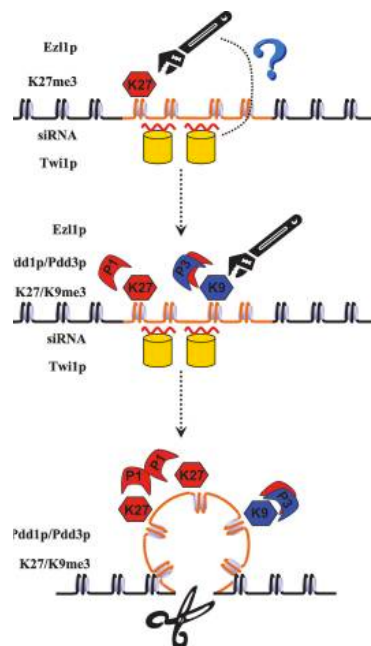


Figure 65: Méthylation ADN et piRNAs chez la souris

Il a été montré un lien chez la souris entre la méthylation ADN et la voie des piRNAs. **A)** Sur ce graphique est représenté le pourcentage de piRNAs qui sont homologues à l'élément transposable IAP en sens (marron) et antisens (bleu). Il y a plus de piRNAs sens en contexte sauvage (WT). Dans un mutant pour la méthyle transférase de l'ADN (Dnmt3L) il y a une augmentation de la quantité de piRNAs sens mais pas antisens (Dnmt3L KO). **B)** Dans les cellules germinales mâles de souris il y a expression de l'élément IAP (Sonde ARN reconnaissant les LTRs de l'élément IAP) en contexte Dnmt3L mutant (panneau du haut) comparé à un contexte sauvage (panneau du bas). [Aravin *et al.*, 2008; Bourc'his *et al.*, 2004]



Recrutement de
Ezh1p via siRNAs

TriMéthylation H3K27
et
compaction de l'ADN

Elimination de
l'ADN compacté

Figure 66: Réarrangement du génome des ciliés et PcG

Chez *Tetrahymena thermophila*, la protéine Ezh1p, l'homologue de E(Z), est recrutée sur les séquences d'ADN à éliminer lors du réarrangement du génome des cilié grâce à des siRNAs et à la protéine Twi1p. Ce recrutement va induire une triméthylation de H3K27, la compaction de la chromatine ciblée et l'excision cette région. [Liu *et al.*, 2007]

(TGS) par Dnmt3L est liée à l'activité de MIWI2 (**Figure 65**) (Bourc'his and Bestor 2004; Aravin, Sachidanandam et al. 2008).

Récemment, il a aussi été montré que les protéines PcG pouvaient se lier à des ARNs non codants (Zhao, Ohsumi et al. 2010) et ainsi permettre de cibler des gènes ou des séquences d'ADN à éliminer comme chez les cilié. Par exemple chez *Tetrahymena thermophila*, la protéine Ezl1p, l'homologue de E(Z), est recrutée sur les séquences d'ADN à éliminer lors du réarrangement du génome grâce à des siRNAs (Liu, Taverna et al. 2007) (**Figure 66**). Les protéines PcG peuvent aussi être impliquées dans la régulation des éléments transposables comme l'élément MLV (endogenous Murine Leukemia Virus) chez la souris (Leeb, Pasini et al. 2010).

Ces données tendent à montrer que la régulation des éléments transposables est étroitement liée à une régulation chromatinienne et à la voie d'ARN interférence mais aucune vision précise du phénomène n'est encore connue chez un autre organisme que *Saccharomyces pombe*.

3) Le modèle d'étude

a) La drosophile

La drosophile est l'un des organismes modèles les plus importants de la génétique classique. La drosophile s'est révélée être un outil de choix dans l'étude du contrôle de l'expression de gènes au cours du développement ou dans la compréhension de l'effet mutagène des éléments transposables. L'entretien des souches s'avère facile, les descendance s'obtiennent en moins de quinze jours et avec un grand nombre d'individus.

La drosophile possède un génome relativement compact avec un nombre limité de séquences répétées et ce génome se répartit sur quatre chromosomes. Non seulement le

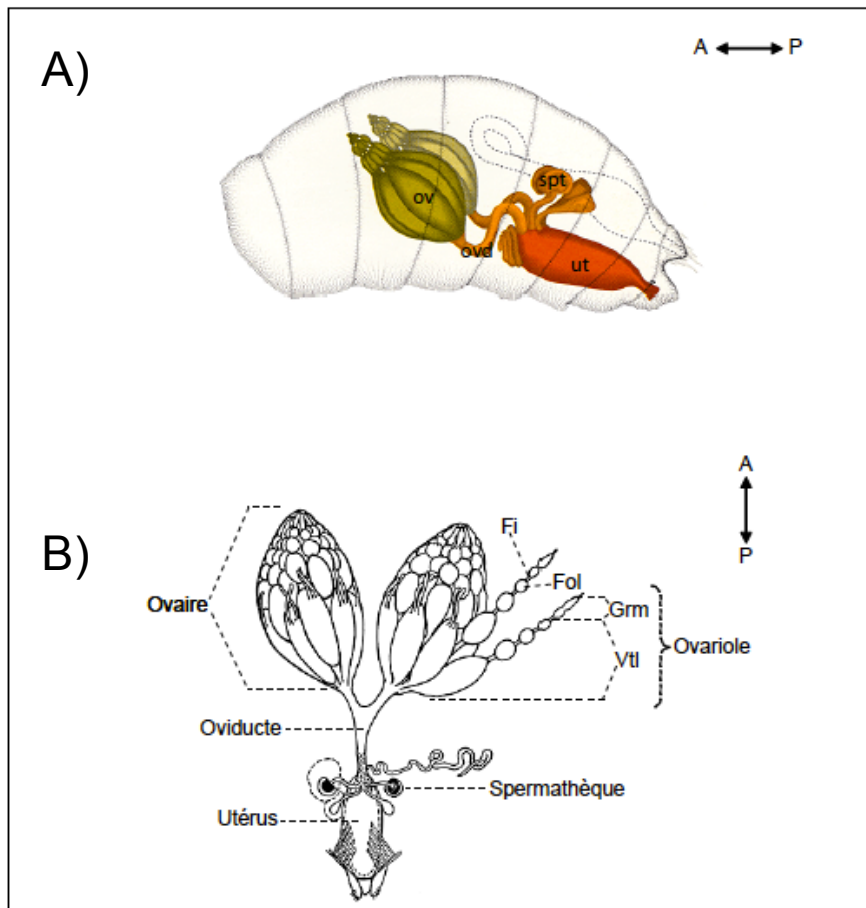


Figure 67: Représentation de l'appareil reproducteur de la drosophile femelle

A) Représentation au sein de l'abdomen de la drosophile adulte.

B) Chaque ovaire contient une vingtaine d'ovarioles. Chaque ovariole est constitué d'une succession de follicules à des degrés de maturité croissants du pôle antérieur vers le pôle postérieur. Le follicule mature sort de l'ovariole et transite par l'oviducte pour arriver dans l'utérus, où il est fécondé par un des spermatozoïdes stockés dans la spermathèque. ov : ovaire ; ovd : oviducte ; spt : spermathèque ; ut : utérus ; Fi : Filament interfolliculaire ; Fol : Follicule ; Vtl : Vitellarium ; A : Antérieur ; P : Postérieur.

[Hartenstein 1993, King 1970]

génomique de la drosophile est entièrement séquencé, du moins les régions euchromatiques, mais de nombreux allèles mutants sont disponibles. Il est donc aisé d'étudier l'effet de certains gènes vis à vis de la mobilisation et de l'expression des éléments transposables. De nombreux outils génétiques sont actuellement couramment utilisés chez la drosophile tels que la transgénèse, la mutagenèse, les systèmes de recombinaison ou même la réalisation de clones de cellules mutantes au sein d'un organisme de génotype sauvage.

L'étude des mécanismes de répression des éléments transposables est surtout réalisée dans l'appareil reproducteur de la drosophile, pour deux raisons principales. En effet, la voie des piRNAs est majoritairement active dans ces tissus et c'est ce tissu qui va donner la descendance et qui doit donc être protégé des effets mutagènes des éléments transposables. En outre, il semblerait que la majorité des éléments transposables soit capable d'être activés dans ces tissus reproducteurs.

b) L'ovogénèse adulte de drosophile

Chez la drosophile, l'ovogénèse se déroule de façon continue tout au long de la vie adulte des femelles, ce qui leur permet, dans des conditions optimales, de pondre jusqu'à 90 oeufs par jours. L'appareil reproducteur femelle de la drosophile est composé d'une paire d'ovaires, des canaux génitaux efférents et de leurs annexes (**Figure 67**) (King 1970). Chaque ovaire contient une vingtaine d'ovarioles, qui sont des structures polarisées où se forment et se développent les follicules ovariens. Chaque ovariole est constituée d'une succession de follicules à des degrés de maturité croissant du pôle antérieur vers le pôle postérieur. Un ovariole contient en moyenne 6 ou 7 follicules qui sont reliés les uns aux autres par des filaments interfolliculaires. Pendant sa maturation, l'ovocyte se développe à l'intérieur du follicule ovarien composé de cellules germinales entourées par un épithélium de cellules somatiques. Au cours de l'ovogénèse, les interactions entre les cellules

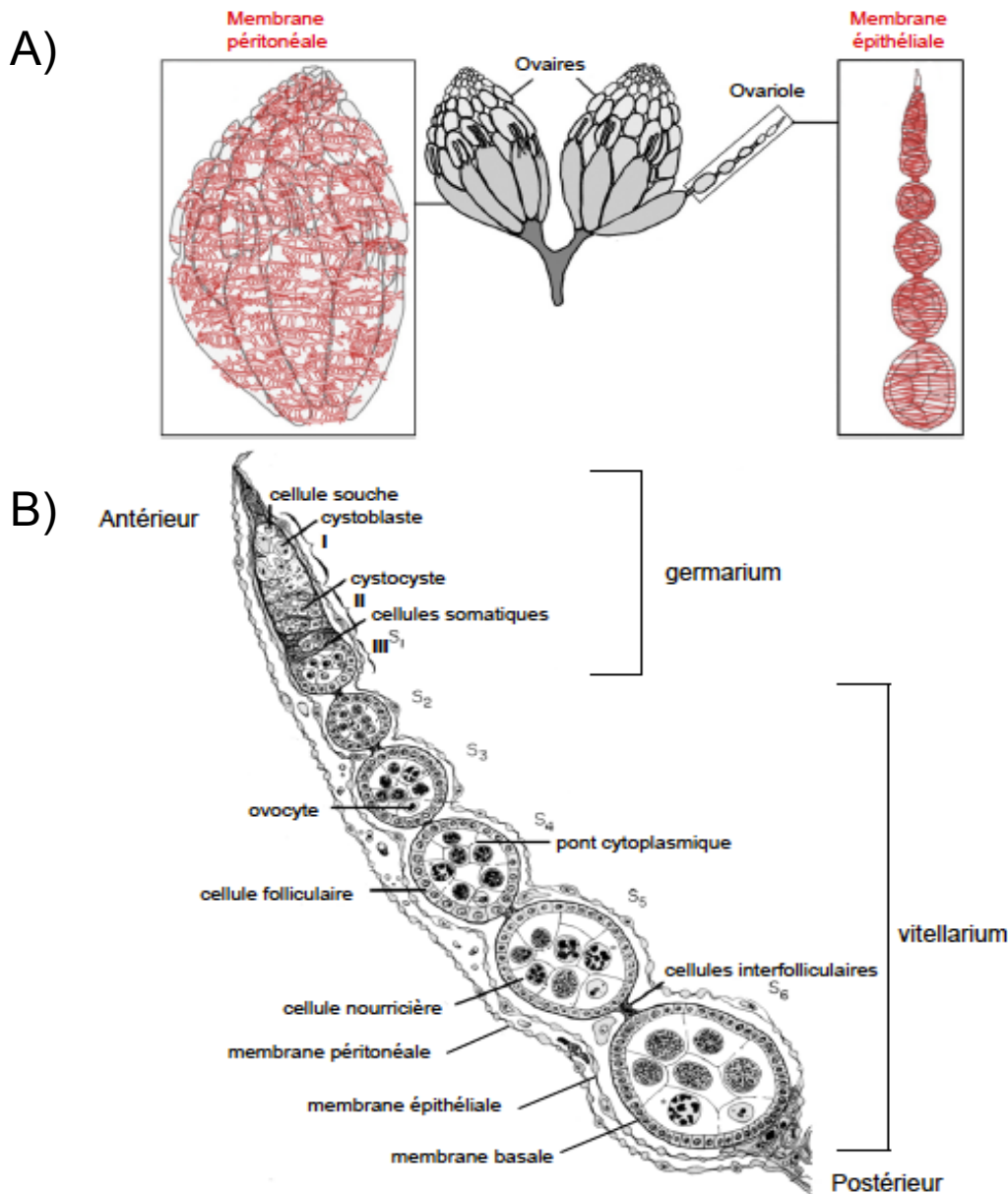


Figure 68: Représentation schématique d'une paire d'ovaire et d'un ovariole

A) Les ovaires sont entourés par la membrane péritonéale et les ovarioles par la membrane épithéliale. Ces deux membranes sont représentées en rouge.

B) Un ovariole est divisé en deux parties : le germarium où se forment les follicules et le vitellarium où les follicules se développent. Les chiffres en caractères romains représentent les trois régions du germarium. S1 à S6 représentent les six premiers stades de développement des follicules. Au total, 14 stades de maturation ont été décrits : le stade 1 correspond au follicule qui bourgeonne du germarium, le stade 14 correspond au follicule mature prêt à être fécondé. [Hudson *et al.*, 2008 (A), King 1970 (B)]

somatiques et les cellules germinales et au sein de chacune de ces deux populations sont essentielles afin d'assurer la formation du follicule et la maturation de l'ovocyte.

Chaque follicule est composé de deux lignages cellulaires : les cellules somatiques ou folliculaires et les cellules germinales. Ces dernières forment un cyste (ou syncytium) de 16 cellules composé d'un ovocyte et 15 cellules nourricières polyploïdes. Les cellules germinales sont interconnectées entre elles par des ponts cytoplasmiques. Le cyste est entouré par une monocouche de cellules somatiques : l'épithélium folliculaire ovarien. Les cellules folliculaires possèdent une polarité apico-basale, leur face apicale est en contact avec la lignée germinale, facilitant ainsi les échanges et la communication entre les deux populations (**Figure 68**). Les cellules folliculaires sont connectées les unes aux autres par des jonctions serrées et par des jonctions d'adhérence. Dès sa formation, le follicule est polarisé selon l'axe antéropostérieur. Cette polarité est marquée par le positionnement de l'ovocyte au pôle postérieur du follicule. Puis, au cours de l'ovogenèse, différents territoires vont être définis au sein de l'épithélium folliculaire, au départ suivant l'axe antéropostérieur et ensuite suivant l'axe dorsoventral.

La production de follicules de façon continue tout au long de la vie de la femelle est assurée par l'activité des cellules souches somatiques et germinales présentes dans le germarium. Le germarium a été divisé en quatre régions morphologiquement distinctes : les régions I, IIa, IIb et III alignées le long de l'axe antéropostérieur (**Figure 69**). Dans la région I se trouve les cellules souches germinales qui se divisent alternativement. A chaque division, elles génèrent une nouvelle cellule souche et un cystoblaste qui va subir quatre divisions à cytokinèse incomplète conduisant à la formation du cyste de 16 cellules dans la région IIa. Des cellules escortes issues des « escort stem cells » entourent le cystoblaste et restent associées avec le cyste en formation (Decotto and Spradling 2005). A la frontière entre les régions IIa et IIb se trouvent les cellules souches somatiques qui produisent les

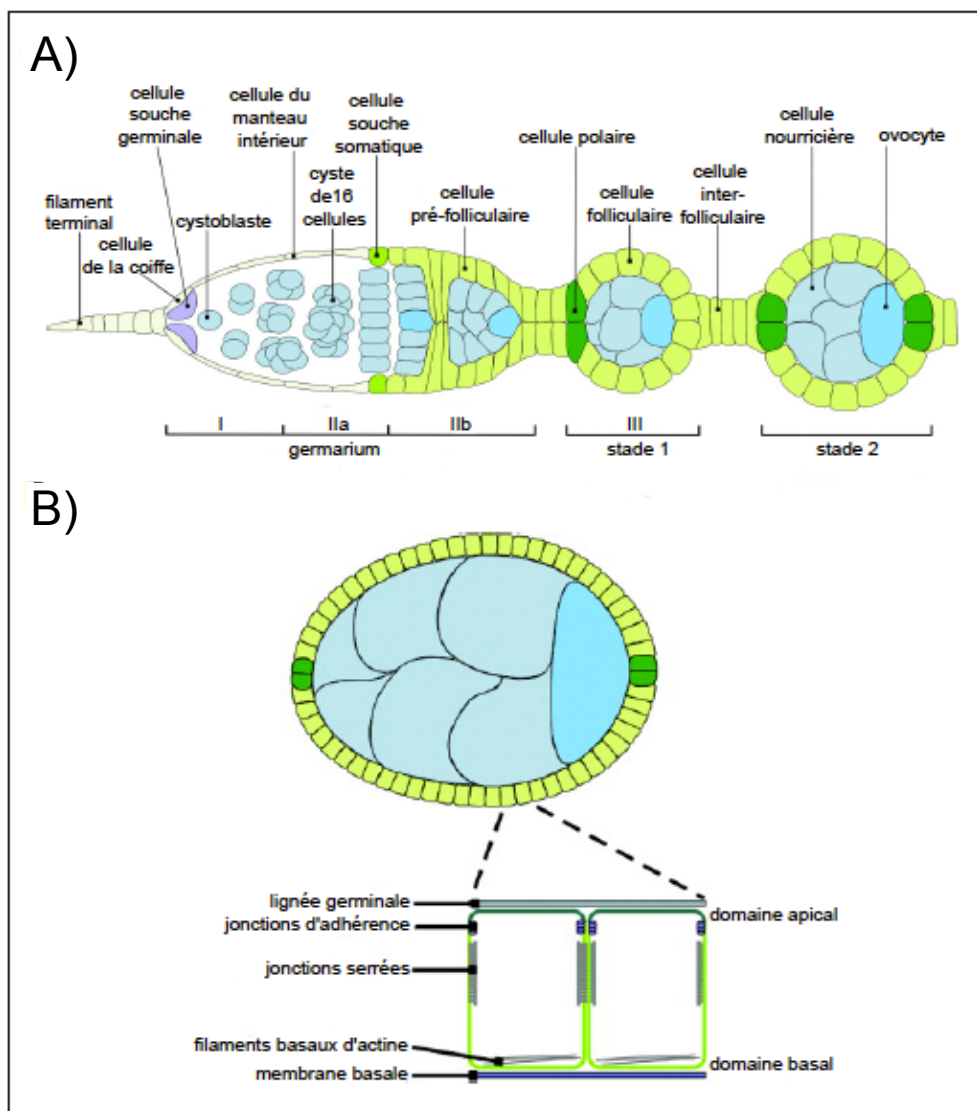


Figure 69: Représentations schématiques d'un germarium et de follicules ovariens

A) Le germarium est divisé en 4 régions : I, IIa, IIb, III. Dans la région I se trouvent les cellules souches germinales (CSGs) qui se divisent asymétriquement pour générer une nouvelle CSG et un cystoblaste. Ce dernier va ensuite subir 4 cycles de division à cytokinèse incomplète conduisant à la formation d'un cyste de 16 cellules : 15 cellules nourricières et 1 ovocyte (région IIa). Les cellules souches somatiques (CSSs) sont localisées à la frontière entre les régions IIa/IIb. Les cellules filles des CSSs migrent entre les cystes afin de les individualiser. Dans la région III, le follicule de stade 1 bourgeonne du germarium et des cellules somatiques s'intercalent les unes entre les autres pour former un filament interfolliculaire.

B) Représentation d'un follicule ovarien composé d'un cyste de 16 cellules germinales dont 1 ovocyte localisé au pôle postérieur. Le cyste germinale est entouré par une monocouche de cellules épithéliales polarisées : les cellules folliculaires, dont la face apicale est en contact avec les cellules germinales et le domaine basal, en contact avec la membrane basale. [Horne-Badovinac et Bilder, 2005]

précurseurs des cellules préfolliculaires. Dans la région IIb, des contacts s'établissent entre le cyste et les cellules préfolliculaires et les cellules escortes vont être éliminées. Le cyste qui avait une forme arrondie adopte alors une forme en lentille. A ce stade, l'ovocyte est déterminé et il est localisé au centre du cyste. Les cellules pré-folliculaires migrent pour entourer le cyste et le séparer du cyste plus jeune localisé en antérieur. L'épithélium se forme à partir d'une transition cellule mésenchymateuse-cellule épithéliale. Les cellules somatiques sont en contact avec la membrane basale ce qui induit une polarisation initiale qui, pour être complète, nécessite le contact avec la lignée germinale (Tanentzapf, Smith et al. 2000). Le follicule qui vient d'être formé est localisé dans la région III du germarium. L'encapsulation du cyste et la sortie du follicule du germarium reposent sur la différenciation des cellules pré folliculaires en cellules folliculaires et en cellules intercystes. Ces dernières sont les précurseurs des cellules polaires et des cellules interfolliculaires.

La formation du follicule est un processus relativement long. En effet, il s'écoule environ 6 jours (à 25°C) entre le moment où la cellule souche germinale se divise pour générer un cystoblaste et la sortie du follicule du germarium (Spradling 1993).

c) Les éléments ZAM et Idefix

c) .1 historique

Au commencement de l'étude de ZAM et Idefix, l'équipe de Chantal Vaury étudiait l'effet mutagène du facteur I sur un modèle utilisant le gène *white*. Ce gène est localisé sur le chromosome X de *Drosophila melanogaster* et code une protéine White (W) impliquée dans la coloration des yeux. Dans une lignée contenant l'allèle *wIR6*, le facteur I est inséré dans le premier intron du gène. Cette insertion provoque une dérégulation de l'épissage du premier intron dû à l'apport par le facteur I d'un nouveau site d'épissage. Deux types de

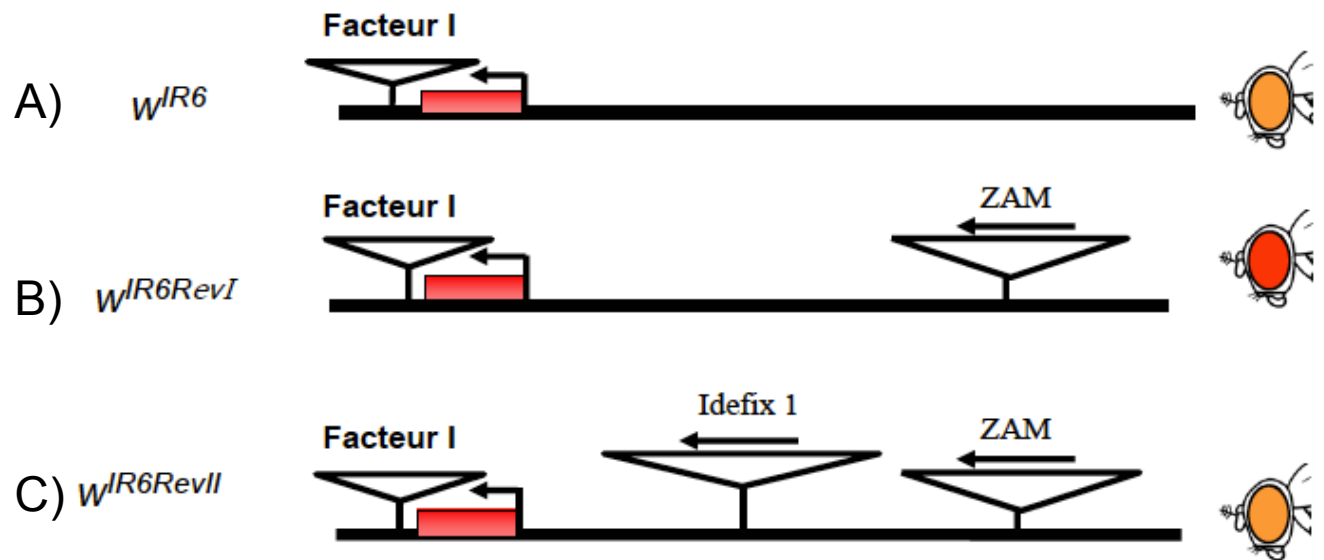


Figure 70: Structure des allèles w^{IR6} , $w^{IR6RevI}$ et $w^{IR6RevII}$ et phénotype des yeux associé à ces allèles

A) La lignée w^{IR6} présente une insertion d'un facteur I dans le premier intron du gène *white* conduisant ainsi à un phénotype présentant des yeux orange. B) La lignée $w^{IR6RevI}$ dérive de w^{IR6} et présente une insertion de l'élément ZAM en amont du gène *white*. La présence de ce retrotransposon induit une réversion du phénotype orange vers le phénotype sauvage. C) La lignée $w^{IR6RevII}$ dérive de $w^{IR6RevI}$ et porte une nouvelle insertion d'un retrotransposon nommé Idefix, 1.7 kb en amont du gène *white*. Cette nouvelle insertion cause un phénotype d'oeil orange. Le rectangle rouge représente le gène *white*. [Conte *et al.*, 2000]



Figure 71: Structure génomique d'*Idefix*

Le retrotransposon à LTR *Idefix* est constitué de deux LTRs (Long Terminal Repeat), d'une région en 5' non traduite (5' UTR) et de trois cadres de lecture, *gag*, *pol* et *env*. [Desset *et al.*, 1999]; [Conte *et al.*, 2002]

transcrits sont alors observés : un transcrit de type sauvage, correctement épissé et un transcrit aberrant. Le messenger de type sauvage traduit en protéine active est moins abondant que dans l'allèle sauvage, ce qui a pour conséquence de modifier la coloration des yeux. Les individus portant l'allèle wIR6 ont des yeux brun-orangé tandis que ceux portant l'allèle de type sauvage sont rouge brique (Lajoinie, Drake et al. 1995) (**Figure 70**).

Dans le but d'identifier des gènes suppresseurs de la mutation wIR6, une mutagenèse a été menée sur la souche wIR6 par mobilisation de l'élément P ou mutagenèse P/M. Un mâle révertant est apparu présentant des yeux rouge brique comme le type sauvage. Il a été appelé Rev (Révertant). La lignée Rev s'est rapidement révélée instable car des individus présentant des couleurs d'œil variables apparaissaient sporadiquement. Les analyses génétiques et moléculaires ont pu mettre en évidence les causes de ces changements phénotypiques. C'est l'insertion de deux rétrotransposon à LTR de *Drosophila melanogaster*, ZAM et Idefix, en amont du gène white qui sont à l'origine des variations de couleur d'œil (Leblanc, Desset et al. 1997; Desset, Conte et al. 1999; Conte, Calco et al. 2000; Conte, Dastugue et al. 2002). Suite à la mutagenèse P/M, un changement génétique s'est produit qui a eu pour conséquence des insertions de ZAM et Idefix au locus white. Les expériences d'hybridation *in situ* sur chromosomes polytènes ont permis de mettre en évidence que ZAM et Idefix ont été amplifiés dans la lignée Rev en comparaison de la lignée parentale wIR6 (Desset, Conte et al. 1999).

c) .2 Structure et expression de ZAM et Idefix

Les éléments ZAM et Idefix sont des rétrotransposons à LTR de *Drosophila melanogaster*. Ils nécessitent une étape de rétrotranscription dans leur cycle de mobilisation et présentent de fortes similitudes avec les rétrovirus de vertébrés (Boeke and

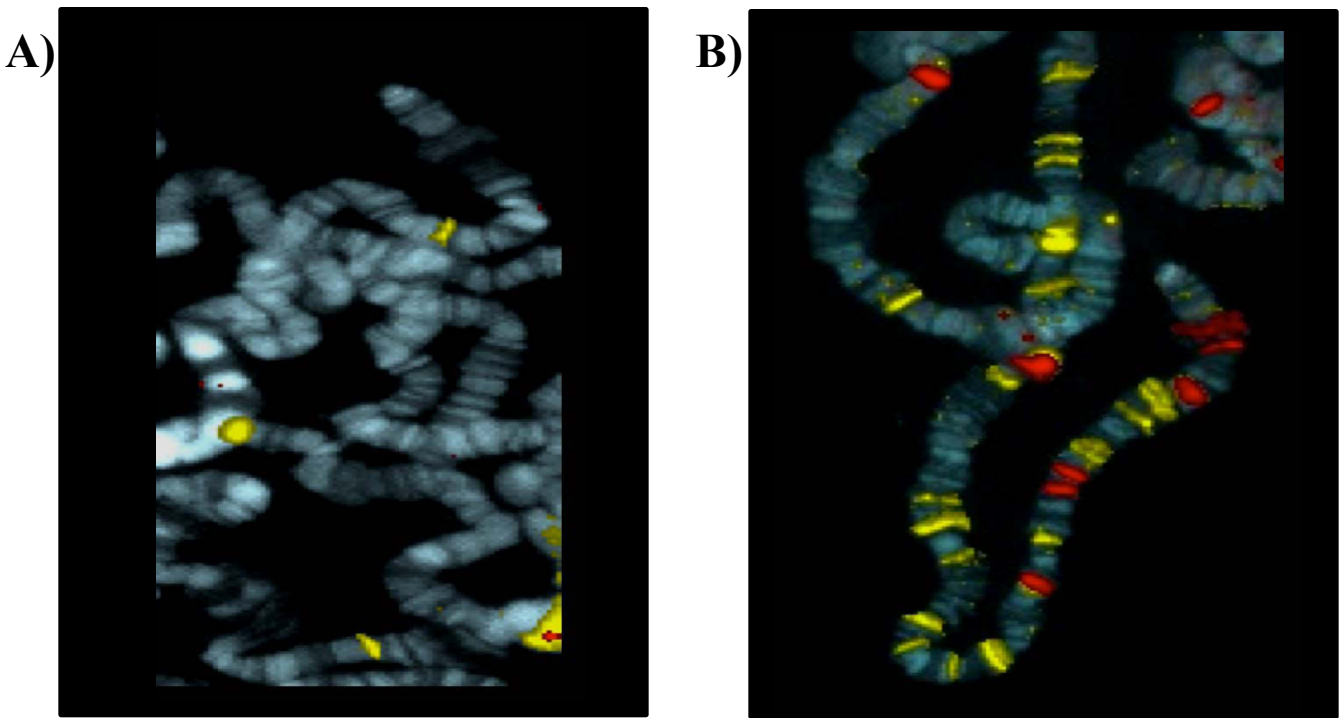


Figure 72 : Distribution génomique des séquences d' *Idefix*.

Observation par FISH (Fluorescent In Situ Hybridization) des chromosomes polytènes des glandes salivaires de deux souches de drosophiles. L' une régule la transposition d' *Idéfix* (A) et l' autre ne la régule pas (B). Le signal d' hybridation dû à la sonde d' *Idefix* est visualisé en jaune (le signal rouge correspond à un autre élément, appelé *ZAM*). [Desset *et al.*, 1999].

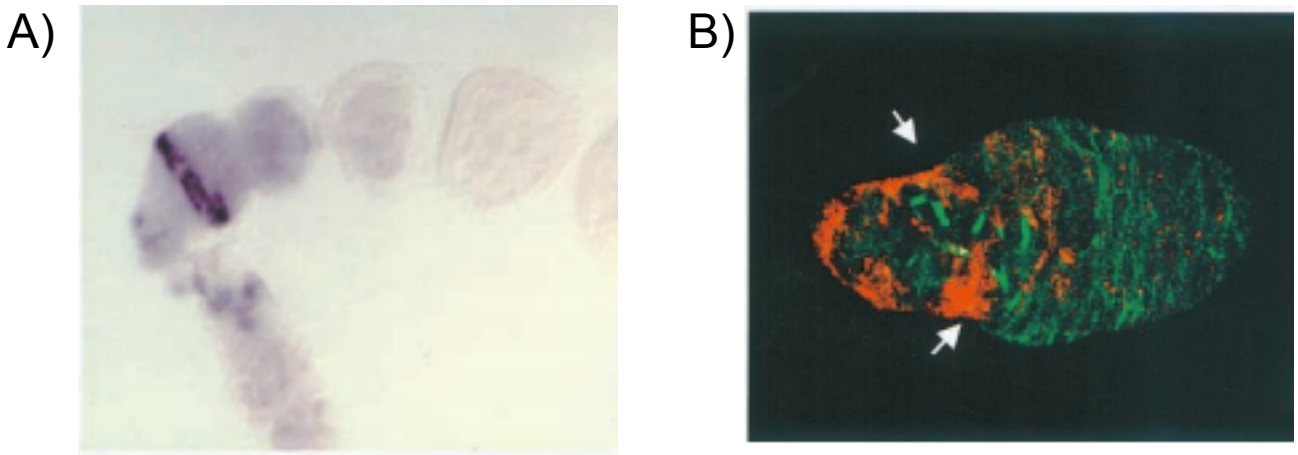


Figure 73: Expression d' *Idefix* en contexte mutant pour COM

A) Hybridation *in situ* avec une sonde reconnaissant le transcrit d' *Idefix* dans un contexte génétique permissif (mutation dans COM). *Idefix* est exprimé précocement dans le germarium. B) Immunomarquage avec un anticorps anti β -gal (rouge) d' un germarium contenant une construction de type LTR(*Idefix*)-LacZ dans un contexte génétique permissif (mutation dans COM). L' expression d' *Idefix* semble être restreinte aux cellules somatiques précoces. [Tcheressiz *et al.*, 2002]

Pickeral 1999). Deux longues régions terminales répétées ou LTR encadrent les phases ouvertes de lecture. Leurs LTRs contiennent la plupart des éléments de contrôle transcriptionnel, tels qu'un promoteur et des séquences régulatrices de type enhancer nécessaires à l'expression. Leurs LTRs sont de longues régions directement répétées, localisées aux extrémités de la forme provirale.

Durant ma thèse, je me suis plus particulièrement intéressé à l'élément Idefix (**Figure 71**). Il est présent en un nombre peu important de copies euchromatiques (**Figure 72**) qui ont conservé leur capacité de transposition dans le génome. En revanche un grand nombre de copies dégénérées sont présentes dans l'hétérochromatine. Il a été montré que l'expression d'Idefix est possible dans les cellules folliculaires (**Figure 73**) (Tcheressiz, Calco et al. 2002), tissu somatique de l'appareil reproducteur et qu'il est régulé par la voie des piRNAs dans ce même tissu (Desset, Buchon et al. 2008). Cette régulation a été démontrée par les expériences suivantes : des transgènes ont été construits selon le modèle décrit figure 74A : « promoteur-GFP-Idefix ». Dans ces transgènes, la GFP correspond au gène rapporteur codant la « Green Fluorescente Protein ». Idefix correspond à un fragment d'Idefix orienté de façon à ce que le brin transcrit corresponde au transcrit sens de cet élément. Dans les lignées transgéniques établies avec ces constructions, aucune expression de la GFP n'est observée dans les cellules somatiques de l'appareil reproducteur de la drosophile (**Figure 74 B**). En revanche l'expression de la GFP est moins réprimée si le fragment d'Idefix est placé en antisens (**Figure 74 C et D**). La présence d'un fragment d'Idefix sens peut donc réprimer l'expression d'un gène rapporteur qui lui est accolé dans les cellules folliculaires de drosophile. D'autre part, des expériences de RNase protection ont montré la présence de petits ARNs ayant les caractéristiques de piRNAs, c'est-à-dire des petits ARNs de 28 à 30 nucléotides correspondant au brin antisens d'Idefix (**Figure 75**). Il a été montré que la répression d'Idefix était dépendante de la protéine PIWI (**Figure**

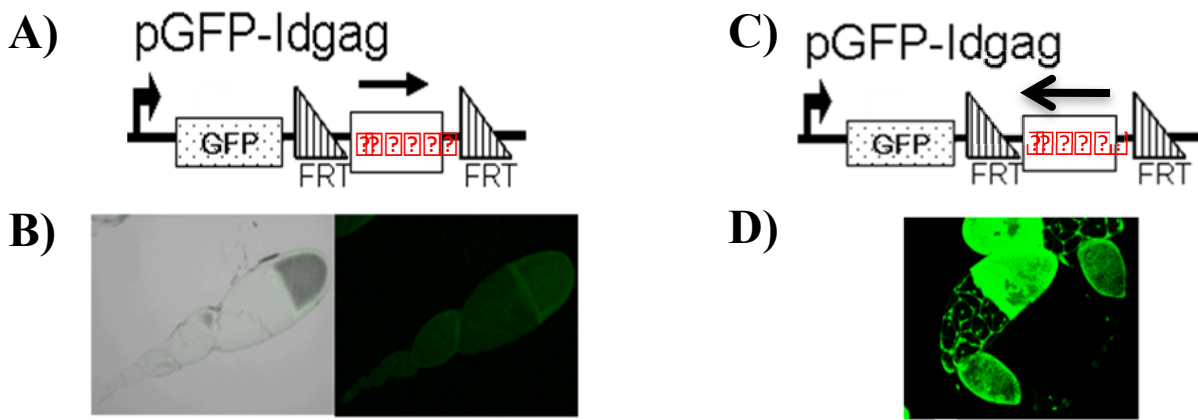


Figure 74: Expression de la GFP à partir de transgènes portant un fragment d'*Idefix* en aval

A) Représentation schématique d'un transgène [Promoteur-GFP-*Idefix*] où *Idefix* est inséré dans le sens de transcription de la GFP, qui sert de gène rapporteur.

B) Observation de l'expression de la GFP dans des ovaires de drosophiles contenant le transgène décrit en A). *Idefix*, inséré dans le sens de transcription de la GFP, permet sa répression.

C) Représentation schématique d'un transgène [Promoteur-GFP-*Idefix*] où *Idefix* est inséré en antisens de la GFP qui sert de gène rapporteur.

D) Observation de l'expression de la GFP dans des ovaires de drosophiles contenant le transgène décrit en C), *Idefix*, inséré en antisens de la GFP, ne permet pas sa répression. [Desset *et al.*, 2008]

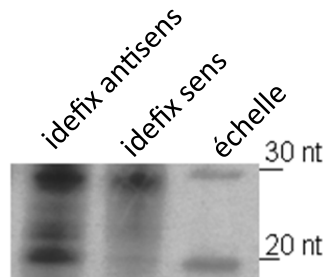


Figure 75: Détection de petits ARNs antisens d'*Idefix*.

Des expériences de RNase protection ont montré la présence dans le génome de drosophile de nombreux petits ARNs antisens d'*Idefix* d'une taille proche de 30 nucléotides. D'autres ARNs plus petits sont aussi détectés. [Desset *et al.*, 2008]

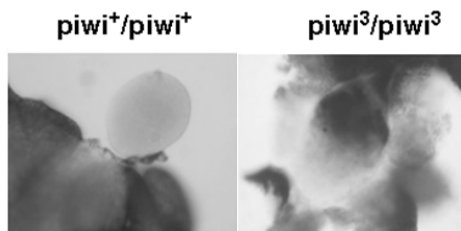


Figure 76: La régulation d'*Idefix* est dépendante de la protéine PIWI

Hybridation *in situ* d'ovaires de larves de drosophiles avec une sonde révélant les transcrits d'*Idefix*. On remarque qu'*Idefix* n'est pas transcrit dans les souches sauvages (gauche). En revanche, dans une souche mutante pour PIWI (droite), il y a transcription d'*Idefix*. [Desset *et al.*, 2008]

76) et que le locus producteur de piRNAs COM (Centre Organisateur de la Mobilisation) (Desset, Meignin et al. 2003) est le master locus régulateur d'Idefix.

Ces expériences montrent qu'Idefix est régulé par la voie des piRNAs dans les cellules somatiques de l'appareil reproducteur et que cette répression est au moins due à un phénomène de type PTGS.

L'objectif de ma thèse a été de démontrer quelles pouvaient être les régulations qui répriment Idefix dans différents types de tissus de la drosophile (somatiques, germinaux, cellules folliculaires) et si cette régulation impliquait du PTGS, du TGS ou les deux.

