

Année 2012

Thèse n° 1949

# THÈSE

pour l'obtention du

## DIPLÔME NATIONAL DE DOCTORAT DÉLIVRÉ PAR L'UNIVERSITÉ BORDEAUX SEGALLEN

**Ecole doctorale Sciences de la vie et de la santé**

**Mention : Sciences, Technologies, Santé**

**Option : Génétique**

**Présentée et soutenue publiquement**

Le 17/12/2012

Par Ilyass ZNIBER

Né(e) le 04/05/1983 à Meknès (Maroc)

### **Développement d'outils pour l'étude *in vivo* de la régulation post-transcriptionnelle chez *Caenorhabditis elegans***

#### **Membres du Jury**

M Benoît ARVEILER,	PU-PH Université de Bordeaux.....	Président
M Jonathan EWBANK,	DR1 CNRS Université de Marseille.....	Rapporteur
M Hervé SEITZ,	CR1 CNRS Université de Montpellier .....	Rapporteur
M Denis DUPUY,	CR1 INSERM Université de Bordeaux .....	Directeur de thèse

## La mort du loup

...  
*Le Loup vient et s'assied, les deux jambes dressées  
Par leurs ongles crochus dans le sable enfoncées.  
Il s'est jugé perdu, puisqu'il était surpris,  
Sa retraite coupée et tous ses chemins pris ;  
Alors il a saisi, dans sa gueule brûlante,  
Du chien le plus hardi la gorge pantelante  
Et n'a pas desserré ses mâchoires de fer,  
Malgré nos coups de feu qui traversaient sa chair  
Et nos couteaux aigus qui, comme des tenailles,  
Se croisaient en plongeant dans ses larges entrailles,  
Jusqu'au dernier moment où le chien étranglé,  
Mort longtemps avant lui, sous ses pieds a roulé.  
Le Loup le quitte alors et puis il nous regarde.  
Les couteaux lui restaient au flanc jusqu'à la garde,  
Le clouaient au gazon tout baigné dans son sang ;  
Nos fusils l'entouraient en sinistre croissant.  
Il nous regarde encore, ensuite il se recouche,  
Tout en léchant le sang répandu sur sa bouche,  
Et, sans daigner savoir comment il a péri,  
Refermant ses grands yeux, meurt sans jeter un cri.*

...  
*A voir ce que l'on fut sur terre et ce qu'on laisse  
Seul le silence est grand ; tout le reste est faiblesse.  
- Ah ! Je t'ai bien compris, sauvage voyageur,  
Et ton dernier regard m'est allé jusqu'au cœur !  
Il disait : " Si tu peux, fais que ton âme arrive,  
A force de rester studieuse et pensive,  
Jusqu'à ce haut degré de stoïque fierté  
Où, naissant dans les bois, j'ai tout d'abord monté.  
Gémir, pleurer, prier est également lâche.  
Fais énergiquement ta longue et lourde tâche  
Dans la voie où le Sort a voulu t'appeler,  
Puis après, comme moi, souffre et meurs sans parler.*

Alfred de Vigny

Je tiens tout d'abord à remercier Monsieur Benoît Arveiler, président de mon jury, et Messieurs Jonhatan Ewbank et Hervé Seitz, mes rapporteurs, de m'avoir fait l'honneur d'être présents dans mon jury pour évaluer le travail que j'ai accompli ces dernières années.

Je voudrais remercier M. Denis Dupuy, mon directeur de thèse, sans qui cette aventure n'aurait pas été possible. Il m'a accordé sa confiance il y a quatre ans de cela. Il m'a enseigné la méthodologie et la rigueur du travail en laboratoire. Il m'a envoyé dans des congrès aux quatre coins du monde pour enrichir mes connaissances. Il m'a permis de travailler sur des sujets de recherche variés allant du fondamental jusqu'au technique... Et il a surtout œuvré pour me permettre de garder le cap lorsque j'ai failli chavirer, quitter le navire et me noyer terrassé par une avalanche de soucis dont j'espère que certains finiront par se régler un jour. Denis, permets moi de te témoigner ma reconnaissance.

Et comme un capitaine à lui seul ne peut manœuvrer un bateau, je voudrais saisir cette occasion pour remercier les autres membres de l'équipe que j'ai côtoyés et à qui je dois beaucoup. Je ne saurais par qui commencer tellement chacun de vous a joué un rôle déterminant dans mon parcours. Mais sachez que l'ordre ici ne traduit d'aucune façon une préférence. Je remercie Rosina Giordiano d'avoir été si agréable, si serviable et si plaisante avec moi, tu m'avais fait promettre que j'irai au bout et voilà j'y suis presque. Je remercie Karine Reborà pour son aide, son soutien, ses connaissances, sa gentillesse... Tu as su à plusieurs occasions me redonner confiance en moi et me faire retrouver le goût de la recherche en me montrant les qualités que tu voyais en moi, tu as aussi été une oreille attentive à qui je pouvais me confier. Nous sommes arrivés au laboratoire au même moment et nous le quittons au même moment alors je saisi cette occasion pour te souhaiter plein d'épanouissement dans tes nouveaux projets professionnels. Je voudrais remercier Esther Marza qui n'a cessé de m'envoyer des bouées de sauvetage à chaque fois que je tombais à l'eau et qui m'a pris sous son aile et a été attentive à mes problèmes et m'a aidé du mieux qu'elle pouvait. J'ai rarement croisé quelqu'un avec autant d'humanité, je te remercie du fond du cœur et te souhaite plein de joie dans ta nouvelle vie parisienne. Tu vois j'ai gardé le cap et tu n'y es pas pour rien ! Je remercie Léo Guignard pour son expertise dans le développement logiciel qui a permis de réaliser une partie de ce travail. Tu as su traduire mes remarques et mes désirs en langage informatique avec une aisance époustouflante entre deux cours. Je te souhaite plein de réussite. Je remercie la nouvelle venue, Cécile Quéré pour sa gentillesse, sa motivation, son aide et son expertise en microscopie qui m'a permis d'avoir de belles images lorsque le logiciel natif du microscope refusait obstinément d'obtempérer. Je te remercie de t'être occupée d'aller chercher mes manuscrits et de les poster pour moi alors que j'ai un pied plâtré (encore une épreuve). Enfin j'adresse tous mes vœux de réussite à Jonathan Millet, le dernier arrivé, que je n'ai pas eu l'occasion de connaître.

Je voudrais remercier nos collaborateurs, Hidehito Kuroyanagi et son équipe au Japon et Cameron Mackereth et Samir Amrane de l'IECB (Bordeaux) qui m'ont permis de réaliser ce travail. Je voudrais par la même occasion remercier les chercheurs et les étudiants de mon unité pour leur aide à un moment ou un autre dans mon parcours et leur gentillesse. Je remercie également Kati Ba-Pierozzi pour la gestion administrative et pour sa bonne humeur. Je remercie mon ancien directeur d'unité et actuel directeur d'institut M. Jean-Jacques Toulmé de m'avoir accueilli dans sa structure. Je remercie enfin M. Jean-Louis Mergny, actuel directeur, de m'avoir ouvert les portes de son bureau lorsque je l'ai sollicité et d'avoir tout mis en œuvre pour régler mes problèmes administratifs lorsque tous mes recours ont échoués.

A présent je voudrais remercier tous mes amis pour le soutien psychologique énorme qu'ils m'ont apporté tout au long de ces années et surtout lorsque le ciel s'est assombri. Merci à vous tous de m'avoir servi de phares et/ou d'escaliers de joie, de gaieté et de réconfort. Un mot, une ligne, une page ne sont pas suffisants pour évoquer tout ce que vous m'avez apporté. Votre présence à mes côtés a été cruciale, croyez moi. Je n'ai pas trouvé un moyen moins injuste pour vous remercier qu'en vous citant par ordre alphabétique alors je vous prie de me pardonner pour ce court inventaire : Nadia Boutayeb, Anne-Sophie Chamaret, Asen Daskalov, Aurore Delotterie, Habiba Elatmani, Camille Gaillard, Jean-Manuel Jimenez, Ariane Kinge, Camille Latapy, Amine Machat, Pauline Maquet, Thomas Moreau, Claire Morel, Chiara Pascali, Veki Petkova... et enfin mes amis restés ou repartis de l'autre côté de la méditerranée parmi lesquels mon ami d'enfance ou dois-je dire mon frère et cousin Ayoub Zniber à qui j'adresse une pensée particulière. Depuis ton retour au Maroc je n'ai malheureusement pas eu l'occasion de partager beaucoup de choses avec toi. Je te prie de me pardonner de ne pas avoir été présent à ton mariage, ce jour-là je n'ai pas pu m'empêcher de partir faire un tour auprès de ton éphémère demeure bordelaise comme pour arracher en vain au destin un moment avec toi. Je te souhaite tout le bonheur du monde dans ta vie et j'espère qu'un jour nous pourrions rattraper le temps perdu.

Je pense à présent à ma grande famille, oncles et tantes pour leur affection, cousins et cousines pour la complicité.

Je pense surtout à mon frère Moufdi et mes sœurs Rim et Dyâa. Mon dieu j'ai les larmes aux yeux en abordant cette partie. Vous êtes ce que je possède de plus précieux au monde après nos parents. Ça fait bientôt 11 ans que je suis parti, mais il n'y a pas un jour sans que vous ne soyez dans mon rêve, dans mes pensées, dans mon cœur, dans mes souvenirs ou dans mes projets. C'est la façon que j'ai trouvé de vous garder près de moi où que je sois. Je n'ai pas eu la chance voir vos familles s'agrandir, je n'ai pas eu le plaisir de voir mes neveux grandir mais à chaque fois que je vous retrouvais je sentais en vous l'amour, tout l'amour, que vous avez pour moi et que vous aviez communiqué à mes neveux. J'aimerais tant vous serrer dans mes bras

si fort et vous dire à quel point je vous aime tous, vous, vos conjoints qui sont tout autant des frères et sœurs pour moi et mes adorables nièces et neveux. Je vous dédie ce travail à vous tous.

Maman, Papa : J'ai souvent eu la chance de vous dire combien je vous aimais et je ne vais pas rater cette occasion pour le réitérer. Je vous aime. C'est grâce à vous que j'ai pu réaliser mes rêves, c'est grâce à vous que j'ai pu aller loin, c'est grâce à vous que j'ai pu prendre mon envol. Vous m'avez accompagné dans cette aventure dès le premier jour pour m'aider à m'installer. Je fêtais tout juste mes dix-huit ans et je me souviens du jour où vous repartiez et que je restais seul pour la première fois. Une sensation d'angoisse, de peur, d'intimidation avait traversé mon corps sous forme de frisson mélangé à de l'excitation, de la détermination et de l'ambition. Je savais néanmoins que je pouvais compter sur vous en tout temps, toute occasion et en toute circonstance. Je pense chaque jour à vous. Vous me manquez et vous me manquez toujours terriblement. Il n'y a pas une seule chose que je ne vous dois pas. Je me rends compte à présent de la chance que j'ai eu de vous avoir comme parents. J'ai eu grâce à vous un nid douillet, une famille épanouie et de l'amour sans limite. Je vous dédie cette thèse. Je t'aime maman, je t'aime papa.

Je m'adresse à moi-même à présent, un témoignage qui me servira à me remémorer cet instant dans le futur si je suis toujours de ce monde, parce que la mémoire est courte. Cette thèse a été conçue dans l'espoir et l'ambition mais je l'ai accouchée dans la douleur. Je devais me résoudre à voir certains de mes projets s'évanouir à jamais mais je m'accrochais à l'espoir de décrocher le doctorat comme ultime récompense. Je voulais alors que cette thèse soit la plus belle, la plus aboutie et la plus studieuse possible. Je me suis donné à la tâche sans repos, sans trêve... et à mon grand désespoir sans résultats. Un échec supplémentaire m'a plongé dans une mélancolie étouffante. Au même moment, j'ai été rattrapé par des soucis de santé et de... papiers. Je ne m'attendais pas à ce dernier cas de figure comme un refoulement de ma part que la décennie a consolidée et quand j'ai réalisé en 2010 que mon séjour risquait de ne pas être renouvelé j'ai commencé à paniquer. Alors pour prendre les devants je décide de demander la naturalisation puisque j'ai été formé sur les bancs de l'université de la république et que ma vie est ici depuis une dizaine d'années. La naturalisation m'a été refusée pour prétexte de travail en CDD autant que salarié de l'inserm. C'est là où je me rendais compte pour la première fois de la schizophrénie de l'administration. Que serais la recherche française sans les agents non titulaires qui y travaillent ? La politique publique a tout fait pour encourager le travail à durée déterminée dans ce domaine et pour ériger cela en modèle incontournable pour tout jeune chercheur en devenir. Depuis 5 ans maintenant que je fréquente les laboratoires de recherche je n'ai assisté qu'à trop peu de titularisations autour de moi. Parmi celles-ci, la titularisation bien tardive à mon goût de mon chef en 2011 qui a une décennie d'expérience et qui dirige sa propre équipe depuis quelques années. Mais pour en revenir à ma naturalisation, voilà qu'on me reproche de

travailler en CDD dans la recherche scientifique. L'administration reste sourde à mes requêtes. Mes craintes initiales se confirment rapidement lorsque vint le temps du renouvellement de mon séjour en 2011, qui coïncide avec la fin de mon contrat, et que celui-ci me fut refusé. En me refusant de prolonger le séjour, je devinais la difficulté de la situation qui se dessinait. Déjà fragilisé pour mille et une raisons avant cela voilà un coup de grâce supplémentaire qui m'est porté. J'ai été désorienté, déboussolé et démuni pendant cette période et c'est là que j'ai bien failli « jeter l'éponge » à jamais par désespoir. C'est là que les gens autour de moi ont joué un rôle crucial et m'ont encouragé à continuer jusqu'à décrocher mon diplôme et c'est grâce à eux que j'ai pu accomplir ce travail aujourd'hui. J'ai essayé de m'accrocher à ma thèse et je me suis dit que j'allais faire valoir mes droits par l'intermédiaire d'un avocat. C'était sans compter sur la lourdeur et la lenteur de l'appareil judiciaire qui font que mon dossier ne sera jugé qu'en décembre 2013 selon les prévisions les plus optimistes. Les honoraires de l'avocat étaient une source de dépense imprévue et conséquente que j'aurais voulu éviter car bientôt je fus frappé par une autre conséquence directe. Je ne pouvais pas prétendre à l'assurance chômage comme tous mes camarades. J'ai dû puiser dans mes économies pour pouvoir continuer à vivre. Je me suis adressé aux directeurs de toutes les administrations concernées sans succès. Je me suis adressé au ministre du travail, au ministre de la recherche et au ministre de l'intérieur sans réponses à ce jour. J'ai dû me résigner à chercher du travail tout en m'accrochant du mieux que je pouvais dans ces conditions à ma thèse. Les espoirs nés des quelques entretiens que j'ai réussi à décrocher au sein de sociétés de biotechnologies se sont rapidement évanouis lorsqu'il fallait aborder les questions administratives. La seule société de séquençage qui voulait me garder quand même me proposait un contrat d'intérim de 6 mois en région grenobloise sous condition de disponibilité immédiate. J'ai dû me résoudre à refuser de donner suite pour pouvoir avoir la chance de soutenir cette thèse. Entre temps, lors de mes incessants allers retours à la préfecture, les agents finissent par admettre qu'une circulaire interne stipule qu'un scientifique étranger n'a pas vocation à rester sur le territoire français à la fin de son contrat et qu'il doit repartir au sein de son université d'origine. Mes chers messieurs de la préfecture et du ministère, je n'ai jamais connu d'autre université que celle de Bordeaux... que faire ? Évidemment je n'allais pas avoir de réponse à cette question. Un an après (à présent donc), un autre effet pervers de cette situation se manifeste. Sans couverture sociale et sans mutuelle je me fracture le pied en rentrant chez moi un soir d'octobre de retour du laboratoire. Je ne pouvais m'empêcher de me plonger dans une tourmente cauchemardesque en ayant des pensées pour les séquelles que je risquais de garder en me fracturant un pied déjà poly traumatisé antérieurement, une pensée pour les conditions dans lequel je devais finir de rédiger ce manuscrit et malheureusement pour les frais médicaux que tout cela allait engendrer. L'administration qui me prive de mes droits d'un côté me rappelle à mes devoirs de l'autre et exige toute fois de ma part de régler les impôts de l'année 2011 avec pénalités de retard. Elle n'est pas ironique la vie ? En tout cas j'espère que le rétablissement sera total peu importe le

reste. J'aperçois cependant une lueur d'espoir pour 2013... espérant que cela ne soit pas qu'un mirage.

Voilà l'une des facettes du contexte dans lequel ce manuscrit a été rédigé. Voilà les épreuves que tu as du surmonter pour pouvoir achever le travail que tu as directement entamé en 2008 et indirectement en 2001. Ta thèse n'est certes pas aussi brillante que tu ne l'avais souhaité, et elle ne traduit pas tout l'effort et le cœur que tu as mis dedans, mais peut être la vie te réserve d'autres réconforts car ce ne serait que justice. N'oublies jamais que sans le soutien des quelques personnes qui t'entourent et pour qui tu comptes, même la soutenance de thèse aurait été impossible. Tu as bien souvent été orgueilleux dans ton passé ! À quoi bon ? Que ces années d'épreuves t'apprennent l'humilité. Si dans le futur tout cela ne devient qu'un lointain mauvais souvenir garde une pensée pour tous ceux qui ont fait que l'impossible soit devenu possible... Nous sommes finalement peu de choses sans les autres !

Pardonnez-moi papa, maman, de vous avoir caché cela. Je ne jugeais pas nécessaire de vous informer de cette situation qui n'aurait eu d'autre conséquence que de vous plonger dans une angoisse malade. J'étais bien tenté de le cacher pour toujours et de garder le silence comme dans le poème que j'ai mis en préface. Je me suis longtemps inspiré de ce loup dans ma vie mais il y a des injustices dont le silence n'est que complice.



L'un des tableaux qui me fascinent et me captivent :

Het Meisje met de Parel  
(La jeune fille à la perle)  
**Johannes Vermeer**

## **Liste des Abréviations**

<b>3'UTR</b> Région 3' Non traduite	<b>GFP</b> <u>G</u> reen <u>f</u> luorescent <u>p</u> rotein
<b>5'UTR</b> Région 5' Non traduite	<b>ITC</b> <u>I</u> sothermal <u>t</u> itration <u>c</u> alorimetry
<b>ADN</b> <u>A</u> cide <u>d</u> ésoxyribo <u>n</u> ucléique	<b>Kb</b> Kilobases
<b>ADNc</b> ADN complémentaire	<b>KDa</b> Kilo Dalton
<b>ARN</b> <u>A</u> cide <u>r</u> ibo <u>n</u> ucléique	<b>L1</b> Stade larvaire 1
<b>ARNdb</b> ARN double brin	<b>L2</b> Stade larvaire 2
<b>ARNi</b> (ou RNAi) ARN interférence	<b>L3</b> Stade larvaire 3
<b>ARNm</b> ARN messenger	<b>L4</b> Stade larvaire 4
<b>ARN pol II</b> ARN polymérase II	<b>miARN</b> micro ARN
<b>ARNr</b> ARN ribosomal	<b>NMD</b> <u>N</u> on-sens <u>m</u> ediated <u>d</u> ecay
<b>ATP</b> <u>A</u> denosine <u>t</u> riphosphate	<b>NGM</b> <u>N</u> ematode <u>g</u> rowth <u>m</u> edia
<b>BMW</b> <u>B</u> ody <u>w</u> all <u>m</u> uscle	<b>Pb</b> Paires de bases
<b>COPAS</b> <u>C</u> omplex <u>o</u> bject <u>P</u> arametric <u>a</u> nalysis and <u>s</u> orting	<b>PCR</b> <u>P</u> olymerase <u>c</u> hain <u>r</u> eaction
<b>CSD</b> <u>C</u> old <u>s</u> hock <u>d</u> omain	<b>PMT</b> photomultiplicateur
<b>DMD</b> <u>D</u> uchenne <u>m</u> uscular <u>d</u> ystrophy	<b>pre-ARNm</b> ARN pré-messager
<b>dsRNA</b> double stranded RNA (ARN double brin)	<b>pre-miARN</b> pré-microARN
<b>EMS</b> <u>E</u> thyl <u>m</u> ethane <u>s</u> ulfonate	<b>pri-miRNA</b> pri-microARN
<b>ENU</b> N-ethyl N-nitroso urée	<b>RFP</b> <u>R</u> ed <u>f</u> luorescent <u>p</u> rotein
<b>EST</b> <u>E</u> xpressed <u>s</u> equence <u>t</u> ags	<b>RISC</b> <u>R</u> NA <u>i</u> nduced <u>s</u> ilencing <u>c</u> omplex
<b>Exp5</b> Exportin 5	<b>RNA-seq</b> séquençage ARN à haut débit
	<b>RRPT</b> <u>R</u> apporteur de la <u>r</u> égulation <u>p</u> ost-transcriptionnelle

**RT-PCR** Reverse transcriptase polymerase chain reaction

**SnRNP** Small nuclear ribonucleoprotein

**SNP** Single nucleotide polymorphism

**stARN** Small temporal RNA

# Développement d'outils pour l'étude *in vivo* de la régulation post- transcriptionnelle chez *Caenorhabditis* *elegans*

---

## TABLE DES MATIERES

<b>I.</b>	<b>Introduction</b>	<b>14</b>
1.1	La régulation post-transcriptionnelle	15
1.2	L'épissage des ARN	17
1.2.1	Découverte	17
1.2.2	Généralités	18
1.2.3	Mécanisme d'épissage	19
1.3	Epissage alternatif	22
1.3.1	Généralités	22
1.3.2	Les différents types d'épissage	23
1.3.3	Epissage alternatif et régulation post-transcriptionnelle	24
1.4	Les microARN	25
1.4.1	Historique	25
1.4.2	Localisation et transcription des miARN	28
1.4.3	Biogénèse des miARN	29
1.4.3.1	Le clivage nucléaire par Drosha	29
1.4.3.2	Export des pré-miARN	30
1.4.3.3	Le clivage par Dicer	30
1.4.3.4	Le complexe RISC et le mécanisme d'action des miARN	31
1.4.4	Les cibles des miARN	32
1.5	L'ARN interférence	33
1.6	<i>Caenorhabditis elegans</i> comme modèle d'étude	35
1.6.1	Historique	35
1.6.2	Caractéristiques	36
1.6.2.1	Anatomie	36
1.6.2.2	Cycle de vie	37
1.6.2.3	Développement post-embryonnaire	38
1.6.2.4	La présence de mâles	38
1.6.3	Génétique de <i>C. elegans</i>	39
1.6.3.1	Les opérons	39
1.6.3.2	L'épissage alternatif	40
1.6.4	Visualisation de la régulation post- transcriptionnelle <i>in vivo</i>	41
1.6.4.1	Etude de l'épissage alternatif par des rapporteurs bicolores	41
1.6.4.1.1	Epissage du gène <i>let-2</i>	41
1.6.4.1.2	Epissage du gène <i>egl-15</i>	45
1.6.4.2	Approche fluorescente pour valider les interactions miARN/UTR	49
<b>II.</b>	<b>Première partie : étude de l'épissage du gène <i>let-2</i></b>	<b>53</b>

<b>Objectif :</b>	<b>54</b>
<b>2.1 Approche par génétique classique</b>	<b>54</b>
2.1.1 Analyse phénotypique par COPAS profiler	54
2.1.2 Automatisation du crible par COPAS biosorter	57
2.1.3 <i>asd-2</i> et phénotypes mutants	63
2.1.4 Les mutants sauvages pour le gène <i>asd-2</i>	64
2.1.5 Séquençage du génome entier	66
2.1.5.1 Les mutations communes	68
2.1.5.2 Les mutations spécifiques dans les régions codantes	69
2.1.6 Discussion	69
<b>2.2 Approche par génétique inverse</b>	<b>71</b>
2.2.1 Modification du COPAS pour l'obtention d'une plateforme adaptée à un crible ARNi à grande échelle	71
2.2.1.1 Le module Reflx natif du COPAS	71
2.2.1.2 Amélioration des performances du module Reflx	72
2.2.1.3 Traitement logiciel des données collectées	74
2.2.2 Criblage systématique des facteurs de l'épissage de <i>let-2</i> par ARNi	77
<b>III. Deuxième partie : étude in vivo de l'impact de la modification de la région cis régulatrice du gène <i>egl-15</i></b>	<b>79</b>
3.1 Mesure d'affinité SUP-12/ <i>egl-15</i> par ITC	80
3.2 Corrélation des résultats <i>in vivo</i> et <i>in vitro</i>	83
<b>IV. Troisième partie : génération de système rapporteur in vivo de régulation post-transcriptionnelle par miARN</b>	<b>85</b>
4.1 Stratégie RRPT	86
4.1.1 Le Recombineering	88
4.1.2 Génération de lignées transgéniques exprimant la cassette RRPT	91
4.1.3 Conclusion	96
<b>V. Discussions et perspectives</b>	<b>97</b>
5.1 Epissage alternatif de <i>let-2</i>	98
5.2 Modification du module Reflx du COPAS	100
5.3 Etude de l'élément cis régulateur de l'épissage du gène <i>egl-15</i>	101
5.4 Génération de système rapporteur de la régulation post-transcriptionnelle <i>in vivo</i> chez <i>C. elegans</i>	102
5.4.1 Clonage par assemblage de Gibson	102
5.4.2 Identifier les raisons de l'expression inattendue	104
5.4.2.1 Vérification d'absence d'activité promotrice de l'opéron linker <i>gpd2/gpd-3</i>	105
5.4.2.2 Vérification de la stabilité des rapporteurs	106
5.4.2.3 Régulation post-transcriptionnelle du gène <i>unc-54</i>	106
<b>VI. Matériels et méthodes</b>	<b>107</b>
6.1 Protocoles de <i>C. elegans</i>	108
6.1.1 Culture et maintenance	108
6.1.2 Synchronisation des vers	108
6.1.3 Génération de mâles	108
6.1.4 Croisement	109
6.1.5 Mutagénèse par EMS	109
6.1.6 Injection des vers	110
6.1.7 Microscopie	110

<b>6.2</b>	<b>Protocoles de biologie moléculaire</b>	<b>110</b>
6.2.1	PCR	110
6.2.2	Préparation d'ADN génomique	110
6.2.3	Production de bactéries électrocompétentes	111
6.2.4	Electroporation d'ADN	111
6.2.5	Construction de cassette RRPT	111
6.2.6	Plasmides rapporteurs de <i>egl-15</i>	113
6.2.7	Sélection des fosmides	113
6.2.8	Extraction des fosmides	114
6.2.9	Recombineering	115
6.2.9.1	Recombineering Dolphin & Hope	115
6.2.9.2	Recombineering Tursun et al.	115
<b>6.3</b>	<b>COPAS Biosorter</b>	<b>117</b>
6.3.1	Analyse des vers au COPAS biosorter	117
<b>VII.</b>	<b>Annexe</b>	<b>119</b>
<b>VIII.</b>	<b>Bibliographie</b>	<b>126</b>

# **I. INTRODUCTION**

## 1.1 LA REGULATION POST-TRANSCRIPTIONNELLE

La régulation de l'expression des gènes est fondamentale pour coordonner la synthèse, l'assemblage et la localisation des complexes macromoléculaires dans les cellules. Cette expression est régulée à divers niveaux [1, 2] (Figure 1).

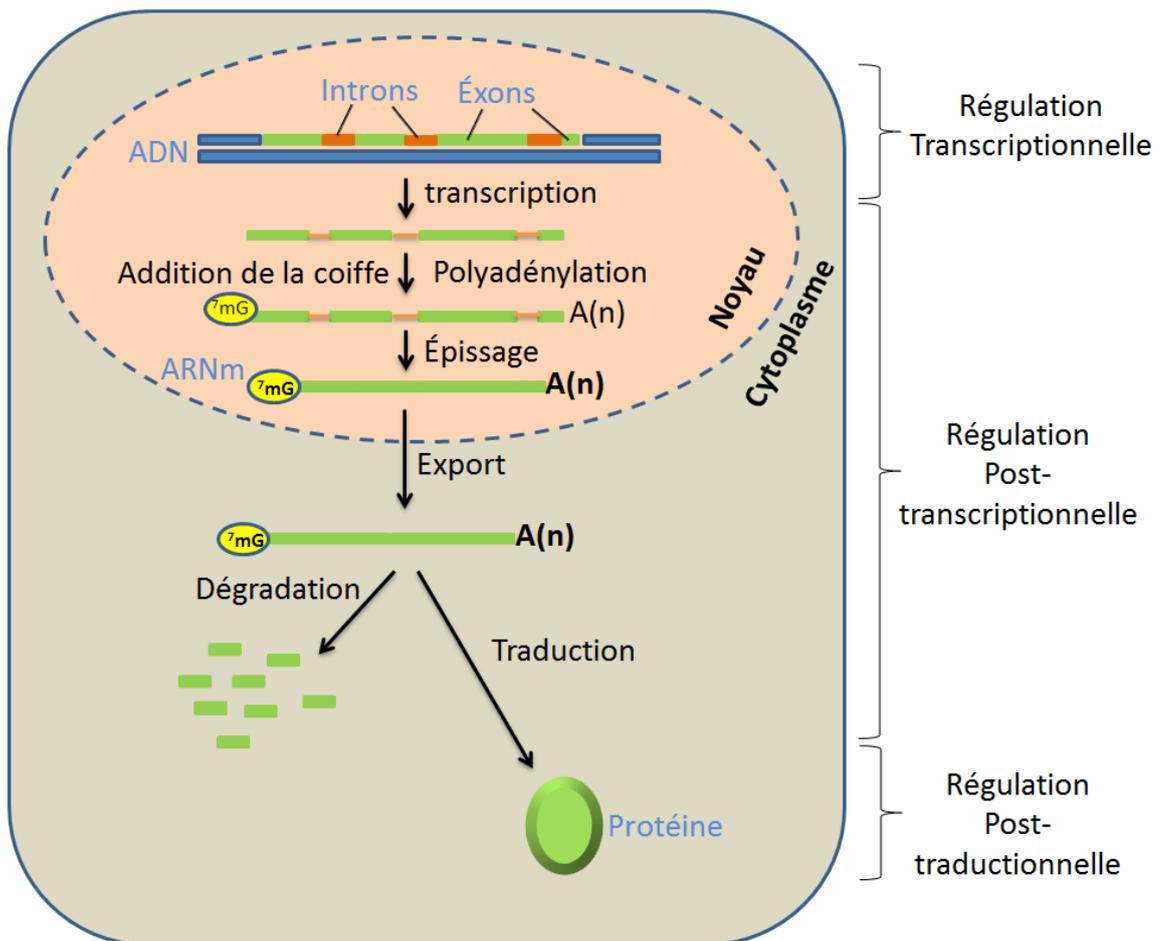


Figure 1 : Les différentes étapes de l'expression génique

La régulation commence dans le noyau où les facteurs de transcription se lient à des séquences spécifiques d'ADN proximales ou distales d'un gène et recrutent les ARN polymérases pour la synthèse des ARN.

Les protéines de liaison à l'ARN s'associent avec l'ARN en cours de synthèse et opèrent diverses modifications comme l'ajout d'une coiffe en 5', l'épissage, l'édition et la poly-adénylation en 3' [1, 3, 4]. Les transcrits sont alors exportés à travers les pores nucléaires vers le cytoplasme où ils vont

être adressés et stockés dans des régions subcellulaires grâce à des signaux de localisation et des protéines de liaison à l'ARN [5]. Les ARNm s'assemblent avec des facteurs de traduction et les ribosomes pour initier la synthèse protéique de manière contrôlée [6]. Enfin, les ARNm sont dégradés par les exonucléases grâce à différentes voies de dégradation [7]. Le sort et la localisation des protéines peuvent à leur tour être contrôlés par la modification spécifique d'acides aminés, le clivage à des sites spécifique par des protéases et enfin la dégradation par le protéasome [8].

Le développement depuis une dizaine d'années d'outils d'analyse à l'échelle génomique comme les puces à ADN ont permis une meilleure compréhension globale des programmes de régulation des gènes. Par exemple, les puces à ADN ont été largement utilisées pour étudier les programmes transcriptionnels en comparant le niveau d'expression moyen des ARN dans divers stades et dans divers types cellulaires [9] ou en cartographiant les sites de liaison des protéine de liaison à l'ADN par immunoprécipitation (ChIP-ChIP) [10]. L'intégration de ces données a permis une description de la régulation transcriptionnelle impliquant un grand nombre de gènes qui contrôlent les réponses physiologiques et développementales [11-13].

Contrairement à la régulation transcriptionnelle, l'architecture globale des systèmes qui régulent les étapes post-transcriptionnelles d'expression des gènes est encore peu connue. Un tel système de régulation post-transcriptionnelle doit être contrôlé par des centaines de protéines de liaison à l'ARN et de microARN (miARN) encodés dans les génomes eucaryotes [14-16]. C'est pourquoi il est important de disposer d'outils et de plateforme adaptés à l'étude de cette régulation à l'échelle génomique.

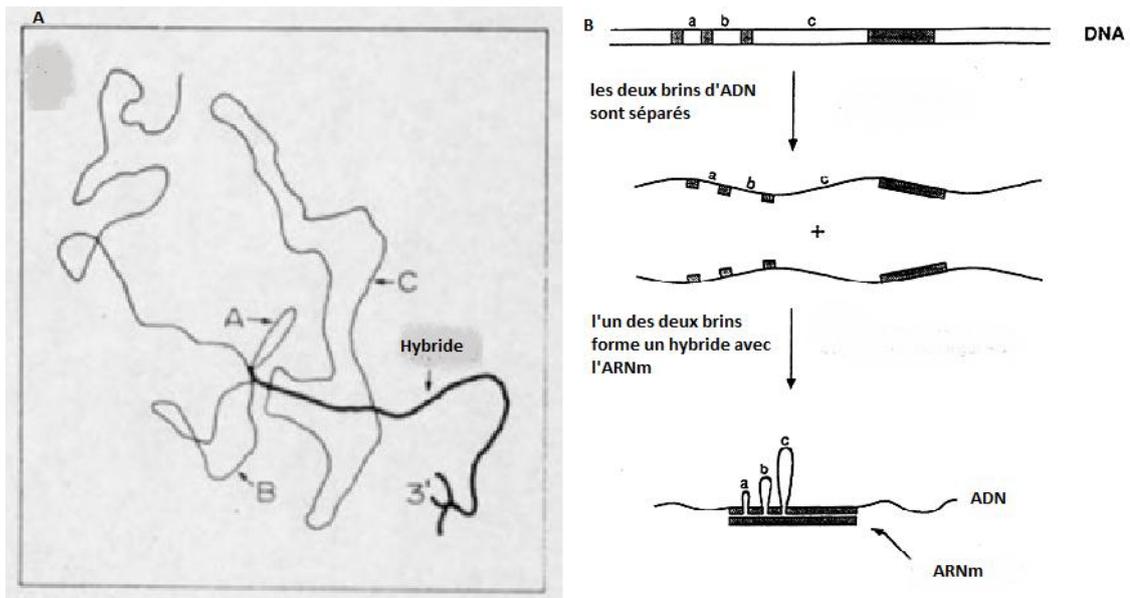
Dans cette thèse, nous allons nous intéresser à deux programmes de la régulation post-transcriptionnelle chez *Caenorhabditis elegans* : l'épissage alternatif et la régulation par les miARN.

## **1.2 L'ÉPISSAGE DES ARN**

### **1.2.1 Découverte**

Au début de la biologie moléculaire, Francis Crick avait proposé qu'il existe une colinéarité entre la séquence d'un gène et le produit ARN de celui-ci. Ceci est vrai chez les procaryotes dont le génome est très compact. En revanche, l'un des premiers indices qui a conduit à infirmer l'universalité de cette hypothèse est venu de la découverte du phénomène d'épissage dans les années 1970. En utilisant des adénovirus comme système expérimental, le chimiste anglais Richard Roberts et le biologiste moléculaire Philip Sharp ont chacun découvert que les gènes peuvent être divisés en plusieurs segments le long du génome [17, 18]. Les deux scientifiques ont observés en microscopie électronique par une technique nommée R-Loop que les molécules d'ARN messagers (ARNm) du gène hexon de l'adénovirus 2 (Ad2) ne s'hybrident pas à un seul fragment d'ADN, mais à quatre segments d'ADN discontinus (Figure 2).

On sait à présent que transcription d'un gène aboutit à la synthèse d'un ARN pré-messager qui est fragmenté puis assemblé au cours de la maturation pour donner l'ARN messager qui permettra la synthèse protéique en aval. Les segments d'ARN qui sont impliqués dans la synthèse protéique sont appelés exons. Les segments d'ARN non codants qui s'intercalent entre deux exons se nomment introns. Durant le phénomène d'épissage qui se produit après la transcription, les introns sont excisés et les exons sont ligués ensemble ou « épissés ». Aujourd'hui, on sait que la discontinuité des gènes et l'épissage ne sont pas une exception mais plutôt la norme chez la plupart des eucaryotes [19].



**Figure 2 : Représentation schématique de l'hybridation entre l'ARNm et l'ADN du gène hexon de l'adénovirus 2**

Hybridation de l'ARNm hexon de l'adénovirus 2 (Ad2) avec le génome de celui-ci en microscopie électronique par technique de R-Loop. Les régions hybridées sont représentées en lignes foncées, et les segments non hybridés sont représentés en lignes fines. A, B et C représentent les 3 introns contenus dans ce gène (A). Représentation schématique de la séparation des deux brins de l'ADN du gène hexon viral et l'hybridation avec son ARNm correspondant, les introns forment des boucles non hybridées a, b et c (B). [Adapté de Berget et al. 1977 [17]]

## 1.2.2 Généralités

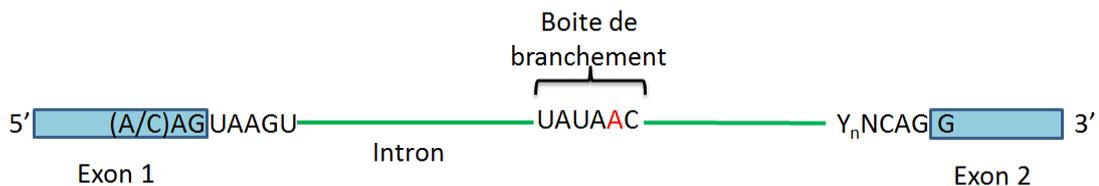
Le nombre et la taille des introns varient de gène en gène et d'espèce en espèce. La taille moyenne des introns chez les mammifères est d'environ 2 kilopaires de bases (Kb) alors que la taille des exons est de 200 paires de bases (pb) [20]. Le gène de la dystrophie musculaire humaine Duchenne (Duchenne Muscular Dystrophy ; DMD) constitue un exemple extrême. Ce gène contient 79 exons et 78 introns qui s'étalent sur 2.5 millions de pb. Quand il est épissé, les 79 exons produisent un ARNm de 14kb qui code pour la Dystrophine [21-23]. La Dystrophine est une protéine du cytosquelette qui participe à l'ancrage de l'actine à la matrice extracellulaire. Un gène d'une telle taille est plus vulnérable aux mutations. La maladie DMD est due dans 65% des cas à des délétions, dans 15 % des cas à des duplications et dans 20% des cas à des mutations ponctuelles [24]. Dans tous les cas, la maladie survient quand la séquence codante du gène est altérée de façon à donner une protéine déficiente. La découverte et l'étude de ce gène est liée à la maladie lourdement handicapante qu'il engendre en cas d'anomalie. Comme la séquence codante de ce gène est trop grande pour espérer pouvoir introduire artificiellement une version fonctionnelle du gène complet, la restauration d'une protéine fonctionnelle en éliminant l'exon muté constitue la piste la plus prometteuse pour le traitement de la maladie. Dans ce but, les études récentes s'intéressent à une manière artificielle de forcer la machinerie à sauter des exons portant la mutation par

l'administration d'oligonucléotides anti-sens, mais l'application de cette technique chez l'humain n'est pas encore réalisable [25]. Cette méthode exploite artificiellement le phénomène d'épissage alternatif présent naturellement chez les eucaryotes. Un chapitre sera spécifiquement consacré à l'épissage alternatif. L'exploitation du saut d'exons à des fins thérapeutiques démontre que la compréhension du mécanisme d'action de l'épissage peut constituer un allié sérieux dans la lutte contre les maladies.

### 1.2.3 Mécanisme d'épissage

L'épissage des pré-ARNm est catalysé à l'intérieur du noyau par de petits ribonucléoprotéides nucléaires appelés snRNP [26, 27]. Quatre snRNP sont impliqués dans l'épissage : U1, U2, U5 et le complexe U4/U6, ils constituent le corps de la machinerie de l'épissage, le spliceosome.

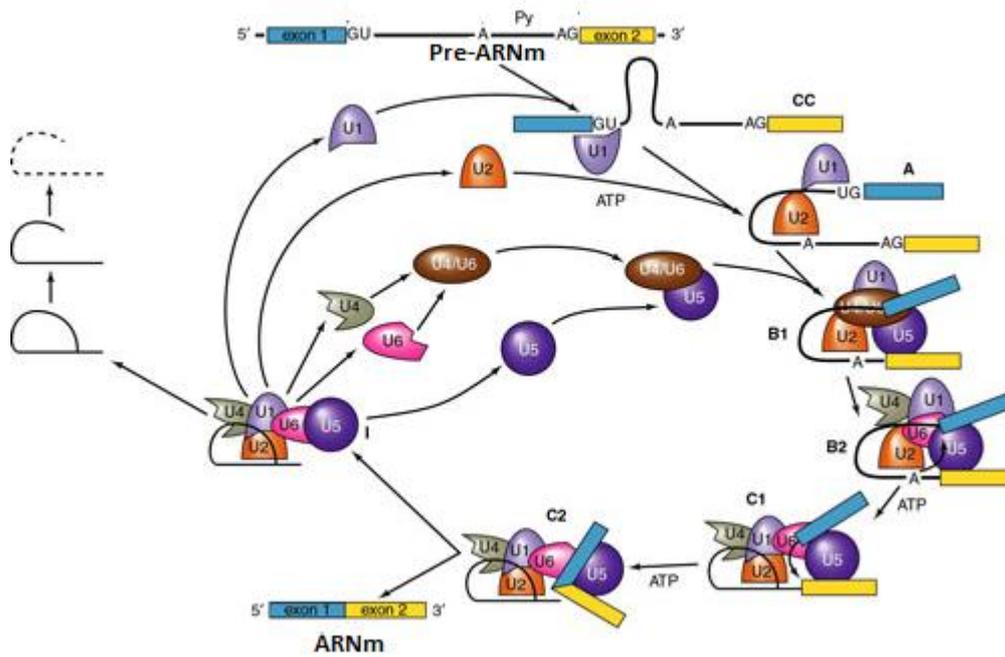
Les jonctions amont et aval de chaque intron se caractérisent par des séquences de nucléotides conservées, qui sont reconnues par la machinerie d'épissage. Il existe également une séquence conservée à l'intérieur de l'intron, appelée "boite de branchement" qui contient une adénosine strictement conservée, essentielle au mécanisme (Figure 3).



**Figure 3 : les séquences introniques et exoniques reconnues par la machinerie de l'épissage**

Les exons sont représentés par les rectangles bleus et l'intron par un trait vert. (A/C) signifie A ou C ; Y<sub>n</sub> Pyrimidine n fois ; N n'importe quel nucléotide. Le A en rouge représente le nucléotide conservé de la boite de branchement.

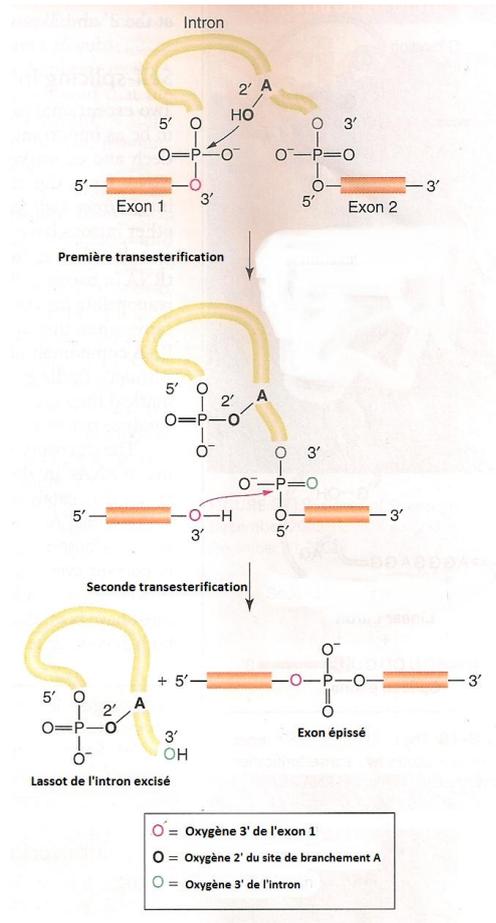
L'ARN contenu dans le complexe U1 interagit avec le site d'épissage en 5' par appariement de type Watson et Crick [28] qui est stabilisé chez les eucaryotes supérieurs par des familles de protéines riches en sérine-arginine et par les protéines du complexe U1. L'ARN du complexe U2 interagit de manière similaire avec le site de branchement situé dans l'intron [29]. Cet appariement est stabilisé par le complexe protéique hétéromérique du complexe U2 [30]. Le snRNP U5 agit au niveau du site donneur en 3' dans une interaction avec le complexe U4/U6 qui permet d'induire la première réaction d'estérification de l'épissage [31] (Figure 4). Les ARN des complexes U sont ubiquitaires et exprimés abondamment [32]. La majorité des organismes étudiés ont plusieurs copies de gènes de snRNA [33].



**Figure 4 : Mécanisme d'action du spliceosome**

[www.uic.edu]

La première étape de l'épissage est une attaque de la liaison phosphodiester à la jonction entre le côté 5' de l'intron et l'exon amont. C'est le 2'-OH du ribose de l'adénosine conservée dans la boîte de branchement qui joue le rôle de nucléophile dans cette réaction qui conduit à la formation d'un intermédiaire cyclique appelé "lasso" par transestérification. Dans cet intermédiaire il y a formation d'une liaison 2',5' phosphodiester entre la guanosine conservée à l'extrémité de l'intron et l'adénosine. Cette dernière est ainsi impliquée dans trois liaisons phosphodiester, en 5' et 3' avec les nucléotides précédents et suivants du pré-ARNm, et en 2' avec l'extrémité 5' de l'intron.



