



Régulation post-transcriptionnelle du gène *unc-54* de *Caenorhabditis elegans* identifiée in vivo par un système de double rapporteurs fluorescents

Cécile Quéré

► To cite this version:

Cécile Quéré. Régulation post-transcriptionnelle du gène *unc-54* de *Caenorhabditis elegans* identifiée in vivo par un système de double rapporteurs fluorescents. Génétique. Université de Bordeaux, 2014. Français. [⟨NNT : 2014BORD0347⟩](#). [⟨tel-01140643⟩](#)

HAL Id: tel-01140643

<https://theses.hal.science/tel-01140643v1>

Submitted on 9 Apr 2015

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

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



HAL Authorization

THÈSE PRÉSENTÉE
pour l'obtention du
DIPLÔME NATIONAL DE DOCTORAT
DÉLIVRÉ PAR L'UNIVERSITÉ DE BORDEAUX

École doctorale Sciences de la vie et de la santé

Mention : Sciences, Technologies, Santé

Option : Génétique

Présentée et soutenue publiquement

Le 17/12/2014

Par Cécile QUÉRÉ

**Régulation post-transcriptionnelle du
gène *unc-54* de *Caenorhabditis elegans*
identifiée *in vivo* par un système de
double rapporteurs fluorescents**

Membres du jury

M Ijsbrand KRAMER	Professeur à l'Université de Bordeaux	Président
Mme Sophie JARRIAULT	DR2 CNRS Université de Strasbourg	Rapporteur
M Hervé SEITZ	CR1 CNRS Université de Montpellier	Rapporteur
M Christophe GROSSET	CR1 INSERM Université de Bordeaux	Membre invité
M Denis DUPUY	CR1 INSERM Université de Bordeaux	Directeur de thèse

À mes grands-pères.

Remerciements

La thèse est l'étape essentielle du début de la formation d'un jeune chercheur, ou d'une jeune chercheuse. C'est un travail qui ne se fait pas seule, alors forcément, il y a beaucoup de personnes à remercier.

Mes remerciements vont en premier à mon directeur, Denis Dupuy qui a su me guider dans mon projet scientifique et toute la communication autour de ce projet. Merci aussi de m'avoir donné la chance de participer à des congrès, d'avoir su me guider dans la rédaction de documents, et de m'inclure dans d'autres projets scientifiques et d'avoir su partager cette curiosité et cet émerveillement constant face à la science et ses multiples possibilités.

Je tiens particulièrement à remercier les membres de mon jury qui m'ont fait l'honneur de donner de leur temps pour lire et juger mes travaux, Isjbrand Kramer, Christophe Grosset, Hervé Seitz et Sophie Jarriault.

Je remercie aussi mes collègues passés et présents, Rosina, Karine, Esther et Ilyass qui m'ont appris toutes les techniques relative à *C. elegans* et aussi beaucoup aidé dans le développement de mon projet le tout dans la joie et la bonne humeur. Je remercie aussi mon collègue Jonathan Millet pour ces deux ans et demi passés ensemble dans le laboratoire et les nombreuses discussions scientifiques. J'ai aussi eu la chance de travailler avec plusieurs stagiaires dont les compétences techniques ont permis d'avancer, alors merci à Matthias, Sophie, Camille, Emilie et Marine pour votre aide.

Plus généralement, je souhaite remercier l'ensemble de mes collègues de l'IECB qui a été un environnement de travail formidable et m'a permis de rencontrer des gens passionnants et passionnés, jamais avarés de conseils et toujours prêts à échanger. Merci à toutes ces personnes pour leur soutien scientifique et personnel, j'espère qu'elles se reconnaîtront même si je ne les nomme pas. Merci aussi aux comités successifs de l'organisation du young scientist symposium de m'avoir donné la chance de participer à cet évènement et aux équipes IGEM pour cette expérience fabuleuse.

Ces trois ans ont été intenses et je me dois de remercier mes proches d'avoir supporté ça et d'avoir cru en moi du début à la fin. D'abord merci à ma maman, pour sa confiance aveugle et éternelle et son soutien sans faille, et bien sûr à Christophe qui en a tout autant. Je remercie aussi mes grand-mères qui sont pour moi une source d'inspiration quotidienne par leur force et leur caractère. Merci aussi au reste de la famille pour leur soutien.

Il faut une mention spéciale rien que pour Nicolas, qui non content de m'avoir subie 5 années à la fac, m'a suivi (bien involontairement) 3 années de plus dans l'aventure de la thèse. Les longues soirées, les doutes, les discussions scientifiques, personnelles, la rédaction, merci d'avoir traversé tout ça avec moi. Aux autres qui n'ont pas souhaité nous suivre dans la belle aventure de la thèse, Flo, Markus, Léa et Virginia, merci de votre soutien et d'être là tous les jours sans faute. A ceux dont ce sera bientôt le tour, Jérémie et Chloé, vous êtes formidables, et je crois en vous comme vous avez cru en moi. Noël-Arnaud, je ne sais pas comment j'aurai réussi à démarrer en face de la page blanche et à gérer mon stress sans toi pour venir me sortir du labo pour aller déjeuner ou rien qu'avec des petits messages. Enfin à tous les autres dont les encouragements ont été constants sur ces dernières années et sans qui les choses seraient bien moins chouettes, Nolwenn, Lionnel, Rill, Raphaëlle, Marie, Clémence, Diane (les deux !!), Pidjey et Cassandra.

Enfin, merci à Thomas d'être là tous les jours, d'avoir cru en moi dès notre rencontre et de m'épauler dans ma progression tant professionnelle que personnelle, et de me prouver que la route est moins difficile quand on est accompagnée

« La vérité pure et simple est très rarement pure et jamais simple. »

Oscar Wilde

Table des matières

Liste des abréviations	10
I. Introduction	12
1. L'expression génétique et sa régulation	13
1. Généralités	13
A. Mécanismes des régulations	13
B. Les maturations post-transcriptionnelles des ARN	15
C. Les régulations post-transcriptionnelles	18
2. Les régulations post transcriptionnelles par les microARN	18
A. Historique et enjeux présents	18
B. Localisation et transcription	20
C. Cibles et fonctionnement	22
2. Le modèle <i>Caenorhabditis elegans</i>	24
1. Utilisation en laboratoire	24
A. Histoire et faits marquants	24
B. Modèle biologique	26
C. Ressources disponibles	26
2. Généralités	28
A. Milieu naturel	28
B. Anatomie	29
C. Cycle de vie	30
3. La génétique de <i>C. elegans</i>	31
1. Organisation du génome	31
A. Les chromosomes	31
B. Les opérons	31
2. Transgénése	34
A. La micro-injection	34
B. Le bombardement	35
C. Devenir de l'ADN exogène	36
3. Les microARN chez <i>C. elegans</i>	39
A. Découverte et connaissances	39
B. Méthodes d'étude	41
4. Génération d'un rapporteur de régulation post-transcriptionnelle	43
1. Fonctionnement théorique	43
2. Sélection des gènes de l'étude	45
II. Objectifs des travaux de thèse	48

III. Matériel et Méthodes	50
1. Biologie moléculaire	51
1. Sélection des séquences régulatrices : promoteur et 3'UTR	51
2. Préparation du matériel génétique	52
A. Préparation d'ADN	52
B. Préparation d'ARN	53
3. PCR	53
A. Rétro-Transcription d'ARN	54
4. Chew back annealing	55
5. Mutagenèse dirigée	56
6. Génération de plasmides	57
2. Protocoles de <i>C. elegans</i>	57
1. Maintenance	57
2. Décontamination et synchronisation des vers	58
3. Transgénèse	59
4. Microscopie	59
A. Coloration au Dil	60
5. Mutagenèse à l'EMS	60
6. Croisements de lignées	61
IV. Résultats	62
1. Utilisation d'un rapporteur de régulation post-transcriptionnelle sous forme d'opéron	63
1. Développement du projet	63
2. Analyse phénotypique des lignées transgéniques	64
A. Détermination du nombre d'individus à analyser	65
B. Avantage du contrôle interne	71
C. Comparaison des fluorescences	72
3. Vérification des différents éléments de la construction	79
A. Suppression de l'opéron linker : co-injection des rapporteurs	79
B. Les protéines fluorescentes présentent le même profil d'expression placées sous le contrôle des mêmes séquences régulatrices	81
2. Caractérisation de la régulation post-transcriptionnelle du gène <i>unc-54</i>	83
1. Identification des neurones présentant une inhibition de la séquence <i>unc-54</i> 3'UTR	83
A. Détermination de la morphologie des neurones	83
B. Confirmation de l'identification par coloration au Dil	84
2. Mutagenèse dirigée sur la séquence <i>unc-54</i> 3'UTR	85
A. Analyse <i>in silico</i> de la séquence et choix des sites cibles	86
B. Résultats des délétions ciblées	86
3. Crible génétique à l'EMS	89

4.	Analyse de l'expression du gène <i>unc-54 in vivo</i>	93
5.	Potentiels gènes cibles de la régulation opérée sur la région <i>unc-54</i> 3'UTR	96
A.	Sélection des gènes candidats	96
B.	Expression du gène <i>osm-9</i> et analyse	97
6.	Analyse du site 2 muté responsable du phénotype observé	100
3.	Biologie moléculaire : clonage par <i>Chew Back Annealing</i>	104
1.	Le choix d'une méthode novatrice	104
2.	Test de l'efficacité de la méthode	105
V.	Discussion	110
1.	Les constructions sous forme d'opéron	111
1.	La cassette RRPT	111
2.	Les applications des constructions en opéron	112
2.	La régulation du gène <i>unc-54</i>	114
1.	La région <i>unc-54</i> 3'UTR subit une régulation post-transcriptionnelle	114
2.	Vers l'élucidation de la voie de régulation	115
3.	Etude du gène <i>osm-9</i>	116
3.	Le choix de la séquence 3'UTR contrôle	116
4.	Conclusion	117
VI.	Bibliographie	120
VII.	Annexes	134
1.	Liste des amorces	135
2.	Liste des plasmides	138
3.	Liste des lignées	140
4.	Alignements de séquence	144

Liste des abréviations

3'UTR : région 3' non traduite

5'UTR : région 5' non transcrite

ADN : Acide désoxyribonucléique nucléique

ADNc : ADN complémentaire

ADNg : ADN génomique

ARN : Acide ribonucléique nucléique

ARNi : ARN interférence

ARNm : ARN messenger

ARNt : ARN de transfert

CRISPR : clustered regularly interspaced short palindromic repeats

EMS : Ethyl methane sulfonate

EST : Expressed sequence tags

GFP : Green Fluorescent Protein (protéine de fluorescence verte)

IRES : Internal Ribosomal entry site

kDa : kiloDalton

miARN : microARN

NGM : Nematode growth media (milieu de croissance pour nématode)

NMD : non-sense mediated decay

PAS : PolyAdenylation Signal (signal de polyadénylation)

PCR : Polymerase chain reaction **RT-PCR** : Retro-transcription PCR

pb : paire de base

kb : kilobase

pré-ARNm : ARN pré-messenger

pré-miARN : pré-microARN

pri-miARN : pri-microARN

RISC : RNA induced silencing complex

RRPT : rapporteur de regulation post-transcriptionnel

siARN : short interfering RNA

SL : splice leader

snRNP : small nuclear ribonucleoprotein (petite ribonucléoprotéine nucléaire)

I. Introduction

1. L'expression génétique et sa régulation

1. Généralités

L'Acide désoxyribonucléique (ADN) est le support de l'information génétique et contient toutes les informations nécessaires à la formation d'un organisme vivant. Une légère variation suffit à générer deux espèces différentes. Outre les différences entre les espèces, des variations génétiques intra espèces conduisent au développement d'individus différents (couleurs des yeux, du poil), parfois plus fragiles ou non viables. La défaillance de l'expression correcte des gènes conduit à l'apparition de pathologies qui peuvent se déclarer à différents stades du développement. Les maladies génétiques ou l'apparition des cancers sont intrinsèquement liées à des défauts d'expression des gènes, qu'il s'agisse d'une absence d'expression ou au contraire, d'une surexpression des gènes concernés. Les gènes présents doivent ainsi être exprimés au bon moment et au bon endroit pour permettre le développement d'un individu.

Chez les eucaryotes, l'expression des gènes doit être régulée en termes de localisation et de temps afin de produire un organisme sain. L'expression des gènes conditionne le devenir des cellules au cours du développement, et les gènes exprimés dans différents tissus ne sont pas les mêmes. Par exemple, l'expression variable de gènes permet de produire des cellules différenciées comme les cellules du cœur, de la peau et du cerveau. Pourtant si ces cellules appartiennent au même organisme, alors elles sont porteuses du même ADN, et par conséquent de la même information génétique. La manière dont l'expression des gènes est régulée au sein du règne du vivant est donc centrale, et à ce si l'ensemble des mécanismes qui la régissent est connu, l'organisation de sa régulation ne l'est pas.

A. Mécanismes des régulations

L'ADN a été reconnu comme support de l'hérédité au début des années 1950 (Avery, Macleod, and McCarty 1944). La structure de la molécule d'ADN est une double hélice, formée de deux brins composés de nucléotides ou bases (Watson and Crick 1953). Si les protéines sont connues depuis les années 1830, l'intermédiaire entre ADN et protéine est découvert dans les années 1960. Cet intermédiaire est l'ARN messager (ARNm) un transcrit complémentaire à la séquence d'ADN (S. Brenner, Jacob, and Meselson 1961). L'expression des gènes est ainsi basée sur un système à 3 niveaux : l'ADN support de l'information, les ARNm et les protéines assurant les fonctions cellulaires.

L'ADN est porteur de l'information génétique, présente sous forme d'unités appelées gènes. La définition actuelle d'un gène correspond à une séquence de l'ADN donnant lieu à une

transcription en ARN. La séquence transcrite en ARN est précédée d'une séquence appelée promoteur, qui permet la fixation de la machinerie de transcription en ARN.

Aujourd'hui, la mécanique générale d'expression des gènes est connue chez les procaryotes et les eucaryotes. Chez les eucaryotes, l'ADN est transcrit en ARN dans le noyau (Figure 1). La transcription, première étape de cette interprétation, est le mécanisme permettant le passage de la molécule d'ADN à l'ARN, intermédiaire indispensable avant la synthèse des protéines. La transcription débute par l'ouverture de la double hélice d'ADN grâce à l'action d'hélicases décrites dès la fin des années 1970 (Kornberg, Scott, and Bertsch 1978; revue de Lohman, Tomko, and Wu 2008). La région du promoteur devenue accessible va alors recruter des facteurs de transcription, formant un complexe de pré-initiation et finalement une ARN polymérase (Barberis and Gaudreau 1998; T. I. Lee and Young 2000; Cosma, Tanaka, and Nasmyth 1999). Les ARN polymérases (Pol) sont au nombre de trois chez les eucaryotes et ont été nommées au moment de leur découverte en fonction de l'ordre dans lequel elles sont purifiées par chromatographie sur colonnes DEAE-Sephadex (Roeder and Rutter 1969). La Pol I est recrutée pour la transcription des ARN ribosomiques et la Pol III pour les ARN courts non codants (ARNt, ARN 5S, snARN...) (Tamura et al. 1996). La Pol II est responsable de la transcription des gènes et des miARN. Une fois la polymérase fixée la phase d'élongation peut commencer. L'accessibilité au promoteur d'un gène et les facteurs qui la régulent constituent une première couche de régulation, les régulations transcriptionnelles.

La polymérase se déplace le long de l'ADN tout en synthétisant le brin d'ARN. Des protéines de liaison à l'ARN se fixent au cours de la synthèse et débute les processus de maturation post-transcriptionnelle : ajout d'une coiffe 7-méthylguanosine en 5', polyadénylation en 3' et édition (Moore 2005; Orphanides and Reinberg 2002). La terminaison de la transcription est déterminée par plusieurs signaux dont souvent le signal de polyadénylation (PAS). L'ARN transcrit est ensuite exporté dans le cytoplasme où il va être pris en charge par les facteurs de traduction et les ribosomes pour la synthèse des protéines. A cette étape, un second niveau de régulation est présent pour médier l'acheminement et la maturation des ARN, les régulations post-transcriptionnelles.