Résultats

Publication 1

Régulation des éléments transposables dans les tissus somatiques

1) Introduction

Lors de mon arrivé au laboratoire l'équipe venait de montrer que les éléments transposables ZAM et Idefix étaient régulés par la voie des piRNAs dans les tissus somatiques qui entourent l'ovaire (cellules folliculaires) et par une voie indépendante de PIWI dans les autres tissus somatiques (embryon, larve, pupes, mouche adulte) (Desset, Buchon et al. 2008). Il est clair que la voie des piRNAs ne semble pas avoir de rôle dans les tissus somatiques en dehors de l'ovaire car la grande majorité des protéines impliquées dans cette voie ne sont pas exprimées dans ces tissus.

En revanche de nombreuses équipes venaient de publier l'existence d'une nouvelle voie de régulation des éléments transposables dans ces tissus somatiques : les endo-siRNAs. Ces petits ARNs double brin endogènes semblent pouvoir réguler l'expression des éléments transposables dans les tissus somatiques de drosophile (Chung, Okamura et al. 2008; Czech, Malone et al. 2008; Ghildiyal, Seitz et al. 2008; Kawamura, Saito et al. 2008; Okamura, Chung et al. 2008).

Chez la levure il a été montré que les petits ARNs endogènes de type siRNAs étaient impliqués dans la formation de l'hétérochromatine. Or il a été montré que les protéines Ago2 et Dcr2 jouent un rôle dans la formation de l'hétérochromatine ainsi que dans la fabrication des endo-siRNAs chez la drosophile.

Mon objectif a été de démontrer si, dans les tissus somatiques dans et en dehors de l'ovaire, il existait un lien entre les régulations de type PTGS et la mise en place d'une structure de chromatine compacte sur les éléments transposables euchromatiques.

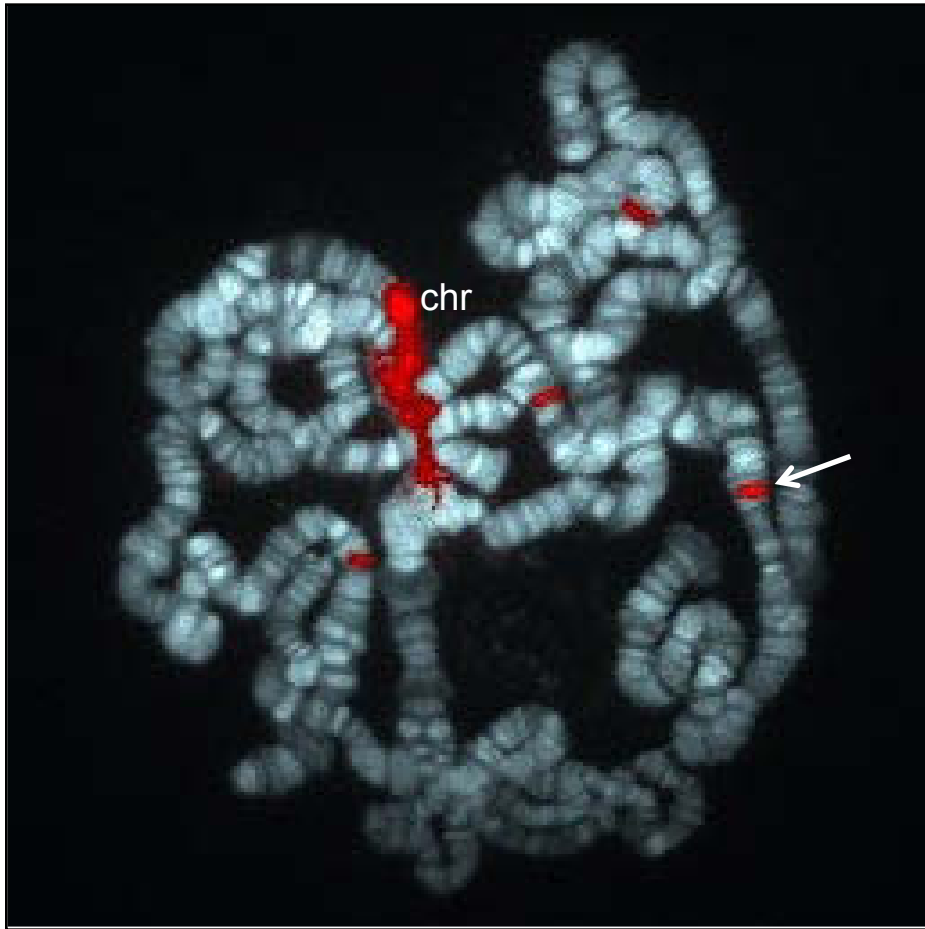


Figure 77 : Distribution génomique des séquences d'*Idefix*.

Observation par FISH (Fluorescent In Situ Hybridization) des chromosomes polytènes des glandes salivaires de drosophiles. Une sonde dirigée contre les éléments Idefix (rouge) montre la présence de copies euchromatiques (flèches) sur les bras des chromosomes et de nombreuses copies hétérochromatiques sur le chromocentre (chr). [Desset *et al.*, 1999].

2) Analyse des éléments transposables euchromatiques

Les éléments transposables sont présents aussi bien dans l'euchromatine que dans l'hétérochromatine (**Figure 77**), il est donc impossible de suivre l'expression ou la structure chromatinienne des éléments transposables euchromatiques uniquement. Toutes les expériences qui sont réalisées sur les éléments endogènes ne peuvent discriminer si les dérégulations observées se passent sur des éléments euchromatiques ou hétérochromatiques. Cette présence ubiquitaire pose un problème du fait que, par exemple, la majorité des clusters de piRNAs se trouvent dans l'hétérochromatine et sont composés d'éléments transposables.

Pour pallier ce problème et afin de pouvoir répondre clairement à mon objectif, j'ai donc utilisé des transgènes situés sur des régions euchromatiques et présents en une seule copie (Desset, Buchon et al. 2008).

Polycomb Group-Dependent, Heterochromatin Protein 1-Independent, Chromatin Structures Silence Retrotransposons in Somatic Tissues Outside Ovaries

J. Dufourt¹, E. Brasset², S. Desset¹, P. Pouchin¹, and C. Vaury^{1,*}

Clermont Université, Université d'Auvergne, BP 10448, F-63000 Clermont-Ferrand, France; Inserm, U 931, F-63001 Clermont-Ferrand, France and CNRS, UMR 6247, F-63001 Clermont-Ferrand, France¹ and Clermont Université, Université d'Auvergne, Institut Universitaire de Technologie, BP 10448, F-63000 Clermont-Ferrand, France; Inserm, U 931, F-63001 Clermont-Ferrand, France and CNRS, UMR 6247, F-63001 Clermont-Ferrand, France²

*To whom correspondence should be addressed. Tel: +33 4-17-81-84. Fax: +33 4-27-61-32.
E-mail: chantal.vaury@u-clermont1.fr

Edited by Osamu Ohara

(Received 20 May 2011; accepted 8 August 2011)

Abstract

Somatic cells are equipped with different silencing mechanisms that protect the genome against retrotransposons. In *Drosophila melanogaster*, a silencing pathway implicating the argonaute protein PIWI represses retrotransposons in cells surrounding the oocyte, whereas a PIWI-independent pathway is involved in other somatic tissues. Here, we show that these two silencing mechanisms result in distinct chromatin structures. Using sensor transgenes, we found that, in somatic tissues outside of the ovaries, these transgenes adopt a heterochromatic configuration implicating hypermethylation of H3K9 and K27. We identified the Polycomb repressive complexes (PRC1 and 2), but not heterochromatin protein 1 to be necessary factors for silencing. Once established, the compact structure is stably maintained through cell divisions. By contrast, in cells where the silencing is PIWI-dependent, the transgenes display an open and labile chromatin structure. Our data suggest that a post-transcriptional gene silencing (PTGS) mechanism is responsible for the repression in the ovarian somatic cells, whereas a mechanism that couples PTGS to transcriptional gene silencing operates to silence retrotransposons in the other somatic tissues.

Key words: retrotransposons; *Drosophila*; heterochromatin; Polycomb

1. Introduction

Due to their ability to transpose to virtually any genomic site, transposable elements (TEs) have the ability to generate deleterious mutations in the host genome. In response, host cells have devised strategies to control TE activity. RNA interference (RNAi), which has a prominent role in this control, is triggered by small RNAs. Several families of small RNAs have now been reported, but it appears that *D. melanogaster* mainly has two RNA silencing mechanisms that repress transposons expression: endo-siRNAs (endogenous small interfering RNAs) and piRNAs. piRNAs (for PIWI-interacting RNAs) are mainly derived from TE antisense strands and are produced from discrete genomic loci.¹

piRNAs are from 26 to 30 nucleotides (nts) in length and have been reported in the reproductive apparatus of *Drosophila*, mice, rats, and humans.^{2,3} Endo-siRNAs are ~21 nts long and associate with the argonaute protein Ago2. Like piRNAs, endo-siRNAs are mainly derived from retrotransposons and other genomic repetitive elements. However, contrasting with piRNAs, endo-siRNAs are likely to be expressed ubiquitously.^{4–7} It is currently proposed that piRNAs and endo-siRNAs have a germline-specific function and a soma-specific function, respectively, in the establishment of TE silencing.⁴

Besides these post-transcriptional regulations implicating RNAi silencing pathways, TEs can also be transcriptionally silenced. A range of chromatin modifications suppress their transcription, including histone

modifications and altered chromatin packing.⁸ Furthermore, in both plants and mammals, DNA methylation on cytosine residues is another important signal that represses TE transcription and provides a mechanism for inheritance of TE silencing.

We recently reported that PIWI is necessary for silencing of ZAM and *Idefix*, two long terminal repeat retrotransposons from *Drosophila melanogaster*, in the ovarian follicle cells, but is dispensable in other somatic lineages throughout fly development.⁹ However, it remains unclear whether a specific chromatin structure is deposited on these elements and in which tissues. Furthermore, since DNA methylation is absent from *D. melanogaster*, nothing is currently known about the mechanisms involved to maintain their silencing through cell divisions. Here, we investigated further the chromatin structures necessary to silence *Idefix*.

Like most transposable families, many copies of *Idefix* are detected within the *Drosophila* genomes. Active copies are generally dispersed on chromosomal arms but many copies corresponding to *Idefix* vestiges also exist and accumulate in the heterochromatic regions. It was then of prime importance to follow the silencing exerted on a copy known to a target of the RNAi pathways. To this aim, we made use of sensor transgenes carrying a GFP (green fluorescent protein) reporter gene followed by an *Idefix* fragment that recapitulates the control exerted on active copies of *Idefix*.⁹ Analysis of the chromatin structure deposited on the transgene could then be easily explored over the GFP sequence that is unique in the genome.

We show that in somatic cells in contact with the germline, the silencing needs to constantly target the *Idefix* sequence to be maintained. By contrast, in the other somatic tissues referred to as the soma, the silencing is stably established even when the TE fragment is excised. Our data further indicate that a compact chromatin structure likely associated with Polycomb (PC)-dependent chromatin structures is deposited.

2. Materials and methods

2.1. *Drosophila* strains and transgenic lines

The following mutant fly stocks were used: ago2⁴¹⁴,¹⁰ ago2^{51B},¹¹ dcr2^[L811fsX],¹² PIWI² and PIWI³,¹³ r2d2¹,¹⁴ and loqs^[f00791].¹⁵ Bl279 (+;

+/+; MKRS hs-Flip/Tm6) comes from Bloomington stock. pc¹ comes from Cavalli's lab.

About 419 bp homologous to the *Idefix gag* coding region (1003–1422) were inserted either in a sense (tGglds) or anti-sense orientation (tGgldas) according to *gfp* transcription within the pUAST-*gfp* vector. Three and six independent transgenic lines were analyzed for tGglds and tGgldas, respectively (Table 1).

2.2. Chromatin immunoprecipitation assays

Chromatin immunoprecipitation (ChIP) was performed using the standard procedures, which was modified as follows: for each immunoprecipitation, three OD₂₆₀ of chromatin in ChIP buffer was pre-incubated in the presence of 80 µl of Protein A-Agarose (PAA) beads (Millipore) for 30 min at 4°C. PAA were removed, antibodies were added (a control in the absence of antibody named Input was included), and samples were incubated overnight at 4°C in a rotating wheel. Then, 60 µl of PAA were added and incubation was continued for 1 h at 4°C. Samples were washed in saline concentration buffer. Chromatin was eluted from PAA in 500 µl of elution buffer (sodium dodecyl sulfate and NaHCO₃) at room temperature for 15 min. The eluate was incubated overnight at 65°C to reverse cross-links and treated by proteinase K for 1 h at 50°C. Samples were phenol–chloroform extracted and ethanol precipitated. DNA was resuspended in 40 µl of 10 mM Tris (pH 7.5).

qPCR (quantitative polymerase chain reaction) (ABI 7300) from at least two biological-independent extractions and three technical replicate was performed using SYBR Green (Sigma) and Platinum Taq DNA Polymerase (Invitrogen). Enrichment was calculated relative to *RpL32* and values were normalized to input measurements; the error is indicated by the standard error of mean (SEM). Antibodies used include: H3K9-di-methylation #07-441, H3K9-tri-methylation #07-523, H3K9/K14-acetylation #06-599, H3K27-di-methylation #06-421, H3K27-tri-methylation #06-449, H4K20-mono-methylation #07-440, H2A-ubi #05-678 from Millipore, heterochromatin protein 1a (HP1a) (C1A9) from DSHB, and PC and PH antibodies were provided by G. Cavalli Lab. The following primers were used uptssF: CCAAGCTTTGCGTACTCG and uptssR: CCGTGGGGTTTGAATTAAC for the upstream *gfp* TSS; gfpF: ACCATTACCTGTCCACACAA and gfpR:

Table 1. Characteristics of tGglds and tGgldas lines in somatic tissues and their homologous piRNA and endo-siRNAs (esiRNAs) reported in the databases^{4,31}

	Number of lines	Silencing in follicle cells	Silencing in larvae	Number of sense piRNAs	Number of antisense piRNAs	Number of sense esiRNAs	Number of antisense esiRNAs
tGgld sense	3	On	Off	0	194	0	0
tGgld antisense	6	On	On	34	0	3	0

CCAGCAGCTGTTACAACTC for the *gfp* gene; and Rpl32F: CCGCTTCAAGGGACAGTATC and Rpl32R: GACAATCTCCTTGCGCTTCT for the control gene *Rpl32*.

2.3. Fluorescent staining and microscopy

Ovaries were dissected and fixed in 5% formaldehyde in phosphate buffer solution for 15 min each hour after the heat shock. GFP was viewed in whole-mount ovaries using the 488 nm filter set of the ZEISS LSM 510 confocal microscope.

2.4. In situ hybridization

FISH (fluorescence in situ hybridization) was performed on larval imaginal discs, as described previously.¹⁶ Detailed coordinates of the PCR fragments used to produce the probes can be provided upon request. Three-dimensional images were acquired on a Leica SP5 confocal microscope using a 63X objective. The colocalization was searched for automatically by using a script written on Imaris XT software (Supplementary Fig. S2). Three hundred and five and 179 nuclei were examined from seven tGgldas and seven tGg^{Dld} larval discs, respectively.

3. Results

3.1. *Idefix* sequence is continuously required for its silencing in the follicle cells but not in other somatic tissues

To characterize further the silencing targeting *Idefix* in the somatic tissues, we used a sensor transgene called tGgldas. In this transgene, the *gfp* reporter gene driven by the UAS promoter was linked to a fragment of the *gag* gene of *Idefix* (Fig. 1A).⁹ When tGgldas is driven by a ubiquitous driver, Actine-Gal4, a GFP signal is detected neither in the follicle cells nor in other somatic tissues throughout development: larvae, pupae, and adults (Fig. 1B). When *Idefix* is excised and a new line established with the sensor transgene lacking *Idefix* sequence, then GFP is fully recovered (see tGg^{Dld} line, Fig. 1B). Since tGgldas is repressed in all the somatic tissues within and outside of the ovaries because of the *Idefix* sequence, we used it to test whether its silencing might switch to an active state as soon as the targeted fragment of *Idefix* has been excised.

We made use of the *Flp*/FRT (Flip/FLP recognition target) system and its selective induction by the heat shock *Flp* (HS-*Flp*) driver. Heat shocks were performed for 1 h on third instar larvae. If tGgldas silencing is stably established in somatic tissues, GFP repression should persist through cell divisions even after the *Idefix* fragment has been flipped out. Alternatively, if the silencing is labile and depends

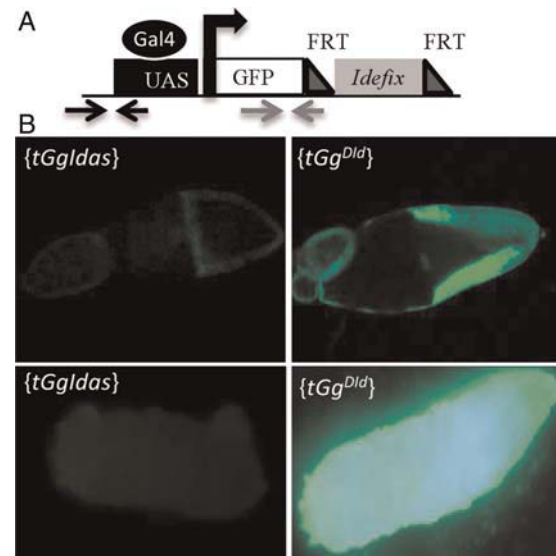


Figure 1. Structure and silencing of tGgldas. (A) Structure of tGgldas: the *gfp* reporter gene is depicted by a white rectangle and its TSS by an arrow. The *Idefix* portion (grey) are inserted between two FRT sites (dark grey triangles). Position of the qPCR primer sets taken in the *gfp* sequence and upstream of the TSS are indicated by arrows. (B) An example of GFP silencing targeting tGgldas in the follicle cells (above) and in larvae (below) is presented on the left panels; expression of the reference line carrying a sensor transgene, tGg^{Dld}, in which the *Idefix* fragment has been excised, is presented on the right.

on a regulatory pathway that needs to constantly target *Idefix* sequence to establish the repression, then the silencing should be lost and the GFP expression recovered during the life of the fly as soon as the *Idefix* sequence is flipped out (Fig. 2A).

In a first set of experiments, we determined the time necessary for the FLP recombinase to excise a DNA fragment flanked by FRT sequences after induction by heat shock. We used a triple transgenic line designated as CF (for control *Flp*). A first transgene is the FLP recombinase under the control of the heat shock promoter (HS-*Flp*). A second transgene (Actin>FRT-*yellow*-FRT>Gal4) carries the ubiquitous actin promoter upstream of a *yellow* gene flanked by FRT sequences and a Gal4 reporter gene placed downstream. When *yellow* is present, Gal4 is not expressed. The third transgene of the CF line carries the UAS promoter placed upstream of the *gfp* reporter gene. If a heat shock is performed in this triple transgenic line, the *yellow* gene flanked by FRT sites is excised. The Gal4 protein is then expressed, and in turn, activates the UAS-*gfp* transgene giving rise to fluorescence. A kinetic of GFP appearance in both the follicle and larval cells at time intervals of 1 h after heat shock was performed, and the time necessary to recover GFP expression was determined in both tissues. We estimated that 15 and 24 h was the beginning of expression of the UAS-*gfp* reporter gene in the follicle cells and the larval tissues, respectively, and a complete expression was recovered after

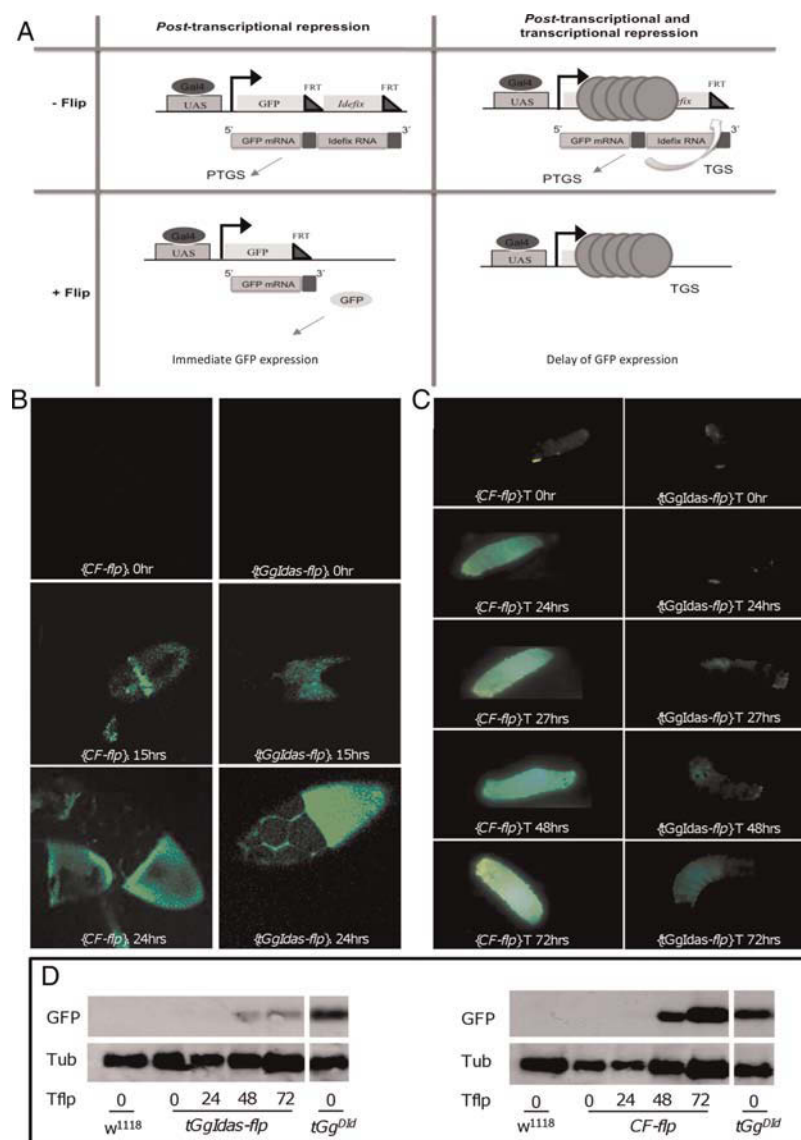


Figure 2. The silencing exerted on tGldas is lost in the follicle cells as soon as the targeted retrotransposon sequences are excised, but remains in the soma. (A) Tested hypothesis: (left) if tGldas silencing is labile and depends on a regulatory pathway that needs to constantly target Idefix sequence to establish the repression, then the silencing should be lost and the GFP expression recovered even during the life of the fly as soon as the Idefix sequences is flipped out; (right) if the silencing is stably established in tissues due to a compact heterochromatic structure able to maintain a cellular memory, GFP repression should persist through cell divisions even after the Idefix fragment has been flipped out. (B) (left column) When the CF line [*HS-Flp*, *Actin*>*FRT*>*yellow*>*FRT*>*Gal4*, *UAS-gfp*] is submitted to a heat shock treatment, GFP expression is detected in the follicle cells as soon as 15 h after treatment (second panel) and a full expression at 24 h (third panel). (right column) When the tGldas line is submitted to the heat shock, GFP is detected as soon as 15 h after treatment and a full expression at 24 h compared with tGg^{Did} (first panel). (C) (left column) When CF larvae are submitted to a heat shock treatment, GFP expression is detected 24 h after treatment; GFP is fully recovered after 48 h (fifth panel). (right column) After heat shock, GFP expression remains faint in tGldas larval tissues even 72 h after heat shock. (D) Western blot analysis for GFP expression in larvae submitted to heat shock kinetics. Western blotting was performed with an antibody that specifically detects GFP. The *w¹¹¹⁸* line with no *gfp* transgene is used as a negative control and tGg^{Did} is used as a positive control for GFP expression. The transgenic lines tGldas (left panel) and the test line CF (right panel) are analyzed at time intervals presented below the figure: 0, 24, 48 and 72 h. Tub is a loading control.

24 h in the follicle cells and 48 h in larvae (Fig. 2B and C, left columns).

Then, we determined how long the silencing of the tGldas transgene persists after *Idefix* has been flipped out. A heat shock was performed on flies or larvae with the genetic backgrounds [*Actin*-*Gal4*/+; tGldas/*HS-Flp*] or [*Actin*-*Gal4*/+; tGldas/TM6].

Ovaries were dissected every hour from the heat-shocked flies and GFP expression examined in the follicle cells. We found that fluorescence was detected in some patches 15 h after the heat shock (Fig. 2B, right column, middle panel) and a strong green signal in the entire follicular epithelium was recovered after 24 h (Fig. 2B, right

column, lower panel). These data indicate that the repression exerted in the follicle cells is reversible as soon as the targeted retrotransposon sequence is excised.

When examined in larvae, GFP expression remained undetectable 24 s after heat shock induction in flies [Actin-Gal4/+; tGldas/HS-Flp]. Fluorescence was hardly detected 48 h after treatment and remained very faint even after 72 h, indicating that a full expression was not recovered (Fig. 2C, right column).

As controls, we verified first that the *Idefix* fragment has been indeed excised after the heat shock treatment (PCR amplifications not shown), and second, that GFP expression is not due to the heat shock treatment, but indeed to *Idefix* excision since no GFP was detected in cells with the [Actin-Gal4/+; tGldas/TM6] genetic background after heat shock.

Then, proteins were extracted from the treated larvae at different time intervals from 0 to 72 h and western blot experiments were performed with an antibody raised against GFP. The reference line established with the tGg^{Did} transgene with no *Idefix* sequence and the w¹¹¹⁸ line with no *gfp* transgene were used as positive and negative controls of GFP expression, respectively. In accordance with results described above, western blots revealed that in [Actin/+; tGldas/hs-flp] larval tissues (denoted tGldas-flp on Fig. 2D), the level of GFP was

hardly detected by 48 or 72 h as opposed to the level of GFP expressed from the CF line 48 or 72 h after heat shock (Fig. 2D, left and right panels, respectively).

Overall, experiments performed on larval tissues indicate that in these somatic tissues, the silencing persists even after the *Idefix* sequence has been excised what contrasts with the silencing exerted in follicle cells.

3.2. Histone methylation marks cover silenced transgenes in somatic tissues outside ovaries

We asked whether specific chromatin structures could mark tGldas within and/or outside of the ovaries. In a first set of experiments, post-translational histone modifications including lysine acetylation and methylation were analyzed. Whereas acetylation is generally linked with gene activity, methylation of histone H3 on Lys9 (H3K9) and Lys27 (H3K27) correlates with gene silencing. We performed ChIP assays using antibodies against H3K9/K14 acetylation, H3K9me3, H3K27me2, H3K27me3, and H4K20me1, followed by a real-time PCR performed with a primer set targeting *gfp* from both tGldas and tGg^{Did} transgenes (Fig. 1A).

In the ovaries, the H3K9me3, H3K27me2, or H3K27me3 marks were found absent from tGldas

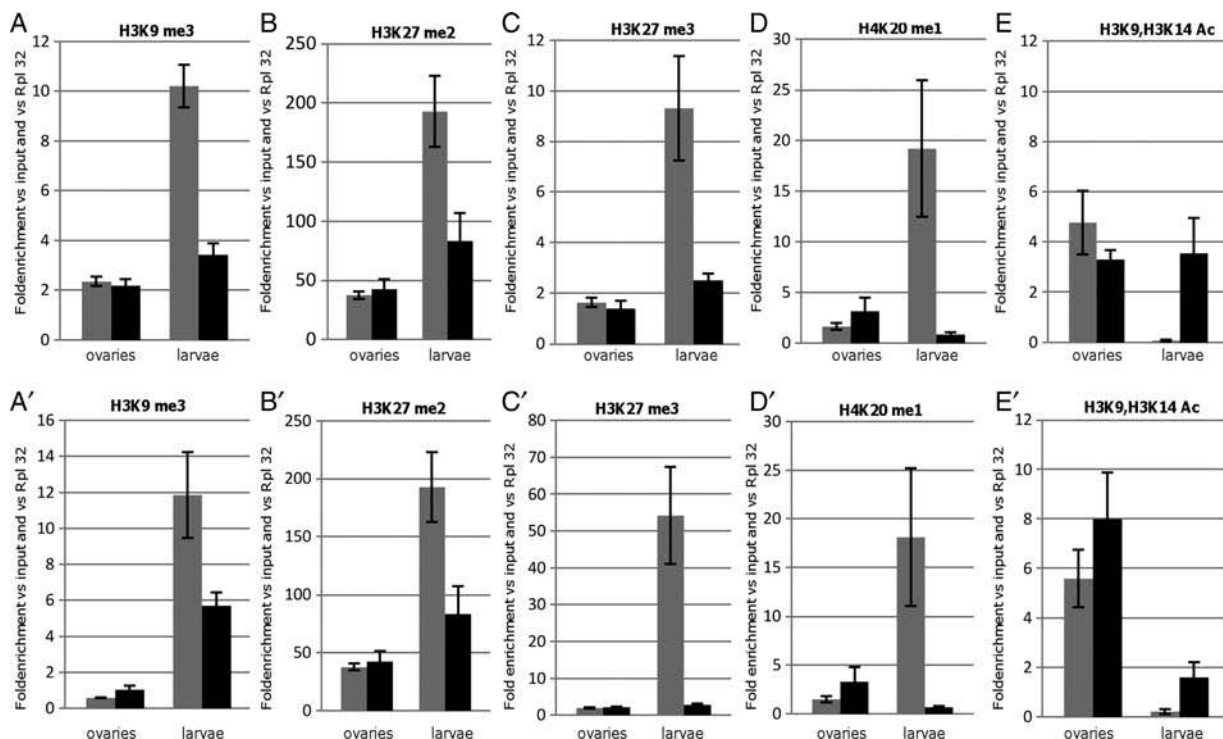


Figure 3. Histone post-translational modifications associated with the silenced sensor transgene in ovaries and larvae: the histograms show relative amounts of H3K9me3 (A and A'), H3K27me2 (B and B'), H3K27me3 (C and C'), H4K20me1 (D and D'), and H3K9/H3K14Ac (E and E'). Each panel displays the results obtained from ovaries (left) and larvae (right) of transgenic lines tGldas (gray bar) and tGg^{Did} (black bar). ChIP was performed on the *gfp* reporter gene (A–D) and upstream of its TTS (A'–D') (see primer sets in Materials and methods). Measurements were normalized to the *Rpl32* gene. Error bars, SEM.

and tGg^{Dld} transgenes (Fig. 3A–C). As expected, when the repressive marks are absent, both of these transgenes displayed an enrichment of H3K9 acetylation (Fig. 3E).

By contrast, a very strong enrichment of the methylated marks was detected on tGgldas in larval tissues (Fig. 3A–C). These repressive chromatin hallmarks depend on the presence of *Idefix* because they are enriched >3–10-fold when the *Idefix* sequence is present compared with tGg^{Dld}. They also depend on tGgldas transcription since they are found absent when it is not activated by the Gal4 driver (data not shown). In contrast, histone marks such as H3K9 acetylation were found completely absent from tGgldas compared with tGg^{Dld} (Fig. 3E).

We further verified that the H4K20me1 mark, generally associated with transcriptional repression and chromosomal condensation, is also enriched on tGgldas when compared with tGg^{Dld} in larval tissues, whereas it is absent in ovaries (Fig. 3D). Finally, we also analyzed the H3K9me2 modification and found no enrichment on tGgldas compared with tGg^{Dld} (Supplementary Fig. S1).

Given the evidence that heterochromatin structures can spread along the chromatin fiber, we wondered whether histone marks associated with the sensor transgenes in the larval tissues could be detected

outside of the transcription start site (TSS). We performed ChIP experiments using primers spanning a region from –440 to –330 bp upstream of the TSS (Fig. 1A). We observed similar results as those obtained with primer targeting *gfp*. On one hand, there is an enrichment of H3K9me3, H3K27me2, H3K27me3, or H4K20me1 on tGgldas in the larval tissues. On the other hand, these repressive marks were found absent from the sensor transgene in the ovaries, whereas an H3K9 acetylation similar to tGg^{Dld} was observed (Fig. 4A'–E').

We conclude that distinct chromatin structures cover tGgldas in larval versus ovarian tissues. Post-translational histone modifications generally linked to repression are only observed in non ovarian somatic tissues.

3.3. PC binding, and not HP1, correlates with histone methylation at silenced transgenes

HP1 has been implicated in TE silencing in ovaries.¹⁷ Thus, we were interested to test whether HP1 could be detected on the silenced transgene either in ovaries or larvae. ChIP experiments revealed that tGgldas could not be immunoprecipitated with an anti-HP1 antibody at levels above background (Fig. 4A and A'). ChIP experiments performed on a region targeted by HP1 (the

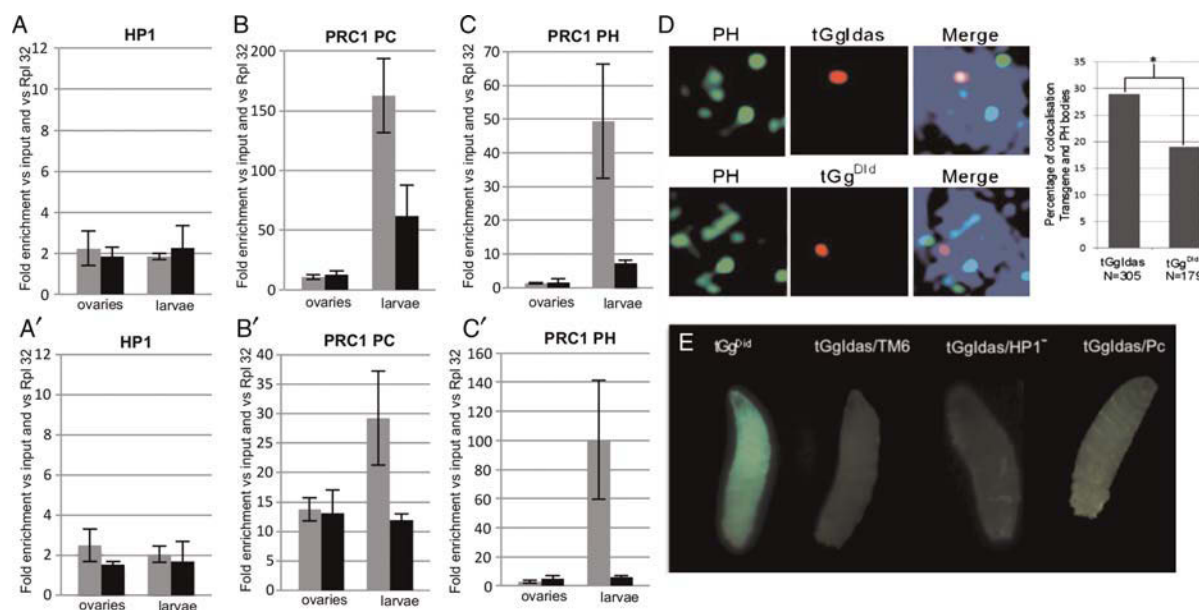


Figure 4. PC and PH, not HP1, are required to silence the sensor transgene in larvae: the histograms show relative amounts of HP1 in A and A', the PC protein in B and B', and PH protein in C and C'. Each panel displays the results obtained from ovaries (left) and larvae (right) of transgenic lines tGgldas (gray bar) and tGg^{Dld} (black bar). ChIP was performed on the *gfp* reporter gene (A–C) and upstream of its TSS (A'–C'). Measurements were normalized to the *Rpl32* gene. Error bars, SEM. (D) The silenced tGgldas transgenes colocalize with Pc-G bodies: characteristic individual nuclei analyzed by FISH-I in larval imaginal discs: Hoechst staining, the Pc-G bodies, the transgene insertion site and the merge of the three channels are shown. The scale bar represents 1 μ m. On the right, quantification of the percentage of nuclei in which the transgene, either tGgldas or tGg^{Dld}, is present in a Pc-G body. Genotype and the total number of nuclei analyzed are indicated below each bar. Asterisk denotes that the colocalization with PH was statistically different between tGg^{Dld} and tGgldas ($P < 0.02$). (E) GFP expression from tGgldas in an heterozygous *Su(var)205*⁵ (noted HP1⁻) and *pc*¹ mutant background (last two larvae on the right) compared with a non-silenced transgene tGg^{Dld} when *Idefix* is excised from tGgldas (first larvae on the left) and to a strong silenced transgene tGgldas in a Pc and HP1 wild-type background (second larvae from the left).

heterochromatic *flam/COM* locus) was used as a positive control for the ChIP experiment (Supplementary Fig. S1). These results indicate that HP1 does not cover tGldas in somatic cells.

Evidence of H3K27 methylation in larval tissues suggested that Polycomb group (Pc-G) proteins might be implicated in the transgene silencing. Indeed, the methylation of H3K27 is a hallmark of the silencing mediated by PRC2, since it is carried out by the Histone Methyl Transferase Enhancer of Zeste, E(z), the catalytic subunit of PRC2. The chromo-domain of the *Drosophila* PC protein from the PC repressive complex 1 (PRC1) binds to the H3 tail peptide trimethylated at K9 and more strongly at K27me2/3. This prompted us to investigate Pc-G proteins occupancy. ChIP assays were performed with antibodies against PC and Polyhomeotic (PH). Again, two distinct patterns of enrichment were observed. A very strong enrichment of both these proteins was detected on the *gfp* gene in larval tissues, whereas neither PC nor PH proteins could be detected on the *gfp* gene in the ovarian follicle cells (Fig. 4B and C).

Given the evidence that heterochromatin structures can spread along the chromatin fiber, ChIP assays were performed using primers taken upstream of the TSS of tGldas (Fig. 1). Results indicated that an enrichment of PC and PH was also detected in larval tissues outside of the TSS (Fig. 4B' and C'). A comparably low amount of PC and PH marks was detected on tGldas and tGg^{Dld} in the follicle cells (Fig. 4B' and C').

To further verify that the binding of Pc-G protein is indeed a hallmark of the chromatin structure associated with the repressed transgene in the somatic tissues outside of the ovaries, we performed three additional tests. First, we investigated whether a colocalization of Pc-G bodies and the targeted transgene might be visualized in larval imaginal discs. A combination of 3-D FISH and immunostaining technique (FISH-I) was used, allowing the detection of the relative localization of the transgene with Pc-G bodies visualized using an anti-PH antibody. The tGldas transgene used in this study is integrated on chromosome 3R at the cytological position 25591560, 937 bp upstream of the *Kay* gene. This site is not a Pc-G binding site reported by Negre *et al.*¹⁸ We found that the colocalizations of tGg^{Dld} on one side, and tGldas on the other, with Pc-G bodies are statistically different, 19 versus 29%, respectively ($P < 0.02$; Fig. 4D). Thus, insertion of tGldas results in a higher incidence of this locus to colocalize with Pc-G bodies than the same transgene without the *Idefix* sequence. Second, we performed similar ChIP assays on tissues extracted from adult carcasses and verified that indeed the same enrichment in Pc-G proteins is observed in these somatic tissues (Fig. 5A) compared with a low H3K9 acetylation (Fig. 5B). Third, we

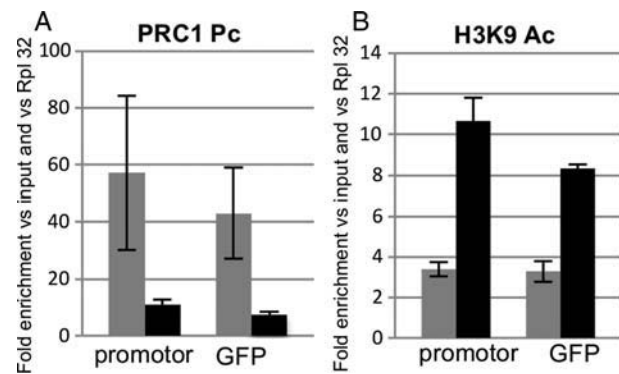


Figure 5. H3K9 acetylation and PC association with the silenced sensor transgene in adults: the histograms show relative amounts of PC (A) and H3K9ac (B) from carcasses of the transgenic lines tGldas (gray bar) and tGgDld (black bar). Each panel displays the results obtained when ChIPs were performed with primer sets taken at the GFP reporter gene (right) and upstream of the TSS (left) (see primer sets in Materials and methods). Measurements were normalized to the *Rpl32* gene. Error bars, SEM.

verified that H2A ubiquitination of lysine 119, which is performed by the PRC1 E3 ubiquitin ligase Sce/RING, is also a histone mark linked to PC silencing of tGldas^{19,20} (Supplementary Fig. S1).