**Figure 5 : Les deux étapes de transesterification lors de l'épissage**

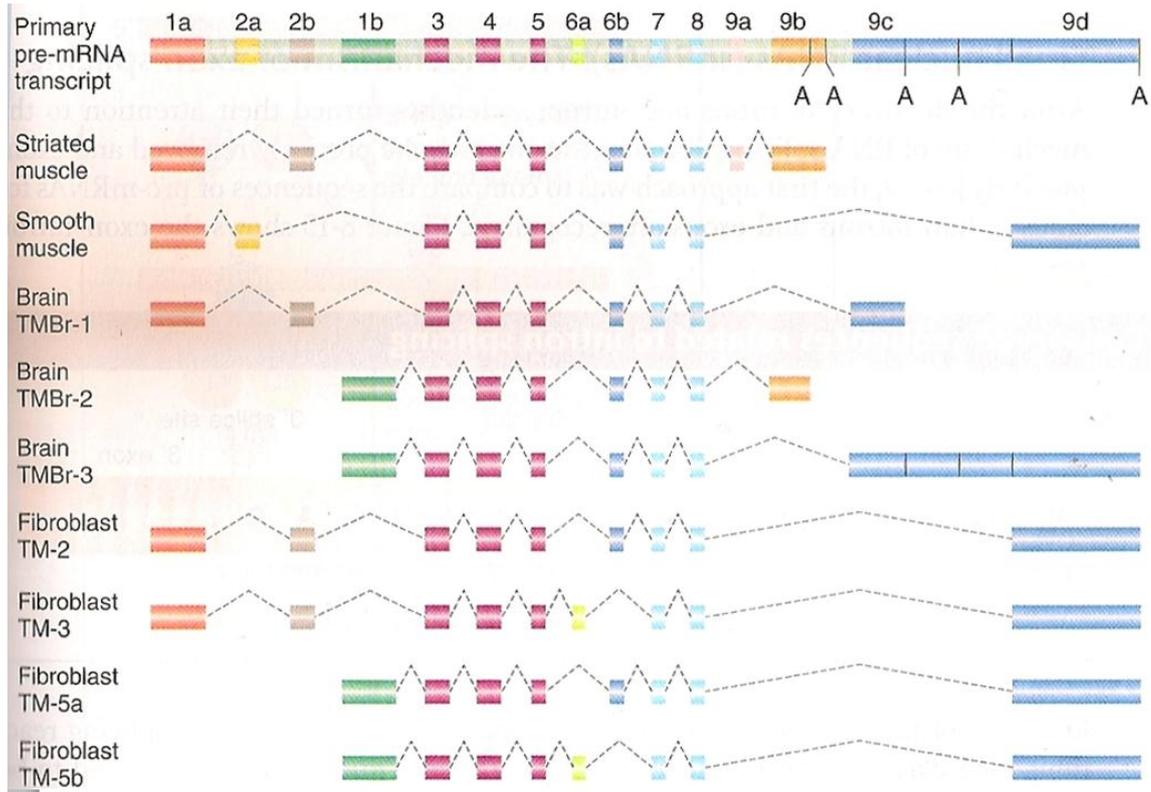
[Genes IX, B. Lewin, ed 2007]

Ce premier transfert libère une extrémité 3'-OH libre sur l'exon en amont. Celle-ci attaque alors à son tour la jonction aval de l'intron par un mécanisme analogue de transestérification ce qui aboutit à la suture des deux exons. L'intron sous forme de lasso est libéré et sera ultérieurement dégradé (Figure 5). A l'issue de ce processus, les jonctions sont faites entre chacun des exons.

## 1.3 EPISSAGE ALTERNATIF

### 1.3.1 Généralités

Chez les organismes eucaryotes multicellulaires, de nombreux gènes sont soumis à l'épissage alternatif. Les introns sont éliminés et les exons sont épissés dans différentes combinaisons pour aboutir à plusieurs ARNm distincts à partir d'un gène unique (Figure 6).



**Figure 6 : l'épissage alternatif de l' $\alpha$ -tropomyosine dans différents tissus**

Les rectangles verts clairs représentent les introns, les autres couleurs représentent les exons. Les signaux de polyadénylation sont représentés par un A. les traits discontinus dans les ARNm représentent les exons qui ont été exclus de l'épissage dans chaque tissu. [Extrait de Lees et al. 1990 [34], Genes IX, B. Lewin, ed 2007]

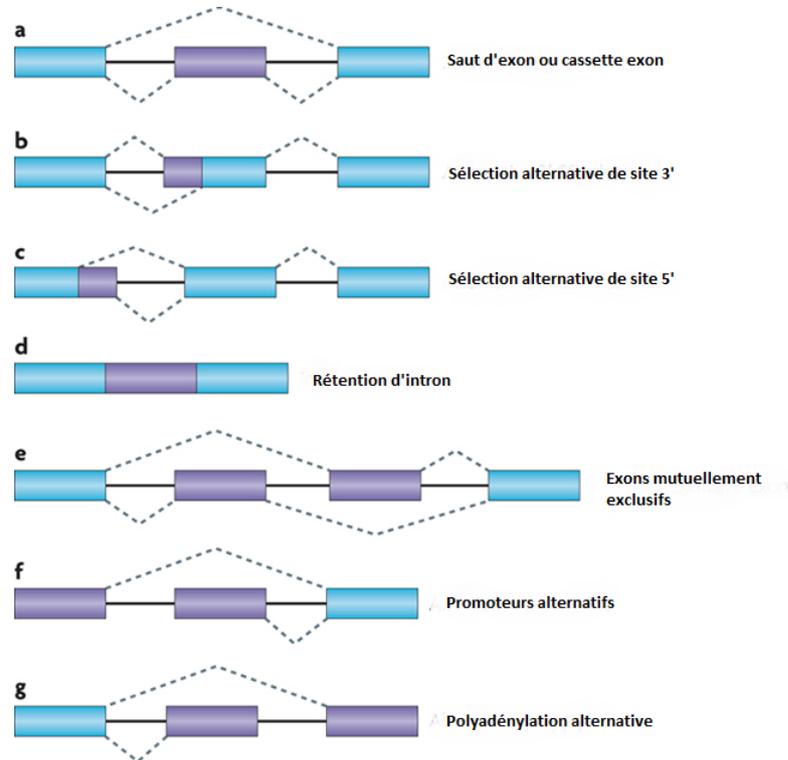
Chez l'humain, des études récentes suggèrent qu'environ 95% des gènes sont soumis à au moins un épissage alternatif [35, 36]. L'épissage alternatif peut être hautement régulé au cours du développement ou de manière tissu spécifique [37-39]. La régulation de l'épissage alternatif contrôle l'endroit où survient l'assemblage d'un spliceosome fonctionnel. Les exons impliqués dans un tel épissage possèdent généralement des séquences consensus faibles aux extrémités 5' et 3' des introns [40]. Les expériences de génétique, de génomiques et de biochimie ont mené à la compréhension de l'importance des éléments en *cis* et de facteurs en *trans* pour le contrôle de la régulation de l'épissage alternatif [37, 41, 42]. La caractérisation des sites d'épissage en 3' et en 5' est le fruit des alignements de séquences des ADN complémentaires (ADNc) et d'EST (expressed sequence tag) ce qui a permis la

prédiction des limites introns/exons montrant l'importance d'une approche computationnelle [43]. Le recours à la génomique comparative et au séquençage des lassos introniques d'ARN ont permis l'amélioration des prédictions pour l'identification du point de branchement [44, 45]. Les séquences des sites d'épissage constitutif ont ainsi pu être décryptées grâce à la caractérisation des éléments *cis* régulateurs. L'inclusion ou l'exclusion alternative des exons semble pilotée par de nombreux éléments régulateurs d'épissage qui fonctionnent comme des répresseurs ou favorisateurs [46]. La majorité des éléments *cis* régulateurs de l'épissage ont été identifiés par des mutagénèses dirigées sur des gènes alternativement épissés, par l'analyse des mutations portées par des gènes impliqués dans des maladies étudiées ou par l'analyse bio-informatiques des séquences disponibles [47-49]. Les éléments *cis* régulateurs favorisant l'épissage semblent importants pour l'épissage constitutif tandis que les éléments *cis* inhibiteurs sont majoritairement responsables de la régulation de l'épissage alternatif [50]. Ces éléments *cis* régulateurs peuvent être présents dans les introns ou dans les exons et jouent un rôle important dans le recrutement de facteurs protéiques en *trans* [51]. Les éléments activateurs de l'épissage en *trans* activent ou suppriment la reconnaissance des sites d'épissage par la machinerie du spliceosome. Pour une meilleure compréhension de la régulation de l'épissage alternatif il est important de pouvoir identifier dans les gènes épissés les éléments *cis* régulateurs impliqués et les facteurs en *trans* qui s'y attachent.

### **1.3.2 Les différents types d'épissage**

Il existe plusieurs types d'épissage alternatif qui peuvent être classifiés en quatre sous-groupes majeurs. Le premier type est le saut d'exon où la cassette exon est soit intégrée soit exclue de l'ARNm (Figure 7 a). Le saut d'exon constitue environ 40% des événements d'épissage alternatif chez les eucaryotes supérieurs mais reste extrêmement rare chez les eucaryotes inférieurs [52, 53]. Le second et le troisième type sont la sélection de sites alternatifs en 3' et 5' (Figure 7 b et c). La sélection d'épissage alternatif en 3' et en 5' comptent respectivement pour 18,4% et 7,9% des événements d'épissage alternatif chez les eucaryotes supérieurs [54]. Le quatrième type est la rétention d'introns (Figure 7 d). La rétention d'introns permet comme son nom l'indique d'inclure un intron dans l'ARNm mature et sa séquence primaire est donc traduite. C'est le type d'épissage alternatif le plus rare chez les vertébrés et les invertébrés comptant pour moins de 5% des événements [52, 53, 55, 56] contrairement aux plantes, champignons et protozoaires où la rétention d'intron est le type le plus prévalent [55]. Parmi les événements d'épissage alternatif plus rares, on retrouve des événements plus complexes qui aboutissent à des transcrits alternatifs dus à des exons mutuellement exclusifs (Figure 7 e), l'utilisation de promoteurs alternatifs (Figure 7 f) et enfin des sites de polyadénylation alternatifs (Figure 7 g) [55, 57, 58]. La polyadénylation alternative permet de contrôler à quel niveau se produit le clivage et la polyadénylation en 3' [59]. La sélection de promoteurs alternatifs aboutit à l'inclusion d'un premier exon différent. C'est le cas du gène *asd-2*

impliqué dans l'épissage alternatif chez *C. elegans* qui sera détaillé plus loin [60]. Enfin, les exons mutuellement exclusifs ne coexistent jamais dans l'ARNm. Uniquement l'un des deux exons est inclus au cours de la maturation de l'ARN. C'est le cas des gènes *let-2* et *egl-15* dont l'épissage alternatif fera l'objet d'une attention particulière dans cette thèse [61, 62].



**Figure 7 : Les différents types de l'épissage alternatif**

Les introns sont représentés par un trait fin tandis que les exons sont représentés en rectangles. Les exons constitutifs sont en bleu alors que les exons alternatifs sont en violet. Les traits discontinus représentent l'épissage retenu.

### 1.3.3 Epissage alternatif et régulation post-transcriptionnelle

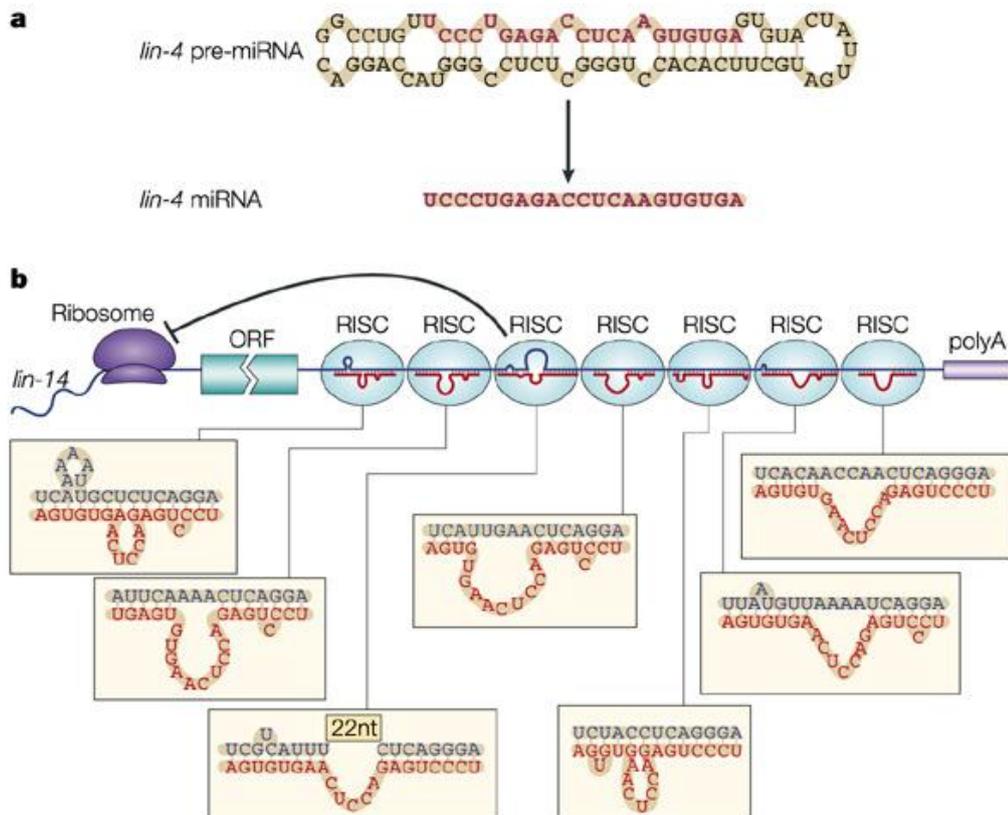
L'épissage alternatif peut également jouer un rôle dans la stabilité des ARN en produisant différents isoformes qui ont différents codons stop. Un isoforme avec un codon stop prématuré est un candidat à la dégradation médiée par le non-sens (non-sens mediated decay ; NMD). La coordination entre l'épissage alternatif et le NMD a été mise en évidence chez *C. elegans* et permet une dégradation rapide du messager contenant un stop prématuré [63]. Une analyse plus détaillée des données de l'épissage alternatif chez les vertébrés montre qu'un tiers des événements d'épissage alternatif aboutit à la production d'isoformes qui sont des substrats pour le NMD [64]. L'épissage alternatif peut donc être également considéré comme un mode de contrôle du niveau d'expression d'un gène.

## 1.4 LES MICROARN

### 1.4.1 Historique

Les microARN (miARN) sont des molécules d'ARN de 20-22 nucléotides qui jouent un rôle de régulation chez les organismes eucaryotes et procaryotes.

*lin-4* et *let-7* ont été les premiers miARN identifiés chez *C. elegans* au cours d'études génétiques d'un défaut de développement larvaire. Les vers présentant des mutations dans ces locus présentent des anomalies de développement au cours de leur transition entre les quatre stades larvaires caractéristiques de *C. elegans*. Le clonage de ces gènes a révélé qu'ils ne codaient pas pour la production de protéines mais que leur produit final était un petit ARN dérivé d'un précurseur ayant une structure en épingle à cheveux (Figure 8 a). Fonctionnellement, *lin-4* découvert en 1993 par Lee *et al.* et *let-7* découvert par Reinhart *et al.* en 2000 sont tous deux des répresseurs post-transcriptionnels d'ARN messager (ARNm). Les produits de ces ARNm jouent un rôle dans la chronologie du mécanisme de développement chez le nématode. [65-68].



**Figure 8 : le miARN *lin-4* et son appariement avec sa cible *lin-14***

Le précurseur en forme de tige boucle et la séquence du miARN *lin-4* mature (a). Complémentarités imparfaites de séquences entre le miARN *lin-4* en rouge et plusieurs régions du 3' UTR du messager de *lin-14* en bleu. RISC : RNA induced silencing complex (b). [He et al. 2004 [69]]

Le gène *lin-4* code un ARN dont la séquence est en partie complémentaire à des sites localisés dans la région 3' non traduite (3' UTR) des ARNm des gènes *lin-14* et *lin-28* [65] (Figure 8 b). La synthèse des protéines LIN-14 et LIN-28 est normalement réprimée par le miARN *lin-4* dans les stades précoces de développement [70]. LIN-14 est une protéine nucléaire nécessaire au bon déroulement des divisions cellulaires spécifiques du premier stade larvaire (L1) de *C. elegans*, la baisse de sa concentration à la fin du stade L1 permet la progression vers le second stade L2 [65, 71, 72]. Il s'est avéré que la complémentarité imparfaite des séquences entre *lin-4* et la région 3'UTR de *lin-14* est essentielle pour maintenir la capacité du gène *lin-4* à contrôler l'expression de la protéine LIN-14 en inhibant la traduction de son ARNm [66, 73, 74]. LIN-28, régulée par *lin-4*, est une protéine comportant des domaines de choc thermique (Cold Shock Domaine, CSD) qui est nécessaire à la mise en place des divisions cellulaires spécifiques à la transition du stade L2 au stade L3 [67]. Il faut attendre l'année 2000 pour la découverte du deuxième gène de miARN : *let-7*. Lui aussi produit un petit ARN de 21 nucléotides impliqué dans la mise en place de la transition du quatrième stade larvaire (L4) au stade adulte [68, 75-77].

La découverte d'uniquement deux miARN en 7 ans de recherches, combinée au fait qu'ils soient tous deux dans un même organisme et impliqués dans le même phénomène biologique aurait pu conduire à penser qu'il s'agissait d'une spécificité de *C. elegans*.

Cependant, une étude de Pasquinelli *et al.* en 2000 a démontré que l'expression d'ARN de *let-7* peut être détectée dans un large éventail d'espèces animales, des vertébrés aux mollusques, en passant par les annélides et les arthropodes... Par la suite trois autres membres de la famille de *let-7* ont été identifiés chez *C. elegans*, une quinzaine chez les humains et un seul chez la drosophile [78-80]. Ces découvertes indiquent l'universalité de leur implication dans la régulation de l'expression des gènes chez les organismes métazoaires, en particulier au cours du développement.

Rapidement, plus d'une centaine de nouveaux gènes exprimant des petits ARN non codants ont été découverts chez l'humain, la souris, la drosophile, le poisson zèbre et le nématode [81-84]. La majorité de ces miARN sont exprimés dans des types cellulaires spécifiques. Le produit d'expression de tous ces gènes font environ 22 nucléotides de long et sont maturés à partir de l'un des deux brins d'un précurseur en épingle à cheveux (Figure 9). Désormais le terme miARN est officiellement utilisé pour désigner les petits ARN de ce type en remplacement du terme stARN (small temporal RNA) qui soulignait la fonction temporelle des deux premiers petits ARN découverts [81-83].

Les approches de clonage systématique ont permis d'identifier les miARN les plus abondants mais les miARN présents dans des cellules moins abondantes ou qui sont spécifiques à des événements physiologiques plus

rare peuvent échapper à l'identification par clonage. Ceci a poussé à l'élaboration de nouvelles approches pour l'identification de nouveaux microARN par bio-informatique. L'une de ces approches consiste à rechercher des homologues aux miARN déjà connus [78, 81-83]. L'inconvénient majeur de cette approche est qu'elle ne permet de découvrir que très peu de nouveaux miARN. Une seconde approche mène les scientifiques à s'intéresser à la présence, à proximité de miARN connus, d'autres structures en épingle à cheveux qui pourraient être des miARN organisés en cluster génomique [82, 85, 86]. Cette stratégie tire profit du fait que certains gènes exprimant des microARN sont organisés en cluster. D'autres approches indépendantes de l'homologie ou de la proximité consistent à identifier des segments génomiques conservés qui sont en dehors de régions codantes et qui sont potentiellement capables de former des structures en tige-boucle [79, 84, 87, 88]. Environ 15000 segments chromosomiques chez les humains remplissent ces critères et sont conservés chez au moins trois espèces de vertébrés [88]. Par la suite, un score est donné aux structures en boucles de ces miARN hypothétiques en fonction de leur profil de conservation et leur complémentarité avec des ARNm qui caractérisent les miARN connus. MiRscan est un algorithme basé sur cette méthode [70], il évalue la présence de miARN dans les structures conservées en tige-boucle en déplaçant une fenêtre de 21 nucléotides le long de la séquence. Un score est ensuite attribué à chaque fenêtre qui rend compte de la similarité de la séquence qui y est contenue avec celles des miARN validés expérimentalement chez *C. elegans* et leurs homologues de *C. briggsae* un autre nématode phylogénétiquement proche [80]. MiRseeker, un autre algorithme se basant sur les mêmes caractéristiques a été utilisé pour identifier les miARN chez les insectes [79].

Les deux algorithmes ont mené à l'identification de nouveaux miARN et ont permis d'estimer le nombre de miARN dans le génome humain entre 200 et 255 [88] tandis que ceux de *C. elegans* ont été estimés entre 103 et 120 [80, 86]. Chez la drosophile, ce nombre est entre 96 et 124 [79]. D'autres algorithmes tels que srnaloop, ERPIN, findMIRNA, miralign, ProMIR ont été développés pour la prédiction chez d'autres organismes, ou se basant sur d'autres propriétés structurales des précurseurs des miARN, ou la spécificité des miARN matures [89].

En 2006, Cummins et ses collègues ont développé une approche expérimentale nommée miRAGE pour identifier d'autres miARN. Ils ont isolé des milliers de petits ARN à partir des cellules colorectales humaines qu'ils ont convertis en ADN complémentaire (ADNc) par transcription inverse. Par la suite, les ADNc sont reliés entre eux en une longue chaîne puis clonés et enfin séquencés. Ensuite les séquences sont analysées bioinformatiquement pour une homologie avec des séquences connues ou pour la capacité à former des structures en tige-boucle. L'analyse de plus de 274000 séquences de petits ARN a permis ainsi de confirmer la présence de 200 miARN déjà connus et 133 nouveaux candidats [90].

Aujourd'hui (Octobre 2012), la base publique de référencement des miARN, nommée miRBase, contient 1600 séquences prédites pour l'humain, 855 pour la souris, 238 pour la mouche et 223 pour le nématode [91]. Le nombre de séquences présentes sur la base de toutes les espèces confondues a plus que quadruplé durant ces cinq dernières années. Une telle augmentation de données a été possible grâce à la démocratisation des technologies de séquençage en profondeur [91].

### **1.4.2 Localisation et transcription des miARN**

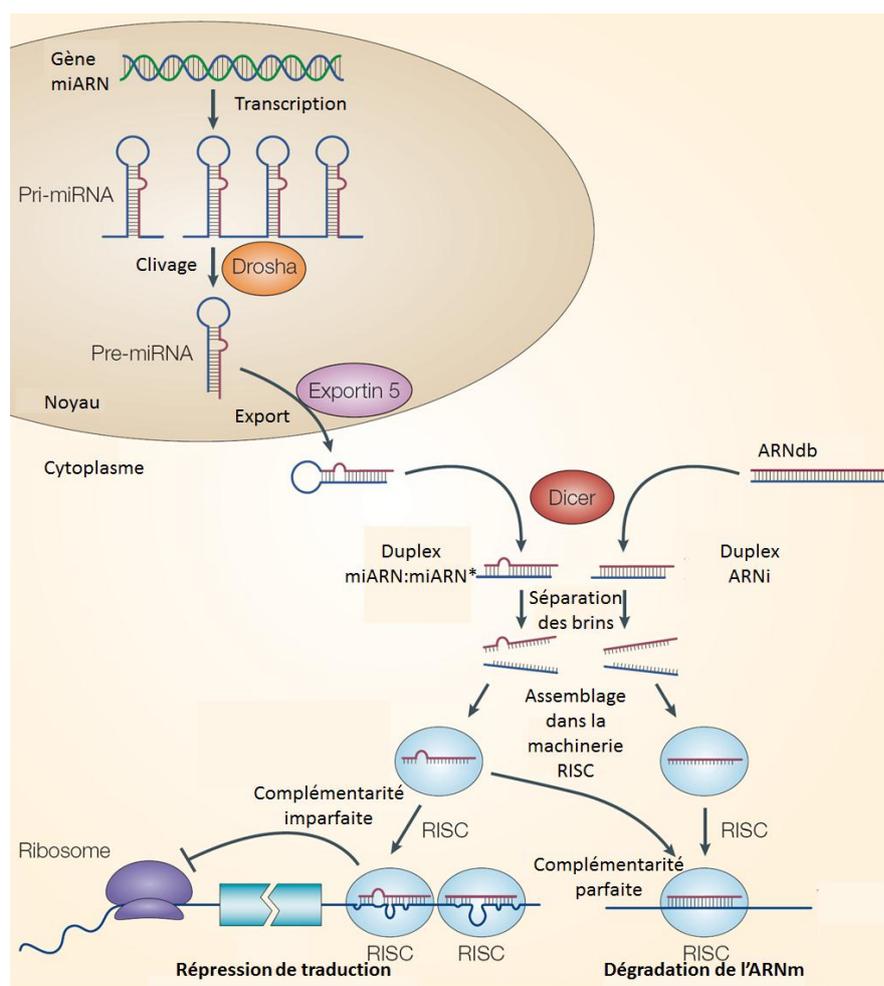
Les premiers microARN identifiés étaient localisés dans les régions intergéniques [81]. Une analyse plus récente sur la localisation génomique de 232 miARN chez les mammifères montre que 70% des miARN sont localisés dans des unités de transcription connues [92]. De plus, plus de deux tiers de ces miARN sont localisés dans des introns dans la même orientation que l'ARN préalablement décrit pour être transcrit à cet endroit. 80% des 117 miARN localisés dans les introns se trouvent dans des unités de transcription codants pour des protéines tandis que les autres se trouvent dans les introns d'ARN non codants[92].

Chez les humains, les miARN sont localisés sur tous les chromosomes excepté le chromosome Y. 50% des microARN sont regroupés en clusters [81, 82] et sont exprimés sous forme d'un transcrit initial polycistronique [93]. Il y a en général deux à trois gènes par cluster [94] et jusqu'à sept gènes dans le cluster le plus grand [95]. Chez la drosophile également, plus de la moitié des miARN connus sont regroupés en clusters [96]. Les précurseurs du cluster de miARN humains miR-23a-miR-27a-mir-24-2 et le miARN isolé miR-21 font respectivement 2.2 et 3.4 kb de long et sont tous les deux non épissés, portent la coiffe en 5' et sont polyadénylés. Par ailleurs, le précurseur de mir-155 contient deux introns, dispose de deux sites distincts de polyadénylation et est épissé alternativement donnant des précurseurs de 0.6 et 1.4 Kb [97, 98].

Les microARN localisés dans les introns partagent les mêmes éléments de régulation que leur gène hôte et sont transcrit dans le même ARN initial. Ces microARN ont donc souvent les mêmes profils d'expression que les gènes dans lesquels ils sont localisés [81, 82, 92]. Pour les miARN transcrits à partir de leur propres promoteur, très peu de transcrits primaires ont été identifiés.

### 1.4.3 Biogénèse des miARN

Contrairement à la majorité des gènes à ARN, les gènes codant les miARN sont transcrits par l'ARN polymérase II [94, 99-101] et forment le premier précurseur des miARN que l'on nomme les pri-miARN et qui comporte une (ou plusieurs) structures en tige-boucle qui abritent la séquence du miARN mature. Lors de la première étape de la maturation, le pri-miARN est coupé par l'enzyme Drosha au niveau nucléaire (Figure 9).



**Figure 9 : Biogénèse des miARN et mécanisme d'action des ARN non codants**  
[extrait de He et al. 2004 [69]]

#### 1.4.3.1 Le clivage nucléaire par Drosha

Les protéines Droshas d'une taille de 130 à 160 KDa contiennent deux domaines catalytiques RNase III et un domaine de liaison à l'ARN double brin [102]. D'autres domaines de fonction inconnue sont localisés dans la moitié N-terminale [103, 104]. Dans le règne animal, un homologue de Drosha est présent dans chaque espèce. Cependant, chez les plantes, aucun gène homologue de Drosha n'a été identifié. [105].

Drosha est une enzyme de type RNase III qui appartient à la famille des ribonucléases spécifiques d'ARN doubles brin. Trois enzymes appartenant à cette famille sont exprimées dans les cellules humaines. L'une est impliquée dans la maturation d'ARNr mitochondrial tandis que les deux autres (Drosha et Dicer) sont nécessaires pour la maturation des miARN.

Drosha coupe le pri-miARN en une structure en tige boucle de soixante-dix nucléotides appelée le pré-miARN. Le pré-miARN dispose d'une terminaison 5' phosphate et de deux nucléotides flottants du côté 3' ce qui est typique des produits d'endonucléases de type RNase III [104, 106]. Les deux nucléotides dépassants de la structure en tige-boucle du côté 3' semblent nécessaires pour le transport des pre-miARN vers le cytoplasme par l'Exportin-5 [107, 108] (Figure 9).

Drosha agit conjointement avec une protéine de liaison à l'ARN double brin Pasha (chez la drosophile) ou son orthologue DGCR8 chez les mammifères et *C. elegans* [109, 110]. Pasha/DGCR8 se lie à la région centrale du domaine RNase III (RIIIDs) de Drosha [102, 109]. Drosha et Pasha/DGCR8 forment un complexe nucléaire de 500 à 650 KDa appelé microProcessor [109-111]. La grande taille de ce complexe est due à une dimérisation de ces composants et à la présence de protéines additionnelles [109]. Les protéines de ce complexe sont toutes requises *in vivo* pour la conversion du pri-miARN en pré-miARN. Une réduction d'expression de Drosha ou Pasha par RNAi conduit à une réduction de pré-miARN et de miARN matures [109-111].

#### **1.4.3.2 Export des pré-miARN**

Lors d'études sur les adénovirus, il a été démontré que l'Exportin-5 (Exp5) est responsable de l'export nucléaire de l'ARN non codant d'adénovirus [112, 113]. La structure de cet ARN non codant qui contient une mini-hélice du côté 5' et 3 nucléotides flottants du côté 3' est essentielle à sa reconnaissance par Exp5 [107, 112]. La similarité de cette structure avec celle du pré-miARN a rapidement conduit à s'intéresser à Exp5 comme responsable du transport nucléocytoplasmique du précurseur des miARN. Effectivement, l'export est dépendant de l'Exportin-5 qui se lie spécifiquement au pré-miARN en présence de Ran-GTP [107, 108] (Figure 9).

Le knock down d'Exp5 a peu d'effet sur le niveau d'expression des pri-miARN et pré-miARN, mais réduit significativement le niveau d'expression de miARN matures et conduit à la rétention nucléaire des pré-miARN [107].

#### **1.4.3.3 Le clivage par Dicer**

Une fois dans le cytoplasme, les pré-miARN sont clivés par Dicer et forment un duplexe miARN d'une vingtaine de nucléotides (Figure 9). Les

protéines Dicer sont des protéines de 200 KDa qui contiennent un domaine ATPase/ARN hélicase, des domaines PAZ, deux domaines catalytiques RNase III et domaine de liaison au niveau C-terminal [102, 103].

Dicer est une protéine hautement conservée dont on trouve un homologue chez la levure (Dcr), un homologue chez l'humain, un chez *C. elegans* (DCR-1), deux chez la drosophile (DCR-1 et DCR-2) et quatre chez *Arabidopsis* (DCL1, DCL2, DCL3, DCL4) [103, 114, 115].

Chaque domaine RNase de Dicer coupe indépendamment un des deux brins d'ARN et génère un duplexe avec deux nucléotides flottants du côté 3' [102]. La structure de son centre catalytique et sa fonction est très similaire à celle de Drosha [104, 116].

Le domaine PAZ de Dicer reconnaît les nucléotides flottants à l'extrémité 3' du pré-miARN. L'homologue de Dicer, Dcr-2 chez la drosophile a besoin d'une hélicase fonctionnelle pour le clivage de l'ARN double brin et l'ATP a un effet très important sur Dcr-2 chez la drosophile et DCR-1 chez *C. elegans in vitro* [117]. En revanche, l'ATP n'est pas requise pour l'homologue humain [118]. Les deux homologues de Dicer chez la drosophile sont impliqués dans deux processus distincts. Dcr-1 fait partie de la machinerie de miARN tandis que Dcr-2 est impliqué dans le mécanisme d'ARN interférence (RNAi) [98, 102, 119, 120] (Page 33).

Comme Drosha, Dicer s'associe à d'autres protéines contenant des domaines de liaison à l'ARN double brin. Dcr-2 forme un hétérodimère avec R2D2 qui contient deux domaines de liaison à l'ARN [117]. Chez les vertébrés et les nématodes, un seul homologue de Dicer est présent et est impliqué dans les deux processus de RNAi et de biogénèse de miARN. Pour cela, d'autres protéines doivent interagir avec Dicer et moduler la spécificité de l'enzyme. Chez *C. elegans*, une protéine similaire à R2D2, RDE-4 a été caractérisée. Cette protéine fait partie d'un complexe composé de Dicer et d'une protéine Argonaute RDE-1 et est essentielle pour le fonctionnement de la machinerie RNAi et non les miARN [121]. Les protéines Dicer interagissent également avec les protéines Argonautes [115].

#### **1.4.3.4 Le complexe RISC et le mécanisme d'action des miARN**

Une fois clivés par Dicer, les miARN matures d'une vingtaine de nucléotides sont incorporés dans un complexe responsable de la répression de l'expression par ARN que l'on appelle RISC (RNA Induced Silencing Complex) [122-124]. Lors de l'assemblage du complexe RISC, l'un des deux brins du duplexe de petit ARN issu du clivage est incorporé après avoir été désapparié par une hélicase tandis que l'autre brin appelé miARN\* est dégradé (Figure 9). Le brin incorporé dans le complexe RISC est celui dont l'extrémité 5' est la moins fortement appariée au côté 3' de l'autre brin du duplexe [70, 125, 126]. Le complexe RISC identifie ses cibles grâce à la

complémentarité entre la vingtaine de nucléotides du petit ARN qu'il transporte et l'ARNm cible (Figure 9). Du fait de l'existence d'une complémentarité imparfaite entre la séquence d'un miARN et sa cible localisée sur l'ARNm, le complexe RISC induit une répression de la traduction. Dans le cas d'une complémentarité parfaite entre le miARN et sa cible, l'ARNm est dégradé [127-129] par l'activité endonucléase du complexe RISC[70]. La coupure s'opère alors au niveau des nucléotides complémentaires aux résidus 10 et 11 du miARN [123, 127, 130, 131].

La voie majeure de régulation par miARN semble la répression de traduction de l'ARNm. Dans les cellules humaines, les miARN inhibent la synthèse de protéines aux stades précoces d'initiation de la traduction et les ARNm ainsi réprimés sont localisés dans des organelles intracellulaires appelée P-bodies [132].

#### **1.4.4 Les cibles des miARN**

Les sites reconnaissance des miARN sont situés dans les régions 3' UTR des gènes cibles [69]. Les régions 3'UTR peuvent posséder plusieurs séquences complémentaires aux miARN [65, 66, 69, 75-77, 104, 133] (Figure 8). La multiplication de ces sites cibles apporte une répression de traduction plus efficace par le complexe RISC [129]. Par ailleurs, les miARN peuvent également se lier à des séquences complémentaires dans le cadre ouvert de lecture de l'ARNm cible ou à son extrémité 5' UTR mais ces cas restent cependant anecdotiques pour le moment (un peu comme l'existence des miRNA entre 1993 et 2000 ?) [129, 134]. Certains petits ARN non codants semblent également capables de cibler non seulement les ARNm mais également l'ADN pour une répression de transcription chez les plantes et les champignons [135-139] dans ces cas, ils sont vraisemblablement associés à des complexes comparables au RISC encore méconnus.

Différentes approches informatiques ont été utilisées pour découvrir les cibles potentielles [140, 141]. La majorité de ces méthodes sont basées sur la connaissance de la nature entre l'appariement des miARN avec leurs gènes cibles. Des programmes ont été développés chez les mammifères et la drosophile pour la prédiction des cibles [142-148] qui se basent sur la complémentarité entre les régions 3'UTR des ARNm cibles et les miARN. Un accent est mis sur l'appariement critique à l'extrémité 5' du miARN avec sa cible comme cela a été démontré pour les miARN préalablement validés [149, 150].

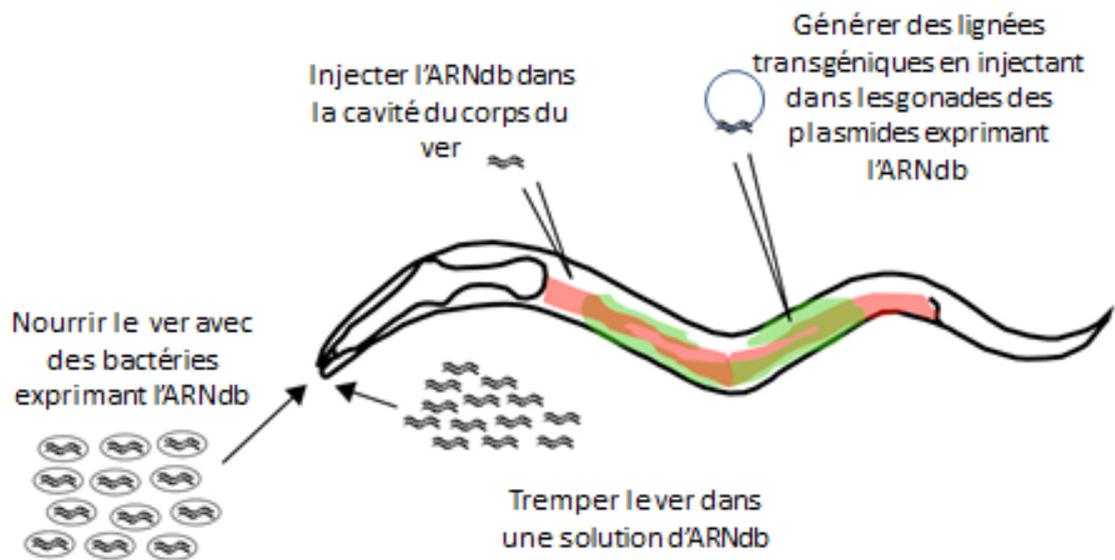
On retrouve parmi les algorithmes développés TargetScan et TargetScanS pour la prédiction chez les vertébrés, le nématode, la drosophile et le poisson [144, 151], DIANA MicroT pour l'humain et la souris [147], miRanda , miRNA-Target prediction et RNAhybrid la drosophile [142, 143, 146, 148, 152].

## 1.5 L'ARN INTERFERENCE

L'ARN anti-sens a été utilisé expérimentalement pour la première fois en 1984 pour induire la répression spécifique de l'expression d'un ARNm dans des cellules en culture par Izant et Weintraub [153, 154]. Le concept initial était qu'un anti-sens s'apparierait sur le messager et empêcherait sa traduction. Rapidement, la technique a été appliquée dans des systèmes variés comme les ovocytes de Xénope [155], ceux de souris [156] et les embryons de drosophile [157]. Chez *C. elegans*, des ARN anti-sens exprimés à partir de transgènes permettaient de bloquer l'expression de deux gènes musculaires *unc-22* et *unc-54* [158]. Cependant, l'inconstance de l'efficacité de l'approche par anti-sens d'un gène à l'autre et d'une application à l'autre dans plusieurs systèmes a conduit à conclure que cette méthodologie n'était pas fiable pour l'inhibition de l'expression des gènes [159]. En 1995, Guo et Kemphues montrent que la micro-injection d'ARN anti-sens correspondant au gène *par-1* cause le phénotype perte de fonction avec efficacité [160]. L'utilisation d'ARN anti-sens pour inhiber la fonction des gènes a rapidement été adoptée chez *C. elegans* à mesure que le nombre de gènes ciblés augmentait [159]. Deux observations capitales ont mené à suggérer que quelque chose de plus que le simple appariement dépendant de la concentration de l'ARN anti-sens était impliqué dans le processus d'interférence chez *C. elegans*. La première venait de Craig Mello qui observait que l'effet de l'interférence était transmis jusqu'à deux générations après l'injection de l'ARN [159]. La seconde vient de la découverte que l'interférence se propage aux autres tissus au-delà du site d'injection [161].

Dans la même étude, Fire et ses collègues découvrent que l'utilisation d'ARN double brin (double strand RNA, dsRNA) était au moins 10 fois plus efficace que les injections d'ARN anti-sens [161] ce qui a valu à Craig Mello et Andrew Fire d'être récompensés par un prix Nobel de médecine en 2006. Cette découverte (le RNAi) a été d'une grande importance car elle a rapidement mené à l'utilisation de l'ARN double brin pour réprimer l'expression des gènes avec une grande efficacité dans une multitude d'organismes incluant les plantes [162], les trypanosomes [163], la drosophile [164], le planaria (des vers plats des marais) [165], les hydres [166] et les embryons de souris [167].

Hormis l'origine exogène de l'ARN double brin introduit pour induire de l'ARN interférence, le mécanisme d'action de l'ARNi est identique à celui des miARN à partir de la prise en charge du dsRNA par DICER jusqu'à son inclusion dans la machinerie RISC (Figure 9). La complémentarité parfaite entre l'ARNi et sa cible en ARNm induit la dégradation de l'ARNm contrairement à la majorité des miARN qui inhibent la traduction (Figure 9).



**Figure 10 : Les différents modes d'administrations de l'ARNdb chez *C. elegans* pour induire l'ARNi**

Chez *C. elegans*, l'efficacité de l'utilisation de l'ARN double brin a rendu possible de nouvelles méthodes pour induire l'ARNi. Cela inclut de tremper simplement les vers dans des solutions de dsRNA [168], nourrir les vers avec des bactéries *E. coli* exprimant l'ARN double brin [169], injecter les dsRNA [161] ou générer des vers transgéniques exprimant l'ARN double brin [168, 170] (Figure 10). Ces nouveaux modes d'administration ont ouvert la porte à une multitude de nouvelles applications expérimentales de l'ARNi comme la réalisation d'un crible génétique à la recherche de mutants résistants à l'ARNi ce qui a conduit à l'identification de l'implication de la protéine RDE-1 (de la famille Argonaute) dans le processus d'interférence [168] et plus globalement à l'inhibition systématique des gènes à l'échelle génomique [171-174].

## **1.6 CAENORHABDITIS ELEGANS COMME MODELE D'ETUDE**

### **1.6.1 Historique**

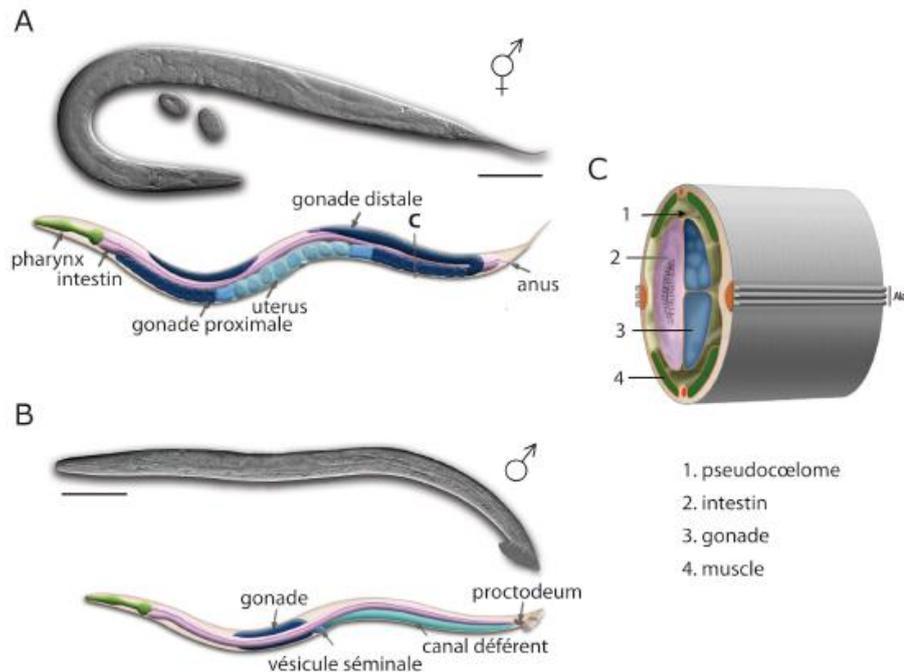
En 1965, Sydney Brenner s'est porté sur *Caenorhabditis elegans* (*C. elegans*) comme modèle d'étude pour étudier le développement et le comportement [175, 176]. Les travaux qu'il a initiés chez cet organisme lui ont valu le prix Nobel de médecine en 2002 conjointement avec Robert Horvitz et John E. Sulston. Ce nématode offre un grand potentiel pour entreprendre des analyses génétiques en partie grâce à son cycle de développement très rapide (3 jours), son mode de reproduction par autofécondation des hermaphrodites permet à un seul ver de se reproduire en cas de dispersion à la recherche de nourriture (pour le maintien des lignées) ou par croisement entre les mâles et les hermaphrodites (études génétiques de complémentarité), sa petite taille (1.5 mm pour un adulte), sa transparence qui permet de visualiser les cellules et les organes *in vivo* sous microscope et permettant un recours à des protéines fluorescentes et la facilité de sa culture et son entretien en laboratoire.

L'espèce *Rhabditis* a été mise en lumière par Dougherty and Calhoun en 1948 [177]. *C. elegans* a cependant été décrit et nommé *Rhabditis elegans* par Maupas au début du XXème siècle [178]. En 1952, Osche l'a placé dans le sous genre *Caenorhabditis* [179] qui par la suite a été élevé au stade d'espèce générique par Dougherty [180]. Le nom a pour racine un mélange gréco-romain : *Caeno* veut dire récent, *Rhabditis* signifie bâtonnet et *elegans* en raison de l'élégance de son mode de locomotion.

Historiquement, deux souches de ce nématode ont été utilisées. La première a été isolée à Bergerac, dont elle porte le nom, par Victor Nigon de l'Université de Lyon [181]. L'autre souche, Bristol, porte également le nom du sol dont elle a été originellement isolée en Angleterre en 1959 dans un compost de champignon [182]. Cependant, à cause de l'infertilité des mâles de la souche Bergerac et la stérilité sensible à la température des hermaphrodites de cette lignée [183], la majorité des études génétiques ont été réalisées sur un dérivé N2 de la lignée Bristol établie dans le laboratoire de Sydney Brenner en 1964.

## 1.6.2 Caractéristiques

### 1.6.2.1 Anatomie



**Figure 11 : Anatomie de *C. elegans* adulte hermaphrodite et mâle**

Image Nomarski et représentation schématique d'un vers hermaphrodite adulte (A). Image Nomarski et représentation schématique d'un vers mâle adulte (B). Vue schématique d'une coupe transversale côté postérieur d'un hermaphrodite adulte (C). [Adapté de [www.wormatlas.org](http://www.wormatlas.org)]

Les nématodes sont tous constitués d'un corps respectant un plan simple qui se compose de deux tubes concentriques séparés par un espace rempli de fluide, le pseudocoelome (Figure 11 C). La forme de l'animal est maintenue par la pression hydrostatique interne grâce à la pression osmotique [184]. Le tube externe est recouvert de collagène et d'une cuticule extracellulaire sécrétée par l'hypoderme. Les muscles du corps sont arrangés en quatre bandes longitudinales attachés à la cuticule à travers une fine couche de l'hypoderme [185]. La contraction des deux muscles subventraux et la relaxation des dorsaux génère un mouvement de locomotion sinusoïdale dans le plan dorso-ventral [186]. Le système nerveux, les gonades, les cellules du coelome et le système sécrétoire sont également présents dans ce tube externe [184]. Le tube interne est composé des muscles du pharynx avec son système nerveux et de l'intestin (Figure 11). Le pharynx bilobé joue le rôle d'une pompe aspirante de nourriture qui envoie les aliments vers l'intestin. Les cellules intestinales entourent un lumen central connecté à l'anus à proximité de la queue. Tous les nématodes muent quatre fois durant leur développement [187]. *C. elegans* ne déroge pas à cette règle de mue (excepté au stade Dauer). Au cours de son cycle de vie, sa taille augmente environ d'un tiers entre chaque mue et continue d'augmenter tout au long

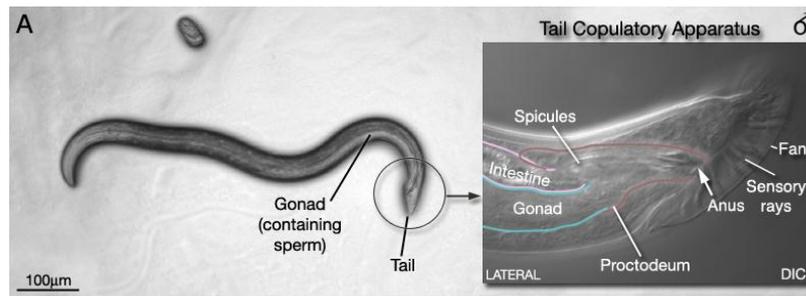


### **1.6.2.3 Développement post-embryonnaire**

Le développement post-embryonnaire débute après l'éclosion. En absence de nourriture, le développement est arrêté jusqu'à ce qu'une source alimentaire soit disponible (Figure 12). Les divisions cellulaires et le programme de développement reprennent lors de la prise alimentaire [195]. Les larves peuvent survivre ainsi jusqu'à dix jours [196, 197]. Dès qu'une source alimentaire est trouvée, le développement reprend d'une manière normale [198]. La première mue se produit au bout de 12h et l'animal passe au stade larvaire 2 (L2). Il se peut qu'avant ce stade, en présence d'une surpopulation, de famine ou de température élevée, les nématodes entrent dans un stade particulier appelé « *Dauer* » (Figure 12). Ce stade commence par une phase intermédiaire L2d qui est capable de reprendre un développement normal vers le troisième stade (L3) en cas de conditions favorables. Si la pénurie continue, L2d entre dans le stade *Dauer* proprement dit [199]. Durant ce stade, la cavité buccale est fermée et le système locomoteur continue à propulser l'animal à la recherche d'une nouvelle source de nourriture. L'animal sort du stade *Dauer* 1h après l'accès à la nourriture. Il commence à se nourrir dans les 2 à 3h suivants et au bout de 10h, le nématode mue vers le dernier stade larvaire L4. Durant le stade L4, la formation des gonades qui a débuté à la fin du stade L3 est achevée. Les bras des gonades s'allongent en migrant vers les extrémités le long du muscle dorsal (body wall muscle). La méiose dans la lignée germinale débute à partir de la mue L3/L4 et les cellules germinales se différencient en spermatozoïdes matures. Lors de la mue vers le stade adulte, la production de sperme s'arrête et les cellules germinales se spécialisent dans la formation exclusive des ovocytes. 45 à 50h après éclosion, le nématode hermaphrodite pond son premier œuf commençant ainsi sa période reproductive qui durera 4 jours [188]. L'adulte hermaphrodite est capable de produire jusqu'à 300 œufs par autofécondation. Au-delà de cette période fertile, il est capable de vivre encore 10 à 15 jours et éventuellement de continuer à se reproduire si il est fécondé par un mâle [200] (Figure 12).

