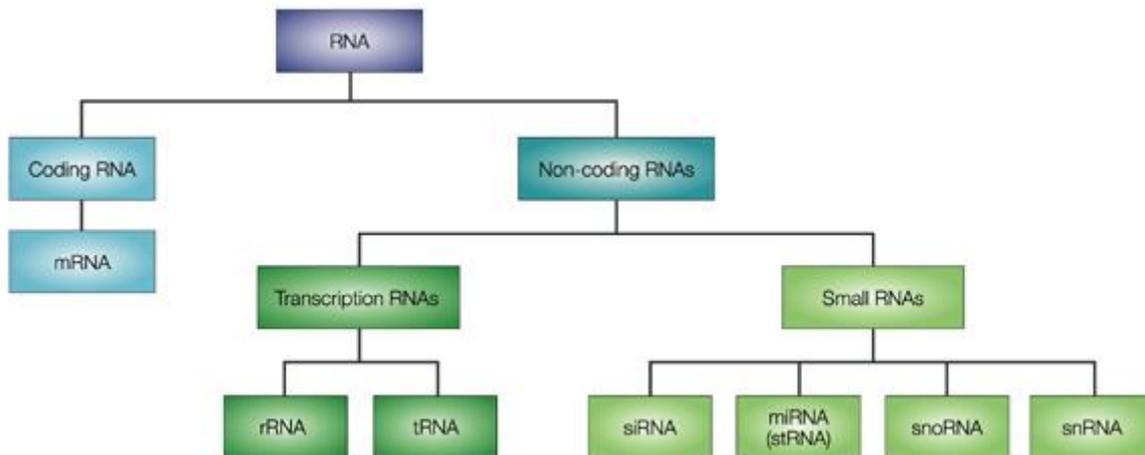


## Les ADN NON CODANTS



### Introduction

Les ARNs non codants constituent une classe d'acides nucléiques de plus en plus étudiée. D'abord pensés comme étant une catégorie anecdotique de produits de la transcription, il apparaît clairement, que sous la forme de petites molécules (**micro ARNs** -ou miARNs-, **siARNs** et **piARNs**) ou de molécules beaucoup plus longues (**long-non coding RNA** ou lncRNAs), ils jouent un rôle crucial dans la régulation épigénétique de l'expression des gènes, dans le contexte général de l'interférence à ARN. Des molécules synthétiques mimant leur action peuvent aisément être produites (siARNs ou mimes de miARNs) et vont être capables d'inhiber l'expression des gènes (voire de miARNs naturels). L'intérêt thérapeutique de leur utilisation est donc évident dans quasiment tous les domaines de la médecine. Cette utilisation nécessite tout de même de détruire plusieurs verrous techniques ou technologiques, ce qui est clairement en passe de se produire.

Ces molécules varient considérablement en taille, **depuis une vingtaine de nucléotides** jusqu'à plusieurs milliers, voire plusieurs centaines de milliers de nucléotides (Mattick, 2001).

Certaines de ces molécules d'ARN jouent un rôle indéniable dans la physiologie cellulaire et sont considérées comme des gènes, connus de longue date, tels par exemple le gène Xist, acteur majeur de l'inactivation d'un des chromosomes X chez les mammifères placentaires femelles (Avner & Heard, 2001; Heard, 2004) ou encore le gène H19, un des gènes soumis à empreinte les plus étudiés (Ripoche et al. 1997).

Le gène H19 produit un ARN non codant fortement exprimé au cours du développement embryonnaire. Découvert il y a plus de 20 ans [1, 11], il fut, avec le gène voisin Igf2 (insulin-like growth factor 2), parmi les premiers gènes décrits comme étant soumis à l'empreinte parentale [2, 3], mécanisme épigénétique qui conduit à une expression monoallélique de ces gènes, dépendante de l'origine parentale de l'allèle. Le gène H19 est ainsi exclusivement exprimé à partir de l'allèle transmis de la mère.

Le locus H19 pourrait donc exercer ses différents rôles soit par son long ARN non codant, soit par l'expression de ce microARN. Une étude récente [9] a montré que, au cours du développement embryonnaire, ce miR-675 est exprimé exclusivement dans le placenta en fin de gestation. Il est totalement réprimé dans les tissus embryonnaires, y compris dans le foie ou le cœur, tissus dans lesquels le précurseur de ce microARN, l'ARN H19, de forme longue, est fortement exprimé. Ceci suggère que la production de ce microARN à partir de la forme

longue de l'ARN H19 est totalement inhibée dans les tissus embryonnaires et le placenta en début de gestation, mais que cette inhibition est un processus dynamique et qu'elle peut être levée en fin de gestation dans le placenta pour permettre son expression.

Plus récemment, le séquençage à haut débit a révélé l'existence de longs ARNs non codants ou lncRNA (long noncoding RNA) de plus de 200 nucléotides, ces molécules sont suspectées jouer un rôle majeur dans la régulation épigénétique de l'expression des gènes.