Finally, to definitively implicate Pc-G protein in the silencing, we aim at analyzing the effect of PC mutant on the silencing exerted on tGldas. However, described mutants of PC are recessive lethals. Moreover, a clonal analysis was not possible due to the FRT sequences present in the transgene. Nevertheless, since *Pc* alleles are found regularly but with low expressivity as dominant phenotypes, transgenic flies tGldas were crossed to *pc*¹/TM6 mutant backgrounds. GFP expression was examined in the larval progeny [tGldas/+; *pc*¹/+] and compared with [tGg^{Dld}/+; +/+]⁺ larvae as a control for GFP expression when the *Idefix* sequences has been flipped out and to [tGldas/+; +/+]⁺ as a positive control for GFP silencing. As exemplified Fig. 4E, a faint but significant expression of GFP was recovered in the *Pc* heterozygous mutant background (Fig. 4E, right). Finally, to verify that HP1 is not implicated in the silencing of tGldas, we examined GFP expression in larvae heterozygous for an HP1 mutant allele, [tGldas/+; Su(var)205⁵/+]. No GFP expression was recovered in these larvae (Fig. 4E).

Overall, the data indicate that a repressive structure implicating Pc-G protein is associated with the sensor transgene tGldas in somatic tissues outside of the ovaries.

3.4. Which signal is required in the soma to target Pc-G proteins to the sensor transgene?

We wondered whether the silencing observed in the somatic cells has the capacity to discriminate one

strand of the sensor transgene from the other. To this end, we established new transgenic lines carrying a transgene, tGglds, similar to tGgldas except that the same *Idefix* fragment has been inserted in a sense orientation. Three independent transgenic lines were established. When the expression of tGglds and tGgldas was examined in the follicle cells where the Piwi pathway has been clearly implicated, we found that both of them were repressed in this tissue (Fig. 6A). However, we found that, unlike tGgldas, tGglds is not targeted to silencing in the soma where a full GFP expression is observed (Fig. 6A) and it is not enriched in repressive chromatin hallmarks such as H3K27me3 as illustrated in Supplementary Fig. S1.

This result indicated that the silencing mechanism involved in the soma is likely to be an RNA silencing pathway able to discriminate between the two mRNA strands produced by the transgenes. Since the silencing of TEs is presumed to be directed by the endo-siRNA pathway in the soma, we investigated whether proteins involved in the endo-siRNA pathway were necessary actors for the silencing exerted on tGgldas. Therefore, we compared GFP expression in homozygous mutant and wild-type flies for *dcr2*, *ago2*, *loq*, and *R2D2*. We found no release of GFP silencing in either of these mutant backgrounds (Fig. 6B).

4. Discussion

In this study, we tested whether the silencing targeting a TE sequence in somatic lineages constantly

needs the presence of the target or alternatively, once established, the silencing is maintained even though the targeted sequence is lost. We further correlated differences observed in the ovarian follicle cells versus other somatic tissues with differences in the chromatin structures established on the target.

We found that, *in vivo*, the *Idefix* sequence is continuously required for its ongoing repression in the follicle cells. Its repression is not heritably maintained through cell divisions. Accordingly, neither the hallmarks of histone post-translational modifications that generally correlate with compact chromatin structure, nor HP1 or Pc-G proteins are linked to the silenced sensor transgenes. This result argues against the current model in which piRNA-PIWI protein complexes silence transposons in ovarian tissues by directing assembly of heterochromatin-like domains.¹⁷ Interestingly, using ovarian somatic stem cells transfected with double-strand RNAs corresponding to Piwi, Haase *et al.*²¹ reported recently that the integrity of the piRNA pathway is essential for the ongoing repression of mobile elements. Consistent with our results, they further provided evidence that, once set by the action of Piwi proteins on chromatin, a transposon silencing cannot be autonomously maintained.

Considered together, these studies promote a model in which, in the follicle cells, TEs are submitted to a post-transcriptional gene silencing (PTGS) which is not coupled to an epigenetic silencing state implicating an HP1-dependent heterochromatin structure.

By contrast, the assembly of a compact chromatin structure marked by the methylation of H3K9 and

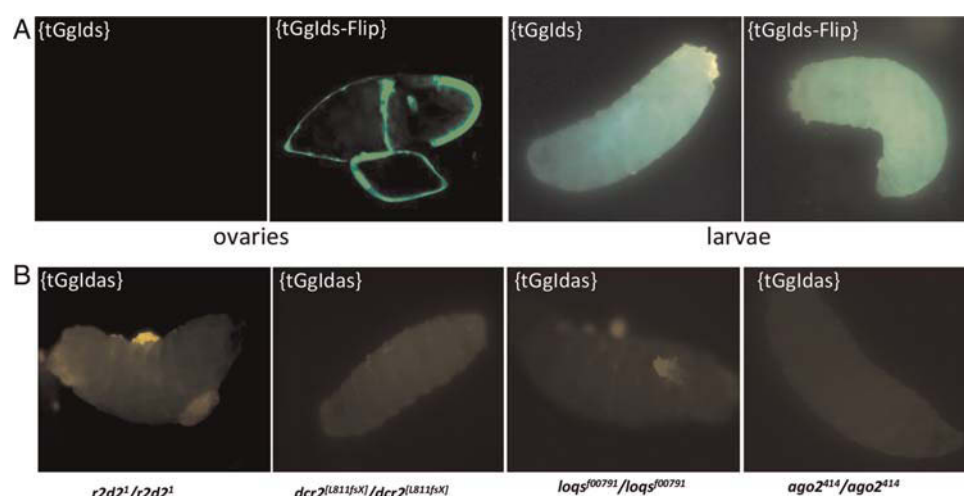


Figure 6. Characteristics of the silencing targeting tGglds (A) and tGgldas (B) transgenes. (A) The silencing targeting the sensor transgene tGglds is active in the ovaries and inactive in the larvae (first and third panels from the left, respectively). Expression of tGglds is driven by the ubiquitous actin-Gal4 driver. Expression of the transgene after flp-recombinase action (*tGglds-Flip*) is presented as a positive expression of the GFP reporter gene in the second and last panels. (B) Mutations of genes involved in the endo-siRNA pathway do not release the silencing exerted on the tGgldas sensor transgene. GFP expression is never recovered in larvae from homozygous mutants for *r2d2*, *dcr2*, *loq*, and *ago2*.

H3K27 targets TEs in the somatic cells outside of the ovaries. Furthermore, our study identifies Pc-G proteins as major actors. The two PRCs, PRC1 and PRC2, are involved in this structure seeing that H3K27me3 mediated by the histone methyl transferase E(z), belonging to PRC2, is detected as well as PC and PH proteins which belong to PRC1. We further demonstrate that this compact chromatin structure has two main consequences on the targeted locus. First, it can spread along the chromatin fiber and invade the genome upstream of the TSS of the sensor transgene. Second, once established, repression of the sensor transgene is then stably maintained throughout cell divisions and does not need the TE sequence to constantly direct the silencing.

Exploiting the mechanism of PC-mediated silencing to repress endogenous retroelements has also been described by several studies performed on mice.^{22,23} Furthermore, in a recent study, the E(z) homolog, EZH, in the green algae *Chlamydomonas reinhardtii*, has been implicated in the repression of both (hetero)euchromatic transgenes and dispersed retrotransposons.²⁴ Therefore, consistent with the present study, all these data point PRC1 and PRC2 as evolutionarily conserved epigenetic factors employed to silence TE and convey the silencing from one generation of cells to the next.

Interestingly, our study further points out that HP1 is not a partner of the chromatin structure associated with tGldas either in the ovaries or in the soma. Although HP1 has been frequently reported as associated with TE,^{17,25} our data are in line with recent data published by the *Drosophila* modENCODE project.²⁶ This project has generated data sets that profile histone modifications, chromosomal proteins, transcription factors, and nucleosome properties across a developmental time course and in multiple cell lines. Their data show that only TE located in heterochromatin are marked by HP1. By contrast, TE inserted in euchromatic regions are not covered by HP1 as found for tGldas. Thus, one can assume that HP1 might be a signature of TE vestiges because of their heterochromatic localization, whereas HP1-independent silencing pathways might act within euchromatic regions to specifically silence full-length TE and prevent their mobilization.

Overall, if our data implicate PRC1 and PRC2 as epigenetic factors employed to silence TE, they fail to explain how the TE sequence is recognized. In our effort to identify the effector mechanism directing Pc-G proteins on the targeted sensor transgene, we found that the silencing machinery active in the soma has the capacity to discriminate between both orientations of the *Idex* sequence present within the sensor transgene. Indeed, tGldas is targeted to silencing, whereas tGlds is not. This result is

consistent with a recognition occurring at the RNA, and not the DNA, level.

However, PRC1 and PRC2 employed in this silencing do not have specificity for only one strand and not the other. Therefore, one has to assume that their binding occurs under the presence of a primary signal which brings the sequence specificity. In our study, we were unable to identify this signal. Kanhere *et al.*²⁷ reported that short RNA from 50 to 200 nts transcribed from the 5' ends of PC target genes play a role in the association of PRC2. Such short RNAs complementary to *Idex* sequences could potentially explain the targeting. Another likely possibility is that the recognition might occur via homologous siRNAs produced by the cell. Epigenetic complexes made of small RNAs together with PRC1 or PRC2 might potentially explain their recruitment to homologous genomic sequences.^{28–30}

Given the evidence that TEs are mainly silenced in the soma via the endo-siRNA pathway, genetic mutants affecting these pathways should have resulted in a loss of the sensor transgene repression. However, no variation in the GFP expression could be detected in these mutant backgrounds. A plausible explanation is that, once established on the transgene, the silent epigenetic state cannot be released even though the targeted fragment is excised or the silencing pathway necessary for the target recognition mutated. Nevertheless, our study pointed out that, like in the follicle cells where piRNAs are involved (Table 1),^{9,31} the primary effectors acting in the soma are likely to be single-stranded molecules since only one strand of the mRNA produced by the sensor transgene is targeted. This is comforted by the list of endo-siRNAs reported in the databases (Table 1). Endo-siRNAs able to target and silence tGldas are reported in the databases, whereas none is complementary to the *Idex* fragment inserted in tGlds.

Overall, in the soma, we propose that the silencing of TE is initiated by a PTGS mechanism which is successively coupled to transcriptional gene silencing. This latter displays all the molecular imprint underlying cell memory and epigenetic inheritance. This epigenetic state, absent in the ovarian follicle cells, accounts for the molecular mechanism controlling maintenance of TE silencing during development.

Acknowledgements: We are grateful to F. Bantignies for helpful discussions and PC and PH antibodies. We thank S. Ronsseray for mutant flies. We thank P. George and C. Fritsch for critical review of the manuscript. We thank F. Pellissier for technical assistance and Flyfacility for transgenic lines (www.fly-facility.com).

Supplementary data: Supplementary data are available at www.dnaresearch.oxfordjournals.org.

Funding

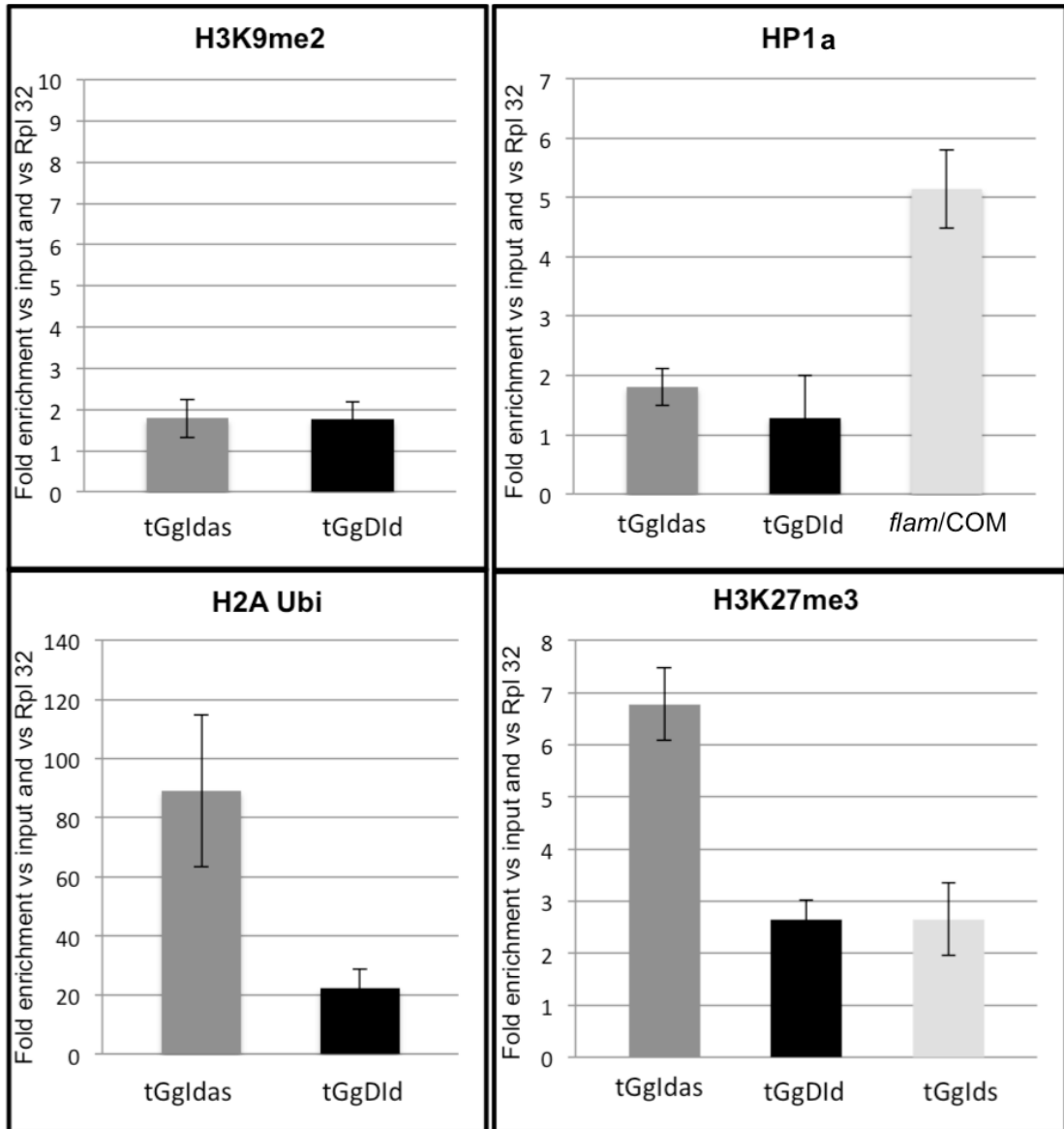
This work was supported by grants from the Association pour la Recherche contre le Cancer (ARC 1139). J.D. received a graduate grant from the Ministère de l'Enseignement Supérieur et de la Recherche (MESR). E.B. and P.P. received a grant from the Région Auvergne.

References

1. Brennecke, J., Aravin, A.A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., et al. 2007, Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*, *Cell*, **128**, 1089–103.
2. Aravin, A., Gaidatzis, D., Pfeffer, S., Lagos-Quintana, M., Landgraf, P., Iovino, N., et al. 2006, A novel class of small RNAs bind to MILI protein in mouse testes, *Nature*, **442**, 203–7.
3. Girard, A., Sachidanandam, R., Hannon, G.J. and Carmell, M.A. 2006, A germline-specific class of small RNAs binds mammalian Piwi proteins, *Nature*, **442**, 199–202.
4. Ghildiyal, M., Seitz, H., Horwich, M.D., Li, C., Du, T., Lee, S., et al. 2008, Endogenous siRNAs derived from transposons and mRNAs in *Drosophila* somatic cells, *Science*, **320**, 1077–81.
5. Chung, W.J., Okamura, K., Martin, R. and Lai, E.C. 2008, Endogenous RNA interference provides a somatic defense against *Drosophila* transposons, *Curr. Biol.*, **18**, 795–802.
6. Czech, B., Malone, C.D., Zhou, R., Stark, A., Schlingeheyde, C., Dus, M., et al. 2008, An endogenous small interfering RNA pathway in *Drosophila*, *Nature*, **453**, 798–802.
7. Kawamura, Y., Saito, K., Kin, T., Ono, Y., Asai, K., Sunohara, T., et al. 2008, *Drosophila* endogenous small RNAs bind to argonaute 2 in somatic cells, *Nature*, **453**, 793–97.
8. Slotkin, R.K. and Martienssen, R. 2007, Transposable elements and the epigenetic regulation of the genome, *Nat. Rev. Genet.*, **8**, 272–85.
9. Desset, S., Buchon, N., Meignin, C., Coiffet, M. and Vaury, C. 2008, In *Drosophila melanogaster* the COM locus directs the somatic silencing of two retrotransposons through both Piwi-dependent and -independent pathways, *PLoS One*, **3**, e1526.
10. Saito, K., Nishida, K.M., Mori, T., Kawamura, Y., Miyoshi, K., Nagami, T., et al. 2006, Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the *Drosophila* genome, *Genes Dev.*, **20**, 2214–22.
11. Xu, K., Bogert, B.A., Li, W., Su, K., Lee, A. and Gao, F.B. 2004, The fragile X-related gene affects the crawling behavior of *Drosophila* larvae by regulating the mRNA level of the DEG/ENAC protein pickpocket1, *Curr. Biol.*, **14**, 1025–34.
12. Lee, Y.S., Nakahara, K., Pham, J.W., Kim, K., He, Z., Sontheimer, E.J., et al. 2004, Distinct roles for *Drosophila* dicer-1 and dicer-2 in the siRNA/miRNA silencing pathways, *Cell*, **117**, 69–81.
13. Lin, H. and 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.
14. Liu, Q., Rand, T.A., Kalidas, S., Du, F., Kim, H.E., Smith, D.P., et al. 2003, R2D2, a bridge between the initiation and effector steps of the *Drosophila* RNAi pathway, *Science*, **301**, 1921–25.
15. Forstemann, K., Tomari, Y., Du, T., Vagin, V.V., Denli, A.M., Bratu, D.P., et al. 2005, Normal microRNA maturation and germ-line stem cell maintenance requires loquacious, a double-stranded RNA-binding domain protein, *PLoS Biol.*, **3**, e236.
16. Brasset, E., Bantignies, F., Court, F., Cheresiz, S., Conte, C. and Vaury, C. 2007, Idefix insulator activity can be modulated by nearby regulatory elements, *Nucleic Acids Res.*, **35**, 2661–70.
17. Klenov, M.S., Lavrov, S.A., Stolyarenko, A.D., Ryazansky, S.S., Aravin, A.A., Tuschl, T., et al. 2007, Repeat-associated siRNAs cause chromatin silencing of retrotransposons in the *Drosophila melanogaster* germline, *Nucleic Acids Res.*, **35**, 5430–38.
18. Negre, N., Hennetin, J., Sun, L.V., Lavrov, S., Bellis, M., White, K.P., et al. 2006, Chromosomal distribution of PcG proteins during *Drosophila* development, *PLoS Biol.*, **4**, e170.
19. Fritsch, C., Beuchle, D. and Muller, J. 2003, Molecular and genetic analysis of the Polycomb group gene sex combs extra/ring in *Drosophila*, *Mech. Dev.*, **120**, 949–54.
20. Wang, H., Wang, L., Erdjument-Bromage, H., Vidal, M., Tempst, P., Jones, R.S., et al. 2004, Role of histone H2A ubiquitination in Polycomb silencing, *Nature*, **431**, 873–78.
21. Haase, A.D., Fenoglio, S., Muerdter, F., Guzzardo, P.M., Czech, B., Pappin, D.J., et al. 2010, Probing the initiation and effector phases of the somatic piRNA pathway in *Drosophila*, *Genes Dev.*, **24**, 2499–2504.
22. Golding, M.C., Zhang, L. and Mann, M.R. 2010, Multiple epigenetic modifiers induce aggressive viral extinction in extraembryonic endoderm stem cells, *Cell Stem Cell*, **6**, 457–67.
23. Leeb, M., Pasini, D., Novatchkova, M., Jaritz, M., Helin, K. and Wutz, A. 2010, Polycomb complexes act redundantly to repress genomic repeats and genes, *Genes Dev.*, **24**, 265–76.
24. Shaver, S., Casas-Mollano, J.A., Cerny, R.L. and Cerutti, H. 2010, Origin of the polycomb repressive complex 2 and gene silencing by an E(z) homolog in the unicellular alga *Chlamydomonas*, *Epigenetics*, **5**, 301–12.
25. Haynes, K.A., Caudy, A.A., Collins, L. and Elgin, S.C. 2006, Element 1360 and RNAi components contribute to HP1-dependent silencing of a pericentric reporter, *Curr. Biol.*, **16**, 2222–27.

26. Riddle, N.C., Minoda, A., Kharchenko, P.V., Alekseyenko, A.A., Schwartz, Y.B., Tolstorukov, M.Y., et al. 2011, Plasticity in patterns of histone modifications and chromosomal proteins in *Drosophila heterochromatin*, *Genome Res.*, **21**, 147–63.
27. Kanhere, A., Viiri, K., Araujo, C.C., Rasaiyaah, J., Bouwman, R.D., Whyte, W.A., et al. 2010, Short RNAs are transcribed from repressed polycomb target genes and interact with polycomb repressive complex-2, *Mol. Cell*, **38**, 675–88.
28. Grimaud, C., Negre, N. and Cavalli, G. 2006, From genetics to epigenetics: the tale of Polycomb group and trithorax group genes, *Chromosome Res.*, **14**, 363–75.
29. Zhang, H., Christoforou, A., Aravind, L., Emmons, S.W., van den Heuvel, S. and Haber, D.A. 2004, The *C. elegans* Polycomb gene SOP-2 encodes an RNA binding protein, *Mol. Cell*, **14**, 841–47.
30. Zhao, J., Sun, B.K., Erwin, J.A., Song, J.J. and Lee, J.T. 2008, Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome, *Science*, **322**, 750–56.
31. Brennecke, J., Malone, C.D., Aravin, A.A., Sachidanandam, R., Stark, A. and Hannon, G.J. 2008, An epigenetic role for maternally inherited piRNAs in transposon silencing, *Science*, **322**, 1387–92.

Supplemental data 1



Supplementary figure 2

```
% ImarisXTColoc (Discs)
% A function to compute the percentage of colocalization in discs
% Works under Imaris 6.4.2 & 7.0.0
function ImarisXTColoc_discs(aImarisID) % ID corresponds to Imaris instance
%% COM initialization
vImarisServer = actxserver('ImarisServer.Server');
vImarisApplication = vImarisServer.GetObject(aImarisID);

%% Data cloning
aType='eTypeUInt16';
vImarisApplication.mDataSet.mType=aType;
aSurpassScene = vImarisApplication.mSurpassScene;
aDataSet=vImarisApplication.mDataSet.Clone;
aBackgroundSubtraction=1;

%% Size variables
sizeX = vImarisApplication.mDataSet.mSizeX;
sizeY = vImarisApplication.mDataSet.mSizeY;
sizeZ = vImarisApplication.mDataSet.mSizeZ;
sizeC = vImarisApplication.mDataSet.mSizeC;
extendMinX = aDataSet.mExtendMinX;
extendMinY = aDataSet.mExtendMinY;
extendMinZ = aDataSet.mExtendMinZ;
extendMaxX = aDataSet.mExtendMaxX;
extendMaxY = aDataSet.mExtendMaxY;
extendMaxZ = aDataSet.mExtendMaxZ;
spacingX = (extendMaxX - extendMinX) / sizeX;
spacingY = (extendMaxY - extendMinY) / sizeY;
spacingZ = (extendMaxZ - extendMinZ) / sizeZ;

% Number of polycombs spots required in a nucleus
nbSpots = 9;

% Brightness threshold for 'red'
brthresh = 70;

% Vector "numberofspots" initialization
aNumberOfSpots = {'Canal', 'Nombre'};

%% User interface
% We have the user select which channel corresponds to the DAPI
prompt={'Enter the "DAPI Channel" number:'};
name='Colocalization parameters';
numlines=1;
defaultanswer={num2str(sizeC)};
for i=1:sizeC
    prompt{i+1} = ['(Channel ',num2str(i),') Estimated size of the objects (um):'];
    if sizeC == 3
        if i == 1 || i == 2
            defaultanswer{i+1}='0.4';
        else
            defaultanswer{i+1}='2';
        end
    else
        defaultanswer{i+1}='6';
    end
end

% We retrieve the answers
answer=inputdlg(prompt,name,numlines,defaultanswer);
if isempty(answer)
    return;
end
dapi=floor(str2double(answer{1})-1); % DAPI Channel
objectsSize = zeros(sizeC,1);
for i=1:sizeC
    objectsSize(i)=str2double(answer{i+1}); % Size of the nuclei
end

% We rectify incorrect answers
if dapi < 0
    dapi = 0;
else if dapi >= sizeC
    dapi = sizeC - 1;
end
end
```



```

% Channels content
if sizeC == 3
    c1 = 1;
    c2 = 2;
    if dapi ~= 2
        c2 = 3;
    end
    if dapi == 0
        c1 = 2;
    end
end

% Waitbar
vProgressTotalCount = 6;
vProgressDisplay = waitbar(0, 'Thresholding');
vProgressCount = 0;

%% Cells Segmentation
%% Thresholding
aDataDNA1D = aDataSet.GetDataVolumeAs1DArray(dapi, 0);
histo = histc(single(aDataDNA1D), 0:255);
P = histo / sum(histo);
P2 = P .* (1:256);
muT = ones(256, 256) * sum(P2(1:256));
omega0 = zeros(256, 256);
omega1 = omega0;
omega2 = omega0;
mu0 = omega0;
mu1 = omega0;
mu2 = omega0;
for i=1:255
    for j=i+1:256
        omega0(i,j) = sum(P(1:i));
        omega1(i,j) = sum(P(i+1:j));
        omega2(i,j) = sum(P(j+1:256));
        mu0(i,j) = sum(P2(1:i)) ./ omega0(i,j);
        mu1(i,j) = sum(P2(i+1:j)) ./ omega1(i,j);
        mu2(i,j) = sum(P2(j+1:256)) ./ omega2(i,j);
    end
end
sigma = omega0 .* (mu0 - muT).^2 + omega1 .* (mu1 - muT).^2 + omega2 .* (mu2 - muT).^2;
[Y, I] = max(sigma);
[~, J] = max(Y);
thresh = I(J);
vImarisApplication.mImageProcessing.ThresholdChannel(aDataSet, dapi, thresh, 0);

vProgressCount = vProgressCount + 1;
waitbar(vProgressCount/vProgressTotalCount, vProgressDisplay, 'Cleaning DAPI')

%% Computing DNA regions
[aR, aG, aB, aA] = aDataSet.GetChannelColor(dapi);
nameC = aDataSet.GetChannelName(dapi);
aSpots{dapi+1} =
vImarisApplication.mImageProcessing.DetectSpotsRegionGrowing(aDataSet, [], dapi, objectsSize(dapi+1), aBackgroundSubtraction, "Quality above 0", 1, 1, 0.0, 1, 1);
aSpots{dapi+1}.mName = nameC;
len = length(aSpots{dapi+1}.GetRadii);
aSpots{dapi+1}.SetColor(aR, aG, aB, aA);
aNumberOfSpots{dapi+1, 1} = nameC;
aNumberOfSpots{dapi+1, 2} = int2str(len);

% Nuclei spots
XYZ = aSpots{dapi+1}.GetPositionsXYZ;
iT = aSpots{dapi+1}.GetIndicesT;
aSpots{dapi+1}.Set(XYZ, iT, ones(len, 1) * objectsSize(dapi+1)/2);
aSurpassScene.AddChild(aSpots{dapi+1});

```



```

% Excluding DNA regions (if small or incorrect shape)
aDataDNA1D = aDataSet.GetDataVolumeAs1DArray(sizeC,0);
aDataDNA = reshape(aDataDNA1D,sizeX,sizeY,sizeZ);
vmin = min(aDataDNA1D); % vmin = background
vmax = max(aDataDNA1D);
nucleusSizeVox = objectsSize(dapi+1) * [ 1/spacingX, 1/spacingY, 1/spacingZ ];
sizeLimit = 1/6*pi * 3*nucleusSizeVox(1)/4 * 3*nucleusSizeVox(2)/4 * 3*nucleusSizeVox(3)/4;
for v=vmin+1:vmax
    I = find(aDataDNA == v);
    if isempty(I) == 0
        [x, y, z] = ind2sub(size(aDataDNA),I);
        if numel(I) < sizeLimit || std(x) > 0.75*nucleusSizeVox(1) ...
            || std(y) > 0.75*nucleusSizeVox(2) || std(z) > nucleusSizeVox(3)
            aDataDNA(aDataDNA == v) = 0;
        end
    end
end

% Removing spots outside DNA regions
[aPositionsXYZ,aIndicesT,aRadii] = aSpots{dapi+1}.Get;
len = length(aRadii);
l=1;
while l <= len
    posx = ceil((aPositionsXYZ(l,1)-extendMinX)/spacingX);
    posy = ceil((aPositionsXYZ(l,2)-extendMinY)/spacingY);
    posz = ceil((aPositionsXYZ(l,3)-extendMinZ)/spacingZ);
    if aDataDNA(posx,posy,posz) == 0
        aPositionsXYZ(l,:) = [];
        aIndicesT(l) = [];
        aRadii(l) = [];
        len = len-1;
    else
        l = l+1;
    end
end
aSpots{dapi+1}.Set(aPositionsXYZ,aIndicesT,aRadii);
aSurpassScene.AddChild(aSpots{dapi+1});

vProgressCount = vProgressCount + 1;
waitbar(vProgressCount/vProgressTotalCount,vProgressDisplay,'Cleaning other channels')

%% Cleaning other channels
for c=0:sizeC-1
    if c ~= dapi
        aData1D = aDataSet.GetDataVolumeAs1DArray(c,0);
        aData = reshape(aData1D,sizeX,sizeY,sizeZ) .* min(1,aDataDNA);
        aDataSet.SetDataVolume(uint16(aData),c,0);
    end
end

% Saturating 'red' (brighter spots) for equivalent visibility in both cases
c = c2-1;
vImarisApplication.mImageProcessing.ThresholdUpperChannel(aDataSet, c, brthresh, 255);

vProgressCount = vProgressCount + 1;
waitbar(vProgressCount/vProgressTotalCount,vProgressDisplay,'Segmentation')

%% Other channels segmentation
for c=0:sizeC-1
    if c ~= dapi
        [aR,aG,aB,aA]=aDataSet.GetChannelColor(c);
        nameC = aDataSet.GetChannelName(c);
        aSpots{c+1} = vImarisApplication.mImageProcessing.DetectSpots2(aDataSet, ...
            [], c, objectsSize(c+1), ...
            aBackgroundSubtraction, ...
            "Quality" above automatic threshold');

        aSpots{c+1}.mName = nameC;
        aSpots{c+1}.SetColor(aR,aG,aB,aA);
        aNumberOfSpots{c+1,1} = nameC;
        aNumberOfSpots{c+1,2} = int2str(length(aSpots{c+1}.GetRadii));
        aSurpassScene.AddChild(aSpots{c+1});
    end
end
vProgressCount = vProgressCount + 1;
waitbar(vProgressCount/vProgressTotalCount,vProgressDisplay,'Colocalization')

```



```

%% Colocalization
if sizeC == 3
    % Spots variables
    p1 = aSpots{c1}.GetPositionsXYZ;
    p2 = aSpots{c2}.GetPositionsXYZ;
    p3 = aSpots{dapi+1}.GetPositionsXYZ;
    iT1 = aSpots{c1}.GetIndicesT;
    iT2 = aSpots{c2}.GetIndicesT;
    iT3 = aSpots{dapi+1}.GetIndicesT;
    r1 = aSpots{c1}.GetRadii;
    r2 = aSpots{c2}.GetRadii;
    r3 = aSpots{dapi+1}.GetRadii;
    len1 = length(r1);
    len2 = length(r2);
    len3 = length(r3);

    % Coloc spots initialization (= red spots)
    n = 0;
    p = zeros(len2,3);

    % Coloc test
    for j=1:len2
        for i=1:len1
            if norm(p1(i,:) - p2(j,:)) < 1 * (r1(i) + r2(j))
                n = n+1;
                p(n,:) = p2(j,:);
            end
        end
    end

    % Delete non coloc spots
    for j=len2:-1:n+1
        p(j,:) = [];
    end

    % Create objects
    if(n > 0)
        coloc = vImarisApplication.mFactory.CreateSpots;
        coloc.mName = 'Coloc';
        coloc.Set(p, zeros(n,1), r2(1)*ones(n,1));
        vImarisApplication.mSurpassScene.AddChild(coloc);
    end
    vProgressCount = vProgressCount + 1;
    waitbar(vProgressCount/vProgressTotalCount,vProgressDisplay,'Finding appropriate spots')

    % Filtering spots (1/nucleus max) and computing mean values
    aSpotsZone1 = zeros(len1, 1);
    aSpotsZone2 = zeros(len2, 1);
    aSpotsZone3 = zeros(len3, 1);
    aColocZones = zeros(n, 1);
    aZones = zeros(vmax-vmin-1,1);
    colocalone = 0;
    mean_polycombs = 0;
    total = 0;
    for i=1:len1
        posx = ceil((p1(i,1)-extendMinX)/spacingX);
        posy = ceil((p1(i,2)-extendMinY)/spacingY);
        posz = ceil((p1(i,3)-extendMinZ)/spacingZ);
        aSpotsZone1(i) = aDataDNA(posx,posy,posz);
    end
    for i=1:len2
        posx = ceil((p2(i,1)-extendMinX)/spacingX);
        posy = ceil((p2(i,2)-extendMinY)/spacingY);
        posz = ceil((p2(i,3)-extendMinZ)/spacingZ);
        aSpotsZone2(i) = aDataDNA(posx,posy,posz);
    end
    for i=1:len3
        posx = ceil((p3(i,1)-extendMinX)/spacingX);
        posy = ceil((p3(i,2)-extendMinY)/spacingY);
        posz = ceil((p3(i,3)-extendMinZ)/spacingZ);
        aSpotsZone3(i) = aDataDNA(posx,posy,posz);
    end
    for i=1:n
        posx = ceil((p(i,1)-extendMinX)/spacingX);
        posy = ceil((p(i,2)-extendMinY)/spacingY);
        posz = ceil((p(i,3)-extendMinZ)/spacingZ);
        aColocZones(i) = aDataDNA(posx,posy,posz);
    end

```



```

end

% Counting coloc spots
for v=vmin+1:vmax
    I = find(aSpotsZone1 == v);
    J = find(aSpotsZone2 == v);
    K = find(aColocZones == v);
    L = find(aSpotsZone3 == v);
    if (numel(I) > nbSpots || numel(K) > 0) && (numel(J) < 2)
        if numel(I) < nbSpots+1
            colocalone = colocalone + 1;
        end
        aZones(v - vmin) = 1;
        mean_polycombs = mean_polycombs + numel(I);
        total = total + 1;
    else
        aColocZones(K) = [];
        aSpotsZone1(I) = [];
        aSpotsZone2(J) = [];
        aSpotsZone3(L) = [];
        p1(I,:) = [];
        p2(J,:) = [];
        p3(L,:) = [];
        p(K,:) = [];
        r1(I) = [];
        r2(J) = [];
        r3(L) = [];
        iT1(I) = [];
        iT2(J) = [];
        iT3(L) = [];
        aDataDNA(aDataDNA == v) = 0;
    end
end
mean_polycombs = mean_polycombs / total;
nombre_coloc = length(aColocZones);
aNumberOfSpots{c1,2} = int2str(length(r1));
aNumberOfSpots{c2,2} = int2str(length(r2));
aNumberOfSpots{dapi+1,2} = int2str(length(r3));
aSpots{c1}.Set(p1, iT1, r1);
aSpots{c2}.Set(p2, iT2, r2);
aSpots{dapi+1}.Set(p3, iT3, r3);
aSurpassScene.AddChild(aSpots{c1});
aSurpassScene.AddChild(aSpots{c2});
aSurpassScene.AddChild(aSpots{dapi+1});

% Add spots to scene if coloc spots exist
if n > 0
    coloc.Set(p, zeros(nombre_coloc,1), r2(1)*ones(nombre_coloc,1));
    aSurpassScene.AddChild(coloc);
end
end

aDataSet.SetDataVolume(uint16(aDataDNA), sizeC, 0);
close(vProgressDisplay);
display(aNumberOfSpots);
display(nombre_coloc);
display(colocalone);
display(mean_polycombs);
display(total);

aType='eTypeUInt8';
vImarisApplication.mDataSet.mType=aType;

end

```

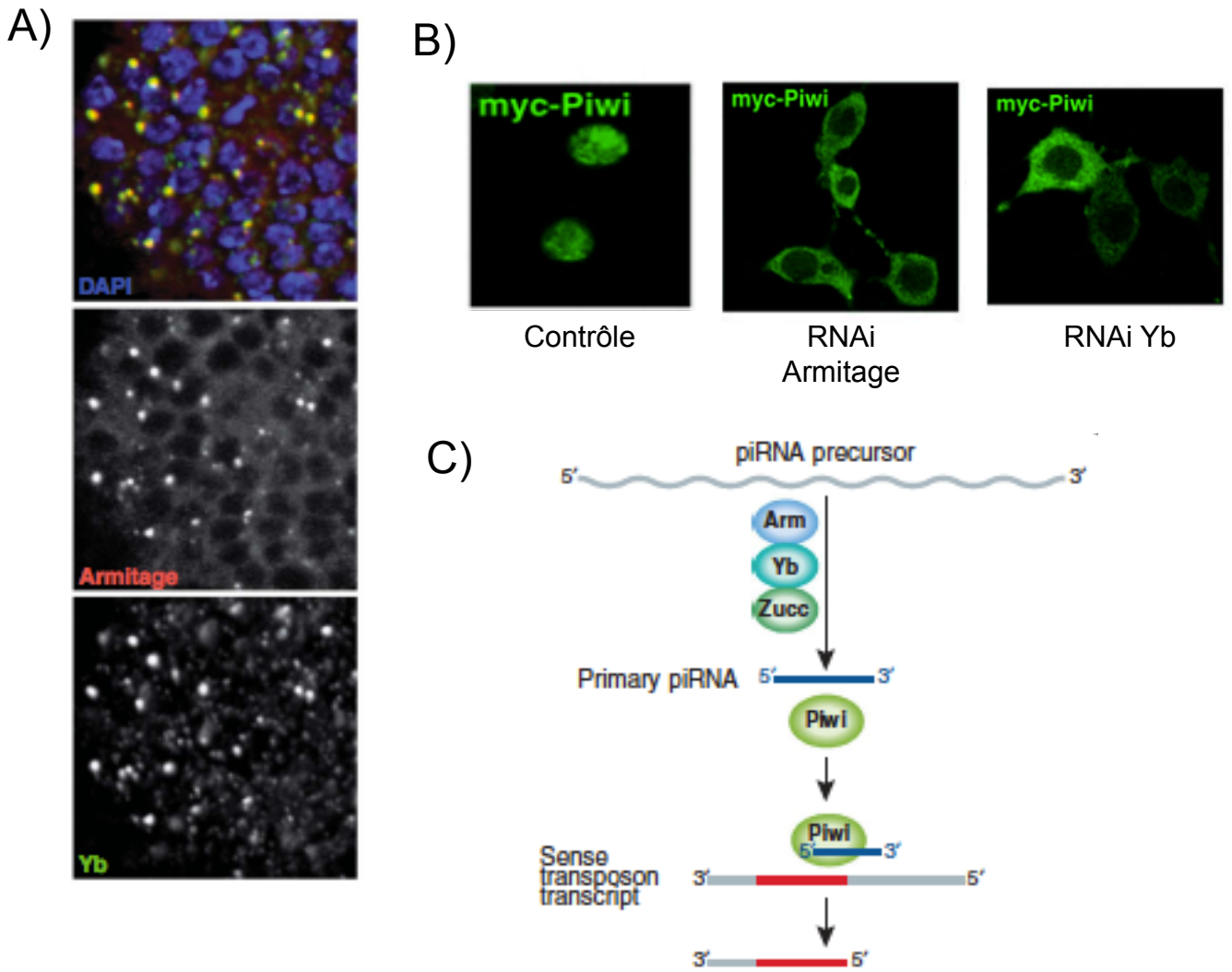


Figure 78: La voie des piRNAs en cellules folliculaires de drosophile

A) Co-immunomarquage des protéines Yb (vert) et Armitage (rouge) en cellules folliculaires. Les protéines colocalisent dans des foyers (L'ADN est marqué en bleu). **B)** En absence des protéines Yb ou Armitage, la protéine PIWI (vert) ne rentre pas dans le noyau des cellules folliculaires en culture (OSS) comparé au contrôle sauvage. **C)** La voie des piRNAs en cellules folliculaires commence par la maturation du transcrite des clusters de piRNAs par trois protéines Armitage, Yb et Zucchini. Cette maturation donne des piRNAs primaires qui vont être chargés dans la protéine PIWI et vont aller cibler et détruire les ARNm des éléments transposables. [Olivieri *et al.*, 2010; Zamore 2010; Saito *et al.*, 2010].