### **1.6.2.4 La présence de mâles**

L'un des aspects attrayants du modèle *C. elegans* réside dans le fait que des mâles peuvent apparaître et sont capables de féconder les hermaphrodites. 0.1% des vers sont des mâles qui apparaissent spontanément à cause de la non-disjonction du chromosome X [201, 202]. Le génotype des mâles est donc 5AA ; X0 car il ne comporte qu'un seul chromosome X (AA pour les différentes paires d'autosomes et X0 pour le chromosome sexuel) alors que le génotype des hermaphrodites est 5AA ; XX. La disponibilité des mâles est intéressante pour réaliser des croisements et procéder à des analyses génétiques. La capacité de *C. elegans* de se reproduire par auto-fécondation et par croisement rend possible des analyses génétiques comme la complémentation et une maintenance facile de lignées. Les mâles diffèrent des hermaphrodites par leur anatomie globale.



**Figure 13 : Anatomie caractéristique du mâle**

Image Nomarski d'un mâle *C. elegans* (gauche) et son corps copulateur en DIC (droite). [www.wormatlas.org]

Sous un stéréomicroscope, le mâle peut être reconnu grâce à son corps fin, ses gonades, contenant du sperme, qui donnent une apparence ventrale blanche et son corps copulateur caractéristique qui se situe au bout de la queue (Figure 13). A âge égal, le mâle est plus petit que l'hermaphrodite [200]. Comme chez l'hermaphrodite, la filiation de chacune des cellules a été réalisée [189, 203, 204]. Dans le cas d'un croisement entre un mâle et un hermaphrodite, la descendance compte 50% de mâles. Les mâles peuvent s'accoupler pendant 6 jours suivant leur entrée au stade adulte et peuvent en théorie engendrer environ 6000 nématodes [200].

### **1.6.3 Génétique de *C. elegans***

La publication du génome entier en 1998 a augmenté la popularité de *C. elegans* comme modèle d'étude [133]. Il est le premier organisme multicellulaire à avoir été pleinement séquencé et à peine le deuxième eucaryote derrière *Saccharomyces cerevisiae*. Le génome de *C. elegans* est de  $100 \times 10^6$  paires de bases (pb) réparties sur 5 paires d'autosomes et une paire de chromosomes sexuels chez l'hermaphrodite. Le génome de *C. elegans* code environ 20000 gènes. Les transcrits chez *C. elegans* sont assemblés à partir d'une moyenne de 6.4 exons (9,7 pour les humains) [205]. La taille moyenne de l'exon chez *C. elegans* est de 218 bases alors que 56% des introns chez *C. elegans* ont une taille inférieure à 100pb [206].

#### **1.6.3.1 Les opérons**

Le trans-splicing des ARNm est un épissage qui nécessite la machinerie du spliceosome excepté le snARN U1 [207]. Il a d'abord été observé chez *C. elegans* et ensuite découvert chez d'autres organismes [208, 209]. Il a abouti à la découverte des opérons chez *C. elegans* car il est utilisé pour le traitement des pré-ARNm polycistroniques [210, 211]. 15% de la totalité des gènes de *C. elegans* sont regroupés en opérons [210]. Le premier opéron à avoir été identifié et étudié chez *C. elegans* fut le *gpd-2/gpd-3* [209] (Figure 14). Dans cette région génomique on retrouve le gène *mai-1*, le gène *gpd-2* et *gpd-3*. Ce cluster de trois gènes constitue un opéron. Les trois gènes partagent un même promoteur ce qui leur permet d'être co-transcrits [211].

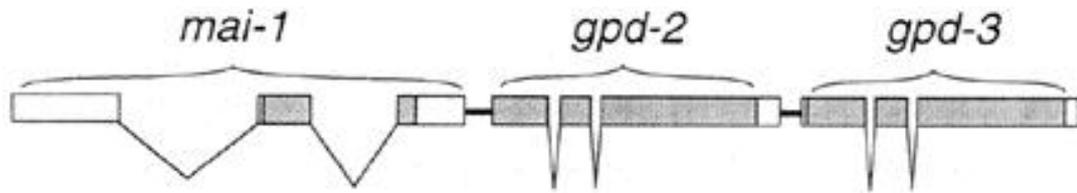


Figure 14 : Organisation de l'opéron polycistronique *mai-1/gpd-2/gpd-3* chez *C. elegans*

Le système opéron chez les bactéries permet une expression simultanée de gènes qui opèrent ensemble. L'ARNm des opérons reste un transcrite unique qui donne plusieurs protéines lors de la traduction. Chez *C. elegans* l'ARN issu d'un opéron est mûré en ARNm distincts pour chacun des gènes. De ce fait, même si ils sont co-transcrits, les produits de gènes peuvent ne pas coexister grâce à la différence de stabilité des messagers [211-213].

### 1.6.3.2 L'épissage alternatif

25 % des gènes subissent un épissage alternatif chez *C. elegans* [214]. Le spliceosome basal responsable de l'épissage est le même chez *C. elegans* que chez les autres eucaryotes [31, 215, 216]. Plus d'une centaine de protéines différentes peuvent faire partie de la composition du spliceosome comme facteurs associés au pré-ARN messager (préARNm) ou aux snRNP [217, 218]. Une séquence consensus en 5' du site d'épissage chez *C. elegans* a été déterminée pour lier par complémentarité parfaite le 5' du snRNA U1. Il s'agit de la séquence AG|guaaguu (les lettres en majuscules représentent la fin de l'exon et le début de l'intron est en minuscule) dans 51% des cas où l'intron débute par un G [219].

L'identification initiale des gènes épissés alternativement chez *C. elegans* dépendait de la comparaison des EST (Expressed sequence Tags) à la séquence génomique qui a permis l'identification de 844 gènes concernés [219]. Ce nombre a augmenté grâce à l'annotation des gènes, de nouvelles EST identifiées et l'analyse des données de SAGE [220]. De nouvelles données permettant l'identification de nouveaux gènes ont été accumulées grâce aux nouvelles technologies de séquençage ARN à haut débit (RNA-seq) ce qui a permis d'identifier 2183 nouveaux événements [214]. Cette étude réalisée par Ramani et ses collègues a permis ainsi de déterminer que 5000 gènes seraient soumis à l'épissage alternatif parmi les 20000 gènes de *C. elegans*. Le fait que le quart des gènes codés par le génome subiraient un épissage alternatif montre que c'est un processus de régulation des gènes majeur chez le nématode. Plusieurs études comparatives entre différentes espèces de nématodes montrent que chez l'espèce *Caenorhabditis*, l'épissage alternatif subit de très fortes pressions de sélection [221-224].

## **1.6.4 Visualisation de la régulation post-transcriptionnelle *in vivo***

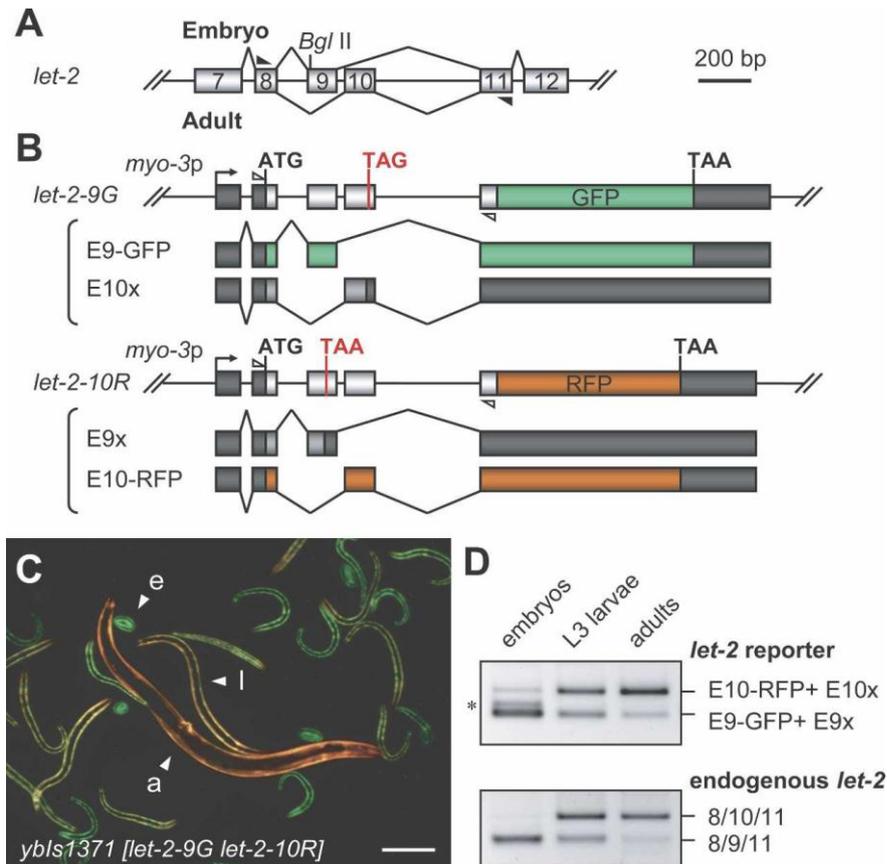
### **1.6.4.1 Etude de l'épissage alternatif par des rapporteurs bicolores**

#### 1.6.4.1.1 Epissage du gène *let-2*

Chez le nématode, une partie de l'épissage alternatif soumis à une régulation au cours du développement [214]. *let-2*, un gène codant pour le collagène de type IV dispose de deux exons mutuellement exclusifs qui génèrent les versions de la protéine spécifiques à l'embryon et à l'adulte [61].

Le gène *let-2* code pour le collagène  $\alpha 2$  de *C. elegans* il contient deux exons alternatifs mutuellement exclusifs 9 et 10 qui sont utilisés différemment au cours du développement (Figure 15 A). Dans les embryons, seul l'ARNm incluant l'exon 9 est exprimé tandis que dans les stades larvaires tardifs et chez l'adulte l'ARNm incluant l'exon 10 prédomine.

Cette régulation est conservée au cours de l'évolution chez les nématodes *C. elegans* et *Ascaris* [61, 225]. Grâce à un système de mini-gènes, Notre collaborateur Hidehito Kuroyanagi (Tokyo, Japan) a réussi à créer une lignée fluorescente bi colore rapportrice de ce processus [60] (Figure 15 C).



**Figure 15 : Epissage alternatif mutuellement exclusif du gène *let-2* régulé au cours du développement.**

Représentation schématique de la région épissée alternativement du gène *let-2* endogène. Les rectangles représentent les exons avec leur numéro indiqué (A). Représentation schématique des deux mini-gènes rapporteurs de l'épissage de *let-2*, *let-2-9G* et *let-2-10R* et leur ARNm. Les ORF prédites sont en vert (*let-2-9G*) en rouge (*let-2-10R*) ou en gris (*let-2-9X* et *let-2-10X*) (B). Image en microscopie à fluorescence de vers transgéniques *ybls1371* exprimant les deux minigènes, (e) montre un embryon, (l) un vers L3 et (a) un adulte (C). Résultats de RT-PCR obtenus avec les amorces représentées par un triangle sur les figures A et B pour les minigènes (haut) et *let-2* endogène (bas) (D). [Ohno *et al.* 2008 [60]]

Pour pouvoir visualiser la sélection mutuellement exclusive des exons du gène *let-2* *in vivo*, deux systèmes de mini-gènes rapporteurs *let-9G* et *let-10R* ont été construits (Figure 15 B). Les mini-gènes portent le fragment génomique de *let-2* de l'exon 8 à l'exon 11 connectés en phase avec l'ADNc des protéines fluorescentes GFP ou RFP. Des codons stops ont été artificiellement introduits dans l'exon 10 pour *let-9G* et dans l'exon 9 de *let-10R*. La construction *let-9G* permet lors de l'inclusion de l'exon 9 d'exprimer un ARNm codant pour une protéine GFP fusion (E9-GFP) tandis que l'inclusion de l'exon 10 aboutit à un ARNm contenant un codon stop en amont de la séquence codante de la GFP. De la même manière, la construction *let-10R* aboutit à la production d'un ARNm exprimant une protéine RFP fusion (E10-RFP) uniquement lors de l'inclusion de l'exon 10, (Figure 15 B). L'expression des deux mini-gènes a été placée sous le contrôle du promoteur du gène *myo-3* spécifique du body wall muscle (BWM) où le gène *let-2* endogène est exprimé [61, 226].

Les rapporteurs *let-2* ont permis de visualiser avec succès le phénomène d'épissage alternatif (Figure 15 C). Comme attendu, l'expression des rapporteurs dans le « body wall muscle » (BWM) a permuté graduellement et presque totalement du vert au rouge au cours du développement du nématode [61]. Les embryons expriment exclusivement la GFP (E9-GFP) tandis que les vers de stade de développement plus tardifs expriment majoritairement la RFP (E10-RFP). La permutation d'usage entre exon 9 et 10 a été confirmée par RT-PCR où l'on voit l'exclusion de l'exon 9 et l'inclusion de l'exon 10 aux stades tardifs du développement (Figure 15 D). Ces résultats montrent que la lignée générée (KH1401) permet de visualiser le phénomène d'exclusion mutuelle des exons du gène *let-2* au cours du développement.

Pour identifier les facteurs régulant l'épissage alternatif du gène *let-2*, l'équipe de notre collaborateur a réalisé une mutagénèse aléatoire sur la lignée KH1401.

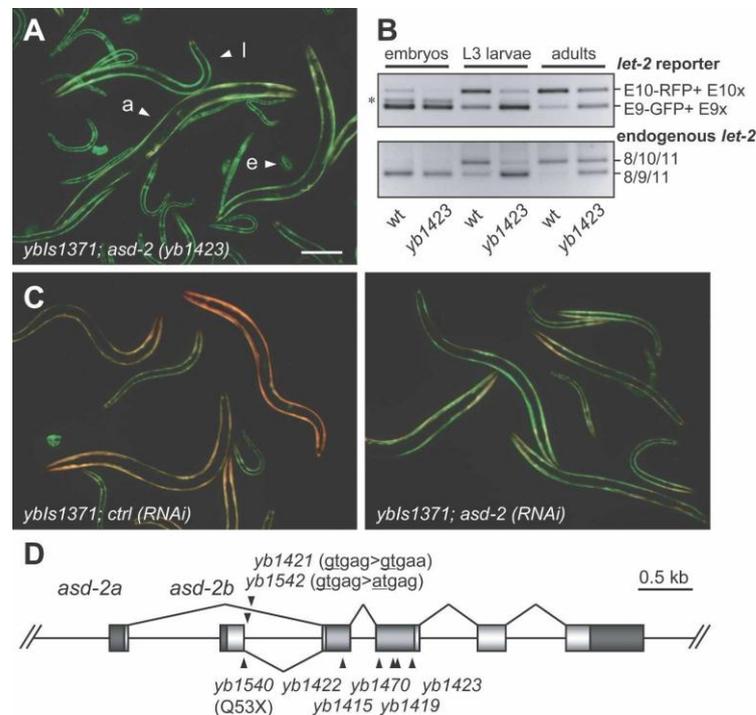
Classiquement, deux types de mutagènes aléatoires sont utilisées dans la plupart des systèmes modèles : les agents biologiques (tels que les plasmides, les virus ou les transposons) dont l'insertion dérègle le fonctionnement des éléments d'ADN fonctionnels, les agents physiques tels que les rayons ultra-violets et les rayons X. ou encore, les agents chimiques (tels que l'éthyle méthane sulfonate (EMS) ou le *N*-ethyl *N*-nitroso urée (ENU)) qui introduisent des mutations ponctuelles ou des délétions dans les séquences génomiques. L'utilisation des facteurs mutagènes biologiques facilite l'identification du locus génomique responsable du phénotype grâce à l'empreinte moléculaire apportée par l'insertion de l'élément. Cependant, la mutagénèse chimique apporte une fréquence de mutagénèse plus élevée et induit un spectre de phénotypes plus large (hypomorphes, hypermorphes, amorphes et néomorphes) très difficile à obtenir par des agents biologiques. Qui ne permettent souvent qu'une perte de fonction.

L'équipe de H. Kuroyanagi a utilisé l'EMS comme agent mutagène sur la souche rapportrice puis a procédé au criblage visuel sous microscope de la descendance F2.

Chaque vers est capable d'engendrer en moyenne une descendance de 300 vers. La mutagénèse des vers P0 touche entre autre les gonades de celui-ci. Sa progéniture hérite d'un spectre de mutations différent selon la cellule germinale retenue lors de la fécondation. *C. elegans* étant un organisme diploïde, la mutagénèse d'un ver P0 engendre deux fois 300 génomes différents. Pour couvrir le plus de génomes possible, il est donc nécessaire de procéder à la mutagénèse du plus grand nombre d'individus possible.

$5 \times 10^4$  génomes ont été ainsi criblés par cette approche à la recherche d'individus présentant une altération de l'expression des rapporteurs

fluorescents. Aucun nématode exprimant la RFP dès les stades précoces n'a été isolé. En revanche, Plusieurs allèles indépendants ont été isolés qui expriment la GFP jusqu'aux stades tardifs du développement (Figure 16 A, D).



**Figure 16 : Mutants *asd-2* déficients en épissage alternatif de *let-2***

Image en microscopie à fluorescence d'un mutant pour le gène *asd-2*, les flèches indiquent un embryon, un L3 et un adulte (A). Résultats de RT-PCR des minigènes (haut) et de *let-2* endogène (bas) pour les différents stades (B). Expérience d'ARNi sur la lignée KH1401 avec le tampon seul (gauche) ou avec l'ARNdb ciblant *asd-2* (droite), la visualisation se fait par microscopie à fluorescence (C). Représentation Schématique du gène *asd-2*. Les exons sont représentés par des rectangles blancs pour l'ORF et foncés pour les UTR. La localisation des mutations portées par les mutants identifiés sont indiquées (D). [Ohno et al. 2008 [60]]

L'analyse par RT-PCR du gène *let-2* confirme que les mutants sont déficients pour l'inclusion de l'exon 10 dans les stades tardifs du développement comme suggéré par la perte de la fluorescence rouge (Figure 16 B). La cartographie des SNP combinée au séquençage génomique de l'ADN ont indiqués que les 8 mutants obtenus portent des allèles d'un gène unique T21G5.5 non caractérisé jusqu'à lors. A la suite de cette étude, le gène T21G5.5 a été nommé *asd-2* pour « alternative splicing defective-2 ». Le gène *asd-2* est lui-même soumis à un épissage alternatif donnant deux isoformes protéiques, ASD-2a et ASD-2b. Seule l'isoforme ASD-2b est impliquée dans l'épissage alternatif de *let-2* (Figure 16 D).

Aucune autre mutation portée sur un autre gène n'a été identifiée lors de cette étude. Il est cependant peu probable qu'aucun autre gène ne participe à l'établissement de la régulation de la transition entre les deux isoformes au cours du développement. En effet, même si la protéine ASD-2 était l'unique effecteur déterminant cette régulation, les gènes affectant sa

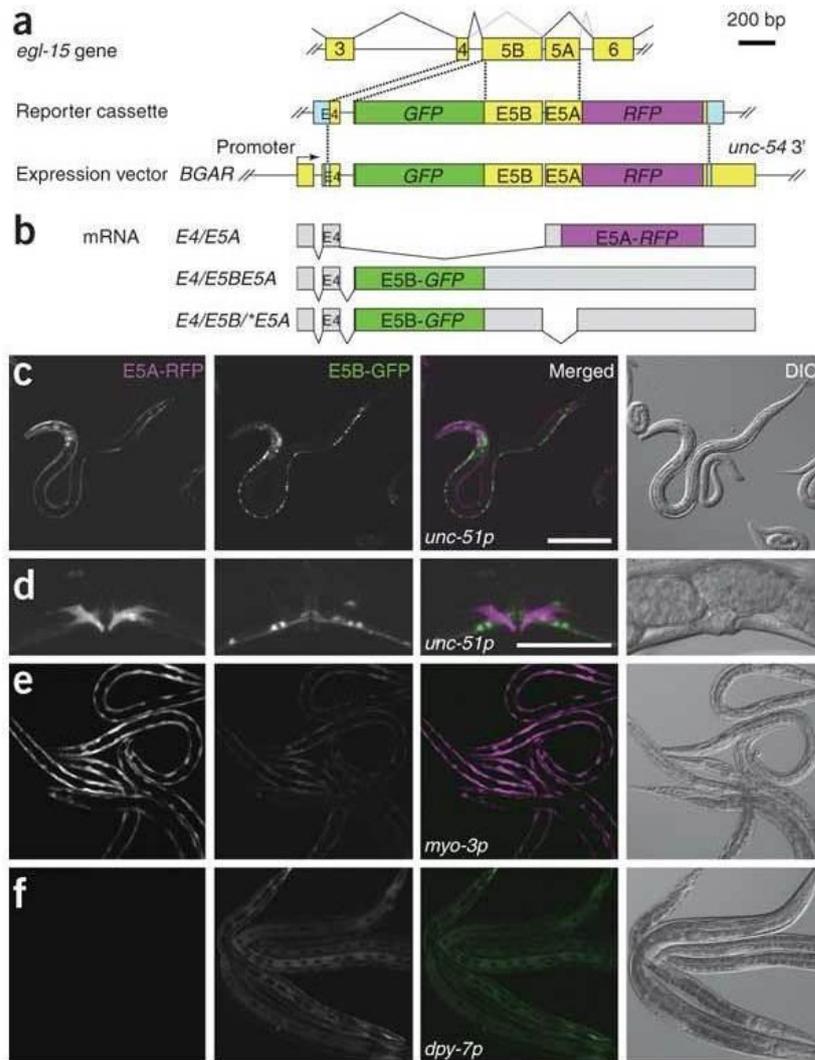
propre expression et/ou son épissage devraient également avoir un effet sur la régulation de *let-2*.

Notre hypothèse de travail était donc que bien que le crible initial effectué par Kuroyanagi et ses collègues semble avoir atteint la saturation, l'utilisation d'un système de détection plus quantitatif devrait nous permettre d'identifier des éléments de régulation plus subtils.

Dans le premier chapitre des résultats je présenterai les efforts que j'ai effectués pour identifier d'autres gènes impliqués dans la régulation de *let-2*. Pour ce faire j'ai d'une part mis en place un crible de mutagenèse automatisé utilisant les propriétés du COPAS-Profiler, et d'autre part modifié le système de bras robotique associé (Module Reflx) pour permettre la réalisation d'un crible par ARN interférence qui a été réalisé par ma collègue Karine Reborá.

#### 1.6.4.1.2 Epissage du gène *egl-15*

Antérieurement à l'étude du gène *let-2*, l'équipe japonaise a apporté une preuve de principe sur la faisabilité d'un rapporteur *in vivo* de l'épissage alternatif [227] (Figure 17). Le gène *egl-15* qui code pour l'unique homologue du récepteur fibroblastique des facteurs de croissances (FGFR) chez *C. elegans* a été choisi comme modèle. Ce gène contient également deux exons mutuellement exclusifs le 5A et le 5B qui sont différentiellement inclus de manière tissu-spécifique [62, 227]. L'équipe a réalisé une construction rapportrice où l'exon 5A est en phase avec la RFP alors que le l'exon 5B est en phase avec la GFP (Figure 17 a). Cette cassette rapportrice a été nommée BGAR.



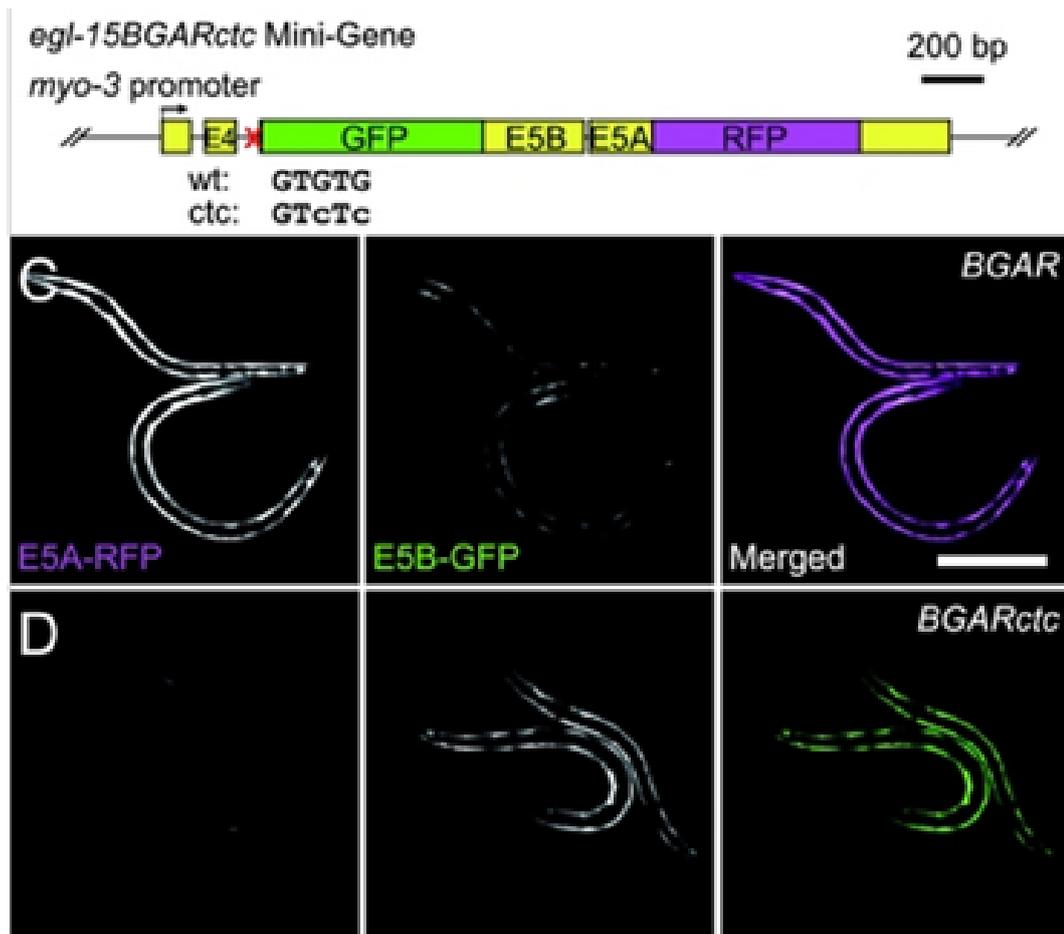
**Figure 17 : Des vers rapporteurs de l'épissage alternatif mutuellement exclusif du gène *egl-15***  
 Construction du rapporteur de l'épissage du gène *egl-15*, E5B-GFP-E5A-RFP nommé BGAR (a). Les ARNm obtenus par l'expression de la cassette BGAR sont représentées en vert lors de l'inclusion de l'exon 5B et en magenta lors de l'inclusion de l'exon 5A (b). Vers transgéniques exprimant la cassette BGAR sous le contrôle du promoteur *unc-51* (c,d) *myo-3* (e) et *dpy-7* (f). [Kuroyanagi et al. 2006 [227]]

L'utilisation de promoteurs tissus spécifiques engendre l'inclusion de l'un des deux exons (Figure 17 c-f). Sur la lignée portant la construction rapportrice, ceci se traduit par la visualisation de la GFP ou de la RFP. Par exemple, sous le contrôle du promoteur *myo-3* qui s'exprime dans le BWM, on voit la fluorescence de la RFP dans la totalité du tissu (Figure 17 e). C'est donc l'exon 5A qui est inclus dans le BWM. En revanche, sous le contrôle du promoteur *dpy-7*, c'est l'exon 5B qui est inclus lors de l'expression hypodermale du rapporteur qui se traduit par la visualisation de la GFP (Figure 17 f).

Cette étude a identifié le gène *asd-1* et *fox-1* comme impliqués dans cet épissage mutuellement exclusif. Ces deux gènes appartiennent à la famille des gènes Fox-1 qui ont été identifiés chez les vertébrés. Les gènes de cette famille ont été identifiés comme des régulateurs de l'épissage alternatif avec

des propriétés de liaison spécifiques à la séquence UGCAUG chez les mammifères [228, 229]. Une séquence UGCAUG conservée entre *C. elegans* et *C. briggsae* est justement présente au niveau de l'intron 4 du gène *egl-15* [227]. La modification de cette séquence en UcgAUG fait en sorte que l'inclusion de l'exon 5B devient systématique alors que celle de l'exon 5A devient indétectable [227]. ASD-1 et FOX-1 se fixent spécifiquement à cette séquence et empêchent la sélection de l'exon 5B par la machinerie de l'épissage qui se reporte sur l'exon 5A. Cette régulation est dépendante du dosage protéique.

Cette étude illustre à nouveau la puissance de l'utilisation de systèmes rapporteurs bicolores *in vivo* chez *C. elegans*. Comme pour l'étude de *let-2* qui a abouti à l'identification d'*asd-2*, l'équipe a été capable de visualiser la modulation de fluorescence. Les causes de cette modulation ont été disséquées et ont permis l'identification de nouveaux facteurs en *trans* impliqués dans l'épissage alternatif chez *C. elegans*, ASD-1 et FOX-1. Ces facteurs en *trans* lient spécifiquement une séquence régulatrice en *cis* hautement conservée (UGCAUG). Une étude ultérieure réalisée avec le même rapporteur a abouti à l'identification de la protéine SUP-12 comme étant affine pour une autre séquence *cis* régulatrice GUGUG voisine de la première et qui est essentielle à l'épissage alternatif du gène *egl-15* [230]. SUP-12 agit de manière coordonnée avec les protéines ASD-1 et FOX-1 pour permettre cet épissage en se fixant sur leur séquences *cis*-régulatrice cible respective.



**Figure 18 : Effet de mutation de séquence GTGTG sur la cassette BGAR**

Le mini-gène rapporteur bicolore BGAR de l'épissage *egl-15* est présenté schématiquement. Les rectangles jaunes représentent les exons de *egl-15*, la GFP est vert et la RFP en magenta. Les traits fins sont des introns et la position de la séquence GTGTG est indiquée par une croix rouge. Le remplacement de la séquence GTGTG par GTcTc donne la cassette BGARctc. Les mini-gènes sont sous le contrôle du promoteur *myo-3* pour une expression dans le BWM. Les animaux transgéniques portant le transgène BGAR (C) et BGARctc (D) sont analysés par microscopie à fluorescence dans le canal rouge E5A-RFP (première colonne) vert E5B-GFP (colonne centrale) et la juxtaposition des deux (dernière colonne). [Kuroyanagi et al. 2007 [230]]

Le remplacement de la séquence GUGUG initiale par la séquence GUcUc entraîne l'inclusion de l'exon 5B au lieu de l'exon 5A dans le BWM (Figure 18 D). Cela se traduit par une fluorescence verte à ce niveau au lieu d'une fluorescence rouge.

Dans le deuxième chapitre des résultats, je présenterai l'usage du système de détection quantitative dans le contexte d'une étude systématique de l'interaction entre le facteur d'épissage SUP-12 et son intron cible dans le gène *egl-15*.

### 1.6.4.2 Approche fluorescente pour valider les interactions miARN/UTR

Chez *C. elegans*, PicTar est le programme de prédilection pour identifier les cibles des 115 miARN qui y sont référencés [231]. La puissance de ce programme de prédiction réside dans le fait que la conservation des séquences entre différents nématodes est également prise en compte [231]. Afin de valider expérimentalement quelques-unes des cibles prédites par la nouvelle version de PicTar, Lall *et al.* ont développés un système de validation *in vivo* (Figure 19).

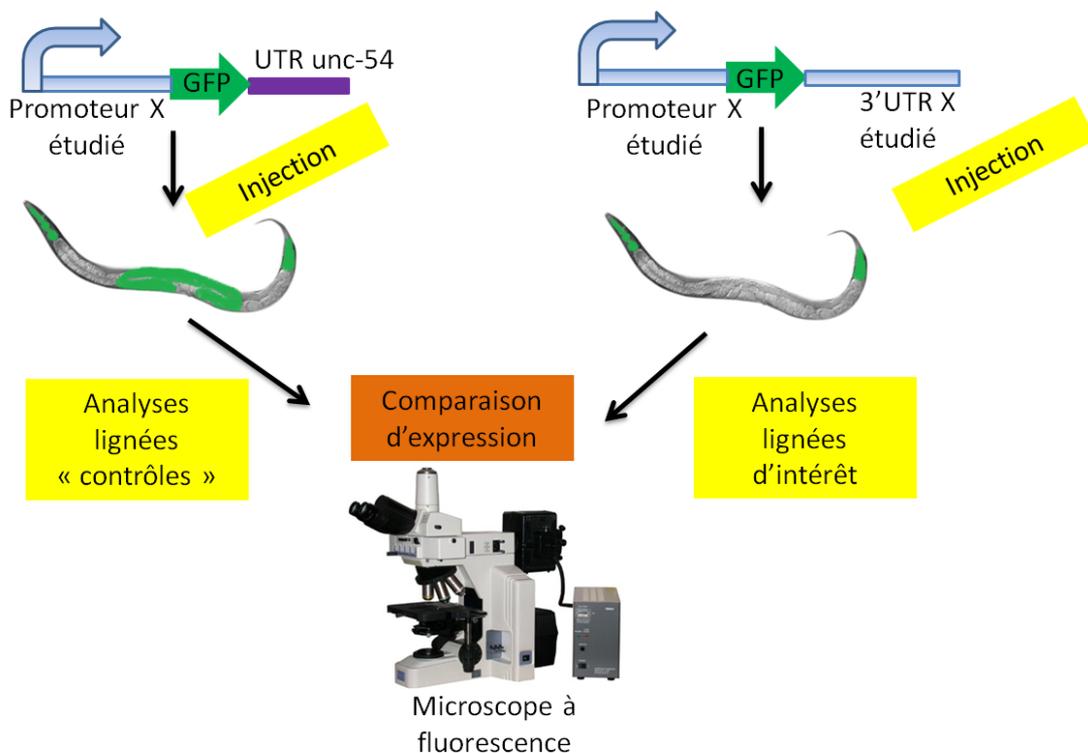
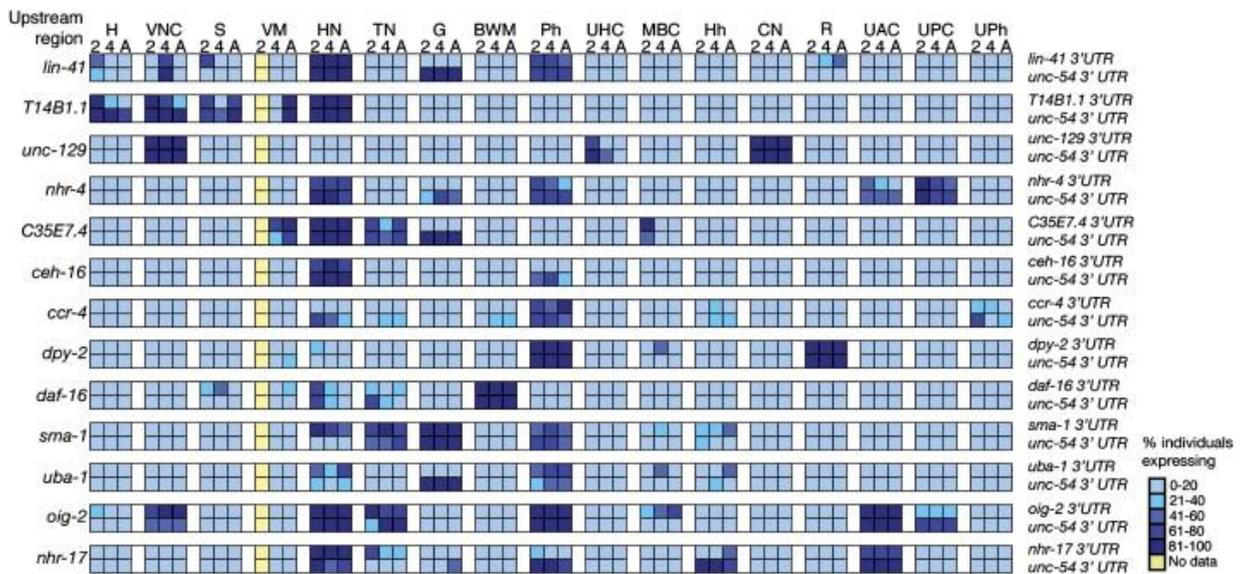


Figure 19 : Approche de validation *in vivo* de la régulation post-transcriptionnelle utilisée par Lall *et al.*

Ils se sont concentrés sur l'étude expérimentale de 13 des cibles prédites du miARN *let-7* chez *C. elegans* par PicTar. Pour chacune des cibles étudiées, l'équipe a construit plusieurs lignées indépendantes dans lesquelles la GFP est sous le contrôle du promoteur endogène du gène étudié et la région 3'UTR de celui-ci ou du gène permissif *unc-54* respectivement. Cette dernière construction sert de contrôle négatif de la régulation post-transcriptionnelle par le miARN *let-7* puisque PicTar ne prédit aucune séquence cible sur ce gène [231].

Les plasmides sont injectés dans la lignée germinale de *C. elegans* pour générer des lignées transgéniques fluorescentes. La fluorescence émise par la construction rapportrice de chacun des gènes étudiés est comparée à celle de la lignée contrôle portant l'UTR *unc-54* systématiquement. La comparaison porte sur la localisation de la fluorescence dans 17 tissus différents au cours de trois stades du développement.

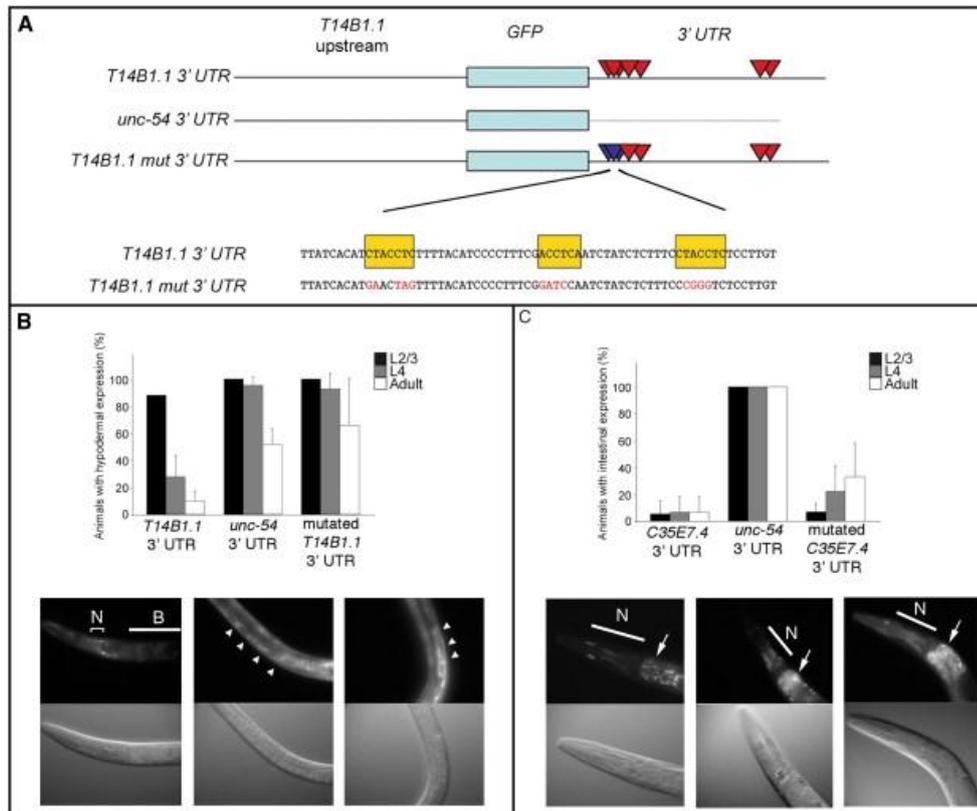
L'équipe a représenté les résultats sous forme de tableau dont chaque case représente une construction donnée à un stade particulier pour un tissu. Puis un code couleur a été établi qui rend compte du pourcentage d'animaux qui expriment la GFP dans le tissu observé (Figure 20). Une différence d'expression de la GFP est observée entre la construction rapportrice et son contrôle négatif qui est attribuée à une régulation post-transcriptionnelle induite par la région 3'UTR du gène cible.



**Figure 20 : Analyses de la régulation post-transcriptionnelle des cibles prédites de *let-7***

Des animaux transgéniques portant les deux constructions distinctes sont analysés. L'expression de la GFP est analysée chez au moins 10 vers. Le pourcentage des animaux exprimant la GFP a été reporté pour différents stades 2 (L2), 4 (L4) et A (adulte) dans différents tissus. Pour chaque gène, la ligne supérieure est le résultat obtenu avec l'UTR spécifique tandis que la ligne du bas résume le résultat pour l'UTR *unc-54*. H : Hypoderme ; VNC : corde nerveuse ventrale ; S : cellules hypodermiques latérales ; VM : muscle de la vulve ; HN : neurones de la tête ; TN : neurones de la queue ; G : intestin ; BWM : muscles du corps ; Ph : pharynx ; UHC : paire de cellules musculaires de la tête ; MBC : cellules du milieu du corps ; Hh : cellules épidermique de la tête ; CN : motoneurons ventaux ; R : cellules adjacentes au rectum ; UAC : paire spécifique des neurones de la tête ; UPC : cellules postérieures ; UPh : neurones du pharynx. [Lall et al. 2006 [231]]

Pour pouvoir déterminer si cette régulation post-transcriptionnelle est due à une répression par miARN, 3 gènes cibles, *T14B1.1*, *C35E7.4* et *oig-2* ont été sélectionnés pour que quelques-unes des séquences complémentaires cibles de *let-7* présentes dans les régions 3'UTR des constructions rapportrices soient mutées. Les mutations introduites invalident les séquences initialement reconnues par le miARN afin que *let-7* ne puisse plus s'y attacher (Figure 21 A).



**Figure 21 : la régulation de deux des trois gènes testés est dépendante de *let-7***

Constructions utilisées pour tester si la régulation du gène *T14B1.1* est dépendante du miARN *let-7*. Les sites cibles de *let-7* sont marqués avec un triangle rouge, les sites mutés avec un triangle bleu. Les mutations (texte en rouge) sont prédites pour perturber l'appariement du 3'UTR de *T14B1.1* avec le 5' de *let-7* (les séquences initiales sont surlignées en jaune) (A). pourcentage des animaux exprimant la GFP dans le tissu indiqué pour les animaux L2/L3 (noir), L4 (gris) et adulte (blanc) pour les gènes *T14B1.1* et *C35E7.4*. Des images en microscopie à fluorescence de vers présentant les différentes constructions sont présentées. Le trait avec la lettre B indique où l'expression hypodermique de la GFP est présente en cas de non régulation. La lettre N pour les neurones, et leur expression GFP est indiquée par des flèches (B et C).

Sur les 3 gènes sélectionnés, seuls deux exhibent une différence de fluorescence entre la construction contrôle et la construction mutée. Il s'agit de *T14B1.1* et *C35E7.4* (Figure 21 B et C).

Les constructions *oig-2* ne montrent aucune différence de fluorescence entre le rapporteur muté et le contrôle. L'inclusion de ce gène dans la liste des cibles potentiel de *let-7* est due à un paramétrage peu restrictif de PicTar. *In vivo*, *oig-2* ne semble pas être régulé par *let-7* et les résultats obtenus par l'équipe démontrent qu'il s'agit dans ce cas d'un faux positif. Sa présence dans la liste des cibles prédites n'avait aucune pertinence biologique et le paramétrage plus restrictif actuel de PicTar a fait que le gène *oig-2* n'est plus prédit comme cible de *let-7*.

Ce dernier exemple souligne les limites des logiciels et des algorithmes de prédiction *in-silico*. La validation expérimentale des prédictions bio-informatiques se révèle être une étape indispensable lors du développement d'un algorithme. Dans le cas des 2000 gènes régulés par miARN chez *C. elegans*, il est essentiel de disposer d'un outil de validation biologique

adaptable à une étude à grande échelle. Or, la stratégie retenue par Lall et ses collègues nécessite de générer deux séries de trois lignées indépendantes de *C. elegans* portant chacune une construction différente. Par ailleurs, le nombre de vers qui peuvent être examinés à l'œil sous microscopie constitue rapidement une limite expérimentale. Enfin le manque de contrôle interne à la lignée étudiée peut entraîner un biais d'appréciation lors de la comparaison de deux lignées indépendantes.

Dans la troisième partie des résultats je présenterai le travail que j'ai effectué pour établir un système de rapporteur bicolore similaire à ceux utilisés dans le contexte de l'épissage alternatif pour faciliter l'étude de la régulation post-transcriptionnelle médiée par les microARN.

Dans chaque chapitre je discuterai les résultats obtenus et les difficultés rencontrées.

**II. PREMIERE PARTIE : ETUDE  
DE L'ÉPISSAGE DU GENE *LET-*  
2**

## **Objectif :**

L'étude initiale menée par notre collaborateur a permis d'identifier la protéine ASD-2b comme impliquée dans la régulation de l'épissage alternatif du gène *let-2* au cours du développement. Cependant, les facteurs permettant l'expression de l'isoforme active de ASD-2 dans les stades tardifs du développement restent non-identifiés.