Une fois matures, les ARNm sont pris en charge par les ribosomes qui assurent l'étape de traduction en protéine. Un troisième niveau de régulation dit post-traductionnelle peut intervenir à cette étape en modifiant le profil d'expression de la protéine.

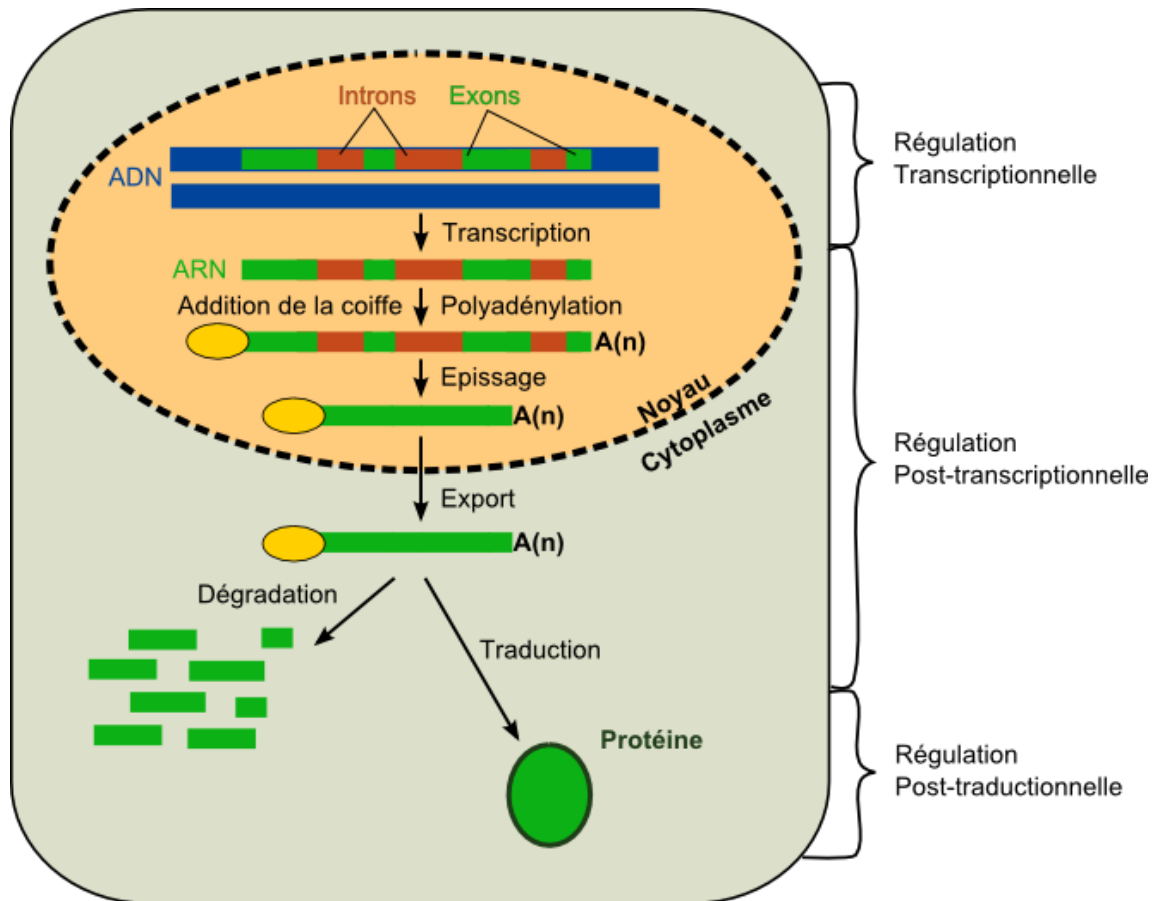


Figure 1 Schéma général de l'expression des gènes et leur régulation.

L'expression des gènes dépend donc de nombreux facteurs qui interagissent pour permettre ou réprimer celle-ci en fonction des besoins. Des niveaux supplémentaires de régulations post-transcriptionnelles et post-traductionnelles permettent d'affiner encore cette expression. Ces régulations fonctionnent sous forme de réseaux qui interagissent et ne sont pas complètement élucidés à ce jour.

Cette organisation aboutit à un flux d'information partant de l'ADN passant par les ARNm et aboutissant à la formation de protéines. Cependant, certains ARN sont transcrits mais non traduits en protéines. Leur maturation conduit à un ARN fonctionnel qui agit au sein de complexes comme par exemple les ARN ribosomiaux, les ARN de transfert (ARNt) ou les microARN (miARN).

B. Les maturations post-transcriptionnelles des ARN

Les régions codantes sont constituées des gènes qui peuvent être organisés indépendamment ou en opéron. Les gènes contiennent l'information nécessaire à la synthèse des protéines. Les cadres de lectures ouverts correspondent à la protéine et commencent classiquement par un codon ATG, correspondant à l'acide aminé méthionine. Ils se terminent par un codon stop ocre (AAT), ambre (TAG) ou opale (TGA). Les ARN sont synthétisés à partir du promoteur du gène,

site de fixation du complexe de transcription. Les maturations post-transcriptionnelles démarrent pendant la phase d'élongation de l'ARN.

L'ARN pré-messager (pré-ARNm) va d'abord subir la modification de son extrémité 5', initialement formée d'un nucléotide 5' triphosphate. L'extrémité triphosphate est clivée par une triphosphatase, puis une guanylyltransférase forme une liaison 5'-5' triphosphate. Finalement une méthyltransférase ajoute les groupements méthyle, formant ainsi la coiffe définitive du pré-ARNm (Banerjee 1980; Shuman 2001).

A la fin de la transcription des pré-ARNm eucaryotes, ceux-ci subissent l'addition de la queue poly(A) (Edmonds et al. 1971; Darnell, Wall, and Tushinski 1971). La région 3'UTR de l'ARNm est clivée suite à un signal (PAS) consensus AAUAAA situé à 10-30 nucléotides du site de fixation de la queue poly(A). L'étape de l'ajout de la queue poly(A) est réalisée par un complexe formé d'une poly-A polymérase et d'une protéine de liaison poly-A (Edmonds 2002).

Dans les années 1970, les techniques de séquençages (Sanger, Nicklen, and Coulson 1977; Maxam and Gilbert 1977) permettent de comparer de courtes séquences d'ADN génomique (ADNg) à des séquences d'ARN rétro-transcrits (ADN codant, ADNc). Le séquençage met rapidement en évidence que les séquences des gènes de l'ADNg présentent les séquences codantes séparées par des régions non codantes. Les séquences codantes constituent les exons, qui sont exportés dans le cytoplasme pour la traduction alors que les introns sont éliminés au cours de la maturation (Gilbert 1978). Chez les eucaryotes, l'épissage est une étape essentielle de la maturation de l'ARN. Les étapes d'épissage commencent également au moment de la synthèse, mais ne s'achèvent qu'après la synthèse de l'ARN (Singh and Padgett 2009). Les introns sont systématiquement éliminés au cours de l'épissage et parfois certains exons aussi, ce qui correspond à l'épissage alternatif. L'épissage alternatif conduit à l'élimination de certains exons ce qui permet de générer différents ARNm à partir du même pré-ARNm et donc différentes protéines.

L'épissage des pré-ARNm est opéré à l'intérieur du noyau par des petits ribonucléoprotéides nucléaires appelés snRNP. Généralement, quatre snRNP sont impliqués dans l'épissage : U1, U2, U5 et le complexe U4/U6, ils constituent les composants principaux de la machinerie d'épissage, le spliceosome (revue de Wahl, Will, and Lührmann 2009; Brody and Abelson 1985). D'autres spliceosomes sont présents chez certains organismes et s'appuient sur des snRNP variants : U11 et U12. Les deux nucléotides en amont et en aval de chaque intron sont des signaux caractéristiques conservées en majorité et identifiées par le spliceosome. Au milieu de l'intron « une boîte de branchement » contient une adénosine strictement conservée et

indispensable au déroulement de la réaction d'épissage conduisant à l'excision de l'intron (Wahl, Will, and Lührmann 2009 ; Figure 2A).

La réaction se déroule en plusieurs étapes (Figure 2B). En premier lieu, l'ARN du complexe U1 interagit avec le site d'épissage situé en 5'. Les protéines de liaison à la boîte de branchement vont plier la molécule d'ARN au niveau de l'adénosine conservée. Ce complexe protéique inclut un facteur de recrutement pour le snRNP U2. Le complexe U2 est responsable de l'identification des nucléotides des extrémités de l'intron. Le complexe U4/U6 rejoint le pré-spliceosome simultanément au complexe U5. Le recrutement des snRNP terminé, le spliceosome est activé par la séparation des complexes U4 et U6 qui provoque le relargage des complexes U1 et U4 et laisse le champ libre pour établir la liaison U6/U2. Les extrémités des exons et la boîte de branchement sont rassemblés par la protéine Prp-8. Deux réactions de transésterification vont permettre d'exciser l'intron qui est relargué sous forme de lasso et dégradé. Les extrémités des exons sont liguées entre elles et les snRNP recyclés dans un nouveau spliceosome.

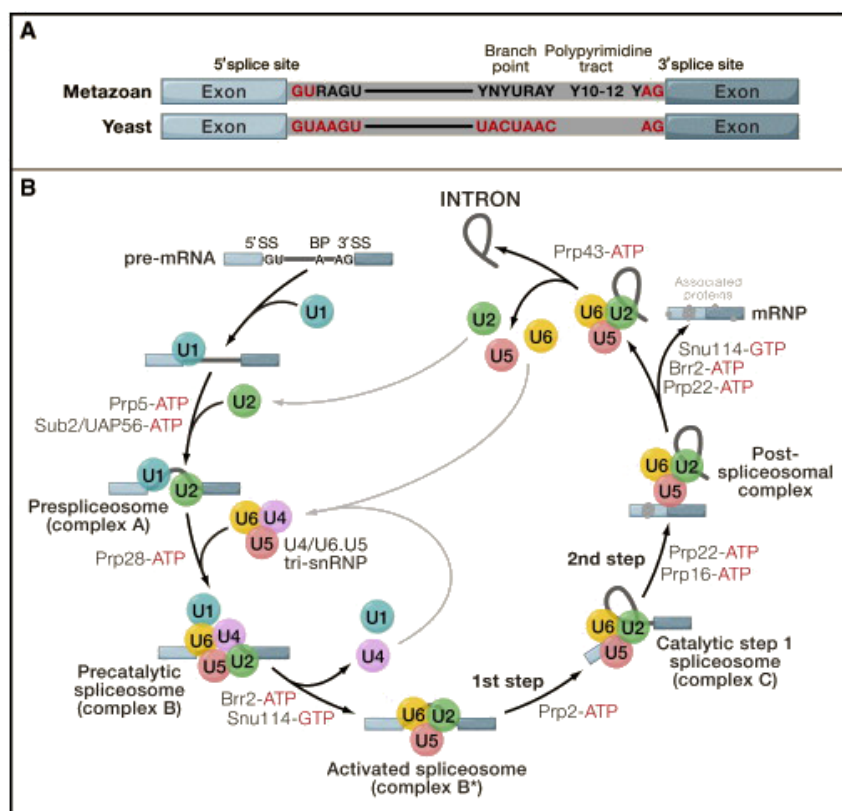


Figure 2 Assemblage du spliceosome et réaction d'épissage chez les eucaryotes (de (Wahl, Will, and Lührmann 2009)). **A** Séquences conservées en 5', 3' et au niveau de la boîte de branchement. **B** Le cycle d'assemblage et désassemblage du spliceosome se fait en plusieurs étapes successives. Les interactions entre snRNP sont décrites au cours de l'excision d'un intron. Huit protéines conservées au cours de l'évolution agissent à des stades spécifiques afin de catalyser les réactions d'interactions ARN-ARN et les changements de conformation des snRNP. Certaines de ces protéines, Prp5, Prp16 et Prp22 ont des fonctions de correction aux étapes où elles sont montrées.

C. Les régulations post-transcriptionnelles

Chez les eucaryotes il est répandu qu'un gène code pour plusieurs protéines grâce à l'épissage alternatif qui permet d'obtenir, à partir d'un seul ARN pré-messager, plusieurs ARN messagers (ARNm) distincts.

Le mécanisme d'épissage est le même que celui décrit précédemment. La régulation de l'épissage alternatif implique donc le contrôle des sites d'assemblage des spliceosomes. Différentes expériences ont permis de mettre en évidence l'importance d'éléments régulateurs en cis et de différents facteurs en trans dans la régulation de l'épissage alternatif. Les éléments cis régulateurs ont un rôle essentiel dans le recrutement des facteurs protéiques en trans. Les éléments en trans peuvent activer ou supprimer la reconnaissance des sites d'épissage par le spliceosome (House and Lynch 2008).

L'épissage alternatif peut conduire à la création d'un codon stop prématuré et orienter le transcrit vers la voie de dégradation médiée par le non-sens (NMD, non-sense mediated decay). Par exemple, l'analyse informatique du devenir des pré-ARNm humain a mis en évidence que les événements d'épissage conduisent une fois sur trois à la génération d'un substrat pour la voie NMD (B. P. Lewis, Green, and Brenner 2003). L'épissage alternatif constitue ainsi une strate de régulation supplémentaire de l'expression des gènes, qui s'ajoute aux mécanismes régulateurs de la transcription des ARN.

En plus de l'épissage alternatif, des régulations peuvent avoir lieu dans les régions non transcrites de l'ARNm, comme en 3'UTR. Les régions 3'UTR de certains transcrits peuvent présenter des sites cibles pour les microARN.

2. Les régulations post transcriptionnelles par les microARN

A. Historique et enjeux présents

La découverte des miARN et des ARN interférents (siARN) a eu lieu autour des années 1990. Ces molécules sont présentes dans le règne animal et végétal, bien qu'elles soient absentes chez certains eucaryotes unicellulaires comme la levure (L. Wang et al. 2014).

Les miARN ont été découverts chez un petit ver appartenant à l'ordre des Rhabditiae. L'organisme *C. elegans* est étudié depuis les années 60 pour son développement très régulier. Le développement des tissus cellulaires s'appuie sur une succession de divisions dont le résultat aboutit à un total de 959 cellules, après la mort cellulaire de 131 cellules. C'est en étudiant les gènes hétérochroniques, responsables du contrôle temporel du développement, que V. Ambros

et G. Ruvkun ont découvert la première régulation liée aux miARN. Les expériences sur des mutants indiquaient que le gène *lin-4* régulait négativement le gène *lin-14* (Ambros 1989). Il a été constaté que *lin-14* code pour une protéine mais que la séquence nécessaire à la régulation par *lin-4* est présente dans la région 3'UTR de l'ARNm (B. Wightman et al. 1991; Bruce Wightman, Ha, and Ruvkun 1993). Le clonage de *lin-4* a révélé que le gène n'encodait pas de protéine, mais la séquence clonée permettait bien le sauvetage du phénotype mutant associé à la délétion du gène. L'équipe de V. Ambros a ensuite mis en évidence que la séquence produisait de courts ARN de 25 et 60 nucléotides capable de former une structure tige-boucle. La collaboration entre G. Ruvkun et V. Ambros a permis de mettre en évidence la complémentarité entre la séquence de la région 3'UTR de l'ARNm de *lin-14* et la séquence des ARN courts produits par *lin-4* (R. C. Lee, Feinbaum, and Ambros 1993; Bruce Wightman, Ha, and Ruvkun 1993). Des expériences au cours desquelles les séquences complémentaires de ces gènes ont été délétées ont supprimé la répression de *lin-14* par *lin-4* mettant ainsi en évidence la première régulation ARN/ARN (R. C. Lee, Feinbaum, and Ambros 1993).