## **1 - Les petits ARN régulateurs**

L'interférence ARN ("*RNA interference*") ou RNAi est une voie de régulation du taux d'ARN messagers traduits ou plus précisément de l'expression post-transcriptionnelle des gènes

### **1.1 - Découverte et première description**

Si les ARN non codants ont des fonctions diverses, un groupe particulier semble se dégager, celui des ARN régulateurs. Ce sont des ARN de taille variable qui sont impliqués dans la régulation de l'expression des gènes. Leur taille va déterminer les mécanismes impliqués. Les premiers à avoir été décrits sont les petits ARN participant à « l'ARN interférence ». L'introduction du terme « ARN interférence » vient des travaux de Fire et Melo (Fire *et al*, 1998) qui ont mis en évidence chez *C. elegans* un mécanisme de régulation de l'expression des gènes par des ARN double brin. Les auteurs ont montré que l'introduction par micro-injection dans les cellules de *C. elegans* d'ARN double brin induit une inhibition de l'expression de tout gène qui contient une région identique en séquence avec celle des ARN double brin. Ils ont aussi établi que ce mécanisme est post-transcriptionnel et affecte l'accumulation des ARN messagers correspondants. Ainsi la micro-injection dans quelques cellules de *C. elegans* est suffisante pour induire un « silencing » aussi appelé PTGS (« Post-Transcriptional Gene Silencing ») qui se réfère à un groupe de mécanismes par lequel l'expression d'un ou plusieurs gènes est régulée négativement ou entièrement supprimée par l'introduction d'un ARN antisens.

Ces observations où l'ARN semble être l'acteur principal d'une régulation génique constituaient clairement une rupture des dogmes de la biologie moléculaire. Ou les protéines sont l'acteur principal de toutes les régulations de l'expression des gènes.

### **1.2 - L'interférence par les siARN**

Le silencing par les siARN a été le premier à être décrit. Celui-ci est impliqué dans la réponse antivirale de la cellule induite par la présence d'ARN double brin dans le cytoplasme. Il est aujourd'hui utilisé comme un outil moléculaire pour le silencing de l'expression des gènes dans les études fonctionnelles.

Comme pour les autres petits ARN régulateurs, le mécanisme d'action de ces petits ARN va reposer sur sa capacité à interagir avec un complexe de protéines régulatrices et à le guider vers un ARN « cible ».

Depuis 1998, il a été montré chez de nombreux eucaryotes que l'ARN double brin pouvait induire une suppression de l'expression génique par des mécanismes spécifiques de

séquence. L'inhibition de l'expression peut avoir lieu à différents niveaux (transcription, traduction, dégradation de l'ARNm).

Initialement, l'appellation d'interférence par l'ARN désignait la dégradation ciblée des ARN messagers induite par l'ARN double brin [1]. Mais il est apparu ensuite que le blocage de la traduction constituait probablement un autre aspect du même mécanisme.

Les petits ARNm et ARNsi ne sont pas codés par la même partie du génome. Les premiers (ARNmi) proviennent de gènes qui leur sont propres et dont ils sont les uniques produits.

Les seconds (ARNsi) dérivent de deux sources, exogène ou endogène.

A-La source exogène correspond soit à l'apport extérieur d'un ARN double brin par injection expérimentale, soit à un apport en provenance de génomes à ARN double brin comme ceux de certains virus à ARN des plantes.

B- La source endogène correspond à une production naturelle par utilisation des parties exoniques (codantes ou non codantes) d'un gène plus étendu.

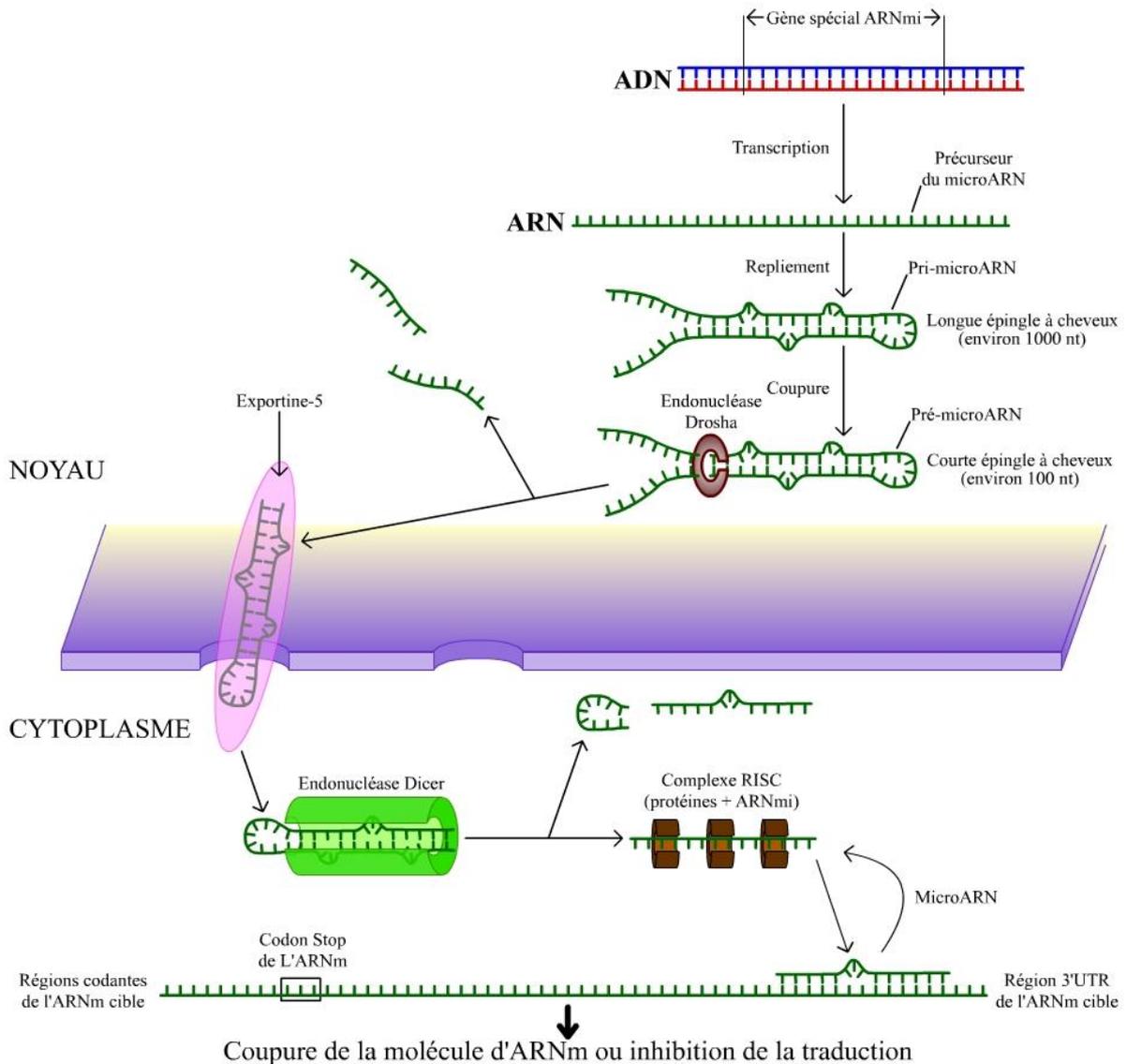
Concernant les ARNm, leur transcrite est d'abord, pour chacun, une molécule d'ARN simple brin d'un millier de nucléotides. Chacune de ces molécules présente, le long de sa séquence, des motifs palindromiques, partiellement complémentaires, qui les conduisent à adopter une structure secondaire en forme d'une longue épingle à cheveux.