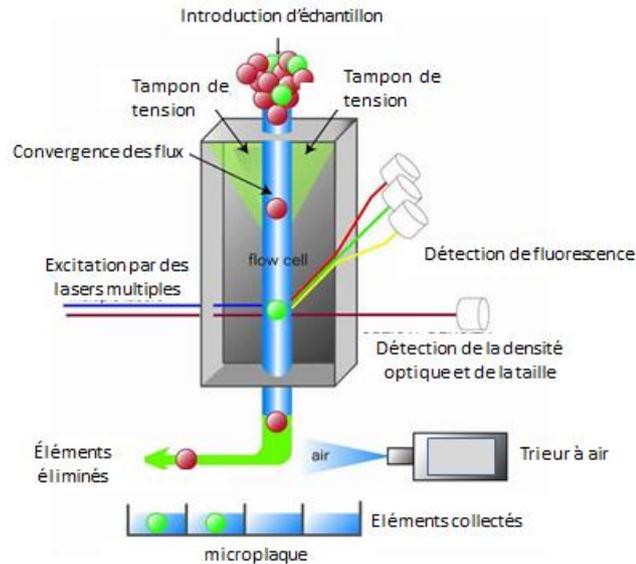
La totalité des mutations identifiées lors de cette étude sont portées sur un seul et même gène ce qui nous amène à penser que le crible a été porté à saturation dans les conditions dans lesquelles il a été mené. Cependant, un crible visuel ne permettrait pas d'identifier des phénotypes plus modérés avec des profils plus discrets.

Nous avons donc décidé de compléter cette étude par une approche à haut débit. Nous avons pour cela utilisé la puissance et la rapidité de traitement de la cytométrie en flux adaptée pour des organismes de petite taille COPAS Biosorter (Complex object parametric analysis and sorting). Cette approche devrait permettre une analyse plus quantitative et une discrimination plus subtile des profils de fluorescence.

## **2.1 APPROCHE PAR GENETIQUE CLASSIQUE**

### **2.1.1 Analyse phénotypique par COPAS profiler**

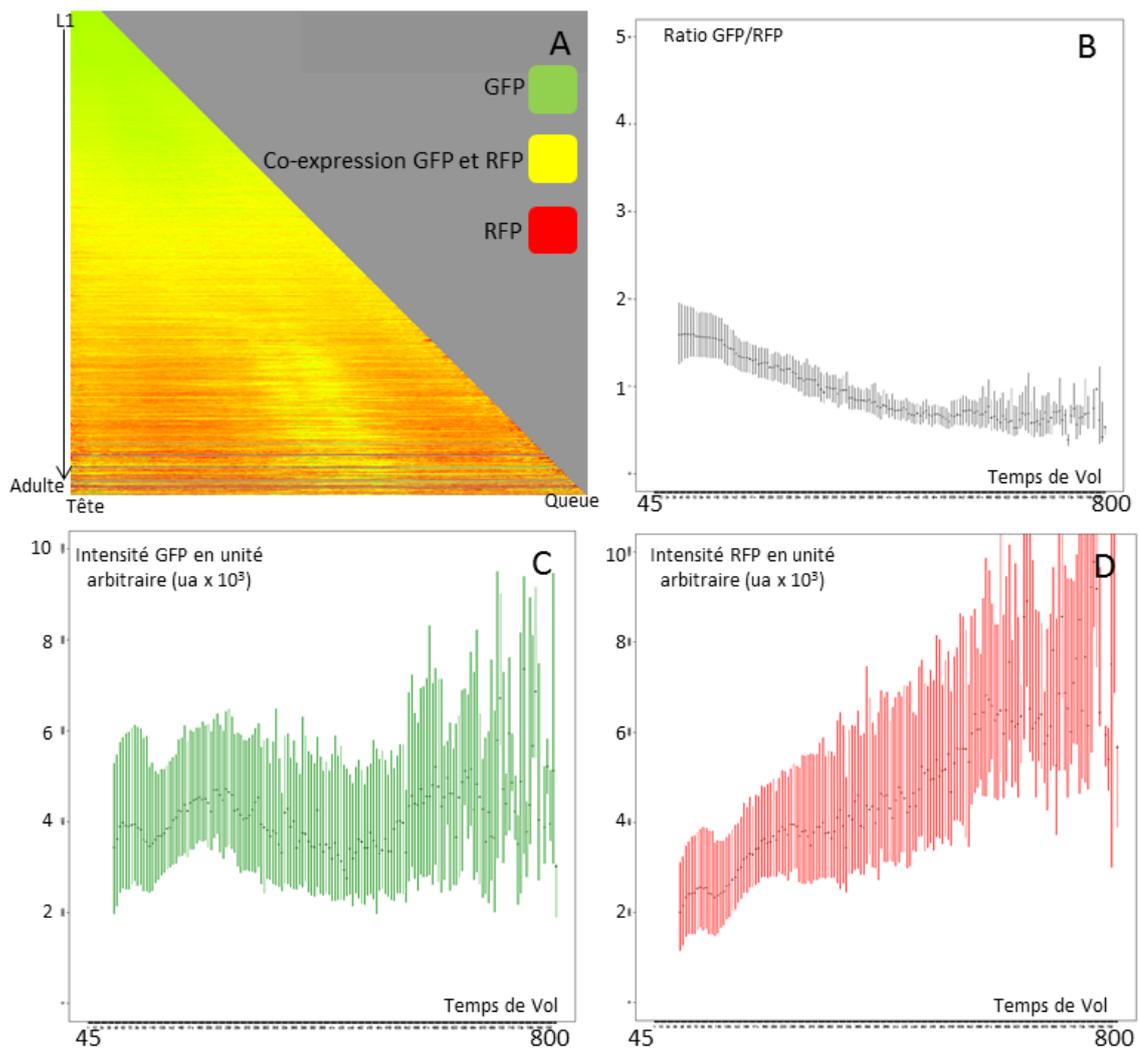
Le COPAS profiler est un instrument de cytométrie en flux adapté pour analyser et isoler des organismes vivants de petite taille comme *C. elegans* sur la base de leur densité optique, leur taille et leur fluorescence [232] (Figure 22).



**Figure 22 : représentation schématique du principe d'analyse et de tri du COPAS**

Les nématodes à analyser sont suspendus dans une solution (eau, S-basal, NGM) puis sont disposés dans la coupe de collecte de l'instrument. La suspension de *C. elegans* sous pression progresse vers la chambre de détection. Une solution tampon permet de maintenir le nématode en extension. Les fluorescences dans le rouge, le vert et le jaune sont analysées grâce à un faisceau laser qui excite les nématodes dans la chambre de détection. En aval de cette analyse, un flux d'air constant pousse les nématodes en direction d'une cuve d'élimination. L'interruption de ce flux d'air, permet d'isoler les animaux en fonction de paramètres déterminés au préalable par l'utilisateur (Figure 22) [233-235].

Les données récoltées sont traitées par un programme développé au sein de notre laboratoire de façon à obtenir une représentation spatio-temporelle des données de fluorescence récoltées.



**Figure 23 : caractéristiques de fluorescence de la souche KH1401 analysée au COPAS**

Les données récoltées après analyse de la souche KH1401 au COPAS sont représentées sous forme de chronogramme bi-canal où chaque ligne représente la fluorescence moyenne des animaux d'une taille donnée pour l'expression de la GFP (vert), de la RFP (rouge) et leur co-expression spatio-temporelle (jaune/orange) depuis le premier stade larvaire L1 jusqu'au stade adulte. Le chronogramme est orienté de sorte que la tête se trouve à gauche et la queue à droite (A). Le ratio moyen de l'expression GFP/mCherry pour les animaux d'une taille donnée est représenté en fonction du temps de vol (TOF) qui est corrélé à la taille (B). L'intensité de l'expression moyenne en unité arbitraire (ua) de la population d'une taille donnée en fonction du temps de vol est représentée par des barres vertes pour la GFP (C) et rouges pour la RFP (D).

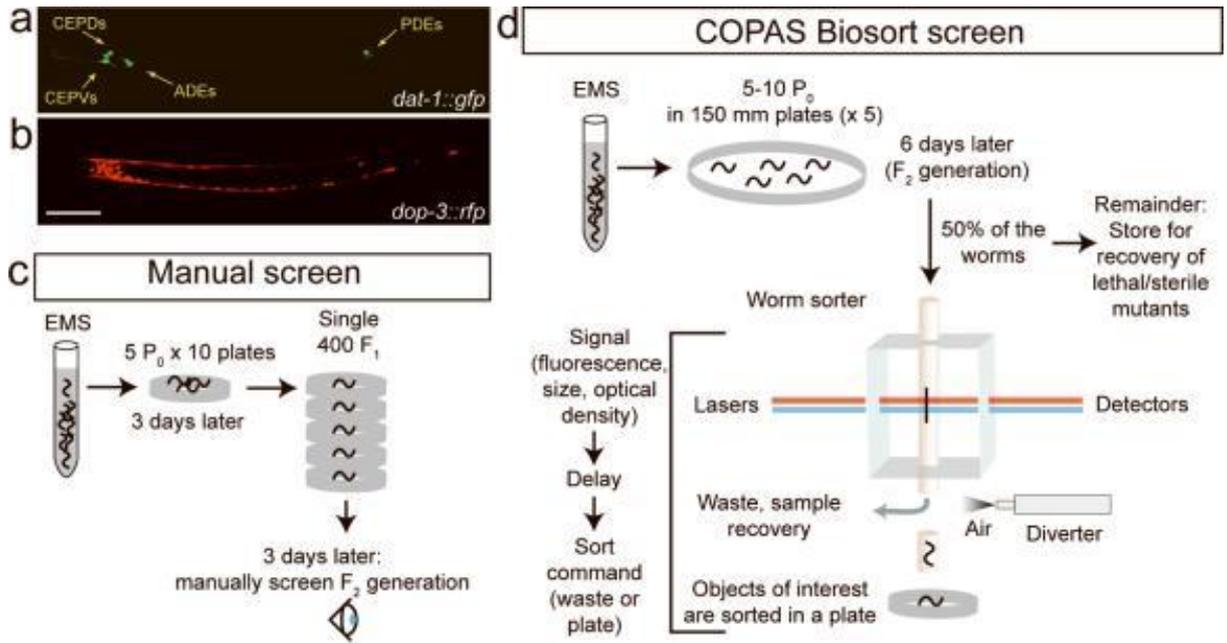
La souche KH1401 a été analysée avec le COPAS afin de déterminer son profil de fluorescence dans le rouge et dans le vert au cours du développement.

Les données collectées par l'analyse de la souche KH1401 sont représentées sous forme de chronogramme [236] (Figure 23 A). Chaque ligne de ce chronogramme représente la moyenne des mesures de fluorescences obtenues le long de l'axe antéro-postérieur des vers de la taille correspondante. Dans cette version du chronogramme, les expressions de la GFP et de la mCherry sont représentées sur la même image en vert et en rouge respectivement. La co-expression de ces deux protéines est traduite par une couleur jaune sur le chronogramme. L'intensité de fluorescence de

la GFP (Figure 23 C) et de la mCherry (Figure 23 D) en fonction de la taille permet de visualiser l'évolution de l'intensité de chacune des deux fluorescences au cours du développement. Le ratio entre les deux expressions (GFP/mCherry) en fonction de la taille est également reporté sur le graphique (Figure 23 B). La représentation des données sous cette forme permet de visualiser le phénomène d'exclusion mutuelle des exons 9 et 10 des mini-gènes rapporteurs de *let-2* au cours du développement. Nous observons la permutation de l'expression de fluorescence du vert aux stades précoces vers le rouge aux stades tardifs conformément aux observations de microscopie (Figure 15 C) [60].

### **2.1.2 Automatisation du crible par COPAS biosorter**

Une étude réalisée en 2008 sur des lignées de *C. elegans* bicolores a montré la possibilité de tirer avantage du COPAS pour automatiser un crible de mutagenèse à l'échelle génomique et a permis d'isoler des mutants en se basant sur la modulation de l'expression de leur fluorescence (Figure 24) [237]. Doitsidou et ses collègues s'intéressent à l'isolation de mutants affectés dans la spécification des neurones dopaminergiques. Pour ce faire, ils ont créé une lignée de nématode où la protéine fluorescente GFP est sous le contrôle du promoteur du gène *dat-1* spécifique des neurones dopaminergiques (Figure 24 a). Comme il n'existe que 8 cellules dopaminergiques sur le millier de cellules que compte *C. elegans*, l'utilisation de *dat-1::gfp* représente un bon rapporteur pour tester la sensibilité du système. Cependant, pour détecter une modulation d'expression fine de fluorescence à l'échelle d'un individu, un contrôle endogène a été nécessaire. Pour cela un transgène exprimant la RFP (*dop-3::rfp*) a été introduit dans le fond génétique de la souche à cribler afin de servir de référence d'intensité de fluorescence (Figure 24 b).

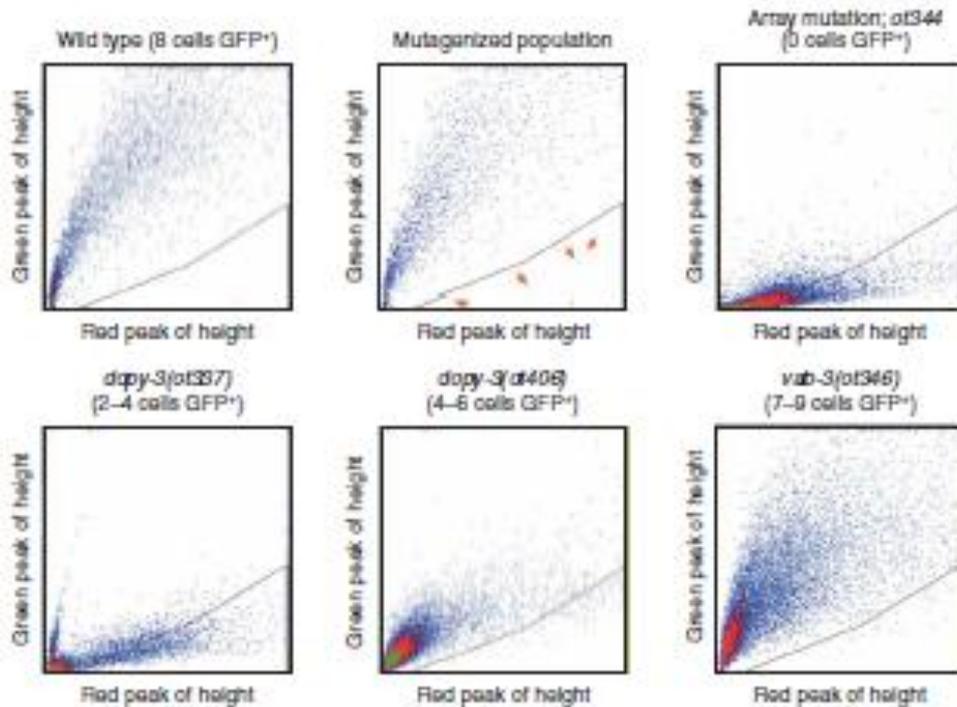


**Figure 24 : Crible de mutants affectés dans la spécification des neurones dopaminergiques**

Image en microscopie à fluorescence d'un vers dont l'expression de la GFP est sous le contrôle du promoteur *dat-1* (a) et de la RFP sous le contrôle du promoteur *dop-3* (b). La représentation schématique des étapes du crible manuel (c). La représentation schématique des étapes du crible automatisé (d). [Doitsiou et al. 2008 [237]]

Lors du crible manuel, il a été nécessaire de disposer au départ 5 vers P0 par boîte de culture sur une dizaine de boîtes, d'individualiser 400 vers de la progéniture F1 sur une boîte unique puis de cribler leur progéniture sous microscopie à fluorescence. Le crible manuel a nécessité 100 jours de travail à plein temps pour une personne rompue à ce genre d'exercice et a permis d'isoler 10 allèles mutants pour 11.000 génomes diploïdes criblés. Soit une fréquence de 1 allèle tous les 10 jours.

En comparaison, le crible automatisé n'a nécessité que 25 jours pendant lesquels le processus est majoritairement autonome et ne nécessite que la vérification du bon fonctionnement de la machine par l'expérimentateur. Dans ce crible, 5 à 10 vers P0 ayants subis la mutagenèse sont disposés sur 10 grandes boîtes de culture pendant 6 jours jusqu'à l'obtention de la génération F2 qui est criblée au COPAS en se focalisant sur l'isolation de mutants ayants un ratio vert/rouge réduit. 80.000 génomes diploïdes ont été criblés et 17 mutants dont certains sont affecté dans l'expression d'un seul neurone ont pu être isolés, soit une moyenne d'un mutant tous les 1,5 jours.



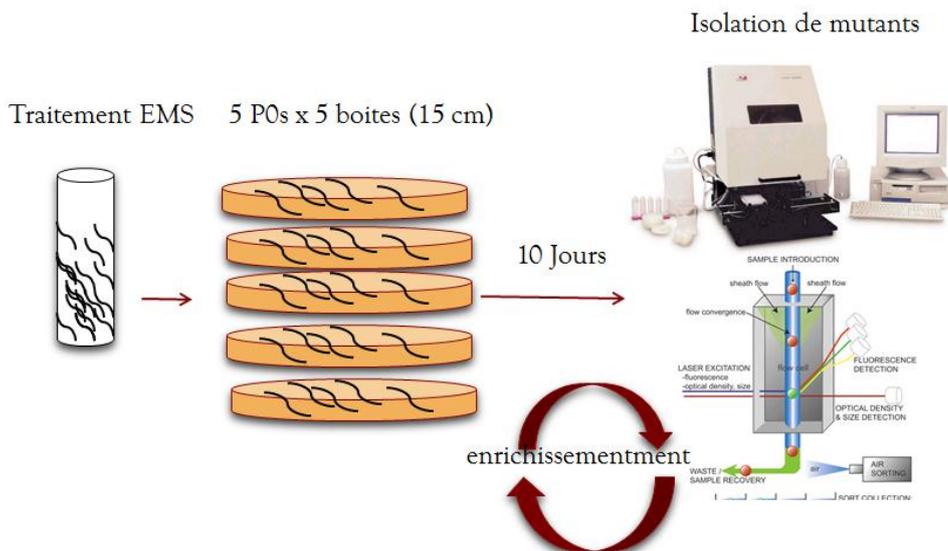
**Figure 25 : Intensité de fluorescence relative entre le rouge et le vert pour les animaux criblés**

La population de vers est représentée en nuage de point où chaque point correspond à un animal unique. La population n'ayant pas subi de mutagenèse est représentée, ainsi que celle ayant subi un traitement EMS, et de populations homogènes de différents mutants isolés. Les flèches rouges indiquent des événements qui se situent dans la fenêtre de sélection (coté inférieur droit e chaque fenêtre) [Doitsidou et al. 2008 [237]]

Un gain de temps considérable en comparaison avec un crible manuel a été possible grâce à l'automatisation de l'approche. 1h d'effort intense est nécessaire pour cribler manuellement environ 60 génomes. Dans le même laps de temps, 50.000 vers (soit environ 2000 génomes par heure) peuvent être criblés en utilisant le COPAS et aboutir efficacement à l'isolation de mutants affectés dans la voie de spécification des neurones dopaminergiques (Figure 25).

Nous avons décidé d'adopter cette méthode pour réaliser une mutagenèse de la lignée bicolore KH1401 afin d'identifier les autres facteurs participants à la régulation de l'épissage alternatif de *let-2*.

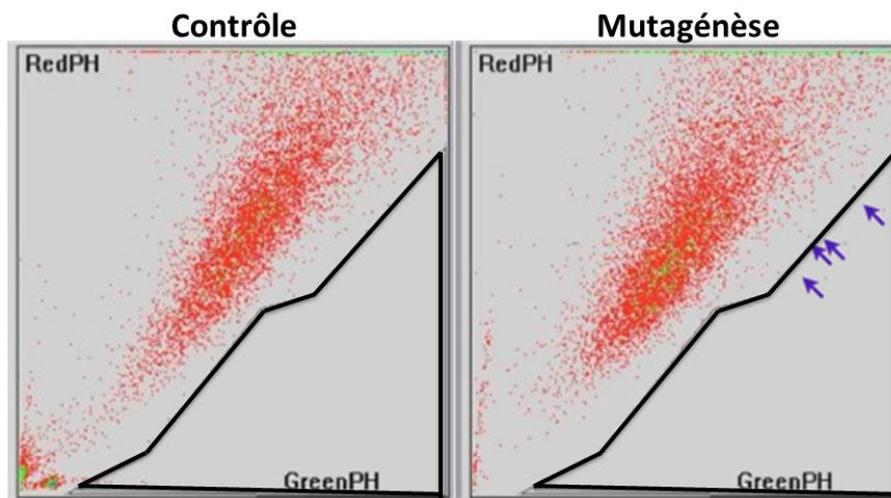
Après exposition à l'EMS pendant 4h des vers KH1401 au stade L4, 5 vers sont disposés sur des boîtes de culture de 15 cm de diamètre. Pendant 10 jours, les vers ayant subi la mutagenèse se développent et engendrent une descendance qui fera l'objet du crible pour les mutants déficients pour l'épissage alternatif de *let-2* (Figure 26).



**Figure 26 : Crible génomique des mutants affectés dans l'épissage de *let-2***

Représentation schématique des étapes de mutagenèse et de sélection des mutants en utilisant le COPAS pour automatiser le crible.

Nous avons procédé au crible de 200 boîtes issues de la mutagenèse EMS. Au début de chaque analyse, la souche KH1401 de référence est analysée au COPAS afin de déterminer la fenêtre de sélection des événements qui seront isolés par la suite.



**Figure 27 : Détermination de la fenêtre de sélection des mutants**

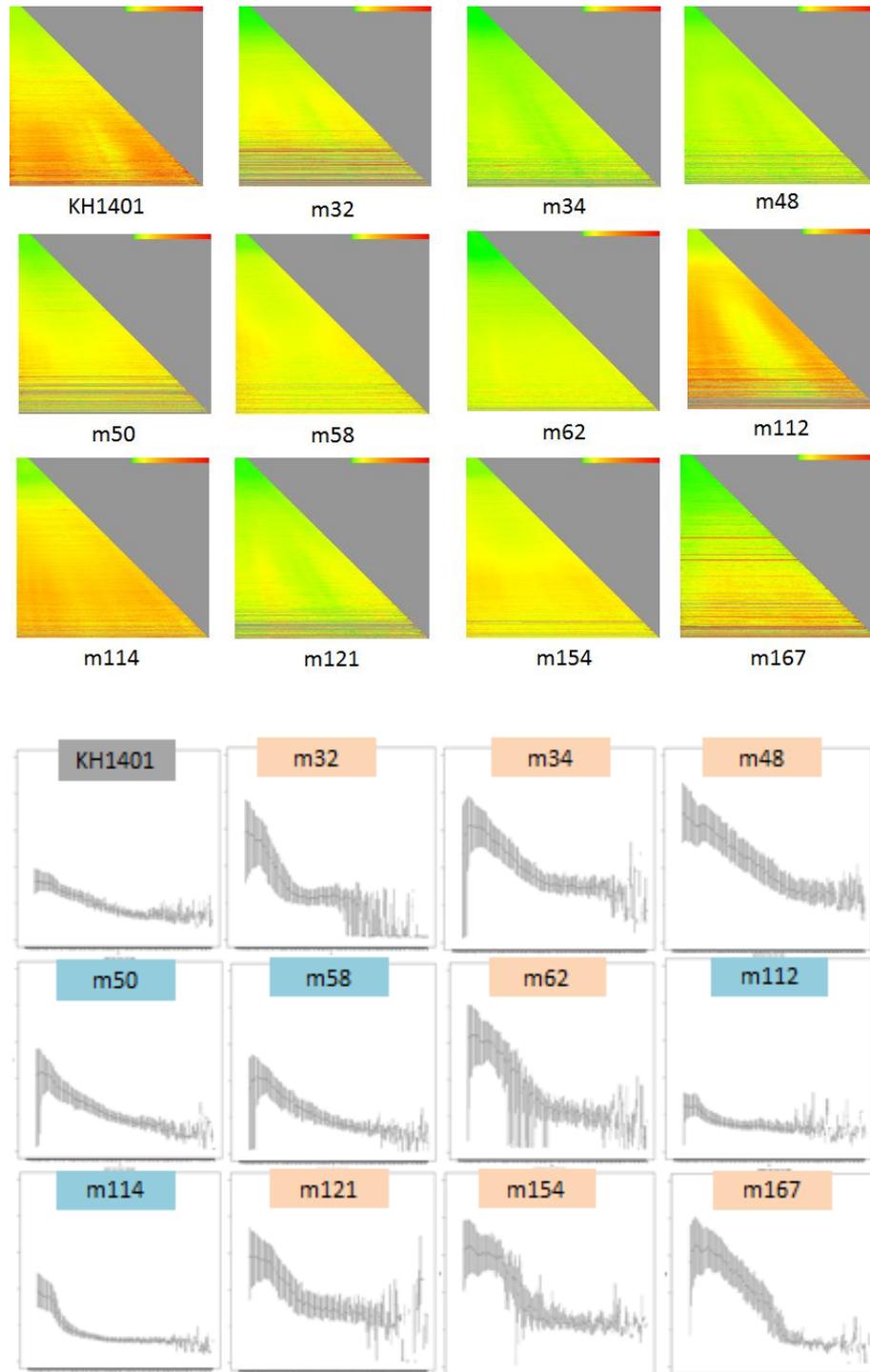
La population de référence KH1401 non traitée à l'EMS (à gauche) et traitées (à droite) sont représentées sous forme de nuage de point où chaque point correspond l'intensité relative du rouge (RedPH ; red peak height) par rapport au vert (GreenPH ; green peak height) pour un animal unique. La fenêtre de sélection établie est représentée par des traits épais. Les flèches mauves montrent des événements sélectionnés par le COPAS.

A la suite de l'analyse de la souche rapportrice KH1401, nous traçons une fenêtre qui exclut la totalité des événements présents lors de cette analyse (Figure 27). Cela permet de minimiser au maximum l'isolation de faux-positifs. Les 200 boîtes analysées contenaient chacune 5 vers diploïdes

P0 qui engendrent une progéniture de ~300 vers, Nous avons donc criblé ainsi  $6 \times 10^5$  génomes pour la modulation de l'expression de fluorescence. Lors des 200 analyses réalisées en moyenne 150000 vers sont analysés par boîte, le COPAS a isolé à chaque traitement une moyenne de 3 évènements qui se situent dans la fenêtre prédéterminée. Les nématodes ainsi isolés sur une nouvelle boîte de culture prolifèrent en vue d'une nouvelle analyse afin de vérifier que l'altération du profil d'expression est transmise à sa descendance.

Sur les 200 boîtes analysées, 163 étaient des faux positifs lors de la seconde analyse après environ 7 à 10 jours le temps d'avoir une population dense sur une boîte de culture de 9 cm (soit environ 10000 à 15000 vers). 26 boîtes des 37 retenues présentaient également un profil sauvage lors d'une troisième analyse et contenaient une population semblable au profil de référence de la souche KH1401. Les 11 boîtes restantes présentaient des phénotypes mutants au-delà d'une quatrième analyse. Un vers unique a été isolé de chacune de ces 11 boîtes et disposé sur une nouvelle boîte où le profil de fluorescence de sa descendance a été à son tour analysée au COPAS. Ainsi nous avons isolés 11 lignées mutantes présentant une déficience d'expression du rouge dans les stades tardifs du développement. Il s'agit, par ordre chronologique d'isolation, des mutants m32, m34, m48, m50, m58, m62, m112, m114, m121, m154 et m167.

Ces mutants montrent un phénotype d'expression de fluorescence différent de celui de la souche rapportrice KH1401 comme le démontrent les chronogrammes respectifs et leur ratio d'expression vert/rouge (Figure 28).



**Figure 28 : Caractéristiques de fluorescence des mutants affectés pour l'épissage de *let-2*.**

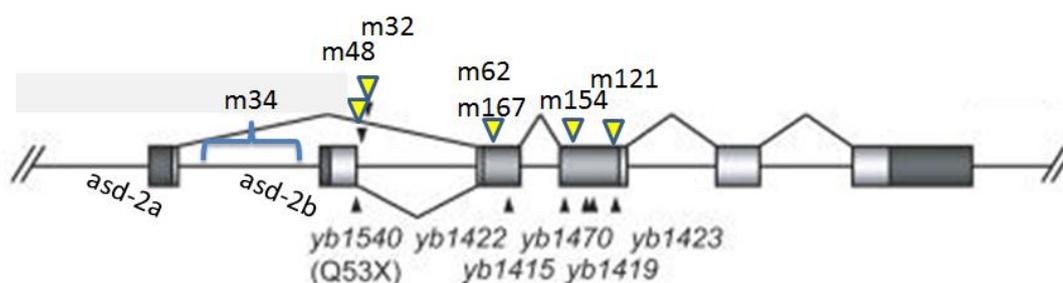
La population de référence KH1401 et celles mutées sont analysées au COPAS. Les chronogrammes bicolores correspondant à chacune des populations sont présentés dans la figure du haut. Les Ratios de l'intensité des fluorescences vert/rouge en fonction de la taille sont présentés dans la figure du bas. Les abscisses et les ordonnées sont identiques à ceux de la Figure 23. Les mutants *asd-2* sont indiqués en orange. Les mutants sauvages pour *asd-2* sont indiqués en bleu.

### 2.1.3 *asd-2* et phénotypes mutants

Comme l'étude préalable menée par notre collaborateur a abouti à l'isolation exclusivement de 8 mutants du gène *asd-2*, Nous avons procédé au séquençage de ce gène pour chacune des 11 lignées isolées lors de notre crible automatique. Nous nous attendions à retrouver des mutants de ce même gène ce qui validerait notre approche.

7 des 11 mutants isolés portent une mutation dans le gène *asd-2*. D'une manière intéressante, nous retrouvons dans notre crible deux mutants précédemment isolés lors du crible manuel réalisé par nos collaborateurs. Il s'agit respectivement des mutants m48 et m32 qui exhibent les mêmes mutations que yb1542 et yb1421 [60]. Dans le mutant m48, la séquence consensus gtag de reconnaissance de site accepteur/donneur d'épissage du premier intron de l'isoforme *asd-2b* a été remplacée par atgag. Le mutant m32 quant à lui a été muté pour cette même séquence et porte la mutation gtgaa.

Les mutants m62 et m167 portent la même mutation au niveau du deuxième exon d'*asd-2b* où la proline en position 87 a été remplacée par une Histidine (P187H). Les mutants m121 et m154 sont mutés dans l'exon 3. L'acide glutamique en position 206 a été remplacé par une glycine chez le mutant 121 (E206G) tandis que la Méthionine originalement en position 170 de la protéine ASD-2b a été remplacée par une Thréonine (M170T). Le mutant m34 quant à lui exhibe une délétion de 300pb au niveau du promoteur de l'isoforme active du gène *asd-2* (Figure 29).



**Figure 29 : Localisation des mutations portées par les mutants *asd-2***

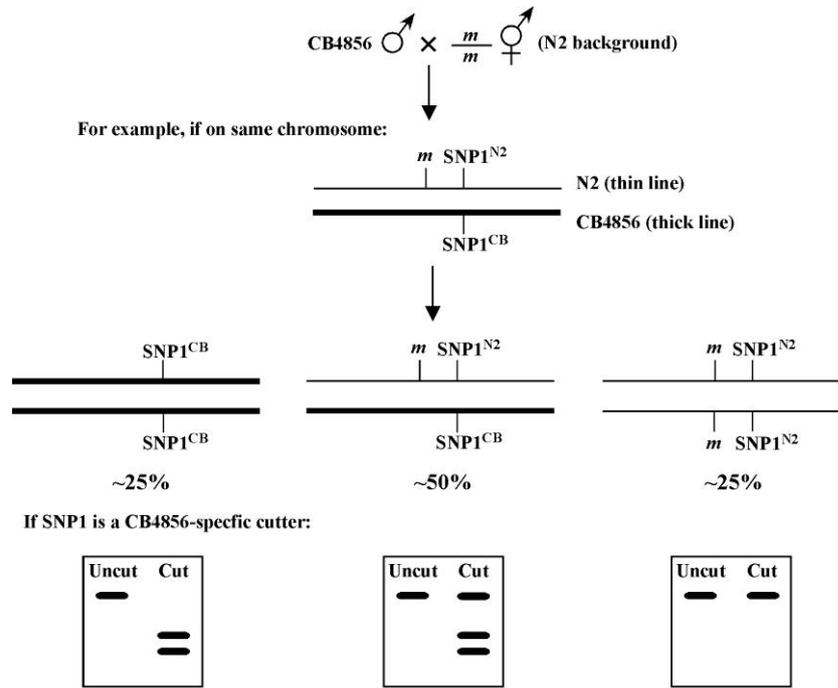
Les mutations portées par les mutants identifiées lors de notre crible sont localisées sur le gène *asd-2* et sont indiquées par une flèche jaune au-dessus de la représentation schématique du gène. Les mutants identifiés par le crible de notre collaborateur H. Kuroyanagi sont localisés par une flèche noire en dessous de la représentation schématique du gène.

### **2.1.4 Les mutants sauvages pour le gène *asd-2***

Les 4 autres mutants m50, m58, m112 et m114 sont sauvages pour le gène candidat *asd-2*. Ce résultat est encourageant puisque lors de l'étude initiale menée par notre collaborateur, la totalité des mutants isolés étaient des allèles du même gène. Or, l'automatisation du crible et la sensibilité du COPAS ont permis d'isoler 4 mutants exhibant un phénotype intéressant bien que sauvages pour l'unique gène impliqué dans l'épissage de *let-2* connu à ce jour.

L'approche classique utilisée pour localiser la région génomique responsable du phénotype consiste à cartographier le génome pour les polymorphismes de nucléotides uniques (SNP).

La cartographie des SNP chez *C. elegans* exploite la haute densité de SNP entre la souche N2 de Bristol et la souche d'Hawaï CB4856 en croisant les deux souches entre elles dont l'une (généralement N2) porte la mutation d'intérêt. Un SNP peut être observé toute les 840 pb entre ces deux souches cousines [238, 239]. Pour localiser une mutation, il est nécessaire de croiser le génome mutant homozygote issu de la souche N2 à un mâle de la souche CB4856 pour obtenir une population hétérozygote. La descendance F1 mutante de celle-ci est isolée et individualisée sur des boîtes de culture séparées. Dans cette population recombinante, les polymorphismes sont distribués de manière aléatoire le long des chromosomes issus de la souche Hawaïenne et celle de Bristol excepté pour les polymorphismes qui entourent la région génomique recherchée créant ainsi un déséquilibre de liaison. Dans ce cas on observe une prédominance des allèles de N2 qui sont proches de la mutation. Une fois ces étapes de croisements et d'isolations de mutants réalisées, différentes approches permettent d'identifier la nature moléculaire des polymorphismes en utilisant une combinaison de PCR et de profil de restriction de digestion (Figure 30) [240] ou en utilisant des sondes d'hybridations [241].



**Figure 30 : Exemple de cartographie d'une mutation liée à un SNP**

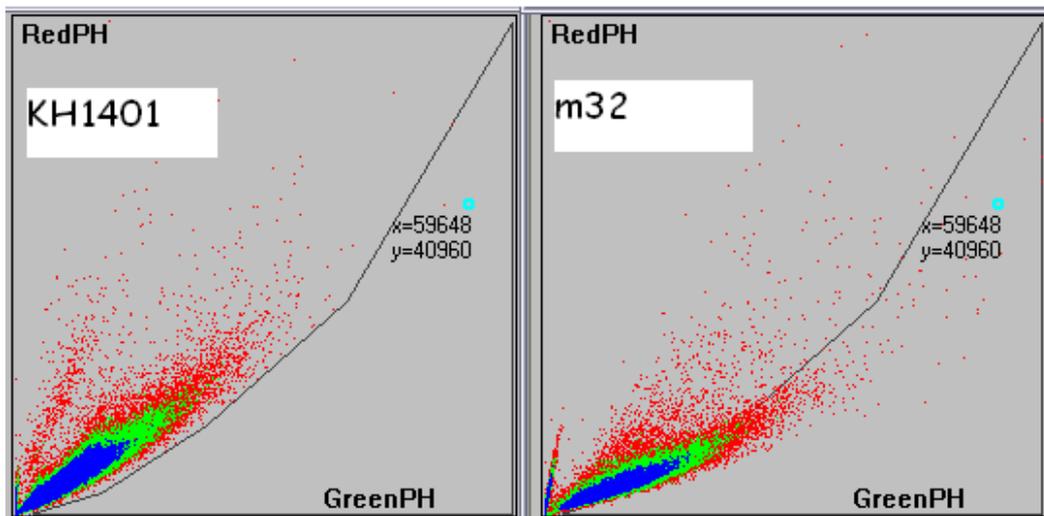
Cartographie d'une mutation portée par la souche N2 (trait fin) liée au SNP1 (single nucleotide polymorphism 1) en la croisant avec la souche CB4856 (Trait épais). La descendance F2 suit une ségrégation mendélienne pour ce polymorphisme. Un exemple de détection par profil de restriction est schématisé pour chacun des génotypes de la descendance F2. Dans cet exemple, le SNP1 permet une coupure spécifique de la souche CB4856 et non de N2. Les descendants F2 hétérozygotes présenteront systématiquement un profil de restriction hétérogène lors de la coupure du SNP1 si la mutation y est associée. [www.wormbook.org]

Bien que la cartographie des SNP permette de localiser finement la région génomique d'intérêt à un intervalle de 200 à 300 kb, il reste encore à identifier le gène responsable dans cet intervalle. Il faut donc procéder au séquençage de cette région afin d'identifier la mutation incriminée.

La réalisation d'un croisement lors de la cartographie génétique nécessite d'identifier les animaux mutants à l'échelle individuelle. Un seul animal, de la souche N2 portant la mutation, est croisé à un seul animal de la souche Hawaï de phénotype sauvage. La capacité de discriminer à l'échelle individuelle entre les deux phénotypes est donc une condition *sine qua non* pour une telle entreprise. Il en va de même lors des étapes d'individualisation de la descendance issue de ce croisement.

Dans notre étude, nous nous heurtons à deux obstacles majeurs à la réalisation de la cartographie. Premièrement, nous ne disposons pas d'une souche Hawaï portant les mini gènes rapporteurs de l'épissage de *let-2*. Ceci nous oblige à identifier la descendance portant le rapporteur et la mutation responsable du phénotype et ayant été croisées à la souche Hawaïenne. Par ailleurs, *C. elegans* est hermaphrodite et capable d'autofécondation. La discrimination entre la descendance ayant effectivement été croisée et celle provenant d'autofécondation n'est possible que lors des étapes ultérieures indiquant l'absence de polymorphisme dans le second cas de figure. Deuxièmement, les profils de fluorescence des mutants que nous avons

isolés se distinguent de celui du sauvage à l'échelle de la population entière mais la lignée présente un spectre très large de profil de fluorescence. Plusieurs milliers d'individus mutants ou sauvage exhibent le même profil de fluorescence relative vert/rouge ce qui rend impossible la discrimination à l'échelle individuelle de phénotypes pourtant distincts à l'échelle de la population (Figure 31).



**Figure 31 : Intensités relatives de la fluorescence rouge par rapport à la verte pour la population KH1401 et m32**

Les populations analysées sont représentées en nuage de point (l'augmentation de la densité est représentée par un passage de la couleur rouge (individu unique) au vert puis au bleu. Chaque point représente le ratio de la fluorescence rouge (RedPH ; red peak height) sur la fluorescence verte (GreenPH ; Green peak height) pour un animal donné. La fenêtre de sélection a été tracée en excluant la totalité des vers de la population de référence. Le pourcentage d'événements inclus dans cette fenêtre est comptabilisé par le logiciel du COPAS.

La fenêtre de sélection exclut la totalité des vers KH1401 de référence. Par exemple dans le cas du mutant m32 bien que la distribution générale de la population soit visiblement différente seuls 6,4 % des 52963 vers de la population se situent hors de la fenêtre d'expression du sauvage. Nous avons donc décidé de ne pas procéder à une cartographie par croisement.

Pour identifier les mutations portées par les 4 derniers mutants, nous avons opté pour le séquençage du génome entier de chacune des lignées afin de s'affranchir de tous les obstacles qui se dressent en procédant à une cartographie classique [242, 243]. Le séquençage du génome entier offre également un gain de temps considérable par rapport à l'approche classique.

### **2.1.5 Séquençage du génome entier**

Au vu des résultats de la mutagenèse obtenus par notre collaborateur lors de son crible manuel où la totalité des mutants obtenus étaient mutés dans le gène *asd-2*, et que lors du crible automatisé que nous avons réalisés, 7 des mutants isolés sur 11 étaient mutés sur le même gène, nous avons émis l'hypothèse que les 4 mutants restants seraient mutés sur un unique second gène impliqué également dans l'épissage de *let-2* et non identifié à ce

jour. Cette hypothèse nous apparaît plus probable que de prédire qu'il s'agit de 4 nouveaux gènes indépendants.

Nous avons procédé au rétrocroisement de chacun des mutants avec la souche parentale KH1401 afin de restreindre le nombre de mutations dues au traitement chimique qui ne sont pas responsables de la modulation de fluorescence. Comme précédemment, nous avons utilisé le COPAS Biosorter pour isoler un individu unique sur une boîte après avoir déterminé une fenêtre d'exclusion des vers KH1401.

Nous avons procédé au séquençage du génome entier du mutant m58 non retro croisé, des mutants m50 et m112 croisés une fois et du mutant m114 croisé deux fois. Nous avons également procédé au séquençage du mutant 114 initial non croisé afin de vérifier la diminution des taux de mutations lors des croisements. La souche parentale KH1401 a également été séquencée. Les mutations présentes sur la souche KH1401 parentale et communes à nos mutants ne seront pas prises en compte dans notre étude puisque elles sont indépendantes de notre mutagénèse. Les données brutes reçues ont été traitées par MAQGene [244].

Pour chaque souche nous avons comptabilisé le nombre de mutations et la nature de celles-ci ainsi que leur localisation dans le génome. Nous avons rapporté ces résultats dans le tableau suivant (Tableau 1).

	mutations	Codants	Non géniques	ARNnc	5'UTR	3'UTR	Splice donneur	Zones non couvertes	Taille moyenne des zones non couvertes en pb
M50 BC1	1164	162	929	9	12	13	1	94	84
M58 BC0	873	154	667	8	7	15	1	674	77
M112 BC1	354	13	307	2	3	2	0	91	76
M114 BC2	657	77	527	3	3	8	0	48	68
M114 BC0	3286	343	2416	10	11	39	1	246	140

**Tableau 1 : Natures et nombres des mutations des souches séquencées**

L'analyse du résultat de croisement du mutant m114 croisé deux fois en comparaison avec les mutations présentes dans la même souche non croisée nous informe que le nombre de mutations a été diminué d'un facteur 5. Le croisement avec la souche parentale est donc utile pour permettre de débarrasser la souche mutante des mutations non liées au phénotype étudié. Cependant, des études récentes montrent que malgré le fait que le croisement contribue à diminuer le nombre de mutations induites lors de la mutagénèse un grand nombre de mutations sont toujours observées même dans les lignées rétro-croisées dû à la variabilité génétique de la souche



### 2.1.5.2 Les mutations spécifiques dans les régions codantes

Hormis ces mutations en amont du gène *F38C2.1* pour lequel aucune fonction n'est connue, le séquençage n'a pas permis d'étayer notre hypothèse de départ à la recherche d'un gène unique. Aussi, nous avons décidé de traiter les mutants indépendamment en s'intéressant aux mutations survenues dans les régions codantes du génome (Tableau 2).

	M50	M58	M112	M114
Mutations dans les régions codantes.	162	154	13	77
Mutations silencieuses	54	42	4	21
Changement propriété A.A.	70	68	3	36
Fonction connue ou prédite sur <a href="http://www.wormbase.org">www.wormbase.org</a>	29	30	0	18

**Tableau 2 : Nombres des mutations dans les régions codantes portées par chacune des souches**

Parmi les gènes mutés dans chacune des souches, nous nous sommes focalisés sur les mutations qui induisent le changement des propriétés physico-chimiques de l'acide aminé original. Nous avons ensuite interrogé la base de données [www.wormbase.org](http://www.wormbase.org) afin de connaître la fonction des gènes mutés à la recherche de gènes précédemment étudiés ou prédits impliqués dans la liaison à l'ARN ou dans le phénomène d'épissage des exons. Aucun des gènes mutés dont la fonction est reportée sur wormbase ne codait pour une protéine appartenant à ces catégories.

### 2.1.6 Discussion

Nous ne sommes pas parvenus à identifier les causes moléculaires du phénotype étudié lors de cette étude pour les mutants m50, m58, m112 et m114 par génétique classique. La méthode employée n'est cependant pas à remettre en cause, nous avons confirmé notre capacité à effectuer des cribles de mutagenèses à haut débit assisté par le COPAS et répliqué les résultats obtenus par nos collaborateurs japonais. Nous avons confirmé l'implication du gène *asd-2* dans ce processus, 7 de nos 11 mutants correspondent à 6 mutations indépendantes (dont deux déjà identifiées lors de l'étude précédente). Cependant, nous n'avons pas pu identifier les lésions

moléculaires responsables de la modulation de fluorescence chez les 4 mutants restants.

Il se peut que dans ces cas, il s'agisse de faux-positifs ou que la mutation ait été perdue au cours des backcross. En effet la figure 28 montre clairement que les ratios d'expression des mutants m112 et m114 sont très proches de celles de la lignée d'origine KH1401.

Il est possible également que la lésion soit portée par l'une des multiples copies des minigènes rapporteurs diminuant légèrement l'expression de la fluorescence. Lors du séquençage du génome entier la séquence de la souche N2 répertoriée dans wormbase est utilisée comme séquence de référence. La séquence du concatémère qui consiste en plusieurs répétitions des minigènes présents dans la souche KH1401 et dans les souches mutées n'est pas incluse dans les résultats de séquençage. Nous n'avons donc pas pu vérifier cette hypothèse. On ne peut pas non plus vérifier cette hypothèse par séquençage classique.

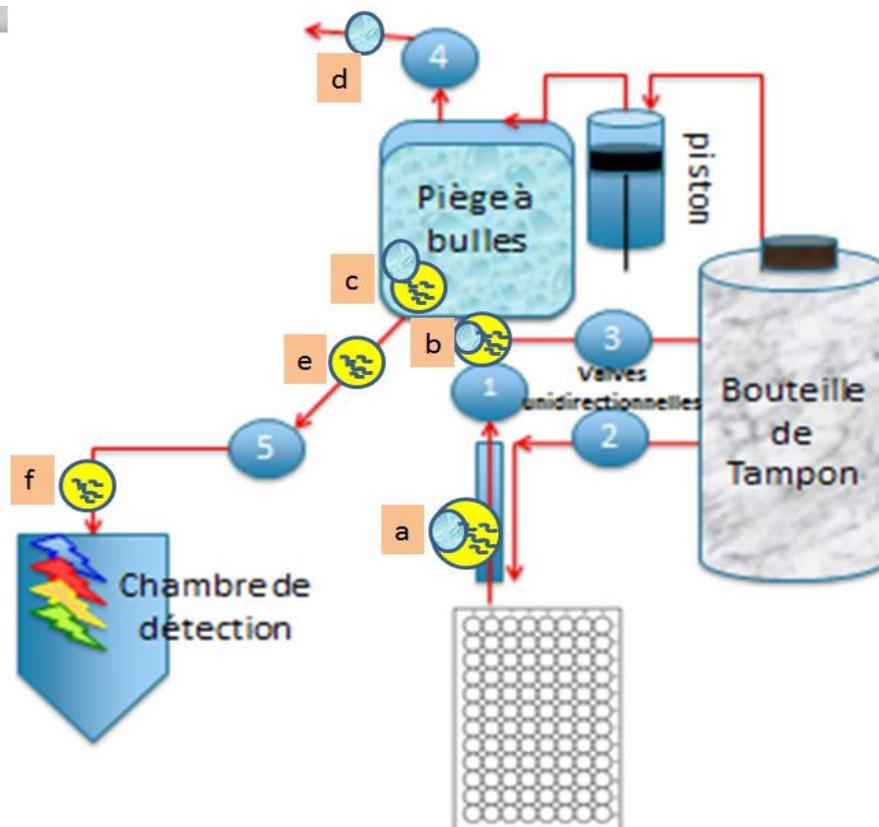
Nous espérons lors de cette étude pouvoir identifier soit une autre protéine de liaison à l'ARN impliquée dans l'épissage du gène *let-2*, soit pouvoir identifier le facteur responsable de l'expression de l'isoforme *asd-2b* dans les stades tardifs du développement dans le BWM. Il est cependant probable que la protéine ASD-2 soit la seule impliquée dans l'inclusion alternative de l'exon 10 du gène *let-2*. Il est également possible que les autres facteurs impliqués soit essentiels et leur mutation induise un phénotype létal chez *C. elegans*. Ce pourrait être également le cas du facteur contrôlant l'expression du gène *asd-2b* et que nous n'aurions pas pu identifier lors de cette étude. D'où l'intérêt d'avoir recours à une stratégie de génétique inverse par crible de banque ARNi pour contourner cette limite. C'est l'étude que mène ma collègue Karine Reborà pour cribler la banque à la recherche de facteurs en *trans* impliqués dans l'épissage alternatif du gène *let-2*.