Quelques années plus tard, la découverte du miARN *let-7*, également responsable du contrôle temporel du développement, ouvre la voie des recherches sur la présence de miARN chez les autres espèces. En effet, la comparaison des données de séquençage indique que le gène *let-7* est extrêmement conservé au sein des chordés et des bilatériens (Pasquinelli et al. 2000).

La découverte des miARN a eu un effet important sur les recherches en génétique, particulièrement sur l'étude de la régulation des gènes. Par exemple chez les plantes, la régulation par ARN interférence a permis d'expliquer les phénomènes de résistance. Les petits ARN inhibiteurs des plantes ont été suspectés chez les plantes, et plus particulièrement dans des plants de tabac (Lindbo et al. 1993). Lorsque la plante est infectée par un virus, l'infection reste dans les feuilles inférieures et ne progresse pas par le méristème jusqu'aux feuilles supérieures. De plus, si la plante se remet de l'infection, elle ne sera plus infectée par le même virus. Ce phénomène n'était pas explicable par les mécanismes classiques d'immunité impliquant la formation d'anticorps comme chez les animaux (Hamilton and Baulcombe 1999; Voinnet and Baulcombe 1997). Des transgènes codant pour des ARNm des plantes ou des ARN viraux ont été utilisés dans les premières transgénèses de plantes. L'expression des ARNm des gènes s'accompagnait souvent de la répression du gène endogène de l'hôte ou du virus et les ARN exogènes semblaient être détruits de manière ciblée. La découverte de petits ARN d'environ 25 nucléotides, complémentaires des ARNm des gènes dont l'activité est éteinte, dans des plants de tomates et de tabac présentant le phénomène de répression associée à la présence de transgènes (viraux ou GFP) confirme le phénomène (Hamilton and Baulcombe 1999).

Les expériences suivantes ont révélé le rôle des miARN dans le développement, les suivantes ont confirmé cette hypothèse, notamment chez les souris, en mettant en évidence la relation entre délétion de miARN (*miR-133a1* et *miR-133a2*) et un phénotype entraînant la mort prématuré des sujets (A. H. Williams et al. 2009). Les miARN sont également dérégulés dans de nombreuses cellules tumorales, de la même manière que les protéines. Ils constituent au même titre que les protéines des marqueurs du développement tumoral (Calin and Croce 2006).

Actuellement, les miARN sont donc au cœur des recherches scientifiques tant pour leurs rôles au cours du développement, que dans des maladies comme les cancers.

Récemment, une étude a mis en évidence la présence des tRF qui sont de petits ARN issus des ARNt (Kumar et al. 2014) dans de nombreux organismes eucaryotes. Ces petits ARN ont été détectés en quantité équivalente aux miARN. Les tRF sont de petits ARN qui dérivent des ARNt et sont classés en trois catégories selon leur origine. Les tRF-5s et tRF-3s dérivent respectivement des extrémités 5' et 3' des ARNt matures, et les tRF-1s dérivent du précurseur des ARNt. Ces tRF semblent se retrouver dans l'ensemble du règne du vivant et présentent de grandes similitudes avec les miARN au niveau de leur maturation, notamment pour tRF-3s dont les 7 premiers résidus sont très conservés. Les tRF s'associent chez l'homme avec les protéines AGO-1, AGO-3 et AGO-4 et la méthode de CLASH a permis de détecter de nombreuses chimères tRF-3s et tRF-5s/ARNm. Ces résultats indiquent qu'ils jouent sans doute un rôle dans la régulation de l'expression des gènes en régulant la traduction des ARNm.

B. Localisation et transcription

Les régions codantes pour les miARN sont localisées dans les introns des gènes codant pour des protéines et/ou rassemblés en clusters. Les miARN sont donc transcrits majoritairement par l'ARN Pol II. Un petit groupe de miARN de certains organismes localisés dans des séquences répétées (centromères, télomères) peuvent être transcrits par la Pol III (Kim, Han, and Siomi 2009).

Les protéines de la voie de biogénèse des miARN

Les protéines Drosha et Pasha agissent sous la forme d'un complexe appelé microProcessor (Denli et al. 2004) responsable d'un premier clivage. La protéine Drosha, une RNase de type III, a été identifiée en 2003 et son rôle du clivage du pri-miARN au pré-miARN a été confirmé (Y. Lee et al. 2003). L'identification de Pasha (DGCR8 chez l'homme) a suivi de peu grâce à l'étude de différents mutants de Drosha (Han et al. 2004). Bien que son rôle exact ne soit pas connu, il est établi que Pasha augmente la stabilité des pré-miARN et *in vivo* l'action conjointe

des deux est nécessaire. La mutation de l'une des deux entraîne une réduction de pré-miARN et des miARN (Yeom et al. 2006).

Les premières étapes de maturation se déroulent dans le noyau des cellules alors que les miARNs agissent dans le cytoplasme. L'exportine 5 est une protéine déjà connue pour son rôle dans le transport de l'ARN non codant d'adénovirus (Gwizdek et al. 2003). La structure de l'ARN non codant des adénovirus présente des similitudes structurelles avec les pré-miARN : une mini-hélice à l'extrémité 5' et 3 nucléotides flottant à l'extrémité 3'. La similitude a conduit à examiner le rôle que la protéine Exp5 joue dans l'export des pré-miARN. Différentes expériences de fixation de pré-miARN *in vitro* ou de RNAi dirigée contre Exp-5 ont permis de confirmer le rôle de la protéine dans l'export des pré-miARN au cytoplasme (Lund et al. 2004).

Une fois dans le cytoplasme, une autre RNase de type III Dicer, achève la maturation du pré-miARN. Dicer a été parmi les premières protéines de la voie de biosynthèse à être identifiée (Bernstein et al. 2001). L'activité de Dicer a été confirmée par un essai *in vitro* après immunoprécipitation de la protéine.

La biogénèse

Ces enzymes agissent en cascade dans la biogénèse des miARN. Le transcrit primaire, le pri-miARN, mesure entre 150 et 200 nucléotides. Il va subir différentes étapes de maturation, par des enzymes spécifiques et conservées entre les espèces (Figure 3). Le pri-miARN adopte une conformation de type tige-boucle (ou épingle à cheveux). Il est pris en charge par Drosha et Pasha. Le microProcessor coupe le pri-miARN, maintenant la structure tige-boucle, en pré-miARN de 60 à 70 nucléotides. Le pré-miARN est alors exporté dans cytoplasme par l'exportin-5 (Exp5). Une fois dans le cytoplasme, les pré-miARN sont pris en charge par l'enzyme Dicer. L'enzyme Dicer réalise une coupure au niveau de la boucle terminale, ce qui a pour effet de générer un duplex de deux brins d'ARN de 22 nucléotides, les miARN.

Pour jouer leur rôle, les miARN doivent être incorporés dans un complexe protéique responsable de la répression des gènes par ARN : le RISC (*RNA induced silencing complex*). Le complexe RISC est initié par les protéines de la famille Argonaute, dès que le duplex d'ARN est pris en charge par ces dernières. Ultérieurement, l'un des deux brins du duplex d'ARN va être éliminé et dégradé, ne laissant qu'un brin, le miARN intégré dans le complexe final miRISC (Carthew and Sontheimer 2009). L'autre brin, appelé brin passager ou miARN*, est dégradé.

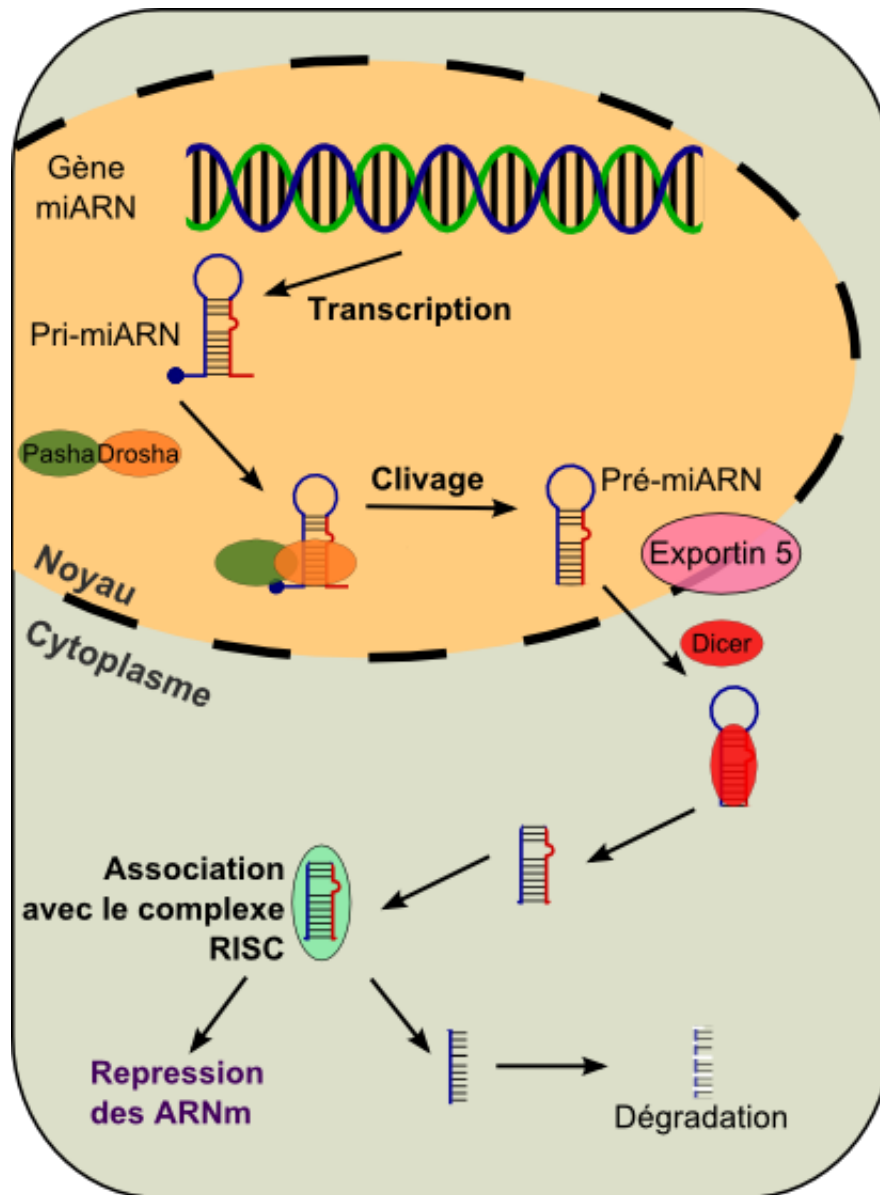


Figure 3 Biogénèse des miARN. Le gène d'un miARN est transcrit en Pri-miARN. L'enzyme Drosha réalise un premier clivage libérant les extrémités 5' et 3', formant un pré-miARN. L'exportin 5 permet le passage du pré-miARN dans le cytoplasme où Dicer réalise le clivage de la boucle apicale. Le duplex est finalement dissocié au moment de son intégration dans le complexe RISC.

Le complexe RISC peut cibler les ARNm grâce à la présence du miARN pouvant s'hybrider avec les régions 3'UTR des ARNm.

C. Cibles et fonctionnement

Les régions 3'UTR des ARNm contiennent des sites cibles pour les miARN (He and Hannon 2004). Une seule région 3'UTR d'un ARNm donné peut ainsi être la cible d'un ou plusieurs miARN. Il n'est pas rare que les séquences 3'UTR contiennent plusieurs sites cibles pour un même miARN. Cela semble renforcer l'efficacité de la répression opérée par le complexe RISC (Doench, Petersen, and Sharp 2003). Les miARN peuvent aussi se fixer à des séquences

complémentaires dans le cadre ouvert de lecture (Shen et al. 2008; Duursma et al. 2008; Forman, Legesse-Miller, and Coller 2008) ou bien à l'extrémité 5' d'un ARNm (Rigoutsos 2009; Grey et al. 2010; I. Lee et al. 2009; Rigoutsos 2009; Xu et al. 2014). De plus en plus d'indices fournis par les études des extraits d'ARN nucléaires purifiés (Z. F. Li et al. 2013; Hwang, Wentzel, and Mendell 2007; Liao et al. 2010; Park et al. 2010) indiquent également les certains miARN seraient adressés au noyau et agirait comme régulateur directement sur l'ADN en formant des triplex (Toscano-Garibay and Aquino-Jarquin 2014)

Le complexe RISC identifie ses cibles grâce à la complémentarité de séquence existant entre le miARN qu'il porte et l'ARNm. Selon la complémentarité des séquences entre le miARN et l'ARNm, l'appariement peut être parfait ou imparfait. En cas d'appariement entre le miARN et l'ARNm, celui-ci est dégradé par l'activité endonucléase du complexe RISC. Si l'appariement est imparfait entre le miARN et sa cible, le complexe RISC va simplement provoquer une répression de la traduction (Bartel 2004). La voie majeure de la régulation par les miARN semble bien être la répression de la traduction. Par exemple, dans les cellules humaines, les miARN inhibent la synthèse de protéines à des stades précoces. Les ARNm ainsi régulés sont stockés dans des organelles intracellulaires, les P-bodies (Pillai et al. 2005).

Avec le développement des technologies de détection, beaucoup de petits ARN sont actuellement découverts et ajoutés aux bases de données malgré l'absence de validation fonctionnelle. Or, il est essentiel de distinguer les petits ARN fonctionnels d'un « bruit » pouvant provenir de la dégradation d'ARN plus longs (Kim, Han, and Siomi 2009). L'enrichissement des bases de données en parallèle du développement de l'outil informatique a permis la création d'algorithmes prédisant les interactions entre ARNm et miARN. Les algorithmes développés comprennent TargetScan (Benjamin P. Lewis, Burge, and Bartel 2005) pour les prédictions chez les vertébrés, la drosophile, le nématode et le poisson zèbre ; DIANA MicroT (Maragkakis et al. 2009) spécifique de l'humain et la souris ; miRanda, miRNA-Target prediction et RNAhybrid (Rehmsmeier et al. 2004). L'algorithme PicTar (Krek et al. 2005; Lall et al. 2006) est le programme de prédilection dans les études de régulation par les miARN chez le nématode *C. elegans*. Sa particularité est de prendre en compte la conservation des séquences entre différentes espèces de nématodes.

La validation expérimentale des données obtenues par des algorithmes prédictifs des interactions ARNm/miARN est devenue cruciale afin de mieux cerner la régulation de l'expression génétique.

2. Le modèle *Caenorhabditis elegans*

1. Utilisation en laboratoire

L'attrait des scientifiques pour l'utilisation de l'espèce *C. elegans* comme modèle biologique a démarré dans les années 1960 et n'a cessé de croître.

A. Histoire et faits marquants

L'introduction de *Caenorhabditis elegans*, ver microscopique de la famille des nématodes, comme organisme modèle a été bien plus tardive que pour d'autres modèles, telle la drosophile, utilisée depuis le début des années 1900. Avant les années 1970, un très petit nombre de publications concernent cet animal.

L'espèce a d'abord été décrite et nommée dans les années 1900 par Emile Maupas (Maupas 1901), biologiste français rattaché à l'Académie des Sciences. Initialement nommé *Rhabditis elegans*, le nom a été actualisé en 1952 pour *Caenorhabditis elegans* d'un mélange de grec et de latin. *Caeno* signifie « récent », *Rhabditis* (son premier nom) signifie bâtonnet et *elegans* lui fut conféré pour sa locomotion élégante.

C'est à la fin des années 1940 que des chercheurs se penchent sur l'intérêt de *C. elegans* dans la recherche (Dougherty and Calhoun 1948) et vont étudier les différentes souches cultivées à l'époque. Initialement, ce sont deux souches qui ont été utilisées pour la culture en laboratoire. Les deux ont été nommées d'après les villes où elles ont été isolées. La première, Bergerac a été isolée par Victor Nigon de l'Université de Lyon (Nigon 1949). La seconde, Bristol, a été isolée d'un compost de champignon un peu plus tard (Dougherty and Hansen 1956). La différence majeure entre ces deux souches concerne leurs croissances. En effet, la souche Bergerac est stérile au-dessus de 18°C quand la souche Bristol peut croître jusqu'à 25°C sans problème. Cette première disparité a immédiatement intéressé les généticiens qui ont constaté qu'ils pouvaient y associer une ségrégation mendélienne et donc un trait génétique (Fatt and Dougherty 1963). C'est finalement à partir de la lignée Bristol qu'ont été isolés les vers de la lignée N2, souche isolée par S. Brenner en 1964.