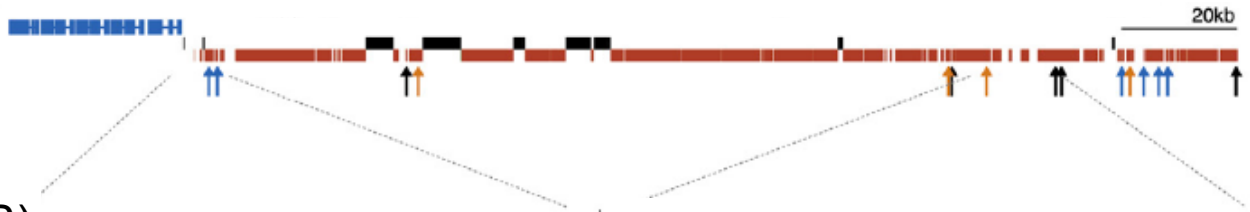
3) Conclusions

De manière très intéressante j'ai pu montrer que dans les tissus somatiques, Idefix était sujet à deux types de régulations. Dans les cellules folliculaires, qui sont des tissus somatiques en contact avec la lignée germinale et où la protéine PIWI est présente, la régulation d'Idefix est seulement de type post-transcriptionnelle. Ces résultats obtenus au début de ma thèse ont aussi été observés par l'équipe de Greg Hannon où il a été montré que la protéine PIWI est continuellement requise pour maintenir les éléments transposables sous silence en culture de cellules dérivées des cellules folliculaires de drosophile, les cellules OSS (Ovarian Somatic Sheet) (Haase, Fenoglio et al. 2010).

Dans ces tissus somatiques la voie des piRNAs est active mais semble être simplifiée (Lau, Robine et al. 2009; Olivieri, Sykora et al. 2010; Saito, Ishizu et al. 2010). En effet, il y a absence des protéines Aubergine et Ago3 et par conséquent pas de *Ping-Pong* possible. La voie des piRNAs somatiques semble ne faire intervenir que peu d'acteurs, les transcrits issus des cluster de piRNAs seraient maturés en petits ARN *via* un complexe qui contient les protéines Armitage, Yb et Zuc situées dans le cytoplasme et formant des foyers (**Figure 78 A**). Ces piRNAs seraient ensuite chargés sur la protéine PIWI cytoplasmique qui transloquerait dans le noyau (**Figure 78 B**) pour aller cibler les ARNm des éléments transposables (**Figure 78 C**).

Ils semblent cependant étrange que la protéine HP1 n'ait pas de rôle dans ces tissus car de nombreuses publications impliquent cette protéine, en ovaires, dans la mise sous silence des éléments transposables (Klenov, Lavrov et al. 2007; Shpiz, Olovnikov et al. 2011; Wang and Elgin 2011). Deux hypothèses peuvent expliquer la différence observée : soit les régulations observées se passent dans la lignée germinale, soit les observations faites par ces équipes concernent les régulations subies par les éléments hétérochromatiques. Les données récentes du projet modENCODE semblent montrer que

A)



B)

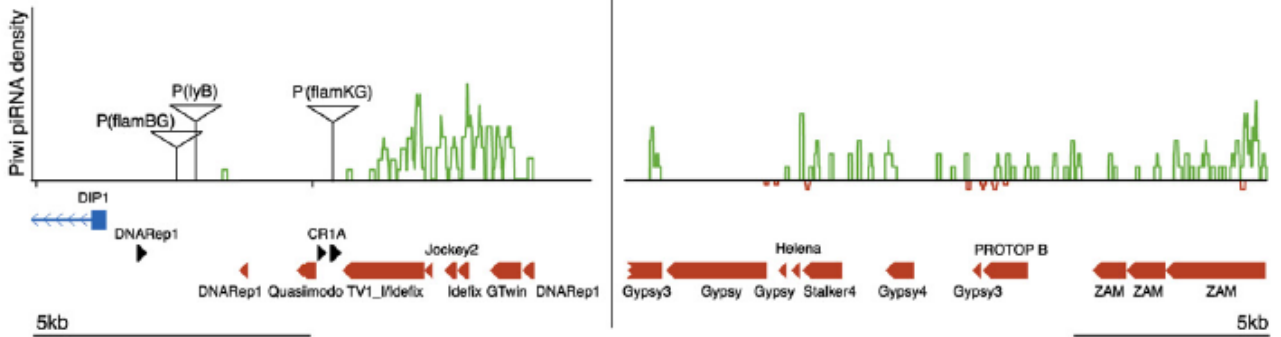


Figure 79: Organisation du cluster de piRNAs *Flam/COM*

A) Le cluster de piRNAs *Flam/COM* est situé dans une région péricentromérique du chromosome X de drosophile. Il est borné du côté distal par le gène *dip1* (bleu). Il est composé par de nombreux élément transposable défectifs insérés en antisens du sens de transcription du locus (rectangles rouges). Les éléments insérés en sens sont représentés par des rectangle noirs. **B)** Focus dur deux zones du cluster où on peut observer qu'il y a de nombreux fragments d'éléments transposables (flèches épaisses rouge et noire). Les pics en vert représentent la densité en piRNAs uniquement identiques aux séquences génomiques présentent dans le cluster *Flam/COM* (précipités avec la protéine PIWI). [Brennecke *et al.*, 2007].

seuls les éléments transposables situés dans des régions hétérochromatiques sont cibles de HP1, alors que les éléments euchromatiques sont recouverts de modifications de chromatine très variées mais potentiellement permissives à la transcription (Riddle, Minoda et al. 2011). De plus, les données de DamID avec HP1 indiquent que cette protéine est présente sur les séquences répétées qui se trouvent préférentiellement au milieu de nombreuses autres séquences répétées (de Wit, Greil et al. 2007). Ce cas de figure est rencontré particulièrement dans l'hétérochromatine, par exemple dans les clusters de piRNAs (**Figure 79**). Il en ressort donc que les analyses faites par ces laboratoires seraient majoritairement focalisées sur les éléments hétérochromatiques et potentiellement non actifs (car délétés) ou régulateurs (cluster de piRNAs). L'utilisation de transgènes euchromatiques présents en une seule copie semblent donc être nécessaire pour analyser les mécanismes de régulations des éléments transposables euchromatiques potentiellement actifs et dont la transposition doit être contrôlée.

Parallèlement à cela, j'ai montré que dans les tissus somatiques en dehors de l'ovaire (larve et carcasse adulte) Idefix est recouvert par une structure chromatinienne de type fermée et impliquant le PRC1, le PRC2 et la marque H3K9me³. Il y aurait donc dans ces tissus une régulation transcriptionnelle sur les éléments transposables qui permettrait de verrouiller leur expression. J'ai aussi pu mettre en évidence que HP1 n'était pas présente sur les transgènes. Il s'agirait donc d'une couverture chromatinienne de type chromatine bleue. Il a été montré que les protéines PcGs ne couvrent pas les mêmes régions que HP1 ce qui est en accord avec les résultats trouvés (de Wit, Greil et al. 2007).

Ces résultats soulèvent une question importante concernant le ciblage des éléments transposables afin de déposer dessus des marques de fermeture de la chromatine. Il est clairement établi que des ARNs peuvent permettre de diriger le dépôt des protéines PcGs et on sait que dans ces tissus somatiques la voie des endo-siRNAs est active. J'ai donc

tenté de muter les gènes impliqués dans la voie des endo-siRNAs mais cela n'a pas permis de lever la répression en tissu somatique. On peut penser que si les transgènes sont enveloppés dans une structure chromatinienne compacte, les voies d'ARN interférences ne seraient pas continuellement requises contrairement à ce qu'il se passe dans les cellules folliculaires. Il serait donc intéressant de rechercher des mutants qui permettraient de lever la répression en tissus somatiques en dehors de l'ovaire. Pour cela un crible RNAi en culture de cellules somatiques pourrait être envisagé.

Dès le début de mon master 2, je me suis aussi intéressé aux régulations des éléments transposables dans la lignée germinale. Ce sont les résultats de ces études que je vais présenter maintenant.

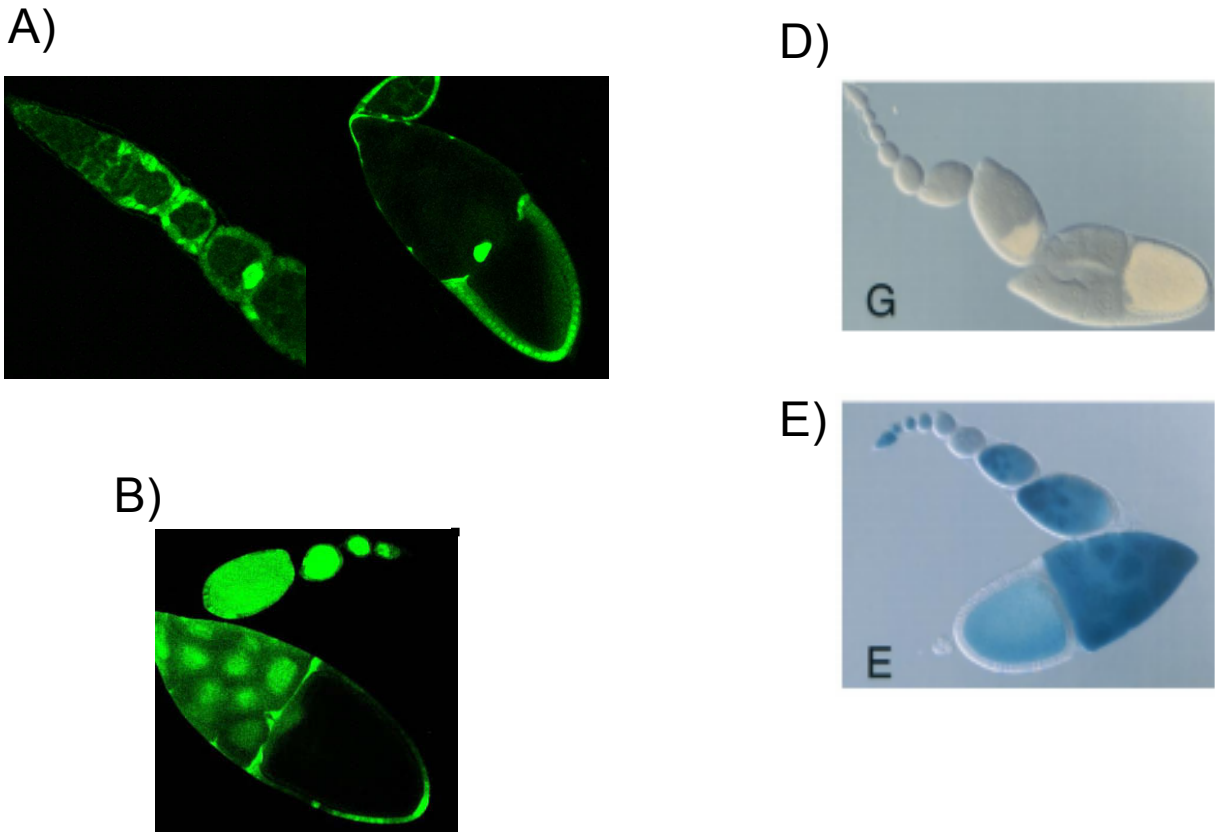


Figure 80: pUASp et pUASt dans les ovaires de drosophile

A) Transgène exprimé dans les ovaires de drosophiles, celui ci est fait à partir d'un vecteur pUASt couplé à la GFP et activé ubiquitairement. **B)** Il s'agit d'un transgène fait à partir d'un vecteur pUASp couplé à la GFP et activé ubiquitairement. On peut remarquer que le pUASt ne peut pas exprimer la GFP dans la lignée germinale contrairement au pUASp (vert). **D)** Transgène exprimé dans les ovaires de drosophiles à partir d'un vecteur pUASt couplé à la b-Galactosidase et activé uniquement dans la lignée germinale. Transgène exprimé dans les ovaires de drosophiles à partir d'un vecteur pUASp couplé à la β -galactosidase et activé uniquement dans la lignée germinale. Seul le vecteur pUASp permet d'exprimer la β -galactosidase dans la lignée germinale (bleu). [Rorth 1998]

Publication 2

Régulation des éléments transposables dans les tissus germinaux

1) Introduction

Les tissus germinaux sont les tissus qui vont permettre de donner la descendance, il est donc primordial de les protéger des effets délétères des éléments transposables. Paradoxalement c'est dans la lignée germinale que doivent se mobiliser les éléments transposables afin de générer de la variabilité génétique. Il est donc important de comprendre comment les organismes ont évolué afin de bien réguler les éléments transposables dans ces tissus mais aussi de leur permettre de se mobiliser. On sait que dans ces tissus la voie des piRNAs est active et comprend beaucoup plus d'acteurs que la voie des piRNAs dans les cellules folliculaires (tissus somatique) où la régulation des éléments transposables est uniquement de type PTGS (Dufourt, Brasset et al. 2011).

Mon objectif a été d'analyser les régulations (TGS, PTGS ou les deux) que peuvent subir les éléments transposables euchromatiques dans la lignée germinale. Afin de répondre à cette question, j'ai utilisé des transgènes que j'ai établis sur le modèle de ceux utilisés dans les tissus somatiques avec cependant une variante majeure : le promoteur minimal de ces transgènes est celui de l'élément P de drosophile. J'ai donc créé des vecteurs de transgénèse à partir du vecteur pUASp capables de s'exprimer dans la lignée germinale, car les transgènes établis précédemment à partir du vecteur pUAS_T (Dufourt, Brasset et al. 2011) ne peuvent s'exprimer dans ce tissu (**Figure 80**) (Rorth 1998; Desset, Buchon et al. 2008; Ni, Zhou et al. 2011). Les transgènes ont été construits selon le modèle décrit figure 74A : « promoteur-GFP-Idefix ». Dans ces transgènes, la GFP correspond au gène rapporteur codant la Green Fluorescente Protein. Idefix correspond à un fragment

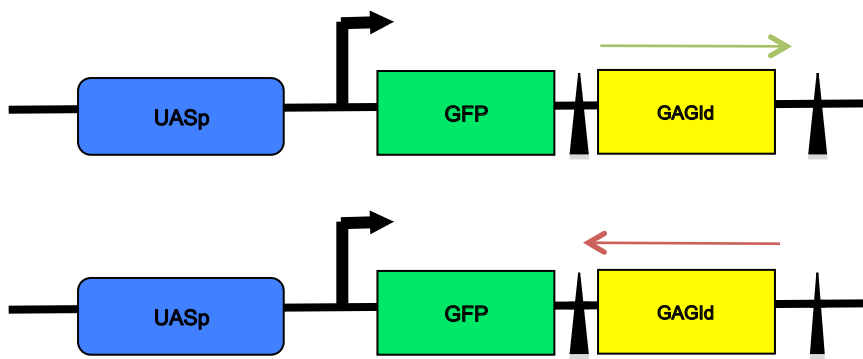


Figure 81: Structure des transgènes utilisés pour l'étude d'Idefix en ligné germinale

Représentation schématique des transgènes [UASp-Promoteur-GFP-*Idefix*] utilisés pour l'étude de la régulation germinale des éléments transposables. Un fragment d'Idefix (rectangle jaune) est inséré en sens (flèche verte) ou antisens (flèche rouge) de transcription de la GFP, qui sert de gène rapporteur (rectangle vert). Les fragments d'Idefix sont encadrés par des séquences FRT (Flip Recognition Target) (triangle noir). Les séquences permettant l'activation de la transcription du transgène sont des UAS (Upstream Activating Sequences) (rectangle bleu).

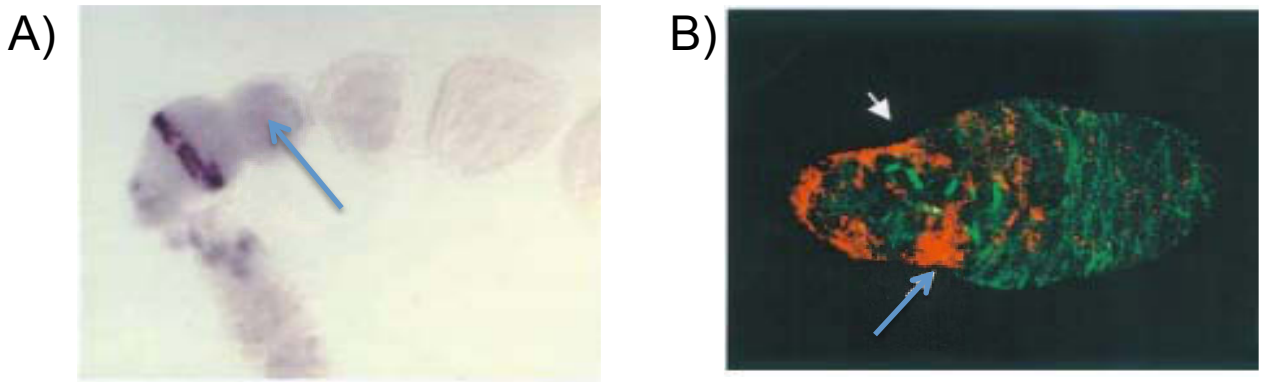


Figure 82: Expression d'Idefix en contexte mutant pour COM

A) Hybridation *in situ* avec une sonde reconnaissant le transcrit d'Idefix dans un contexte génétique permissif (mutation dans COM). Idefix est exprimé précocement dans le gerarium. **B)** Immunomarquage avec un anticorps anti β -gal (rouge) d'un gerarium contenant une construction de type LTR(Idefix)-LacZ dans un contexte génétique permissif (mutation dans COM). L'expression d'Idefix semble se situer dans des cellules germinales (flèches bleues). [Tcheressiz *et al.*, 2002] ☐

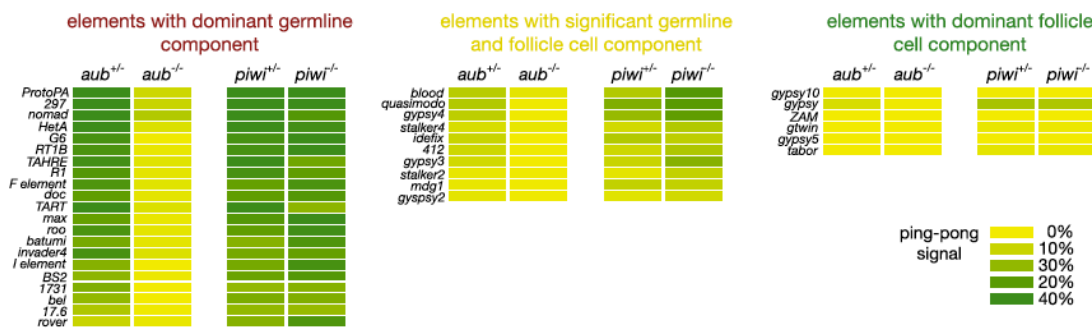


Figure 83: Trois classes d'éléments transposables régulés par les piRNAs

Les éléments transposables (ET) peuvent être classés en trois catégories reflétant leur lieu d'expression. La première comprend les ET qui sont exprimés majoritairement dans la lignée germinale (gauche) et ont un fort signal Ping-Pong. La deuxième les ET qui sont exprimés dans tout l'ovaire (centre) et la troisième les ET exprimés en cellules folliculaires uniquement (droite). Idefix est dans la deuxième catégorie [Malone *et al.*, 2009] ☐

d'Idefix orienté de façon à ce que le brin transcrit corresponde au transcrit sens ou antisens de cet élément (**Figure 81**).

Il peut paraître surprenant d'utiliser Idefix comme élément transposable dans l'étude des régulations en lignée germinale car cet élément transposable a été décrit comme étant somatique et régulé dans les tissus somatiques. Toutefois il existe plusieurs raisons intéressantes d'utiliser Idefix comme rapporteur de régulation en lignée germinale. En effet il a été montré que la régulation d'Idefix était conservée même dans les cellules ou les tissus où l'élément endogène n'est pas exprimé (Desset, Buchon et al. 2008; Dufourt, Brassat et al. 2011). De plus Idefix est exprimé dans la lignée germinale mâle de drosophile (Tcheressiz, Calco et al. 2002) et il semblerait que son expression dans le germarium des tissus reproducteurs femelles de drosophile ne soit pas forcément restreinte aux tissus somatiques (**Figure 82**). Enfin, récemment, le séquençage massif des piRNAs différenciellement exprimés entre la lignée germinale et les cellules folliculaires a permis de classer les éléments transposables en trois catégories : les éléments uniquement somatiques, les éléments uniquement germinaux et des éléments présents à la fois dans les tissus somatiques et germinaux (Li, Vagin et al. 2009; Malone, Brennecke et al. 2009). Cette classification a été faite en fonction de la quantité de paires de ping-pong de piRNAs, qui est le reflet d'une activité des éléments en tissu germinale, seul tissu où ce mécanisme peut avoir lieu (**Figure 83**).

An in vitro RNAi assay demonstrates a hierarchical rule in piRNA post-transcriptional silencing repressing transposable element in the germ line of *Drosophila melanogaster* with a small window of repression sensitivity in the dividing cysts

J. Dufourt, E. Brasset, N. Gueguen, C. Dennis, P. Pouchin and C. Vaury*

Address:

Clermont Université, Université d'Auvergne, BP 10448, F-63000 Clermont-Ferrand, France

Inserm, U1103, F-63001 Clermont-Ferrand, France

CNRS, UMR6293, F-63001 Clermont-Ferrand, France

* Corresponding author

Tel: (33) 4 73 17 81 84

Fax: (33) 4 73 27 61 32

E-mail: Chantal.VAURY@u-clermont1.fr

Abstract:

The PIWI-interacting RNAs (piRNAs) and the so-called piRNA pathway silence transposable elements (TE) in the germ line and protect many animals against their deleterious effect. In *Drosophila* germ line, piRNAs, encoded by heterochromatic clusters and transposons, undergo a target-dependent amplification, the ping-pong cycle. Genes involved in this germ line pathway have been reported to be necessary for piRNA biogenesis, nuage formation or compact chromatin structure. Despite this, the successive steps of the PIWI-mediated target silencing remain elusive as well as the distinct regulations exerted on either the piRNA clusters or their targets. Here, we engineered transgenes providing a convenient read-out of the silencing exerted on Idefix, a *Drosophila* retrotransposon, and distinguish its regulation from that of the heterochromatic piRNA cluster. We show that Aub, Spn-E, Squ, Vasa and Krimp should be impaired early in the developing cysts to release silencing but, thereafter, their presence is dispensable to the later stages of oogenesis. By contrast, Ago3, PIWI and Mael are continuously necessary. When the nuage component Armi is absent, the transcriptional silencing is lost but a repression at a translational level persists. Furthermore, prolonging the molecular characterization of the piRNA pathways, we present for the first time genetic evidence for piRNA biogenesis differing in the germ cell and in their associated somatic follicle cells.

Our data further point to a post transcriptional silencing and an absence of transcriptional silencing targeting TE in the germ line. Finally, we identify a small window during cysts mitosis in which PIWI and HP1 expression are impaired and the targeted sensor-transgene de-repressed.

Introduction:

Transposable elements (TE) are major constituent of eukaryotic genomes. Since their activity represents a constant threat for genome stability, organisms have developed protection mechanisms that prevent TE from producing proteins required for their mobilization cycle. Yet, to colonize genomes, TE has to escape these defense mechanisms and mobilize in cells that will ensure their propagation to the next generation. Therefore, the germ line development constitutes both a critical stage at which the defense mechanisms has to be active to maintain germ-line integrity, and a suitable window at which any escape in the control will result in heritable new TE-insertions. Extensive studies performed in *Drosophila* and mice have identified the piRNA pathway that implicates small interfering RNAs associated to PIWI proteins, the piRNAs, as the important pathway for mounting an effective defense in the germ line [1-3]. In *Drosophila*, an ovary consists of 15 to 18 ovarioles that each contains a series of egg chambers at progressively more-advanced stages of oogenesis [4]. The stem cells of the germ line (GSC) and the somatic follicle cell precursors reside in a region called the germarium at the anterior tip of each ovariole [5]. During GSC division, one daughter cell remains in the niche and continues to divide as a GSC. The daughter cell called cystoblast undergoes four cycles of mitotic division to form interconnected cysts of successively 2, 4, 8 and 16 germ cells. When the mature egg chamber leaves the germarium, the germ line consists of the oocyte and 15 nurse cells surrounded both by a monolayer of somatic follicle cells. Both germ cells and somatic follicle cells possess a functional piRNA pathway but they differ in the piRNAs biogenesis. In both types of cells, a pool of primary piRNAs is processed from presumed long single stranded transcripts and target transposable elements. These long transcripts are produced from discrete genomic loci (piRNAs clusters), which typically reside in pericentric heterochromatin enriched in TE or their relics. In the germ line only, these primary piRNAs produced typically from anti-sense transcripts target

transposon sense-transcripts what results in a secondary pool of piRNAs [3]. Secondary sense piRNAs may enhance cleavage of anti-sense piRNAs, and reciprocally, resulting in an amplification of the piRNAs production called the ping-pong cycle. The three PIWI proteins - PIWI, Aub and Ago3 - are major players of this ping-pong pathway and have been shown to have a crucial role in gonadal development. By contrast, somatic ovarian follicle cells only express primary piRNAs produced from piRNAs clusters. Among the three PIWI proteins, only PIWI is used in a pathway that lacks the ping-pong amplification cycle [6-8].

Several screens have uncovered factors required for piRNA biogenesis. In the germ line, mutants of the nuage components, such as *vasa*, *maelstrom (mael)*, *armitage (armi)*, *squash (squ)*, and *krimper (krimp)* exhibit defects in piRNA production and de-repression of transposons [9-13]. This identified the nuage as a critical germ line structure for piRNA production. Similarly, in the follicle cells, the proteins Armi, Zucchini (Zuc), Yb and the TUDOR domain protein Vreteno are required for piRNA biogenesis in a perinuclear cytoplasmic structure called the Yb body [6-8, 14].

Overall, the piRNA pathway active somatic cells and germ cells implicate a post-transcriptional silencing (PTGS) targeting mRNAs encoded by TE [14, 15].

In mice, the role of PIWI and piRNAs seems to extend to transcriptional silencing (TGS). *Mili* and *Miwi2* mutants fail to establish *de novo* methylation of TE sequences [16, 17]. In *Drosophila*, PIWI and piRNAs have also been reported to prevent TE mobilisation through effects on chromatin. PIWI, whose localisation is mainly nuclear, has been shown to interact with the heterochromatin protein HP1a [18]. Recently, Rangan et al. reported that heterochromatin formation is required on piRNA clusters for piRNA transcription and germline silencing of TE [19]. However, since piRNAs clusters harbour fragments and relics of TE whose sequence can hardly be distinguished from the corresponding active TE copies,

further investigations are necessary to elucidate whether heterochromatic structures are deposited on the silenced TE targets or on their regulatory piRNA clusters, or on both.

We recently reported sensor-transgenes whose expression is submitted to the silencing pathway targeting the *Drosophila* retrotransposon, *Idefix*. We have shown that these transgenes adopt a heterochromatic configuration with an hypermethylation of H3K9 and -K27, and with Polycomb Repressive Complexes (PRC1 and 2) in the somatic tissues outside of the ovaries. By contrast, they display an open and labile chromatin structure in the somatic follicle cells suggesting that the piRNA pathway acting at a post transcriptional level is the only mechanism used to silence the TE-targets in these cells [15].

Here, we used new sensor-transgenes to elucidate germ line components from the PTGS and/or TGS pathways affecting TE silencing because transgene *tGgId* previously describe couldn't be express in the germline even with a strong germline specific driver [15, 20, 21]. An *Idefix*-sensor called *pGgId*, driven by an UASp promoter active in the germ line was engineered. Combining RNAi with immunofluorescent antibodies and ChIP experiments, we found that the silencing targeting *pGgId* only depends on a PTGS mechanism implicating the piRNA pathway. Further, we provide several lines of evidence that transcriptional silencing does not impair *pGgId* expression in the germ line: i) the *Idefix*-sensor is not embedded in a compact heterochromatin structure implicating H3K9me3 and HP1 marks but rather in an active chromatin structure; ii) the silencing is not epigenetically transmitted since it is lost as soon as the targeted sequences is lost. Although germ line silencing is very efficient, we identify a small window of the germarial development in which the silencing of sensor-transgenes is weakened. Further, our experiments establish that addition of the *Idefix*-sensor in the *Drosophila* genome has a more general impact than its sole control. It also enhances repression of endogenous functional *Idefix*.

Results:

1) An engineered transgene enlightens the silencing pathway exerted on TE targets in the germ cells.

We attempted to generate a *Drosophila* reporter system that closely resembles natural germinal piRNA targets and, at the same time, provides a convenient read-out. Transgenic lines carrying transgenes named *pGgIds* or *pGgIdas* were established. These transgenes display a *gfp* reporter gene (G) linked to a fragment of the *gag* gene of *Idefix* either in a sense (*gIds*) or an anti-sense orientation (*gIdas*). The *Idefix* fragment is flanked by FRT sequences (Fig. 1a). *pGgIds* and *pGgIdas* were driven from their *UASp* promoter (p) by two different drivers, ie: the ubiquitous *Actine-Gal4* (Fig. 1b, s1 and as1) and the germ line specific α *tubuline-Gal4* (*atub-Gal4*) (Fig. 1b, s2 and as2). Whatever the driver, no GFP signal, both at the RNA and protein levels, was detected from the *pGgIds* transgene in the germline throughout oogenesis (Fig. 1b, line s1 and s2). The silencing affecting *pGgIdas* was found less efficient because a faint GFP expression was occasionally recovered in the nurse cells with both drivers. We also noticed that a stronger GFP signal was detected in the early stages of oogenesis within the germarium (Fig. 1b, line as1). When *Idefix* is excised thanks to the FRT sequences, GFP RNA and protein signals are fully recovered from all the resulting *pGg^{Dld}* transgenes (Fig. 1b, lines s1, as1, s2 & as2, second and fourth panels). As expected, when *pGg^{Dld}* lines lacking the *Idefix* sequence are driven by the germ line specific driver, *atub-Gal4*, GFP expression is never recovered in the somatic follicular cells (dotted line Fig 1b, line as2 fourth panel).

We used quantitative RT-PCR with primers targeting *gfp* to measure the expression of the sensor-transgenes (see Methods for details). The *gfp* primers that are unique in the transgenic lines provide suitable tools to distinguish transcripts specifically produced by the *Idefix*-sensors. When driven by *atub-Gal4*, expression from both *pGgIds* and *pGgIdas* is very low

although *pGgIds* displays a 2-fold higher expression than *pGgIds* (p value=0,0071). After *Idefix* has been flipped out, RNAs increase more than 13-fold (p value=0,0007) and 10-fold (p value=0,0038) the amount of RNAs encoded by *pGgIds* and *pGgIds* respectively (Fig. 1c).

We conclude that the *Idefix* fragment present within *pGgIds* and *pGgIds* is target of the germ line repression.

Thus, we screened for components from PTGS pathways involved in this germ line silencing. Lines called pValium 20 or 22 in which RNAi transgenes can be induced in the germ line were used to disrupt either the ping-pong mechanism (Ago3, Aub, PIWI) or nuage components (Sqh, Krimp, Mael, Armi, Spn-E, Zuc, Vasa), or the miRNA pathway (Hen1, Drosha) [22-24]. A first series of crosses was performed with *pGgIds* driven by *αtub-Gal4*. We found that *pGgIds* is de-repressed in PIWI, Ago3- and Mael-RNAi. By contrast, Aub-, Zuc-, Spn-E-, Squ-, Krimp-, Armi-, Hen1-, and Drosha-RNAi as well as vasa-RNAi failed to de-repress the sensor transgene (Fig. 2a).

Since disruption of genes known as major components of the piRNA pathway such as *Aub*, *Vasa*, *Armi* or *Spn-E*, did not release the silencing, we wondered whether their expression was impaired at the wrong moment during the developing ovary. Indeed, driven by the *αtub-Gal4* driver; their activity was impaired when the developing egg chambers at stage 3 sort out of the germarium. Therefore, we decided to drive the RNAi-lines from the *nanos-Gal4* (*nos-Gal4*) driver whose activity starts early in the germarium when the cysts divide. We found that Aub-, Squ-, Krimp-, SpnE- and Vasa-RNAi driven by this earlier driver lead to de-repression of *pGgIds* (Fig. 2d). *PIWI*-RNAi lines could not be analyzed because the corresponding females displayed atrophic ovaries (data not shown). Armi-, Zuc-, Hen1- and Drosha-RNAi had no effect (data not shown unless for Armi in fig 2d).

Quantitative RT-PCR with primers targeting *gfp* (see Methods for details) confirmed these results (Fig 2b, c, e, f). However, we found one exception. While Armi-RNAi did not permit

to recover GFP fluorescence from *pGglds*, we found that de-repression could be visualized at the RNA level with a high increase of *gfp* RNAs (Fig. 2b and e).

Overall, our data confirm that components of the germ line piRNA pathway and the nuage are essential for the *pGglds* silencing. They also identify two categories of genes from the piRNA pathway: genes whose expression is required early in the germ line development and genes whose expression can be provided at any time along the egg chamber development. Our data further indicate that when Armi is absent, the transcriptional silencing is lost but a repression at a translational level persists.

The piRNA pathway does not direct heterochromatic structures on the silenced Idefix - sensors in the germ cells.

De-repression of TE by chromatin mutants might originate from modifications affecting chromatin deposited either on the TE or on its regulatory heterochromatic loci or on both. Therefore, *pGglds* and *pGglds* whose reporter gene *gfp* is a target of the piRNA pathway are suitable tools to analyze their associated chromatin structure.

Disruption of *Su(Var)205* (encoding *HPI*), *Su(Var)3-9*, *Rhino*, *Corto*, *Su(Var) 3-7*, *CENP-C*, *Ph-p*, *dSetdb1* and *Cuff* was performed in a secondary set of cross performed with the tub-Gal4 driver. A clear release of silencing was observed in the *Rhino*-, *HPI*- and *Cuff*-RNAi lines (Fig 2a) and not for *Corto*, *Su(Var) 3-7*, *CENP-C*, *Ph-p* and *dSetdb1*. Females issued from the tested *Su(Var)3-9*-RNAi line revealed to had surprising ovaries phenotypic defect (Sup data Fig 1).

When driven by *nos-Gal4*, the same results were obtained for the tested RNAi-lines disrupting proteins involved in compact chromatin structures (data not shown unless for *dSETdb1* (Fig. 2d). *Su(Var)3-9*- and *HPI*-RNAi lines gave rise to sterile females with an

ovarian phenotype of egg chambers arrested at an early and identical developmental stage and atrophic gonads respectively (Sup data Fig 2).

These data indicated that a compact chromatin structure implicating Cuff, Rhino and HP1 is necessary for the silencing of Idefix-sensors however they did not help to discriminate whether these chromatin structures were associated to the piRNA clusters or to their targets. Thus we performed ChIP experiments to analyze the proteins deposited on the sensors.

Modification of histone H3 were tested with H3K4me2, H3K36me3, H3K14ac as marks of active chromatin, and H3K9me2 and H3K9me3 as heterochromatic marks. Immunoprecipitation of HP1a, the heterochromatin protein 1, was added. H3K9me2, H3K9me3 marks and HP1a were found absent from *pGgIds* and *pGgIdas* transgenes (Fig. 3a red). As expected when the repressive marks are absent, both of these transgenes displayed an enrichment of H3K9/14ac, H3K36me3 and H3K4me2 (Fig. 3a green).

To confirm these results, we investigated whether HP1a enriched regions and the targeted transgene co-localize (or not) in nurse cell nuclei. A combination of 3-D FISH and immunostaining technique (FISH-I) using an anti-HP1 anti-body was used. No co-localizations could be detected (Fig. 3b). Since the silencing targeting the sensor transgenes does not implicate a compact heterochromatic structure that might convey the silencing from one generation of cells to the next, we thought to test whether *pGgIds* silencing switches to an active state as soon as the targeted *Idefix* fragment has been excised. Using of the *Flp*/FRT system and its selective induction by the heat shock *Flp* driver (HS-*Flp*), we determined how long the silencing of *pGgId* persists after *Idefix* has been flipped out. A one hour heat shock was performed on three days females with the genetic backgrounds [HS-*flp*;Actine>*yellow*>Gal4;UASp-*gfp*], [HS-*Flp*;Actine-Gal4;*pGgIds*], [HS-*Flp*;Actine-Gal4;*pGgIdas*]. The first line was used to estimate the time required for a *UAS-gfp* transgene to be fully expressed in the germ line after Gal4 induction (Fig. 3 a and [15]) The second

and third lines were used to estimate the time required for *pGglds* and *pGgldas* to lose silencing after the *Idefix* fragment inserted in a sense or antisense orientation respectively has been lost.

Ovaries were dissected every hour from the heat shock and GFP expression was examined in the germ cells. We found that 7 hours after Gal4 induction by the heat shock were necessary to visualize GFP expression from a control transgene *pGFP* (Fig. 3c, first line). Similarly, 7 hours were also necessary to recover GFP expression from *pGglds* and *pGgldas*. In the three lines, GFP fluorescence increases progressively to reach a strong green signal recovered in all the follicles at 24 hrs (Fig. 3). These data show that the germ line repression exerted on *pGgld* is lost as soon as the targeted retrotransposon sequence is excised.

Overall, these experiments indicate that the *pGglds* and *pGgldas* silencing is not coupled to a compact heterochromatic state able to be epigenetically maintained whenever the targeted sequence is lost. They further indicate that proteins associated to compact chromatin structures are indeed necessary for the silencing but they are likely to be necessary for the regulatory piRNA clusters and not their targets.

The presence of *Idefix*-sensors in the genome impacts on the endogenous *Idefix* repression in the germ line.

The ping-pong model has been proposed to silence TE in the germ line. In this model, complementary relationships between sense and anti-senses RNAs produced by TE mRNAs and piRNA clusters ensure an increase of piRNAs populations and consequently of TE silencing. Therefore, we wondered whether the addition of an *Idefix*-sensor transgene driven by *αtub-Gal4* within a genome could also impact the silencing exerted on endogenous *Idefix* retrotransposons. We carried out qRT/PCR experiments and measured transcript levels from endogenous *Idefix* elements in the ovaries of the transgenic lines *pGglds* compared to

pGgDlds line in which the Idefix fragment has been flipped out. We also analyzed transcripts produced by other TE expressed in the germ line such as the I element or HetA. Although no significant variation between I elements and HetA transcription was noticed, a clear decrease in Idefix transcripts was observed in *pGglds* compared to *pGgDlds* (Fig. 5a). We also performed the same experiment using the ubiquitous driver *Actine-Gal4* and found no variation in the I-element and HetA expression again. In this analysis, ZAM was added as a retrotransposon whose expression is restricted to the somatic follicle cells. No variation in its expression was observed either (Fig. 5a).

Thus, expression of transgenes that display sequence homology with Idefix helps repression of endogenous *Idefix* in the germ cells.

Then, we wondered whether the addition of two complementary fragments carried by the sensor-transgenes, *pGglds* + *pGgldas*, might become actors of the ping pong cycle and contribute to reinforce the target silencing. To this hand, we examined GFP expression in flies with the following genotype [*Actine-Gal4/Cy*; *pGglds/ pGgldas*]. Contrasting with the homozygous *pGgldas* line in which *pGgldas* is not fully repressed in the germ line of 2 days-old females as exemplified Fig. 4b, we found that no GFP could be detected in the ovarian germ cells of the double transgenic line. Thus, the presence of *pGglds* enhances the silencing exerted on *pGgldas*. Contrasting with the germ line, an opposite effect was revealed in the follicle cells. All along the ovariole, a strong GFP signal was detected in the follicle cells where both complementary sensor-transgenes are expressed (arrows fig.4b, bottom).

This experiment evidences genetically the existence of the two distinct piRNA pathways for the control of TE in ovarian somatic cells versus germ cells: the presence of two complementary targeted sequences enhances silencing in germ cells where the ping pong mechanism is active whereas the presence of both impairs repression in the somatic follicle cells where no ping-pong cycle exists.