## 2.2 APPROCHE PAR GENETIQUE INVERSE

### 2.2.1 Modification du COPAS pour l'obtention d'une plateforme adaptée à un crible ARNi à grande échelle

#### 2.2.1.1 Le module Reflx natif du COPAS

L'un des objectifs du laboratoire est de conduire des cribles de génétique inverse à l'échelle génomique, c'est pourquoi nous nous sommes équipé d'une version du COPAS disposant d'un bras motorisé nommé Reflx capable d'analyser de manière autonome les échantillons présents dans une plaque multi-puits (jusqu'à 384 puits).



**Figure 33 : Configuration native du module Reflx du COPAS**

Les flèches rouges indiquent le sens de progression de l'échantillon et du tampon. Les valves unidirectionnelles sont représentées par des cercles numérotés. L'échantillon est représenté par une goutte jaune contenant des nématodes, les bulles sont représentées par un cercle de bulles. a,b,c,d,e et f représentent les différentes étapes de progression de l'échantillon.

Dans sa configuration native le système utilise 5 valves unidirectionnelles pour transférer les échantillons séquentiellement depuis la plaque multi-puits vers un piège à bulles (Bubble trap) qui permet

d'éliminer les bulles d'air de l'échantillon avant son passage devant la cellule de détection où elles seraient reconnues comme des objets (Figure 33).

Lors du traitement des échantillons, l'aiguille plonge dans le puits, la valve numéro 1 s'ouvre pour laisser passer le liquide aspiré par le piston afin de le diriger vers le piège à bulles. À la fin de l'aspiration, la valve 2 s'ouvre pour injecter du liquide dans le puits et resuspendre les objets sédimentés pour permettre leur aspiration (Figure 33 a). Le piston arrête l'aspiration, les valves 1 et 2 se referment, et la valve 3 s'ouvre (Figure 33 b) pour permettre la progression de l'échantillon vers le piège à bulles (Figure 33 c) où les bulles d'air s'échappent à l'air libre après ouverture de la valve numéro 4 (Figure 33 d). Les valves numéro 3 et 4 se referment (Figure 33 e) et la valve numéro 5 s'ouvre. Le piston commence alors l'expulsion de l'échantillon vers la chambre de détection où il sera analysé (Figure 33 f). La valve numéro 5 se referme ensuite et le bras motorisé se positionne au-dessus du puits suivant.

Dans cette configuration, le traitement de chaque puits nécessite 38 secondes. Le traitement d'une plaque de 96 puits entière exige donc un temps de 62 minutes.

Lors du congrès international de *C. elegans* 2009 qui s'est tenu à l'université UCLA (Los Angeles, USA), j'ai discuté avec différentes personnes utilisant le module Reflex du COPAS qui m'ont fait part d'un problème de « carry over » des vers d'un puits vers le suivant. Je me suis rapproché du stand de la société « Union Biometrica » (qui a conçu et commercialise le COPAS) présente à l'occasion du congrès pour recueillir leur solution pour ce problème. La solution qui m'a été proposée est :

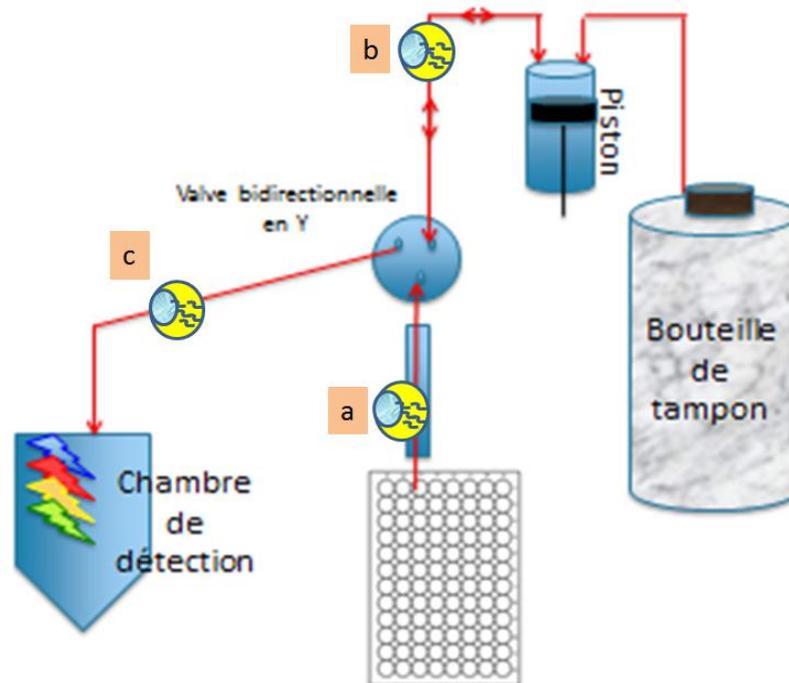
*« Vous devez intercaler deux puits vides entre chaque échantillon à tester afin de pallier au problème »* Union Biometrica, communication personnelle au Congrès International de *C. elegans* 2009 (Los Angeles, USA).

Cela aurait eu pour effet de tripler la quantité de plaque 96 puits nécessaire à la réalisation d'une étude à une échelle génomique. Par ailleurs, nous aurions dû adapter nous même les plaques à ce nouveau format et référencer les nouvelles positions pour chaque clone à partir des banques génomiques disponibles pour la communauté scientifique. Nous avons donc décidé de modifier le module Reflex afin d'augmenter sa vitesse de traitement et corriger le problème de « carry over ».

### **2.2.1.2 Amélioration des performances du module Reflex**

Nous avons changé la configuration matérielle de la machine pour ne garder que les étapes strictement nécessaires : aspiration, refoulement et traitement.

Pour cela, nous avons décidé d'éliminer les cinq valves unidirectionnelles. Elles ont été remplacées par une valve unique en Y qui permet de transférer en seulement deux étapes les échantillons depuis le puits vers la chambre de détection. La première entrée de la valve est reliée à l'aiguille d'aspiration, la deuxième au piston et la troisième à la chambre de détection (Figure 34).

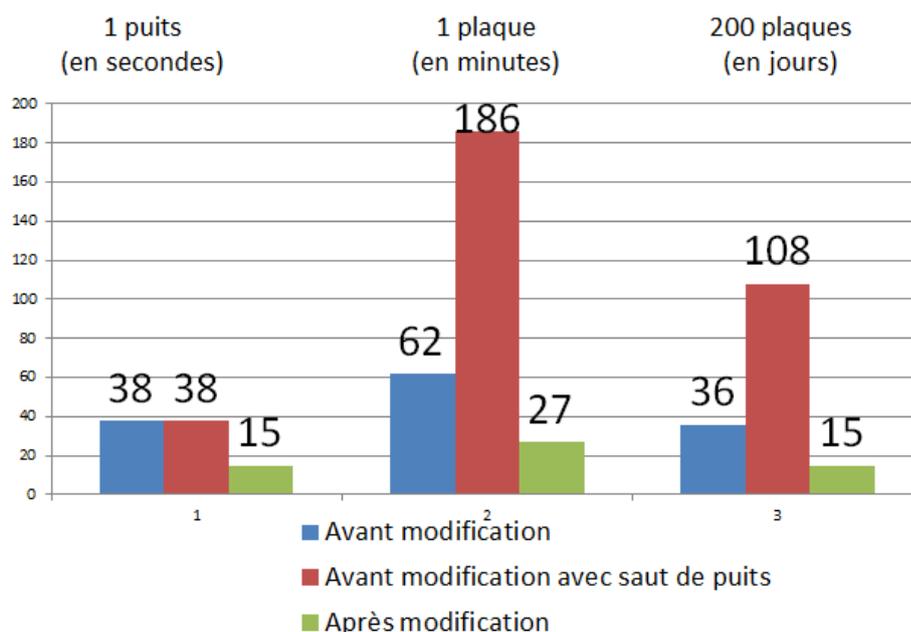


**Figure 34 : Configuration du module Reflx du COPAS après modification**

Les flèches rouges indiquent le sens de progression de l'échantillon et du tampon. Les valves unidirectionnelles sont représentées par des cercles numérotés. L'échantillon est représenté par une goutte jaune contenant des nématodes, les bulles sont représentées par un cercle de bulles. a,b,c,d,e et f représentent les différentes étapes de progression de l'échantillon.

Le piège à bulles a également été éliminé. Dans cette configuration, les bulles d'air seront transportées conjointement avec les vers et passeront par la chambre de détection. Elles seront détectées comme des événements par le cymomètre, mais leurs caractéristiques optiques distinctes de celles des vers permettront de les éliminer à posteriori par traitement informatique.

Désormais, 15 secondes suffisent à la machine pour analyser la totalité de l'échantillon contenu dans chaque puits. Les nouvelles performances du COPAS sont comparées à celles de la version native (Figure 35).



**Figure 35 : Performance du module Reflx du COPAS avant et après modification**

Dans la nouvelle configuration, le module Reflx du COPAS est 2,5 fois à 7 fois plus rapide que dans la version native. Une étude à l'échelle génomique comptant 200 plaques peut désormais être effectuée en quinze jours.

Des modifications aussi importantes du matériel nécessitent une validation de chacune des étapes, de l'aspiration jusqu'à l'analyse.

### **2.2.1.3 Traitement logiciel des données collectées**

Dans la configuration initiale, la valve numéro 2 permettait un apport en solution tampon qui venait resuspendre les vers sédimentés au fond du puits. L'apport en solution tampon étant éliminé dans la nouvelle configuration, il était donc nécessaire de vérifier que la totalité des nématodes sont aspirés. Pour cela, chaque puits d'une plaque a été rempli de 1000 nématodes dans un volume final de 150 µl puis la plaque est analysée. A la fin de l'analyse, la plaque est inspectée à la loupe binoculaire. Les tests préliminaires ont montrés que plusieurs dizaines de vers restaient en suspension dans une goutte d'environ 10 µl au fond de chaque puits ce qui correspond à une perte de 1% de l'échantillon étudié. Nous avons résolu ce problème en ajoutant, une pointe de cône de pipette de 2 µl au tube d'aspiration pour aspirer la totalité de l'échantillon sans aucune autre modification. Avec cette modification, moins de 5 vers persistent par puits dans un volume de liquide inférieur à 1µl.

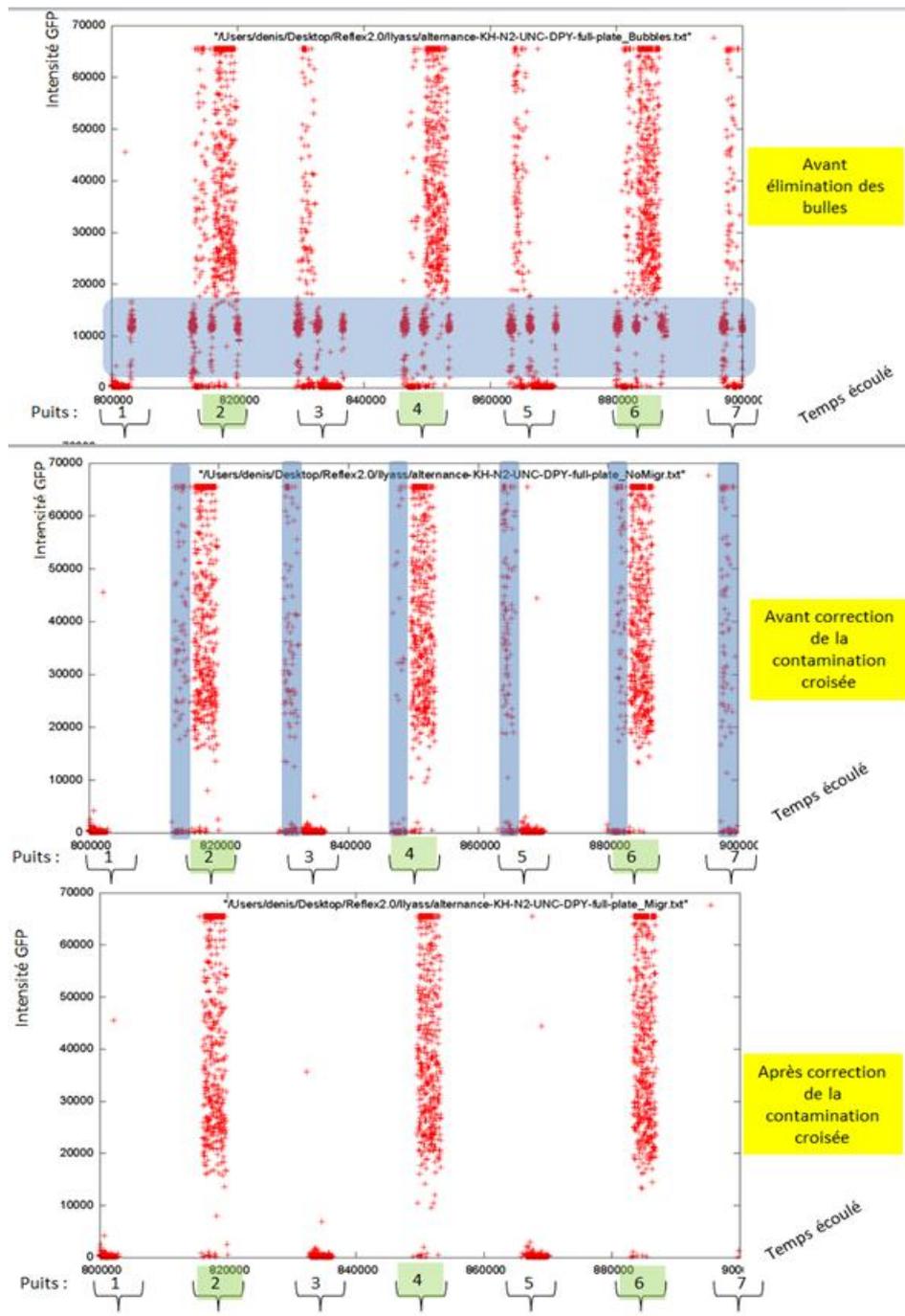
L'étape d'aspiration de l'échantillon est donc satisfaisante. Il faut à présent valider les performances du logiciel développé pour éliminer les bulles.

Lors d'une étude préalable du comportement des bulles d'air lors de leur passage devant le détecteur, l'attention de mon directeur de thèse a été portée sur le profil atypique de celles-ci. En effet, les bulles disposent d'un profil de fluorescence qui représente une courbe en cloche (courbe de Gauss) bien différent de celui d'un nématode. De plus, les bulles d'air ne représentent aucune fluorescence dans le canal jaune. Ces deux paramètres ont été utilisés pour créer un modèle mathématique qui permet de distinguer automatiquement une bulle d'un ver. Grâce au logiciel mis en place par le développeur informatique de l'équipe Léo Guignard, les objets qui sont identifiés comme des bulles d'air sont écartés des analyses. Le rôle du bocal de dé-bullage qui a été éliminé dans la nouvelle configuration a été remplacé avec succès par le traitement logiciel. L'analyse à posteriori des données et leur traitement informatique a permis d'économiser un temps non négligeable.

Pour tester si notre modification a effectivement résolu le problème de « carry-over », nous avons interposés dans chaque puits d'une plaque entière deux souches de *C. elegans*. L'une est la souche sauvage (N2) non fluorescente et l'autre (KH1401) fluoresce fortement dans le vert. La plaque est analysée au COPAS sous sa nouvelle configuration (Figure 36).

Nous observons qu'une zone de mélange des échantillons contenant à la fois des vers GFP positifs et des N2 est délimitée par les deux premières vagues de bulles au début de chaque puits (Figure 36). En revanche les données présentes entre la deuxième et la dernière vague de bulles contient exclusivement des échantillons correspondant à leurs puits respectifs.

L'attribution informatique des données de la zone de mélange aux puits correspondants est théoriquement possible dans le cas où des caractéristiques physiques bien distinctes sont exhibées par chaque échantillon. Ici il s'agit de la fluorescence dans le vert spécifique à chaque souche. Or, lors d'un crible à grande échelle, nous ne connaissons pas d'avance les caractéristiques spécifiques des échantillons présents dans chaque puits ou bien leurs caractéristiques sont similaires et ne permettent donc pas la discrimination entre les objets venant du puits précédent ou suivant.



**Figure 36 : Correction de la contamination croisée par le logiciel**

Une fenêtre de 7 puits issue d'une plaque 96 puits est présentée ici, les puits contiennent soit des vers N2 non fluorescents (en blanc) soit des vers KH1401 fluorescents (en vert). Chaque croix correspond à un animal unique positionné en fonction de l'intensité de GFP détectée et le temps auquel il est passé devant le détecteur (proportionnel à la position du puits). La zone surlignée en bleu dans la figure du haut correspond aux bulles d'air détectées par la chambre de détection, la zone surlignée dans le graphique du milieu correspond aux zones de contamination croisées survenant à intervalle régulier.

Nous avons décidé d'éliminer les données présentes dans cet intervalle survenant à temps fixe lors du traitement de chaque puits qui contient un mélange d'échantillons de deux puits successifs. Les données éliminées correspondent à une perte de 7.5% des échantillons de chaque puits.

La modification du COPAS combinée au développement d'un logiciel adapté a permis d'augmenter significativement la vitesse d'analyse des données dans le format 96 puits nécessaire aux projets de criblage systématiques que nous souhaitons mettre en œuvre. Nous avons par ailleurs, réussi à surmonter le problème de carry-over qui est problématique lors des études à haut débits utilisant des banques génomiques préexistantes.

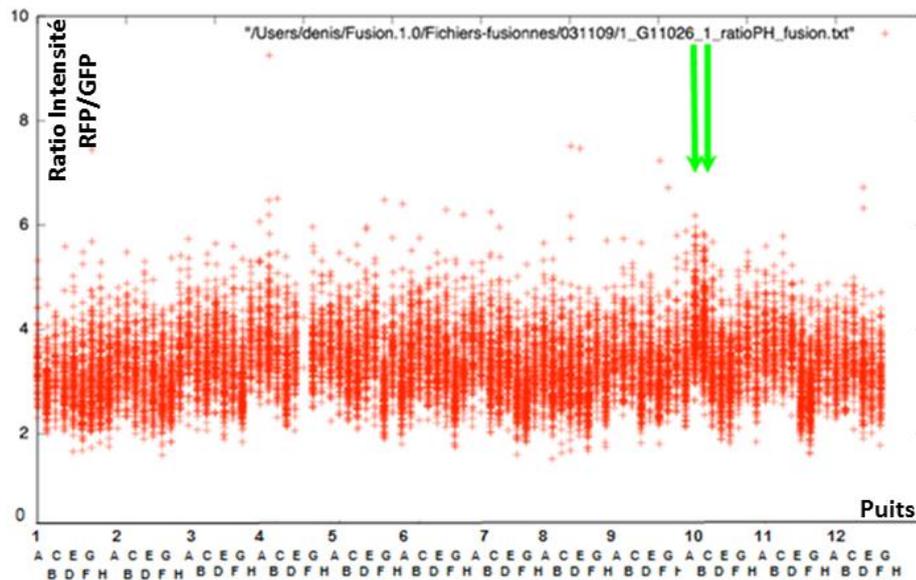
### **2.2.2 Criblage systématique des facteurs de l'épissage de *let-2* par ARNi**

La disponibilité de l'outil RNAi pour la déplétion de l'expression des gènes a permis une multitude d'études à l'échelle génomique chez *C. elegans* [246-252].

Une étude complémentaire à la mutagenèse aléatoire pour l'étude de l'épissage alternatif du gène *let-2* décrite plus haut a été menée par ma collègue Karine Rébora. L'approche de génétique inverse qu'elle utilise en optant pour un crible d'ARNi à échelle génomique sur la souche KH1401 présente deux avantages majeurs. L'ARNi permet de contourner les phénotypes létaux qui n'auraient pas pu être sélectionnés lors de la mutagenèse aléatoire dont la sélection s'opère au-delà de la troisième génération. Le second avantage inhérent à la génétique inverse réside dans le fait que l'observation d'un phénotype se traduit par l'identification immédiate du gène ciblé.

Pour cette étude, K. Rebora a utilisé la librairie RNAi d'ORFeome disponible pour *C. elegans* [253, 254]. Cette librairie contient des bactéries exprimant l'ARNdb ciblant 10953 gènes de *C. elegans*. Les bactéries sont disposées dans des plaques 96 puits et la librairie complète en compte 140. La lignée KH1401 sera disposée dans cette étude dans des plaques 96 puits et sera nourrie par les bactéries exprimants un ARN double brin. La déplétion de l'expression d'un gène impliqué dans le phénomène de l'épissage alternatif de *let-2* devrait se traduire par une modulation de fluorescence vert/rouge. Cette modulation va être mesurée par cytométrie en flux. L'utilisation du module Reflx du COPAS permettra l'automatisation du processus pour la réalisation d'une étude à haut débit. La disposition en plaque 96 puits permet de tirer pleinement avantage des modifications apportées au COPAS. L'optimisation de celui-ci a été faite spécifiquement pour des plaques de ce format. L'étude à échelle génomique constitue un test grandeur nature pour la machine.

Toute la banque génomique d'ARNi a été testée lors de cette étude. Le résultat pour chaque plaque est ensuite analysé pour déterminer si les animaux d'un puits particulier exhibent un phénotype intéressant. Ce phénotype se traduit par un ratio plus élevé dû à la modulation de fluorescence.



**Figure 37 : Crible RNAi sur une plaque 96 puits de l'épissage de *let-2***

La plaque 96 puits contenait environ 200 vers KH1401 par puits. Chaque puits reçoit des bactéries RNAi spécifiques. Le ratio de fluorescence Vert/Rouge est reporté sur le graphe pour chaque animal (croix rouge) en fonction du puits. L'ARNi ayant servi à nourrir les animaux d'un puits exhibant une modulation de fluorescence est retenu pour validation (flèches vertes).

Voici l'exemple des données traitées pour une des 140 plaques de la banque dont nous disposons (Figure 37). Sur les 96 puits, les ARNi ayant induit une modulation de fluorescence sont retenus comme gènes potentiellement impliqués dans l'épissage et sont sélectionnés pour des tests de confirmation ultérieurs. Dans l'exemple que nous vous présentons, 2 puits sur 96 font partie de ceux sélectionnés. Moins d'une demi-heure a été nécessaire pour accumuler ces données grâce aux modifications que nous avons apporté au bras automatisé du COPAS. Les résultats préalables obtenus par ma collègue Karine Rébora lors du crible ARNi des facteurs de l'épissage de *let-2* ont permis la sélection d'une centaine de gènes candidats dont l'invalidation exhibe un phénotype intéressant.

Nous avons comparé la liste de gènes mutés chez les mutants que nous avons isolés à ceux des candidats intéressants sélectionnés lors des résultats préalables obtenus lors du crible génomique ARNi. Cependant, aucun gène ne constitue un candidat en commun entre nos deux études.

Cette étude à grande échelle constitue une validation grandeur nature des modifications que nous avons réalisées. Esther Marza, Post-doctorante ayant rejoint le laboratoire, a également pu utiliser cette plateforme pour un crible ARNi dans le cadre de son projet d'étude de la régulation génétique de la réponse du stress du Réticulum Endoplasmique Chez *C. elegans*.

Le schéma du prototype que nous avons construit a inspiré une nouvelle génération de module Reflx qui est maintenant disponible commercialement.

**III. DEUXIEME PARTIE : ETUDE  
*IN VIVO* DE L'IMPACT DE LA  
MODIFICATION DE LA REGION  
CIS REGULATRICE DU GENE  
*EGL-15***

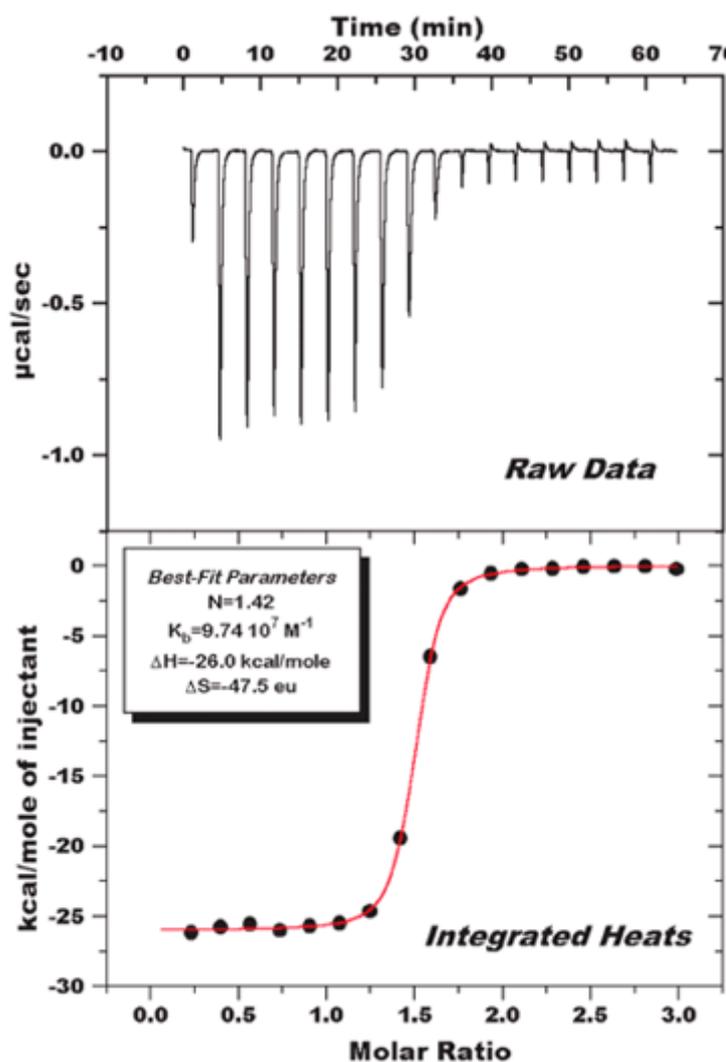
## **Objectif :**

L'étude initiale menée par notre collaborateur a permis d'identifier la protéine SUP-12 comme facteur d'épissage essentiel se fixant sur la séquence GUGUG du gène *egl-15* [230].

Afin de mieux comprendre le rôle de cette séquence *cis*-régulatrice pour l'épissage du gène *egl-15*, Cameron Mackereth (IECB, Bordeaux, France) a initié une étude *in vitro* entre cette séquence WT ou portant des mutations et le domaine de liaison de l'ARN de la protéine SUP-12 par mesure de calorimétrie par titration isothermale. Nous allons générer des lignées de vers rapporteurs de l'épissage *egl-15* et portant ces mutations dans la cassette BGAR afin de vérifier si il y a une corrélation entre les résultats obtenus *in vitro* et ceux qu'on observe *in vivo*.

### **3.1 MESURE D'AFFINITE SUP-12/EGL-15 PAR ITC**

La calorimétrie par titration isothermale ou ITC (Isothermal Titration Calorimetry) est une technique qui permet la mesure des paramètres d'interaction entre des biomolécules [255, 256]. Il peut s'agir d'une interaction protéine/protéine, protéine/acide nucléiques, antigène/anticorps, enzyme/ligand, etc... L'interaction entre les biomolécules libère ou absorbe de l'énergie sous forme de chaleur. Cette caractéristique thermodynamique peut être directement mesurée par ITC. La mesure de cette chaleur permet une détermination précise de l'affinité de liaison ( $K_d$ ), la stœchiométrie de liaison ( $n$ ), l'enthalpie ( $\Delta H$ ) et l'entropie ( $\Delta S$ ) fournissant ainsi un profil thermodynamique complet de l'interaction moléculaire en une seule expérience [256].

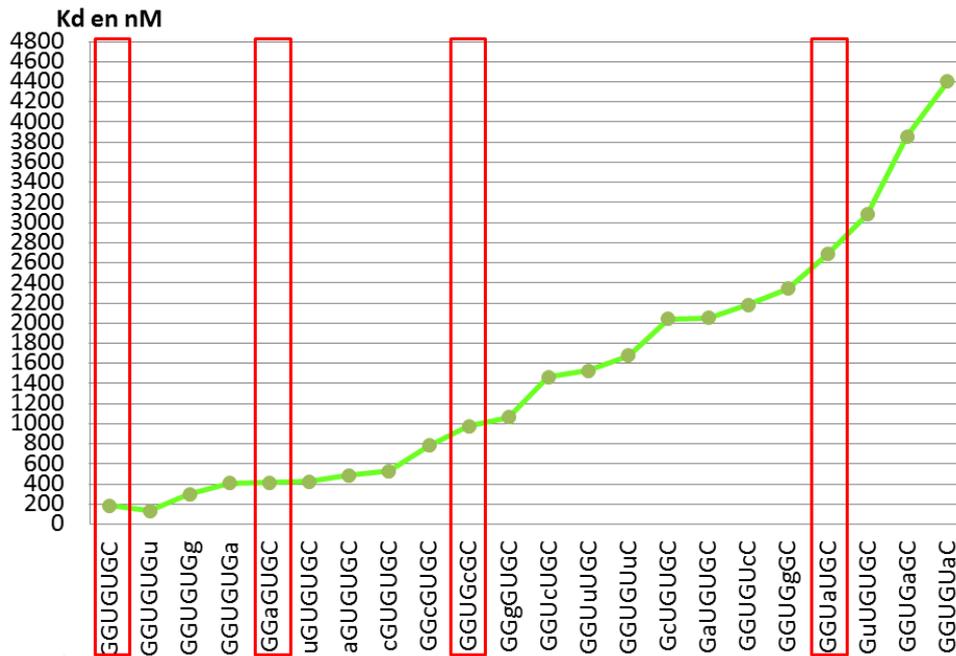


**Figure 38 : Données typiques générées par ITC**

La quantité de chaleur émise ou absorbée lors d'une interaction est mesurée en fonction du temps. Chaque pic correspond à l'injection d'une quantité déterminée de ligand (graphique du haut). La courbe des paramètres thermodynamiques de liaison peut donc être obtenue et la stœchiométrie ( $n$ ), la constante d'association ( $K_d = 1/K_b$ ), l'enthalpie ( $\Delta H$ ) et l'entropie ( $\Delta S$ ) de cette interaction peuvent être calculés (courbe du bas). [Source fabricant GE Healthcare [www.microcal.com](http://www.microcal.com)]

Dans une expérience typique d'ITC, une solution d'une biomolécule, est titrée dans une solution de son partenaire de liaison. La chaleur libérée lors de leur interaction ( $\Delta H$ ) est suivie au cours du temps. Chaque pic correspond à une variation de chaleur associée à l'injection d'un petit volume d'échantillon dans la cellule de réaction de l'ITC (Figure 38, haut). Des quantités successives de ligand sont titrées dans la cellule ITC et la quantité de chaleur absorbée ou dégagée est directement proportionnelle à la quantité de molécules qui interagissent. Lorsque le système arrive à saturation, le signal de chaleur diminue ou augmente pour atteindre l'équilibre. Une courbe de liaison peut donc être obtenue en fonction de la quantité de chaleur mesurée pour chaque titration en fonction du ratio des molécules en interaction dans la cellule de détection (Figure 38, bas).

Afin d'étudier la spécificité de la séquence de liaison de SUP-12 sur l'ARN pré-messager du gène *egl-15*, l'équipe de notre collaborateur Cameron Mackereth (Bordeaux, France) a mesurée cette interaction par ITC pour déterminer les valeurs thermodynamiques de cette association. Par la suite, des mutations ponctuelles ont été introduites dans la séquence GGTGTGC pour déterminer quels acides nucléiques sont déterminants pour l'affinité de liaison. Pour cela, chacune des bases de la séquence a été remplacée par l'une des trois autres possibles et la constante d'affinité a été mesurée.



**Figure 39 : Constante d'affinité mesurée *in vitro* pour chaque mutation ponctuelle induite**

La séquence GGUGUGC est mutée pour chacune des bases. Pour chaque position les 3 substitutions sont effectuées et le Kd de la nouvelle séquence est mesuré en nM puis reporté sur ce graphique. Les mutations induites sont représentées en lettre minuscules. La séquence sauvage et les substitutions retenues pour être testées *in vivo* sont encadrées en rouge.

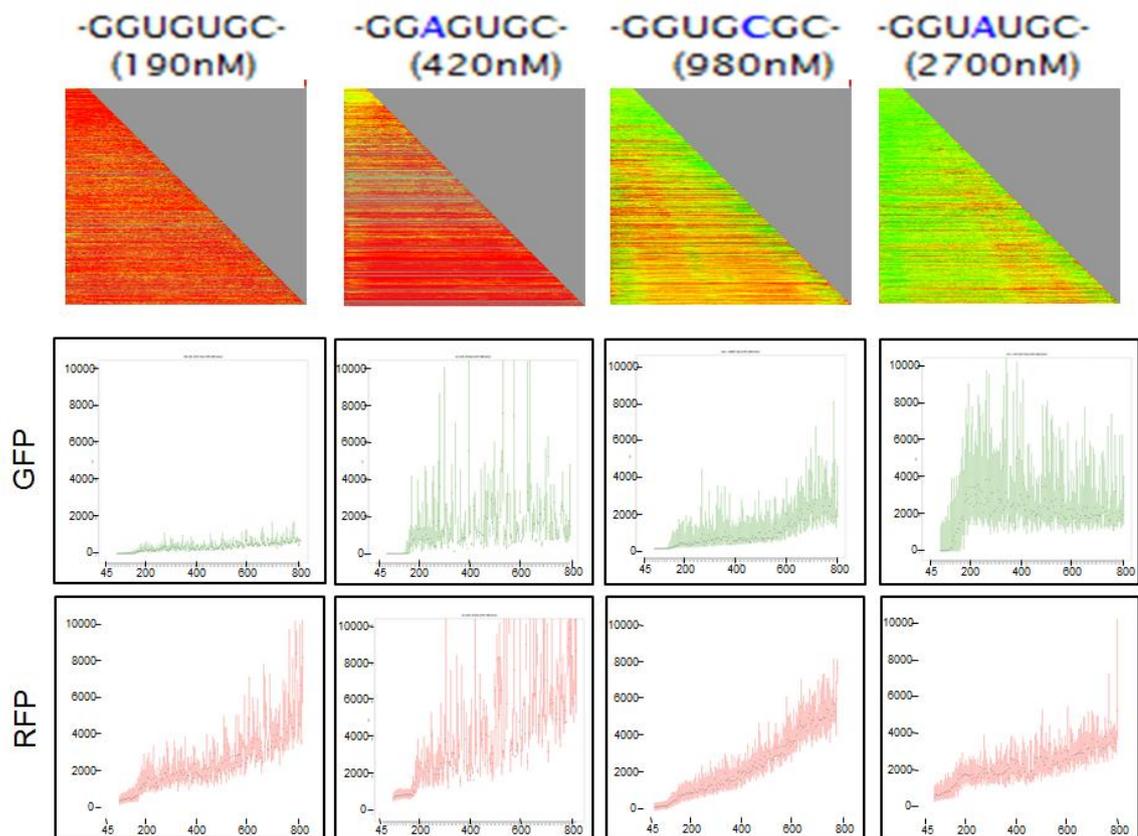
Les résultats obtenus pour les 21 séquences étudiées ont permis de confirmer que le motif GUGUG était optimal pour la fixation de SUP-12 (Figure 39). De plus, les résultats obtenus par nos collaborateurs montrent qu'une mutation ponctuelle dans la séquence de liaison de SUP-12 modifie l'affinité de l'interaction sur une large gamme allant de 190 nM à plus de 4000 nM selon la nature et la position de celle-ci.

Nous avons tiré avantages de cette large gamme d'affinité en étudiant plus particulièrement 3 substitutions : GGaGTGC, GGTGcGC et GGTaTGC dont la constante d'association est respectivement de 420 nM, 980 nM et 2700 nM à comparer avec 190 nM de la séquence sauvage GGUGUGC (Figure 39). Ces mutations sont représentatives de l'ensemble du spectre d'affinité précédemment déterminé. Nous avons décidés de vérifier si les résultats observés *in vitro* ont un impact *in vivo* sur l'épissage alternatif du gène *egl-15* chez *C. elegans*.

### 3.2 CORRELATION DES RESULTATS *IN VIVO* ET *IN VITRO*

Afin de valider *in vivo* les résultats obtenus *in vitro* par Cameron Mackereth et son équipe, nous avons généré les lignées transgéniques de *C. elegans* portant les rapporteurs BGAR respectifs pour chacune des mutations ponctuelles sélectionnées. La cassette BGAR portant la séquence sauvage permet l'inclusion de l'exon 5A sous le contrôle du promoteur *myo-3* et se traduit par l'expression exclusive de la RFP dans le BWM tout au long du développement (Figure 18 C). Nous avons généré également la lignée portant la cassette BGAR sauvage pour qu'elle puisse servir de référence.

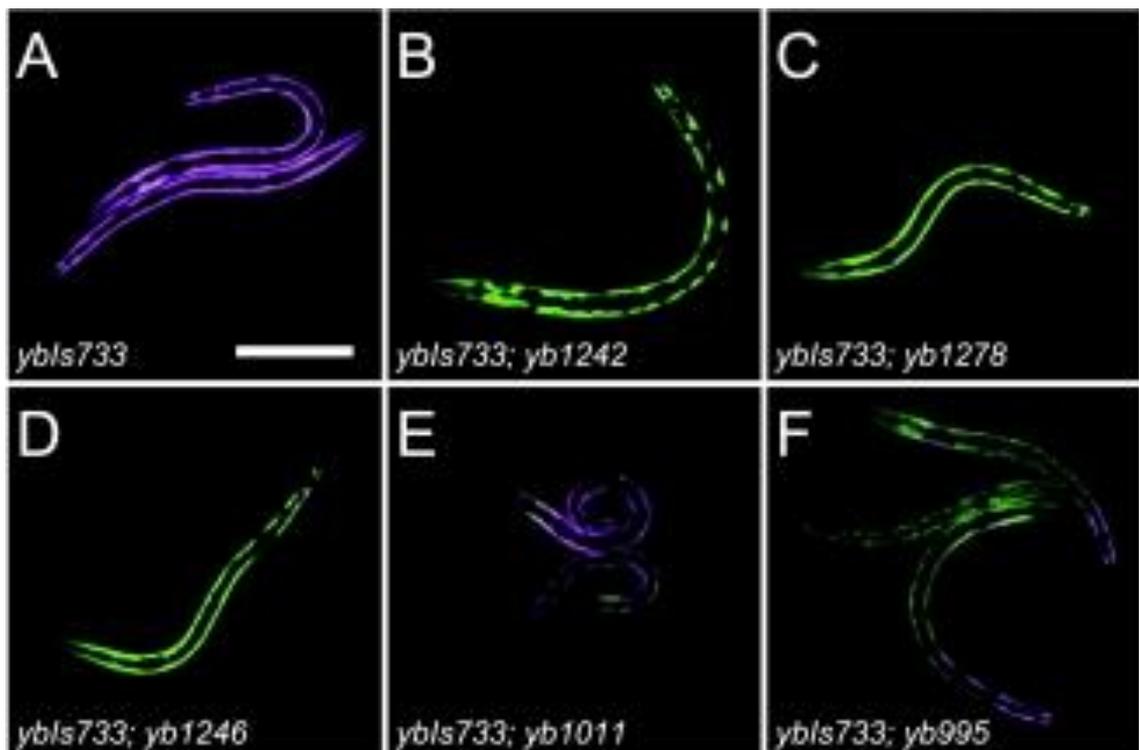
Chaque lignée a été analysée au COPAS pour définir son profil d'expression dans le rouge, dans le vert et son chronogramme bicolore pour déterminer l'expression spatio-temporelle de la GFP et de la RFP. Nous voulions voir si l'épissage alternatif du transgène est directement corrélé au Kd mesuré pour chaque mutation.



**Figure 40 : fluorescences *in vivo* des lignées portant les mutations de séquence GTGTG**

Les chronogrammes bicolores caractéristiques pour chaque lignée sont présentés en haut. L'intensité moyenne de fluorescence dans le vert et dans le rouge en fonction de la taille des animaux sont reportées sur les graphiques respectifs du bas. Les abscisses et les ordonnées sont identiques à ceux représentés dans Figure 23. La séquence sauvage et les séquences mutées (en bleu) ainsi que leur affinité calculées par ITC sont indiquées.

Les résultats obtenus sur les lignées transgéniques *in vivo* sont cohérents avec ceux obtenus *in vitro* par l'équipe de Cameron Mackereth. Moins la séquence est affine pour la protéine SUP-12 plus l'expression de la GFP (l'inclusion de l'exon 5B) augmente (Figure 40). Nous observons également que l'expression de la GFP est progressive en fonction de l'affinité de SUP-12 pour la séquence *cis*-régulatrice ce qui se traduit par l'obtention d'animaux qui expriment à la fois le rouge et le vert dans le BWM comme cela a été observé lors de l'étude préalable menée par nos collaborateurs japonais sur les phénotypes induits par différentes mutations portées par le gène *sup-12* qui se lie à cette séquence (Figure 41).



**Figure 41 : mutants *sup-12* déficients pour l'épissage alternatif correct de la cassette BGAR**  
 Images de fluorescence de la RFP (magenta) et de la GFP (vert). (A) expression normale de la cassette. (B et C) mutants déficients pour l'épissage alternatif de la cassette. (D à F) mutants chimériques pour l'épissage alternatif. [Kuroyanagi et al. 2007 [230]]

Ceci confirme notre capacité à mener des études quantitatives de l'épissage alternatif en particulier et de la régulation post-transcriptionnelle plus globalement *in vivo* chez *C. elegans*.

Une publication issue de cette collaboration portant comme titre « Nucleic-acid binding by splicing factor SUP-12 and isoform control in nematode muscle » est en cours de rédaction.

**IV. TROISIEME PARTIE :  
GENERATION DE SYSTEME  
RAPPORTEUR *IN VIVO* DE  
REGULATION POST-  
TRANSCRIPTIONNELLE PAR  
MIARN**

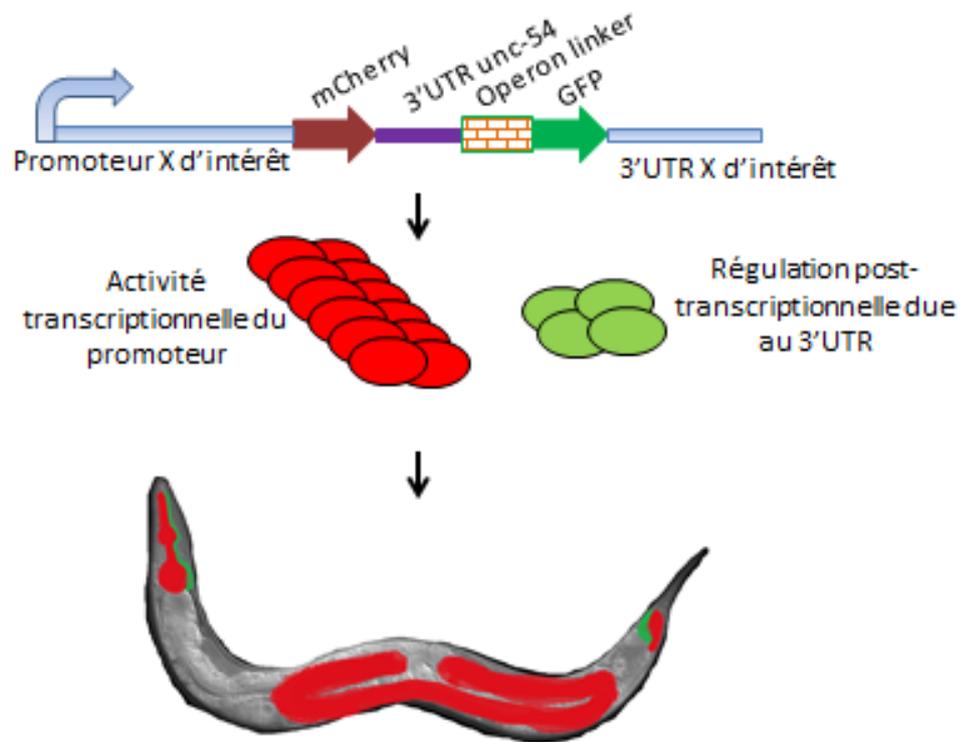
## Objectif :

Lors de l'étude préalable présentée plus haut, Lall et ses collègues ont procédé à la génération et à la comparaison de l'expression de fluorescence de plusieurs lignées indépendantes de *C. elegans* portant chacune une construction différente dans le but d'étudier la régulation post-transcriptionnelle médiée par la région 3'UTR des gènes cibles du microARN *let-7*. Cependant, le nombre de vers qui peuvent être examinés à l'œil sous microscope constitue rapidement une limite expérimentale et le manque de contrôle interne à la lignée étudiée peut entraîner un biais d'appréciation lors de la comparaison de deux lignées indépendantes [231].

Nous avons décidé de développer une stratégie adaptée à une validation à grande échelle des cibles des miARN de *C. elegans* prédites par PicTar. Nous allons générer des lignées bicolores de *C. elegans* dont la fluorescence GFP est rapportrice de la régulation post-transcriptionnelle due à la région 3'UTR étudiée tandis que la fluorescence de la protéine mCherry (Rouge) sert de contrôle interne et de référence transcriptionnelle. L'utilisation du cytomètre en flux adapté au nématode (COPAS ; Union Biometrica) en aval permettrait l'automatisation de l'étude comparative et l'analyse de plusieurs centaines de vers à la seconde.

### 4.1 STRATEGIE RRPT

Nous avons décidé d'utiliser une double fluorescence, la GFP (vert) qui sera liée à l'UTR étudié, et la mCherry (rouge) sous le contrôle de l'UTR permissif du gène *unc-54*. Cette double fluorescence permet de disposer d'un contrôle interne à chaque lignée [237]. Dans notre construction, les deux protéines fluorescentes seront co-exprimées sous forme d'opéron, séparés par la séquence intermédiaire de l'opéron *gpd-2/gpd3* endogène à *C. elegans* [209]. Cette cassette bicolore sera nommée cassette RRPT pour Rapporteur de la régulation post-transcriptionnelle (Figure 42).