Sydney Brenner introduisit *C. elegans* comme organisme modèle dans les années 1970 afin d'étudier la différenciation cellulaire et le développement d'un organisme pluricellulaire. Il publie en 1973 et 1974 les premiers articles sur le développement et le comportement (Sydney Brenner 1973; S. Brenner 1974). La méthodologie pour obtenir des mutants et les analyser y est décrite avec beaucoup de précision, ainsi que les nombreux atouts de cet eucaryote pour la science. En effet, le potentiel de *C. elegans* est remarquable pour les généticiens, puisque

l'espèce est hermaphrodite, ce qui permet de démarrer une population à partir d'un seul individu, et présente une apparition spontanée de mâles ce qui permet tout de même les croisements. Grâce à son temps de développement inférieur à 3 jours et au grand nombre d'œufs pondus par hermaphrodite, une grande population peut être générée rapidement, à partir de quelques individus.

Sa petite taille, inférieure à 1,5 mm ainsi que sa transparence permettent des études de microscopie sur l'animal entier, sans recourir à la dissection. Les propriétés chimiques (pourcentage de bases GC) de l'ADN de *C. elegans* est également publiée par S. Brenner et J. Sulston cette année-là (J. E. Sulston and Brenner 1974).

J. Sulston, un collaborateur de S. Brenner, a travaillé principalement sur le lignage cellulaire en cartographiant les divisions successives qui se produisent au cours du développement. Il a ainsi établi les bases du lignage des cellules et également observé le phénomène de mort programmée des cellules au cours du développement (J. E. Sulston 1976).

Le phénomène de mort programmée a ensuite été étudié par un troisième collaborateur, R. Horvitz. Il identifie les premiers gènes responsables de la mise en route du processus de mort cellulaire programmée ainsi que les conséquences sur l'organisme (Ellis and Horvitz 1986).

Ces travaux trouveront rapidement des échos au sein de la communauté scientifique, et les phénomènes de régulations et de développement identifiés chez le nématode se vérifieront chez les mammifères. C'est en 2002 que ces travaux sont finalement récompensés par un prix Nobel.

D'autres prix Nobel suivront celui-ci. En 2006, A. Fire et C. Mello reçoivent également ce prix pour leur découverte de l'ARN interférence qui permet d'éteindre les gènes de manière spécifique (Andrew Fire et al. 1998). En 2008, c'est un prix Nobel de chimie qui récompense conjointement chimistes et biologistes. Parmi eux M. Chalfie, qui introduisit avec succès la GFP chez *C. elegans* créant ainsi le premier rapporteur fluorescent *in vivo* (Chalfie et al. 1994).

Les récompenses des prix Nobel et les avancées qui ont été permises par ces travaux d'envergure ont été complétées par un autre exploit technique. En effet, en 1998, le génome de *C. elegans* est le premier génome d'un organisme pluricellulaire à être entièrement séquencé (Consortium 1998). Ce travail a ouvert la voie au séquençage du génome humain, réalisé en 2003.

Les avancées réalisées grâce à *C. elegans*, ne se limitent pas qu'aux laboratoires terrestres. En 2003, embarqués dans la navette spatiale Columbia pour des expériences, des descendants furent retrouvés vivants, dans les débris. D'autres souches ont été renvoyées dans l'espace en

2009, pour des expériences sur l'impact de l'absence de gravité sur le développement musculaire (Adenle, Johnsen, and Szewczyk 2009; Y. Honda et al. 2014).

B. Modèle biologique

Au départ, *C. elegans* est utilisé par S. Brenner qui en établit la première carte génétique (S. Brenner 1974). Suivent, peu de temps après, la cartographie intégrale du lignage cellulaire de la fertilisation de l'œuf au stade adulte (J. E. Sulston and Horvitz 1977; John E. Sulston, Horvitz, and Kimble 1988) et la cartographie du réseau neuronal reconstituée à partir de coupes d'animaux analysées en microscopie électronique (J. G. White et al. 1986). Ces études sont rendues possibles par la transparence de l'animal qui permet le suivi des divisions cellulaires.

La mise en œuvre des mêmes approches génétiques permet d'étudier le développement embryonnaire, la détermination du sexe et le développement larvaire (Cassada and Russell 1975; Hodgkin and Brenner 1977; Vanderslice and Hirsh 1976). L'un des premiers gènes étudié a été celui codant pour une chaîne de myosine lourde des cellules musculaires *unc-54*. La mutation du gène *unc-54* entraîne l'apparition d'un phénotype de type *uncoordinated* qui empêche l'animal de se mouvoir correctement. La protéine UNC-54 d'un poids moléculaire de 210 kDa compte parmi les plus abondantes et a donc été facilement analysée en SDS-PAGE (Epstein, Waterston, and Brenner 1974).

Aujourd'hui le nématode *C. elegans* est devenu un organisme modèle prisé, au même titre que la drosophile, la souris ou le rat. Les champs d'études se sont développés et comprennent toujours la Génétique et le Développement auxquels s'ajoutent le Vieillissement, les Neurosciences, l'Immunologie, l'Evolution et l'Ecologie. Son utilisation se justifie d'autant plus qu'environ 38% des protéines humaines ont un orthologue chez *C. elegans* (Berglund et al. 2007).

En plus de la capacité à survivre sur une longue période sans nourriture, les *C. elegans* survivent également à la congélation jusqu'à -80°C. Ces atouts majeurs font de ce petit ver l'un des eucaryotes les plus faciles d'utilisation pour la génétique.

C. Ressources disponibles

L'essor de la recherche scientifique autour des nématodes a été suivi par un développement conséquent des ressources qui a su s'appuyer sur les innovations technologiques des dernières décennies. Actuellement, la grande majorité des données disponibles est partagée grâce à des interfaces sur internet. Ce chapitre présente de manière non exhaustive les principales plateformes concernant *C. elegans*. Une revue de 2007 (Antoshechkin and Sternberg 2007)

répertorie en détail les bases de données, leurs rôles et incluent aussi les ressources pour les autres nématodes.

Les données du séquençage du génome de *C. elegans*, les données concernant son anatomie, ainsi que les différents protocoles expérimentaux créés sont accessibles par trois sites qui interagissent entre eux.

Le site *WormBase* (www.wormbase.org) met à disposition la séquence du génome de *C. elegans* ainsi que celle d'autres nématodes. Cette base de données est mise à jour au fur et à mesure des publications scientifiques afin d'actualiser les informations sur un gène et sa (ses) protéine(s). Les gènes codant pour des ARN non traduits sont également répertoriés et fournissent dans ce cas la séquence de l'ARN fonctionnel. La base de données est exploitée par des bases plus généralistes telles que *Ensembl*, *GenBank* et *Uniprot*. Les informations de base comme la séquence du gène, sont complétées selon les connaissances par d'autres données telles que les profils d'expression, les publications principales, les phénotypes mutants connus et les données d'homologie. Il est également possible de faire une recherche d'expression de gènes à partir d'un élément anatomique.

L'anatomie du vers est extrêmement bien documentée, et chaque cellule qui compose l'organisme est identifiée. Toutes les informations anatomiques concernant *C. elegans* sont regroupées sur le site *WormAtlas* (www.wormatlas.org) qui est mis à jour de manière régulière. Les informations sur les expressions des gènes spécifiques de tissus sont régulièrement ajoutées. Le site met également à disposition une banque de données d'images de microscopie facilitant l'identification des structures ainsi que les méthodes permettant d'aboutir à l'identification d'une structure d'intérêt.

L'ensemble des méthodes et des connaissances générales concernant le nématode sont rassemblées sur le site du *WormBook* (www.wormbook.org) qui succède aux livres précédemment édités *The nematode Caenorhabditis elegans* et *C. elegans II* (1997). Le site compte ainsi plus de 130 chapitres résumant tout ce qu'il faut savoir de la culture du nématode jusqu'aux protocoles détaillés pour l'analyse comportementale. De nouveaux chapitres sont régulièrement ajoutés, d'autres sont mis à jour en tenant compte des avancées techniques et scientifiques actuelles.

Ces trois sites sont connectés entre eux, et la communauté de chercheurs est fédérée autour d'un forum de discussion (<http://forums.wormbase.org/index.php>) ainsi qu'une lettre d'information *The Worm Breeder's gazette* où les contributions sont toujours les bienvenues. Il existe aussi *Textpresso* un moteur de recherche spécialisé dans les publications concernant *C. elegans*.

L'ensemble de ces données facilement accessibles a sans doute favorisé l'émergence d'un projet novateur. Le projet *OpenWorm* (www.openworm.org), démarré en 2011, vise à créer le premier organisme virtuel en utilisant pour cela les informations disponibles.

Les données dématérialisées telles que les séquences et les protocoles ne sont pas les seules ressources dont disposent les chercheurs. Egalement accessibles *via* des interfaces web, plusieurs centres procurent du matériel physique pour les chercheurs.

La première ressource concerne les souches de nématodes. Le *Caenorhabditis Genetics Center* (CGC) stocke et distribue les lignées et assure la coordination de la nomenclature des gènes et des lignées. D'autres banques plus spécifiques existent également pour la distribution de souches : NemaGENETAG (Vallin et al. 2012) distribue des lignées contenant des insertions pour le transposon *Mos-1* ; les consortia *C. elegans* Gene Knockout Consortium, National Bioresource Project of Japan et *C. elegans* Knockout Consortium ont généré plus 6700 lignées mutantes (The *C. elegans* Deletion Mutant Consortium 2012) ; et le BC *C. elegans* Gene Expression Consortium génère et distribue des lignées portant des transgènes du type promoteur::GFP pour les études de profil d'expression.

Outre les lignées, le matériel nécessaire à leur obtention peut également être fourni suite à l'aboutissement de projets de grande envergure. L'ORFeome est une banque de vecteur contenant les cadres ouverts de lecture (*open reading frame*, ORF) couvrant 65% des gènes. Le Promoterome est une banque de vecteur contenant les promoteurs de 6 000 gènes de *C. elegans*. Le *Fire Lab Vector Kit* propose des vecteurs d'expression spécifiques à *C. elegans* pour l'expression de rapporteurs, le test de séquences d'ARN interférents ou le marquage des introns et des exons.

2. Généralités

A. Milieu naturel

Les souches de *C. elegans* actuellement utilisées en laboratoire comme « sauvage » ont été isolées du sol, et/ou à proximité de fruits en décomposition. De ce fait, il a souvent été décrit que *C. elegans* vit dans le sol et se nourrit de bactéries ou autres micro-organismes et ce, dans toutes les régions du monde. Néanmoins, le fait que d'autres nématodes soient connus comme espèces parasitiques des gastéropodes (Morand, Wilson, and Glen 2004) ainsi que des études écologiques, ont établi un lien entre *C. elegans* et les escargots, particulièrement *Helix asperga* (Caswell-Chen et al. 2005).

Lors des isolations à partir de l'environnement extérieur, il est commun d'obtenir des animaux de stade *dauer* qui est un stade de latence au cours duquel l'animal arrête son développement et ne se nourrit pas, son pharynx est fermé (Wolkow and Hall ; voir Cycle de vie). En poursuivant son développement une fois mis en présence de nourriture, la larve de stade *dauer* reprend son développement vers un adulte fertile. Adulte, *C. elegans* est un nématode qui mesure adulte environ 1 mm de long pour environ 60 µm de diamètre.

L'espèce compte deux sexes, mâle (0,1% de la population) et hermaphrodite. Ces deux sexes sont facilement identifiables, compte tenu des différences anatomiques, visibles à la loupe binoculaire.

B. Anatomie

Le développement de *C. elegans* est remarquable puisqu'il conduit invariablement au même nombre de cellules entre les individus de même sexe. Les hermaphrodites sont formés de 959 cellules somatiques, parmi lesquelles on dénombre 302 neurones. Les mâles, dont l'apparence varie notamment à l'extrémité de la queue, sont formés de 1031 cellules.

L'organisation interne de *C. elegans* est assez simple (Figure 4). Le corps se compose de deux tubes concentriques séparés par le pseudocœlome. A l'intérieur, se trouvent l'intestin, les gonades et l'utérus. A l'extérieur, l'animal est recouvert par une cuticule, recouvrant l'hypoderme, les muscles et des cellules nerveuses.

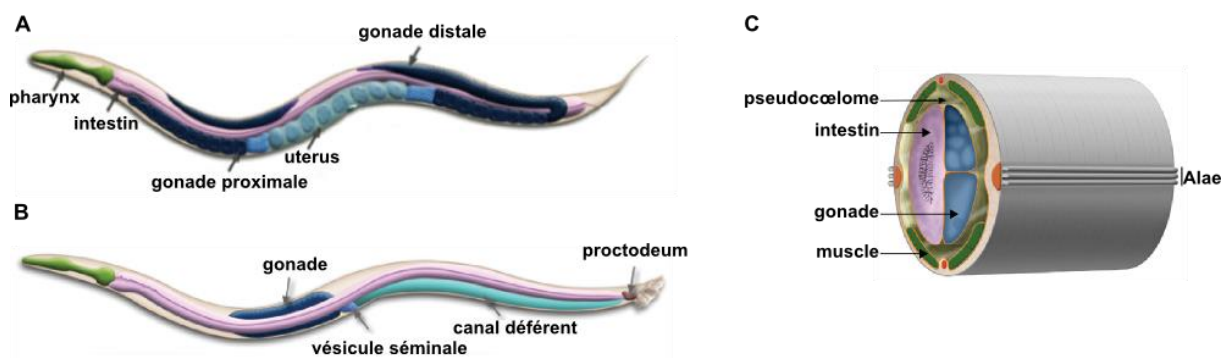


Figure 4 Anatomie du nématode *Caenorhabditis elegans* (adapté du WormAtlas (Lints and Hall 2009)). A Vue schématique d'un hermaphrodite adulte (latéral gauche). **B** Vue schématique d'un mâle adulte (latéral gauche). **C** Schéma d'une coupe transversale côté postérieur d'un individu adulte.

L'information génétique est supportée par 5 paires de chromosomes autosomes et une paire de chromosomes sexuels. Le sexe de l'animal est déterminé génétiquement par le nombre de chromosomes X. Les mâles n'ont qu'un seul chromosome X (XO), les hermaphrodites en possèdent une paire (XX).

C. Cycle de vie

Un hermaphrodite peut pondre jusqu'à 300 œufs autofécondés au cours de sa vie reproductive. Le développement débute par l'embryogenèse qui se déroule dans un œuf *in utero* les premières heures puis *ex utero* jusqu'à l'éclosion. L'éclosion est suivie de 4 stades larvaires L1, L2, L3 puis L4 marqués par une mue entre chacun d'entre eux. La dernière mue marque la transition au stade adulte où l'individu commence à pondre des œufs (Figure 5). La rapidité de ce cycle est dépendante de la température environnante. Les animaux se développent plus lentement à basse température (12-15°C) qu'à haute température (22-25°C), mais cela n'a aucun impact sur leur morphologie finale. Par exemple, à 22°C il faut un peu plus de 2 jours (environ 58h) pour que les œufs atteignent le stade adulte.