Cette structure peut être divisée en trois parties : le « corps » qui est formé de séquences plus ou moins complémentaires présentes sur le simple brin, se disposant en vis-à-vis par appariement, la « tête » (en forme de boucle) et les « jambes », simple brin de l'épingle qui portent les séquences n'ayant pas trouvé de complémentaires. Se produit alors la série de transformations suivante.

- L'enzyme Drosha (ribonucléase de type III) y effectue une première série de coupures qui réduit cette structure en une courte molécule d'une centaine de nucléotides, toujours en épingle à cheveux (*short hairpin RNA* : shRNA), mais amputée de ses « jambes ». L'exportine-5 la transporte alors dans le cytoplasme.
- Là, une seconde ribonucléase de type III, appelée Dicer, achève le travail commencé par l'enzyme Drosha : par une deuxième série de coupures, elle élimine la « tête » de l'épingle à cheveux tandis que d'autres protéines qui lui sont associées (de type hélicase) dissocient le petit ARNd (double brin) subsistant, d'une vingtaine de nucléotides, en deux ARN simple brin. L'enzyme Dicer clive et sépare également les longs ARN double brin des ARNsi en morceaux d'environ 21 nucléotides.
- L'un des simples brins obtenus est le microARN (ARNmi) ou le petit ARN interférent (ARNsi). Il s'associe à un complexe protéique en formant un nouveau complexe appelé RISC (*RNA Induced Silencing Complex*). C'est en cet équipage qu'il se porte sur sa cible : l'ARNm auquel il s'apparie.
- En cas d'appariement parfait (ARNsi), l'ARNm est détruit et il n'y a pas traduction. Les RISC qui sont à l'origine de cette destruction restent ensuite parfaitement fonctionnels, ce qui leur permet d'opérer de nouveau sur d'autres ARNm de même

spécificité. C'est cette réutilisation qui les rend particulièrement offensifs. Mais même lorsque l'appariement est imparfait (ARNmi), il y a tout de même barrage à la traduction de l'ARNm.

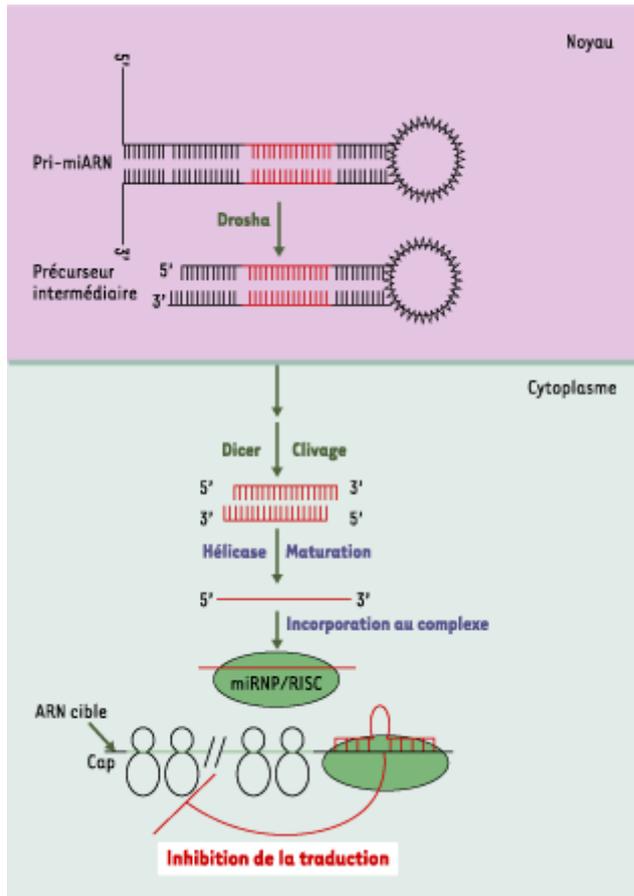
- À travers l'exemple des ARNmi, la figure ci-dessous résume le mode d'action de ces petits ARN.
- 



### 1-3 Les microARN

Les microARN sont des ARN non codants de 21 à 25 nucléotides qui contrôlent l'expression génique au niveau post-transcriptionnel. Plusieurs centaines de gènes codant pour des microARN ont été identifiés chez les animaux, et une quarantaine chez les plantes. Certains de ces gènes sont conservés entre espèces et parfois même entre phylums. Ces microARN règlent l'expression génique en s'appariant avec des ARNm cibles dont ils sont partiellement complémentaires. Cette hybridation réprime la traduction de la protéine correspondante ou clive l'ARNm cible au milieu du site de fixation du microARN. Ce dernier mécanisme est très similaire à celui mis en œuvre lors de l'interférence par l'ARN.