Germ line repression of TE is much sensitized in a patch of cells early in the germarium.

Although no GFP could be detected in the nurse cells when pGgIds and pGgIdas driven by actine-Gal4 were present, a small patch of cells located in the anterior tip of the germarium remained fluorescent (Fig. 4c). This suggested that the piRNA pathway might be less efficient in this specific group of cells. Through immunohistology experiments, we better characterized the cells and found that GFP expressed from *pGgIdas* co-localizes with *gfp* mRNA and the germ line marker Vasa (Fig. 5, first line and sup data fig 3). An accurate examination indicated that co-staining between GFP and Vasa anti-bodies is only observed in dividing cysts either of 4, 8 or 16 cells at stage 2a of the germarium development (data not shown).

Since the piRNA pathway silences pGgId sensors, we wondered whether *PIWI*, *Aub* and *Ago3* were expressed in these cells. Co-immunostaining experiments performed with Aub- or ago3-antibodies showed that both proteins are expressed throughout the germarium including in cysts where GFP is detected (Fig. 5a, line 3 and 4). In contrast, although the PIWI protein was also detected, we found that in cells expressing GFP, the amount of Piwi was very low because hardly detected by the Piwi antibody (Fig. 5a, line 2 and sup data 4).

Recently, it has been reported that H3K9me3, HP1 and H4K20me3 are found in repressive chromatin centers (RCC) that are required in the germ line for piRNAs production [19]. Accordingly, by immunostaining experiments using a H3K9me3 antibody, we found H3K9me3 signals in prominent and discrete foci from the cystoblast to the late stages of germarium development (Fig. 5b line 1). To our surprise, immunostaining experiments using HP1 antibody revealed that HP1 signal is absent from RCC in the early stages of the germarium development including cells where *pGgIdas* repression is inactive and PIWI expression is low (Fig. 5b line 2 and sup data fig4). At later stages of the germarium

development starting at stage 2b, HP1 expression is then detected and co-localizes to RCC with H3K9me3 (Fig. 5b line 2 and sup data fig 6, 7).

Thus, the silencing of *Idefix-sensors* is weakened in the dividing cysts when Piwi and HP1 proteins are deficient. Our data further indicate that this silencing is impaired whenever Aub, Ago3 are expressed and H3K9me3 detected in RCC.

To further understand the reasons why the *Idefix-sensors* are de-repressed in this specific lineage, we established a transgenic line in which *piwi* expression could be artificially driven in all the germ cells including the dividing cysts. A transgene (*UASp-piwi*) in which the cDNA encoded by *piwi* was placed under the control of the UASp promoter (Fig. 5a) was over expressed in the germ cells by *nos-Gal4*. Surprisingly, ovaries expressing *nos-Gal4/UASp-piwi* together with *pGglds* or *pGgldas* displayed the same phenotype as ovaries without *nos-Gal4/UASp-piwi*. Piwi expression could not be detected in the 4 to 8-cells cysts and GFP expression due to a lack of *pGglds* repression was observed (Fig. 5b). These results suggested that *piwi* expression is low in this specific germ cell lineage because of a post-transcriptional or post-translational regulation.

Material and Method

Drosophila strains and transgenic lines

All experiments were performed at 22°C. The used fly strains were Bl279 (+ ;+/+; MKRS hs-Flip/Tm6), Act5C-Gal4 (3954, 4414, 25708), nos-Gal4 (4937) and Ay-Gal4(4412) comes from Bloomington stock. RNAi lines were from the Bloomington stock Centre: ago3 (35232), armi (35343), aub (35201), cenp-c (35480), corto (35456), cuff (35318), drosha (35233), dSetdb1 (34803), hen-1 (35203), krimp (35230, 35231), mael (35202), ph-p (35207), spn-e (35303), squ (35574), su(var)205 (33400), zuc (35227, 35228), piwi (33724), su(var)3-9 (33401), su(var)3-7 (35368), w (35573, 33644), rhino (34071), vasa (34950). α Tub-Gal4 is a kind gift from V.Mirouse. About 419 bp homologous to the Idefix gag coding region (1003–1422) were inserted either in a sense (pGgIds) or anti-sense orientation (pGgIdas) according to gfp transcription within the pUASp-gfp vector. Six independent transgenic lines were analyzed for pGgIds and pGgIdas.

Fluorescent In Situ Hybridization

Gfp probe were cloned into pGemt-easy using following primer gfp-probe_for: TAGATGGTGATGTTAATGGGC and gfp-probe_rev: GTTTGTATAGTTCATCCATGCC. FISH was done as follow: Ovaries were dissected in PBT (PBS-0.2% Tween) on ice, fixed with 4% formaldehyde/PBT at RT for 30 min, rinsed three times with PBT fixed with 4% formaldehyde/PBT at RT for 10 min and rinsed three times with PBT. After permeabilisation (1 h in PBS-0.3% triton at 300rpm) prehybridation was done as follow: 10 min HYB-(Formamide 50%, SSC 5X, Tween 0.02%, QSP H2O)/PBT 1:1 300rpm, 10 min Hyb- 300rpm, 1h HYB+ (yeast tRNA 0.1 μ g/ μ l, heparin 0.25mg/ml, QSP HYB-) 60°C 300rpm. Hybridization was done overnight at 60°C with 1ug of RNA UTP-Dig labeled probe. Ovaries were rinsed two time in HYB-, HYB-/PBT and PBT. Blocking was done with TNB

(Perkin-Elmer TSA kit) 1h and immunodetection with anti-Dig-HRP during 1.5 h in TNB 0.2% tween, rinsed three time in PBT and incubated 10 min with TSA-Cy3 in 1/25 amplification diluent (Perkin-Elmer). Then ovaries were rinsed three time and stain with DAPI. For RNA visualization RNaseH treatment for 30min was done before TSA amplification to destroyed RNA/DNA hybridization. For DNA staining visualization RNase A treatment for 1 hours was done before the second step of fixation to destroyed RNA.

Immuno-cytochemistry

Ovaries were dissected in PBS on ice, fixed with 4% formaldehyde/PBS at RT for 15 min and rinsed three times with PBT (PBS with 0.2%Triton X-100). After blocking in BBT (PBT with 0.1% BSA) for 2 h at RT, the ovaries were incubated with primary antibodies in BBT overnight at 4°C. After three PBT washes, secondary antibodies were incubated 2 h at RT. Primary antibodies used were α -Aub and α -Ago3 (1:500, rabbit; Brennecke *et al*, 2007), α -Piwi (1:1000, mouse; Saito *et al*, 2006), α -HP1 (1:100, mouse C1a9 from Developmental Studies Hybridoma Bank), α -Vasa (1:100, rat from Developmental Studies Hybridoma Bank) and α - H3K9-tri-methylation (#07-523 Millipore). GFP was viewed in whole-mount ovaries using the 488 nm filter set of the LEICA SP5 confocal microscope. Apparition experiment was done as previously describe in Dufourt *et al* 2011.

Microscope Analysis and treatment

Immunostaining analysis was performed on LEICA SP5 confocal microscope. 3D reconstruction and colocalisation was done by Imaris software. Script was used to count colocalisation (Sup data Fig. 8).

Chromatin immunoprecipitation assays

Chromatin immunoprecipitation (ChIP) was performed using the standard procedures, which was modified as follows: for each immunoprecipitation, three OD260 of chromatin in ChIP buffer was pre-incubated in the presence of 80 ml of Protein A-Agarose (PAA) beads (Millipore) for 30 min at 48C. PAA were removed, antibodies were added (a control in the absence of antibody named Input was included), and samples were incubated overnight at 48C in a rotating wheel. Then, 60 ml of PAA were added and incubation was continued for 1 h at 48C. Samples were washed in saline concentration buffer. Chromatin was eluted from PAA in 500 ml of elution buffer (sodium dodecyl sulfate and NaHCO₃) at room temperature for 15 min. The eluate was incubated overnight at 65°C to reverse cross-links and treated by proteinase K for 1 h at 50°C. Samples were phenol–chloroform extracted and ethanol precipitated. DNA was resuspended in 40 ml of 10 mM Tris (pH 7.5). Antibodies used include: H3K9-di-methylation #07-441, H3K9-tri-methylation #07-523, H3K14-acetylation #07-353, H3K4-di-methylation #07-030, H3K36-tri-methylation #07-549, normal Mouse IgG #12-371 and normal Rabbit IgG #12-370 from Millipore, Heterochromatic Protein 1a (HP1a) (C1A9) from DSHB. For primer and PCR analysis see QrtPCR description.

QrtPCR analysis

First strand cDNA was obtained by using random primers on Trizol extracted total ovarian RNA from 2- to 3-day-old flies. Quantitative PCR was performed using Roche FastStart SYBR Green Master on LC480. Steady-state RNA levels were calculated from the threshold cycle for amplification using the $2^{-\Delta \Delta C_T}$ method (Livak and Schmittgen, 2001). Throughout rp49 was used for the normalization Average levels and standard deviations were calculated from at least four biological replicates according to Livak and Schmittgen (2001). For RNAi

screening fold enrichments are in comparison with a pGIds/RNAi-*white* control sample. For each experiment control sample and rp49 were present to avoid technical error.

Primers for QPCR analysis were

rp49_for: GACGCTTCAAGGGACAGTATCTG

rp49_rev: AAACGCGGTTCTGCATGAG

Idefix_for : AAACAAATCGTGGCAGGAAG

Idefix_rev : GCTTCTTTGGTTGGTCTGGA

ZAM_for: CTACGAAATGGCAAGATTAATTCCACTTCC

ZAM_rev: CCCGTTTCCTTTATGTCGCAGTAGCT

I element_for: CAAAAACAACAATACCGCTAAT

I element_rev: AGCAGGTTGCCGTCTCTTGTA

HeT-A_for: CGCGCGGAACCCATCTTCAGA

HeT-A_rev: CGCCGCAGTCGTTTGGTGAGT

GFP_for : TACCTGTCCACACAATCTGC

GFP_rev : ATCCATGCCATGTGTAATCC

Acknowledgements:

This work was supported by grants from the Association pour la Recherche contre le Cancer (ARC 1139). J.D. received a graduate grant from the Ministère de l'Enseignement Supérieur et de la Recherche (MESR) and Association contre le cancer (ARC). EB and P.P. received a grant from the Région Auvergne. We thank M. Miller, F. Pellissier for technical assistance, Flyfacility for transgenic lines (www.Fly-facility.com), the Bloomington *Drosophila* Stock Center, V. Mirouse for vectors and drivers, M. Siomi and G. Hannon for antibodies.

Figure legends

Figure 1: Structure and silencing of *pGgIds* and *pGgIdas*.

(a) Structure of *pGgIds* and *pGgIdas*: The minimal promotor and the Gal4 target sequences (UASp), the *gfp* reporter gene and the *Idefix* portion are depicted by black, white and grey rectangles respectively. Arrows above the rectangle indicate orientation of *Idefix*. The two FRT sites flanking *Idefix* are indicated as black triangles. A black arrow indicates TSS. (b) An example of silencing targeting *pGgIds* and *pGgIdas* in ovaries. *gfp* silencing is presented at the RNA (first line) and protein (third line) levels. *gfp* RNA (second line) and protein (fourth line) are recovered when the *Idefix* fragment is excised (*pGg^{DIds}* and *pGg^{DIdas}*). Two drivers are tested: the ubiquitous driver *Actine-Gal4* (s1-as1) and the germline driver *αtub-Gal4* (s2-as2). No *gfp* expression in somatic cells is recovered when *αTub-Gal4* is used (white arrow). (c) Changes in steady-state levels of transcripts encoded by *Idefix*-sensors driven by the *αTub-Gal4*. *pGgIds* (1), *pGg^{DIds}* (2), *pGgIdas* (3), and *pGg^{DIdas}* (4). Experimental quadruplicate was done with primer in *gfp* sequences (see primer set in Materials and methods, mean ± SD and standart Ttest $p < 0.05^*$, $p < 0.01^{**}$, $p < 0.001^{***}$).

Figure 2: Comprehensive analysis of the involvement of silencing pathway members in the germline.

(a-d). Shown is GFP fluorescence as readout for *Idefix* silencing for genetically identified genes associated to silencing pathways. pValium22 RNAi lines were used, or pValium20 RNAi lines when pValium22 were unavailable. The transgene *pGgIds* was driven with *αTub-Gal4* (a) or *nos-Gal4* (d). (b, c, e, f). QrtPCR was performed for each RNAi background and normalized to RNAi against *white* (the blue line represents no enrichment) with *gfp* primer set. The transgene *pGgIds* was driven with *αTub-Gal4* (b, e) or *nos-Gal4* (c, f) (n=4 biological experiments. Error bars represent SEM)

Figure 3: Characteristics of the silencing exerted on *Idefix-sensors* in the germ cells.

a) Idefix sensors display an opened chromatin structure: Chromatin immunoprecipitation (ChIP) assay in *pGgIds* ovary driven by α *Tub-Gal4* normalised to IgG controls. Open chromatin marks (green) H3K14ac, H3K36me³, H3K4me², and no enrichment of close chromatin marks (red) H3K9me², H3K9me³ and heterochromatic protein HP1 are observed. (n=3 independent ChIP experiments). Error bars SD. (b) *Idefix-sensors* do not localize in enriched HP area: DNA immunoFISH in nurse cells (left panel) showing no co-localisation between HP1 (blue) and the *pGgIds* DNA sequence (red). Below: schematic representation of the distribution of red dots in blue area (dot in bright red and bright blue are co-localised). (c) *Idefix* sequence from *Idefix-sensors* is continuously required for their silencing in the germ cells: (first row) The Ct line [*HS-Flp*, *Actin.FRT.yellow.FRT.Gal4*, *UASp-gfp*] is used to determine the time necessary for the *FLP* recombinase to excise a DNA fragment flanked by FRT sequences after induction by heat shock, and to detect GFP expression from the flipped transgene. In the germ cells, detection of GFP starts at 7 hrs after the heat shock treatment (second panel) and a full expression is recovered at 24 hrs (fourth panel). When the *Idefix* fragment is flipped out from *pGgIds* or *pGgIdas*, GFP expression (Second and third row respectively) is detected at 7 hrs after treatment and a full expression recovered at 24 hrs.

Figure 4: Ping Pong assays

(a) qRTPCR was done in *pGgIds* (black bar) and *pGg*^{DIds} (grey bar) background driven by α tub-Gal4 (upper panel) or Act-Gal4 (lower panel) with primer set against *Idefix*, *I Element*, *ZAM* and *Heta* and normalized on *Rp49* transcript level (n=4 biological experiments, see primer set in Materials and methods, mean $\pm$ SD and standart Ttest p<0.05*, p<0.01**, p<0.001***). (b) GFP expression in *pGgIds* homozygote (upper panel), *pGgIdas* homozygote

(middle panel) or pGgIds/pGgIdas heterozygote (lower panel) line driven by homozygous Act-Gal4. GFP level in germline was indicated by arrowhead and in follicular cell by arrow.

(c) Focus on the region framed by a white rectangle in (b) lower panel. GFP signal was shown in green and DNA (Hoechst) in purple. A white arrow showed cells where GFP is always not repressed even in pGgIdas and pGgIds/pGgIdas heterozygote line. Schematic representation of GFP positive cell (green) in pGgIds/pGgIdas heterozygote background driven by Act-Gal4 (gsc: germline stem cell, cb: cystoblast, ssc: somatic stem cell, fc: follicular cell, nc: nurse cell).

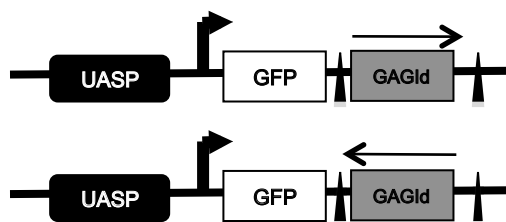
Figure 5: Sensitive germline window

(a) Immunocytochemical analysis of Aub, AGO3, Vasa and PIWI (white and red) protein localization in pGgIdas (GFP signal in green) background germlarium. (b) Immunocytochemical analysis of RCC, in pGgIdas background germlarium (GFP signal in green). H3K9me³ and HP1 (upper panels red and blue respectively). Immunocytochemical analysis of PIWI (red) and HP1 (blue) (lower panel). (c) Schematic structure of pTP transgene, UASP enhancer (black box) activate P element minimal promoter (arrow) and transcribe a fusion mRNA which give rise to a fusion proteins containing Tandem dimer Tomato fluorescent protein (white box) and PIWI protein (grey box). (d) Immunocytochemical analysis of GFP in pGgIdas (GFP signal in green) where pTP is expressed (red).

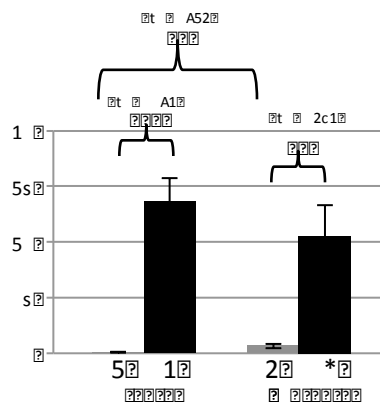
1. Aravin, A., et al., *A novel class of small RNAs bind to MILI protein in mouse testes*. Nature, 2006. **442**(7099): p. 203-7.
2. Girard, A., et al., *A germline-specific class of small RNAs binds mammalian Piwi proteins*. Nature, 2006. **442**(7099): p. 199-202.
3. Brennecke, J., et al., *Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila*. Cell, 2007. **128**(6): p. 1089-103.
4. King, R.C., *The meiotic behavior of the Drosophila oocyte*. Int Rev Cytol, 1970. **28**: p. 125-68.
5. Spradling, A.C., *Germline cysts: communes that work*. Cell, 1993. **72**(5): p. 649-51.
6. Olivieri, D., et al., *An in vivo RNAi assay identifies major genetic and cellular requirements for primary piRNA biogenesis in Drosophila*. EMBO J, 2010. **29**(19): p. 3301-17.
7. Saito, K., et al., *Roles for the Yb body components Armitage and Yb in primary piRNA biogenesis in Drosophila*. Genes Dev, 2010. **24**(22): p. 2493-8.
8. Lau, N.C., et al., *Abundant primary piRNAs, endo-siRNAs, and microRNAs in a Drosophila ovary cell line*. Genome Res, 2009. **19**(10): p. 1776-85.
9. Pane, A., K. Wehr, and T. Schupbach, *zucchini and squash encode two putative nucleases required for rasiRNA production in the Drosophila germline*. Dev Cell, 2007. **12**(6): p. 851-62.
10. Cook, H.A., et al., *The Drosophila SDE3 homolog armitage is required for oskar mRNA silencing and embryonic axis specification*. Cell, 2004. **116**(6): p. 817-29.
11. Lim, A.K. and T. Kai, *Unique germ-line organelle, nuage, functions to repress selfish genetic elements in Drosophila melanogaster*. Proc Natl Acad Sci U S A, 2007. **104**(16): p. 6714-9.
12. Pek, J.W., A.K. Lim, and T. Kai, *Drosophila maelstrom ensures proper germline stem cell lineage differentiation by repressing microRNA-7*. Dev Cell, 2009. **17**(3): p. 417-24.
13. Vagin, V.V., et al., *The RNA interference proteins and vasa locus are involved in the silencing of retrotransposons in the female germline of Drosophila melanogaster*. RNA Biol, 2004. **1**(1): p. 54-8.
14. Haase, A.D., et al., *Probing the initiation and effector phases of the somatic piRNA pathway in Drosophila*. Genes Dev, 2010. **24**(22): p. 2499-504.
15. Dufourt, J., et al., *Polycomb group-dependent, heterochromatin protein 1-independent, chromatin structures silence retrotransposons in somatic tissues outside ovaries*. DNA Res, 2011. **18**(6): p. 451-61.
16. Aravin, A.A., et al., *A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice*. Mol Cell, 2008. **31**(6): p. 785-99.
17. Bourc'h, D. and T.H. Bestor, *Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L*. Nature, 2004. **431**(7004): p. 96-9.
18. Brower-Toland, B., et al., *Drosophila PIWI associates with chromatin and interacts directly with HP1a*. Genes Dev, 2007. **21**(18): p. 2300-11.
19. Rangan, P., et al., *piRNA production requires heterochromatin formation in Drosophila*. Curr Biol, 2011. **21**(16): p. 1373-9.
20. Desset, S., et al., *In Drosophila melanogaster the COM locus directs the somatic silencing of two retrotransposons through both Piwi-dependent and -independent pathways*. PLoS One, 2008. **3**(2): p. e1526.
21. Rorth, P., *Gal4 in the Drosophila female germline*. Mech Dev, 1998. **78**(1-2): p. 113-8.

22. Yang, Z., et al., *HEN1 recognizes 21-24 nt small RNA duplexes and deposits a methyl group onto the 2' OH of the 3' terminal nucleotide*. Nucleic Acids Res, 2006. **34**(2): p. 667-75.
23. Horwich, M.D., et al., *The Drosophila RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC*. Curr Biol, 2007. **17**(14): p. 1265-72.
24. Lee, Y., et al., *The nuclear RNase III Drosha initiates microRNA processing*. Nature, 2003. **425**(6956): p. 415-9.

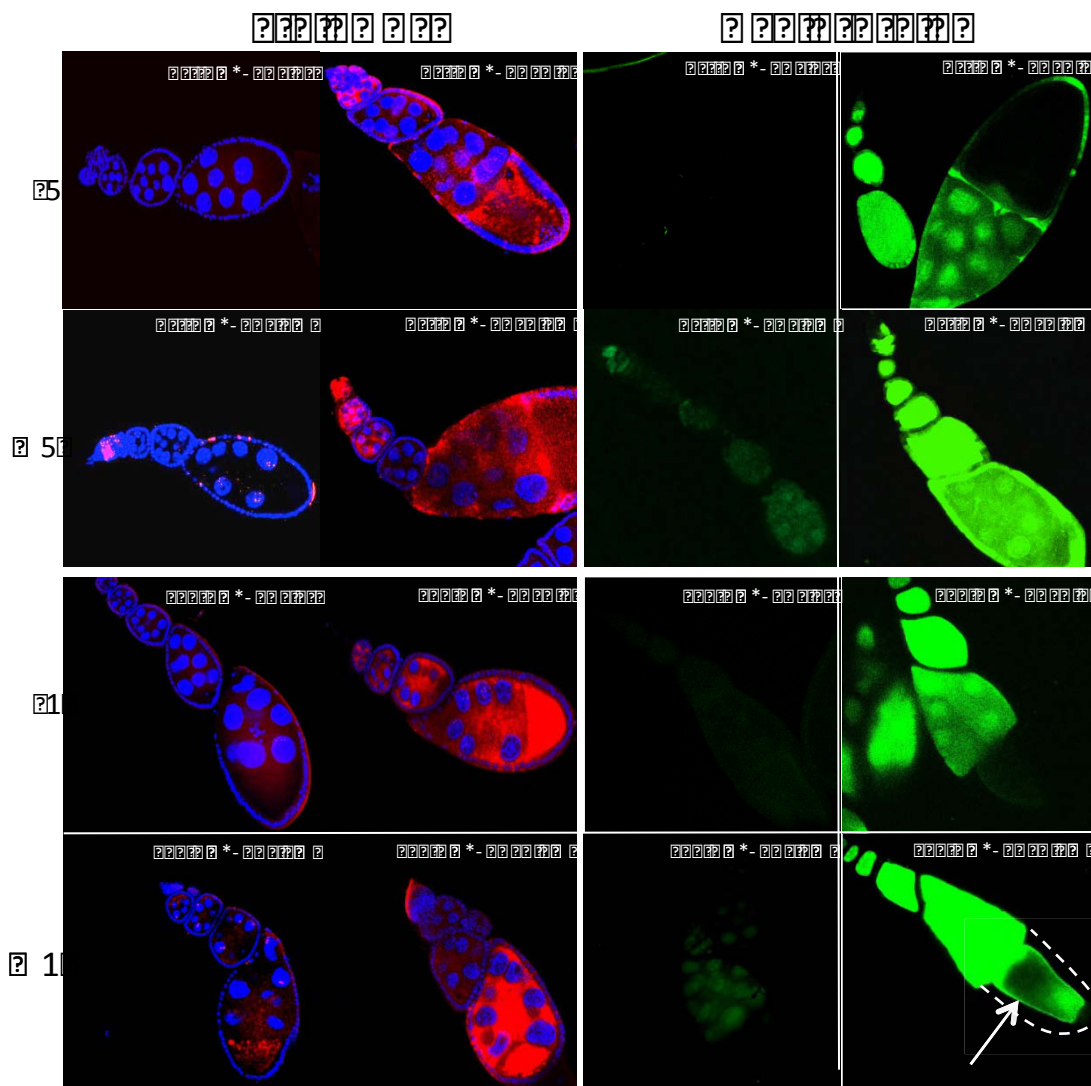
a



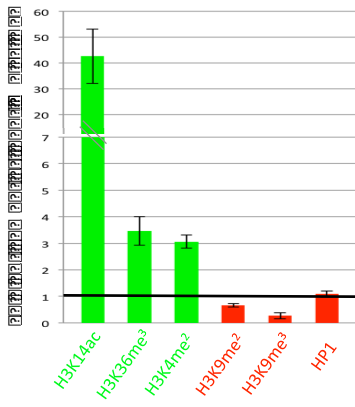
c



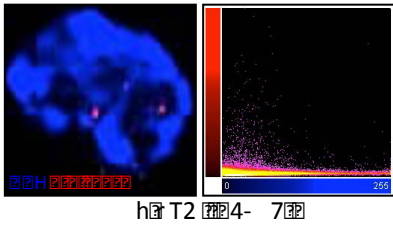
b



?

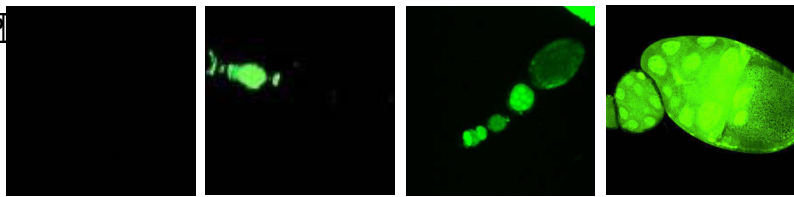


?

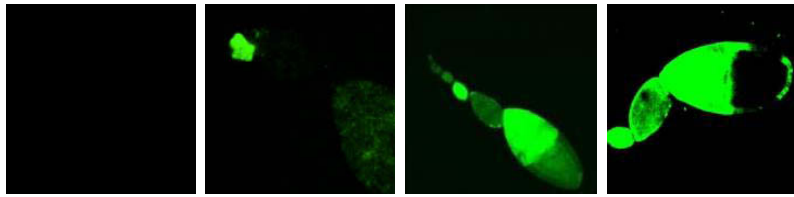


??

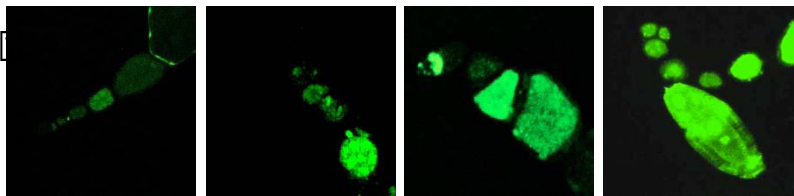
??



?



?

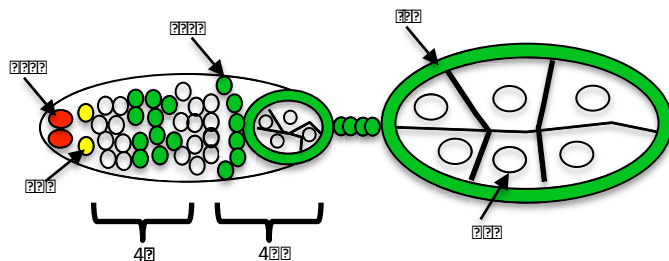
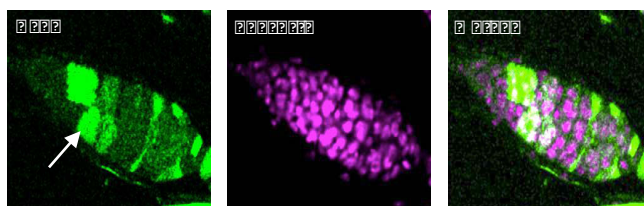
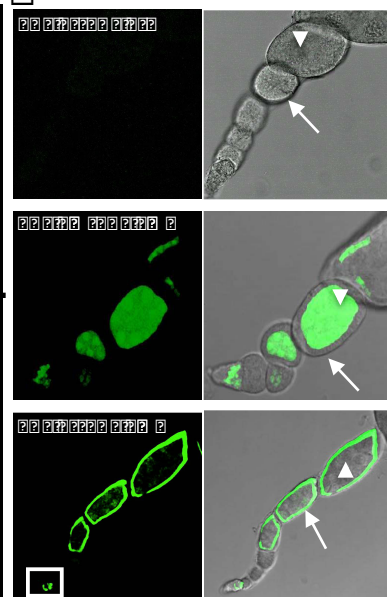
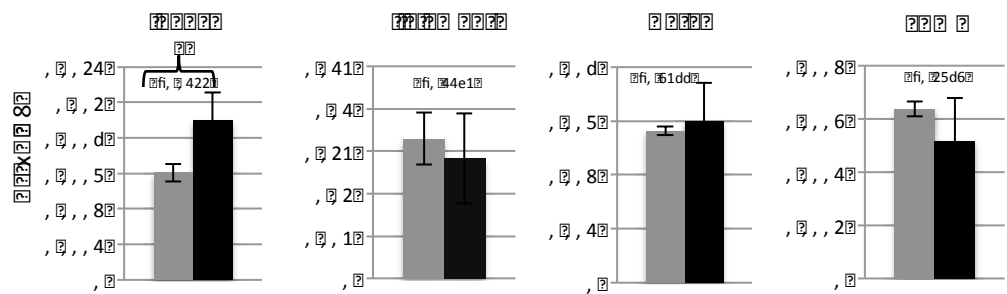
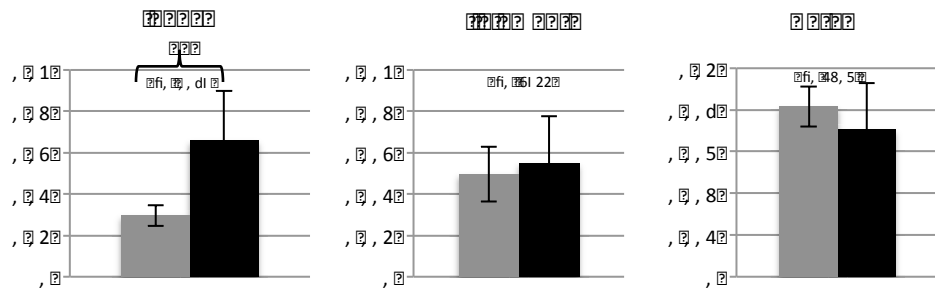


???

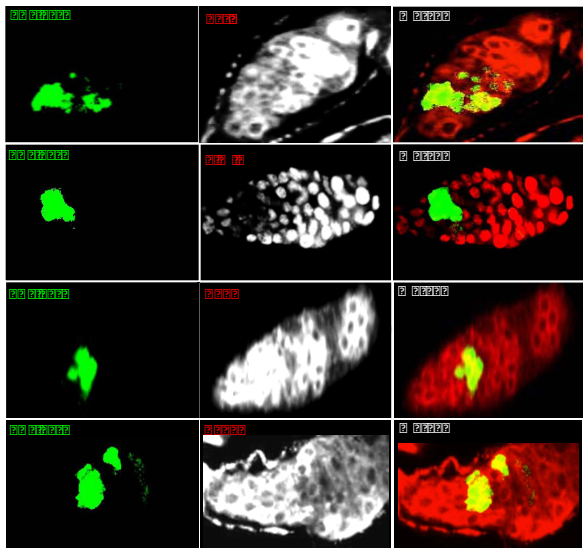
h???

H7???

ST???



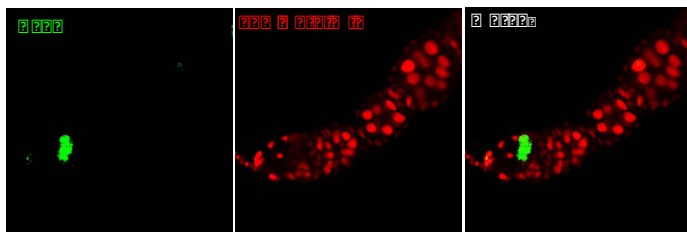
?



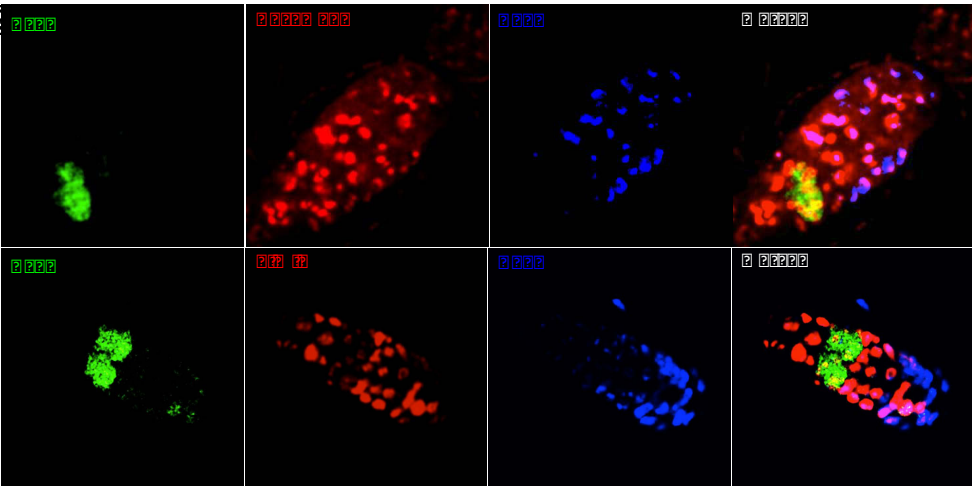
??



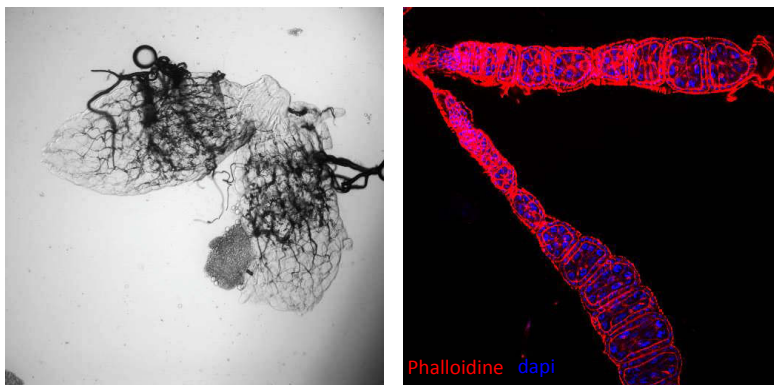
??



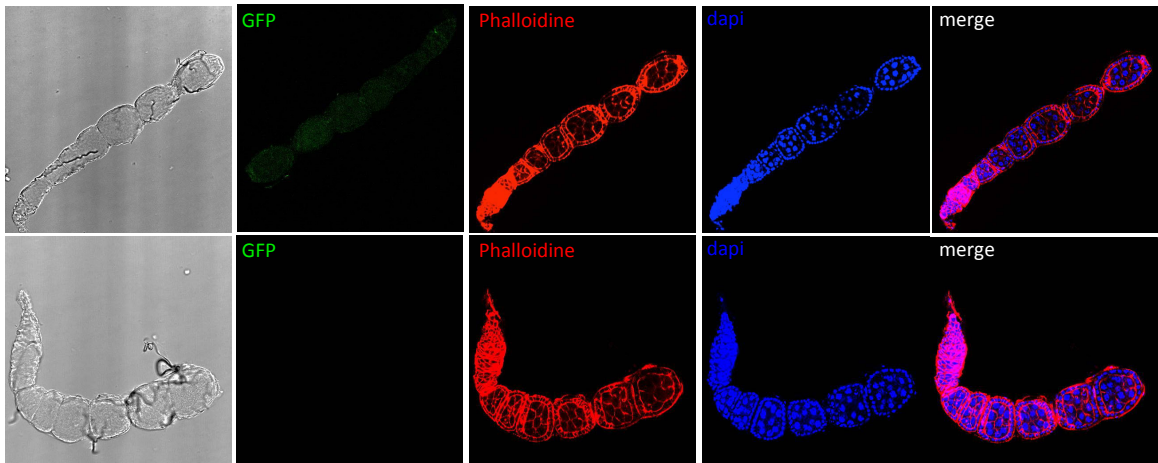
?



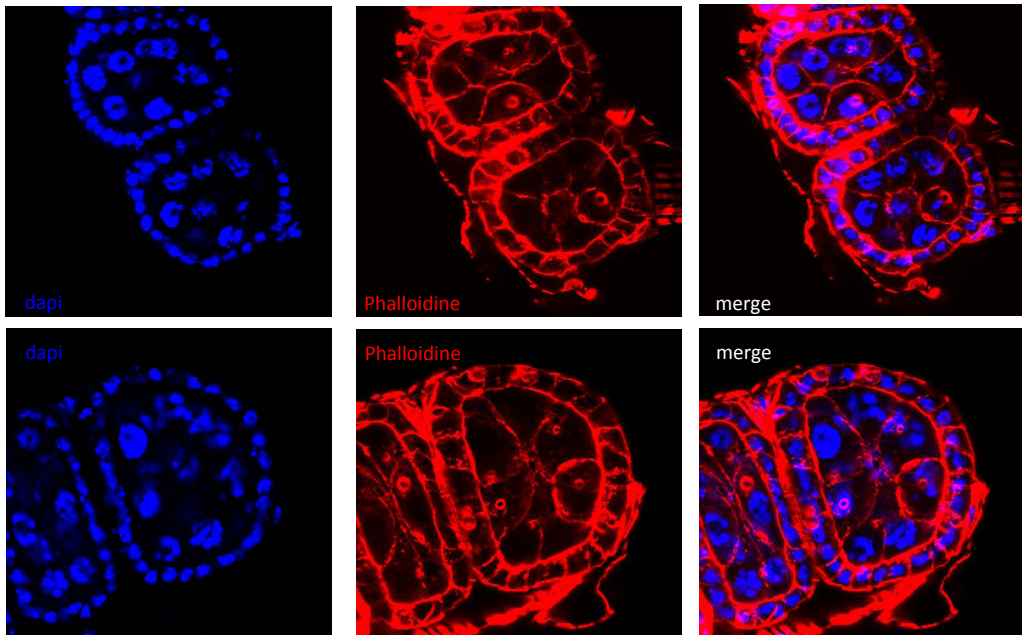
a



b



c

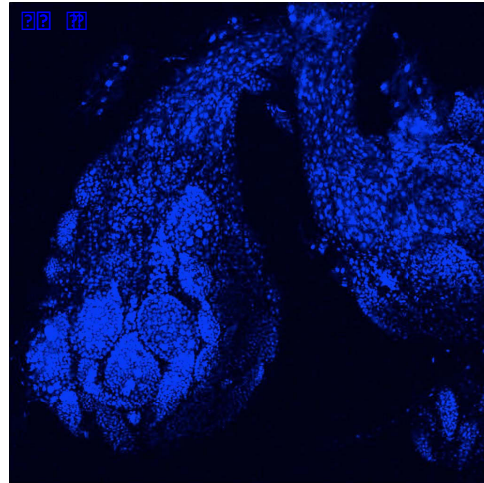
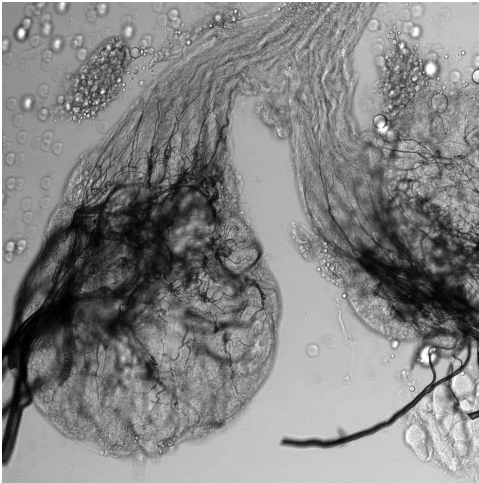


Supplemental data 1

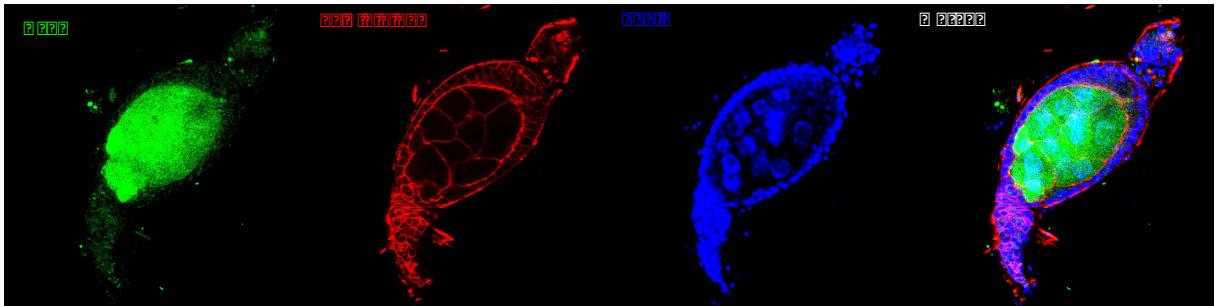
RNAi against Su(var)3.9 in pGgIds drive by α Tub-Gal4 background

- a) Ovaries overview (left), confocal analysis with phalloidin (red) and DAPI (blue) staining
- b) Confocal analysis with phalloidin (red) and DAPI (blue) staining and pGgIds GFP (green)
- c) Zoom on latest stages present in this background with phalloidin (red) and DAPI (blue) staining

?