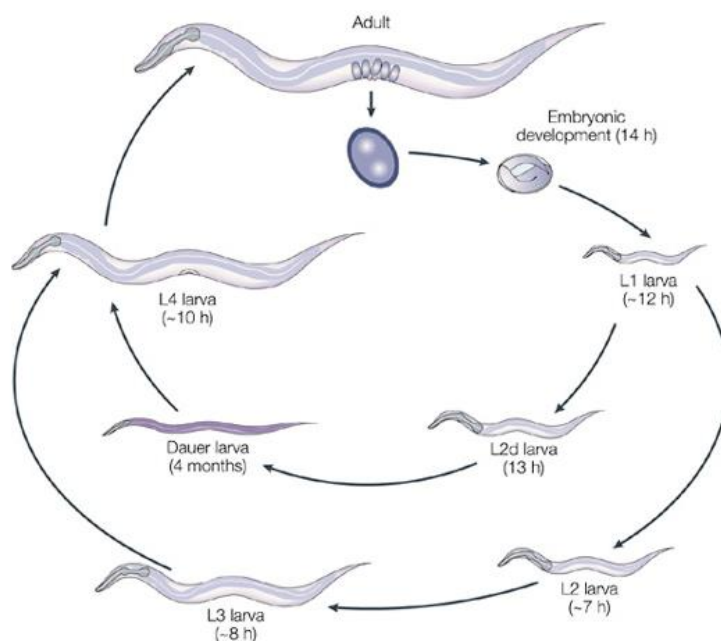


Figure 5 Cycle de vie de *C. elegans*. La boucle extérieure correspond au cycle de vie dans des conditions optimales. La boucle intérieure illustre le développement en cas de conditions de privation de nourriture. Pour chaque stade figure la durée du stade de développement. Source (**Jorgensen and Mango 2002**)

Avec un rythme de croissance soutenu, l'espèce épuise vite les ressources du milieu sur lequel elle croît. C'est pourquoi plusieurs mécanismes permettent de mettre le développement en pause quand le milieu s'appauvrit. Si la larve de stade L1 ne dispose pas des ressources nécessaires, le développement s'arrête et la larve peut entrer dans le stade *Dauer* au lieu du stade L2. Le stade *Dauer* (qui signifie durée en allemand) est un stade larvaire au cours duquel les animaux sont en diapause et peuvent perdurer pendant plusieurs mois sans nourriture. Dès que des conditions favorables surviennent, les animaux quittent ce stade, et retournent à un développement au stade larvaire L4.

3. La génétique de *C. elegans*

1. Organisation du génome

A. Les chromosomes

Le génome de *C. elegans* est composé de 100 mégabases réparties sur six paires de chromosomes, dont cinq paires d'autosomes. Ces chromosomes supportent l'expression d'environ 20 000 gènes.

Les chromosomes de *C. elegans* sont holocentriques (ils ne possèdent pas de centromères), au cours des divisions cellulaires les kinétochores sont dispersés le long de l'ensemble du chromosome. Les chromosomes sont ainsi attachés sur toute leur longueur aux microtubules. Ce mécanisme est également présent chez d'autres organismes, notamment chez certaines plantes. Il est néanmoins peu représenté chez les organismes modèles généralement utilisés en laboratoire (Melters et al. 2012).

L'absence de centromère sur les chromosomes de *C. elegans* permet que n'importe quelle séquence d'ADN serve de centromère. De l'ADN introduit dans l'organisme sera ainsi répliqué et transmis, ce qui permet de générer des lignées transgéniques non intégrées. L'ADN introduit est présent sous la forme d'un concatémère extrachromosomique (Mello et al. 1991). A ce jour, le phénomène permettant la formation des concatémères chez *C. elegans* n'est pas connu.

B. Les opérons

La particularité des nématodes par rapport à bon nombre d'autres espèces animales est de présenter des gènes exprimés sous forme d'opérons.

Le génome de *C. elegans* compte au moins 1 000 opérons, dont la longueur varie de 2 à 8 gènes. Ceci représente donc environ 15% du génome de *C. elegans* (Blumenthal et al. 2002). Historiquement, c'est l'opéron *gdp-2/gdp-3* qui a été le premier à être identifié et étudié (Huang and Hirsh 1989). L'identification a été permise par une différence dans l'extrémité 5' de l'ARNm. La région génomique comporte en fait 3 gènes : *mai-1*, *gpd-2* et *gpd-3* qui forme un opéron (Spieth et al. 1993).

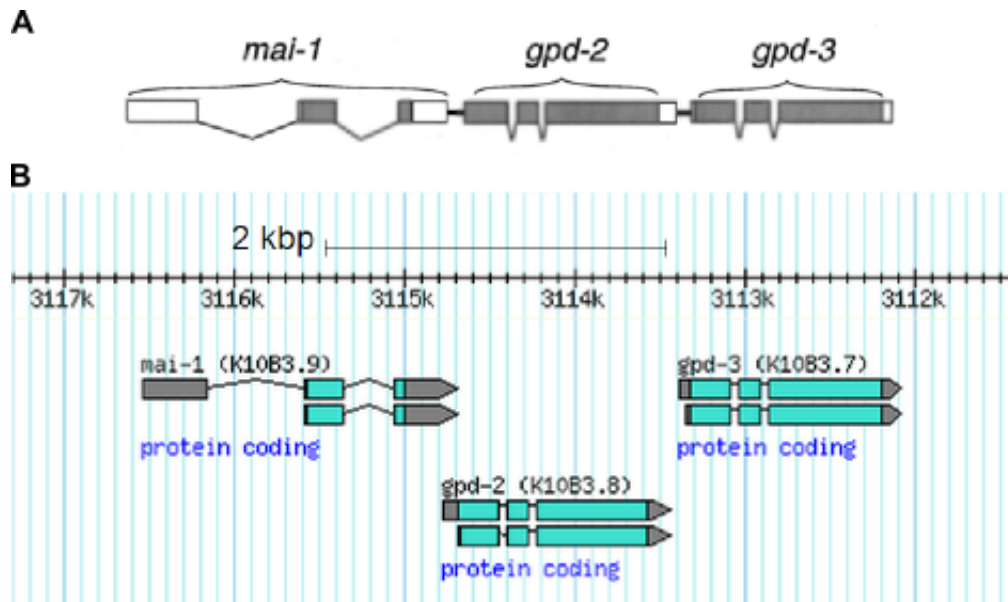


Figure 6 Organisation de l'opéron polycistronique *mai-1/gpd-2/gpd-3* de *C. elegans*. A Représentation adaptée de la publication de 1993 (Spieth et al. 1993). B Région génétique de l'opéron issue du site Wormbase. En gris les séquences non transcrites apparaissent.

L'opéron *mai-1/gpd-2/gpd-3* donne lieu à un transcrit unique qui donne plusieurs protéines lors de la traduction. Ce mécanisme diffère néanmoins de celui des bactéries qui permet également l'expression simultanée de plusieurs gènes. Chez *C. elegans* l'ARN issu d'un opéron est maturé en plusieurs ARNm indépendants pour chacun des gènes. Cela permet de maintenir une expression différentielle des gènes placés sous le contrôle d'un même promoteur, grâce à la stabilité variable des ARNm de chaque gène (Paulsen, Capowski, and Strome 1995; Spieth et al. 1993; Zorio et al. 1994).

Les régions 3'UTR des ARNm des gènes subissent aussi l'épissage alternatif, et un gène peut donc avoir plusieurs séquences 3'UTR associées. La différence dans les régions 3'UTR a un impact direct sur la stabilité de l'ARNm. Dans l'étude de M. Mangone (2010), il a été mis en évidence une corrélation entre la place du gène dans l'opéron et la diminution de la taille des régions 3'UTR, du nombre d'isoformes, ainsi que de la fréquence de 3'UTR sans signal de polyadénylation (PAS) (Figure 7).

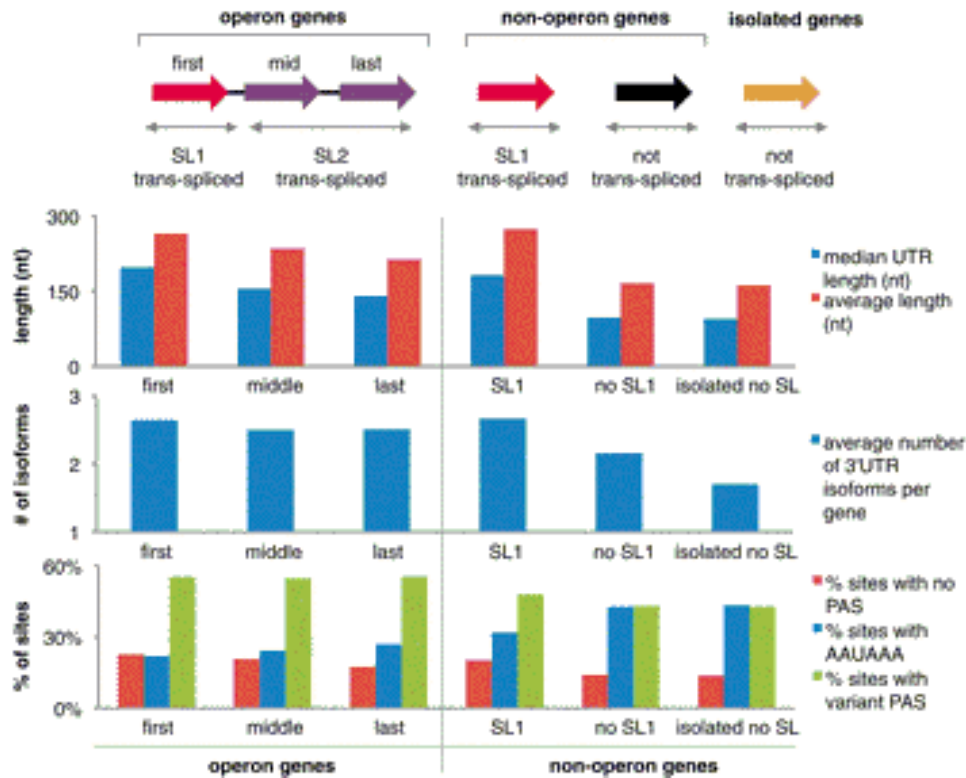


Figure 7 Comparaison des 3'UTR des opérons entre les ARNm trans-épissés ou non (M. Mangone et al. 2010). En haut : représentation schématique des gènes en opéron (gauche, n = 574), hors-opéron (milieu, n = 4348), et isolés (droite, n = 2098). Les premiers gènes de l'opéron (rouge) sont trans-épissés SL1 ; les gènes suivants (violet) sont trans-épissés avec d'autres SLs (SL2 à SL12). Les gènes hors opérons sont trans-épissés SL1 (rouge, n = 3530) ou non trans-épissés (noir, n = 818). Les gènes isolés, sans autres séquences codantes dans les 2kb avoisinants ne sont pas trans-épissés (orange, n = 2098). En bas : tailles, nombre d'isoformes et sites PAS des 3'UTRs des gènes en opéron et hors opéron.

Le trans-épissage, qui consiste à substituer la région 5'UTR du pré-ARNm par une séquence courte d'environ 22 nucléotides appelée *splice-leader* (SL), par opposition au cis-épissage concernant les exons, est un mécanisme d'épissage qui requiert également la machinerie du spliceosome. Environ 70% des gènes de *C. elegans* sont concernés par ce mécanisme qui est donc loin d'être anecdotique. La séquence SL est fournie par un ARN d'environ 100 nucléotides, le SL ARN. Le trans-épissage est réalisé par la machinerie du spliceosome, de manière similaire à l'excision des lassos, sauf que c'est le complexe U2 qui initie la réaction. U1 n'est pas nécessaire à la réaction de trans-épissage. La moitié des gènes de *C. elegans* sont trans-épissés avec le *splice-leader* SL1 qui se branche à l'extrémité 3' de la région 5', juste avant le codon d'initiation de la traduction. Les autres gènes subissant un trans-épissage ne faisant pas appel à SL1, sont situés dans des pré-ARNm polycistroniques générés par les opérons.

Le phénomène de trans-épissage génère des pré-ARNm dont l'épissage alternatif n'a pas encore eu lieu, y compris à partir d'un pré-ARNm polycistronique. Chez *C. elegans*, la proportion des gènes subissant un épissage alternatif est estimée à 25%.

2. Transgénèse

La transgénèse consiste à introduire un ou plusieurs gènes sans modifier le fond génétique de l'organisme. Il existe plusieurs méthodes pour faire pénétrer de l'ADN exogène dans un organisme. Pour *C. elegans*, il en existe deux utilisées communément. Chacune présente des avantages et des spécificités, et les deux requièrent un équipement dédié.

A. La micro-injection

La micro-injection a été la première méthode mise au point. A l'aide d'un microscope et d'une aiguille montée sur un micromanipulateur il est possible de réaliser une injection de l'ADN directement dans l'animal adulte.

L'obtention d'animaux transgénique par cette procédure est liée à la physiologie de *C. elegans*. Sa lignée germinale est formée d'un syncytium située dans les gonades distales de l'animal (Figure 8A et B). Le syncytium contient les noyaux des cellules germinales qui partagent le même cytoplasme (Lints and Hall 2009). A l'extrémité de la gonade distale se trouvent les cellules germinales qui permettent la production des noyaux des ovocytes en devenir. Les noyaux sont maintenus par la lame basale gonadale et sont connectés entre eux par le rachis (Figure 8B), un canal central. Les noyaux progressent le long du bras de la gonade et sont progressivement encapsulés par une membrane cellulaire. L'encapsulation produit des ovocytes encore immatures. Leur maturation va s'effectuer en même temps que leur progression dans la gonade. A la fin de la gonade, ils traverseront la spermathèque où ils sont fécondés.

Au cours de la micro-injection, l'ADN exogène est injecté dans les gonades distales. Les propriétés de la maturation des ovocytes vont permettre que l'ADN injecté soit progressivement incorporé dans les ovocytes en formation (Figure 8C). L'ovocyte mature contenant l'ADN exogène poursuit son parcours et est fécondé dans la spermathèque. Les embryons résultants sont ainsi transgéniques.

La micro-injection est une méthode efficace de transgénèse chez le nématode. Néanmoins, elle présente des limites techniques. En effet, l'injection étant réalisée manuellement, sa réussite et son efficacité dépendent de la précision du manipulateur. Le nombre d'individus injectés est également dépendant de l'expérimentateur et de sa précision technique.

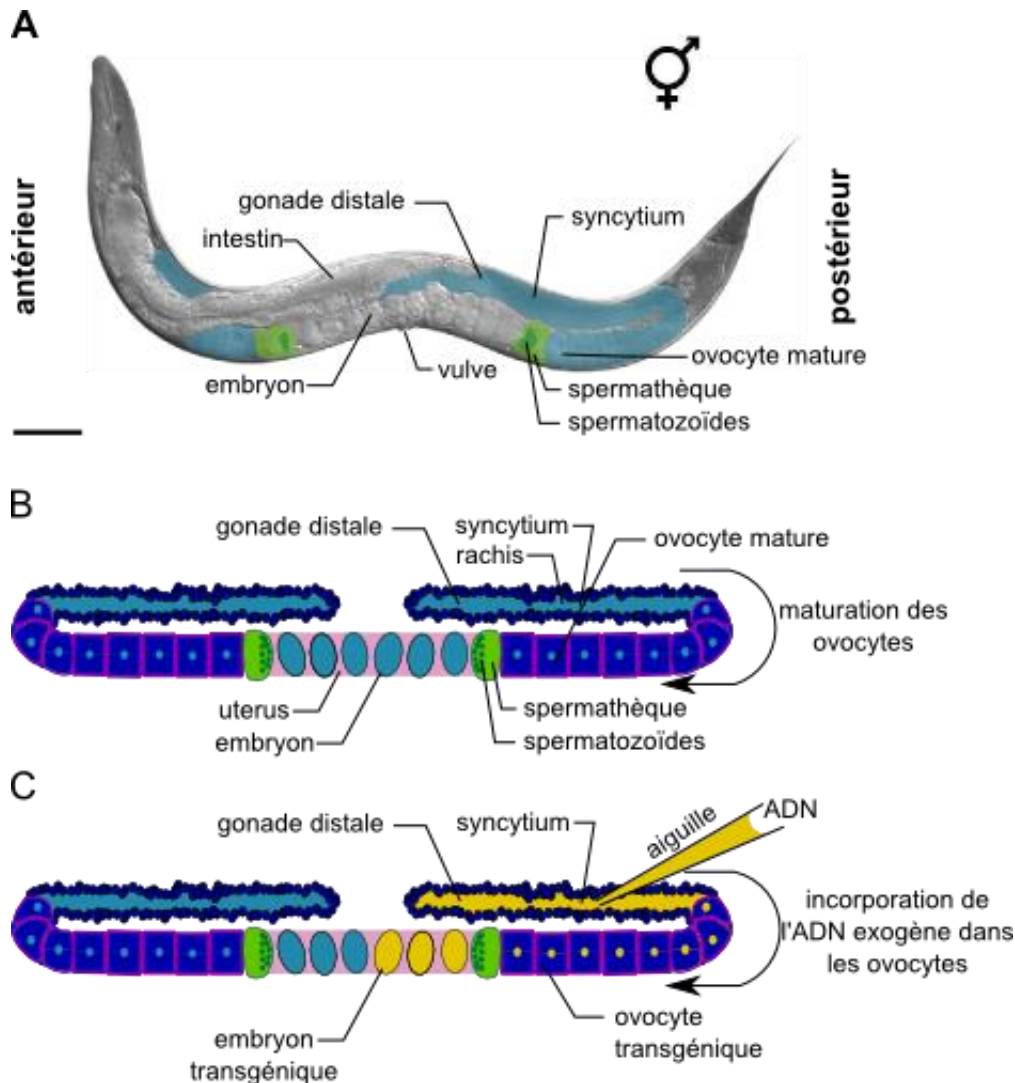


Figure 8 Transgénèse par micro-injection. **A** Image Normarski (barre d'échelle 50μm). Les couleurs signalent l'appareil reproducteur de *C. elegans*. En bleu les gonades et ovocytes, en vert spermathèque et spermatozoïdes. **B** Vue schématique de l'ensemble de l'appareil reproducteur. Les deux gonades sont symétriques par rapport à l'utérus (rose). **C** Schéma de la transgénèse par micro-injection. L'ADN exogène (jaune) est injecté à l'aide d'une micro-aiguille dans le syncytium d'un jeune hermaphrodite. Il est encapsulé dans les ovocytes au cours de la maturation et conduit à la production d'embryons transgéniques, après fécondation dans la spermathèque.