Le pri-microARN double brin (Pri-miARN), transcrit à partir du gène, donne naissance à un ARN double brin (précurseur intermédiaire) par action d'une ribonucléase Droscha. Après une exportation active du précurseur dans le cytoplasme, une forme double brin du microARN est produite par l'action de la protéine Dicer. Une étape de maturation supplémentaire, à laquelle participe vraisemblablement une hélicase, est ensuite nécessaire avant que le microARN simple brin mature se fixe sur le complexe miRNP/RISC (*microribonucleoprotein/RNA-induced silencing complex*). Ce complexe guide ensuite le microARN vers la région 3' non codante (3'UTR) de l'ARNm cible. Après hybridation imparfaite entre le microARN et la cible, la traduction de la protéine correspondante est inhibée.



Les gènes des miARN sont généralement transcrits par l'ARN polymérase II qui produit un transcrit coiffé et polyadénylé, le pri-miARN. Ce pri-miARN possède des structures en tige-boucle qui vont être clivées à leur base par la RNase III Droscha associée à DGCR8 générant les pre-miARN, précurseurs de 70 nucléotides en tige-boucle. Après avoir été exporté dans le cytoplasme, le pre-miARN est coupé par Dicer pour donner un miARN double brin d'environ 22 nucléotides. Un de ces deux brins miARN est incorporé au complexe RISC et permet de guider le complexe effecteur sur l'ARNm cible. Le complexe miRISC va modifier le devenir de l'ARNm de plusieurs manières à travers une inhibition de la traduction de l'ARNm cible mais également sa déadénylation et son décoiffage conduisant à sa dégradation par les exonucléases de la cellule.

### **1.3 – L’interférence par les piARN**

Les « PIWI-interacting small RNAs » ou piARN sont la classe la moins bien caractérisée des petits ARN régulateurs liés à l’interférence ARN. Ils ont pourtant un rôle crucial de « gardien du génome » en maintenant une répression de l’ensemble des éléments transposables. Le manque de connaissance sur ce mécanisme provient sans doute du fait que ce mécanisme est restreint aux cellules de lignée germinale, même si quelques travaux suggèrent qu’ils pourraient avoir des fonctions non germinales chez l’Aplysie (Rajaseethupathy *et al*, 2012), la drosophile (Rouget *et al*, 2010) ou les mammifères (Watanabe *et al*, 2011). En dépit de cette spécialisation, cette voie de régulation a été observée très tôt dans l’histoire de l’interférence ARN.

D’une taille plus importante que celle des siARN et miARN (23-32 nucléotides). Ces petits ARN étant dérivés de séquences d’éléments répétés dans le génome, ils furent alors nommés dans un premier temps « repeat-associated small interfering RNAs » ou rasiARN.

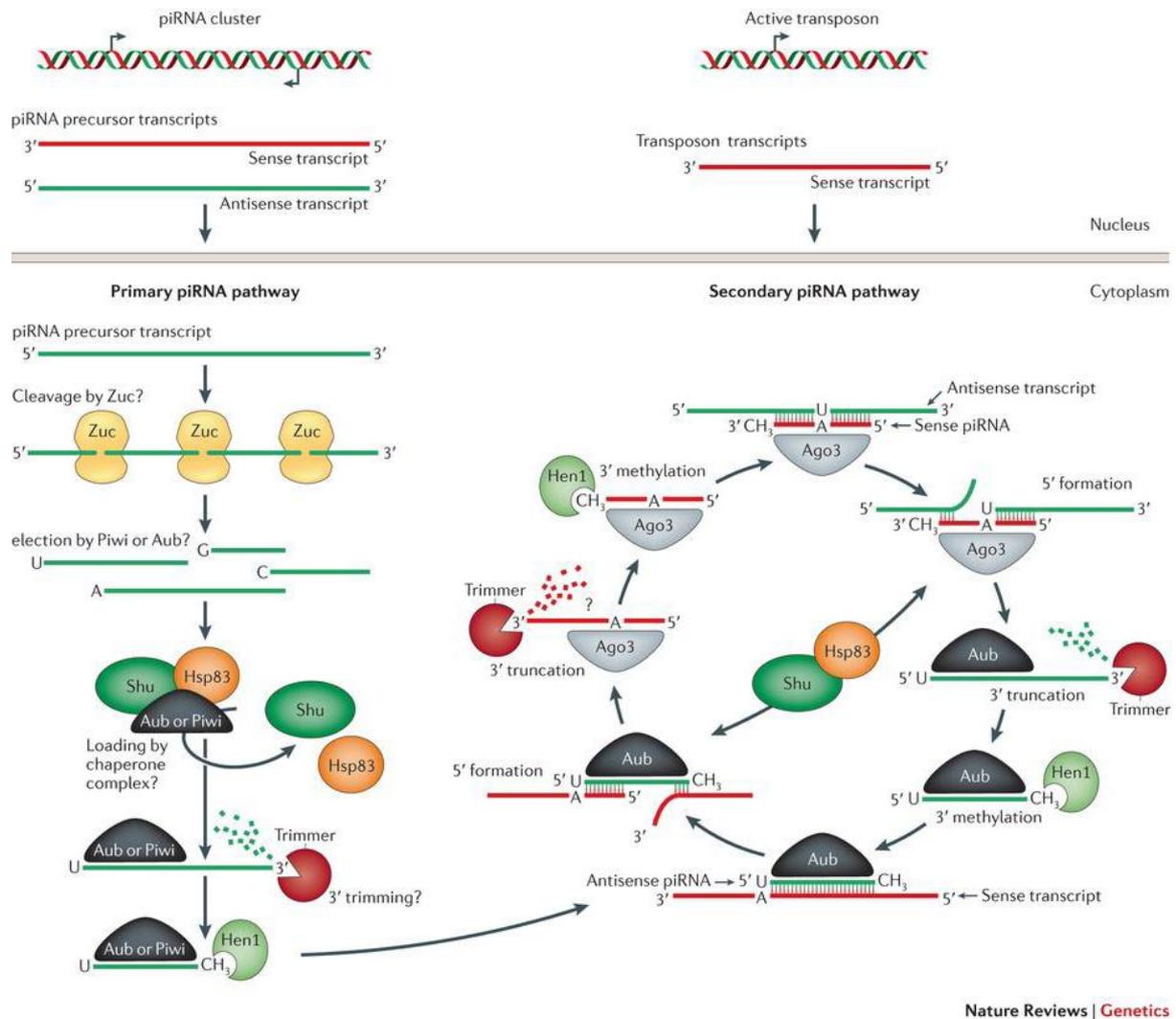
Les piARN sont codées à partir de longues séquences intergéniques dans le génome (> 100 000 bases) aussi appelées piARN clusters et souvent très riches en éléments transposables. La majorité des piARN sont issus du brin antisens des séquences des éléments transposables permettant un silencing via l’hybridation avec le transcrit des transposons (Saito et Siomi, 2010).