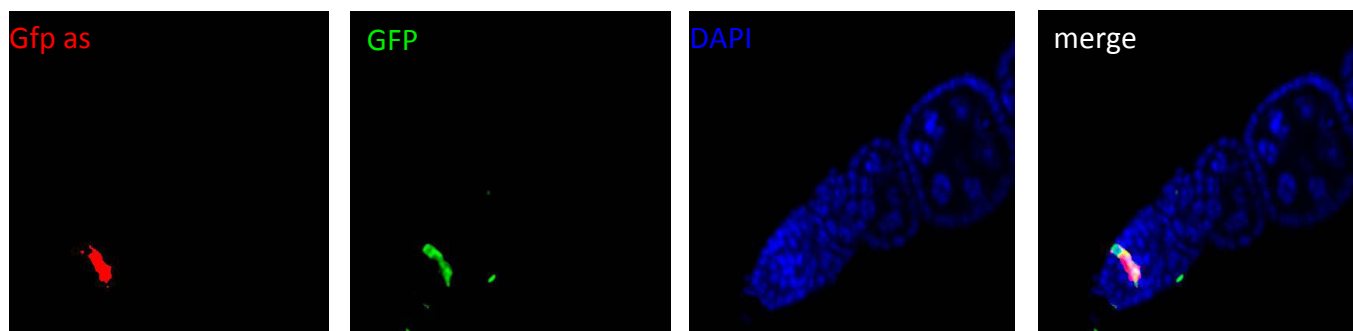
?



?????? ???? ? ? ?

RNAi against Su(var)205 in pGgIds drive by nos-Gal4 background

- a) Ovaries overview (left), confocal analysis with DAPI (blue) staining (right)
- b) Confocal analysis with phalloidin (red) and DAPI (blue) staining and pGgIds GFP (green)

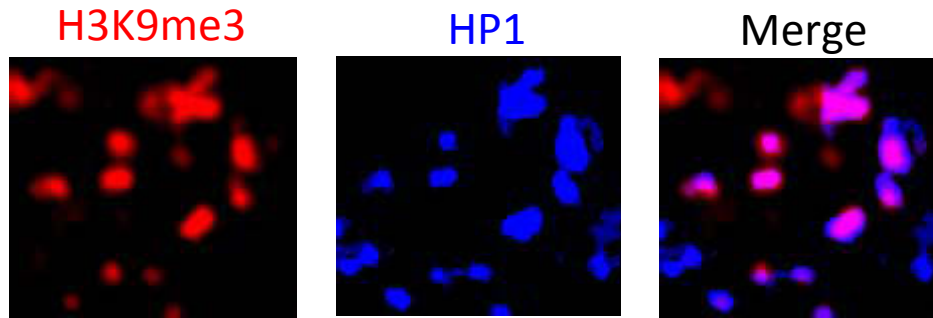


Supplemental Figure 3

RNA immuno-FISH of pGgIdas driven by Act-Gal4. Antisense *gfp* probe are visualize in red, GFP protein in green and DNA in blue.

Supplemental data Figure 4 : Germarium movie of pGgIdas2 GFP (green) and PIWI (red) staining.

Supplemental data Figure 5 : Germarium movie of pGgIdas2 GFP (green), PIWI (red) and HP1 (blue) staining



Supplemental Figure 6

Double Immunostaining between RCC with H3K9me3 antibody (red) and HP1 (blue) in *Drosophila* germarium. In purple are shown colocalisation.

Supplemental data Figure 7: Germarium movie of pGgIdas2 GFP (green), H3K9me3 (blue) and HP1 (red) staining

Supplemental Figure 8

Imaris script for localisation of pGgIds and pGgIdas DNA and HP1 proteins

```
% ImarisXTColoc
% A function to compute the percentage of colocalization
% between different cells.
function ImarisXTColoc_jeremy(aImarisID,varargin) % ID corresponds
to Imaris instance
vImarisServer = actxserver('ImarisServer.Server');
vImarisApplication = vImarisServer.GetObject(aImarisID);

%% Data cloning
aType='eTypeUInt16';
vImarisApplication.mDataSet.mType=aType;
aSurpassScene = vImarisApplication.mSurpassScene;
aDataSet=vImarisApplication.mDataSet;% Clone; %% Current dataset
aBackgroundSubtraction=1;

%% Size variables
sizeX = vImarisApplication.mDataSet.mSizeX;
sizeY = vImarisApplication.mDataSet.mSizeY;
sizeZ = vImarisApplication.mDataSet.mSizeZ;
sizeC = vImarisApplication.mDataSet.mSizeC;
extendMinX = aDataSet.mExtendMinX;
extendMinY = aDataSet.mExtendMinY;
extendMinZ = aDataSet.mExtendMinZ;
extendMaxX = aDataSet.mExtendMaxX;
extendMaxY = aDataSet.mExtendMaxY;
extendMaxZ = aDataSet.mExtendMaxZ;

% Vector "numberofspots" initialization
%aNumberOfSpots = zeros((sizeC-1)*(SizeC-2)/2 + 1,1);
aNumberOfSpots = {'\','Nombre'};

%% User interface
% We have the user select which channel corresponds to the DNA
numlines=1;
name='Colocalization parameters';
prompt={'Enter the "DNA Channel" number:'};
defaultanswer={num2str(sizeC)};
prompt{2} = 'Enter the "Blue Channel" number:';
defaultanswer{2}='1';
prompt{3} = 'Enter the "Red Channel" number:';
defaultanswer{3}='2';
prompt{4} = 'Dot size:';
defaultanswer{4}='2';
prompt{5} = 'Size of small cells:';
defaultanswer{5}='6';
prompt{6} = 'Min Volume of nurses: (um^3)';
defaultanswer{6}='10000.00';
prompt{7} = 'Max Volume of nurses: (um^3)';
```



```

defaultanswer{7}='100000.00';
prompt{8} = 'Keep only nurses ? (1 = yes, 0 = no (small cells))';
defaultanswer{8}='1';
prompt{9} = 'Automatic spots ?';
defaultanswer{9}='1';

% We retrieve the answers
answer=inputdlg(prompt,name,numlines,defaultanswer);
if isempty(answer)
    return;
end
dapi=floor(str2double(answer{1})-1); % DNA Channel
blue=floor(str2double(answer{2})-1); % Blue Channel
red=floor(str2double(answer{3})-1); % Red Channel
dotsize=str2double(answer{4});
size=str2double(answer{5});
volumeMin=answer{6};
volumeMax=answer{7};
big = floor(str2double(answer{8}));
auto = floor(str2double(answer{9}));

% We rectify incorrect answers
if dapi < 0
    dapi = 0;
else if dapi >= sizeC
    dapi = sizeC - 1;
end
end
if red < 0
    red = 0;
else if red >= sizeC
    red = sizeC - 1;
end
end
if blue < 0
    blue = 0;
else if blue >= sizeC
    blue = sizeC - 1;
end
end
if big > 1
    big = 1;
else if big < 0
    big = 0;
end
end
if auto > 1
    auto = 1;
else if auto < 0
    auto = 0;
end
end

%% Dapi

```



```

aDataBlue1D = aDataSet.GetDataVolumeAs1DArray(blue,0);
aDataRed1D = aDataSet.GetDataVolumeAs1DArray(red,0);
aDataDapi1D = aDataSet.GetDataVolumeAs1DArray(dapi,0);
aDapi =
vImarisApplication.mImageProcessing.DetectSurfaces(aDataSet,[],dapi,
2,size,1,0,'"Volume" above 10 um^3');
aMaskDNA =
aDapi.GetMask(extendMinX,extendMinY,extendMinZ,extendMaxX,extendMaxY
,extendMaxZ,sizeX,sizeY,sizeZ,0);
maskDNA = aMaskDNA.GetDataVolumeAs1DArray(0,0);
maskDNA = uint16(maskDNA);
aDataBlue1D = aDataBlue1D .* maskDNA;
aDataRed1D = aDataRed1D .* maskDNA;
aDataDapi1D = aDataDapi1D .* maskDNA;
aDapi.mName = 'DAPI';
aDapi.mVisible = false;
aSurpassScene.AddChild(aDapi);

%% Big Blue
if big == 1
    aSurface =
vImarisApplication.mImageProcessing.DetectSurfaces(aDataSet,[],dapi,
2,size,1,0,['"Volume" between ' volumeMin ' um^3 and ' volumeMax '
um^3']);
    aSurface.mName = 'Nurses';
    aSurface.mVisible = false;
    aSurpassScene.AddChild(aSurface);
    aMask =
aSurface.GetMask(extendMinX,extendMinY,extendMinZ,extendMaxX,extendM
axY,extendMaxZ,sizeX,sizeY,sizeZ,0);
    mask = aMask.GetDataVolumeAs1DArray(0,0);
    mask = uint16(mask);
    aDataBlue1D = aDataBlue1D .* mask;
    aDataRed1D = aDataRed1D .* mask;
    aDataDapi1D = aDataDapi1D .* mask;
end
aDataBlue = reshape(aDataBlue1D,sizeX,sizeY,sizeZ);
aDataRed = reshape(aDataRed1D,sizeX,sizeY,sizeZ);
aDataDapi = reshape(aDataDapi1D,sizeX,sizeY,sizeZ);
aDataSet.SetDataVolume(uint16(aDataBlue),blue,0);
aDataSet.SetDataVolume(uint16(aDataRed),red,0);
aDataSet.SetDataVolume(uint16(aDataDapi),dapi,0);

%% Small Red
% aData1D = aDataSet.GetDataVolumeAs1DArray(red,0);
% histo = histc(single(aData1D), 0:255);
% P = histo / sum(histo);
% P2 = P .* (1:256);
% omega0 = zeros(256, 1);
% omega1 = omega0;
% mu0 = omega0;
% mu1 = omega0;
% for i=1:256
% omega0(i) = sum(P(1:i));

```



```

% omega1(i) = sum(P(i+1:256));
% mu0(i) = sum(P2(1:i)) ./ omega0(i);
% mu1(i) = sum(P2(i+1:256)) ./ omega1(i);
% end
% sigma = omega0 .* omega1 .* (mu1 - mu0) .* (mu1 - mu0);
% [~, thresh] = max(max(sigma));
%vImarisApplication.mImageProcessing.MedianFilterChannel(aDataSet, red, [3 3 1]);
%vImarisApplication.mImageProcessing.ThresholdChannel(aDataSet, red, thresh, 0);

% Segmentation
[aR,aG,aB,aA]=aDataSet.GetChannelColor(red);
nameC = aDataSet.GetChannelName(red);
if auto == 1
    aSpots{1} =
vImarisApplication.mImageProcessing.DetectSpots2(aDataSet,[],red,dot
size,aBackgroundSubtraction,['"Quality" above automatic threshold
"Intensity Center Ch=' num2str(blue+1) '" above 120.000']);
else
    aSpots{1} =
vImarisApplication.mImageProcessing.DetectSpots2(aDataSet,[],red,dot
size,aBackgroundSubtraction,['"Intensity Center Ch=' num2str(red+1)
'" above 120.000' ' "Intensity Center Ch=' num2str(blue+1) '" above
120.000']);
end
aSpots{1}.mName = 'Coloc';
aSpots{1}.SetColor(aR,aG,aR,aA);
aNumberOfSpots{1,1} = 'Nb of red dots in blue';
aNumberOfSpots{1,2} = int2str(length(aSpots{1}.GetRadii));
if auto == 1
    aSpots{2} =
vImarisApplication.mImageProcessing.DetectSpots2(aDataSet,[],red,dot
size,aBackgroundSubtraction,'"Quality" above automatic threshold');
else
    aSpots{2} =
vImarisApplication.mImageProcessing.DetectSpots2(aDataSet,[],red,dot
size,aBackgroundSubtraction,['"Intensity Center Ch=' num2str(red+1)
'" above 120.000']);
end
aSpots{2}.mName = nameC;
aSpots{2}.SetColor(aR,aG,aB,aA);
aNumberOfSpots{2,1} = 'Nb of red dots';
aNumberOfSpots{2,2} = int2str(length(aSpots{2}.GetRadii));
aSurpassScene.AddChild(aSpots{1});
aSurpassScene.AddChild(aSpots{2});

%% Conclusion
aType='eTypeUInt8';
vImarisApplication.mDataSet.mType=aType;
display(aNumberOfSpots);

end

```


A)

	Silencing in nurse cell	Number of sense piRNAs in w1118 Ovaries	Number of antisense piRNAs in w1118 Ovaries	Number of sense piRNAs 3mm in w1118 Ovaries	Number of antisense piRNAs 3mm in w1118 Ovaries
Idefix	+	220	1114	703	3224
Gag fragment from pGgId	+	12	104	48	289

B)

	Silencing in nurse cell	Number of sense piRNAs from w1118 Embryo 0-2hrs	Number of antisense piRNAs from w1118 Embryo 0-2hrs	Number of sense piRNAs 3mm from w1118 Embryo 0-2hrs	Number of antisense piRNAs 3mm from w1118 Embryo 0-2hrs
Idefix	+	47	299	175	800
Gag fragment from pGgId	+	3	18	11	60

Figure 84: piRNA ciblant Idefix

A) Quantité de piRNAs ciblant, l'élément Idefix complet (ligne du haut) ou le fragment inséré dans les transgènes pGgIds et pGgIdas (ligne du bas). Les piRNAs ont été extraits du séquençage des petits ARNs des ovaires de la lignée de drosophile W1118 du papier Brennecke et al 2008 grâce au logiciel Nucbase. **B)** Quantité de piRNAs ciblant, l'élément Idefix complet (ligne du haut) ou le fragment inséré dans les transgènes pGgIds et pGgIdas (ligne du bas). Les piRNAs ont été extraits du séquençage des petits ARNs des embryons précoces (0-2h) de la lignée de drosophile W1118 du papier Brennecke et al 2008 grâce au logiciel Nucbase. L'orientation est définis en fonction de l'orientation de l'ARNm d'Idefix. les transgènes pGgIds et pGgIdas ont été inséré dans cette même lignée W1118. (3mm: 3 mismatch autorisé)

2) Conclusions

Nous avons montré dans une étude précédente que les éléments transposables étaient sujet à une régulation de type PTGS dans les cellules somatiques qui entourent la lignée germinale dans l'ovaire de drosophile, alors que dans les autres des tissus somatiques de drosophile une régulation impliquant TGS et PTGS est impliquée. Afin de savoir si un mécanisme de type PTGS ou TGS régule les éléments transposables situés dans l'euchromatine et potentiellement actifs dans la lignée germinale de drosophile, nous avons construit des lignées de drosophile contenant des transgènes pGgIds et pGgIdas sur le modèle des lignées contenant les transgènes tGgIds et tGgIdas. En effet ces derniers ne peuvent pas s'exprimer dans la lignée germinale (Rorth 1998; Dufourt, Brasset et al. 2011)(Desset, Buchon et al. 2008). En exprimant le transgène pGgIds soit de manière ubiquitaire (Act-Gal4) soit spécifiquement dans la lignée germinale (α Tub-Gal4) nous avons pu observer que celui-ci est sujet à une répression. Cette répression peut s'observer aussi bien sur les ARNs du transgène que sur l'expression de la protéine GFP. Le fragment d'Idefix permet cette répression car dans les lignées tGg^{DIds} où ce fragment a été excisé grâce aux séquences FRT, on recouvre l'expression du transgène aussi bien au niveau ARN que protéique. Avec le transgène pGgIdas une répression similaire peut être observée avec cependant une légère nuance. En effet en présence du fragment antisens d'Idefix, il existe bien une répression au niveau ARN et protéique mais celle-ci est plus faible qu'en présence du fragment sens. Ceci a été observé pour huit sites d'insertions différents aussi bien en sens qu'en antisens, cette diminution de la répression étant plus particulièrement observée dans les stades précoces.

La différence de répression peut être corrélée à la quantité de piRNAs sens et antisens pouvant cibler le fragment d'Idefix. En effet il y a quantitativement plus de piRNAs antisens que de sens (**Figure 84**) ce qui peut entraîner une plus forte répression du

fragment d'Idefix orienté en sens. De plus une accumulation des piRNAs au cours de la croissance des follicules pourrait expliquer que la répression en antisens semble complète dans les stades plus tardifs.

Finalement on peut observer une faible quantité de transcrits par rt-QPCR avec des amorces amplifiant une région du gène de la GFP, indiquant que la transcription des transgènes n'est pas complètement verrouillée en présence du fragment d'Idefix, aussi bien en sens qu'en antisens.

En conclusion ces lignées transgéniques subissent une répression grâce à la présence d'un fragment d'élément transposable et peuvent servir d'outils afin d'étudier les régulations germinales des éléments transposables euchromatiques.

Afin de savoir quelles pouvaient être les voies de régulations permettant de mettre sous silence Idefix, nous avons effectué un crible RNAi de divers gènes codant des protéines pouvant être impliquées dans la régulation des éléments transposables. Les résultats ont montré que trois classes de protéines sont impliquées dans la mise sous silence du transgène : les protéines Argonaute de type PIWI, les protéines présentes dans le nuage et certaines protéines hétérochromatiques. On peut organiser ces trois classes en deux catégories : d'une part les protéines impliquées précocement dans la régulation et non requises tout au long du développement du follicule, et d'autre part les protéines requises durant tout le développement folliculaire. Les protéines PIWI, AGO3, HP1, Cuff, Rhino et Maelstrom sont requises précocement dans la répression et leur présence est aussi nécessaire tout au long du développement du follicule. En effet en exprimant les constructions RNAi dirigées contre les ARNm des gènes codant ces protéines de manière tardive (α Tub-Gal4) ou précoce (nos-Gal4) dans le développement du follicule, on peut observer une dérégulation du transgène pGgIds. Il peut paraître aisé de comprendre

pourquoi la protéine Ago3 est continuellement requise du fait que c'est la seule protéine argonaute qui permette de maintenir le Ping-Pong en s'associant aux piRNAs secondaires.

La maturation des grands transcrits provenant des clusters de piRNAs pourrait être en partie réalisée par Squash (Pane, Wehr et al. 2007), grâce à son activité endonucléase afin de donner des précurseurs des piRNAs. Ceux-ci une fois fixés par la protéine PIWI seraient maturés par une protéine à activité exonucléase 3'-5' qui pourrait être Maelstrom. En effet c'est la seule protéine connue présente dans le nuage et ayant une putative activité exonucléase (Zhang, Xiong et al. 2008). La maturation des précurseurs liés à PIWI grâce à Maelstrom pourrait expliquer le fait que ces protéines soient requises tout au long du développement du follicule.

En ce qui concerne les trois autres protéines qui sont requises tout au long du développement du follicule, il s'agit de protéines connues pour leur présence sur l'hétérochromatine. Cuff et Rhino semblent importantes pour l'activation des clusters de piRNAs et sont capables d'interagir. Ces protéines permettraient donc un maintien de l'activité des clusters de piRNAs tout au long du développement de la lignée germinale (Klattenhoff, Xi et al. 2009; Pane, Jiang et al. 2011). La protéine HP1 a largement été impliquée dans la régulation des éléments transposables en lignée germinale (Klenov, Lavrov et al. 2007; Shpiz, Olovnikov et al. 2011; Wang and Elgin 2011). Ces équipes ont montré que les éléments transposables, en ovaires, sont recouverts par une structure chromatinienne contenant HP1 et H3K9me³. En revanche leurs études n'ont pas permis de discriminer si cette couverture par HP1 était présente sur les éléments euchromatiques ou sur les éléments hétérochromatiques et donc potentiellement sur les clusters de piRNAs. Très récemment il a été montré que des foyers contenant dSetdb1, H3K9me³, H4K20me³ et HP1 étaient nécessaires à l'activation des clusters de piRNAs. Ces "repressive chromatin centers" (RCC) seraient les centres activateurs des clusters de piRNAs

(Rangan, Malone et al. 2011). L'absence de HP1 pourrait donc diminuer l'activation de ceux-ci. Les protéines habituellement connues comme permettant la répression de l'activité transcriptionnelle seraient dans ce cas là des activateurs de la transcription dans les régions hétérochromatiques.

Contrairement aux différentes protéines décrites précédemment, les protéines Aub, Vasa, Krimp, Squash et Spn-E semblent être requises précocement mais pas continuellement au cours du développement du follicule ovarien. Aucun effet n'est observé si le RNAi est exprimé tardivement (α tub-Gal4) alors que l'extinction précoce (nos-Gal4) des ARNm des gènes codant ces protéines entraîne une dérégulation du transgène pGgIds. Vasa et Aubergine sont des protéines clef dans la formation du nuage (Findley, Tamanaha et al. 2003). En effet leur présence semble indispensable pour la mise en place de cette structure. La présence de diméthyl-arginines (sDMAs) sur ces protéines peut permettre le recrutement plus tardif des protéines à domaines Tudor dans le nuage. Il a aussi récemment été montré qu'un complexe impliquant Vasa, Aub et Spn-E est requis lors des premières divisions mitotiques des cellules germinales. En effet ces trois protéines seraient à l'origine du bon déroulement de la condensation des chromosomes mitotiques en permettant le recrutement de la protéine Barren, une sous unité du complexe condensin I, sur les régions péri-centromériques (Pek and Kai 2011). On peut donc proposer que ces protéines pourraient être activateurs des clusters de piRNAs durant les mitoses (région I à IIa du germarium), expliquant ainsi l'importance précoce des protéines Vasa, Aub et Spn-E.

Les protéines Squash et Zucchini ont dans un premier temps été montrées comme des nucléases ARNs importantes dans la production des rasi-RNAs. Mais récemment il a été suggéré que Zucchini serait une phospholipase plutôt qu'une nucléase et que son rôle serait majoritairement restreint aux cellules folliculaires. La seule endonucléase importante pour la maturation des piRNAs en lignée germinale pourrait donc être Squash. Ceci

expliquerait le fait que le RNAi dirigé contre *zuc* n'ait pas d'effet contrairement à celui dirigé contre *squ* dans la lignée germinale.

La protéine Krimper est une composante majeure du nuage, cette protéine à domaine Tudor est fortement exprimée dans les cellules souches germinales et le germarium. Sa localisation dans le nuage est dépendante de la présence de Aub et Vasa mais pas de celle de Maelstrom. En revanche Krimper est nécessaire à la bonne localisation de Maelstrom (Lim and Kai 2007). Son rôle pourrait donc être important dans la formation du nuage et la localisation d'autres protéines expliquant son implication précoce dans la régulation du transgène.

Il est intéressant de remarquer que la protéine Armitage semble requise pour le maintien de la répression des ARNm du transgènes pGgIds aussi bien précocement que tardivement. En revanche aucune expression de GFP n'a été observée. En détruisant l'ARNm du gène codant la protéine Armi il y a donc une dérégulation au niveau transcriptionnel, mais une régulation de type traductionnelle semble persister. On peut envisager que les ARNm du transgène subissent une répression de leur traduction. En effet en 2004, il a été montré que la protéine Armitage est requise pour la bonne formation d'un complexe de type RISC (Tomari, Du et al. 2004), son absence pourrait donc entraîner un mauvais alignement du RISC sur sa cible. Ce mauvais alignement entraînant non pas une dégradation des ARNm mais une modulation de la traduction. Une deuxième hypothèse peut être envisagée, la protéine Armitage pourrait être requise pour la sortie du noyau des ARNm des éléments transposables. En effet les récentes études sur la localisation de cette protéine, en cellules folliculaires, montrent qu'elle semble localisée très proche de la membrane nucléaire (Lau, Robine et al. 2009; Olivieri, Sykora et al. 2010; Saito, Ishizu et al. 2010). Afin d'en savoir un peu plus sur le rôle de la protéine Armitage des expériences

d'hybridation *in situ* fluorescente sur les ARNm des transgènes, dans un contexte mutant pour cette protéine, sont en cours.

Finalement de façon très surprenante, en exprimant un transgène RNAi, tardivement (α Tub-Gal4), contre les ARNm du gène *su(var)3-9*, codant la HMT permettant la méthylation H3K9, nous avons pu remarquer un défaut d'ovogénèse. Ce phénotype semble correspondre à un arrêt de l'ovogénèse, entraînant une accumulation des follicules au stade 4-5. Ce résultat est surprenant car les études décrites précédemment (Yoon, Lee et al. 2008) et observées au laboratoire sur des mutants du gène *su(var)3-9*, ne rapportent pas de défaut phénotypiques visibles.

Nous avons donc mis en évidence que trois classes de protéines étaient importantes pour la régulation des transgènes en lignée germinale et qu'on pouvait les classer en deux catégories : les protéines requises précocement et celles qui sont requises tout au long du développement du follicule. Les protéines présentes dans chacune des catégories semblent correspondre à la hiérarchie des protéines impliquées dans la formation du nuage comme montré par l'équipe de Siomi MC (Nagao, Sato et al. 2011).

De nombreuses études ont suggéré que les éléments transposables dans les tissus germinaux étaient régulés grâce à du TGS couplé à du PTGS, nous nous sommes donc intéressés aux modifications de la chromatine qui recouvrait les transgènes pGgIds. Nous avons pu montrer que les transgènes se trouvaient dans une structure de type chromatine ouverte avec un enrichissement en marques telles que H3K4me² et H3K14ac. L'enrichissement en marque H3K36me³ indique qu'il y a une transcription active. Parallèlement la présence de la protéine HP1 ainsi que de marques de chromatine fermée telles que H3K9me³ et H3K9me² n'a pas pu être décelée. Afin de confirmer cette absence de régulation transcriptionnelle nous avons utilisé la technique de réapparition de la GFP

(Dufourt, Brasset et al. 2011). Avec cette technique nous avons vu qu'il n'existe pas de délai dans l'apparition de la GFP lorsqu'on excise le fragment d'Idefix du transgène comparé à son contrôle.

Ces résultats, couplés au fait que des protéines impliquées dans la voies des piRNAs sont continuellement requises au cour du développement démontrent que le transgène pGgIds utilisé qui est le reflet de la régulation que peuvent subir les éléments transposables euchromatiques et donc potentiellement actifs, est régulé uniquement par du PTGS.

L'équipe de Ruth Lehmann propose que : « l'hétérochromatine protège la lignée germinale en activant les clusters de piRNAs », et nous proposons pour compléter que seule une mise sous silence post-transcriptionnelle protège la lignée germinale des éléments transposables euchromatiques. La perte de HP1 sur les éléments transposables observée par certaines équipes et interprétée comme une régulation de type TGS (Klenov, Lavrov et al. 2007; Shpiz, Olovnikov et al. 2011; Wang and Elgin 2011) pourrait être en réalité le reflet d'une mauvaise activation des clusters de piRNAs en absence de cette protéine. Ces clusters étant majoritairement composés de séquences d'éléments transposables, la perte de HP1 sur ceux-ci entrainerait leur extinction. Contrairement à cela, la perte de la régulation de type PTGS des éléments transposables euchromatiques entrainerait une dérégulation de ceux ci. En conclusion, seule l'utilisation de séquences spécifiques comme les transgènes utilisés lors de cette étude permet de discriminer la régulations mise en place sur les éléments répétés euchromatiques de celle des éléments hétérochromatiques.

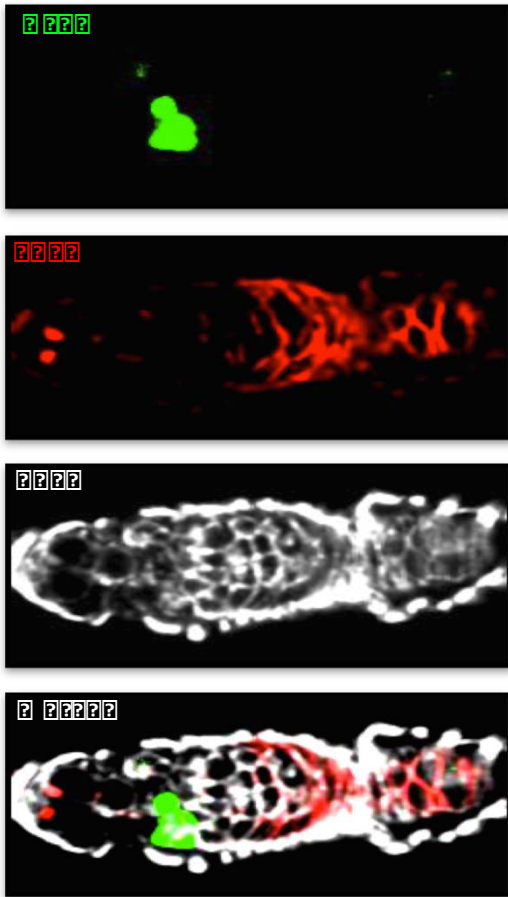
La régulation des transgènes pGgIds étant de type PTGS, on peut imaginer que les ARNm dégradés de ceux-ci pourraient entrer dans le mécanisme Ping-Pong et permettre d'amplifier cette régulation. En effet, le transgène sens pourrait être dégradé en piRNAs sens qui seraient pris en charge par Ago3 et ainsi capables d'amplifier le Ping-Pong dans la

lignée germinale. La présence du transgène sens est capable d'augmenter la répression du transgène antisens ainsi que de diminuer la quantité de transcrits d'Idefix endogènes mais pas celle d'autres éléments tels que Zam, Heta et I. Cette démonstration moléculaire vient confirmer le modèle Ping-Pong qui avait jusque-là été démontré uniquement *via* le séquençage des piRNAs. De manière aussi intéressante, on a pu observer que le transgène pGgId sens ne peut éteindre le transgène antisens dans les cellules folliculaires. Ceci pourrait être dû à l'absence d'Ago3 et donc de mécanisme d'amplifications des piRNAs (Ping-Pong) dans ces cellules.

En présence du transgène sens on a pu observer quelques cellules germinales Vasa positives dans le germarium dans lesquelles on n'observait pas de perte de l'expression de la GFP. Ces cellules ne sont jamais éteintes dans la lignée pGgIdas et sont situées dans les régions I et IIa du germarium. On peut observer que cette dérégulation concerne 4 à 16 cellules. Les protéines Vasa, Aub et Ago3 sont exprimées dans la région concernée par la dérégulation, en revanche la protéine PIWI y est faiblement exprimée. La région où la dérégulation est observée correspond aux cystes en division mitotique. En effet il a été décrit que lors de la division asymétrique des cellules souches germinales (GSC), il va y avoir formation d'un cystoblaste, où PIWI est absent, ainsi que d'une nouvelle cellule souche exprimant PIWI. L'expression de PIWI va ensuite être recouverte progressivement durant les divisions mitotiques et retrouver son niveau maximal au stade 16 cellules (Cox, Chao et al. 2000).

De manière assez surprenante des dérégulations de rapporteurs de l'expression d'éléments transposables dans le germarium sont décrites dans la littérature. En effet, lors d'une répression d'un transgène de type pLacZ par du TSE (Trans Silencing Effect) on peut observer une répression exceptée dans le germarium (Josse, Teyssset et al. 2007; Josse, Maurel-Zaffran et al. 2008; Todeschini, Teyssset et al. 2010). Plus récemment un transgène

A) Act-Gal4>pGgIdas



B) Act-Gal4> pGg^{DId}

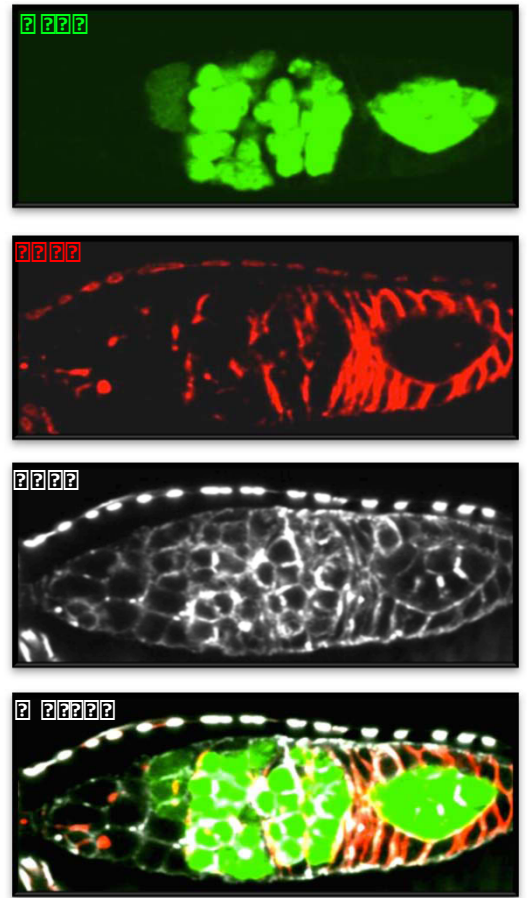


Figure 85: Expression des transgènes pGgIdas et pGg^{DId} avec un activateur ubiquitaire Act-Gal4
A) Le transgène pGgIdas présente une faiblesse de répression dans le germarium (vert). **B)** Le même transgène, mais où le fragment du GAG d'Idefix a été excisé (pGg^{DId}) exprime la GFP dans toute les cellules germinales (vert) mais ne peut s'exprimer précocement dans les cellules souches germinales (point rouge) marquées par un anticorps anti-Phusome (1B1). L'actine est marquée en blanc.

constitué du LTR de Heta suivi du gène rapporteur codant pour la β -galactosidase montre une baisse de la répression dans le germarium en contexte hétérozygote mutant pour Spn-E ou PIWI (Shpiz, Olovnikov et al. 2011). Il semblerait donc qu'il existe dans le germarium une fenêtre précoce où la voie des piRNAs est moins efficace dans des contextes sensibilisés. Cependant les transgènes pGgId ne pouvant s'exprimer qu'à partir de la deuxième division mitotique des cellules souches germinales (4 cellules) (**Figure 85**), on ne peut pas exclure que la dérégulation soit plus précoce que celle observée. C'est pourquoi nous avons entrepris une collaboration avec l'équipe de Stéphane Ronsseray, qui dispose d'outils pouvant s'exprimer plus précocement, afin de définir exactement le stade de dérégulation.

En conclusion, dans un environnement normal il n'y a pas de fuite de l'expression des éléments transposables dans le germarium. On peut donc proposer que les éléments transposables soient très fortement régulés par du PTGS grâce à la voie des piRNAs en lignée germinale. En revanche, dans des contextes sensibilisés pouvant mimer un stress, on peut imaginer qu'une expression de ces éléments serait possible dans le germarium ce qui permettrait de générer de la variabilité dans la descendance et donc une possible adaptation.

On peut aussi observer que la protéine HP1 n'est pas présente à ces stades et pourrait être donc aussi impliquée dans cette sensibilité. Cette protéine étant présente dans les RCC ("Repressive Chromatin Centers") elle pourrait lors de son expression permettre une activation maximale des clusters de piRNAs. En revanche, dans les cellules où il y a une faiblesse de répression, les RCC sont bien formés (H3K9me³) mais ils ne pourraient que faiblement activer les clusters, favorisant un possible relâchement dans la répression des éléments transposables.

Publication 3

Logiciel d'analyse des banques de données de séquençages

1) Introduction

La mise au point des techniques de séquençage exhaustif haut débit des ARNs depuis ces cinq dernières années a engendré la création d'un nombre considérable de bases de données. Le séquençage massif a donc généré une grande quantité d'information mais aussi une grande complexité dans le traitement de ces informations. En effet il est très difficile de rechercher manuellement des informations dans des fichiers contenant plusieurs millions de séquences. Afin de pallier à ce problème nous avons mis au point un logiciel qui permet de localiser les séquences contenues dans ces bases de données sur des gènes d'intérêts.

On sait que les petits ARNs sont capables de cibler des ARNm afin de les détruire. Il est aussi connu que la complémentarité entre ces petits ARNs et leurs cibles est importante pour indiquer le type de dégradation que vont subir ces dernières. Par exemple dans la voies des miRNAs, la reconnaissance d'une cible peut se faire malgré la présence de nucléotides non complémentaires (excepté dans la seed). C'est pourquoi nous avons voulu inclure un système de recherche permettant d'autoriser ou non la présence de nucléotides non complémentaires (mismatch). Enfin il a été montré récemment dans la voie des piRNAs qu'un nombre minimal de 16 nucléotides consécutif d'un piRNA est capable de reconnaître sa cible et de la dégrader (Rouget, Papin et al. 2010) nous avons donc intégré un outil de recherche de blocs de séquences pouvant cibler les gènes d'intérêts.

NucBase, an efficient and handleable tool to extract sequences from large data.

Dufourt J [#], Pouchin P[#], Brassat E, and Vaury C^{*}

Address:

Clermont Université, Université d'Auvergne, BP 10448, F-63000 CLERMONT-FERRAND,
France

Inserm, U 931, F-63001 CLERMONT-FERRAND, France

CNRS, UMR 6247, F-63001 CLERMONT-FERRAND, France

[#] These authors contributed equally to this work.

^{*} Corresponding author

Tel: (33) 4 17 81 84

Fax: (33) 4 27 61 32

E-mail: Chantal.VAURY@u-clermont1.fr

I Introduction

Deep sequencing (next-generation sequencing) has recently emerged as powerful technologies providing unprecedented insights into biological systems (Metzker 2010). With automated sequencing technologies becoming routinely used, the acquisition of millions of sequence reads per experiment leads researchers to a subsequent challenge of data analysis. Among databases available for sequence analysis, databases reporting sequences of small noncoding RNA have been abundant during the past ten years. Indeed, small RNA including siRNA and miRNA came to be important players in the animal transcriptome regulation to destroy exogenous RNA or to regulate mRNA translational activity respectively (Kim, Han et al. 2009). A third class of small RNAs appears to regulate transposable element particularly in the germline cells of many organisms: the Piwi interacting RNAs (piRNAs). Bound to proteins of the argonaute PIWI family, these small RNAs recognize and destroy mRNA transcripts encoded by transposable elements preventing their propagation. In 2007, the GJ Hannon laboratory published the first large database of piRNAs from *Drosophila* with approximately 17000 reads (Brennecke, Aravin et al. 2007). In 2008, more than a million reads was reported in a database from deep sequencing piRNAs (Brennecke, Malone et al. 2008). piRNAs are involved in transposable elements silencing in the germline of mice, rats, *drosophila* and other animal species (Carmell, Girard et al. 2007; Desset, Buchon et al. 2008) (Aravin, Gaidatzis et al. 2006; Girard, Sachidanandam et al. 2006). Despite the amount of available data it still remains obscure whether the recognition of the transposon mRNA by the piRNA requires a perfect identity or if a mismatch is allowed like in miRNA pathway recognition. Also, it is still unclear if there is a seed with a minimal number of consecutive nucleotides required for this recognition like miRNA regulation or gene regulation by the

piRNA pathway (Rouget, Papin et al. 2010).

Here, we report NucBase, a new algorithm designed to search complementary sequences in large nucleic acids databases. NucBase rapidly identifies nucleic motives with strict identity to input sequences. This software also allows introducing mismatches or searching for seed sequences as targets. Results are exported in a Generic Feature Format 3 (GFF3) output format, which allows fast annotation on Gbrowse viewer. This software can be locally run on any Windows, Linux or Mac OS computer with 32-bit architecture compatibility. The software is freely available for download at [...].