B. Le bombardement

Le bombardement est une méthode plus récente que la micro-injection. Elle permet de réaliser la transgénèse sur un grand nombre d'individus en une seule expérience. Le bombardement fait appel à la biolistique, une technique initialement développée chez les plantes. Elle repose sur l'utilisation d'un canon à ADN. Ainsi, une seule expérience de transgénèse peut être conduite sur une population entière de nématodes (plusieurs dizaine de milliers). La population, composée de jeunes adultes, est littéralement bombardée par de petites billes (1,5 à 3μm de diamètre) en tungstène ou en or recouverte de l'ADN exogène (Wilm et al. 1999).

A ses débuts, cette méthode a été utilisée avec des marqueurs phénotypiques visuels, tel que *rol-6(su1006)* (Wilm et al. 1999) qui confère un défaut de locomotion entraînant la rotation du vers sur lui-même. Très vite, la difficulté de localiser et d'identifier les mutants dans la génération F1 est survenue en raison de la taille de la population rendant difficile la manipulation. Les expériences suivantes se sont donc appuyées sur le sauvetage de mutations afin d'opérer la sélection des animaux transgéniques. Les marqueurs utilisés dans ces cas-là ont été les mutations *pha-1(e2123tc)* entraînant la létalité à une température de 25°C, et ultérieurement, *unc-119(ed3)* entraînant à la fois un défaut locomoteur et une incapacité à entrer dans le stade *Dauer* dans des conditions de stress environnemental (Praitis et al. 2001). L'effet de ces mutations peut être annulé par l'introduction d'une copie sauvage du gène. En utilisant une copie sauvage du gène, les vers transgéniques redeviennent résistants au stress et sont ainsi sélectionnés au cours du temps tout en permettant l'élimination des F1 non transgéniques.

Ces dernières années la sélection des nématodes par antibiotiques a été grandement développée. A ce jour, trois antibiotiques peuvent être utilisés pour la sélection des animaux transgéniques : le G418 (Giordano-Santini et al. 2010), la puromycine (Semple, Garcia-Verdugo, and Lehner 2010; Semple, Biondini, and Lehner 2012) et l'hygromycine B (Radman, Greiss, and Chin 2013), tous compatibles avec la méthode du bombardement.

Le mécanisme aboutissant à la formation d'animaux transgéniques *via* le bombardement n'est pas connu. Il est supposé que les microbilles pénètrent l'animal au niveau des gonades permettant l'intégration de l'ADN dans les ovocytes selon un mécanisme similaire à celui de la micro-injection. La taille de la population bombardée rend impossible le suivi des individus pour établir un lien entre la position des microbilles et la génération d'animaux transgéniques.

C. Devenir de l'ADN exogène

Les deux méthodes de transgénèse présentées ci-dessus permettent la génération d'animaux transgéniques. L'ADN exogène introduit par une méthode ou une autre va connaître les mêmes transformations. L'ADN exogène va être assemblé en un concatémère extrachromosomique rassemblant plusieurs copies des transgènes introduits (Mello et al. 1991; Stinchcomb et al. 1985). Le nombre de copies contenues dans un concatémère extrachromosomique est variable selon la méthode. Chaque concatémère ainsi formé est néanmoins unique et contient un nombre de copies variables de chaque transgène. Les individus F1 générés sont donc porteurs de concatémères différents et indépendants.

Le concatémère extrachromosomique est traité comme un chromosome et est transmis à une partie de la descendance. Le taux de transmission peut varier de 10 à 90% selon les lignées

(Hirsh et al. 1985). Il peut aussi être perdu au cours de divisions cellulaires, donnant lieu à l'observation d'un effet mosaïque (Mello et al. 1991). En effet, si le transgène est perdu au cours d'une division tôt dans le développement, il est possible qu'il ne soit pas exprimé dans l'ensemble de l'organisme. Par conséquent, les lignées transgéniques porteuses d'un concatémère extrachromosomique doivent être entretenues afin de maintenir le transgène dans une majorité d'individus si les marqueurs de transgénèse utilisés ne sont pas sélectifs.

Les marqueurs de transgénèse permettent de détecter les animaux transgéniques au sein d'une population. Depuis 2010, la sélection d'animaux transgénique a été simplifiée par le développement de la sélection avec des antibiotiques. Actuellement sont disponibles la sélection sur G418 (Giordano-Santini et al. 2010), puromycine (Semple, Garcia-Verdugo, and Lehner 2010) et hygromycine (Greiss and Chin 2011; Radman, Greiss, and Chin 2013). La sélection avec les antibiotiques a permis de contourner deux difficultés de sélection des animaux transgéniques : le temps d'entretien des lignées pour un marqueur phénotypique visuel non sélectif et le fond génétique mutants des animaux dont la sélection s'opère par sauvetage. L'ensemble des marqueurs de sélection disponibles et employés chez *C. elegans* est présenté Tableau 1, avec les avantages de chacun.

	Compatible avec tout génotype	Sélection non-manuelle
Sauvetage de mutants		
<i>dpy-20</i>	×	×
<i>sup-7(st5)</i>	×	×
<i>pha-1</i>	×	✓*
<i>unc-119</i>	×	✓*
Phénotype dominants		
<i>unc-22 antisense</i>	✓	×
<i>rol-6</i>	✓	×
Marqueurs fluorescents	✓	×
Sélection antibiotique		
NeoR	✓	✓
PuroR	✓	✓
HygR	✓	✓

Figure 9 Marqueurs disponibles pour la transgénèse de *C. elegans* (adapté de (Cornes et al. 2014)). Les propriétés principales des marqueurs courant sont résumées. (*) le phénotype dépend des conditions environnementales. **Sauvetage de mutants** : *dpy-20* sauve *dpy-20(e1282ts)* entraînant un phénotype dumpy (Clark, Lu, and Horvitz 1994), *sup-7(st5)* sauve la stérilité dépendante de la température des mutants *tra-3(e1107)* (Andrew Fire 1986), *unc-119* sauve *unc-119(ed3)* les défauts de locomotions et d'entrée en stade *dauer* (Maduro and Pilgrim 1995), *pha-1* empêche la létalité embryonnaire à 25°C de *pha-1(e2123)* at 25°C (Granato, Schnabel, and Schnabel 1994). (suite de la légende p38)

Phénotypes dominants : *rol-6* provoque un phénotype *roller* (Kramer et al. 1990), *unc-22* confère un phénotype *twitcher* (A. Fire et al. 1991). **Rapporteurs fluorescents** : protéines fluorescentes comme la GFP (Green Fluorescent Protein), mCherry ou dsRed (red fluorescent protein) exprimées en fusion avec une autre protéine ou sous le contrôle d'un promoteur. **Sélection antibiotique** : une cassette de résistance confère un avantage sélectif aux animaux transgéniques en présence de l'antibiotique : NeoR : G418 ; PuroR : puromycine ; HygR: hygromycine B.

L'autre moyen de contourner la difficulté du maintien d'une lignée porteuse d'un concatémère extrachromosomique est de générer des animaux porteurs de transgènes intégrés à leur ADN génomique. Dans le cas du bombardement, le transgène est intégré spontanément dans l'ADN à des fréquences allant de 1/3 à 1/8 des lignées obtenues (Praitis et al. 2001; Reece-Hoyes et al. 2007). Cette intégration reste néanmoins aléatoire et le nombre de copies n'est pas non plus contrôlé. La micro-injection conduit plus rarement à l'obtention de lignées ayant spontanément intégré le transgène. Les injections directement dans les ovocytes tendent à favoriser le phénomène, mais rendent le travail d'injection d'autant plus laborieux (Andrew Fire 1986). La technique faisant appel au transposon *Mos-1* issu de la drosophile permet de générer des lignées avec intégration du transgène. Cette technique permet de contrôler le nombre de copies intégrées dans le génome, ainsi que le *locus* (Robert and Bessereau 2007).

Il existe aussi des protocoles permettant de favoriser l'intégration des transgènes à partir des lignées porteuses de concatémère extrachromosomique. Ceux-ci s'appuient sur des méthodes de cassure de l'ADN qui favoriseront la recombinaison et donc l'insertion du transgène d'intérêt (Evans 2006).

Récemment une nouvelle méthode d'édition du génome a été mise au point, le système CRISPR/Cas9 (Tzur et al. 2013; Friedland et al. 2013; S. Waaijers et al. 2013). Le système se base sur un mécanisme de défense bactérien, le *Clustered Regularly Interspaced Short Palindromic Repeats* (CRISPR) et les protéines associées (Cas) qui détruisent les ADN et/ou ARN exogènes. Un brin d'ARN peut être utilisé comme guide de l'endonucléase Cas9 pour générer une cassure double brin à un lieu précis. La cassure générée est réparée en utilisant l'ADN homologue co-injecté, ce qui permet de générer des délétions ou des insertions très précises. Cette méthode semble pouvoir être couplée à la sélection par antibiotique (Friedland et al. 2013). Plusieurs protocoles ont abouti à l'établissement de lignées intégrée grâce au système CRISPR/Cas9 (Selma Waaijers and Boxem 2014), les premières indications que la méthode serait efficace par *feeding*, qui consiste à nourrir les animaux avec des bactéries exprimant les ARN nécessaires ont récemment été publiées (Liu et al. 2014).

L'obtention de lignées transgéniques chez *C. elegans* est un procédé simple, et les outils disponibles permettent aujourd'hui de manipuler aisément et efficacement des lignées transgéniques.

3. Les microARN chez *C. elegans*

A. Découverte et connaissances

Comme nous l'avons vu plus haut (Historique et enjeux présents), les premiers miARN ont été découverts chez *C. elegans* au début des années 1990. La première mise en évidence est la régulation par le miARN *lin-4* des ARNm des gènes *lin-14* puis *lin-28* (R. C. Lee, Feinbaum, and Ambros 1993; Moss, Lee, and Ambros 1997; Bruce Wightman, Ha, and Ruvkun 1993). Ensuite, les rôles du miARN *let-7* ont été mis en évidence dans la régulation du développement des cellules de l'hypoderme en réprimant les ARNm des gènes *lin-41*, *daf-12*, *hbl-1* et *pha-4* (Abrahante et al. 2003; Großhans et al. 2005; Slack et al. 2000). Ces deux miARN sont impliqués dans le même réseau de régulation responsable du bon déroulement temporel du développement (Figure 10). Le rôle du miARN *lisy-6* dans la définition de l'asymétrie des neurones ASEL et ASER a aussi été identifié parmi les premiers (Johnston and Hobert 2003).

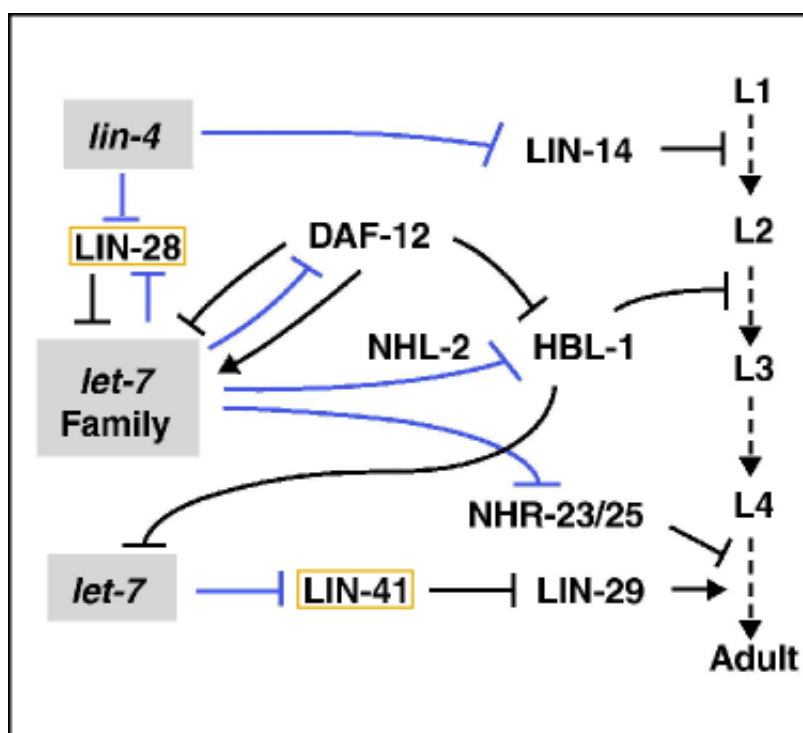


Figure 10 Réseau de régulation des gènes et miARN au cours du développement de *C. elegans* (Ambros 2011). Les miARN (rectangles gris) *lin-4* et de la famille de *let-7* régulent les étapes des différents stades larvaires, décrits à droite de la figure. Les miARN régulent de manière post-transcriptionnelle les ARNm cibles par des interactions directes (lignes bleues) avec les séquences 3'UTR. Les ARNm des protéines LIN-28 et LIN-41 (cadre jaune) sont des cibles de miARN.

Rapidement, des expériences de séquençages des ARN courts ont été entreprises afin d'identifier d'autres miARN. La purification des petits ARN était suivie de la génération d'une banque d'ADNc en vue d'un séquençage. Les séquences des ADNc étaient ensuite comparées à la séquence des gènes ou des EST (*expressed sequence tags*) de *C. elegans*. La localisation au sein du génome permet de distinguer les miARN des autres petits ARN interférant en vérifiant la capacité de la séquence à former une structure tige-boucle. Cette méthode a permis une première identification d'une centaine de miARN (Ambros et al. 2003) dont certains conservés entre plusieurs espèces d'insectes et vertébrés. En 2006 le nombre de miARN identifiés par séquençage est porté à 115 répartis dans 63 familles (Ruby et al. 2006). Les familles de miARN sont fondées sur la base de la conservation de la séquence en 5', appelée séquence graine. La séquence graine est la partie du miARN qui permet l'hybridation avec l'ARNm. La conservation de la séquence graine des miARN semblent être liée à la redondance du système qui permet de supprimer la plupart des miARN sans obtenir de phénotypes visibles (Miska et al. 2007). Actuellement, plus de 200 miARN ont été identifiés et annotés chez *C. elegans* sur la base de données d'expression, de séquençage et par inférence informatique (Wormbase http://www.wormbase.org/resources/gene_class/mir#01--10).