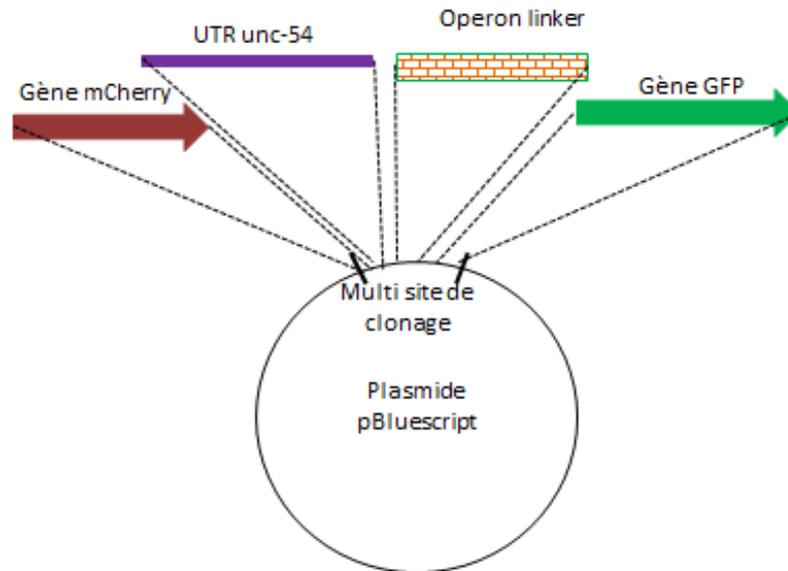


**Figure 42 : Exemple d'expression *in vivo* de la cassette RRPT**

La cassette RRPT est insérée en lieu et place de la séquence codante du gène X. La protéine mCherry transcrite est représentée en cercles rouges tandis que la GFP est en cercles verts. Dans cet exemple, le promoteur X induit une expression dans le pharynx, les gonades et les cellules postérieures. Une régulation post-transcriptionnelle induite par l'UTR du gène X inhibe l'expression de la GFP dans les gonades.

Nous avons procédé à la construction de la cassette RRPT par les techniques classiques de clonage. Le multi-site de restriction du plasmide de clonage pBluescript a été utilisé pour accueillir les quatre éléments de notre rapporteur (Figure 43) [la carte du plasmide pBS-RRPT généré est représentée Figure 53].

La cassette rapportrice de la régulation post-transcriptionnelle (RRPT) a été clonée dans le vecteur de clonage pBluescript (SK-) (pBS (SK-)) grâce à son multi site de restriction. La mCherry a été amplifiée par PCR en ajoutant le site de restriction SacI en amont et NotI en aval. La PCR du 3'UTR du gène *unc-54* contient les sites NotI en amont et XbaI en aval. La PCR de l'opéron linker *gpd-2/gpd-3* est bordée par les sites XbaI et SmaI. L'amplification de la GFP permet d'introduire SmaI et PstI. Les différents fragments générés pour être insérés comportent des sites de restriction utilisés pour digérer les uns les autres, la seule combinaison possible consiste à insérer en premier lieu l'opéron linker *gpd-2/gpd-3* puis la mCherry suivie de l'UTR *unc-54* et enfin la GFP.



**Figure 43 : construction de la cassette RRPT par clonage classique**

Les quatre fragments de la cassette RRPT sont clonés dans le multi site de clonage du plasmide pBluescript. La mCherry (rouge) a été clonée comme fragment SacI/NotI ; l'UTR unc-54 (magenta) a été inséré comme fragment NotI/XbaI ; l'opéron linker XbaI/SmaI et la GFP SmaI/PstI.

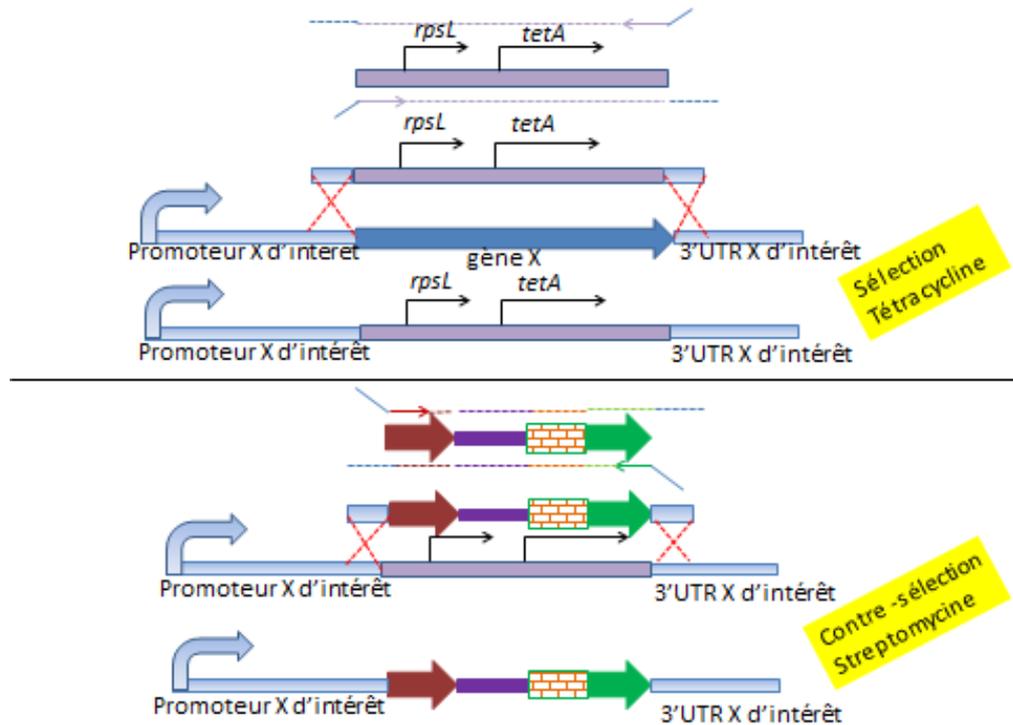
#### 4.1.1 Le Recombineering

En 2006, Dolphin et Hope publient une technique de génération de rapporteurs fluorescents chez *C. elegans* [257] grâce à la recombinaison par l'intermédiaire de la Red-recombinase [258]. Cette dernière reconnaît une homologie entre deux séquences et opère un « crossing over » entre les deux. De cette manière, la première séquence se retrouve insérée en lieu et place de la seconde (Figure 44 et Figure 46)

L'utilisation de cette approche tire l'avantage de la disponibilité d'une librairie complète de clones de fosmides génomiques de *C. elegans*. L'insertion de notre cassette RRPT dans le fosmide permet de préserver le contexte génomique du gène étudié avec ses régions 5' et 3' UTR sur plusieurs Kb au cas où des éléments régulateurs soient présents dans les structures chromatiniennes distantes. Le profil d'expression qui résulte d'une telle construction mimera avec une meilleure résolution l'expression endogène du gène.

La technique de Dolphin et Hope permet l'insertion sans cicatrice d'un rapporteur en deux étapes de recombinaison et contre-sélection. La première recombinaison a pour objectif d'insérer le gène de résistance à la tétracycline en lieu et place de la région codante du gène cible. L'homologie reconnue par la Red-recombinase est ajoutée artificiellement par les amorces lors de

l'amplification du gène de résistance par PCR (une homologie de 50 nucléotides est introduite). Les événements de recombinaison sont sélectionnés sur un milieu contenant de la Tétracycline (Figure 44, haut). Les colonies capables de pousser sont celles ayant reçues le gène de résistance. A la deuxième étape de recombinaison, la cassette RRPT flanquée de part et d'autre d'une séquence d'homologie de 50nt avec le site d'insertion est insérée en lieu et place du gène de résistance à la Tétracycline. Les événements de cette deuxième recombinaison sont contre-sélectionnés par la Streptomycine qui est toxique en présence du gène *rpsL* (Figure 44, bas).



**Figure 44 : Principe de recombineering selon Hope et Dolphin**

La cassette *rpsL tetA* est amplifiée par PCR à l'aide d'amorces portant une queue homologue au site d'insertion, l'amplification attendue est représentée en trait discontinu. La croix rouge représente l'évènement de recombinaison. La première étape de recombinaison (haut) est sélectionnée par l'addition de tétracycline. La seconde recombinaison est sélectionnée par l'addition de Streptomycine (bas). La cassette RRPT est donc insérée en deux étapes par recombineering.

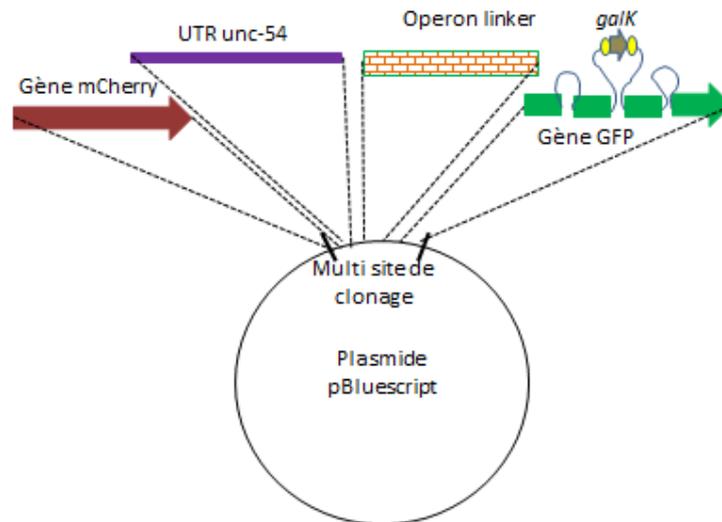
En deux étapes de recombinaisons, la cassette RRPT est censée être insérée en lieu et place de la région codante du gène cible sous le contrôle de son promoteur et ses régions régulatrices.

Le recours systématique à deux étapes de recombinaison pour réaliser une sélection et une contre-sélection s'avère être chronophage et peu adaptable à la réalisation de construction en routine.

Nous avons également testé un second protocole de recombineering permettant la génération d'un fosmide rapporteur pour *C. elegans* en une seule étape [259].

Dans ce nouveau protocole, la GFP utilisée contient le gène de sélection *galK* logé dans l'un de ces introns. Le gène *galK* est délimité de part et d'autre par une séquence reconnue par la Flp recombinase. L'expression de cette dernière permet d'exciser le gène *galK*. La cicatrice de cette excision est située dans l'intron de la GFP et n'a donc aucun effet sur l'expression de la cassette.

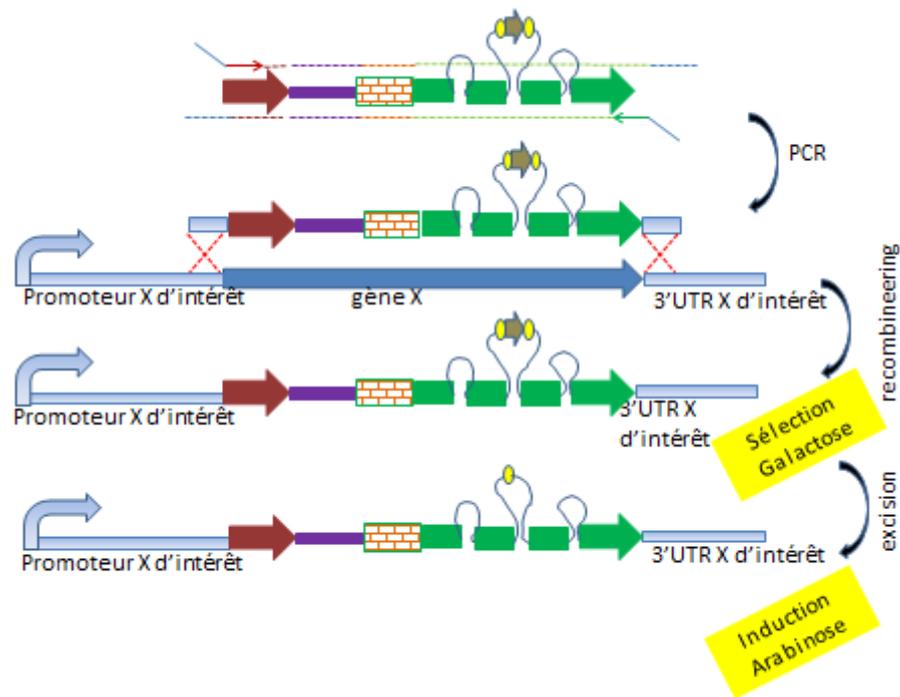
Nous avons remplacé la GFP présente dans la version initiale de la cassette bi-cistronique par la GFP poly-exonique portant le gène *galK* (Figure 45).



**Figure 45 : Génération de la Cassette RRPT (*galK*)**

La cassette RRPT (*galK*) a été construite par clonage dans le multi site de restriction du plasmide pBlueScript en utilisant les mêmes sites de restriction que lors de la construction de la cassette RRPT initiale. La GFP contient 4 exons représentés par des rectangles verts, les introns par des traits bleus. Le gène *galK* est représenté en marron et les sites d'excision reconnus par la Flp recombinase en jaune.

Les régions d'homologie entre la cassette bi-cistronique RRPT (*galK*) et la séquence du fosmide est rajoutée de part et d'autre du rapporteur grâce aux queues flottantes des amorces d'amplification. La recombinaison de cette cassette se fait également grâce à la Red-recombinase. Cet évènement confère l'autotrophie au galactose grâce à l'expression du gène *galK*. Ce gène est ensuite excisé simplement par la Flp recombinase qui est induite par l'addition d'arabinose (Figure 46). De cette façon, un seul évènement de recombinaison est nécessaire à l'insertion de notre cassette RRPT dans le gène cible.



**Figure 46 : Principe de recombineering selon Tursun et ses collègues**

La cassette RRPT (galk) est amplifiée par PCR avec des amorces ayant des queues flottantes homologues aux sites d'insertion. Les séquences régulatrice du gène X sont en bleu claire, la séquence codante en bleu foncé. La croix rouge représente la recombinaison homologue. Le phénomène de recombinaison est sélectionné par Galactose. L'excision du gène galk (flèche marron) est induite par addition d'Arabinose et la Fip recombinase reconnait et excise ce qui se trouve entre les séquences spécifiques (ronds jaunes).

#### 4.1.2 Génération de lignées transgéniques exprimant la cassette RRPT

Nous avons retenu dans un premier temps la stratégie de recombineering pour générer des fosmides rapporteurs pour trois des gènes cibles du microARN *let-7* précédemment testés dans l'étude de 2006 menée par Lall et ses collègues [231]. Nous avons choisi ces trois gènes *lin-41*, *uba-1* et *nhr-17* pour débiter notre étude car ce sont les trois qui montrent une régulation post-transcriptionnelle forte au niveau du pharynx et de l'intestin respectivement.

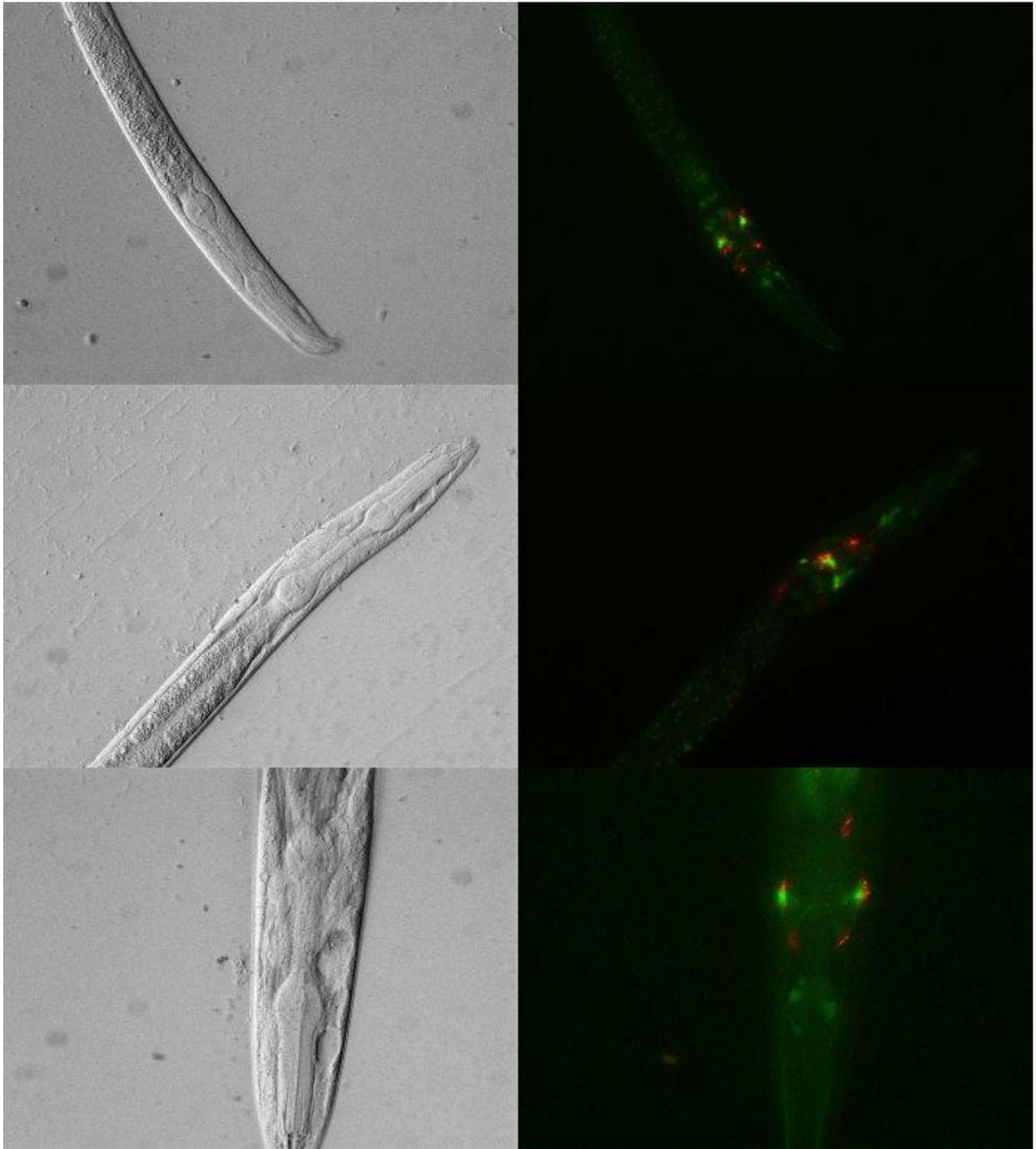
Nous avons donc créé les fosmides respectifs contenant la cassette RRPT (Figure 55). Nous les avons par la suite injectés dans les gonades de *C. elegans* afin de générer des lignées de nématodes rapportrices *in vivo* pour leur régulation post-transcriptionnelle. Nous observons que la cassette rapportrice que nous avons créée s'exprime *in vivo* dans les lignées transgéniques générées (Figure 47, Figure 48 et Figure 49).

Dans les lignées rapportrices pour le gène *nhr-17*, Nous observons une expression de la mCherry, qui dans la cassette est le rapporteur de l'activité transcriptionnelle, dans certains neurones de la tête qui ne sont pas identifiés individuellement au cours de cette étude, durant tous les stades post embryonnaires. La GFP qui est le rapporteur de l'expression post-transcriptionnelle est exprimée dans des neurones de la tête, différents de ceux exprimant la mCherry, dans tous les stades. Nous observons également une expression dans le pharynx durant le stade adulte (Figure 47).

Chez les vers transgéniques rapporteurs pour le gène *lin-41*, on observe une expression de la mCherry dans les neurones de la tête et dans les neurones du pharynx aux différents stades du développement. En revanche, la GFP n'est exprimée que dans le pharynx dans les stades larvaires et cette expression est absente au stade adulte (Figure 48).

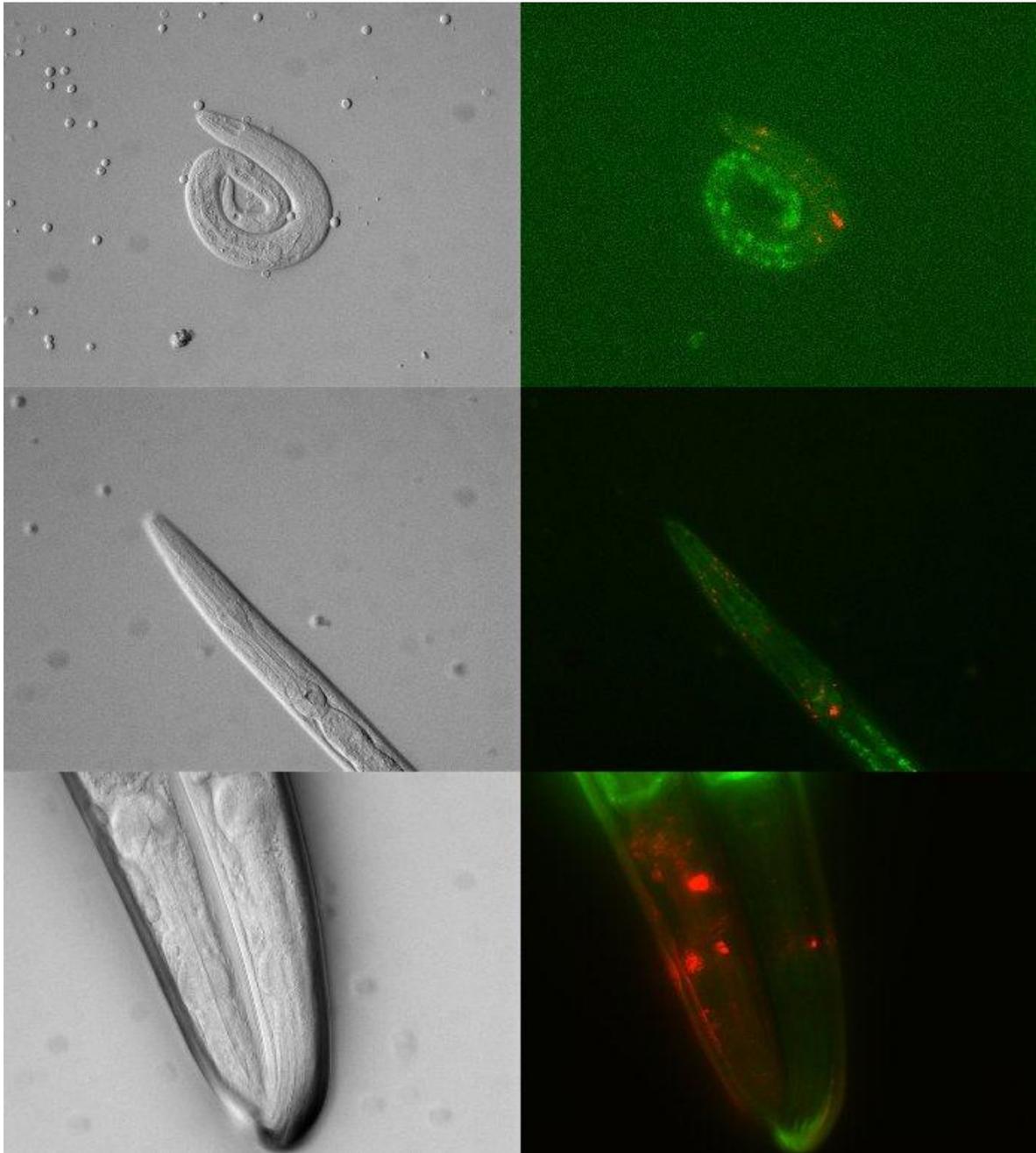
Chez les vers transgéniques rapporteurs du gène *uba-1*, nous observons une expression de la mCherry limitée à quelques neurones de la tête. La GFP en revanche est exprimée dans plusieurs neurones de la tête différents de ceux exprimant la mCherry et dans l'intestin de l'animal (Figure 49).

Nous nous attendions à ce que la fluorescence dans le rouge soit le reflet de l'activité transcriptionnelle du promoteur du gène étudié. La GFP est sensée quant à elle représenter la régulation post-transcriptionnelle du gène étudié liée à son 3'UTR. Cependant, les territoires d'expression de la GFP et de la mCherry ne semblent pas cohérents avec nos attentes. L'expression de la GFP dans des organes où nous ne voyons pas d'expression de la mCherry semble donc être inattendue pour un messager bi-cistronique issu de la cassette RRPT.



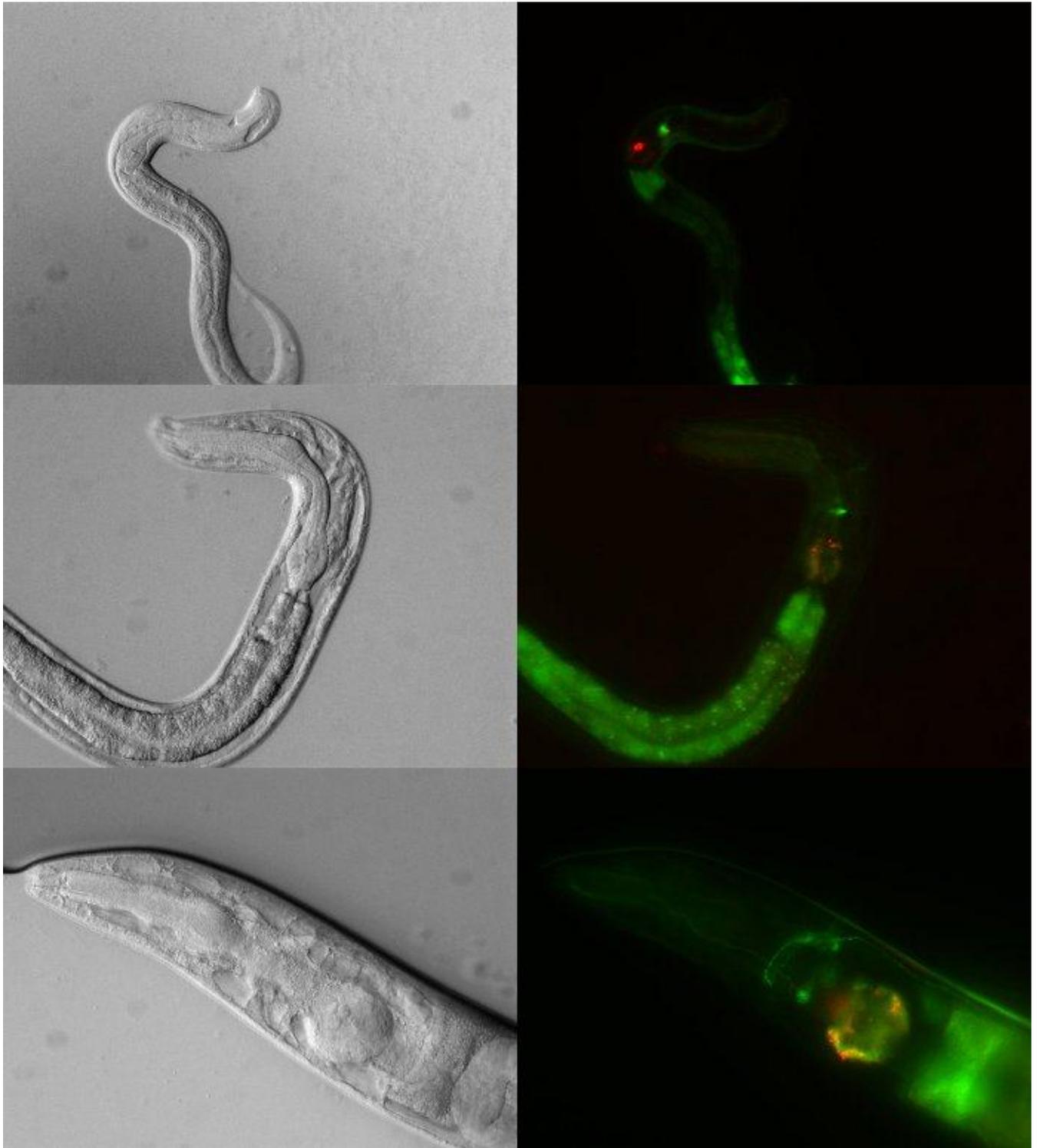
**Figure 47 : Lignée Rapportrice de la régulation Post-transcriptionnelle du gène *nhr-17***

La colonne de gauche représente des images de microscopie prises en DIC avec une exposition de 4 ms. La colonne de droite représente la juxtaposition entre l'expression de la GFP détectée avec une exposition de 500 ms et la mCherry détectée avec une exposition de 2s. Toutes les images ont été prises avec un objectif 40X. La ligne du haut représente un animal au stade L2, la ligne du milieu représente la partie antérieure d'un animal du stade L4 et la ligne du bas la tête d'un animal au stade adulte.



**Figure 48 : Lignée Rapportrice de la régulation Post-transcriptionnelle du gène *lin-41***

La colonne de gauche représente des images de microscopie prises en DIC avec une exposition de 4 ms. La colonne de droite représente la juxtaposition entre l'expression de la GFP détectée avec une exposition de 500 ms et la mCherry détectée avec une exposition de 1s. Toutes les images ont été prises avec un objectif 40X. La ligne du haut représente un animal au stade L1, la ligne du milieu représente la partie antérieure d'un animal du stade L3 et la ligne du bas la tête de deux vers au stade adulte.



**Figure 49 : Lignée Rapportrice de la régulation Post-transcriptionnelle du gène *uba-1***

La colonne de gauche représente des images de microscopie prises en DIC avec une exposition de 4 ms. La colonne de droite représente la juxtaposition entre l'expression de la GFP détectée avec une exposition de 500 ms et la mCherry détectée avec une exposition de 2s. Toutes les images ont été prises avec un objectif 40X. La ligne du haut représente un animal entier au stade L2, la ligne du milieu représente la partie antérieure d'un animal au stade L4 et la ligne du bas la tête d'un animal au stade adulte.

L'expression de la cassette RRPT dans les lignées de vers que nous avons générés s'est avérée non conforme à nos attentes. D'autres lignées réalisées par ma collègue Cécile Quéré portant la cassette RRPT exhibent le même problème.

### **4.1.3 Conclusion**

Lors de cette étude, nous avons souhaité générer une cassette adaptable à un système à haut-débit afin de valider expérimentalement les prédictions de régulation des gènes par les miARN via leur extrémité 3'UTR.

Dans toutes les lignées que nous avons générées, nous avons observé une expression de la GFP qui n'est pas co-localisée avec la mCherry. Notre rapporteur de la régulation post-transcriptionnelle ne semble donc pas se comporter comme nous l'avions envisagé. Pour comprendre l'origine de ce problème nous avons testé une deuxième approche pour pouvoir poursuivre l'étude que nous avons entreprise de la régulation post-transcriptionnelle chez *C. elegans*.

## **V. DISCUSSIONS ET PERSPECTIVES**

## 5.1 EPISSAGE ALTERNATIF DE *LET-2*

L'étude de l'épissage de *let-2* a servi de validation à notre stratégie expérimentale, il nous est possible de mener des études similaires sur d'autres lignées rapportrices bicolores. La majorité des mutants que nous avons isolés sont des allèles du gène *asd-2* qui est impliqué dans cet épissage. Et d'une manière intéressante deux des mutants isolés (m32 et m48) sont identiques aux mutants yb1421 et yb1542 qui ont été isolés par l'équipe de notre collaborateur au Japon ce qui tend à confirmer que le crible est efficace et qu'il atteint la saturation.

Nous n'avons en revanche pour l'instant pas réussi à identifier la mutation responsable des phénotypes des quatre mutants sauvages pour le gène *asd-2*. L'hypothèse qui nous paraissait la plus probable, étant donné la saturation du crible initial et le résultat de notre crible, est que ces 4 mutants sont des allèles d'un même gène. Nous avons vérifié cette hypothèse en analysant les résultats du séquençage du génome entier. Nous nous sommes intéressés aux mutations localisées dans les séquences codantes des gènes mais aucun gène n'est muté en commun chez les quatre mutants. Nous nous sommes donc intéressés aux régions non codantes afin de voir si un ARN non codant par exemple serait communément muté. D'une manière intéressante nous avons observé que les quatre mutants étaient mutés dans une région se situant dans un intervalle de 300 pb localisées à plus ou moins 4 Kb en amont du gène *F38C2.1*. La taille du génome de *C. elegans* est de  $100 \times 10^6$  pb et la mutagenèse EMS que nous avons réalisé a induit plusieurs centaines de mutations distribuées aléatoirement le long de chacun des génomes. La probabilité que 4 mutations soient aléatoirement localisées dans un intervalle de 300pb est donc extrêmement faible. Cela fait du gène *F38C2.1* un candidat intéressant à tester. Les 4 mutations peuvent être localisées dans une région régulatrice ou dans le promoteur de ce gène affectant son expression ce qui modifierait directement ou indirectement l'épissage de *let-2*.

Le gène *F38C2.1* est à ce jour sans fonction connue dans les bases de données. Nous avons effectué un alignement de séquence de ce gène contre la banque du NCBI afin de voir si un homologue est présent chez d'autres espèces pour lequel une fonction est connue mais l'alignement ne donne pas de résultat. Il serait donc intéressant de poursuivre notre étude par une étude fonctionnelle du gène *F38C2.1* afin de vérifier son implication dans l'épissage du gène *let-2*. Il serait intéressant par exemple de générer des lignées qui portent la GFP sous le contrôle de la région intergénique en amont de ce gène afin de vérifier son territoire d'expression et vérifier si elle

est co-localisée avec *let-2* dans le BWM. Il serait également utile d'injecter dans les mutants la séquence sauvage du gène avec un sa région intergénique non mutée ou sous le contrôle d'un promoteur spécifique du BWM tel que *myo-2* et voir si cela restaure le phénotype sauvage. Le gène *F38C2.1* pourrait être un régulateur du gène *asd-2* dont seul l'isoforme *asd-2b* est impliquée dans l'épissage de *let-2*. Il serait intéressant dans ce cas de vérifier la quantité d'ARNm *asd-2b* par PCR quantitative présente dans chacun des mutants.

Par ailleurs, bien que notre approche ne nous ait pas apporté d'informations nouvelles sur la mise en place des réseaux de régulation de l'épissage, l'utilisation de cette démarche sur d'autres lignées rapportrices devrait permettre de dévoiler les facteurs impliqués dans cette régulation. Le quart des gènes de *C. elegans* sont prédits pour être épissés alternativement. Les exemples d'épissages alternatifs étudiés dans cette thèse sont de types mutuellement exclusifs qui sont peu fréquents. Le type d'épissage le plus fréquent est le « cassette exon ». Or, la génération de lignées de *C. elegans* rapportrice de cet épissage semble plus aisée. Il suffirait de générer des minigènes dans lesquels la GFP s'exprime (ou au contraire ne s'exprime plus) lors de l'inclusion de l'exon. Ces minigènes seraient systématiquement co-injectés avec le gène de la RFP qui serait associé à un exon constitutivement exprimé et servirait de contrôle endogène [237]. Une telle approche permettra de conduire une étude systématique *in vivo* de l'épissage alternatif chez *C. elegans* et permettra de confirmer ou d'infirmer les prédictions avancées par Ramani et ses collègues qui stipulent que plus de 5000 gènes de *C. elegans* seraient en proie à de l'épissage alternatif [214].

## **5.2 MODIFICATION DU MODULE REFLX DU COPAS**

Nous avons modifié le COPAS Biosorter et nous avons été capables d'éliminer le problème de « carry-over » et de diminuer significativement le temps nécessaire au traitement d'une plaque 96 puits faisant de cette version du COPAS une plateforme adaptée à une étude à haut débit.

Karine Rébora a utilisé le module pour cribler la banque génomique RNAi à la recherche de facteurs impliqués dans l'épissage alternatif du gène *let-2*. Esthèr Marza, une chercheuse ayant rejoint le laboratoire pour finaliser une étude initiée dans l'équipe d'Eric Chevet (Bordeaux, France) a utilisé le COPAS pour cribler la banque RNAi à la recherche de facteurs impliqués dans le stress du réticulum endoplasmique chez *C. elegans*.

Ces deux études ont été l'occasion pour nous de tester et de valider à grande échelle les modifications que nous avons apporté au cytomètre. La ponction de certains composants présents dans la version native n'a pas altéré la fonctionnalité du COPAS et a permis de drastiquement diminuer le temps nécessaire pour mener une étude à échelle génomique. Nous pouvons désormais tirer pleinement profit de l'automatisation qu'offre le module Reflx du COPAS dans un temps plus court pour mener nos cribles futurs tels qu'un crible systématique pour identifier les facteurs impliqués dans la régulation post-transcriptionnelle médiée par la région 3'UTR des lignées bicolores rapportrices.

Une version commerciale inspirée de notre modèle est désormais proposée par le fabricant Union Biometrica. Le nouveau module Reflx est de plus en plus adopté par les laboratoires réalisant des cribles à haut débit pour s'affranchir des problèmes de « carry-over » et pour le gain de temps qu'il offre.

### **5.3 ETUDE DE L'ÉLÉMENT CIS RÉGULATEUR DE L'ÉPISSAGE DU GÈNE *EGL-15***

Nous avons montré lors de cette étude une corrélation entre l'affinité de la protéine SUP-12 à la région *cis* régulatrice *egl-15* et l'épissage alternatif de ce dernier.

Nous avons pu générer des lignées transgéniques de *C. elegans* portant différentes mutations ponctuelles sur la région *cis* régulatrice du gène *egl-15*. Nous avons montré après l'analyse quantitative de ces lignées grâce au COPAS que l'inclusion d'un exon et l'exclusion de l'autre est directement corrélée à la constante d'affinité de SUP-12 pour son site de fixation sur l'ARN.

Cette étude a été pour nous l'occasion de confirmer qu'une mutation ponctuelle dans un élément *cis*-régulateur entraînant une modulation de fluorescence peut être détectée par le COPAS même quand la modulation est subtile. Or, l'une des perspectives que nous envisageons pour l'étude de la régulation post-transcriptionnelle par les miARN est de muter les éléments en *cis* où ils se fixent et traquer une modulation de fluorescence qui serait la conséquence d'une restauration de l'expression de la fluorescence verte dans les endroits où elle est initialement réprimée. L'utilisation du COPAS pour cette étude permettrait une quantification que l'analyse microscopique ne permet pas.

## **5.4 GENERATION DE SYSTEME RAPPORTEUR DE LA REGULATION POST-TRANSCRIPTIONNELLE IN VIVO CHEZ C. ELEGANS**

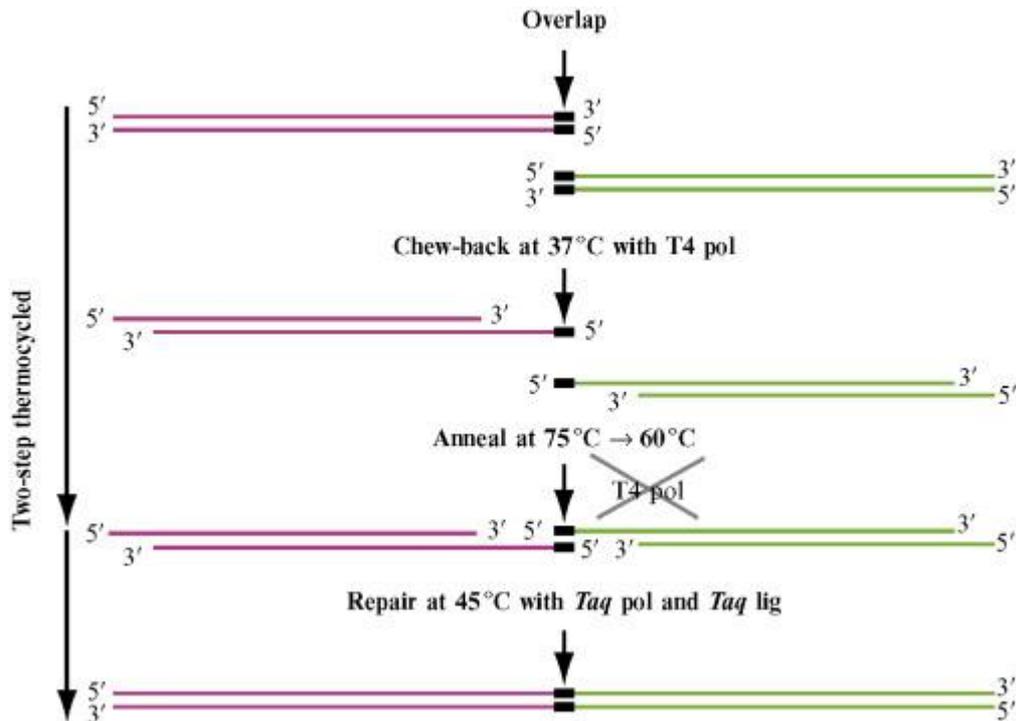
Lors de cette étude nous avons généré une cassette rapportrice de la régulation post-transcriptionnelle. Cette cassette nous l'avons intégré en lieu place de la séquence codante du gène cible en utilisant la technique de recombineering. Nous avons été capables de générer des fosmides rapporteurs pour trois gènes cibles de *let-7*.

Cependant, la technique de recombineering s'est avérée peu efficace. Seules 6 expériences sur Cinquante-deux réalisées ont donné lieu à des recombinaisons. En outre, plus de deux semaines sont nécessaires entre la première étape de recombineering et l'obtention des fosmides rapporteurs.

### **5.4.1 Clonage par assemblage de Gibson**

L'approche par recombineering a été jugée trop lente et peu adaptée à la génération de systèmes rapporteurs à haut débit car nous souhaitons tester environ 10% des gènes prédits pour être régulés par des microARN par PicTar. Il faudrait donc pouvoir réaliser plus de 200 constructions différentes en peu de temps.

Une approche beaucoup plus rapide et plus efficace que le recombineering a été adoptée au laboratoire. Il s'agit de la technique d'assemblage *in vitro* de Gibson (ou technique de « Chew back annealing ») [260, 261]. La réalisation de la fusion *in vitro* permet de s'affranchir des étapes intermédiaires de transformations dans des bactéries et extraction d'ADN pour la réalisation de la PCR en aval.



**Figure 50 : Technique d'assemblage Gibson**

Deux fragments d'ADN (Magenta et Vert) portant une séquence homologue (trait noir épais) sont soudés ensemble *in vitro* pour donner une seule molécule en une réaction en deux étapes. La première réaction consiste à digérer les extrémités 3' par la T4 polymérase puis les deux séquences complémentaires sont appariées. Les extrémités sont ensuite complétées puis liguées par la Taq polymérase et la Taq ligase respectivement [Gibson et al. 2011 [260]]

Cécile Quéré, étudiante en thèse ayant rejoint le laboratoire a utilisé cette technique pour générer dans un premier temps des cassettes rapportrices de deux gènes supplémentaires *lin-28* et *ZK757.1*. L'homologie nécessaire à la réalisation des constructions est apportée par les queues flottantes des amorces utilisées pour amplifier chaque fragment puis ceux-ci sont ensuite ligués en un seul fragment par la méthode Gibson (Figure 51).

Désormais, l'utilisation de la technique de l'assemblage selon Gibson a permis de considérablement diminuer le temps nécessaire (une seule journée est nécessaire pour réaliser les ligations) à la génération des constructions rapportrices. Cette technique est mieux adaptée à la réalisation des constructions en routines nécessaire à l'étude à échelle génomique que nous souhaitons réaliser.

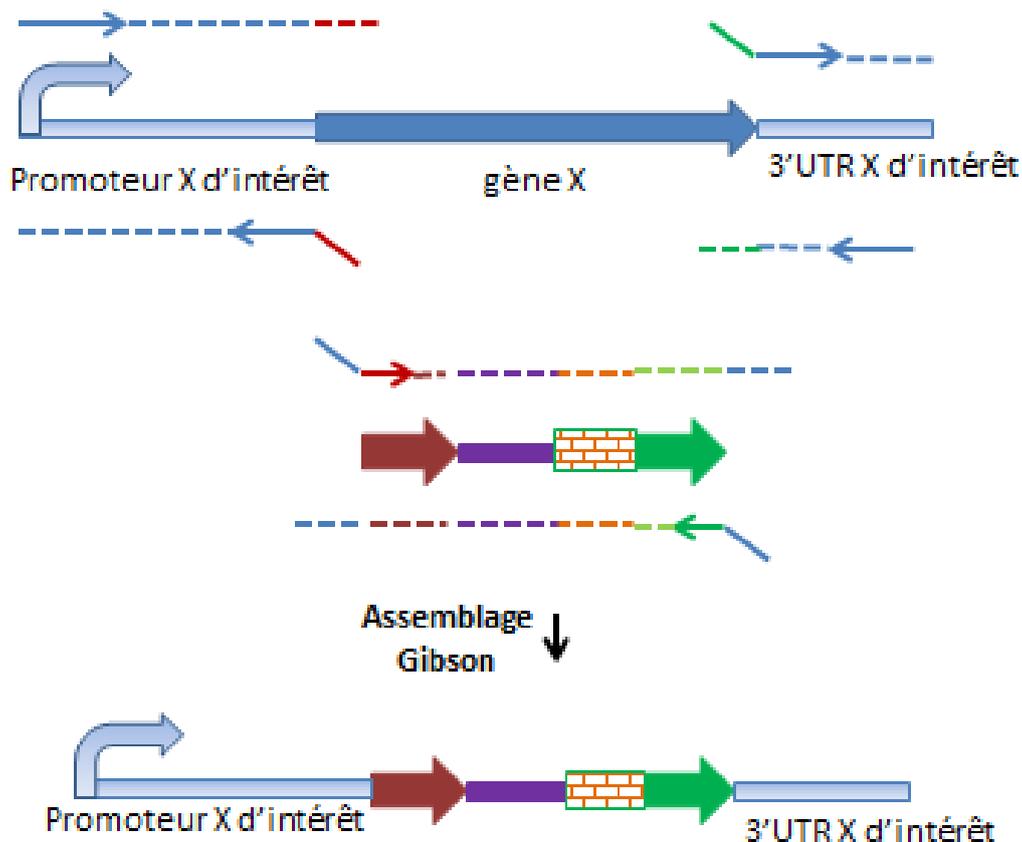


Figure 51 : Construction rapportrice du gène X en utilisant l'assemblage Gibson

Les problèmes d'expressions que nous avons observés dans le contexte des rapporteurs RRPT devront cependant être résolus pour pouvoir disposer d'un rapporteur adapté à notre étude et pouvoir la mener à terme.

#### 5.4.2 Identifier les raisons de l'expression inattendue

La divergence des territoires d'expression de la GFP et la mCherry *in vivo* est problématique pour l'utilisation de la stratégie RRPT que nous avons initiée. La première hypothèse que nous avons émise concerne un éventuel problème intrinsèque à la construction que nous avons choisi. L'utilisation d'un système bicistronique a été retenue pour engendrer une quantité équimolaire des deux ARNm de la GFP et la mCherry lors de la transcription. Si le *trans* épissage de la cassette RRPT ne se produit pas correctement ou si une activité promotrice endogène à « l'opéron linker » fait en sorte d'exprimer la GFP indépendamment du promoteur étudié cela se traduirait par un déséquilibre de la quantité des messagers et aboutir à une expression inattendue. Une étude fonctionnelle de l'opéron linker *gpd-2/gpd-3* montre cependant que celui-ci ne contient pas d'activité promotrice et ne permet pas à lui seul d'engendrer l'expression du gène situé en aval [211].

#### 5.4.2.1 Vérification d'absence d'activité promotrice de l'opéron linker *gpd2/gpd-3*

Afin de discriminer entre ces deux hypothèses nous avons testé une stratégie de co-injection consistant à co-injecter deux produits de PCR indépendants où la mCherry et la GFP sont chacune sous le contrôle d'un même promoteur du gène étudié et de l'UTR *unc-54* et l'UTR du gène respectivement. Les lignées générées expriment également la GFP et la mCherry dans des territoires différents (Figure 52).



**Figure 52 : Lignée rapportrice de la régulation post-transcriptionnelle du gène T10B11.6**

Le panel du haut montre une image d'un ver au stade L1 en DIC avec exposition de 5ms (image de droite) ou en juxtaposition de la GFP avec une exposition de 500ms et mCherry avec exposition de 1s (image de gauche). Le panel du bas représente deux adultes côte à côte où l'on voit la tête de l'un et la partie centrale de l'autre en DIC avec une exposition de 5ms (image de gauche) et juxtaposition de la GFP avec une exposition de 500ms et la mCherry avec une exposition de 1s (image de gauche). Toutes les images ont été prises avec un objectif 40X.

Il semble donc que les différences d'expression entre les deux rapporteurs ne soient pas liées à leur inclusion dans une unique cassette bicistronique RRPT mais il semble donc de plus en plus probable que l'explication de nos observations provienne soit d'une différence de stabilité des rapporteurs soit d'une régulation post-transcriptionnelle divergente entre l'UTR *unc-54* et les UTR endogènes que nous avons testés.