II Software description

NucBase is a small and easy-to-use program developed in C++, with a Qt graphical interface, whose aim is to align millions of reads on one or several target sequences, and their reverse complement counterparts. For this purpose, it first computes the Burrows-Wheeler transform (BWT) (Burrows and Wheeler 1994) of the target, as well as its suffix array, through the fast and lightweight LibDivSufSort library (Mory,Y). These two elements, which form what is called an FM-Index, are then used to find all the positions where the read matches exactly (Ferragina and Manzini 2000). It is also possible to find read matches with, at most, K mismatched characters. The method used for this search is the same as for exact matches, except that for each nucleotide of the read, we also look for all the other possible bases (Zhang, Mukherjee et al. 2003). A third alternative exists when looking for a read: one can specify the size S of what will be called a “seed” and the software will look for all the occurrences of at least S consecutive nucleotides from the original read in the sequence. The possibility to find “seeds” is a unique feature based on the work of Martine Simonelig

(Rouget, Papin et al. 2010), which allows more flexibility in the search. Finally, the software only outputs matching reads by default, but it is possible to also print non-matching reads.

If we consider a string S which ends with a unique character that should be the lowest in the lexicographic order, then, by computing all the cyclic permutations of S and ordering them alphabetically, we obtain the Burrows-Wheeler Transform of S , $BWT(S)$: it is the last column of this matrix (figure1). Moreover, if we call $F(S)$ the first column of the matrix, then the suffix array, $SA(S)$, designates the vector containing the original positions of each character in $F(S)$. A particular property of the transformed sequence is that the n^{th} occurrence of a character in $BWT(S)$ is the n^{th} occurrence of said character in $F(S)$. Therefore, as suggested by Ferragina and Manzini (Ferragina and Manzini 2000), searching a word of size M in the original string of size N is easy and fast (complexity: $O(M)$). As shown in figure 2, we know all the adjacent positions of the word last character in $F(S)$. Since $BWT(S)$ is obtained through cyclic permutations, the characters in $BWT(S)$ are the ones preceding those in $F(S)$. Thus, by counting the number of occurrences of the penultimate character in $BWT(S)$ at the specified positions, we can determine its positions in $F(S)$.

As an input, NucBase needs one file containing the reads, as well as a sequence either in a FASTA or multi-FASTA file, or directly pasted into the software main window. The file containing the reads needs to be in a format similar to the one used by NCBI Geo Datasets: it should be a tab-separated table where the first column contains the reads and the others, numbers. There are two main output formats: General Feature Format (GFF3) and Geo Dataset tables. By default, the software will produce two table files containing the number of matches found in the sequence and its reverse complement, as well as a GFF3 file containing the positions of each read in both directions of the sequence. It is also possible to produce a global summary table giving the total number of matches for each read, as well as a FASTA file containing each sequence and its reverse complement, but with an asterisk where a

character matched. There are also two options that depend on the input database file: if the file contains labels, the output files will be named appropriately, and if it contains degenerate bases the software will expand the corresponding consensus. This last feature allows the user to generate a database from simple sequences: a single word can stand for thousands of nucleic acids.

III Results

Comparison with other software

Contrary to SOAP (Li, Li et al. 2008), BWA (Li and Durbin, 2009) or BowTie (Langmead and Salzberg 2012), NucBase has a graphical interface which allows any user to align reads on sequences. Moreover, it makes an exhaustive search, and finds all the reads matches, even with substitutions. The possibility to use previous results as new databases and the exhaustive search make it an appropriate tool for biologists who do not want to lose potential matches and are willing to trade some computation time in exchange for exhaustive results and simplicity.

Flamenco piRNA cluster analysis

To identify the potential piRNA produced by the flamenco cluster (*flam*) all piRNAs were first mapped to flamenco sequence (X:21495823,21684449). Among 701 853 piRNA of the Brennecke database 2008, 85420 piRNA matching with flamenco/COM locus were identified. Among them, 70795 piRNA are produced from the sense genomic strand of *flam* and 28 418 from the antisense genomic strand whereas 13793 piRNA can be produced from both strands.

All these 85420 piRNA can be aligned to the original genomic sequence to identify the start and the end positions of all piRNAs.

If 2 mismatches are allowed, 134040 piRNAs matching with flamenco/COM sequence are identified with 90444 piRNA produced from the sense genomic strand and 43596 from the antisense genomic strand. The software gives the information of the number of occurrence inside the referential sequence and/or inside the genome.

When piRNA whose sequence is unique in the flamenco/COM locus were searched in release 5 of the Drosophila genome, 85420 piRNA were sorted with 235 unique antisense piRNA and 8450 unique sense piRNA.

We searched to identify the piRNA regulators of the ZAM retrotransposon produced by the flamenco/COM cluster. All the genomic piRNAs were first mapped to the canonical sequence of ZAM and 419 antisense piRNA were found. These 419 antisense piRNA were then matched to the flam/COM sequence what resulted in the identification of 359 piRNA encoded by this piRNA cluster with 100% identity with ZAM. This result confirms that the major piRNA cluster regulating ZAM mobilization is indeed the flam/COM locus since only 60 piRNAs identical to ZAM are produced by other genomic loci.

Acknowledgment

This work was supported by grants from the Association pour la Recherche contre le Cancer (ARC 1139). J.D. received a graduate grant from the Ministère de l'Enseignement Supérieur et de la Recherche (MESR) and from the Association pour la Recherche contre le Cancer. E.B. and P.P. received a grant from the Région Auvergne. We are grateful to Y.Renaud and S.Jensen for helpful discussions.

Figure 1: The Burrows-Wheeler transform of a string is obtained by computing all the cyclic permutations of a string and then sorting those by alphabetic order: the last column of the matrix contains the BWT, while the suffix array gives the original position of the characters present in the first column.

Figure 2: We look for the string “CAT” in the sequence “ATTGCCATCGAA”. The last character is a ‘T’: there are 3 in the sequence, but only two are preceded by an ‘A’. The ‘A’s that precede a ‘T’ are the last 2 ‘A’s of the BWT: we find them in the first column and only one is preceded by a ‘C’, which is the first occurring in the BWT. The suffix array gives us its position in the sequence. We found “CAT” at position 6.

Figure 3: NucBase graphical interface allows the user to select a database containing reads (using Geo Datasets format) and a folder or a file containing the sequence(s) of interest. It is also possible to directly copy and paste a sequence in the main window. Before starting the software, the user needs to specify the column of the database to process and the options to use.

- Aravin, A., D. Gaidatzis, et al. (2006). "A novel class of small RNAs bind to MILI protein in mouse testes." Nature **442**(7099): 203-207.
- Brennecke, J., A. A. Aravin, et al. (2007). "Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*." Cell **128**(6): 1089-1103.
- Brennecke, J., C. D. Malone, et al. (2008). "An epigenetic role for maternally inherited piRNAs in transposon silencing." Science **322**(5906): 1387-1392.
- Burrows, M. and D. J. Wheeler (1994). A block-sorting lossless data compression algorithm.
- Carmell, M. A., A. Girard, et al. (2007). "MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline." Dev Cell **12**(4): 503-514.
- Desset, S., N. Buchon, et al. (2008). "In *Drosophila melanogaster* the COM locus directs the somatic silencing of two retrotransposons through both Piwi-dependent and -independent pathways." PLoS One **3**(2): e1526.
- Ferragina, P. and G. Manzini (2000). Opportunistic data structures with applications. Proceedings of the 41st Annual Symposium on Foundations of Computer Science, IEEE Computer Society: 390.
- Girard, A., R. Sachidanandam, et al. (2006). "A germline-specific class of small RNAs binds mammalian Piwi proteins." Nature **442**(7099): 199-202.
- Kim, V. N., J. Han, et al. (2009). "Biogenesis of small RNAs in animals." Nat Rev Mol Cell Biol **10**(2): 126-139.
- Langmead, B. and S. L. Salzberg (2012). "Fast gapped-read alignment with Bowtie 2." Nat Methods.
- Li, H., J. Ruan, et al. (2008). "Mapping short DNA sequencing reads and calling variants using mapping quality scores." Genome Res **18**(11): 1851-1858.
- Li, R., Y. Li, et al. (2008). "SOAP: short oligonucleotide alignment program." Bioinformatics **24**(5): 713-714.
- Metzker, M. L. (2010). "Sequencing technologies - the next generation." Nat Rev Genet **11**(1): 31-46.
- Mori, Y Bibliothèque LibDivSufSort; <http://code.google.com/p/libdivsufsort/>
- Rouget, C., C. Papin, et al. (2010). "Maternal mRNA deadenylation and decay by the piRNA pathway in the early *Drosophila* embryo." Nature **467**(7319): 1128-1132.
- Zhang, N., A. Mukherjee, et al. (2003). Approximate Pattern Matching Using the Burrows-Wheeler Transform. DCC, IEEE Computer Society: 458.

SA	Permutations		SA	F	Ordered permutations	BWT
1	ATTGCCATCGAA\$	Sort ➔	13	\$	\$ATTGCCATCGAA	A
2	TTGCCATCGAA\$A		12	A	A\$ATTGCCATCGA	A
3	TGCCATCGAA\$AT		11	A	AA\$ATTGCCATCG	G
4	GCCATCGAA\$ATT		7	A	ATCGAA\$ATTGCC	C
5	CCATCGAA\$ATTG		1	A	ATTGCCATCGAA\$	\$
6	CATCGAA\$ATTGC		6	C	CATCGAA\$ATTGC	C
7	ATCGAA\$ATTGCC		5	C	CCATCGAA\$ATTG	G
8	TCGAA\$ATTGCCA		9	C	CGAA\$ATTGCCAT	T
9	CGAA\$ATTGCCAT		10	G	GAA\$ATTGCCATC	C
10	GAA\$ATTGCCATC		4	G	GCCATCGAA\$ATT	T
11	AA\$ATTGCCATCG		8	T	TCGAA\$ATTGCCA	A
12	A\$ATTGCCATCGA		3	T	TGCCATCGAA\$AT	T
13	\$ATTGCCATCGAA		2	T	TTGCCATCGAA\$A	A

Figure 1: Burrows-Wheeler Transform and Suffix Array construction

Searching CAT

SA	F	Ordered permutations	BWT	SA	F	Ordered permutations	BWT	SA	F	Ordered permutations	BWT
13	\$	\$ATTGCCATCGAA	A	13	\$	\$ATTGCCATCGAA	A	13	\$	\$ATTGCCATCGAA	A
12	A	A\$ATTGCCATCGA	A	12	A	A\$ATTGCCATCGA	A	12	A	A\$ATTGCCATCGA	A
11	A	AA\$ATTGCCATCG	G	11	A	AA\$ATTGCCATCG	G	11	A	AA\$ATTGCCATCG	G
7	A	ATCGAA\$ATTGCC	C	7	A	ATCGAA\$ATTGCC	C	7	A	ATCGAA\$ATTGCC	C
1	A	ATTGCCATCGAA\$	\$	1	A	ATTGCCATCGAA\$	\$	1	A	ATTGCCATCGAA\$	\$
6	C	CATCGAA\$ATTGC	C	6	C	CATCGAA\$ATTGC	C	6	C	CATCGAA\$ATTGC	C
5	C	CCATCGAA\$ATTG	G	5	C	CCATCGAA\$ATTG	G	5	C	CCATCGAA\$ATTG	G
9	C	CGAA\$ATTGCCAT	T	9	C	CGAA\$ATTGCCAT	T	9	C	CGAA\$ATTGCCAT	T
10	G	GAA\$ATTGCCATC	C	10	G	GAA\$ATTGCCATC	C	10	G	GAA\$ATTGCCATC	C
4	G	GCCATCGAA\$ATT	T	4	G	GCCATCGAA\$ATT	T	4	G	GCCATCGAA\$ATT	T
8	T	TCGAA\$ATTGCCA	A	8	T	TCGAA\$ATTGCCA	A	8	T	TCGAA\$ATTGCCA	A
3	T	TGCCATCGAA\$AT	T	3	T	TGCCATCGAA\$AT	T	3	T	TGCCATCGAA\$AT	T
2	T	TTGCCATCGAA\$A	A	2	T	TTGCCATCGAA\$A	A	2	T	TTGCCATCGAA\$A	A

Figure 2: FM-Index search

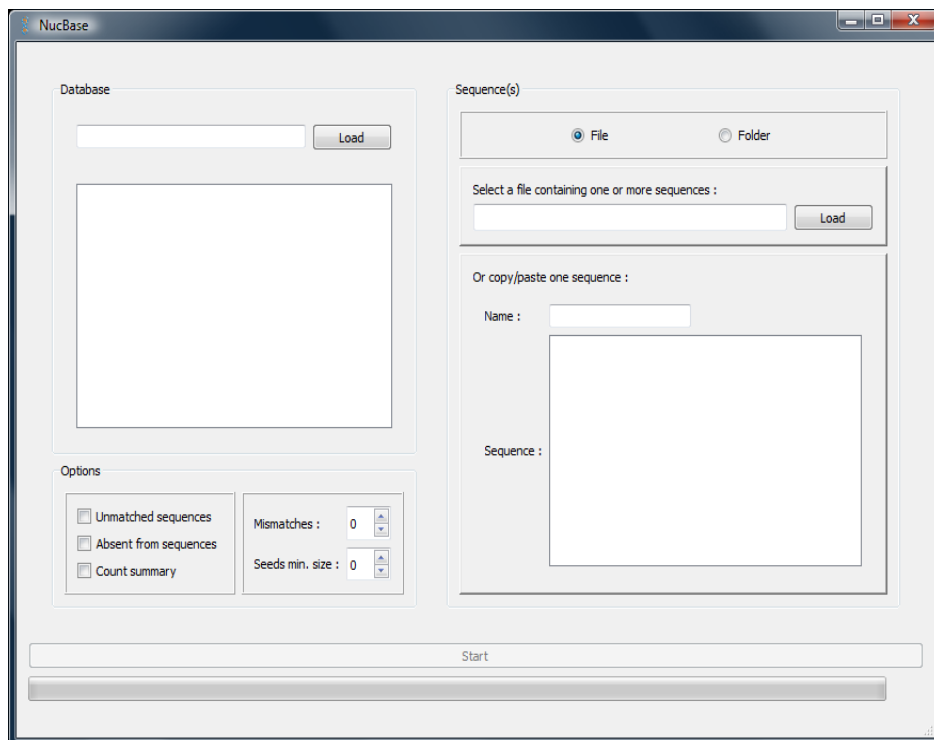


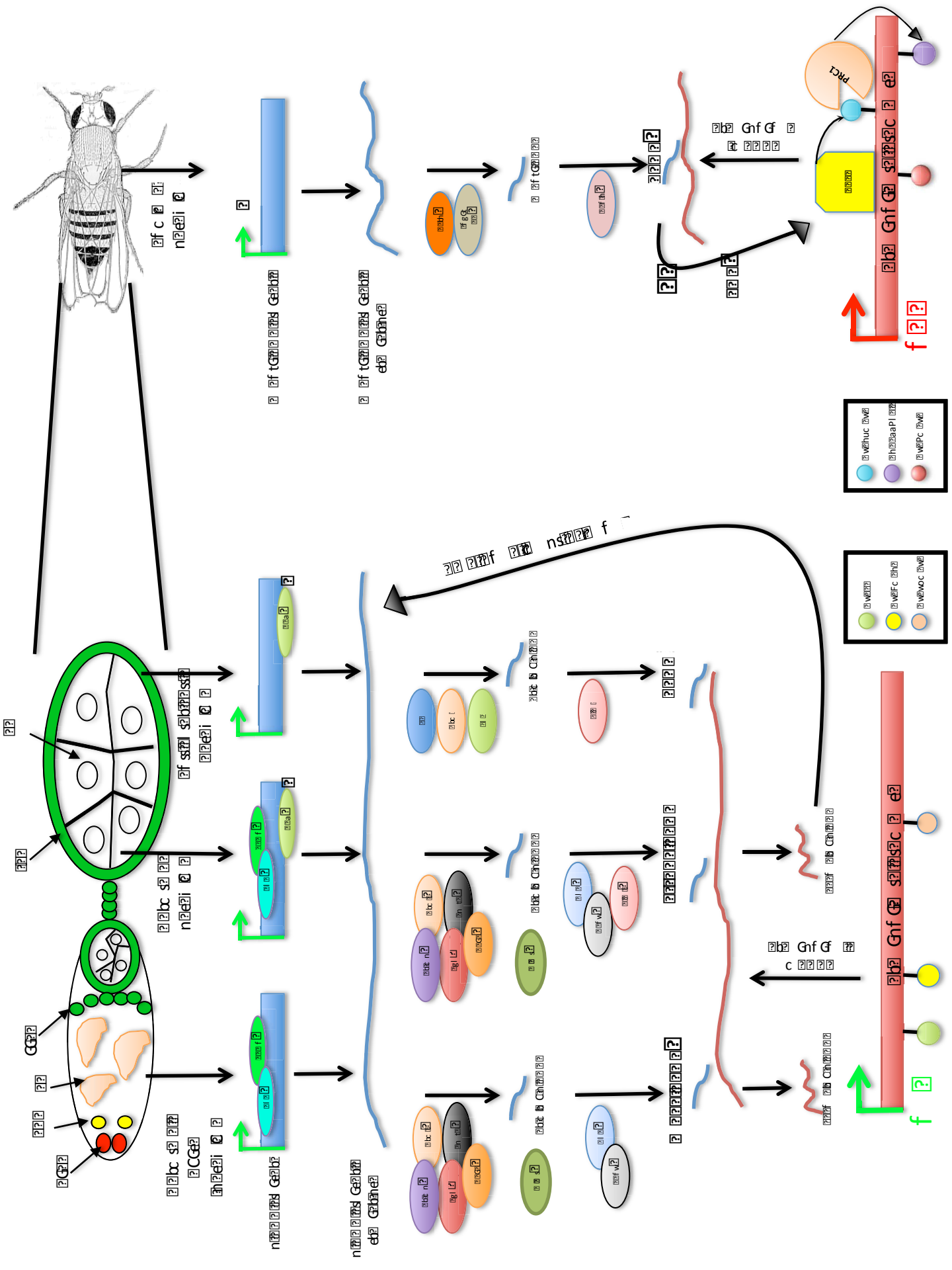
Figure 3: User Interface

2) Conclusion

Nucbase est le premier logiciel de traitement de base de données de séquences nucléotidiques accessible aux personnes n'ayant que très peu de connaissances en informatique. En effet les autres logiciels comparables disponibles sont tous destinés à des bioinformaticiens ou des informaticiens. Nucbase est un logiciel qui est capable de traiter très rapidement plusieurs giga de données et ce avec une approche non statistique. Les résultats obtenus avec ce logiciel ne sont pas basés sur des scores ou des probabilités mais sur des résultats exacts. Cette approche permet d'extraire des données de manière beaucoup plus précise.

Conclusion

En conclusion de ces travaux de thèse j'ai pu mettre en évidence que les éléments transposables étaient soumis à diverses régulations. Ce travail a donc permis de mettre en évidence une régulation différentielle dans les tissus somatiques, en fonction du type de tissu. Mais au delà du type tissulaire, il s'agit peut être de la localisation du tissu qui est importante. En effet les cellules folliculaires sont en contact avec la lignée germinale et peuvent donc transmettre leurs éléments transposables à la lignée germinale *via* une diffusion passive ou *via* une infection par les éléments de type rétrovirus comme ZAM, Idefix ou Gypsy. Ces cellules ne vont pas perdurer dans le temps car après la ponte de l'œuf elles vont dégénérer et former une gaine protectrice pour l'embryon, appelée le chorion. La voie des piRNAs qui régule les éléments transposables en cellules folliculaires semble assez robuste mais ne nécessite pas énormément d'acteurs. En effet si des éléments échappent à la voie de régulation dans ces tissus deux possibilités sont envisageables ; la première est que les éléments transposables vont aller s'insérer dans un nouveau locus au sein même de ces cellules ce qui ne risque pas de poser de problèmes majeurs du fait qu'elles vont dégénérer; la deuxième est qu'ils vont pouvoir passer dans la lignée germinale. Si les éléments sont capables d'envahir la lignée germinale ils se retrouvent la face à un deuxième niveau de régulation plus sophistiqué impliquant de nombreux acteurs. On ne peut pas oublier non plus que les éléments transposables sont un formidable moteur évolutif, et pour que ce bénéfice ait lieu il faut que la variabilité génique qu'ils apportent se passe dans la lignée germinale afin d'être transmise à la descendance. La présence d'une fenêtre de sensibilité très précoce dans le développement des cellules germinales ainsi que la régulation uniquement de type PTGS offre donc la meilleure alternative. En effet ce système permet une forte répression mais qui en cas de stress, provoqué par exemple par un changement environnemental, laisse une opportunité à



la mobilisation des éléments transposables. Cette mobilisation est alors le vecteur d'une possible adaptation.

En revanche dans les tissus somatiques qui vont former l'organisme adulte, il est difficile d'imaginer quel pourrait être le bénéfice apporté par la mobilisation des éléments transposables. En effet les variations géniques apportées ne pourraient pas être transmises à la descendance. On peut en revanche, facilement imaginer que les mutations que pourraient provoquer ces éléments, seraient à l'origine de pathologies variées telles l'apparition de tumeurs. Par exemple l'insertion d'un élément transposable dans un gène suppresseur de tumeur pourrait entraîner une dérégulation de ce dernier. Dans ces tissus j'ai montré que les éléments transposables subissent une régulation qui couple du PTGS et du TGS. L'action conjointe de ces deux types de régulation entraîne une mise sous silence complète des éléments permettant ainsi de sauvegarder l'intégrité du génome somatique. Il semble donc finalement logique que la sélection naturelle ait permis de conserver des mécanismes de régulations tissulaires différents comme montré sur le schéma ci-dessous qui tente de récapituler l'ensemble de mes résultats et de les intégrer dans les connaissances actuelles des voies de régulation des éléments transposables chez la drosophile.

Références

Bibliographiques

A

- Aasland, R. and A. F. Stewart (1995). "The chromo shadow domain, a second chromo domain in heterochromatin-binding protein 1, HP1." *Nucleic Acids Res* **23**(16): 3168-3173.
- Ahlquist, P. (2006). "Parallels among positive-strand RNA viruses, reverse-transcribing viruses and double-stranded RNA viruses." *Nat Rev Microbiol* **4**(5): 371-382.
- Amariglio, N. and G. Rechavi (1993). "Insertional mutagenesis by transposable elements in the mammalian genome." *Environ Mol Mutagen* **21**(3): 212-218.
- Anand, A. and T. Kai (2011). "The tudor domain protein Kumo is required to assemble the nuage and to generate germline piRNAs in *Drosophila*." *EMBO J*.
- Ang, Y. S., S. Y. Tsai, et al. (2011). "Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network." *Cell* **145**(2): 183-197.
- Aravin, A., D. Gaidatzis, et al. (2006). "A novel class of small RNAs bind to MILI protein in mouse testes." *Nature* **442**(7099): 203-207.
- Aravin, A. A., M. S. Klenov, et al. (2004). "Dissection of a natural RNA silencing process in the *Drosophila melanogaster* germ line." *Mol Cell Biol* **24**(15): 6742-6750.
- Aravin, A. A., R. Sachidanandam, et al. (2008). "A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice." *Mol Cell* **31**(6): 785-799.
- Arcot, S. S., Z. Wang, et al. (1995). "Alu repeats: a source for the genesis of primate microsatellites." *Genomics* **29**(1): 136-144.

B

- Babiarz, J. E., J. G. Ruby, et al. (2008). "Mouse ES cells express endogenous shRNAs, siRNAs, and other Microprocessor-independent, Dicer-dependent small RNAs." *Genes Dev* **22**(20): 2773-2785.
- Bagchi, A., C. Papazoglu, et al. (2007). "CHD5 is a tumor suppressor at human 1p36." *Cell* **128**(3): 459-475.
- Bannister, A. J., P. Zegerman, et al. (2001). "Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain." *Nature* **410**(6824): 120-124.
- Bantignies, F. and G. Cavalli (2011). "Polycomb group proteins: repression in 3D." *Trends Genet* **27**(11): 454-464.
- Bantignies, F., V. Roure, et al. (2011). "Polycomb-dependent regulatory contacts between distant Hox loci in *Drosophila*." *Cell* **144**(2): 214-226.
- Bartel, D. P. (2009). "MicroRNAs: target recognition and regulatory functions." *Cell* **136**(2): 215-233.
- Batzler, M. A. and P. L. Deininger (2002). "Alu repeats and human genomic diversity." *Nat Rev Genet* **3**(5): 370-379.
- Bedford, M. T. and S. G. Clarke (2009). "Protein arginine methylation in mammals: who, what, and why." *Mol Cell* **33**(1): 1-13.
- Beisel, C. and R. Paro (2011). "Silencing chromatin: comparing modes and mechanisms." *Nat Rev Genet* **12**(2): 123-135.
- Belancio, V. P., A. M. Roy-Engel, et al. (2010). "All y'all need to know 'bout retroelements in cancer." *Semin Cancer Biol* **20**(4): 200-210.
- Bennetzen, J. L. (1996). "The contributions of retroelements to plant genome organization, function and evolution." *Trends Microbiol* **4**(9): 347-353.

- Biemont, C. (1992). "Population genetics of transposable DNA elements. A *Drosophila* point of view." *Genetica* **86**(1-3): 67-84.
- Biessmann, H. and J. M. Mason (1992). "Genetics and molecular biology of telomeres." *Adv Genet* **30**: 185-249.
- Biessmann, H., K. Valgeirsdottir, et al. (1992). "HeT-A, a transposable element specifically involved in "healing" broken chromosome ends in *Drosophila melanogaster*." *Mol Cell Biol* **12**(9): 3910-3918.
- Boeke, J. D. and O. K. Pickeral (1999). "Retrosuffling the genomic deck." *Nature* **398**(6723): 108-109, 111.
- Bohnsack, M. T., K. Czapinski, et al. (2004). "Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs." *RNA* **10**(2): 185-191.
- Borchert, G. M., W. Lanier, et al. (2006). "RNA polymerase III transcribes human microRNAs." *Nat Struct Mol Biol* **13**(12): 1097-1101.
- Boswell, R. E. and A. P. Mahowald (1985). "tudor, a gene required for assembly of the germ plasm in *Drosophila melanogaster*." *Cell* **43**(1): 97-104.
- Bourc'his, D. and T. H. Bestor (2004). "Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L." *Nature* **431**(7004): 96-99.
- Branco, M. R. and A. Pombo (2006). "Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations." *PLoS Biol* **4**(5): e138.
- Brasher, S. V., B. O. Smith, et al. (2000). "The structure of mouse HP1 suggests a unique mode of single peptide recognition by the shadow chromo domain dimer." *EMBO J* **19**(7): 1587-1597.
- Brennecke, J., A. A. Aravin, et al. (2007). "Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*." *Cell* **128**(6): 1089-1103.
- Brookes, E., I. de Santiago, et al. (2012). "Polycomb Associates Genome-wide with a Specific RNA Polymerase II Variant, and Regulates Metabolic Genes in ESCs." *Cell Stem Cell* **10**(2): 157-170.
- Brower-Toland, B., S. D. Findley, et al. (2007). "*Drosophila* PIWI associates with chromatin and interacts directly with HP1a." *Genes Dev* **21**(18): 2300-2311.
- Bucheton, A. (1990). "I transposable elements and I-R hybrid dysgenesis in *Drosophila*." *Trends Genet* **6**(1): 16-21.

C

- Callebaut, I. and J. P. Mornon (1997). "The human EBNA-2 coactivator p100: multidomain organization and relationship to the staphylococcal nuclease fold and to the tudor protein involved in *Drosophila melanogaster* development." *Biochem J* **321** (Pt 1): 125-132.
- Callinan, P. A. and M. A. Batzer (2006). "Retrotransposable elements and human disease." *Genome Dyn* **1**: 104-115.
- Cao, R., L. Wang, et al. (2002). "Role of histone H3 lysine 27 methylation in Polycomb-group silencing." *Science* **298**(5595): 1039-1043.
- Cao, S., H. Bendall, et al. (2003). "The high-mobility-group box protein SSRP1/T160 is essential for cell viability in day 3.5 mouse embryos." *Mol Cell Biol* **23**(15): 5301-5307.
- Carmell, M. A., A. Girard, et al. (2007). "MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline." *Dev Cell* **12**(4): 503-514.

- Chan, C. S., L. Rastelli, et al. (1994). "A Polycomb response element in the Ubx gene that determines an epigenetically inherited state of repression." *EMBO J* **13**(11): 2553-2564.
- Chen, Y., A. Pane, et al. (2007). "Cutoff and aubergine mutations result in retrotransposon upregulation and checkpoint activation in *Drosophila*." *Curr Biol* **17**(7): 637-642.
- Chintapalli, V. R., J. Wang, et al. (2007). "Using FlyAtlas to identify better *Drosophila melanogaster* models of human disease." *Nat Genet* **39**(6): 715-720.
- Chung, W. J., K. Okamura, et al. (2008). "Endogenous RNA interference provides a somatic defense against *Drosophila* transposons." *Curr Biol* **18**(11): 795-802.
- Conte, C., V. Calco, et al. (2000). "Impact of multiple insertions of two retroelements, ZAM and Idefix at an euchromatic locus." *Genetica* **109**(1-2): 53-59.
- Conte, C., B. Dastugue, et al. (2002). "Coupling of enhancer and insulator properties identified in two retrotransposons modulates their mutagenic impact on nearby genes." *Mol Cell Biol* **22**(6): 1767-1777.
- Cook, H. A., B. S. Koppetsch, et al. (2004). "The *Drosophila* SDE3 homolog armitage is required for oskar mRNA silencing and embryonic axis specification." *Cell* **116**(6): 817-829.
- Cordaux, R. (2008). "The human genome in the LINE of fire." *Proc Natl Acad Sci U S A* **105**(49): 19033-19034.
- Cost, G. J. and J. D. Boeke (1998). "Targeting of human retrotransposon integration is directed by the specificity of the L1 endonuclease for regions of unusual DNA structure." *Biochemistry* **37**(51): 18081-18093.
- Cowell, I. G., R. Auccott, et al. (2002). "Heterochromatin, HP1 and methylation at lysine 9 of histone H3 in animals." *Chromosoma* **111**(1): 22-36.
- Cowieson, N. P., J. F. Partridge, et al. (2000). "Dimerisation of a chromo shadow domain and distinctions from the chromodomain as revealed by structural analysis." *Curr Biol* **10**(9): 517-525.
- Cox, D. N., A. Chao, et al. (2000). "piwi encodes a nucleoplasmic factor whose activity modulates the number and division rate of germline stem cells." *Development* **127**(3): 503-514.
- Cremer, T. and C. Cremer (2001). "Chromosome territories, nuclear architecture and gene regulation in mammalian cells." *Nat Rev Genet* **2**(4): 292-301.
- Cremer, T. and M. Cremer "Chromosome territories." *Cold Spring Harb Perspect Biol* **2**(3): a003889.
- Cuthbert, G. L., S. Daujat, et al. (2004). "Histone deimination antagonizes arginine methylation." *Cell* **118**(5): 545-553.
- Czech, B., C. D. Malone, et al. (2008). "An endogenous small interfering RNA pathway in *Drosophila*." *Nature* **453**(7196): 798-802.
- Czech, B., R. Zhou, et al. (2009). "Hierarchical rules for Argonaute loading in *Drosophila*." *Mol Cell* **36**(3): 445-456.
- Czermin, B., R. Melfi, et al. (2002). "*Drosophila* enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites." *Cell* **111**(2): 185-196.

D

- Daniel, J. A., M. G. Pray-Grant, et al. (2005). "Effector proteins for methylated histones: an expanding family." *Cell Cycle* **4**(7): 919-926.

- de Wit, E., F. Greil, et al. (2007). "High-resolution mapping reveals links of HP1 with active and inactive chromatin components." *PLoS Genet* **3**(3): e38.
- Decotto, E. and A. C. Spradling (2005). "The Drosophila ovarian and testis stem cell niches: similar somatic stem cells and signals." *Dev Cell* **9**(4): 501-510.
- Dekker, J., K. Rippe, et al. (2002). "Capturing chromosome conformation." *Science* **295**(5558): 1306-1311.
- Deng, W. and H. Lin (2002). "miwi, a murine homolog of piwi, encodes a cytoplasmic protein essential for spermatogenesis." *Dev Cell* **2**(6): 819-830.
- Desset, S., N. Buchon, et al. (2008). "In Drosophila melanogaster the COM locus directs the somatic silencing of two retrotransposons through both Piwi-dependent and -independent pathways." *PLoS One* **3**(2): e1526.
- Desset, S., C. Conte, et al. (1999). "Mobilization of two retroelements, ZAM and Idefix, in a novel unstable line of Drosophila melanogaster." *Mol Biol Evol* **16**(1): 54-66.
- Desset, S., C. Meignin, et al. (2003). "COM, a heterochromatic locus governing the control of independent endogenous retroviruses from Drosophila melanogaster." *Genetics* **164**(2): 501-509.
- Diederichs, S. and D. A. Haber (2007). "Dual role for argonautes in microRNA processing and posttranscriptional regulation of microRNA expression." *Cell* **131**(6): 1097-1108.
- Dietzl, G., D. Chen, et al. (2007). "A genome-wide transgenic RNAi library for conditional gene inactivation in Drosophila." *Nature* **448**(7150): 151-156.
- Duan, Z., M. Andronescu, et al. (2010). "A three-dimensional model of the yeast genome." *Nature* **465**(7296): 363-367.
- Dufourt, J., E. Brasset, et al. (2011). "Polycomb group-dependent, heterochromatin protein 1-independent, chromatin structures silence retrotransposons in somatic tissues outside ovaries." *DNA research : an international journal for rapid publication of reports on genes and genomes* **18**(6): 451-461.
- Dufourt, J., E. Brasset, et al. (2011). "Polycomb group-dependent, heterochromatin protein 1-independent, chromatin structures silence retrotransposons in somatic tissues outside ovaries." *DNA Res* **18**(6): 451-461.
- Dupressoir, A. and T. Heidmann (1997). "Expression of intracisternal A-particle retrotransposons in primary tumors of oncogene-expressing transgenic mice." *Oncogene* **14**(24): 2951-2958.

E

- Eissenberg, J. C., T. C. James, et al. (1990). "Mutation in a heterochromatin-specific chromosomal protein is associated with suppression of position-effect variegation in Drosophila melanogaster." *Proc Natl Acad Sci U S A* **87**(24): 9923-9927.
- Enderle, D., C. Beisel, et al. (2011). "Polycomb preferentially targets stalled promoters of coding and noncoding transcripts." *Genome Res* **21**(2): 216-226.
- Esteller, M. (2011). "Non-coding RNAs in human disease." *Nat Rev Genet* **12**(12): 861-874.

F

- Fanti, L., M. Berloco, et al. (2003). "Chromosomal distribution of heterochromatin protein 1 (HP1) in *Drosophila*: a cytological map of euchromatic HP1 binding sites." *Genetica* **117**(2-3): 135-147.
- Filion, G. J., J. G. van Bommel, et al. (2010). "Systematic protein location mapping reveals five principal chromatin types in *Drosophila* cells." *Cell* **143**(2): 212-224.
- Findley, S. D., M. Tamanaha, et al. (2003). "Maelstrom, a *Drosophila* spindle-class gene, encodes a protein that colocalizes with Vasa and RDE1/AGO1 homolog, Aubergine, in nuage." *Development* **130**(5): 859-871.
- Fire, A., S. Xu, et al. (1998). "Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*." *Nature* **391**(6669): 806-811.
- Fischle, W., V. Kiermer, et al. (2001). "The emerging role of class II histone deacetylases." *Biochem Cell Biol* **79**(3): 337-348.
- Flynt, A. S. and E. C. Lai (2008). "Biological principles of microRNA-mediated regulation: shared themes amid diversity." *Nat Rev Genet* **9**(11): 831-842.
- Forneris, F., C. Binda, et al. (2005). "Histone demethylation catalysed by LSD1 is a flavin-dependent oxidative process." *FEBS Lett* **579**(10): 2203-2207.
- Forstemann, K., M. D. Horwich, et al. (2007). "*Drosophila* microRNAs are sorted into functionally distinct argonaute complexes after production by dicer-1." *Cell* **130**(2): 287-297.
- Forstemann, K., Y. Tomari, et al. (2005). "Normal microRNA maturation and germ-line stem cell maintenance requires Loquacious, a double-stranded RNA-binding domain protein." *PLoS Biol* **3**(7): e236.
- Francis, N. J., R. E. Kingston, et al. (2004). "Chromatin compaction by a polycomb group protein complex." *Science* **306**(5701): 1574-1577.
- Francis, N. J., A. J. Saurin, et al. (2001). "Reconstitution of a functional core polycomb repressive complex." *Mol Cell* **8**(3): 545-556.
- Friedman, R. C., K. K. Farh, et al. (2009). "Most mammalian mRNAs are conserved targets of microRNAs." *Genome Res* **19**(1): 92-105.
- Fuks, F. (2003). "[DNA methyltransferases: from chromatin remodeling to cancer]." *Med Sci (Paris)* **19**(4): 477-480.
- Fuks, F. (2005). "DNA methylation and histone modifications: teaming up to silence genes." *Curr Opin Genet Dev* **15**(5): 490-495.

G

- Gasior, S. L., T. P. Wakeman, et al. (2006). "The human LINE-1 retrotransposon creates DNA double-strand breaks." *J Mol Biol* **357**(5): 1383-1393.
- Ghildiyal, M., H. Seitz, et al. (2008). "Endogenous siRNAs derived from transposons and mRNAs in *Drosophila* somatic cells." *Science* **320**(5879): 1077-1081.
- Gilbert, N., S. Lutz-Prigge, et al. (2002). "Genomic deletions created upon LINE-1 retrotransposition." *Cell* **110**(3): 315-325.
- Girard, A., R. Sachidanandam, et al. (2006). "A germline-specific class of small RNAs binds mammalian Piwi proteins." *Nature* **442**(7099): 199-202.
- Golding, M. C., L. Zhang, et al. (2010). "Multiple epigenetic modifiers induce aggressive viral extinction in extraembryonic endoderm stem cells." *Cell Stem Cell* **6**(5): 457-467.

- Grandbastien, M. A., C. Audeon, et al. (1994). "Functional analysis of the tobacco Tnt1 retrotransposon." *Genetica* **93**(1-3): 181-189.
- Greer, E. L., T. J. Maures, et al. (2010). "Members of the H3K4 trimethylation complex regulate lifespan in a germline-dependent manner in *C. elegans*." *Nature* **466**(7304): 383-387.
- Gregory, R. I., K. P. Yan, et al. (2004). "The Microprocessor complex mediates the genesis of microRNAs." *Nature* **432**(7014): 235-240.
- Grimaud, C., F. Bantignies, et al. (2006). "RNAi components are required for nuclear clustering of Polycomb group response elements." *Cell* **124**(5): 957-971.

H

- Haase, A. D., S. Fenoglio, et al. (2010). "Probing the initiation and effector phases of the somatic piRNA pathway in *Drosophila*." *Genes Dev* **24**(22): 2499-2504.
- Haase, A. D., L. Jaskiewicz, et al. (2005). "TRBP, a regulator of cellular PKR and HIV-1 virus expression, interacts with Dicer and functions in RNA silencing." *EMBO Rep* **6**(10): 961-967.
- Han, J., Y. Lee, et al. (2006). "Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex." *Cell* **125**(5): 887-901.
- Handler, D., D. Olivieri, et al. (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**(19): 3977-3993.
- Hartig, J. V., S. Esslinger, et al. (2009). "Endo-siRNAs depend on a new isoform of loquacious and target artificially introduced, high-copy sequences." *EMBO J* **28**(19): 2932-2944.
- Hassig, C. A., J. K. Tong, et al. (1998). "A role for histone deacetylase activity in HDAC1-mediated transcriptional repression." *Proc Natl Acad Sci U S A* **95**(7): 3519-3524.
- Hay, B., L. Ackerman, et al. (1988). "Identification of a component of *Drosophila* polar granules." *Development* **103**(4): 625-640.
- Hay, B., L. Y. Jan, et al. (1988). "A protein component of *Drosophila* polar granules is encoded by vasa and has extensive sequence similarity to ATP-dependent helicases." *Cell* **55**(4): 577-587.
- Haynes, S. R., D. Johnson, et al. (1991). "The *Drosophila* Hrb87F gene encodes a new member of the A and B hnRNP protein group." *Nucleic Acids Res* **19**(1): 25-31.