L'informatique a également permis d'évaluer la présence des miARN au sein de boucles de régulation. Par exemple, le miARN *let-7* est au centre d'un réseau de régulation formé par une boucle de régulation négative avec LIN-28. D'autres miARN ont la capacité d'interagir directement avec des facteurs de transcription. Une boucle de régulation négative est définie lorsque le facteur de transcription qui régule l'expression du miARN est lui-même régulé par ce miARN. Sur la base de 940 facteurs de transcription prédits et 115 miARN d'identifié, vingt-trois boucles de régulations ont pu être détectées *in vitro* chez *C. elegans* (Martinez et al. 2008). L'identification des boucles a été permise par l'analyse des données issues d'un crible *Yeast-1-Hybrid* permettant de détecter les interactions entre facteurs de transcription et régions en amont des miARN et la prédiction des régulations des facteurs de transcriptions par les miARN. Plusieurs de ces boucles ont été validées *in vivo*. Les régulations négatives que les miARN assurent sur l'expression des facteurs de transcription indiquent que ceux-ci sont au cœur des réseaux de régulation de l'expression génétique.

A ce jour, d'autres régulations similaires exercées par les miARN ont pu être déterminées. Le miARN *lsy-6* régule l'asymétrie gauche droite des neurones ASE en fixant l'ARNm du facteur de transcription *cog-1* (Cochella and Hobert 2012; Johnston and Hobert 2003). La famille de *mir-51* est impliquée dans plusieurs voies de régulation (J. L. Brenner, Kemp, and Abbott 2012). Certains miARN de la famille de *mir-51* interfèrent avec d'autres voies de régulation

déjà connues et modifient les phénotypes mutants comme *mir-52* dont la mutation contre les effets de celle de *mir-240/786*, nécessaire pour le rythme de défécation (Miska et al. 2007).

Le *mir-52* régule joue également un rôle dans la régulation exercée par *mir-1* et supprime les effets de sa mutation (J. L. Brenner, Kemp, and Abbott 2012). Le *mir-1* régule l'activité de UNC-29 et UNC-63 présentes dans les jonctions neuromusculaires, plus particulièrement dans les récepteurs acétylcholine. La mutation de *mir-1* augmente la résistance des animaux au levamisol (Simon et al. 2008).

B. Méthodes d'étude

La cartographie des régulations opérées par les miARN, s'appuie sur des approches complémentaires : la prédiction de ces régulations grâce à l'informatique et la validation expérimentale (Shiu, Zhuang, and Hunter 2014; Grosswendt et al. 2014).

En 2006, des logiciels de prédictions des cibles des miARN, tel que Pictar (Krek et al. 2005) sont disponibles et testés. L'équipe de Fabio Piano a utilisé ce logiciel pour identifier les séquences conservées des régions 3'UTR et les comparer entre trois espèces de nématodes (*C. elegans*, *C. briggsae*, et *C. remanei*) et ainsi prédire que 10% des gènes de *C. elegans* seraient cibles des miARN. Afin de valider ces prédictions expérimentalement, les séquences des régions 5' et 3' d'une partie des cibles prédites ont été clonées respectivement en amont et en aval de la GFP (Lall et al. 2006). Un contrôle était réalisé par le clonage de la séquence permissive *unc-54* 3'UTR à la place de la région 3'UTR cible. Dans cette étude, la région *unc-54* 3'UTR a été choisie car son expression est possible dans l'hypoderme et le muscle à tous les stades de développement, et Pictar n'y prédit aucune régulation liée à *let-7* ou à d'autres miARN. La méthode a été testée sur des cibles potentielles et avérées du miARN *let-7*.

Des lignées indépendantes porteuses des constructions contenant la GFP en fusion avec la région promotrice du gène et les séquences 3'UTR du gène ou de *unc-54* 3'UTR ont été obtenues et analysées. La fluorescence a été examinée au microscope afin de déterminer dans quels tissus la GFP était présente. Ces résultats ont été récapitulés sous forme d'un tableau où un code couleur exprime le pourcentage d'individus exprimant la GFP dans le tissu considéré (Figure 11). Les régulations spatio-temporelles sont mises en évidence par un changement d'expression de la fluorescence entre les lignées générées (porteuses de la région 3'UTR endogène ou contrôle).

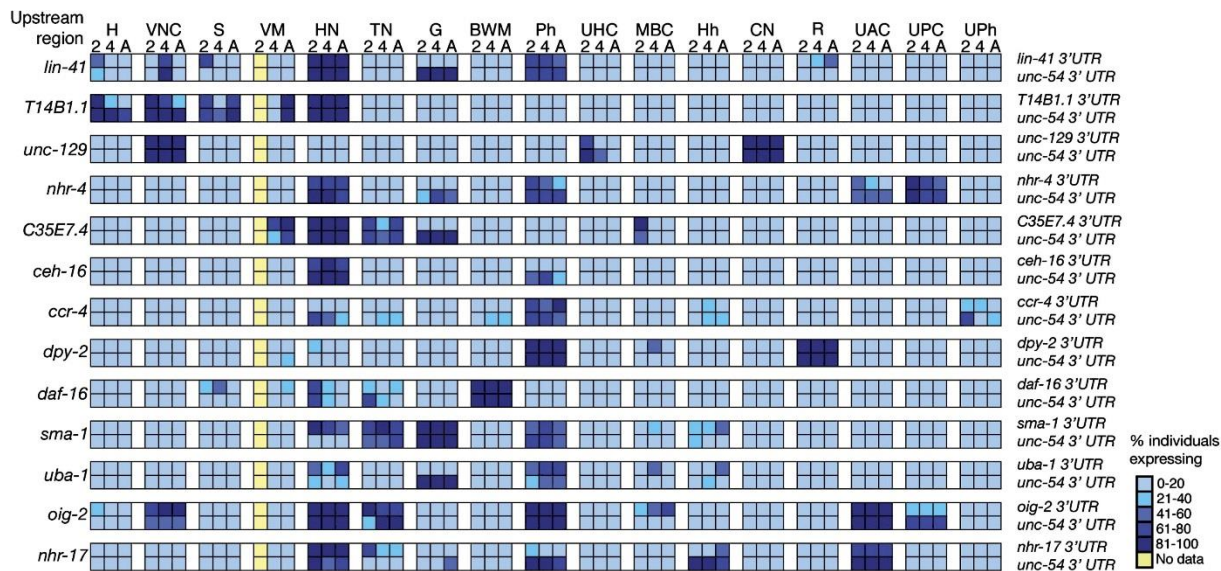


Figure 11 Analyse de l'expression des rapporteurs des cibles prédites de *let-7* publié par (Lall et al. 2006). Le tableau dénombre le pourcentage d'individus exprimant la GFP dans les différents tissus, et à différents stades, L2/L3 (« 2 »), L4 (« 4 ») et Adulte (« A »). H, hypoderme ; VNC, corde nerveuse centrale ; S, cellules hypodermiques latérales ; VM, muscle de la vulve ; HN, neurones antérieurs ; TN, neurones postérieurs ; G, intestins ; BWM, muscles du corps ; Ph, pharynx ; UHC, paire de cellules musculaires de la tête ; MBC, corps cellulaires du corps ; Hh, cellules musculaires ou hypodermiques de la tête ; CN, neurones de la corde ventrale ; R, cellules adjacentes à l'intestin postérieur ou au rectum ; UAC, paire de neurones de la tête ; UPC, cellules postérieures ; UPh, neurones du pharynx.

Parmi les gènes présents dans le tableau ci-dessus, seul *lin-41* est une cible connue et documentée du miARN *let-7*. Les autres sont des cibles prédites par l'algorithme Pictar, avant (*C35E7.4*, *ceh-16*, *uba-1* et *oig-2*) ou après (reste des gènes), l'augmentation de la stringence de l'algorithme par l'ajout des séquences de *C. remanei*. Les expressions de la GFP observées avec les régions 3'UTR des gènes concernées étaient en accord avec les données antérieures si disponibles (immunomarquage, western blot...) sur l'expression du gène concerné. Avec la région 3'UTR provenant du gène *unc-54*, l'expression de la GFP a été significativement augmentée pour 7 des 12 gènes inclus dans l'étude, en accord avec une régulation médiée par la région 3'UTR. Les auteurs émettent aussi l'hypothèse d'autres régulations existantes après avoir observé une fluorescence lorsque la GFP est suivie de la région 3'UTR endogène non reproductible avec *unc-54 3'UTR*.

Les données ont été recueillies en comparant des lignées exprimant deux transgènes différents indépendamment. L'analyse de la fluorescence a été menée sur l'ensemble de l'organisme sans identification précise des cellules marquées. Par exemple, la fluorescence a été observée dans des neurones antérieurs mais ces neurones n'ont pas été identifiés sur chaque individu. Il est ainsi possible que des différences d'expression minimales soient passées inaperçues, par exemple l'expression dans un nombre de neurones différents ou bien dans des paires proches mais

distinctes. Il serait donc extrêmement intéressant de pouvoir exprimer les deux constructions dans un même individu, l'une portant le 3'UTR endogène, l'autre un 3'UTR contrôle.

4. Génération d'un rapporteur de régulation post-transcriptionnelle

1. Fonctionnement théorique

L'idée de construire un transgène sous forme d'opéron a donc été mise en œuvre en 2008. Afin de déterminer la contribution des régions 3'UTR de différents gènes dans la régulation de leur expression, l'équipe de G. Seydoux a utilisé un système de transgène sous forme d'opéron en utilisant la région entre *gpd-2* et *gpd-3* premier opéron étudié dans l'organisme (Merritt et al. 2008). Deux protéines fluorescentes, la GFP et la mCherry, sont exprimées sous le contrôle d'un seul promoteur cloné en amont, et suivie respectivement par la région 3'UTR à étudier et par une région 3'UTR contrôle.

S'appuyant sur les données publiées en 2006 et 2008, un système de rapporteurs de régulation post-transcriptionnelle (RRPT) a été développé au laboratoire. Le but était d'obtenir une construction adaptée à la validation expérimentale à grande échelle des cibles de miARN prédites par les algorithmes tels que PicTar ou TargetScan. En générant des lignées bicolores, il sera possible d'observer *in vivo* l'expression spatio-temporelle du gène d'intérêt ainsi qu'une éventuelle régulation post transcriptionnelle. Ce projet a été initié au sein du laboratoire par Ilyass Zniber qui a réalisé au cours de son travail de thèse l'assemblage de la cassette telle que présentée ci-après.

Le système repose donc sur une cassette dont les éléments centraux restent constants et autour de laquelle il suffit de cloner les séquences du promoteur et du 3'UTR du gène dont la régulation est à étudier (Figure 12). Le promoteur du gène entraîne l'expression de deux protéines fluorescentes, la GFP et la mCherry. La mCherry est suivie de la région *unc-54* 3'UTR comme contrôle tandis que la GFP est suivie par la région 3'UTR du gène d'intérêt. Un pré-ARNm polycistronique est ainsi généré avant d'être maturé en deux ARNm indépendants. Si des différences d'expression apparaissent entre la GFP et la mCherry elles seront liées à une régulation de la région 3'UTR du gène.

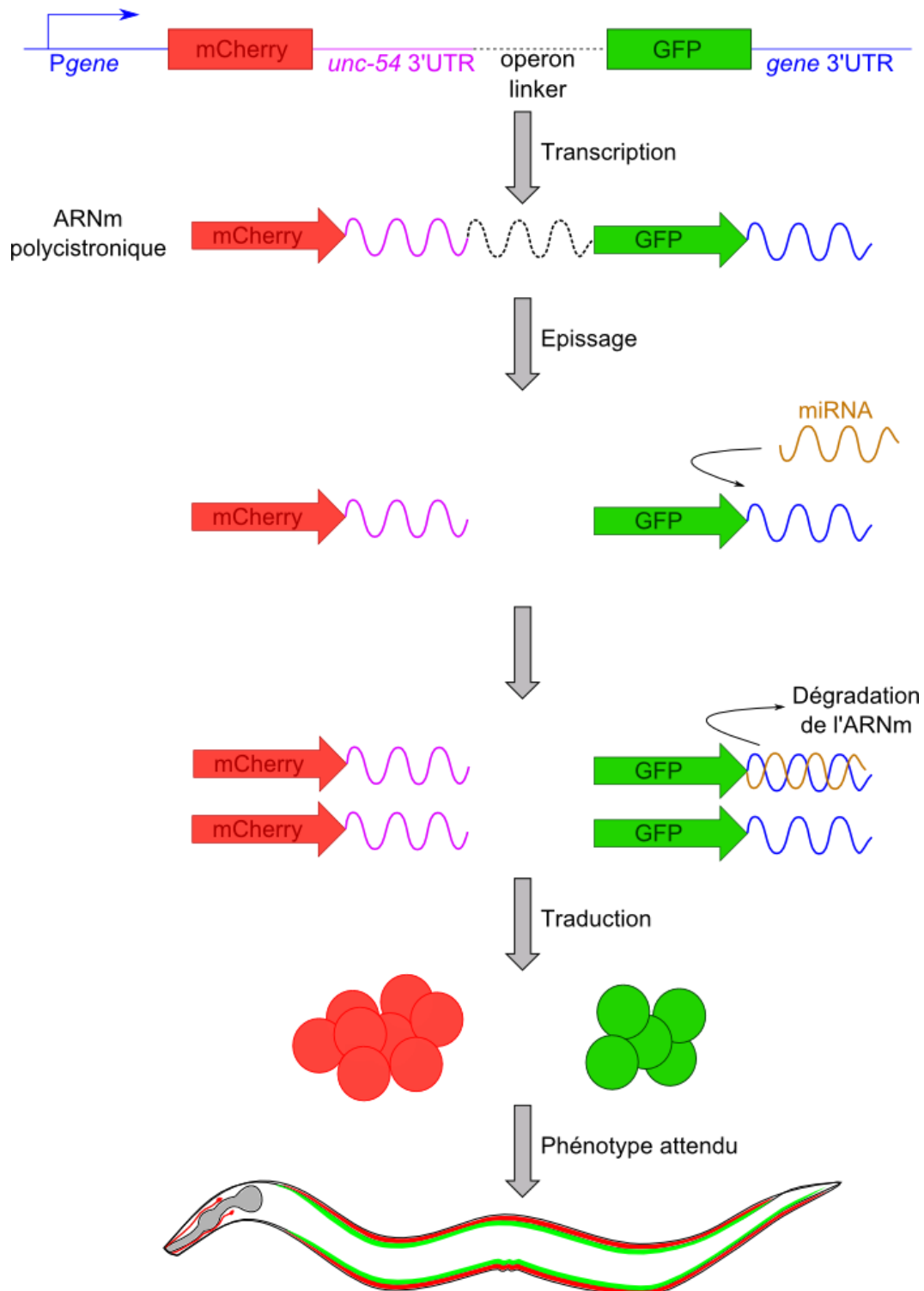


Figure 12 Fonctionnement de la cassette de Rapporteur de Régulation Post-Transcriptionnelle (RRPT). L'expression de la cassette engendre un ARNm polycistronique. Après trans-épissage, l'ARNm de la GFP peut être soumis à une régulation post-transcriptionnelle diminuant ainsi la quantité de protéine fluorescente générée, dans le temps ou dans des tissus spécifiques. Le phénotype attendu est donc par exemple l'expression de la mCherry dans le muscle de l'animal et ses neurones, alors que l'expression de la GFP est cantonnée au muscle et éteinte dans les neurones.

Le développement d'un rapporteur de régulation post-transcriptionnelle bicolore présente de nombreux avantages pour valider expérimentalement les régulations opérées par miARN. La cassette fournit un contrôle interne ce qui évite la génération de lignées contrôle et diminue de

fait le nombre d'individus à analyser pour chaque gène. La comparaison entre les régions 3'UTR endogène et contrôle est faite sur le même individu ce qui élimine les erreurs d'interprétations de localisation des fluorochromes entre deux paires de neurones très proches par exemple.