Contrairement à la biogénèse des siARN et miARN, le précurseur piARN est un ARN simple brin dont la maturation va être indépendante d’enzymes de la famille des RNases III de type Dicer ou Drosha (Vagin *et al*, 2006).

L’autre différence majeure avec les siARN et miARN est que les piARN ne sont pas associés aux protéines de la sous-famille Argonaute qui sont exprimées de manière ubiquitaire mais aux protéines de la sous-famille PIWI. La sous-famille PIWI est principalement exprimée dans la lignée germinale.

L’interaction des piARN avec les protéines PIWI permet la formation du complexe piRISC qui participe à deux mécanismes distincts. Un mécanisme primaire qui va permettre le silencing des éléments transposables. Au niveau transcriptionnel et un second qui va permettre l’amplification de la réponse contre ces éléments via la génération de nouveaux piARN à partir des transcrits cibles.

Le transcrit précurseur piARN ou pre-piARN est transcrit à partir d’un piARN cluster. Cet ARN est exporté et mûré dans le cytoplasme. La première étape de la maturation sera être coupée pour donner des « piRNA like small RNA » qui sont les piARN non encore disponibles pour le complexe RISC.



Exemple des deux voies de biogénèse des piARN chez *Drosophila melanogaster*.

Dans la voie de synthèse primaire, le précurseur issu des gènes piARN est clivé par l'endonucléase Zucchini (Zuc). Aidés par deux protéines chaperonnes Shutdown (Shu) et Heat shock protein 83 (Hsp83), les fragments sont incorporés dans les protéines Piwi ou Aubergine. De préférence les fragments aux extrémités 5'U sont incorporés.

Une fois incorporé, le précurseur piARN continue d'être maturé en subissant un raccourcissement de son extrémité 3' par une exonucléase inconnue puis une méthylation de cette même extrémité par l'enzyme Hen1. Cela clôt la biogénèse primaire.

Si le piARN est associé avec la protéine Aubergine (mais pas Piwi) celui-ci peut participer à la synthèse secondaire. En reconnaissant le transcrit cible, Aubergine permet sa coupure. L'extrémité 5' libérée sur le transcrit cible va alors servir de nouveau précurseur et sera incorporée à la protéine Ago3. Ago3 permet le même mécanisme de coupure donnant à son tour des précurseurs pour de nouveaux piARNs à Aubergine permettant l'amplification de la réponse.

## 2 – Les longs ARN non codants (lncRNA)

### 2.1 – Première description et caractéristiques

La première description d'un long ARN non codant ayant une activité de régulation de l'expression génique est celle de *Xist* (Bursani *et al*, 1991 ; Brown *et al*, 1991 ; Brockdorff *et al*, 1992). Ce dernier est impliqué dans l'inactivation du chromosome X surnuméraire chez les femelles des mammifères. *Xist* est un ARN de 17 kb exprimé à partir du locus XIC (X-inactivation center) du chromosome X inactif.

L'expression de *Xist* est suffisante à elle seule pour permettre l'induction en *cis* d'un silencing complet du chromosome X (Wutz *et al*, 2000). Cette première description a permis d'identifier quelques caractéristiques des longs ARN non codants qui ont été reprises dans les travaux plus récents.