#### **5.4.2.2 Vérification de la stabilité des rapporteurs**

Afin de confirmer cette hypothèse, des constructions où la mCherry et la GFP sont sous le contrôle d'un même promoteur et d'un même UTR ont été réalisées par ma collègue Cécile Quéré et co-injectées pour générer des lignées rapportrices afin de vérifier leur expression. Les résultats préalables de cette étude montrent que la GFP et la mCherry sont parfaitement co-localisées et que donc la stabilité des rapporteurs n'est pas en cause. L'origine du problème pourrait venir donc de l'UTR *unc-54* qui ne serait pas parfaitement permissif dans les régions d'expression des gènes que nous avons testés.

#### **5.4.2.3 Régulation post-transcriptionnelle du gène *unc-54***

L'UTR supposé permissif du gène *unc-54* n'est vraisemblablement pas dénué d'impact sur les événements de la régulation post-transcriptionnelle de l'ARNm dans les zones d'expression des gènes étudiés. Malgré le fait que la région 3'UTR du gène *unc-54* a été utilisée à de nombreuses reprises comme contrôle négatif pour l'étude de l'impact de la régulation des régions 3' non traduites dans différentes études [231, 260-262] serait-il possible qu'une activité régulatrice soit présente dans nos conditions ?

Il serait intéressant de réaliser des études fonctionnelles de l'UTR *unc-54* grâce aux constructions que nous avons préalablement réalisées. La séquence de L'UTR peut être raccourcie jusqu'à retrouver une expression de la GFP co-localisée avec celle de la mCherry. Cela pourrait aboutir à identifier spécifiquement une région de l'UTR portant une séquence responsable d'une régulation post-transcriptionnelle. Par ailleurs, il serait également possible d'utiliser les lignées déjà générées pour réaliser une mutagenèse aléatoire afin de restaurer une co-localisation de la GFP et la mCherry. Le génome entier des lignées exhibant ce phénotype peut être séquencé afin d'identifier les gènes impliqués dans cette régulation post-transcriptionnelle.

Si l'UTR « contrôle » d'*unc-54* s'avère inadéquat comme contrôle endogène, afin de poursuivre notre étude, il faudra le remplacer dans la cassette RRPT par l'UTR endogène auquel on aurait muté les sites de fixation prédits des miARN pour vérifier si une régulation post-transcriptionnelle leur est imputée. L'autre alternative serait d'identifier une région 3' non traduite qui soit réellement neutre du point de vue de la régulation post-transcriptionnelle par exemple parmi celles des gènes exprimés de façon ubiquitaire dans toutes les cellules de l'organisme.

## **VI. MATERIELS ET METHODES**

## **6.1 PROCOLES DE *C. ELEGANS***

### **6.1.1 Culture et maintenance**

Les vers sont maintenus sur des boîtes de Petri sur un milieu de croissance de nématodes (nematode growth Medium ; NGM) inoculé avec les bactéries *E. coli* OP50 (auxotrophe pour l'Uracile) comme source de nourriture. La préparation de 1L de solution NGM agar nécessite 3 g de NaCl (Fischer Scientific) 2.5 g de peptone (BD), 12 g d'agar (Acros organics), 1 ml Cholestérol 95% (5 mg/ml) (Acros organics) et 975 ml d'eau. Le mélange est autoclavé puis mis à température ambiante sous agitation par aimant afin d'éviter sa gélification jusqu'à atteindre 50°C. 1 ml d'une solution CaCl<sub>2</sub> (1M) (Fischer Scientific), 1 ml de MgSO<sub>4</sub> (1M) (Fischer Scientific), 25 ml d'une Solution Potassium Phosphate (1M) sont ajoutés. 1L de la solution Potassium phosphate est obtenue par un mélange de 132 ml de K<sub>2</sub>HPO<sub>4</sub> (1M) et 868 ml de KH<sub>2</sub>PO<sub>4</sub> (1M) (Acros Organics) stériles par filtration. Le NGM est ensuite dispensé sur de boîte de Petri de 9 cm de diamètre et laissés à température pendant au moins 2 jours pour un séchage efficace. Quelques gouttes de culture OP50 sont ajoutées puis étalées sur la boîte où les bactéries vont se développer à température ambiante pendant 3-4 jours. Les boîtes peuvent ensuite être stockées en chambre froide (4°C) pendant plusieurs semaines.

### **6.1.2 Synchronisation des vers**

Des boîtes NGM contenant essentiellement des vers adultes avec des œufs dans leur gonades sont suspendus dans du M9. 1L de solution M9 est composé de 3 g de KH<sub>2</sub>PO<sub>4</sub>, 5.7 g de Na<sub>2</sub>HPO<sub>4</sub>, 5 g de NaCl et 1 ml de MgSO<sub>4</sub> (1M). Les nématodes sont culotés pour éliminer le surnageant et une solution de soude (0.5 M) et d'Hypochlorite de Sodium (5%) est ajoutée puis la suspension est vortexée à plusieurs reprises. Ce traitement tue les vers et dégrade leur cuticule mais les œufs ne sont pas sensibles au traitement. Lorsque la majorité des vers sont dégradés, les œufs sont culotés par centrifugation pendant 3 min à 5000 x g et le surnageant est éliminé. Les œufs dans le culot sont lavés trois fois avec du M9 pour éliminer les traces résiduelles d'hypochlorite de Sodium. Les œufs sont ensuite déposés sur des boîtes NGM pour permettre leur éclosion et développement.

### **6.1.3 Génération de mâles**

Il est possible de générer des mâles en ayant recours à des chocs thermiques. Mettre environ 10 vers hermaphrodites au stade L4 sur des boîtes 9 cm. Préparer une dizaine de boîtes et les mettre à l'étuve à 32°C

pendant 6h. Sortir les boites et les laisser à température ambiante pour permettre au vers d'engendrer une progéniture. Cribler les boites à la recherche de mâles sous la loupe binoculaire et les transférer sur une petite boite de 3 cm avec quelques L4 hermaphrodites pour maintenir un grand nombre de mâles dans la descendance.

#### **6.1.4 Croisement**

Pour procéder au croisement entre des mâles et des hermaphrodites il est nécessaire de disposer d'une culture riche en mâles (voir ci-dessus). Déposer 4 hermaphrodites jeunes L4 avec une dizaine de mâles jeunes adultes (ou fin de L4) sur une boite NGM de 3 cm. Le recours à une petite surface favorise les chances d'accouplement. Cribler la boite après 1 semaine pour vérifier la présence de mâles dans la descendance. Un taux élevé de mâles dans la descendance (théoriquement 50 %) est indicatif d'un croisement réussi. La totalité des L4 peuvent ne pas s'accoupler et la descendance sera donc hétérogène.

#### **6.1.5 Mutagenèse par EMS**

Plusieurs boites (3, 4 boites de 9 cm) contenant des nématodes synchronisés sont maintenues jusqu'à ce que le stade L4 soit fraîchement atteint (réaliser une mutagenèse plus tôt engendrera moins de génomes indépendants et la réaliser plus tard sera inefficace). Resuspendre les animaux dans du M9 et les récupérer dans un tube de 15 ml. Précipiter les jeunes L4 en centrifugeant 1 min à vitesse maximale et éliminer le surnageant. Laver deux fois le culot de vers avec du M9 pour le débarrasser des bactéries résiduelles et éliminer tout le surnageant excepté 2 ml. Sous hotte chimique préparer une solution d'Ethyle methanesulfonate (EMS) avec des pipettes, des tubes et une poubelle dédiée à cause des risques associés à ce composé chimique. 20 µl de la solution EMS commerciale sont ajoutées à 2 ml de solution M9 pour créer une solution EMS à 0.1M. Le tube est retourné plusieurs fois jusqu'à ce que la solution soit homogène et limpide. Les 2 ml de la solution EMS (0.1 M) sont ajoutés aux 2 ml de suspension de vers L4 pour une solution mutagène finale EMS à 0.05 M. Le tube est parafilmé et ensuite mis sous agitation à 20°C pendant 4h. Les vers sont par la suite culotés par centrifugation et le surnageant de la solution de mutagenèse est éliminé dans une poubelle liquide dédiée. Les vers sont lavés 3 fois au M9 pour éliminer toute trace résiduelle d'EMS puis culottés à nouveau. Les vers sont ensuite transférés avec une pipette pasteur sur une boite NGM avec des bactéries OP50 pour leur permettre de récupérer pendant 2h. 5 à 10 vers L4 (génération P0) traités à l'EMS sont ensuite repiqués sur une grande boite 15 cm de NGM avec des OP50. On les laisse se reproduire pendant 10 à 15 jours sur leur boite et les vers contenus sont ensuite criblés au COPAS pour isoler des mutants.

### **6.1.6 Injection des vers**

L'injection des vers a été réalisée telle que décrite dans le protocole de Mello et ses collègues en 1991 [262] en utilisant le marqueur de sélection antibiotique développé au sein de notre laboratoire par Rosina Giordano [263]. De jeunes adultes sont immobilisés dans une goutte d'huile sur lame d'agarose puis une solution d'ADN est injectée dans la partie syncytiale distale des gonades. Les vers sont ensuite soumis à la sélection sur des boîtes NGM supplémentées de l'antibiotique G-418. La liste des lignées générées et leur génotype est présentée en Annexe.

### **6.1.7 Microscopie**

Les vers sont déposés sur des lames recouvertes d'un tapis d'une solution d'agarose à 5%. Une goutte d'azoture de sodium à 10 mM est déposée sur le tapis et joue le rôle d'anesthésique. Un ver est ensuite déposé sur la goutte et une lamelle est déposée par-dessus. Les photos de fluorescence et Nomarski sont prises avec le microscope droit Axio imager Z1 (Zeiss) avec un objectif 10X, 40X ou 63X en immersion.

## **6.2 PROTOCOLES DE BIOLOGIE MOLECULAIRE**

### **6.2.1 PCR**

Les amorces utilisées ont toutes été commandées chez Sigma Aldrich. La liste des amorces utilisées et leur séquence est présentée en Annexe. L'enzyme utilisée pour l'amplification est la fusion Taq DNA polymérase (New England Biolabs) pour les PCR de vérification et La FusionTaq 6X (Finnzyme) pour les clonages et recombineering, suivant les recommandations des fabricants. Pour les PCR d'amplification le programme utilisé est le suivant : 1 cycle de dénaturation initiale de 5 min, suivi de 35 cycles d'amplification avec une dénaturation de 20 sec, hybridation des amorces de 20 sec et élongation d'une durée dépendante de la taille du fragment. Et un cycle final d'élongation de 5 min. les produits de PCR sont vérifiés sur gel d'agarose par électrophorèse et sont purifiés sur colonne QIAquick PCR purification kit (QIAGEN). La PCR pour amplifier les cassettes de sélection et RRPT dans le protocole de recombineering selon Dolphin et Hope a été menée telle que décrite dans leur protocole [257].

### **6.2.2 Préparation d'ADN génomique**

L'extraction de l'ADN génomique des nématodes est réalisée avec le kit NucleoSpin® Tissue (Macherey-Nagel). 3 à 4 boîtes de culture de nématodes

de 9cm de diamètres contenant essentiellement des vers adultes sont rincées avec de l'eau mQ stérile. La suspension de vers est récupérée dans des tubes de 50ml et déposées à 4 °C pour les laisser sédimenter pendant 15 à 60 min. le surnageant est éliminé et le culot de vers est suspendu dans 500µl de tampon de pré-lyse T1 contenu dans le kit. Pour fragiliser les tissus, la suspension est congelée pendant une nuit à -80 °C. Les vers sont décongelés par la suite et 50 µl de protéinase K sont ajoutés pour permettre la lyse. Le tout est incubé pendant 18 à 24h à 56 °C. L'échantillon est ensuite vortexé et incubé à 56 °C à plusieurs reprises pour dégrader les débris au maximum. Les tubes sont ensuite centrifugés pour culotter les débris qui sont susceptibles d'obstruer la membrane du Kit de purification. A partir de cette étape, la suite du protocole se déroule selon les recommandations du fabricant.

### **6.2.3 Production de bactéries électrocompétentes**

Les bactéries sont inoculées dans un Erlenmeyer contenant 500ml de solution de culture bactérienne L.B et sont mises en culture sous agitation à 32 °C ou 37 °C jusqu'à ce que la densité optique atteigne une valeur de 0.8. La culture est transvasée dans des tubes de 50ml à 4 °C pendant 2h puis les bactéries sont culottées par centrifugation à 4000 x g à 4 °C. Le surnageant est éliminé et le culot subit des bains successifs d'eau froide à 4 °C suivis de centrifugation pendant 15min à 4000 x g à 4 °C. Le culot subit ensuite deux bains successifs de glycérol 20% à 4 °C et centrifugé à 4000 x g 4 °C pendant 15 min. La totalité du surnageant est éliminé excepté 500µl où le culot est resuspendu. Des aliquotes de 100µl sont réparti dans des tubes 1,5ml et peuvent servir immédiatement pour l'électroporation ou être stockés à -80 °C.

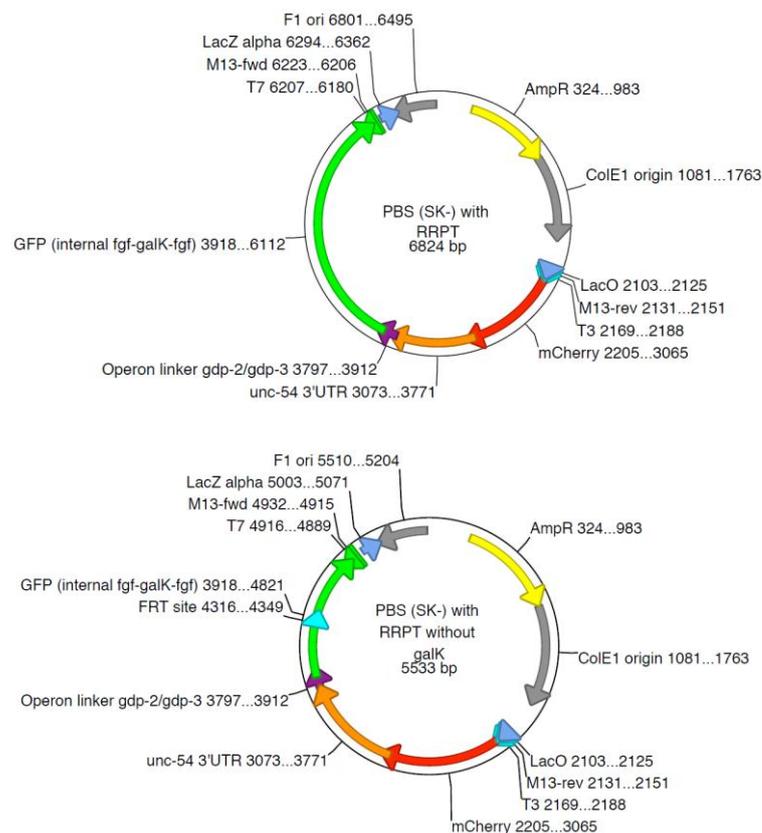
### **6.2.4 Electroporation d'ADN**

L'ADN sous forme de cassette amplifiée par PCR (100 – 150 ng), dans un plasmide (100 pg) ou dans un fosmide (50 ng) dans 1 à 2µl sont introduits dans les bactéries compétentes par électroporation à 2.3Kv dans des cuvettes destinées à cet effet. Les bactéries sont ensuite transférées dans 1ml du milieu L.B sans sélection et mises en culture de 20min à 3h sous agitation pour récupération. Ces cultures sont ensuite étalées sur des boites contenant du L.B agar additionnées de la sélection appropriée et sont mises pour incubation à 32 – 37 °C entre 18 et 48h.

### **6.2.5 Construction de cassette RRPT**

La cassette rapportrice de la régulation post-transcriptionnelle (RRPT) a été clonée dans le vecteur de clonage pBluescript (SK-) (pBS (SK-)) grâce à son multi site de restriction. La mCherry a été amplifiée par PCR en ajoutant le site de restriction SacI en amont et NotI en aval (amorces 143/64). La PCR

du 3'UTR *unc-54* contient les sites *NotI* en amont et *XbaI* en aval (amorces 65/84). La PCR de l'opéron linker *gpd-2/gpd-3* les sites *XbaI* et *SmaI* (amorces 67/68) et celle de la GFP permet d'introduire *SmaI* et *PstI* (amorces 69/144). Les produits de PCR sont purifiés et digérés par les enzymes adéquats (NEB biolabs) pour être insérés un par un dans le multi site de clonage du plasmide pBS (SK-) digéré par les mêmes enzymes et purifié. Les différents fragments insérés comportent des sites de restriction utilisés pour digérer les uns les autres, la seule combinaison possible consiste à insérer en premier lieu l'opéron linker *gpd-2/gpd-3* puis la mCherry suivie de l'UTR *unc-54* et enfin la GFP (Figure 53). Les fragments sont ensuite ligués par la T4 ligase (NEB biolabs). Le produit de ligation est ensuite transformé dans des bactéries compétentes DH5alpha. Les bactéries transformées sont sélectionnées sur des boîtes LB agar additionnées d'Ampicilline (Amp ; 100 µg/ml). L'insertion est ensuite vérifiée par PCR. La construction de la cassette RRPT finale a ensuite été vérifiée par séquençage (la cassette est amplifiée avec les amorces 77/78 localisées sur le pBS ; les amorces 64, 65, 67, 68, 69, 77, 78, 84, 143 et 144 pour le séquençage).



**Figure 53 : Carte du plasmide pBS-RRPT**

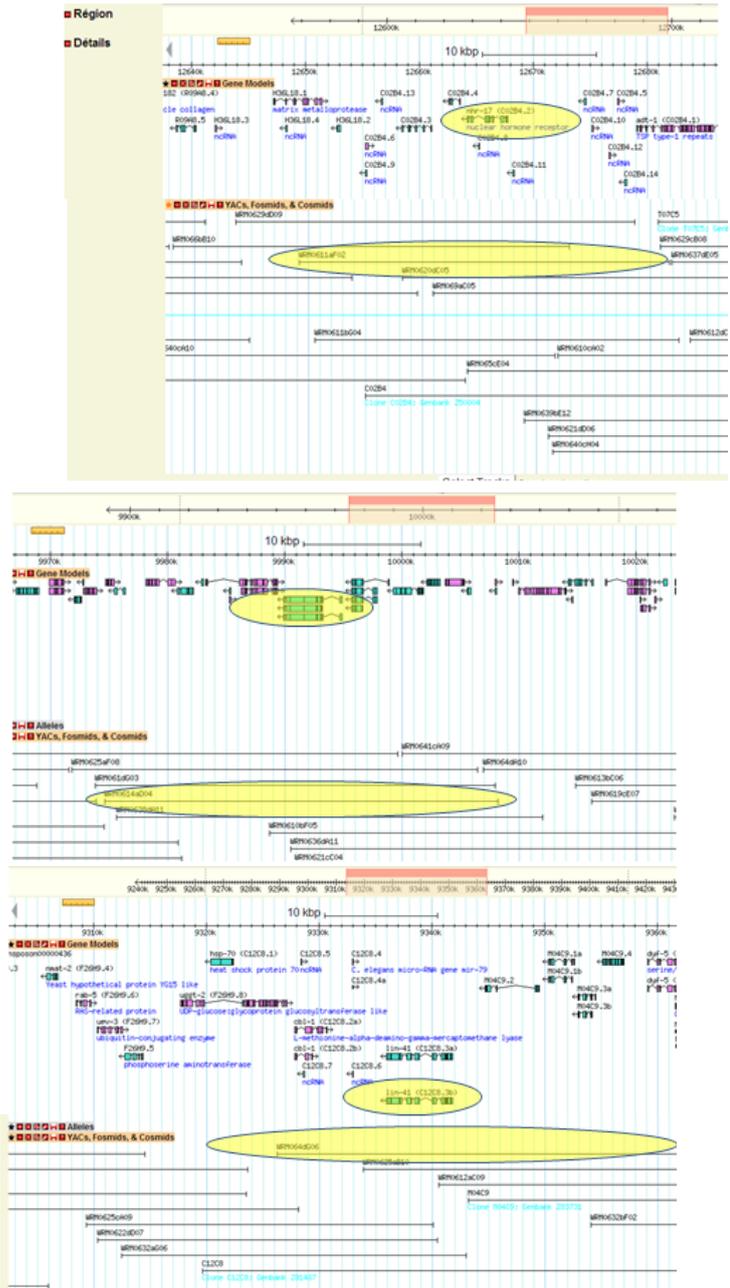
La carte du plasmide pBluescript (SK-) annotée portant la cassette rapportrice RRPT avec le gène de sélection *galK* inclus dans la GFP (plasmide du haut) ou excisé (plasmide du bas).

### **6.2.6 Plasmides rapporteurs de *egl-15***

Le vecteur d'expression BGAR contenant le rapporteur de l'épissage alternatif de *egl-15* [227] provient de notre collaborateur Hidehito Kuroyanagi (Tokyo, Japon). Les mutations dirigées ont été réalisées par notre collaborateur Cameron Mackereth (Bordeaux, France). Avant injection des plasmides chez *C. elegans* ceux-ci ont été envoyés pour séquençage (amorce 315) pour vérifier la présence de la mutation induite chez MilleGene (Labrège, France).

### **6.2.7 Sélection des fosmides**

Le choix entre les différents fosmides disponibles portant le gène d'intérêt a été fait grâce à l'explorateur de [www.wormbase.org](http://www.wormbase.org). Le fosmide portant le gène en milieu de séquence a été retenu (Figure 54). Les fosmides sont commercialisés en France par Biovalley (Marne-la-vallée, France).



**Figure 54 : Choix des fosmides portant le gène d'intérêt dans l'explorateur wormbase**

La position relative du gène par rapport aux différents fosmides est présentée dans ces captures d'écran. Le fosmide WRM0611aF02 portant le gène *nhr-17* en milieu de séquence a été retenu (figure du haut). Le fosmide WRM0614aD04 portant le gène *uba-1* en milieu de séquence a été sélectionné (figure du milieu). Le fosmide WRM064dG06 portant le gène *lin-41* a été sélectionné. Les gènes et les fosmides sélectionnés sont surlignés en jaune.

## 6.2.8 Extraction des fosmides

Les fosmides étant en copie unique par bactérie, il est nécessaire de préparer un grand volume de culture pour les extraire. Une colonie de bactérie contenant le fosmide d'intérêt est mise en culture dans 600ml de milieu LB contenant le marqueur de sélection approprié. La culture est ensuite incubée sous agitation jusqu'à ce que la densité optique soit comprise entre 0.8 et 1. L'extraction des fosmides se fait par la suite grâce

au kit Nucleobond AX Xtra Midi/Maxi (MACHEREY-NAGEL) suivant les recommandations du fabricant.

## **6.2.9 Recombineering**

### **6.2.9.1 Recombineering Dolphin & Hope**

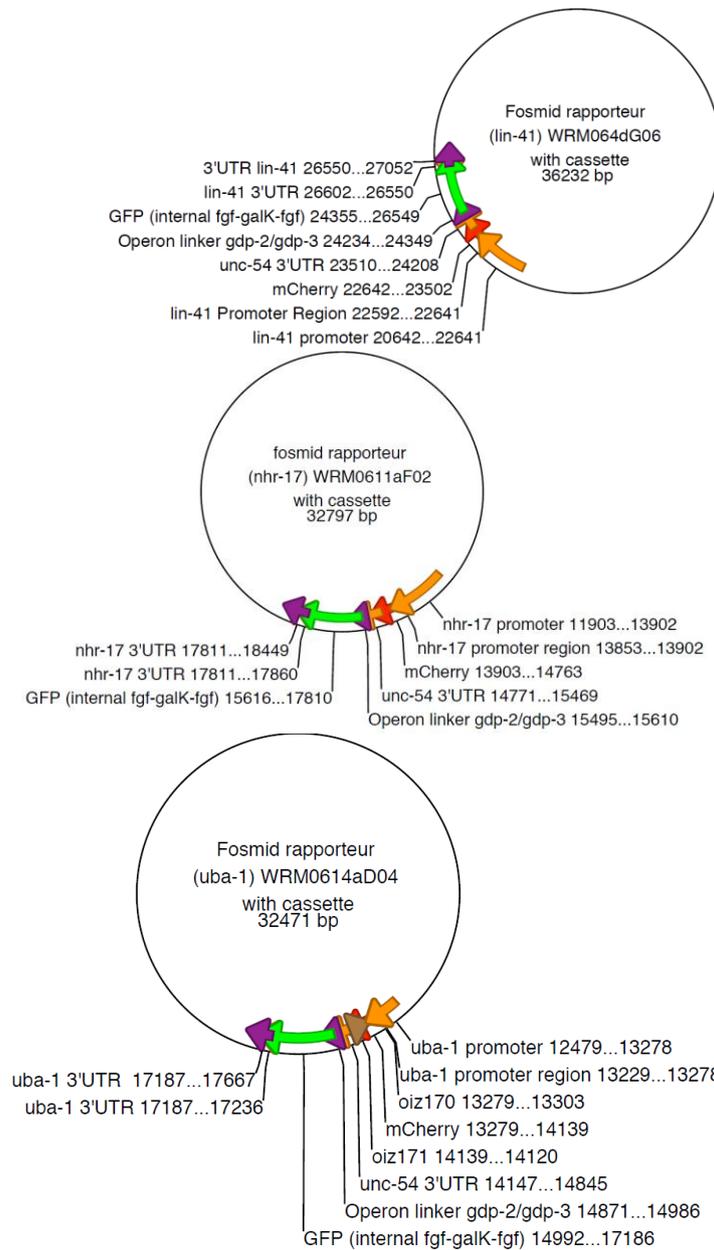
Une colonie bactérienne EL350 contenant le fosmide d'intérêt est mise en culture dans 1 ml dans du milieu SOB[-mg] additionné de Chloramphénicol (Cm, 12.5 µl/ml) à 32 °C à 220 r.p.m.) pendant une nuit. La culture O/N est ensuite utilisée pour inoculer 100 ml du même milieu avec Cm (12.5 µl/mg) jusqu'à ce que la densité optique atteigne 0.6. 50 ml de la culture est ensuite transférée dans un Erlenmeyer 500 ml stérile préchauffé à 42°C sous agitation pendant 20 min supplémentaires afin d'induire l'expression de la Red recombinase. Chacune des deux cultures est ensuite mise sur glace à 4°C et des bactéries électro compétentes sont préparées. Les aliquotes de 100 µl exprimant la Red recombinase ou non (contrôle) sont électroporés avec 100-150 ng de la cassette RT de sélection et contre sélection *rpsL tetA* amplifiée par PCR et portant une homologie de 50nt avec le site d'insertion du fosmide (couples d'amorces 71/72, 73/74 et 75/76 pour *uba-1*, *lin-41* et *nhr-17* respectivement). Les bactéries sont ensuite resuspendues dans 1 ml de milieu SOB [-mg] 2 à 3h, les bactéries sont ensuite culotées et resuspendues dans 100 µl de milieu. Les échantillons sont ensuite étalés sur des boîtes LB agar contenant de la Tétracycline (Tc ; 5 µg/ml). Les boîtes sont ensuite mises à l'étuve pendant 48 à 72 h à 32°C pour l'apparition de colonies.

Une colonie est ensuite mise en culture dans du milieu SOB[-mg] pour préparer des bactéries électro compétentes. La cassette bicistronique RRPT est amplifiée par PCR et porte une homologie de 50nt avec le site d'insertion dans le fosmide (couple d'amorces 123/124, 164/165, 166/167 pour *nhr-17*, *lin-41* et *uba-1* respectivement). L'utilisation des amorces 63, 70, 141, 142 a été envisagée pour qu'elles soient universelles à la deuxième étape de recombinaison en gardant une cicatrice de 50 nt de RT de part et d'autre de la cassette RRPT). 100 à 150 ng du produit PCR sont électroporés dans les bactéries compétentes issues de la première étape de recombineering induites ou non induites pour la Red recombinase. Le protocole se poursuit par la suite comme décrit pour la première étape excepté que les bactéries sont étalées sur des boîtes contenant du Chloramphénicol (Cm ; 12.5 µg/ml) et de la Streptomycine (Str ; 500 µg/ml).

### **6.2.9.2 Recombineering Tursun et al.**

Le recombineering s'opère dans la souche bactérienne *E. coli* SW105 qui peut être obtenue chez Biological Resource Branch NCI (<http://recombineering.ncifcrf.gov>). Une colonie de la souche SW105 transformées préalablement avec le fosmide approprié est mise en culture à

32°C dans 10 ml du milieu LB (Cm 12.5 µg/ml) pendant une nuit. 1 ml de cette culture est inoculé le lendemain dans un Erlenmeyer de 1L stérile contenant 100 ml de LB (Cm 12.5 µg/ml). La culture est mise sous agitation à 32°C jusqu'à ce que la densité optique soit comprise entre 0.4 et 0.6. L'Erlenmeyer est ensuite incubé pendant 20 min à 42°C dans un bain Marie préchauffé pour induire l' $\lambda$  Red Recombinase. L'Erlenmeyer est ensuite transféré immédiatement dans de la glace à 4°C pendant 20 min afin de préparer des bactéries compétentes. Les bactéries sont ensuite transformées par électroporation avec 50 ng du produit de PCR purifié de la cassette RRPT contenant la galK avec l'homologie au site d'insertion (un aliquote de bactéries est transformé avec de l'eau sert de contrôle négatif). Les bactéries sont ensuite récupérées dans 1.8 ml de LB dans des tubes de 2 ml et sont laissés 1h à température ambiante sous agitation. Les bactéries sont ensuite culottées par centrifugation (1 min à 19000 x g) et le surnageant est éliminé. Le culot est lavé avec 1 ml d'eau mQ stérile afin d'éliminer tout le LB résiduel et culoté à nouveau. Le culot est ensuite déposé sur des boîtes de culture avec un milieu minimal (Cm 12.5 µg/ml) supplémenté de Galactose comme seule source nutritionnelle. Les boîtes sont ensuite incubées à 32°C. Les colonies ne sont apparues sur les boîtes qu'au bout d'une à deux semaines d'incubation. L'insertion de la cassette RRPT au locus est vérifiée par PCR dans les colonies ayant poussées. Celles-ci sont ensuite re-striées sur une boîte Galactose (Cm 12.5 µg/ml) et incubées 4 à 5 jours supplémentaires car lors des vérifications, toutes les colonies ne contiennent pas la cassette RRPT. L'insertion de la cassette RRPT est à nouveau vérifiée par PCR et la colonie correspondante est incubée à 32°C dans 1 litre de LB (Cm 12.5 µg/ml) jusqu'à ce que la densité optique soit entre 0.4 et 0.6. La culture est ensuite divisée en deux dans des Erlenmeyers stériles et 1/100 de volume d'une solution 10% Arabinose est ajoutée à l'une tandis que l'autre sert de contrôle négatif de non excision. Les deux cultures sont incubées sous agitation à 32°C pendant 3h (peuvent également être incubées O/N à température ambiante) et les fosmides sont extraits grâce par le protocole de Maxi-prep. L'excision du gène galK est vérifiée par PCR sur les fosmides et se traduit par une diminution de taille de 1.3 Kb entre les fosmides issues de la culture induite et ceux de la culture non induite. La cassette RRPT contenue dans le fosmide est ensuite séquencée (Figure 55).



**Figure 55 : Cartes des fosmides recombinés portant la cassette RRPT**

La cassette RRPT remplace la phase codante du gène étudié dans son fosmide. La position des éléments composant la cassette est annotée.

## 6.3 COPAS BIOSORTER

### 6.3.1 Analyse des vers au COPAS biosorter

Une à plusieurs boîtes contenant une population dense (environ 20000 vers) non affamée sont resuspendues dans de l'eau. La suspension est ensuite versée dans la coupe à échantillon du COPAS. Quand la suspension est trop concentrée en vers, ceux-ci sont dilués pour avoir un nombre de

vers analysés inférieur à 300 vers/seconde. Les paramètres PMT pour les lignées rapportrices de *let-2* et *egl-15* sont de 550 pour le Green et 700 pour le Red. Les lignées portant la cassette RRPT sont analysées avec des PMT de 650 pour le Green et 950 pour le Red).

## **VII. ANNEXE**

## Liste des amorces

Numéro d'amorce	Auteur	Nom d'amorce	SEQUENCE	tm
1	iz	lin41_P_F	GGACTAGTGGGCCCGTGTAGAGAGTGGGAACG	82
2	iz	lin41_P_Gr_R	CTTTACTCATTTCACITTTTCCAAGTCTGA	67
3	iz	T14B1.1_P_F	GGACTAGTGGGCCCTTTAATTTGCATTTCAAC	75
4	iz	T14B1.1_P_Gr_R	CTTTACTCATTCTCTGAAATTGTCGTATAA	63
5	iz	unc129_P_F	GGACTAGTGGGCCCTTATCAACATGTTCCCAT	80
6	iz	unc129_P_Gr_R	CTTTACTCATTCTTCTGCTTCTTCCAA	70
7	iz	nhr4_P_F	CTCTAGAGGGCCCGCATTGAAACGCATAGGTG	80
8	iz	nhr4_P_Gr_R	CTTTACTCATTTACACTGAAATCAAAAATA	62
9	iz	C35E7.4_P_F	GGACTAGTGGGCCCTCTCTGTTTCTGAAATAAC	75
10	iz	c35E7.4_P_Gr_R	CTTTACTCATGATTGAATCAAGTTGGGGA	70
11	iz	ceh16_P_F	CTCTAGAGGGCCCGCAGCCTTGTAGGCAGGCAGG	83
12	iz	ceh16_P_Gr_R	CTTTACTCATTCTGAAAAATAAATAAAA	60
13	iz	ccr4_P_F	GGACTAGTGGGCCCGTCGATAGCTCGTGTAAAC	81
14	iz	ccr4_P_Gr_R	CTTTACTCATCTTGTCTGATCTTCTCTCGCT	70
15	iz	dpy2_P_F	CTCTAGAGGGCCCAATATAGCAGTATTCTGA	71
16	iz	dpy2_P_Gr_R	CTTTACTCATCGCTGAAAGTTTCATATTG	67
17	iz	sma1_P_F	GGACTAGTGGGCCCTTTCTCCGGTCTTCATT	80
18	iz	sma1_P_Gr_R	CTTTACTCATGATAAAGCTTCACAATATTA	62
19	iz	uba1_P_F	GGACTAGTGGGCCCTAGTACGCCCTGGAAGCCG	84
20	iz	uba1_P_Gr_R	CTTTACTCATTAAATCACGAAAGATCAGAAC	65
21	iz	oig2_P_F	GGACTAGTGGGCCCACTGGATACAATTATTTTT	73
22	iz	oig2_P_Gr_R	CTTTACTCATTCTGAAACTAACCTTTTAGA	62
23	iz	nhr17_P_F	GGCTAGCGGGCCCTCATTACTATATATAAAGC	72
24	iz	nhr17_P_Gr_R	CTTTACTCATTACTGAAATGAAATAATGAA	62
25	iz	lin41_U_Gr_F	ATACAAATAGACACTTTCTTCTGCTCTTT	63
26	iz	lin41_U_R	GGGGTACCGCACAGGGAAGGGGCGGAGC	86
27	iz	T14B1.1_U_Gr_F	ATACAAATAGCTGTTATCTCTCTCCAGA	67
28	iz	T14B1.1_U_R	GGGGTACCGCCGCGATTGTTCCGACCA	85
29	iz	unc129_U_Gr_F	ATACAAATAGGTTTATTATTTTCTTCTCGT	60
30	iz	unc129_U_R	GGGGTACCTTTAACTGACCTCTAAAATG	66
31	iz	nhr4_U_Gr_F	ATACAAATAGTGTGTTAATTTTCGGTTTTG	64
32	iz	nhr4_U_R	GGGGTACCAAAAAAAGGAATTTATTTT	66
33	iz	C35E7.4_U_Gr_F	ATACAAATAGTCTTTGTTTAGTGTGTAC	59
34	iz	C35E7.4_U_R	GGGGTACCTAGAATGTTTTTTTCTGA	68
35	iz	ceh16_U_Gr_F	ATACAAATAGAATATTGATTTTTGTTAGGC	62
36	iz	ceh16_U_R	GGGGTACCGAAAGACCTTCTACGAAAAA	71
37	iz	ccr4_U_Gr_F	ATACAAATAGGTGAGAAAATATACTTTTTT	59
38	iz	ccr4_U_R	GGGGTACCAATTTTATGTCGATTTCAA	69
39	iz	dpy2_U_Gr_F	ATACAAATAGTTGATTCTCTCTGCATCCTT	66
40	iz	dpy2_U_R	GGGGTACCGGTATACAGTCGAATGGGCG	77
41	iz	sma1_U_Gr_F	ATACAAATAGATACCTCACCACACGCTGAT	68

42	iz	sma1_U_R	GGGGTACCATTTTGGAGATGAATAACATA	66
43	iz	uba1_U_Gr_F	ATACAAATAGTTATTATGATTTTTCAATGT	58
44	iz	uba1_U_R	GGGGTACCATTTTGTCAACTTGTTA	66
45	iz	oig2_U_Gr_F	ATACAAATAGATTCTGCTAGATTTCTTTAA	60
46	iz	oig2_U_R	GGGGTACCACACTTCATAAAAAGTCTTCA	68
47	iz	nhr17_U_Gr_F	ATACAAATAGAAAATTACATATTTATGATG	57
48	iz	nhr17_U_R	GGGGTACCCTTGCTGACAGCTCAAATGT	75
49	iz	unc54_U_Rd_F	ATTGTATAAGGTCCAATTACTCTTCAACAT	64
50	iz	unc54_U_R	CGGGGCCCGGTACCAACAGTTATGTTTGGTATA	79
51	iz	GFP_F	ATGAGTAAAGGAGAAGAACT	51
52	iz	GFP_R	CTATTTGTATAGTTCATCCA	49
53	iz	RFP_F	ATGAGTAAAGTCTCAAAGGGTGAAGAAG	66
54	iz	RFP_R	CTTATACAATTCATCCATGC	54
55	iz	Adaptator_F	TCGACAACCTTTATTATACAAAAGTTGCCCGGTACCAGATCTCGAGGGCCCACTAC	85
56	iz	Adaptator_R	TGGGCCCTCGAGATCTGGTACCGGGGACAACCTTTGTATAATAAAGTTG	84
57	iz	Operon Linker_F	GGGGTACCTTCATCTTATTGAATCTGCC	70
58	iz	Operon Linker_R	CTTTGAGACTTTACTCATGATGCGTTGAAGCAGTTTCC	77
59	iz	Unc-54(XhoI)_R	CCGCTCGAGAAACAGTTATGTTTGGTATA	70
60	iz	DD03(8026-8045)_F	AAGAAGTGGCTGATCTCAGC	61
61	iz	T3 prom Primer_R	ATTAACCCTCACTAAAGGGA	57
62	iz	Unc54_U_Gr_F	TGGATGAATATACAAATAGGTCCAATTACTCTTCAACAT	72
63	iz	mCherry-RTcassette_F	GGAGCTCTCGCTGTCGAGATATGACGGTGTTCACAAAGTTCCCTAAATTTACTTTTGGATGGTCTCAAAGGGTGAAGAAGATA	88
64	iz	mCherry_R (NotI)	AGCGGCCGCTTATACAATTCATCCATGC	76
65	iz	unc54_F (NotI)	AGCGGCCGCTCCAATTACTCTTCAACAT	78
66	iz	unc54_R (XbaI)	GTCTAGAAACAGTTATGTTTGGTATA	56
67	iz	Operon Linker_F (XbaI)	GTCTAGATTATCTTATTGAATCTGCC	63
68	iz	Operon Linker_R (SmaI)	GTCAGTCCCGGGATGCGTTGAAGCAGTTTCC	82
69	iz	GFP_F (SmaI)	GTCAGTCCCGGGATGAGTAAAGGAGAAGAACT	74
70	iz	GFP-RTcassette_R	GGTCTGCAGATGATAAGCTGTCAACATGAGAATTGGTCTGAGATGGCGGACGCGATGGCTATTTGTATAGTTCATCCATGCCA	93
71	iz	uba1-RT_F	GTTTTTCATCAATTAATAAACTATTTTCAGTGGTTCTGATCTTTCGATGATTATCGCTGTCGAGATATGACGGTG	85
72	iz	uba1-RT_R	ATGAAGACGGGACAGAGACACATACAACATTGAAAAATCATAATAAGATGATAAGCTGTCAAACATGAG	84
73	iz	lin41-RT_F	GCCCTTCCACTGATAATTAATCAACCTTTTCAGACTGGAAAAAGTGAATCGCTGTCGAGATATGACGGTG	87
74	iz	lin41-RT_R	GATTATTTAAAAACAATCTGAAATGGGTAAGAGCAAGAAGAAAGTGTGATGATAAGCTGTCAAACATGA	82
75	iz	nhr17-RT_F	CTTTTGCCCGTTTCTATGTAATAAACTCCAATTCATTATTTTCATTTTCAGTATCGCTGTCGAGATATGACGGTG	85
76	iz	nhr17-RT_R	TTAAAGTACTAGAAGCTGGTTGTAAGTGCATCATAAATATGTAATTTGATGATAAGCTGTCAAACATGAG	81
77	iz	pL440_F	GTTTTCCAGTCACGACGTT	64
78	iz	PL440_R	TGGATAACCGTATTACCGCC	63
79	iz	mcherry::unc54_R	GTAATTGGACCTTATACAATTCATCCATGC	68
80	iz	unc54UTR::Linker_R	TCAATAAGATGAATCTAGACACAGTTATGTTTGGTATA	68
81	iz	RT-cassette-F	TCGCTGTCGAGATATGACGGTG	70
82	iz	RT-cassette-R	GATGATAAGCTGTCAAACATGAG	60
83	iz	unc54UTR-(NotI)-F	AGATGCGGCCGCTCCAATTACTCTTCAACAT	80
84	iz	84 unc54UTR-(XbaI)-R	GCGTCTAGAAACAGTTATGTTTGGTATA	63
116	iz	116 cassette-RT seq 5'_R	TAGCAGCAGCCATAGTGAC	64
117	iz	117 cassette-RT seq 3'_F	CGGCCTCAACCTACTACTGG	64

121	iz	RT-cassette-F (NotI)	AGCGGCCGCTCGTGTGCGAGATATGACGGTG	85
122	iz	RT-cassette-R (XbaI)	GCGTCTAGATGATAAGCTGTCAAACATGAG	70
123	iz	nhr-17/mCh_F	CTTTTCCCCGTTTCTATGTAAAACTCCAATTCATTATTTTCATTTAGTAATGGTCTCAAAGGGTGAAGAAGATA	83
124	iz	Nhr-17/GFP_R	TTAAAGTGACTAGAAGCTGGTTGTAAGTTGCATCATAAATATGTAATTTCTATTTGTATAGTTTCATCCATGCCA	81
125	iz	Lin41 5'Prom région_F	GAAAGAGCGGCATTTTCTG	64
126	iz	Lin41 5'Codon région_R	GTCCCGAGCCAAGATGATTA	64
127	iz	Lin41 3'Codon region_F	TATCGCCAGAAAACATGTCA	64
128	iz	Lin41 3'UTR region_R	GGAGTGGGGAAAAAGAAAG	64
129	iz	asd-2 primer 1_F	GTCCACAATTTCCCATCACT	61,3
130	iz	asd-2 primer 2_F	AAAAGTTCGCCTGTAGTGGA	61,3
131	iz	asd-2 primer 3_F	TTCTCCACCACCAATGAT	64,1
132	iz	asd-2 primer 4_F	TTTTTATCCTTGACAAAATTGAGA	63,3
133	iz	asd-2 primer 5_F	TCCAAAATGGAAAAATCGAG	61,4
134	iz	asd-2 primer 6_F	GACTGATTGACAGCTGTGTCG	63,6
135	iz	asd-2 primer 1_R	CGATCTTCCCTTGCTGAAG	63,7
136	iz	asd-2 primer 2_R	ATTGGGCTCCGAGTCTTCT	64
137	iz	asd-2 primer 3_R	TTGCGCTACCTGACAAAATG	63,7
138	iz	asd-2 primer 4_R	TCCACGTCCACGTACCATAA	63,8
139	iz	asd-2 primer 5_R	CAGTTGCACCAGCAGAAGAG	63,9
140	iz	asd-2 primer 6_R	TGGAATGATGATTGGATGATG	63,2
141	iz	mCherry/(RT)_R	CCGTCATATCTCGACAGCGACTTATACAATTCATCCATGC	79,6
142	iz	GFP/(RT)_F	ATGTTTGACAGCTTATCATCATGAGTAAAGGAGAAGAACT	73,1
143	iz	mCherry (SacI)_F	GGAGCTCATGGTCTCAAAGGGTGAAGA	73,5
144	iz	GFP(PstI)_R	GGTCTGCAGCTATTTGTATAGTTTCATCCA	68,3
150	iz	mog1_1_F	AGACCTCCGGGAAACCTTTA	63,6
151	iz	mog1_2_F	GAGAAAAGCAATGGCAGGAG	63,8
152	iz	mog1_3_F	AGCTGATGAAATGGGAGTGG	64
153	iz	mog1_4_F	AGCGAATCAACGAACTGGTC	64,2
154	iz	mog1_5_F	CCACAATGCAGCTCGTCTTA	64
155	iz	mog1_6_F	CGAAACAATCACGCAAGAA	63,7
156	iz	mog1_1_R	GCACGGATTTTGGAAATCAT	63,5
157	iz	mog1_2_R	TCTTCAGATTCAGCCCTGCT	64
158	iz	mog1_3_R	GCGTGAGCTGTCCAGTGTTA	64,2
159	iz	mog1_4_R	TTCTCGCATGAAACAGTTCG	64
160	iz	mog1_5_R	CACCCATCAAATTCGAGT	63,6
161	iz	mog1_6_R	TCGGCTCTCTCGTTCTGAT	64
162	iz	let2-intron6_F	GTAGGACCTTGTCTTTGG	56,1
163	iz	let2-intron12_R	ATCGGAATTTAGAAAACTA	52,3
164	iz	lin-41/mCh_F	GCCCTTCCACTGATAAATTAATCAACCTTTTCAGACTTGGAAAAAGTAAAATGGTCTCAAAGGGTGAAGAAGATA	85,4
165	iz	lin-41/GFP_R	GATTATTTTAAAAACAATCTGAAATGGGTAAGAGCAAGAAGAAAGTGTCTATTTGTATAGTTTCATCCATGCCA	81,7
166	iz	uba-1/mCh_F	GTTTTTCATCAATTAATAACTATTTTCAGTGGTTCTGATCTTTCGTGATTAATGGTCTCAAAGGGTGAAGAAGATA	82,7
167	iz	uba-1/GFP_R	ATGAAGACGGGACAGAGACACATACAACATTGAAAAATCATAATAACTATTTGTATAGTTTCATCCATGCCA	84
168	iz	GFP_F (SacI)	GGAGCTCATGAGTAAAGGAGAAGAACT	65,9
169	iz	GFP_R (NotI)	AGCGGCCGCTATTTGTATAGTTTCATCCATGCCA	79,3
170	iz	mCherry_F (SmaI)	GTCAGTCCCGGATGGTCTCAAAGGGTGAAGAAGATA	80,7
171	iz	mCherry_R (PstI)	GGTCTGCACCTTATACAATTCATCCATGC	69,5

179	IZ	GFP_R (SacI)		71,1
313	IZ	nhr-17/GFP_F	CTTTTGCCCGTTTCTATGTAAAACTCCAATTCATTATTTTCATTCAGTAATGAGTAAAGGAGAAGAA	80,8
314	IZ	nhr-17/mCh_R	TTAAAGTGA CTAGAAGCTGGTTGTAAGTTGCATCATAAATATGTAATTTTTATACAATTCATCCATG	78,4



DUD0108	C.e	N2	N2+/+ ; dudEx95[Puba-1::gfp::3'UTR uba-1 ; Prps-27::neo::3'UTR unc-54]	injection
DUD0116	C.e	N2	N2+/+ ; dudEx103[Plin-41::gfp::3'UTR lin-41 ; Prps-27::neo::3'UTR unc-54]	injection
DUD0117	C.e	N2	N2+/+ ; dudEx104[Plin-41::gfp::3'UTR lin-41 ; Prps-27::neo::3'UTR unc-54]	injection
DUD0118	C.e	N2	N2+/+ ; dudEx105[Plin-41::gfp::3'UTR lin-41 ; Prps-27::neo::3'UTR unc-54]	injection
DUD0119	C.e	N2	N2+/+ ; dudEx106[Plin-41::gfp::3'UTR lin-41 ; Prps-27::neo::3'UTR unc-54]	injection
DUD0120	C.e	N2	N2+/+ ; dudEx107[Plin-41::gfp::3'UTR lin-41 ; Prps-27::neo::3'UTR unc-54]	injection
DUD0121	C.e	N2	N2+/+ ; dudEx108[Plin-41::gfp::3'UTR lin-41 ; Prps-27::neo::3'UTR unc-54]	injection
DUD0122	C.e	N2	N2+/+ ; dudEx109[Plin-41::gfp::3'UTR lin-41 ; Prps-27::neo::3'UTR unc-54]	injection
DUD0123	C.e	N2	N2+/+ ; dudEx110[Plin-41::gfp::3'UTR lin-41 ; Prps-27::neo::3'UTR unc-54]	injection

## **VIII. BIBLIOGRAPHIE**

1. Maniatis, T. and R. Reed, *An extensive network of coupling among gene expression machines*. Nature, 2002. **416**(6880): p. 499-506.
2. Orphanides, G. and D. Reinberg, *A unified theory of gene expression*. Cell, 2002. **108**(4): p. 439-51.
3. Moore, M.J., *From birth to death: the complex lives of eukaryotic mRNAs*. Science, 2005. **309**(5740): p. 1514-8.
4. Dreyfuss, G., V.N. Kim, and N. Kataoka, *Messenger-RNA-binding proteins and the messages they carry*. Nat Rev Mol Cell Biol, 2002. **3**(3): p. 195-205.
5. St Johnston, D., *Moving messages: the intracellular localization of mRNAs*. Nat Rev Mol Cell Biol, 2005. **6**(5): p. 363-75.
6. Gebauer, F. and M.W. Hentze, *Molecular mechanisms of translational control*. Nat Rev Mol Cell Biol, 2004. **5**(10): p. 827-35.
7. Parker, R. and H. Song, *The enzymes and control of eukaryotic mRNA turnover*. Nat Struct Mol Biol, 2004. **11**(2): p. 121-7.
8. Jensen, O.N., *Interpreting the protein language using proteomics*. Nat Rev Mol Cell Biol, 2006. **7**(6): p. 391-403.
9. Harbison, C.T., et al., *Transcriptional regulatory code of a eukaryotic genome*. Nature, 2004. **431**(7004): p. 99-104.
10. Lee, T.I., et al., *Transcriptional regulatory networks in Saccharomyces cerevisiae*. Science, 2002. **298**(5594): p. 799-804.
11. Ihmels, J., et al., *Rewiring of the yeast transcriptional network through the evolution of motif usage*. Science, 2005. **309**(5736): p. 938-40.
12. Ihmels, J., et al., *Revealing modular organization in the yeast transcriptional network*. Nat Genet, 2002. **31**(4): p. 370-7.
13. Segal, E., et al., *Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data*. Nat Genet, 2003. **34**(2): p. 166-76.