Heitz E. 1928. Das Heterochromatin der Moose. I. Jahrb. Wiss. Bot. 69:762–818

- Herranz, H. and S. M. Cohen (2010). "MicroRNAs and gene regulatory networks: managing the impact of noise in biological systems." *Genes Dev* **24**(13): 1339-1344.
- Horwich, M. D., C. Li, et al. (2007). "The *Drosophila* RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC." *Curr Biol* **17**(14): 1265-1272.
- Hsiao, W. L., S. Gattoni-Celli, et al. (1986). "Effects of 5-azacytidine on expression of endogenous retrovirus-related sequences in C3H 10T1/2 cells." *J Virol* **57**(3): 1119-1126.

- Huang, H., Q. Gao, et al. (2011). "piRNA-associated germline nuage formation and spermatogenesis require MitoPLD profusogenic mitochondrial-surface lipid signaling." *Dev Cell* **20**(3): 376-387.
- Hutvagner, G., J. McLachlan, et al. (2001). "A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA." *Science* **293**(5531): 834-838.

I

- Ibanez-Ventoso, C., M. Vora, et al. (2008). "Sequence relationships among C. elegans, D. melanogaster and human microRNAs highlight the extensive conservation of microRNAs in biology." *PLoS One* **3**(7): e2818.

J

- Jenuwein, T. and C. D. Allis (2001). "Translating the histone code." *Science* **293**(5532): 1074-1080.
- Jiang, F., X. Ye, et al. (2005). "Dicer-1 and R3D1-L catalyze microRNA maturation in Drosophila." *Genes Dev* **19**(14): 1674-1679.
- Josse, T., C. Maurel-Zaffran, et al. (2008). "Telomeric trans-silencing in Drosophila melanogaster: tissue specificity, development and functional interactions between non-homologous telomeres." *PLoS One* **3**(9): e3249.
- Josse, T., L. Teyssset, et al. (2007). "Telomeric trans-silencing: an epigenetic repression combining RNA silencing and heterochromatin formation." *PLoS Genet* **3**(9): 1633-1643.
- Jurka, J. (1997). "Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons." *Proc Natl Acad Sci U S A* **94**(5): 1872-1877.
- Jurka, J., P. Klonowski, et al. (1998). "Mammalian retroposons integrate at kinkable DNA sites." *J Biomol Struct Dyn* **15**(4): 717-721.
- Justice, C. M., Z. Den, et al. (2001). "Phylogenetic analysis of the Friedreich ataxia GAA trinucleotide repeat." *J Mol Evol* **52**(3): 232-238.

K

- Kawamura, Y., K. Saito, et al. (2008). "Drosophila endogenous small RNAs bind to Argonaute 2 in somatic cells." *Nature* **453**(7196): 793-797.
- Kawaoka, S., N. Izumi, et al. (2011). "3' end formation of PIWI-interacting RNAs in vitro." *Mol Cell* **43**(6): 1015-1022.
- Ketting, R. F., S. E. Fischer, et al. (2001). "Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans." *Genes Dev* **15**(20): 2654-2659.
- Khvorova, A., A. Reynolds, et al. (2003). "Functional siRNAs and miRNAs exhibit strand bias." *Cell* **115**(2): 209-216.
- Kim, D. H., M. A. Behlke, et al. (2005). "Synthetic dsRNA Dicer substrates enhance RNAi potency and efficacy." *Nat Biotechnol* **23**(2): 222-226.
- Kim, D. H., P. Saetrom, et al. (2008). "MicroRNA-directed transcriptional gene silencing in mammalian cells." *Proc Natl Acad Sci U S A* **105**(42): 16230-16235.

- Kim, H. G., K. C. Kim, et al. (2011). "Gene silencing in HIV-1 latency by polycomb repressive group." *Virol J* **8**: 179.
- Kim, K., Y. S. Lee, et al. (2007). "Conversion of pre-RISC to holo-RISC by Ago2 during assembly of RNAi complexes." *RNA* **13**(1): 22-29.
- Kim, M., N. J. Krogan, et al. (2004). "The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II." *Nature* **432**(7016): 517-522.
- Kim, V. N. (2004). "MicroRNA precursors in motion: exportin-5 mediates their nuclear export." *Trends Cell Biol* **14**(4): 156-159.
- Kim, Y. K. and V. N. Kim (2007). "Processing of intronic microRNAs." *EMBO J* **26**(3): 775-783.
- King, R. C. (1970). "The meiotic behavior of the *Drosophila* oocyte." *Int Rev Cytol* **28**: 125-168.
- Kingston, R. E. and G. J. Narlikar (1999). "ATP-dependent remodeling and acetylation as regulators of chromatin fluidity." *Genes Dev* **13**(18): 2339-2352.
- Kirino, Y., N. Kim, et al. (2009). "Arginine methylation of Piwi proteins catalysed by dPRMT5 is required for Ago3 and Aub stability." *Nat Cell Biol* **11**(5): 652-658.
- Kirino, Y. and Z. Mourelatos (2007). "The mouse homolog of HEN1 is a potential methylase for Piwi-interacting RNAs." *RNA* **13**(9): 1397-1401.
- Kirino, Y. and Z. Mourelatos (2007). "Mouse Piwi-interacting RNAs are 2'-O-methylated at their 3' termini." *Nat Struct Mol Biol* **14**(4): 347-348.
- Klattenhoff, C., H. Xi, et al. (2009). "The *Drosophila* HP1 homolog Rhino is required for transposon silencing and piRNA production by dual-strand clusters." *Cell* **138**(6): 1137-1149.
- Klein, H. L. and T. D. Petes (1984). "Genetic mapping of Ty elements in *Saccharomyces cerevisiae*." *Mol Cell Biol* **4**(2): 329-339.
- Klenov, M. S., S. A. Lavrov, et al. (2007). "Repeat-associated siRNAs cause chromatin silencing of retrotransposons in the *Drosophila melanogaster* germline." *Nucleic Acids Res* **35**(16): 5430-5438.
- Kuff, E. L. and J. W. Fewell (1985). "Intracisternal A-particle gene expression in normal mouse thymus tissue: gene products and strain-related variability." *Mol Cell Biol* **5**(3): 474-483.
- Kuramochi-Miyagawa, S., T. Kimura, et al. (2004). "Mili, a mammalian member of piwi family gene, is essential for spermatogenesis." *Development* **131**(4): 839-849.
- Kuramochi-Miyagawa, S., T. Kimura, et al. (2001). "Two mouse piwi-related genes: miwi and mili." *Mech Dev* **108**(1-2): 121-133.
- Kurosaki, T., A. Ninokata, et al. (2006). "Evolutionary scenario for acquisition of CAG repeats in human SCA1 gene." *Gene* **373**: 23-27.
- Kuzmichev, A., K. Nishioka, et al. (2002). "Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein." *Genes Dev* **16**(22): 2893-2905.
- Kwon, S. H., L. Florens, et al. (2010). "Heterochromatin protein 1 (HP1) connects the FACT histone chaperone complex to the phosphorylated CTD of RNA polymerase II." *Genes Dev* **24**(19): 2133-2145.
- Kwon, S. H. and J. L. Workman (2008). "The heterochromatin protein 1 (HP1) family: put away a bias toward HP1." *Mol Cells* **26**(3): 217-227.

L

- Lachner, M., D. O'Carroll, et al. (2001). "Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins." *Nature* **410**(6824): 116-120.
- Lajoinie, O., M. E. Drake, et al. (1995). "Aberrant pre-mRNA maturation is caused by LINE insertions into introns of the white gene of *Drosophila melanogaster*." *Nucleic Acids Res* **23**(20): 4015-4022.
- Lander, E. S., L. M. Linton, et al. (2001). "Initial sequencing and analysis of the human genome." *Nature* **409**(6822): 860-921.
- Langst, G. and P. B. Becker (2004). "Nucleosome remodeling: one mechanism, many phenomena?" *Biochim Biophys Acta* **1677**(1-3): 58-63.
- Lanzuolo, C., V. Roure, et al. (2007). "Polycomb response elements mediate the formation of chromosome higher-order structures in the bithorax complex." *Nat Cell Biol* **9**(10): 1167-1174.
- Lau, N. C., N. Robine, et al. (2009). "Abundant primary piRNAs, endo-siRNAs, and microRNAs in a *Drosophila* ovary cell line." *Genome Res* **19**(10): 1776-1785.
- Leblanc, P., S. Desset, et al. (1997). "Invertebrate retroviruses: ZAM a new candidate in *D.melanogaster*." *EMBO J* **16**(24): 7521-7531.
- Lee, J., K. Han, et al. (2008). "Chromosomal inversions between human and chimpanzee lineages caused by retrotransposons." *PLoS One* **3**(12): e4047.
- Lee, J. S. and A. Shilatifard (2007). "A site to remember: H3K36 methylation a mark for histone deacetylation." *Mutat Res* **618**(1-2): 130-134.
- Lee, R. C., R. L. Feinbaum, et al. (1993). "The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*." *Cell* **75**(5): 843-854.
- Lee, T. I. and R. A. Young (2000). "Transcription of eukaryotic protein-coding genes." *Annu Rev Genet* **34**: 77-137.
- Lee, Y., C. Ahn, et al. (2003). "The nuclear RNase III Drosha initiates microRNA processing." *Nature* **425**(6956): 415-419.
- Lee, Y., I. Hur, et al. (2006). "The role of PACT in the RNA silencing pathway." *EMBO J* **25**(3): 522-532.
- Lee, Y., K. Jeon, et al. (2002). "MicroRNA maturation: stepwise processing and subcellular localization." *EMBO J* **21**(17): 4663-4670.
- Lee, Y., M. Kim, et al. (2004). "MicroRNA genes are transcribed by RNA polymerase II." *EMBO J* **23**(20): 4051-4060.
- Leeb, M., D. Pasini, et al. (2010). "Polycomb complexes act redundantly to repress genomic repeats and genes." *Genes Dev* **24**(3): 265-276.
- LeRoy, G., J. T. Weston, et al. (2009). "Heterochromatin protein 1 is extensively decorated with histone code-like post-translational modifications." *Mol Cell Proteomics* **8**(11): 2432-2442.
- Li, C., V. V. Vagin, et al. (2009). "Collapse of germline piRNAs in the absence of Argonaute3 reveals somatic piRNAs in flies." *Cell* **137**(3): 509-521.
- Li, Z., T. Tatsuke, et al. (2011). "Identification and characterization of Polycomb group genes in the silkworm, *Bombyx mori*." *Mol Biol Rep*.
- Liang, L., W. Diehl-Jones, et al. (1994). "Localization of vasa protein to the *Drosophila* pole plasm is independent of its RNA-binding and helicase activities." *Development* **120**(5): 1201-1211.

- Lieberman-Aiden, E., N. L. van Berkum, et al. (2009). "Comprehensive mapping of long-range interactions reveals folding principles of the human genome." *Science* **326**(5950): 289-293.
- Lim, A. K. and T. Kai (2007). "Unique germ-line organelle, nuage, functions to repress selfish genetic elements in *Drosophila melanogaster*." *Proc Natl Acad Sci U S A* **104**(16): 6714-6719.
- Lim, D. H., C. T. Oh, et al. (2011). "The endogenous siRNA pathway in *Drosophila* impacts stress resistance and lifespan by regulating metabolic homeostasis." *FEBS Lett* **585**(19): 3079-3085.
- Lin, H. and A. C. Spradling (1997). "A novel group of pumilio mutations affects the asymmetric division of germline stem cells in the *Drosophila* ovary." *Development* **124**(12): 2463-2476.
- Liu, H., E. H. Cheng, et al. (2007). "Bimodal degradation of MLL by SCFSkp2 and APC^{Cdc20} assures cell cycle execution: a critical regulatory circuit lost in leukemogenic MLL fusions." *Genes Dev* **21**(19): 2385-2398.
- Liu, Q., T. A. Rand, et al. (2003). "R2D2, a bridge between the initiation and effector steps of the *Drosophila* RNAi pathway." *Science* **301**(5641): 1921-1925.
- Liu, Y., S. D. Taverna, et al. (2007). "RNAi-dependent H3K27 methylation is required for heterochromatin formation and DNA elimination in *Tetrahymena*." *Genes Dev* **21**(12): 1530-1545.
- Lomberk, G., D. Bensi, et al. (2006). "Evidence for the existence of an HP1-mediated subcode within the histone code." *Nat Cell Biol* **8**(4): 407-415.
- Lower, R., K. Boller, et al. (1993). "Identification of human endogenous retroviruses with complex mRNA expression and particle formation." *Proc Natl Acad Sci U S A* **90**(10): 4480-4484.
- Lu, J., G. Getz, et al. (2005). "MicroRNA expression profiles classify human cancers." *Nature* **435**(7043): 834-838.
- Luger, K., T. J. Rechsteiner, et al. (1997). "Characterization of nucleosome core particles containing histone proteins made in bacteria." *J Mol Biol* **272**(3): 301-311.
- Lund, E., S. Guttinger, et al. (2004). "Nuclear export of microRNA precursors." *Science* **303**(5654): 95-98.

M

- Ma, J. B., Y. R. Yuan, et al. (2005). "Structural basis for 5'-end-specific recognition of guide RNA by the *A. fulgidus* Piwi protein." *Nature* **434**(7033): 666-670.
- MacAlpine, H. K., R. Gordan, et al. (2010). "*Drosophila* ORC localizes to open chromatin and marks sites of cohesin complex loading." *Genome Res* **20**(2): 201-211.
- Macrae, I. J., K. Zhou, et al. (2006). "Structural basis for double-stranded RNA processing by Dicer." *Science* **311**(5758): 195-198.
- Maestre, J., T. Tchenio, et al. (1995). "mRNA retroposition in human cells: processed pseudogene formation." *EMBO J* **14**(24): 6333-6338.
- Mahillon, J. and M. Chandler (1998). "Insertion sequences." *Microbiol Mol Biol Rev* **62**(3): 725-774.
- Mahowald, A. P. (1971). "Polar granules of *Drosophila*. 3. The continuity of polar granules during the life cycle of *Drosophila*." *J Exp Zool* **176**(3): 329-343.
- Maison, C., D. Bailly, et al. (2002). "Higher-order structure in pericentric heterochromatin involves a distinct pattern of histone modification and an RNA component." *Nat Genet* **30**(3): 329-334.

- Malone, C. D., J. Brennecke, et al. (2009). "Specialized piRNA pathways act in germline and somatic tissues of the *Drosophila* ovary." *Cell* **137**(3): 522-535.
- Marlor, R. L., S. M. Parkhurst, et al. (1986). "The *Drosophila melanogaster* gypsy transposable element encodes putative gene products homologous to retroviral proteins." *Mol Cell Biol* **6**(4): 1129-1134.
- Martinez, A. M., B. Schuettengruber, et al. (2009). "Polyhomeotic has a tumor suppressor activity mediated by repression of Notch signaling." *Nat Genet* **41**(10): 1076-1082.
- McClintock, B. (1956). "Controlling elements and the gene." *Cold Spring Harb Symp Quant Biol* **21**: 197-216.
- Meehan, R. R., C. F. Kao, et al. (2003). "HP1 binding to native chromatin in vitro is determined by the hinge region and not by the chromodomain." *EMBO J* **22**(12): 3164-3174.
- Mills, A. A. (2010). "Throwing the cancer switch: reciprocal roles of polycomb and trithorax proteins." *Nat Rev Cancer* **10**(10): 669-682.
- Miyoshi, K., T. Miyoshi, et al. (2010). "Molecular mechanisms that funnel RNA precursors into endogenous small-interfering RNA and microRNA biogenesis pathways in *Drosophila*." *RNA* **16**(3): 506-515.
- Miyoshi, K., H. Tsukumo, et al. (2005). "Slicer function of *Drosophila* Argonautes and its involvement in RISC formation." *Genes Dev* **19**(23): 2837-2848.
- Mochizuki, K., N. A. Fine, et al. (2002). "Analysis of a piwi-related gene implicates small RNAs in genome rearrangement in tetrahymena." *Cell* **110**(6): 689-699.
- Morlando, M., M. Ballarino, et al. (2008). "Primary microRNA transcripts are processed co-transcriptionally." *Nat Struct Mol Biol* **15**(9): 902-909.
- Morse, B., P. G. Rothers, et al. (1988). "Insertional mutagenesis of the myc locus by a LINE-1 sequence in a human breast carcinoma." *Nature* **333**(6168): 87-90.
- Muller, J., C. M. Hart, et al. (2002). "Histone methyltransferase activity of a *Drosophila* Polycomb group repressor complex." *Cell* **111**(2): 197-208.

N

- Nagao, A., K. Sato, et al. (2011). "Gender-Specific Hierarchy in Nuage Localization of PIWI-Interacting RNA Factors in *Drosophila*." *Front Genet* **2**: 55.
- Nagy, P. L., J. Griesenbeck, et al. (2002). "A trithorax-group complex purified from *Saccharomyces cerevisiae* is required for methylation of histone H3." *Proc Natl Acad Sci U S A* **99**(1): 90-94.
- Napoli, C., C. Lemieux, et al. (1990). "Introduction of a Chimeric Chalcone Synthase Gene into *Petunia* Results in Reversible Co-Suppression of Homologous Genes in trans." *Plant Cell* **2**(4): 279-289.
- Narlikar, G. J., H. Y. Fan, et al. (2002). "Cooperation between complexes that regulate chromatin structure and transcription." *Cell* **108**(4): 475-487.
- Ni, J. Q., R. Zhou, et al. (2011). "A genome-scale shRNA resource for transgenic RNAi in *Drosophila*." *Nat Methods* **8**(5): 405-407.
- Nielsen, A. L., M. Oulad-Abdelghani, et al. (2001). "Heterochromatin formation in mammalian cells: interaction between histones and HP1 proteins." *Mol Cell* **7**(4): 729-739.
- Nielsen, S. J., R. Schneider, et al. (2001). "Rb targets histone H3 methylation and HP1 to promoters." *Nature* **412**(6846): 561-565.
- Noordermeer, D., E. de Wit, et al. (2011). "Variegated gene expression caused by cell-specific long-range DNA interactions." *Nat Cell Biol* **13**(8): 944-951.

O

- O'Loghlen, A., A. M. Munoz-Cabello, et al. (2012). "MicroRNA Regulation of Cbx7 Mediates a Switch of Polycomb Orthologs during ESC Differentiation." *Cell Stem Cell* **10**(1): 33-46.
- Ohara, T., Y. Sakaguchi, et al. (2007). "The 3' termini of mouse Piwi-interacting RNAs are 2'-O-methylated." *Nat Struct Mol Biol* **14**(4): 349-350.
- Okamura, K., W. J. Chung, et al. (2008). "The Drosophila hairpin RNA pathway generates endogenous short interfering RNAs." *Nature* **453**(7196): 803-806.
- Okamura, K., N. Liu, et al. (2009). "Distinct mechanisms for microRNA strand selection by Drosophila Argonautes." *Mol Cell* **36**(3): 431-444.
- Olivieri, D., M. M. Sykora, et al. (2010). "An in vivo RNAi assay identifies major genetic and cellular requirements for primary piRNA biogenesis in Drosophila." *EMBO J* **29**(19): 3301-3317.

P

- Pal-Bhadra, M., U. Bhadra, et al. (2002). "RNAi related mechanisms affect both transcriptional and posttranscriptional transgene silencing in Drosophila." *Mol Cell* **9**(2): 315-327.
- Pal-Bhadra, M., B. A. Leibovitch, et al. (2004). "Heterochromatic silencing and HP1 localization in Drosophila are dependent on the RNAi machinery." *Science* **303**(5658): 669-672.
- Pane, A., P. Jiang, et al. (2011). "The Cutoff protein regulates piRNA cluster expression and piRNA production in the Drosophila germline." *EMBO J* **30**(22): 4601-4615.
- Pane, A., K. Wehr, et al. (2007). "zucchini and squash encode two putative nucleases required for rasiRNA production in the Drosophila germline." *Dev Cell* **12**(6): 851-862.
- Parker, J. S., S. M. Roe, et al. (2005). "Structural insights into mRNA recognition from a PIWI domain-siRNA guide complex." *Nature* **434**(7033): 663-666.
- Paro, R. and D. S. Hogness (1991). "The Polycomb protein shares a homologous domain with a heterochromatin-associated protein of Drosophila." *Proc Natl Acad Sci U S A* **88**(1): 263-267.
- Pasini, D., A. P. Bracken, et al. (2008). "Regulation of stem cell differentiation by histone methyltransferases and demethylases." *Cold Spring Harb Symp Quant Biol* **73**: 253-263.
- Pasquinelli, A. E., B. J. Reinhart, et al. (2000). "Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA." *Nature* **408**(6808): 86-89.
- Patil, V. S. and T. Kai (2010). "Repression of retroelements in Drosophila germline via piRNA pathway by the Tudor domain protein Tejas." *Curr Biol* **20**(8): 724-730.
- Pek, J. W. and T. Kai (2011). "A role for vasa in regulating mitotic chromosome condensation in Drosophila." *Curr Biol* **21**(1): 39-44.
- Phalke, S., O. Nickel, et al. (2009). "Retrotransposon silencing and telomere integrity in somatic cells of Drosophila depends on the cytosine-5 methyltransferase DNMT2." *Nat Genet* **41**(6): 696-702.
- Piacentini, L., L. Fanti, et al. (2003). "Heterochromatin protein 1 (HP1) is associated with induced gene expression in Drosophila euchromatin." *J Cell Biol* **161**(4): 707-714.

- Piacentini, L., L. Fanti, et al. (2009). "Heterochromatin protein 1 (HP1a) positively regulates euchromatic gene expression through RNA transcript association and interaction with hnRNPs in *Drosophila*." *PLoS Genet* **5**(10): e1000670.
- Plasterk, R. H. (2002). "RNA silencing: the genome's immune system." *Science* **296**(5571): 1263-1265.
- Ponting, C. P. (1997). "Tudor domains in proteins that interact with RNA." *Trends Biochem Sci* **22**(2): 51-52.
- Pray-Grant, M. G., J. A. Daniel, et al. (2005). "Chd1 chromodomain links histone H3 methylation with SAGA- and SLIK-dependent acetylation." *Nature* **433**(7024): 434-438.
- Pullirsch, D., R. Hartel, et al. (2010). "The Trithorax group protein Ash2l and Saf-A are recruited to the inactive X chromosome at the onset of stable X inactivation." *Development* **137**(6): 935-943.

Q

- Qiao, D., A. M. Zeeman, et al. (2002). "Molecular characterization of hiwi, a human member of the piwi gene family whose overexpression is correlated to seminomas." *Oncogene* **21**(25): 3988-3999.

R

- Rangan, P., C. D. Malone, et al. (2011). "piRNA production requires heterochromatin formation in *Drosophila*." *Curr Biol* **21**(16): 1373-1379.
- Reinhart, B. J., F. J. Slack, et al. (2000). "The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*." *Nature* **403**(6772): 901-906.
- Reiter, L. T., T. Murakami, et al. (1996). "A recombination hotspot responsible for two inherited peripheral neuropathies is located near a mariner transposon-like element." *Nat Genet* **12**(3): 288-297.
- Riddle, N. C., A. Minoda, et al. (2011). "Plasticity in patterns of histone modifications and chromosomal proteins in *Drosophila* heterochromatin." *Genome Res* **21**(2): 147-163.
- Rorth, P. (1998). "Gal4 in the *Drosophila* female germline." *Mech Dev* **78**(1-2): 113-118.
- Rose, M. R. and W. F. Doolittle (1983). "Molecular biological mechanisms of speciation." *Science* **220**(4593): 157-162.
- Rouget, C., C. Papin, et al. (2010). "Maternal mRNA deadenylation and decay by the piRNA pathway in the early *Drosophila* embryo." *Nature* **467**(7319): 1128-1132.
- Rountree, M. R., K. E. Bachman, et al. (2001). "DNA methylation, chromatin inheritance, and cancer." *Oncogene* **20**(24): 3156-3165.
- Roy, A. M., M. L. Carroll, et al. (2000). "Potential gene conversion and source genes for recently integrated Alu elements." *Genome Res* **10**(10): 1485-1495.
- Ruby, J. G., C. H. Jan, et al. (2007). "Intronic microRNA precursors that bypass Drosha processing." *Nature* **448**(7149): 83-86.

S

- Saito, K., H. Ishizu, et al. (2010). "Roles for the Yb body components Armitage and Yb in primary piRNA biogenesis in *Drosophila*." *Genes Dev* **24**(22): 2493-2498.
- Saraiya, A. A. and C. C. Wang (2008). "snoRNA, a novel precursor of microRNA in *Giardia lamblia*." *PLoS Pathog* **4**(11): e1000224.
- Sarot, E., G. Payen-Groschene, et al. (2004). "Evidence for a piwi-dependent RNA silencing of the gypsy endogenous retrovirus by the *Drosophila melanogaster* flamenco gene." *Genetics* **166**(3): 1313-1321.
- Schaefer, M., T. Pollex, et al. (2010). "RNA methylation by Dnmt2 protects transfer RNAs against stress-induced cleavage." *Genes Dev* **24**(15): 1590-1595.
- Scheuermann, J. C., A. G. de Ayala Alonso, et al. (2010). "Histone H2A deubiquitinase activity of the Polycomb repressive complex PR-DUB." *Nature* **465**(7295): 243-247.
- Schlesinger, M. B. and T. Formosa (2000). "POB3 is required for both transcription and replication in the yeast *Saccharomyces cerevisiae*." *Genetics* **155**(4): 1593-1606.
- Schoenfelder, S., T. Sexton, et al. (2010). "Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells." *Nat Genet* **42**(1): 53-61.
- Schuettengruber, B. and G. Cavalli (2009). "Recruitment of polycomb group complexes and their role in the dynamic regulation of cell fate choice." *Development* **136**(21): 3531-3542.
- Schuettengruber, B., M. Ganapathi, et al. (2009). "Functional anatomy of polycomb and trithorax chromatin landscapes in *Drosophila* embryos." *PLoS Biol* **7**(1): e13.
- Schuettengruber, B., A. M. Martinez, et al. (2011). "Trithorax group proteins: switching genes on and keeping them active." *Nat Rev Mol Cell Biol* **12**(12): 799-814.
- Schwartz, Y. B., T. G. Kahn, et al. (2006). "Genome-wide analysis of Polycomb targets in *Drosophila melanogaster*." *Nat Genet* **38**(6): 700-705.
- Schwarz, D. S., G. Hutvagner, et al. (2003). "Asymmetry in the assembly of the RNAi enzyme complex." *Cell* **115**(2): 199-208.
- Seitz, H. (2009). "Redefining microRNA targets." *Curr Biol* **19**(10): 870-873.
- Sexton, T., F. Bantignies, et al. (2009). "Genomic interactions: chromatin loops and gene meeting points in transcriptional regulation." *Semin Cell Dev Biol* **20**(7): 849-855.
- Sexton, T., E. Yaffe, et al. (2012). "Three-dimensional folding and functional organization principles of the *Drosophila* genome." *Cell* **148**(3): 458-472.
- Shareef, M. M., C. King, et al. (2001). "Drosophila heterochromatin protein 1 (HP1)/origin recognition complex (ORC) protein is associated with HP1 and ORC and functions in heterochromatin-induced silencing." *Mol Biol Cell* **12**(6): 1671-1685.
- Shpiz, S., I. Olovnikov, et al. (2011). "Mechanism of the piRNA-mediated silencing of *Drosophila* telomeric retrotransposons." *Nucleic Acids Res* **39**(20): 8703-8711.
- Siebold, A. P., R. Banerjee, et al. (2010). "Polycomb Repressive Complex 2 and Trithorax modulate *Drosophila* longevity and stress resistance." *Proc Natl Acad Sci U S A* **107**(1): 169-174.
- Singh, P. B., J. R. Miller, et al. (1991). "A sequence motif found in a *Drosophila* heterochromatin protein is conserved in animals and plants." *Nucleic Acids Res* **19**(4): 789-794.
- Siomi, M. C., T. Mannen, et al. (2010). "How does the royal family of Tudor rule the PIWI-interacting RNA pathway?" *Genes Dev* **24**(7): 636-646.

- Smothers, J. F. and S. Henikoff (2001). "The hinge and chromo shadow domain impart distinct targeting of HP1-like proteins." *Mol Cell Biol* **21**(7): 2555-2569.
- Spradling, A. C. (1993). "Germline cysts: communes that work." *Cell* **72**(5): 649-651.
- Steiner, F. A., S. W. Hoogstrate, et al. (2007). "Structural features of small RNA precursors determine Argonaute loading in *Caenorhabditis elegans*." *Nat Struct Mol Biol* **14**(10): 927-933.
- Sun, B., J. Hong, et al. (2008). "Molecular basis of the interaction of *Saccharomyces cerevisiae* Eaf3 chromo domain with methylated H3K36." *J Biol Chem* **283**(52): 36504-36512.

T

- Tam, O. H., A. A. Aravin, et al. (2008). "Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes." *Nature* **453**(7194): 534-538.
- Tan, C. H., T. C. Lee, et al. (2002). "Ziwi, the zebrafish homologue of the *Drosophila* piwi: co-localization with vasa at the embryonic genital ridge and gonad-specific expression in the adults." *Mech Dev* **119 Suppl 1**: S221-224.
- Tanentzapf, G., C. Smith, et al. (2000). "Apical, lateral, and basal polarization cues contribute to the development of the follicular epithelium during *Drosophila* oogenesis." *J Cell Biol* **151**(4): 891-904.
- Tanizawa, H., O. Iwasaki, et al. (2010). "Mapping of long-range associations throughout the fission yeast genome reveals global genome organization linked to transcriptional regulation." *Nucleic Acids Res* **38**(22): 8164-8177.
- Tcheressiz, S., V. Calco, et al. (2002). "Expression of the Idefix retrotransposon in early follicle cells in the germarium of *Drosophila melanogaster* is determined by its LTR sequences and a specific genomic context." *Mol Genet Genomics* **267**(2): 133-141.
- Todeschini, A. L., L. Teyssset, et al. (2010). "The epigenetic trans-silencing effect in *Drosophila* involves maternally-transmitted small RNAs whose production depends on the piRNA pathway and HP1." *PLoS One* **5**(6): e11032.
- Tolhuis, B., E. de Wit, et al. (2006). "Genome-wide profiling of PRC1 and PRC2 Polycomb chromatin binding in *Drosophila melanogaster*." *Nat Genet* **38**(6): 694-699.
- Tomari, Y., T. Du, et al. (2004). "RISC assembly defects in the *Drosophila* RNAi mutant armitage." *Cell* **116**(6): 831-841.
- Tomari, Y. and P. D. Zamore (2005). "Perspective: machines for RNAi." *Genes Dev* **19**(5): 517-529.
- Tsutsumi, A., T. Kawamata, et al. (2011). "Recognition of the pre-miRNA structure by *Drosophila* Dicer-1." *Nat Struct Mol Biol* **18**(10): 1153-1158.
- Tyagi, S. and W. Herr (2009). "E2F1 mediates DNA damage and apoptosis through HCF-1 and the MLL family of histone methyltransferases." *EMBO J* **28**(20): 3185-3195.

V

- Vagin, V. V., M. S. Klenov, et al. (2004). "The RNA interference proteins and vasa locus are involved in the silencing of retrotransposons in the female germline of *Drosophila melanogaster*." *RNA Biol* **1**(1): 54-58.

- Vagin, V. V., A. Sigova, et al. (2006). "A distinct small RNA pathway silences selfish genetic elements in the germline." *Science* **313**(5785): 320-324.
- Vagin, V. V., J. Wohlschlegel, et al. (2009). "Proteomic analysis of murine Piwi proteins reveals a role for arginine methylation in specifying interaction with Tudor family members." *Genes Dev* **23**(15): 1749-1762.
- Vakoc, C. R., S. A. Mandat, et al. (2005). "Histone H3 lysine 9 methylation and HP1gamma are associated with transcription elongation through mammalian chromatin." *Mol Cell* **19**(3): 381-391.
- Vaucheret, H., C. Beclin, et al. (1998). "Transgene-induced gene silencing in plants." *Plant J* **16**(6): 651-659.
- Vermaak, D., S. Henikoff, et al. (2005). "Positive selection drives the evolution of rhino, a member of the heterochromatin protein 1 family in *Drosophila*." *PLoS Genet* **1**(1): 96-108.
- Volpe, A. M., H. Horowitz, et al. (2001). "*Drosophila* rhino encodes a female-specific chromo-domain protein that affects chromosome structure and egg polarity." *Genetics* **159**(3): 1117-1134.
- Volpe, T. A., C. Kidner, et al. (2002). "Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi." *Science* **297**(5588): 1833-1837.

W

- Wang, G. K., J. Q. Zhu, et al. (2010). "Circulating microRNA: a novel potential biomarker for early diagnosis of acute myocardial infarction in humans." *Eur Heart J* **31**(6): 659-666.
- Wang, S. H. and S. C. Elgin (2011). "*Drosophila* Piwi functions downstream of piRNA production mediating a chromatin-based transposon silencing mechanism in female germ line." *Proc Natl Acad Sci U S A* **108**(52): 21164-21169.
- Watanabe, T., S. Chuma, et al. (2011). "MITOPLD is a mitochondrial protein essential for nuage formation and piRNA biogenesis in the mouse germline." *Dev Cell* **20**(3): 364-375.
- Watanabe, T., Y. Totoki, et al. (2008). "Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes." *Nature* **453**(7194): 539-543.
- Wysocka, J., C. D. Allis, et al. (2006). "Histone arginine methylation and its dynamic regulation." *Front Biosci* **11**: 344-355.

X

- Xue, Y., X. Bai, et al. (2000). "*Saccharomyces cerevisiae* RAI1 (YGL246c) is homologous to human DOM3Z and encodes a protein that binds the nuclear exoribonuclease Rat1p." *Mol Cell Biol* **20**(11): 4006-4015.

Y

- Yang, Z., Y. W. Ebright, et al. (2006). "HEN1 recognizes 21-24 nt small RNA duplexes and deposits a methyl group onto the 2' OH of the 3' terminal nucleotide." *Nucleic Acids Res* **34**(2): 667-675.

- Yin, H. and H. Lin (2007). "An epigenetic activation role of Piwi and a Piwi-associated piRNA in *Drosophila melanogaster*." *Nature* **450**(7167): 304-308.
- Yoon, J., K. S. Lee, et al. (2008). "dSETDB1 and SU(VAR)3-9 sequentially function during germline-stem cell differentiation in *Drosophila melanogaster*." *PLoS One* **3**(5): e2234.

Z

- Zamparini, A. L., M. Y. Davis, et al. (2011). "Vreteno, a gonad-specific protein, is essential for germline development and primary piRNA biogenesis in *Drosophila*." *Development* **138**(18): 4039-4050.
- Zhang, D., H. Xiong, et al. (2008). "Functional insight into Maelstrom in the germline piRNA pathway: a unique domain homologous to the DnaQ-H 3'-5' exonuclease, its lineage-specific expansion/loss and evolutionarily active site switch." *Biol Direct* **3**: 48.
- Zhang, Y. and D. Reinberg (2001). "Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails." *Genes Dev* **15**(18): 2343-2360.
- Zhang, Z., J. Xu, et al. (2011). "Heterotypic piRNA Ping-Pong requires qin, a protein with both E3 ligase and Tudor domains." *Mol Cell* **44**(4): 572-584.
- Zhao, J., T. K. Ohsumi, et al. (2010). "Genome-wide identification of polycomb-associated RNAs by RIP-seq." *Mol Cell* **40**(6): 939-953.
- Zhao, J., B. K. Sun, et al. (2008). "Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome." *Science* **322**(5902): 750-756.
- Zhao, T., T. Heyduk, et al. (2000). "Heterochromatin protein 1 binds to nucleosomes and DNA in vitro." *J Biol Chem* **275**(36): 28332-28338.
- Zhao, Z., G. Tavoosidana, et al. (2006). "Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions." *Nat Genet* **38**(11): 1341-1347.
- Zhong, H., H. R. Wang, et al. (2010). "Targeting Smad4 links microRNA-146a to the TGF-beta pathway during retinoid acid induction in acute promyelocytic leukemia cell line." *Int J Hematol* **92**(1): 129-135.
- Zhou, R., B. Czech, et al. (2009). "Processing of *Drosophila* endo-siRNAs depends on a specific Loquacious isoform." *RNA* **15**(10): 1886-1895.

Résumé

Les éléments transposables (ET) sont des séquences d'ADN trouvées chez tous les organismes vivants, et capables de se déplacer d'un site chromosomique à un autre. Ils sont une source de mutations et doivent donc être finement contrôlés par leurs hôtes. Afin de parer à leur mobilisation, les génomes ont mis en place des mécanismes de régulation (RNAi) impliquant des petits ARNs dont les siRNAs en sont la composante la plus connue. Récemment il a été mis en évidence chez la drosophile deux nouvelles classes de petits ARNs appelés piRNAs et endo-siRNAs. Les piRNAs contrôlent spécifiquement les ET dans les tissus reproducteurs qui comprennent des cellules germinales et des cellules somatiques : les cellules folliculaires. Les endo-siRNAs quant à eux, contrôlent ces ET dans les tissus somatiques. Il a été montré au laboratoire qu'Idefix, un retrotransposon à LTR, était régulé par un mécanisme post-transcriptionnel (PTGS). Celui-ci implique la voie des piRNAs et sa composante majeure, la protéine PIWI, dans les tissus reproducteurs de drosophiles. En revanche dans le reste des tissus de drosophiles cette régulation ne dépend pas de la protéine PIWI. Durant ma thèse je me suis intéressé à savoir si en plus de ce contrôle de type PTGS, il existait une régulation de type transcriptionnelle (TGS) appliquée sur les ET de drosophile dans les différents tissus aussi bien somatiques que germinaux. En étudiant les régulations que subissent divers transgènes composés d'un gène rapporteur et de divers fragments d'ET, j'ai montré que seule une régulation de type post-transcriptionnelle permettait de réguler les éléments transposables dans la lignée germinale femelle de drosophile. Cette régulation ayant une faiblesse précoce dans le développement des ovaires pouvant entraîner une mobilisation des éléments transposables dans certaines conditions sensibilisées. En revanche dans les tissus somatiques j'ai montré qu'une régulation transcriptionnelle s'ajoutait à la répression de type PTGS pour réprimer les ET. Cependant cette régulation transcriptionnelle présente une spécificité tissulaire puisqu'elle est observée dans les tissus somatiques de larves de drosophiles et absente dans les cellules somatiques folliculaires de l'ovaire. En conclusion divers systèmes de régulation mettent sous silence les éléments transposables en fonctions de la balance bénéfique/problèmes qu'ils apportent pour un tissu donné.

Transposable elements (TEs) are DNA sequences found in all living organisms, and able to move from one chromosomal site to another. They are source of mutations and therefore must be finely controlled by their hosts. To counteract their mobilization, genomes have developed regulatory mechanisms (RNAi) involving small RNAs including the best-known siRNAs. Recently two novel classes of small RNAs called piRNAs and endo-siRNAs have been reported in *Drosophila*. The piRNAs specifically trigger TE repression in reproductive tissues, composed by germ cells and somatic follicular cells. The endo-siRNAs control those in somatic tissues. It has been shown by our group that Idefix, a LTR retrotransposon, is regulated by a posttranscriptional mechanism (PTGS). It implicates the piRNAs pathway and its major component, the PIWI protein, in reproductive tissues of *Drosophila*. By contrast, in the other *Drosophila* tissues, the regulation does not depend on the PIWI protein. During my PhD, I was interested to know if in addition to this PTGS, a transcriptional control (TGS) was necessary to control *Drosophila* TE in both the somatic and germinal tissues. By studying the regulations of sensor-transgenes carrying a reporter gene (GFP) and various fragments of ET acting as a target of the silencing pathways, I have shown that the post-transcriptional silencing is the only regulatory pathway targeting transposable elements in the *Drosophila* female germline. This regulation has a weakness early in the development of the ovaries that can lead to a mobilization of transposable elements under certain sensitized conditions. In somatic tissues I have shown that a transcriptional regulation is coupled to the PTGS. However, this transcriptional regulation has tissue specificity because it is only observed in somatic tissues of *Drosophila* outside of the ovaries, a PTGS with no TGS targeting TE in the somatic cells of the ovarian follicle.