Les longs ARN non codants sont transcrits par la polymérase II, sont coiffés et peuvent être polyadénylés et épissés. Par convention ils ont une taille supérieure à 200 nucléotides qui peut atteindre plusieurs centaines de kilobases. De plus, comme leur nom l'indique, ces transcrits ne codent pas de protéines, mais il est difficile de donner un sens absolu à cette définition puisqu'un ARN de grande taille contient nécessairement des cadres ouverts de lecture (Dinger *et al*, 2008).

On retrouve des longs ARN non codants dans l'ensemble des tissus, certains étant exprimés de façon tissu spécifique (Mercer *et al*, 2008). Leur localisation dans la cellule va aussi varier et, probablement, reflète leur fonction. Un exemple bien étudié est « Gomafu » ou RNCR2 (Retinal Non Coding RNA 2). Cet ARN non codant s'exprime uniquement dans les cellules du système nerveux (Ishii *et al*, 2006) et s'accumule dans le noyau au niveau des « nuclear speckles » (Sone *et al*, 2007). Cette localisation bien spécifique correspond à sa fonction proposée de régulateur de l'épissage (Tsuiji *et al*, 2011).

Compte tenu de la diversité potentielle des longs ARN non codants, de leurs fonctions, de leurs mécanismes d'action et de leurs interactions avec leurs cibles, plusieurs classifications des longs ARN non codants ont été proposées, le plus souvent en fonction de leurs positionnements par rapport aux gènes codants déjà identifiés dans le génome.

- **Selon leur taille** : les long ARN non codants (lncARN) de taille supérieure à 200pb se distinguent des très long ARN non codants (vlincARN et macro lncARN) de taille supérieure à 10kb (Guenzl and Barlow, 2012; St Laurent et al., 2016).
- **Selon leur localisation** dans une région particulière du génome : certains sont par exemple des pseudogènes qui correspondent à d'anciens gènes qui ont perdu leur potentiel codant au cours de l'évolution. Les lncARN dérivés de pseudogènes peuvent réguler l'expression du gène codant en agissant comme des éponges de miRNA, comme c'est le cas de PTENP1, un lncARN pseudogène dérivé du gène suppresseur de tumeur PTEN. La présence de PTENP1 régulerait l'expression de PTEN par compétition au niveau de la fixation de miRNA sur des séquences identiques aux deux gènes, et ainsi jouerait un rôle protecteur contre les cancers.
- Les expressions de ces deux gènes ont été retrouvées diminuées dans les cancers de la prostate et du colon (Poliseno et al., 2010). D'autres par association aux télomères des

chromosomes, les TERRA (telomeric repeat containing RNA) favorisent la réponse aux dommages de l'ADN télomérique et par conséquent, retardent la sénescence (Porro et al., 2010).

- **Selon leur association avec des éléments génomiques connus** : certains lncARN sont associés à des enhanceurs, il s'agit de longs transcrits de séquences enhanceurs appelés eLncARN. Ils régulent l'expression de gènes avoisinants en agissant en cis sur leurs enhanceurs pour recruter des facteurs de transcription. Par exemple, le eLncARN Evf2 possède les doubles fonctions d'activateur ou de répresseur des gènes à proximité selon les protéines avec lesquelles il s'associe et il joue un rôle critique dans le développement du cerveau (Bond et al., 2009).
- **D'autres lncARN sont associés à des promoteurs et sont appelés des PALR** (Promoter-associated long RNA). Par exemple, des dommages de l'ADN induisent l'expression du PALR du gène CCND1 (Cycline D1) permettant la répression de l'expression des gènes adjacents (Song et al., 2012).
- **Selon leur localisation subcellulaire** : la majorité des lncARN sont retrouvés dans le noyau mais certains sont localisés dans le cytoplasme ou encore dans les mitochondries (voir paragraphe Localisation cellulaire)
- **Selon leur fonction** : de nombreuses fonctions de régulations sont attribuées au lncARN et sont détaillées dans le paragraphe Mécanismes d'action.
- **Selon leur organisation chromosomique par rapport aux gènes adjacents** : il s'agit d'une classification en fonction de la localisation chromosomique des lncARN qui prend en compte leur position par rapport au gène codant annoté le plus proche. A partir de cette classification, 6 types de lncARN sont définis dans la figure suivante (Figure 11) (Moran et al., 2012).

**1-Classe 1 = lncARN intergénique** (LincARN) situé à plus de 1 kb du gène codant le plus proche indépendamment du sens de transcription de celui-ci. La distance minimale de 1kb entre le lncARN et le gène adjacent correspond à la taille moyenne d'un promoteur proximal et a été choisi arbitrairement pour définir la classe de lncARN intergénique (Guttman et al., 2009; Moran et al., 2012).

**2- Classe 2 = lncARN divergent**. Le sens de sa transcription diverge du gène codant le plus proche (brin d'ADN opposé) et se situe à moins d'1kb de celui-ci sans toutefois le chevaucher.

**3-Classe 3 = lncARN convergent**. Le sens de sa transcription converge vers le gène codant le plus proche (brin d'ADN opposé) et se situe à moins d'1kb de celui-ci sans toutefois le chevaucher.

**4-Classe 4 et 5 = lncARN chevauchant en partie ou entièrement la séquence d'un autre gène** et sa transcription est orientée dans le même sens (4) que celui-ci (même brin d'ADN) ou en sens inverse (brin d'ADN opposé) (5)

5- **lncARN intronique** chevauchant un intron d'un autre gène (la double flèche indique que le lncARN peut être orienté dans les deux sens). Cette classe de lncARN a été moins étudiée que les lincARN et les lncARN-AS durant ces dernières années car ils ont longtemps été considérés comme des produits d'épissage alternatif d'ARNm. De plus, ils semblent moins stables que les ARNm et les lncARNAS (Ayupe et al., 2015).