14. Hieronymus, H. and P.A. Silver, *A systems view of mRNP biology*. Genes Dev, 2004. **18**(23): p. 2845-60.
15. Keene, J.D., *RNA regulons: coordination of post-transcriptional events*. Nat Rev Genet, 2007. **8**(7): p. 533-43.
16. Mata, J., S. Marguerat, and J. Bahler, *Post-transcriptional control of gene expression: a genome-wide perspective*. Trends Biochem Sci, 2005. **30**(9): p. 506-14.
17. Berget, S.M., C. Moore, and P.A. Sharp, *Spliced segments at the 5' terminus of adenovirus 2 late mRNA*. Proceedings of the National Academy of Sciences, 1977. **74**(8): p. 3171-3175.
18. Chow, L.T., et al., *An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA*. Cell, 1977. **12**(1): p. 1-8.
19. Harrison, R.b.T.J., *Spliced segments at the 5' terminus of adenovirus 2 late mRNA*. Susan M. Berget, Claire Moore and Phillip A. Sharp; *An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA*. Louise T. Chow, Richard E. Gelinias, Thomas R. Broker and Richard T. Roberts. Reviews in Medical Virology, 2000. **10**(6): p. 355-371.
20. Koralewski, T.E. and K.V. Krutovsky, *Evolution of Exon-Intron Structure and Alternative Splicing*. PLoS ONE, 2011. **6**(3): p. e18055.
21. Duan, D., *Duchenne muscular dystrophy gene therapy: Lost in translation?* Res Rep Biol, 2011. **2011**(2): p. 31-42.
22. Kunkel, L.M., *2004 William Allan Award address. Cloning of the DMD gene*. Am J Hum Genet, 2005. **76**(2): p. 205-14.
23. Koenig, M., et al., *Complete cloning of the Duchenne muscular dystrophy (DMD) cDNA and preliminary genomic organization of the DMD gene in normal and affected individuals*. Cell, 1987. **50**(3): p. 509-17.
24. Aartsma-Rus, A., et al., *Entries in the Leiden Duchenne muscular dystrophy mutation database: an overview of mutation types and paradoxical cases that confirm the reading-frame rule*. Muscle Nerve, 2006. **34**(2): p. 135-44.

25. Takeda, S., [*Exon skipping therapy for Duchenne muscular dystrophy by using antisense Morpholino*]. *Rinsho Shinkeigaku*, 2009. **49**(11): p. 856-8.
26. Padgett, R.A., et al., *Splicing of Messenger RNA Precursors*. *Annual Review of Biochemistry*, 1986. **55**(1): p. 1119-1150.
27. Maniatis, T. and R. Reed, *The role of small nuclear ribonucleoprotein particles in pre-mRNA splicing*. *Nature*, 1987. **325**(6106): p. 673-678.
28. Zhuang, Y. and A.M. Weiner, *A compensatory base change in U1 snRNA suppresses a 5' splice site mutation*. *Cell*, 1986. **46**(6): p. 827-835.
29. Parker, R., P.G. Siliciano, and C. Guthrie, *Recognition of the TACTAAC box during mRNA splicing in yeast involves base pairing to the U2-like snRNA*. *Cell*, 1987. **49**(2): p. 229-239.
30. Gozani, O., R. Feld, and R. Reed, *Evidence that sequence-independent binding of highly conserved U2 snRNP proteins upstream of the branch site is required for assembly of spliceosomal complex A*. *Genes & Development*, 1996. **10**(2): p. 233-243.
31. Wahl, M.C., C.L. Will, and R. Lührmann, *The Spliceosome: Design Principles of a Dynamic RNP Machine*. *Cell*, 2009. **136**(4): p. 701-718.
32. Dahlberg, J.E. and E. Lund, *Structure and expression of U-snRNA genes*. *Molecular Biology Reports*, 1987. **12**(3): p. 139-143.
33. Lund, E. and J.E. Dahlberg, *Differential accumulation of U1 and U4 small nuclear RNAs during Xenopus development*. *Genes & Development*, 1987. **1**(1): p. 39-46.
34. Lees-Miller, J.P., L.O. Goodwin, and D.M. Helfman, *Three novel brain tropomyosin isoforms are expressed from the rat alpha-tropomyosin gene through the use of alternative promoters and alternative RNA processing*. *Mol Cell Biol*, 1990. **10**(4): p. 1729-42.
35. Pan, Q., et al., *Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing*. *Nat Genet*, 2008. **40**(12): p. 1413-5.
36. Wang, E.T., et al., *Alternative isoform regulation in human tissue transcriptomes*. *Nature*, 2008. **456**(7221): p. 470-6.

37. Calarco, J.A., M. Zhen, and B.J. Blencowe, *Networking in a global world: establishing functional connections between neural splicing regulators and their target transcripts*. RNA, 2011. **17**(5): p. 775-91.
38. Hartmann, B. and J. Valcarcel, *Decrypting the genome's alternative messages*. Curr Opin Cell Biol, 2009. **21**(3): p. 377-86.
39. Licatalosi, D.D. and R.B. Darnell, *RNA processing and its regulation: global insights into biological networks*. Nat Rev Genet, 2010. **11**(1): p. 75-87.
40. Lopez, A.J., *Alternative splicing of pre-mRNA: developmental consequences and mechanisms of regulation*. Annu Rev Genet, 1998. **32**: p. 279-305.
41. Hallegger, M., M. Llorian, and C.W. Smith, *Alternative splicing: global insights*. FEBS J, 2010. **277**(4): p. 856-66.
42. Barash, Y., et al., *Deciphering the splicing code*. Nature, 2010. **465**(7294): p. 53-9.
43. Yeo, G. and C.B. Burge, *Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals*. J Comput Biol, 2004. **11**(2-3): p. 377-94.
44. Kol, G., G. Lev-Maor, and G. Ast, *Human-mouse comparative analysis reveals that branch-site plasticity contributes to splicing regulation*. Hum Mol Genet, 2005. **14**(11): p. 1559-68.
45. Gao, K., et al., *Human branch point consensus sequence is yUnAy*. Nucleic Acids Research, 2008. **36**(7): p. 2257-2267.
46. Matlin, A.J., F. Clark, and C.W. Smith, *Understanding alternative splicing: towards a cellular code*. Nat. Rev. Mol. Cell Biol., 2005. **6**: p. 386-398.
47. Fairbrother, W.G., et al., *Predictive Identification of Exonic Splicing Enhancers in Human Genes*. Science, 2002. **297**(5583): p. 1007-1013.
48. Wang, Z., et al., *Systematic identification and analysis of exonic splicing silencers*. Cell, 2004. **119**(6): p. 831-45.
49. Zhang, X.H.-F. and L.A. Chasin, *Computational definition of sequence motifs governing constitutive exon splicing*. Genes & Development, 2004. **18**(11): p. 1241-1250.

50. Wang, Z. and C.B. Burge, *Splicing regulation: From a parts list of regulatory elements to an integrated splicing code*. RNA, 2008. **14**(5): p. 802-813.
51. Chasin, L., *Searching for splicing motifs*, in *Alternative splicing in the postgenomic era*, B.J. Blencowe and B.R. Graveley, Editors. 2007, Landes Biosciences: Austin, Texas. p. 85–106.
52. Alekseyenko, A.V., N. Kim, and C.J. Lee, *Global analysis of exon creation versus loss and the role of alternative splicing in 17 vertebrate genomes*. RNA, 2007. **13**(5): p. 661-70.
53. Sugnet, C.W., et al., *Transcriptome and genome conservation of alternative splicing events in humans and mice*. Pac Symp Biocomput, 2004: p. 66-77.
54. Keren, H., G. Lev-Maor, and G. Ast, *Alternative splicing and evolution: diversification, exon definition and function*. Nat Rev Genet, 2010. **11**(5): p. 345-55.
55. Kim, E., A. Goren, and G. Ast, *Alternative splicing: current perspectives*. Bioessays, 2008. **30**(1): p. 38-47.
56. Sakabe, N. and S. de Souza, *Sequence features responsible for intron retention in human*. BMC Genomics, 2007. **8**(1): p. 59.
57. Ast, G., *How did alternative splicing evolve?* Nat Rev Genet, 2004. **5**(10): p. 773-782.
58. Black, D.L., *MECHANISMS OF ALTERNATIVE PRE-MESSENGER RNA SPLICING*. Annual Review of Biochemistry, 2003. **72**(1): p. 291-336.
59. Proudfoot, N.J., *Ending the message: poly(A) signals then and now*. Genes Dev, 2011. **25**(17): p. 1770-82.
60. Ohno, G., M. Hagiwara, and H. Kuroyanagi, *STAR family RNA-binding protein ASD-2 regulates developmental switching of mutually exclusive alternative splicing in vivo*. Genes & Development, 2008. **22**(3): p. 360-374.
61. Sibley, M.H., et al., *Genetic identification, sequence, and alternative splicing of the Caenorhabditis elegans alpha 2(IV) collagen gene*. The Journal of Cell Biology, 1993. **123**(1): p. 255-264.

62. Goodman, S.J., et al., *Alternative splicing affecting a novel domain in the C. elegans EGL-15 FGF receptor confers functional specificity*. Development, 2003. **130**: p. 3757-3766.
63. Morrison, M., K.S. Harris, and M.B. Roth, *smg mutants affect the expression of alternatively spliced SR protein mRNAs in Caenorhabditis elegans*. Proc Natl Acad Sci U S A, 1997. **94**(18): p. 9782-5.
64. Lewis, B.P., R.E. Green, and S.E. Brenner, *Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans*. Proc Natl Acad Sci U S A, 2003. **100**(1): p. 189-92.
65. Lee, R.C., R.L. Feinbaum, and V. Ambros, *The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14*. Cell, 1993. **75**(5): p. 843-54.
66. Wightman, B., I. Ha, and G. Ruvkun, *Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans*. Cell, 1993. **75**(5): p. 855-62.
67. Moss, E.G., R.C. Lee, and V. Ambros, *The cold shock domain protein LIN-28 controls developmental timing in C. elegans and is regulated by the lin-4 RNA*. Cell, 1997. **88**(5): p. 637-46.
68. Slack, F.J., et al., *The lin-41 RBCC gene acts in the C. elegans heterochronic pathway between the let-7 regulatory RNA and the LIN-29 transcription factor*. Mol Cell, 2000. **5**(4): p. 659-69.
69. He, L. and G.J. Hannon, *MicroRNAs: small RNAs with a big role in gene regulation*. Nat Rev Genet, 2004. **5**(7): p. 522-31.
70. Bartel, D.P., *MicroRNAs: genomics, biogenesis, mechanism, and function*. Cell, 2004. **116**(2): p. 281-97.
71. Ruvkun, G., et al., *Molecular genetics of the Caenorhabditis elegans heterochronic gene lin-14*. Genetics, 1989. **121**(3): p. 501-16.
72. Ruvkun, G. and J. Giusto, *The Caenorhabditis elegans heterochronic gene lin-14 encodes a nuclear protein that forms a temporal developmental switch*. Nature, 1989. **338**(6213): p. 313-9.
73. Olsen, P.H. and V. Ambros, *The lin-4 regulatory RNA controls developmental timing in Caenorhabditis elegans by blocking LIN-14 protein synthesis after the initiation of translation*. Dev Biol, 1999. **216**(2): p. 671-80.

74. Ha, I., B. Wightman, and G. Ruvkun, *A bulged lin-4/lin-14 RNA duplex is sufficient for Caenorhabditis elegans lin-14 temporal gradient formation*. Genes Dev, 1996. **10**(23): p. 3041-50.
75. Reinhart, B.J., et al., *The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans*. Nature, 2000. **403**(6772): p. 901-6.
76. Lin, S.Y., et al., *The C elegans hunchback homolog, hbl-1, controls temporal patterning and is a probable microRNA target*. Dev Cell, 2003. **4**(5): p. 639-50.
77. Abrahante, J.E., et al., *The Caenorhabditis elegans hunchback-like gene lin-57/hbl-1 controls developmental time and is regulated by microRNAs*. Dev Cell, 2003. **4**(5): p. 625-37.
78. Pasquinelli, A.E., et al., *Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA*. Nature, 2000. **408**(6808): p. 86-9.
79. Lai, E.C., et al., *Computational identification of Drosophila microRNA genes*. Genome Biol, 2003. **4**(7): p. R42.
80. Lim, L.P., et al., *The microRNAs of Caenorhabditis elegans*. Genes Dev, 2003. **17**(8): p. 991-1008.
81. Lagos-Quintana, M., et al., *Identification of novel genes coding for small expressed RNAs*. Science, 2001. **294**(5543): p. 853-8.
82. Lau, N.C., et al., *An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans*. Science, 2001. **294**(5543): p. 858-62.
83. Lee, R.C. and V. Ambros, *An extensive class of small RNAs in Caenorhabditis elegans*. Science, 2001. **294**(5543): p. 862-4.
84. Bartel, D.P. and C.Z. Chen, *Micromanagers of gene expression: the potentially widespread influence of metazoan microRNAs*. Nat Rev Genet, 2004. **5**(5): p. 396-400.
85. Seitz, H., et al., *Imprinted microRNA genes transcribed antisense to a reciprocally imprinted retrotransposon-like gene*. Nat Genet, 2003. **34**(3): p. 261-2.

86. Ohler, U., et al., *Patterns of flanking sequence conservation and a characteristic upstream motif for microRNA gene identification*. RNA, 2004. **10**(9): p. 1309-22.
87. Ambros, V., et al., *MicroRNAs and other tiny endogenous RNAs in C. elegans*. Curr Biol, 2003. **13**(10): p. 807-18.
88. Lim, L.P., et al., *Vertebrate microRNA genes*. Science, 2003. **299**(5612): p. 1540.
89. Kim, V.N. and J.W. Nam, *Genomics of microRNA*. Trends Genet, 2006. **22**(3): p. 165-73.
90. Cummins, J.M., et al., *The colorectal microRNAome*. Proc Natl Acad Sci U S A, 2006. **103**(10): p. 3687-92.
91. Kozomara, A. and S. Griffiths-Jones, *miRBase: integrating microRNA annotation and deep-sequencing data*. Nucleic Acids Research, 2011. **39**(suppl 1): p. D152-D157.
92. Rodriguez, A., et al., *Identification of mammalian microRNA host genes and transcription units*. Genome Res, 2004. **14**(10A): p. 1902-10.
93. Lee, Y., et al., *MicroRNA maturation: stepwise processing and subcellular localization*. EMBO J, 2002. **21**(17): p. 4663-70.
94. Calin, G.A., et al., *Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers*. Proc Natl Acad Sci U S A, 2004. **101**(9): p. 2999-3004.
95. He, L., et al., *A microRNA polycistron as a potential human oncogene*. Nature, 2005. **435**(7043): p. 828-33.
96. Aravin, A.A., et al., *The small RNA profile during Drosophila melanogaster development*. Dev Cell, 2003. **5**(2): p. 337-50.
97. Cai, X., C.H. Hagedorn, and B.R. Cullen, *Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs*. RNA, 2004. **10**(12): p. 1957-66.
98. Lee, Y., et al., *MicroRNA genes are transcribed by RNA polymerase II*. EMBO J, 2004. **23**(20): p. 4051-60.
99. Lagos-Quintana, M., et al., *Identification of tissue-specific microRNAs from mouse*. Curr Biol, 2002. **12**(9): p. 735-9.

100. Krichevsky, A.M., et al., *A microRNA array reveals extensive regulation of microRNAs during brain development*. RNA, 2003. **9**(10): p. 1274-81.
101. Miska, E.A., et al., *Microarray analysis of microRNA expression in the developing mammalian brain*. Genome Biol, 2004. **5**(9): p. R68.
102. Filipowicz, W., et al., *Post-transcriptional gene silencing by siRNAs and miRNAs*. Curr Opin Struct Biol, 2005. **15**(3): p. 331-41.
103. Lamontagne, B., et al., *The RNase III family: a conserved structure and expanding functions in eukaryotic dsRNA metabolism*. Curr Issues Mol Biol, 2001. **3**(4): p. 71-8.
104. Lee, Y., et al., *The nuclear RNase III Drosha initiates microRNA processing*. Nature, 2003. **425**(6956): p. 415-9.
105. Kurihara, Y. and Y. Watanabe, *Arabidopsis micro-RNA biogenesis through Dicer-like 1 protein functions*. Proc Natl Acad Sci U S A, 2004. **101**(34): p. 12753-8.
106. Basyuk, E., et al., *Human let-7 stem-loop precursors harbor features of RNase III cleavage products*. Nucleic Acids Res, 2003. **31**(22): p. 6593-7.
107. Yi, R., et al., *Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs*. Genes Dev, 2003. **17**(24): p. 3011-6.
108. Lund, E., et al., *Nuclear export of microRNA precursors*. Science, 2004. **303**(5654): p. 95-8.
109. Han, J., et al., *The Drosha-DGCR8 complex in primary microRNA processing*. Genes Dev, 2004. **18**(24): p. 3016-27.
110. Denli, A.M., et al., *Processing of primary microRNAs by the Microprocessor complex*. Nature, 2004. **432**(7014): p. 231-5.
111. Gregory, R.I., et al., *The Microprocessor complex mediates the genesis of microRNAs*. Nature, 2004. **432**(7014): p. 235-40.
112. Gwizdek, C., et al., *Exportin-5 mediates nuclear export of minihelix-containing RNAs*. J Biol Chem, 2003. **278**(8): p. 5505-8.
113. Brownawell, A.M. and I.G. Macara, *Exportin-5, a novel karyopherin, mediates nuclear export of double-stranded RNA binding proteins*. J Cell Biol, 2002. **156**(1): p. 53-64.

114. Kim, J.K., et al., *Functional Genomic Analysis of RNA Interference in C. elegans*. Science, 2005. **308**(5725): p. 1164-1167.
115. Carmell, M.A. and G.J. Hannon, *RNase III enzymes and the initiation of gene silencing*. Nat Struct Mol Biol, 2004. **11**(3): p. 214-8.
116. Zhang, H., et al., *Human Dicer preferentially cleaves dsRNAs at their termini without a requirement for ATP*. EMBO J, 2002. **21**(21): p. 5875-85.
117. Liu, Q., et al., *R2D2, a bridge between the initiation and effector steps of the Drosophila RNAi pathway*. Science, 2003. **301**(5641): p. 1921-5.
118. Zhang, H., et al., *Single processing center models for human Dicer and bacterial RNase III*. Cell, 2004. **118**(1): p. 57-68.
119. Pham, J.W., et al., *A Dicer-2-dependent 80s complex cleaves targeted mRNAs during RNAi in Drosophila*. Cell, 2004. **117**(1): p. 83-94.
120. Pham, J.W. and E.J. Sontheimer, *The Making of an siRNA*. Mol Cell, 2004. **15**(2): p. 163-4.
121. Tabara, H., et al., *The dsRNA binding protein RDE-4 interacts with RDE-1, DCR-1, and a DExH-box helicase to direct RNAi in C. elegans*. Cell, 2002. **109**(7): p. 861-71.
122. Hammond, S.M., et al., *An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells*. Nature, 2000. **404**(6775): p. 293-6.
123. Elbashir, S.M., W. Lendeckel, and T. Tuschl, *RNA interference is mediated by 21- and 22-nucleotide RNAs*. Genes Dev, 2001. **15**(2): p. 188-200.
124. Martinez, J., et al., *Single-stranded antisense siRNAs guide target RNA cleavage in RNAi*. Cell, 2002. **110**(5): p. 563-74.
125. Khvorova, A., A. Reynolds, and S.D. Jayasena, *Functional siRNAs and miRNAs exhibit strand bias*. Cell, 2003. **115**(2): p. 209-16.
126. Schwarz, D.S., et al., *Asymmetry in the Assembly of the RNAi Enzyme Complex*. Cell, 2003. **115**(2): p. 199-208.
127. Hutvagner, G. and P.D. Zamore, *A microRNA in a multiple-turnover RNAi enzyme complex*. Science, 2002. **297**(5589): p. 2056-60.

128. Zeng, Y., E.J. Wagner, and B.R. Cullen, *Both natural and designed micro RNAs can inhibit the expression of cognate mRNAs when expressed in human cells*. Mol Cell, 2002. **9**(6): p. 1327-33.
129. Doench, J.G., C.P. Petersen, and P.A. Sharp, *siRNAs can function as miRNAs*. Genes Dev, 2003. **17**(4): p. 438-42.
130. Llave, C., et al., *Endogenous and silencing-associated small RNAs in plants*. Plant Cell, 2002. **14**(7): p. 1605-19.
131. Llave, C., et al., *Cleavage of Scarecrow-like mRNA targets directed by a class of Arabidopsis miRNA*. Science, 2002. **297**(5589): p. 2053-6.
132. Pillai, R.S., et al., *Inhibition of translational initiation by Let-7 MicroRNA in human cells*. Science, 2005. **309**(5740): p. 1573-6.
133. *Genome sequence of the nematode C. elegans: a platform for investigating biology*. Science, 1998. **282**: p. 2012-2018.
134. Cuellar, T.L. and M.T. McManus, *MicroRNAs and endocrine biology*. J Endocrinol, 2005. **187**(3): p. 327-32.
135. Mette, M.F., et al., *Transcriptional silencing and promoter methylation triggered by double-stranded RNA*. EMBO J, 2000. **19**(19): p. 5194-201.
136. Hamilton, A., et al., *Two classes of short interfering RNA in RNA silencing*. EMBO J, 2002. **21**(17): p. 4671-9.
137. Hall, I.M., et al., *Establishment and maintenance of a heterochromatin domain*. Science, 2002. **297**(5590): p. 2232-7.
138. Volpe, T.A., et al., *Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi*. Science, 2002. **297**(5588): p. 1833-7.
139. Grewal, S.I. and J.C. Rice, *Regulation of heterochromatin by histone methylation and small RNAs*. Curr Opin Cell Biol, 2004. **16**(3): p. 230-8.
140. Lai, E.C., *Predicting and validating microRNA targets*. Genome Biol, 2004. **5**(9): p. 115.
141. Brown, J.R. and P. Sanseau, *A computational view of microRNAs and their targets*. Drug Discov Today, 2005. **10**(8): p. 595-601.

142. Stark, A., et al., *Identification of Drosophila MicroRNA targets*. PLoS Biol, 2003. **1**(3): p. E60.
143. Enright, A.J., et al., *MicroRNA targets in Drosophila*. Genome Biol, 2003. **5**(1): p. R1.
144. Lewis, B.P., et al., *Prediction of Mammalian MicroRNA Targets*. Cell, 2003. **115**(7): p. 787-798.
145. Rajewsky, N. and N.D. Socci, *Computational identification of microRNA targets*. Developmental Biology, 2004. **267**(2): p. 529-535.
146. REHMSMEIER, M., et al., *Fast and effective prediction of microRNA/target duplexes*. RNA, 2004. **10**(10): p. 1507-1517.
147. Kiriakidou, M., et al., *A combined computational-experimental approach predicts human microRNA targets*. Genes & Development, 2004. **18**(10): p. 1165-1178.
148. John, B., et al., *Human MicroRNA Targets*. PLoS Biol, 2004. **2**(11): p. e363.
149. Vella, M.C., et al., *The C. elegans microRNA let-7 binds to imperfect let-7 complementary sites from the lin-41 3'UTR*. Genes Dev, 2004. **18**(2): p. 132-7.
150. Doench, J.G. and P.A. Sharp, *Specificity of microRNA target selection in translational repression*. Genes Dev, 2004. **18**(5): p. 504-11.
151. Pfeffer, S., et al., *Identification of virus-encoded microRNAs*. Science, 2004. **304**(5671): p. 734-6.
152. Brennecke, J., et al., *Principles of microRNA-target recognition*. PLoS Biol, 2005. **3**(3): p. e85.
153. Izant, J.G. and H. Weintraub, *Constitutive and conditional suppression of exogenous and endogenous genes by anti-sense RNA*. Science, 1985. **229**(4711): p. 345-52.
154. Izant, J.G. and H. Weintraub, *Inhibition of thymidine kinase gene expression by anti-sense RNA: a molecular approach to genetic analysis*. Cell, 1984. **36**(4): p. 1007-15.

155. Harland, R. and H. Weintraub, *Translation of mRNA injected into Xenopus oocytes is specifically inhibited by antisense RNA*. J Cell Biol, 1985. **101**(3): p. 1094-9.
156. Strickland, S., et al., *Antisense RNA directed against the 3' noncoding region prevents dormant mRNA activation in mouse oocytes*. Science, 1988. **241**(4866): p. 680-4.
157. Rosenberg, U.B., et al., *Production of phenocopies by Kruppel antisense RNA injection into Drosophila embryos*. Nature, 1985. **313**(6004): p. 703-6.
158. Fire, A., et al., *Production of antisense RNA leads to effective and specific inhibition of gene expression in C. elegans muscle*. Development, 1991. **113**(2): p. 503-14.
159. Grishok, A. and C.C. Mello, *RNAi (Nematodes: Caenorhabditis elegans)*. Adv Genet, 2002. **46**: p. 339-60.
160. Guo, S. and K.J. Kemphues, *par-1, a gene required for establishing polarity in C. elegans embryos, encodes a putative Ser/Thr kinase that is asymmetrically distributed*. Cell, 1995. **81**(4): p. 611-20.
161. Fire, A., *Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans*. Nature, 1998. **391**: p. 806-811.
162. Waterhouse, P.M., M.W. Graham, and M.B. Wang, *Virus resistance and gene silencing in plants can be induced by simultaneous expression of sense and antisense RNA*. Proc Natl Acad Sci U S A, 1998. **95**(23): p. 13959-64.
163. Ngo, H., et al., *Double-stranded RNA induces mRNA degradation in Trypanosoma brucei*. Proc Natl Acad Sci U S A, 1998. **95**(25): p. 14687-92.
164. Kennerdell, J.R. and R.W. Carthew, *Use of dsRNA-mediated genetic interference to demonstrate that frizzled and frizzled 2 act in the wingless pathway*. Cell, 1998. **95**(7): p. 1017-26.
165. Sanchez Alvarado, A. and P.A. Newmark, *Double-stranded RNA specifically disrupts gene expression during planarian regeneration*. Proc Natl Acad Sci U S A, 1999. **96**(9): p. 5049-54.
166. Lohmann, J.U., I. Endl, and T.C. Bosch, *Silencing of developmental genes in Hydra*. Dev Biol, 1999. **214**(1): p. 211-4.

167. Wianny, F. and M. Zernicka-Goetz, *Specific interference with gene function by double-stranded RNA in early mouse development*. Nat Cell Biol, 2000. **2**(2): p. 70-5.
168. Tabara, H., et al., *The rde-1 gene, RNA interference, and transposon silencing in C. elegans*. Cell, 1999. **99**(2): p. 123-32.
169. Timmons, L. and A. Fire, *Specific interference by ingested dsRNA*. Nature, 1998. **395**(6705): p. 854-854.
170. Tavernarakis, N., et al., *Heritable and inducible genetic interference by double-stranded RNA encoded by transgenes*. Nat Genet, 2000. **24**(2): p. 180-3.
171. Fraser, A.G., et al., *Functional genomic analysis of C. elegans chromosome I by systematic RNA interference*. Nature, 2000. **408**(6810): p. 325-330.
172. Simmer, F., et al., *Genome-Wide RNAi of [C. elegans](http://www.w3.org/1999/xlink) Using the Hypersensitive *rrf-3* Strain Reveals Novel Gene Functions*. PLoS Biol, 2003. **1**(1): p. e12.
173. Gonczy, P., et al., *Functional genomic analysis of cell division in C. elegans using RNAi of genes on chromosome III*. Nature, 2000. **408**(6810): p. 331-6.
174. Piano, F., et al., *RNAi analysis of genes expressed in the ovary of Caenorhabditis elegans*. Curr Biol, 2000. **10**(24): p. 1619-22.
175. Brenner, S., Genetics, 1974. **77**: p. 71-94.
176. BRENNER, S., *THE GENETICS OF BEHAVIOUR*. British Medical Bulletin, 1973. **29**(3): p. 269-271.
177. Dougherty, E.C. and H.G. Calhoun, *Possible significance of free-living nematodes in genetic research*. Nature, 1948. **161**(4079): p. 29.
178. Maupas, E., *Modes et formes de reproduction des nematodes*. Archives de Zoologie Experimentale et Generale, 1901. **8**: p. 463-624.
179. Osche, G., *Systematik und phylogenie der gattung Rhabditis (Nematoda)*. Zoologische Jahrbucher - Abt. Zool. Phys. der Tiere, 1952. **81**: p. 190-280.

180. Dougherty, E.C., *The genera and species of the subfamily Rhabditinae Micoletzky, 1922 (Nematoda): a nomenclatorial analysis, including an addendum on the composition of the family Rhabditidae Orley, 1880.* J Helminthol, 1955. **29**(3): p. 105-52.
181. Nigon, V., *Les modalites de la reproduction et le determinisme du sexe chez quelques nematodes libres.* Annales de Sciences Naturelles - Zool. Biol. Anim., 1949. **11**: p. 1-132.
182. Nicholas, W.L., E.C. Dougherty, and E.L. Hansen, *AXENIC CULTIVATION OF CAENORHARDITIS BRIGGSÆ (NEMATODA: RHABDITIDAE) WITH CHEMICALLY UNDEFINED SUPPLEMENTS; COMPARATIVE STUDIES WITH RELATED NEMATODES\**. Annals of the New York Academy of Sciences, 1959. **77**(2): p. 218-236.
183. Fatt, H.V. and E.C. Dougherty, *Genetic Control of Differential Heat Tolerance in Two Strains of the Nematode Caenorhabditis elegans.* Science, 1963. **141**(3577): p. 266-267.
184. White, J., *4 The Anatomy.* 1988. 1988.
185. Moerman, D.G. and A. Fire, *Structure, Function, and Development. In C. elegans Volume II*, D.L. Riddle, et al., Editors, Cold Spring Harbor Laboratory Press. p. 417-470.
186. Driscoll, M. and J. Kaplan, *Mechanotransduction, in C. elegans II. 2nd edition.*, D.L. Riddle, T. Blumenthal, and B.J. Meyer, Editors. 1997, Cold Spring Harbor Laboratory Press: Cold Spring Harbor (NY).
187. Bird and Bird, *The structure of nematodes.* 1991, San Diego, CA: Academic Press.
188. Byerly, L., R.C. Cassada, and R.L. Russell, *The life cycle of the nematode Caenorhabditis elegans: I. Wild-type growth and reproduction.* Developmental Biology, 1976. **51**(1): p. 23-33.
189. Sulston, J.E., et al., *The embryonic cell lineage of the nematode Caenorhabditis elegans.* Dev Biol, 1983. **100**(1): p. 64-119.
190. Deppe, U., et al., *Cell lineages of the embryo of the nematode Caenorhabditis elegans.* Proceedings of the National Academy of Sciences, 1978. **75**(1): p. 376-380.
191. Wood, W.B., *8 Embryology.* 1988. 1988.

192. William, B.W., *8 Embryology*. Cold Spring Harbor Monograph Archive; Volume 17 (1988): The Nematode *Caenorhabditis elegans*, 1988.
193. Bucher, E.A. and G. Seydoux, *Gastrulation in the nematode Caenorhabditis elegans*. Seminars in Developmental Biology, 1994. **5**(2): p. 121-130.
194. Chin-Sang, I.D. and A.D. Chisholm, *Form of the worm:: genetics of epidermal morphogenesis in C. elegans*. Trends in Genetics, 2000. **16**(12): p. 544-551.
195. Ambros, V., *Control of developmental timing in Caenorhabditis elegans*. Current Opinion in Genetics & Development, 2000. **10**(4): p. 428-433.
196. Johnson, T.E., et al., *Arresting development arrests aging in the nematode Caenorhabditis elegans*. Mechanisms of Ageing and Development, 1984. **28**(1): p. 23-40.
197. Kao, G., et al., *ASNA-1 Positively Regulates Insulin Secretion in C. elegans and Mammalian Cells*. Cell, 2007. **128**(3): p. 577-587.
198. Slack, F. and G. Ruvkun, *Temporal pattern formation by heterochronic genes*. Annu Rev Genet, 1997. **31**: p. 611-34.
199. Donald, L.R., *12 The Dauer Larva*. Cold Spring Harbor Monograph Archive; Volume 17 (1988): The Nematode *Caenorhabditis elegans*, 1988.
200. Jonathan, H., *9 Sexual Dimorphism and Sex Determination*. Cold Spring Harbor Monograph Archive; Volume 17 (1988): The Nematode *Caenorhabditis elegans*, 1988.
201. Ward, S. and J.S. Carrel, *Fertilization and sperm competition in the nematode Caenorhabditis elegans*. Developmental Biology, 1979. **73**(2): p. 304-321.
202. Hodgkin, J. and T. Doniach, *Natural Variation and Copulatory Plug Formation in Caenorhabditis elegans*. Genetics, 1997. **146**(1): p. 149-164.
203. Sulston, J.E. and H.R. Horvitz, *Post-embryonic cell lineages of the nematode, Caenorhabditis elegans*. Dev Biol, 1977. **56**(1): p. 110-56.

204. Sulston, J.E., D.G. Albertson, and J.N. Thomson, *The Caenorhabditis elegans male: postembryonic development of nongonadal structures*. Dev Biol, 1980. **78**(2): p. 542-76.
205. Pruitt, K.D., T. Tatusova, and D.R. Maglott, *NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins*. Nucleic Acids Res, 2007. **35**(Database issue): p. D61-5.
206. Lander, E.S., *Initial sequencing and analysis of the human genome*. Nature, 2001. **409**: p. 860-921.
207. Hannon, G.J., P.A. Maroney, and T.W. Nilsen, *U small nuclear ribonucleoprotein requirements for nematode cis- and trans-splicing in vitro*. J Biol Chem, 1991. **266**(34): p. 22792-5.
208. Krause, M. and D. Hirsh, *A trans-spliced leader sequence on actin mRNA in C. elegans*. Cell, 1987. **49**(6): p. 753-61.
209. Huang, X.Y. and D. Hirsh, *A second trans-spliced RNA leader sequence in the nematode Caenorhabditis elegans*. Proc Natl Acad Sci U S A, 1989. **86**(22): p. 8640-4.
210. Blumenthal, T., et al., *A global analysis of Caenorhabditis elegans operons*. Nature, 2002. **417**(6891): p. 851-4.
211. Spieth, J., et al., *Operons in C. elegans: polycistronic mRNA precursors are processed by trans-splicing of SL2 to downstream coding regions*. Cell, 1993. **73**(3): p. 521-32.
212. Zorio, D.A., et al., *Operons as a common form of chromosomal organization in C. elegans*. Nature, 1994. **372**(6503): p. 270-2.
213. Paulsen, J.E., E.E. Capowski, and S. Strome, *Phenotypic and molecular analysis of mes-3, a maternal-effect gene required for proliferation and viability of the germ line in C. elegans*. Genetics, 1995. **141**(4): p. 1383-98.
214. Ramani, A.K., et al., *Genome-wide analysis of alternative splicing in Caenorhabditis elegans*. Genome Res, 2011. **21**(2): p. 342-8.
215. Moore, M.J. and N.J. Proudfoot, *Pre-mRNA processing reaches back to transcription and ahead to translation*. Cell, 2009. **136**(4): p. 688-700.

216. Thomas, J., et al., *The spliceosomal snRNAs of Caenorhabditis elegans*. Nucleic Acids Res, 1990. **18**(9): p. 2633-42.
217. Jurica, M.S. and M.J. Moore, *Pre-mRNA splicing: awash in a sea of proteins*. Mol Cell, 2003. **12**(1): p. 5-14.
218. Valadkhan, S. and Y. Jaladat, *The spliceosomal proteome: at the heart of the largest cellular ribonucleoprotein machine*. Proteomics, 2010. **10**(22): p. 4128-41.
219. Kent, W.J. and A.M. Zahler, *Conservation, regulation, synteny, and introns in a large-scale C. briggsae-C. elegans genomic alignment*. Genome Res, 2000. **10**(8): p. 1115-25.
220. Ruzanov, P., S.J. Jones, and D.L. Riddle, *Discovery of novel alternatively spliced C. elegans transcripts by computational analysis of SAGE data*. BMC Genomics, 2007. **8**: p. 447.
221. Barberan-Soler, S. and A.M. Zahler, *Alternative splicing and the steady-state ratios of mRNA isoforms generated by it are under strong stabilizing selection in Caenorhabditis elegans*. Mol Biol Evol, 2008. **25**(11): p. 2431-7.
222. Irimia, M., et al., *Widespread evolutionary conservation of alternatively spliced exons in Caenorhabditis*. Mol Biol Evol, 2008. **25**(2): p. 375-82.
223. Li, Y., et al., *Global genetic robustness of the alternative splicing machinery in Caenorhabditis elegans*. Genetics, 2010. **186**(1): p. 405-10.
224. Rukov, J.L., et al., *High qualitative and quantitative conservation of alternative splicing in Caenorhabditis elegans and Caenorhabditis briggsae*. Mol Biol Evol, 2007. **24**(4): p. 909-17.
225. Pettitt, J. and I.B. Kingston, *DEVELOPMENTALLY-REGULATED ALTERNATIVE SPLICING OF A NEMATODE TYPE-IV COLLAGEN GENE*. Developmental Biology, 1994. **161**(1): p. 22-29.
226. Graham, P.L., et al., *Type IV Collagen Is Detectable in Most, but Not All, Basement Membranes of Caenorhabditis elegans and Assembles on Tissues That Do Not Express It*. The Journal of Cell Biology, 1997. **137**(5): p. 1171-1183.

227. Kuroyanagi, H., et al., *Transgenic alternative-splicing reporters reveal tissue-specific expression profiles and regulation mechanisms in vivo*. Nat Methods, 2006. **3**(11): p. 909-15.
228. Jin, Y., *A vertebrate RNA-binding protein Fox-1 regulates tissue-specific splicing via the pentanucleotide GCAUG*. EMBO J., 2003. **22**: p. 905-912.
229. Auweter, S.D., *Molecular basis of RNA recognition by the human alternative splicing factor Fox-1*. EMBO J., 2006. **25**: p. 163-173.
230. Kuroyanagi, H., et al., *The Fox-1 family and SUP-12 coordinately regulate tissue-specific alternative splicing in vivo*. Mol Cell Biol, 2007. **27**(24): p. 8612-21.
231. Lall, S., et al., *A Genome-Wide Map of Conserved MicroRNA Targets in C. elegans*. Current Biology, 2006. **16**(5): p. 460-471.
232. Pulak, R., Methods Mol. Biol., 2006. **351**: p. 275-286.
233. Boyd, W.A., et al., *A high-throughput method for assessing chemical toxicity using a Caenorhabditis elegans reproduction assay*. Toxicol Appl Pharmacol, 2010. **245**(2): p. 153-9.
234. Boyd, W.A., et al., *Medium- and high-throughput screening of neurotoxicants using C. elegans*. Neurotoxicol Teratol, 2010. **32**(1): p. 68-73.
235. Smith, M.V., et al., *A Discrete Time Model for the Analysis of Medium-Throughput *C. elegans* Growth Data*. PLoS ONE, 2009. **4**(9): p. e7018.
236. Dupuy, D., et al., *Genome-scale analysis of in vivo spatiotemporal promoter activity in Caenorhabditis elegans*. Nat Biotech, 2007. **25**(6): p. 663-668.
237. Doitsidou, M., et al., *Automated screening for mutants affecting dopaminergic-neuron specification in C. elegans*. Nat Meth, 2008. **5**(10): p. 869-872.
238. Wicks, S.R., et al., *Rapid gene mapping in Caenorhabditis elegans using a high density polymorphism map*. Nat Genet, 2001. **28**(2): p. 160-164.

239. Wicks, S.R., et al., *Rapid gene mapping in Caenorhabditis elegans using a high density polymorphism map*. Nat. Genet., 2001. **28**: p. 160-164.
240. Davis, M.W., et al., *Rapid single nucleotide polymorphism mapping in C. elegans*. BMC Genomics, 2005. **6**(1): p. 118.
241. Flibotte, S., et al., *Rapid High Resolution Single Nucleotide Polymorphism–Comparative Genome Hybridization Mapping in Caenorhabditis elegans*. Genetics, 2009. **181**(1): p. 33-37.
242. Sarin, S., et al., *Analysis of Multiple Ethyl Methanesulfonate-Mutagenized Caenorhabditis elegans Strains by Whole-Genome Sequencing*. Genetics, 2010. **185**(2): p. 417-430.
243. Sarin, S., et al., *Caenorhabditis elegans mutant allele identification by whole-genome sequencing*. Nat Meth, 2008. **5**(10): p. 865-867.
244. Bigelow, H., et al., *MAQGene: software to facilitate C. elegans mutant genome sequence analysis*. Nat Meth, 2009. **6**(8): p. 549-549.
245. Weber, K.P., et al., *Whole genome sequencing highlights genetic changes associated with laboratory domestication of C. elegans*. PLoS ONE, 2010. **5**(11): p. e13922.
246. Morton, D.G., W.A. Hoose, and K.J. Kemphues, *A Genome-wide RNAi Screen for Enhancers of Par Mutants Reveals New Contributors to Early Embryonic Polarity in Caenorhabditis elegans*. Genetics, 2012.
247. Sun, Y., et al., *A genome-wide RNAi screen identifies genes regulating the formation of P bodies in C. elegans and their functions in NMD and RNAi*. Protein Cell, 2011. **2**(11): p. 918-39.
248. Poole, R.J., et al., *A Genome-Wide RNAi Screen for Factors Involved in Neuronal Specification in Caenorhabditis elegans*. PLoS Genet, 2011. **7**(6): p. e1002109.
249. Nelson, M.D., et al., *A bow-tie genetic architecture for morphogenesis suggested by a genome-wide RNAi screen in Caenorhabditis elegans*. PLoS Genet, 2011. **7**(3): p. e1002010.
250. Struwe, W.B., et al., *Modeling a congenital disorder of glycosylation type I in C. elegans: a genome-wide RNAi screen for N-glycosylation-dependent loci*. Glycobiology, 2009. **19**(12): p. 1554-62.

251. Sugimoto, A., *High-throughput RNAi in Caenorhabditis elegans: genome-wide screens and functional genomics*. Differentiation, 2004. **72**(2-3): p. 81-91.
252. Sonnichsen, B., et al., *Full-genome RNAi profiling of early embryogenesis in Caenorhabditis elegans*. Nature, 2005. **434**(7032): p. 462-469.
253. Ceron, J., et al., *Large-scale RNAi screens identify novel genes that interact with the C. elegans retinoblastoma pathway as well as splicing-related components with synMuv B activity*. BMC Dev Biol, 2007. **7**: p. 30.
254. Rual, J.F., et al., *Toward improving Caenorhabditis elegans phenome mapping with an ORFeome-based RNAi library*. Genome Res, 2004. **14**(10B): p. 2162-8.
255. Cooper, A., *Microcalorimetry of protein-protein interactions*. Methods Mol Biol, 1998. **88**: p. 11-22.
256. Falconer, R.J. and B.M. Collins, *Survey of the year 2009: applications of isothermal titration calorimetry*. J Mol Recognit, 2011. **24**(1): p. 1-16.
257. Dolphin, C.T. and I.A. Hope, *Caenorhabditis elegans reporter fusion genes generated by seamless modification of large genomic DNA clones*. Nucleic Acids Research. **34**(9): p. e72.
258. Copeland, N.G., N.A. Jenkins, and D.L. Court, *Recombineering: a powerful new tool for mouse functional genomics*. Nat Rev Genet, 2001. **2**(10): p. 769-79.
259. Tursun, B., et al., *A Toolkit and Robust Pipeline for the Generation of Fosmid-Based Reporter Genes in *C. elegans**. PLoS ONE, 2009. **4**(3): p. e4625.
260. Gibson, D.G., *Enzymatic assembly of overlapping DNA fragments*. Methods Enzymol, 2011. **498**: p. 349-61.
261. Gibson, D.G., et al., *Enzymatic assembly of DNA molecules up to several hundred kilobases*. Nat Methods, 2009. **6**(5): p. 343-5.
262. Mello, C., et al., *Efficient gene transfer in C.elegans: extrachromosomal maintenance and integration of transforming sequences*. EMBO J, 1991. **10**(12): p. 3959 - 3970.

263. Giordano-Santini, R., et al., *An antibiotic selection marker for nematode transgenesis*. Nat Methods, 2010. **7**(9): p. 721-3.