## **Localisation cellulaire**

Les ARNnc sont présents chez tous les eucaryotes, des organismes unicellulaires aux mammifères supérieurs.

Contrairement aux ARNm matures majoritairement localisés dans le cytoplasme, les lncARN sont retrouvés :

- dans le noyau au niveau de la chromatine ou dans le nucléoplasme : la plupart des lncARN sont nucléaires
- dans le cytoplasme : certains sont exclusivement cytoplasmiques
- dans les mitochondries : le génome mitochondrial code un certain nombre de lncARN dont lncND5, lncND6 et lncCyt b, qui semblent impliqués dans la stabilité des ARN mitochondriaux adjacents (ND5, ND6 et Cyt b) (Rackham et al., 2011).
  - Des changements dans l'environnement, comme un stress oxydatif, peuvent aussi influencer la localisation subcellulaire des lncARN (Giannakakis et al., 2015).
  - Les modifications de compartiments cellulaires existent aussi chez les lncARN, par exemple, les protéines HuR (Hu-Antigen R) et GRSF1 (GRich RNA Sequence Binding Factor 1) régissent la localisation cytoplasmique et mitochondriale du lncARN RMRP (RNA Component Of Mitochondrial RNA Processing Endoribonuclease) codé par l'ADN nucléaire mais ayant des fonctions clés dans les mitochondries (Noh et al., 2016).
  - Ainsi, il semble évident que la répartition de lncARN au sein d'une cellule influe sur leur fonction.

## **Leur fonction**

**a.** Les lncARN peuvent inactiver de grandes régions génomiques voir des chromosomes entiers en faisant intervenir des modulateurs épigénétiques. (Exemple du lncARN XIST détaillé précédemment).

**b.** Ils peuvent jouer le rôle d'enhancer en induisant la transcription en cis ou en trans. Par exemple, le lncARN Evf2 transcrit dans la région enhancer de ses cibles, interagit avec le facteur de transcription DLX (Distal-Less Homeobox) et la protéine de liaison méthyl-CpG 2 (MECP2) pour réguler de manière épigénétique en trans l'expression des gènes Dlx5 et Dlx6 dans le développement des interneurons GABAergiques (Bond et al., 2009).

**c.** Ils peuvent agir comme un leurre aux protéines régulatrices, telles que les facteurs de transcription et les modificateurs de la chromatine, bloquant ainsi leur liaison à l'ADN. Par exemple, le lncARN MALAT1 (Metastasis-Associated Lung Adenocarcinoma Transcript 1) est fortement exprimé dans les neurones et régule la formation des synapses en modulant l'expression des gènes impliqués dans la formation et / ou le maintien des synapses (Bernard et al., 2010). MALAT1 interagit avec les facteurs d'épissage SR (protéines riches en serine et arginine) et module leur phosphorylation et leur liaison à l'ADN, modifiant le schéma d'épissage d'un sous-ensemble de préARNm endogènes (Tripathi et al., 2010).

**d.** Ils peuvent servir de signaux moléculaires pour activer ou réduire l'expression des gènes en signalant des voies de régulation. Pour le cas de Xist, la présence du lncARN transcrit sur la surface du chromosome X déclenche l'inhibition de la transcription des gènes présents sur

ce chromosome suite au recrutement des complexes RNP répressifs ayant une activité méthyl transférase sur les histones.

**e.** Ils peuvent guider des protéines, en général des modulateurs de la chromatine vers des sites spécifiques. (Exemple du lncARN HOTAIR détaillé précédemment).

**f.** Ils peuvent agir comme des échafaudages moléculaires, liant différentes protéines et formant des complexes ribonucléoprotéiques (RNP), qui affectent également l'expression des gènes. (Exemple du lncARN HOTAIR détaillé précédemment).

**g.** Ils interagissent avec des enzymes, telles que des kinases, régulant ainsi leur activité catalytique et modifiant leur signalisation. Par exemple, à la suite d'un stress énergétique, le lncARN NBR2 (neighbor of BRCA1 gene 2) interagit avec l'AMPK (AMP-activated protein kinase) et favorise son activité, augmentant ainsi la fonction anti-tumorale de l'AMPK (Liu et al., 2016).

**h.** Ils peuvent moduler l'épissage alternatif d'autres transcrits. Par exemple, MALAT1 décrit précédemment (c) comme un leurre, qui par son action sur les facteurs d'épissage, modifie l'épissage de certains pré-ARNm (Tripathi et al., 2010). Un autre exemple, le lncARN MIAT (Myocardial Infarction Associated Transcript ou GOMAFU) présente une séquence répétée en tandem (UACUAC) similaire à la séquence BPS (intron branch point séquence) des ARNm permettant de lier le facteur d'épissage SF1 (Splicing Factor 1) (Ishii et al., 2006). Ainsi, MIAT se fixe au facteur d'épissage SF1 avec une plus grande affinité ce qui affecte la cinétique de la réaction d'épissage des ARNm en modifiant la concentration locale des facteurs d'épissage dans le noyau (Tsuiji et al., 2011).

**i.** Ils servent d'éponge à miRNA empêchant ainsi leur effet de répresseur traductionnel et conduisant à l'augmentation du taux de la protéine cible. Il existe une catégorie de lncARN séquestreurs de miRNA : les lncARN circulaires (circRNA) dont les extrémités 5' et 3' sont liées de façon covalentes ou les ARN endogènes concurrents (ceARN). Par exemple, le ceARN linc-MD1 (Muscle Differentiation 1) spécifique du muscle, agit comme une éponge pour deux miRNA différents, miR133 et miR-135. Ces miRNA régulent respectivement l'expression des facteurs de transcription qui activent l'expression de gènes spécifiques du muscle. De cette manière, Linc-MD1 contrôle le moment de la différenciation des myoblastes (Cesana et al., 2011). Un autre exemple est le circRNA, CDR1as (ou ciRS-7, Cerebellar degeneration-related autoantigen 1 antisens) qui contient plus de 70 sites de liaison pour le miR-7 et agit comme une éponge de ce dernier, entraînant une augmentation du taux d'expression de la cible de miR-7 (Hansen et al., 2013).

**j.** Ils peuvent agir sur l'activité d'une protéine ou modifier sa localisation cellulaire en formant des complexes moléculaires avec celles-ci. Voir l'exemple (g)

**k.** En ciblant des ARNm, ils peuvent inhiber ou au contraire augmenter leur traduction par les ribosomes. C'est le cas de l'ARNm Zeb2 (Zinc finger E-box-binding homebox 2) dont la traduction peut avoir lieu uniquement en présence du lncARN Zeb2NAT (Zeb2 Natural Antisens Transcript) transcrit en anti-sens du gène Zeb2 (Beltran et al., 2008). Zeb2NAT chevauche le site de reconnaissance de l'épissage de l'intron du pré-ARNm qui contient une séquence IRES (Internal Ribosome Entry Site). Cette séquence IRES permet l'initiation de la traduction. Ainsi, en évitant l'épissage de cette séquence, Zeb2 peut être traduit en protéine et son taux considérablement augmenté. Un autre exemple de lncARN agissant sur la traduction : il existe deux lncARN, BC1 (brain cytoplasmic RNA 1) et BC200 (200 nt brain cytoplasmic RNA), qui peuvent se lier à eIF4A (eukaryotic translation initiation factor 4A), PABP (poly(A)- binding protein) et d'autres facteurs pour bloquer l'assemblage du complexe d'initiation de la traduction requis (Lin et al., 2008; Muddashetty et al., 2002).

**l.** Certains lncRNA peuvent être transférés à d'autres cellules par des vésicules extracellulaires pour agir sur des molécules à l'intérieure de ses vésicules ou dans d'autres cellules voisines. Des lncARN exosomiques comme Exo2 ou Exo4 sont préférentiellement empaquetés dans des vésicules extracellulaires et peuvent moduler de manière fonctionnelle la viabilité cellulaire par des interactions directes avec la L-lactate déshydrogénase B dans des cellules infectées par le papillomavirus (Hewson et al., 2016). Mais ces lncRNA extracellulaires restent largement inconnus, et la nécessité de davantage de recherches pour caractériser le contenu des vésicules et leurs effets biologiques devient évidente.

**m.** Ils peuvent être précurseurs de miRNA et être clivés par des RNases. Par exemple, le lncARN H19 impliqué dans l'empreinte génétique est aussi le précurseur du miR-675 (Bartolomei et al., 1991). Il peut inhiber l'activation et la prolifération des cellules souches hématopoïétiques, en servant de précurseur du miR-675, qui cible le récepteur Igf1r (insulin-like growth factor 1 receptor) dans le placenta (Keniry et al., 2012).

**En conclusion**, de par leur structure secondaire, les lncARN peuvent se lier à l'ADN, l'ARN, aux miRNA et aux protéines, et agir sur la structure ou les propriétés de ces molécules. Les lncARN se retrouvent impliqués dans la totalité des mécanismes de régulation, ce qui fait d'eux des régulateurs génomiques puissants. L'interrogation fonctionnelle du nombre croissant de nouvelles séquences de lncARN annotées est un défi de taille pour la recherche ces prochaines années. La perturbation de l'expression des lncARN cibles soit par sur-expression, soit par sous-expression (knockdown=KD) est une approche courante pour étudier leur rôle biologique dans la cellule.

### **Implication des lncARN dans les pathologies humaines**

Compte tenu du large éventail de fonctions que révèlent les lncARN dans tous les processus biologiques, il est logique qu'ils soient impliqués dans des pathologies. Des études d'association à l'échelle du génome ont permis d'associer des SNP comme facteurs de risques dans l'étiopathologie de la quasi totalité des grandes pathologies humaines, comme la maladie d'Alzheimer (MA) (Lambert et al., 2013) ou encore la MP (Nalls et al., 2014). De plus, ces études d'association génomique à grande échelle ont révélé que 88% des SNP associés aux maladies humaines se trouvent en dehors des gènes codants des protéines (Brodie et al., 2016; Hindorff et al., 2009). Outre les études GWAS, des analyses ciblées de l'expression de lncARN d'intérêt par RT-PCRq ou de puces à ADN ont permis d'identifier des lncARN impliqués dans divers processus comme la neurogenèse ou la différenciation neuronale (Ng et al., 2012), mais aussi dans des pathologies.

Une analyse des niveaux d'expression de 90 lncARN a mis en évidence une surexpression de différents lncARN, dont le lncARN SNHG1 (Small nucleolar RNA host gene 1), dans les cerveaux de 20 patients souffrant de la MP en comparaison avec 10 patients témoins (Kraus et al., 2017). Puis afin de mieux élucider le rôle de ces lncARN candidats dans les mécanismes qui sous tendent la maladie, des études fonctionnelles sont mises en place. La plupart des études sur l'implication des lncARN dans les pathologies humaines portent sur les cancers (Brunner et al., 2012; Gibb et al., 2011).