Environ 2 000 gènes de *C. elegans* sont régulés par les miARN. La mise au point d'un rapporteur de régulation post-transcriptionnelle facilement utilisable serait un outil précieux pour les études à venir.

2. Sélection des gènes de l'étude

Nous avons en premier lieu choisi de nous appuyer sur l'étude publiée en 2006 et donc de poursuivre avec des cibles du miARN *let-7*. En premier lieu, nous avons sélectionné le gène *lin-41*, gène hétérochronique nécessaire au développement cellulaire post-embryonnaire dont la régulation par le miARN *let-7* est connue et documentée (Abrahante et al. 2003; Cevec, Thibaudeau, and Plavec 2010; Großhans et al. 2005; Slack et al. 2000). Ensuite, nous avons inclus deux autres gènes de l'étude, les gènes *nhr-17* et *uba-1* codant respectivement pour un récepteur nucléaire (Sluder et al. 1999) et une enzyme d'activation de la voie ubiquitine (Jones et al. 2002; Kulkarni and Smith 2008). Le gène *nhr-17* parce qu'il semble bien cible d'une régulation post-transcriptionnelle et inversement, le gène *uba-1* qui avait donné des résultats semblant mettre en évidence l'absence de régulation post-transcriptionnelle. A ces trois gènes, nous avons ajouté le gène *ZK757.1*, cible prédite du miARN *let-7*, dont la fonction reste inconnue. Finalement, le gène *lin-28* code pour une protéine cytoplasmique indispensable au déroulement du développement au cours des stades larvaires. La protéine LIN-28 est particulièrement exprimée au cours des stades larvaires L1 et L2, puis son expression diminue progressivement aux stades L3 et L4 (Moss, Lee, and Ambros 1997; Seggerson, Tang, and Moss 2002). Ce gène a été inclus dans l'étude afin d'avoir un contrôle supplémentaire puisqu'il joue un rôle crucial dans la régulation du miARN *let-7* et fait partie de sa boucle de régulation. Le miARN *lin-4* cible l'ARNm de *lin-28* et diminue son expression ce qui permet de lever l'inhibition de l'expression du miARN *let-7*. La région *lin-28* 3'UTR contient également un site potentiel de fixation du miARN *let-7*.

Le choix de la région *unc-54* 3'UTR comme contrôle repose sur plusieurs années d'utilisation de celui-ci. La région est utilisée pour la première fois dans les années 1990 pour permettre la stabilisation d'un rapporteur transgénique Lac-Z (A. Fire, Harrison, and Dixon 1990) et par la suite mis à disposition dans le *Fire Lab Vector Kit*. Il a depuis été utilisé dans de nombreuses études afin de permettre l'expression de différentes constructions, principalement dans des études d'expression des gènes (Yan et al. 2009; Zou et al. 2012; Korswagen et al. 1997; Reece-

Hoyes et al. 2007; J. Wang and Barr 2005; Hunt-Newbury et al. 2007). Au cours de ces études, la région *unc-54* 3'UTR a été employée en fusion avec la GFP et l'expression de celle-ci a pu être détectée dans une grande variété de tissus. *unc-54* 3'UTR est donc considérée à juste titre comme largement permissive.

II. Objectifs des travaux de thèse

De nombreux travaux s'attachent à décortiquer les réseaux de régulations post-transcriptionnels existants. Les travaux entrepris dans le laboratoire avant mon arrivée avaient permis le développement de la sélection des animaux transgéniques sur G418 et l'assemblage d'un rapporteur de régulation post-transcriptionnelle (cassette RRPT).

Le couplage des deux systèmes a rapidement permis l'obtention de lignées transgéniques porteuses la cassette RRPT et résistantes au G418, et maintenues sur des boîtes NGM G418 tout au long des travaux.

Mon objectif était de confirmer et valider par microscopie que la cassette RRPT fonctionnait comme attendu avant d'étendre cette stratégie à d'autres gènes potentiellement cibles de régulations post-transcriptionnelles. La validation fonctionnelle de la cassette RRPT s'est appuyée sur l'analyse de son expression sous le contrôle des séquences régulatrices de cinq gènes, dont 2 sont régulés de manière avérée par les miARN.

Au cours de ce travail nous avons analysé le comportement de la cassette RRPT et d'autres systèmes de rapporteurs, ce qui a conduit à l'identification d'une régulation post-transcriptionnelle s'opérant sur la région *unc-54* 3'UTR que nous avons essayé de caractériser.

III. Matériel et Méthodes

Les lignées générées au cours de cette étude résultent d'un cheminement expérimental commun (Figure 13). Les méthodes expérimentales décrites dans ce chapitre détaillent ces différentes étapes.

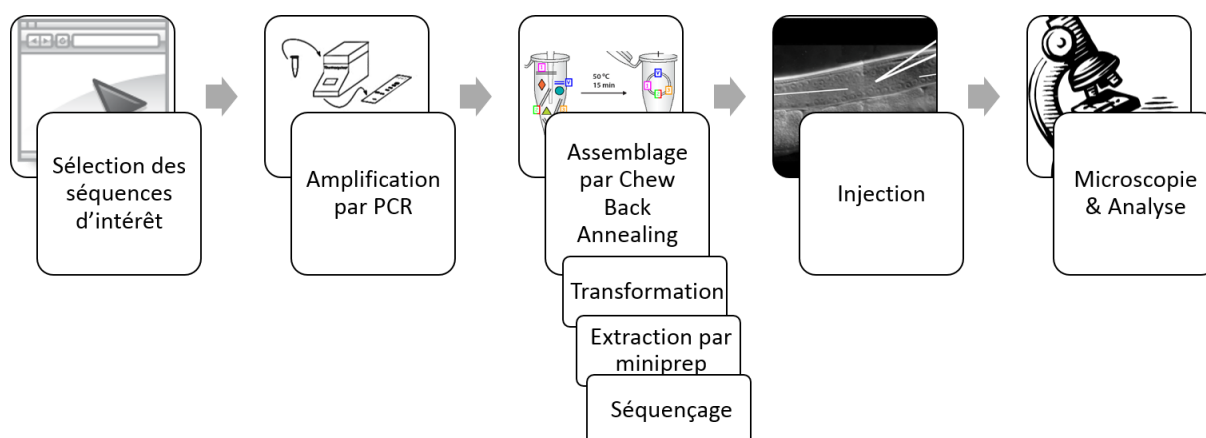


Figure 13 Procédure expérimentale générale pour la génération des lignées transgéniques de *C. elegans*.

1. Biologie moléculaire

1. Sélection des séquences régulatrices : promoteur et 3'UTR

Cinq gènes ont été sélectionnés pour valider le fonctionnement de la cassette RRPT : *lin-41*, *nhr-17*, *ZK757.1*, *uba-1*, et *lin-28*. Trois de ces gènes ont été inclus sur la base d'une étude préalablement publiée (Lall et al. 2006) qui avait pour but d'examiner l'efficacité de l'algorithme Pictar dans ses prédictions de régulation par les miARNs. Le premier gène choisi est *lin-41*, cible confirmée du miARN *let-7* et classé 16^{ème} par Pictar dans la liste des cibles de *let-7*. Les gènes *nhr-17* et *ZK757.1* sont aussi prédits comme cibles par Pictar respectivement 38^{ème} et 35^{ème} sur la liste des cibles potentielles de *let-7*. Le gène *uba-1* était dans l'étude de 2006, a été retiré de la liste des cibles prédites de *let-7* lorsque la stringence de l'algorithme a été augmentée (Lall et al. 2006) en y ajoutant la comparaison aux séquences d'autres nématodes afin de déterminer les séquences conservées, souvent signe de mécanismes de régulation également conservés. Les résultats de 2006 mettaient en évidence l'absence d'une régulation post-transcriptionnelle sur *uba-1* il a donc été inclus comme contrôle négatif. Enfin, le gène *lin-28* est aussi donné par Pictar comme cible de *let-7* (26^{ème} position), et est également la cible avérée du miARN *lin-4*. Le gène *unc-54* dont la région 3'UTR est utilisée comme contrôle est exprimé à tous les stades du développement dans le muscle et Pictar ne prédit aucune régulation sur cette région par un miARN.

Pour chaque gène sélectionné, il faut donc disposer des séquences situées en amont et en aval de la séquence codante. Les séquences régulatrices utilisées dans les différents plasmides ont été déterminées *in silico*. Selon les gènes, les séquences utilisées pour les promoteurs proviennent de l'ensemble de la région intergénique jusqu'à la fin de la région 3'UTR du gène précédent ou de la partie proximale de cette région. La plupart du temps, 2kb suffisent à reproduire l'activité du promoteur. Pour certains gènes, il faut une séquence plus longue en raison d'éléments de régulation situés plus en amont. Si des données d'expression faisaient état de la nécessité d'englober des régions supplémentaires, la longueur des séquences promotrices a été augmentée en conséquence jusqu'à 4kb.

Les séquences des régions 3'UTR ont été sélectionnées en aval de la séquence codante du gène, immédiatement après le codon stop. Si disponible, la séquence est vérifiée sur la plateforme UTRome (Marco Mangone et al. 2007). La plateforme UTRome regroupe les données issues de séquençage des régions 3'UTR de *C. elegans*. Partant du constat qu'environ 10% des gènes de *C. elegans* subissaient un épissage alternatif des derniers exons et avaient donc des régions 3'UTR variables pour un même gène, le clonage des différentes séquences 3'UTR a été réalisée pour une grande partie des gènes et les séquences ainsi obtenues compilées sur la plateforme. La plateforme s'appuie également sur la publication d'autres sources de données telles que WormBase, PicTar le ModENCODE Consortium, et est actualisée par les nouvelles séquences ou les données d'expression générées par le *pipeline* de clonage des laboratoires participants. Dans le cas où plusieurs séquences 3'UTR sont disponibles, la séquence la plus exprimée (à tous les stades) ou la plus longue (si les données d'expression sont absentes) a été choisie comme séquence à cloner dans les plasmides. Si aucune séquence 3'UTR n'était connue pour le gène, la séquence de 750pb en aval du gène a été utilisée.

2. Préparation du matériel génétique

A. Préparation d'ADN

L'ADN génomique des *C. elegans* de souche N2 ou transgéniques est extrait à l'aide du kit NucleoSpin Tissue (Machery Nigél). Deux ou plusieurs boîtes de diamètre 9cm contenant une population de nématodes majoritairement adultes sont rincées avec de l'eau milliQ pour récupérer les vers dans des tubes de 2ml. Les tubes sont placés dans la glace afin de sédimenter les animaux. Après élimination du surnageant, les nématodes sont resuspendus dans 500µl de tampon pré-lyse T1 du kit. La suspension est ensuite placée à -20°C pour plusieurs heures ou toute la nuit afin de fragiliser les tissus. Après décongélation, 50µl de protéinase K sont ajoutés pour l'étape de lyse. L'ensemble est incubé pendant 18h à 56°C. L'échantillon est ensuite vortexé

et remplacé à 56°C jusqu'à ce qu'il n'y ait plus de débris visibles. Une étape supplémentaire de centrifugation est nécessaire pour culoter les débris afin de ne pas obstruer la membrane du kit.

L'échantillon est chargé sur la colonne à cette étape et la suite du protocole se déroule selon les recommandations du fabricant. L'élution est réalisée avec 100µl du tampon d'élution du kit préalablement chauffé à 70°C après 3 minutes d'incubation sur la colonne.

L'ADN ainsi extrait est stocké à -20°C et une partie est aliquotée à une concentration de 100 ng/µl pour les applications ultérieures de PCR.

B. Préparation d'ARN

L'ARN des *C. elegans* peut également être extrait pour vérifier les niveaux d'expression de certains gènes et transgènes par une PCR semi-quantitative. Le protocole utilisé ici a été adapté des protocoles d'extraction classique et de celui fourni avec le kit RNeasy Mini Kit (QIAGEN). Le matériel doit être préalablement décontaminé au DEPC ou bien autoclavé 2 fois afin d'éliminer toutes les RNases.

Les populations utilisées sont d'abord synchronisées selon le protocole décrit plus bas (Décontamination et synchronisation des vers) afin d'obtenir des ARN totaux spécifique des différents stade de développement. Les animaux sont lavés 3 fois dans du M9 dans des tubes Falcons 15 ml. Lors de la dernière centrifugation le culot final (environ 100µL) est repris dans 1mL de Trizol et transféré dans un tube propre. Les échantillons sont ensuite congelés dans de l'azote liquide et conservés à -80°C. Les échantillons subissent ensuite 5 cycles successifs de chauffe à 37°C suivi par 30s de vortex et une congélation dans l'azote liquide. Les échantillons sont déposés dans la glace 30s puis agités avec un vortex 30s, pendant 3 minutes. Les échantillons sont laissés au repos à température ambiante 5min avant l'ajout de chloroforme. La solution est homogénéisée par inversion des tubes. Les échantillons sont ensuite centrifugés 15min à 4°C et le surnageant transféré dans un tube propre et additionné d'un volume d'éthanol 70% froid.

La suite du protocole est conduite avec le kit QIAGEN. L'échantillon est chargé sur la colonne et celle-ci subit plusieurs cycles de lavages par centrifugations successives conformément aux recommandations du fournisseur. L'ARN est finalement élué dans 50 µl d'eau *RNAse-free* et conservé à -80°C jusqu'à utilisation.

3. PCR

Les amorces sont définies sur la base des séquences disponibles sur le site Wormbase et conçues pour avoir majoritairement une température d'hybridation entre 50° et 60°C. Les amorces

employées proviennent du fournisseur Sigma Aldrich. La liste des amorces ainsi que leurs séquences sont disponibles en annexe. Pour les amplifications classiques les amorces ont une longueur de 17 à 22nts. Pour les amplifications spécifiques à la méthode du *chew back annealing* elles sont composées de 40 à 60nts (18 à 22nts hybridant la séquence à l'extrémité 3' de l'amorce et 20 à 40 nts de « queue flottante » à l'extrémité 5').

Selon les fragments à amplifier différentes polymérases sont employées. Pour les fragments de 100pb à 2kb la Taq polymérase ou la Q5 (NEB) sont employées. Pour les autres fragments, la Phusion est employée.

Les matrices ADN sont ajoutées selon les quantités suivantes selon les recommandations du fournisseur (New England Biolabs) : ADN génomique total 100ng, ADN plasmidique 10 à 25ng.

Le programme est composé des mêmes étapes :

Dénaturation : 5 min à 98°C

35 cycles d'amplification :

 dénaturation 20 secondes à 98°C

 hybridation 20 secondes des amorces à la température adéquate

 élongation d'une durée dépendante de la taille du fragment (1 min pour 1kb avec la Taq polymérase, 20 sec pour 1kb avec la Phusion ou la Q5)

Elongation finale 5 min à 72°C.

Les produits obtenus par PCR sont ensuite vérifiés sur gel d'agarose par électrophorèse avant purification. La purification des fragments est réalisée avec le kit QIAquick purification kit (QIAGEN) selon les recommandations du fournisseur. Les produits purifiés sont revérifiés sur gel d'agarose par électrophorèse et dosés au nanodrop.

A. Rétro-Transcription d'ARN

Le protocole de rétro-transcription se déroule en 2 étapes. 500ng d'ARN sont mis en présence de 0,5 µM primer R (spécifique aux gènes cibles), 0,25nM dNTP et incubés 5 min à 70°C et mis dans la glace. 4µL d'un mix contenant 2µL de tampon RT, 0,5µL de Murine RNase Inhibitor et 1µL MuLV RT sont ajoutés avant d'incuber 50 min à 42°C puis 5 min à 80°C.

Les ADNc ainsi générés peuvent être conservés à -20°C ou utilisés directement comme matrice PCR.

4. Chew back annealing

Le Chew Back Annealing est une méthode de clonage développée en 2009 (Gibson et al. 2009; Gibson 2011) et utilisée dans le laboratoire pour la génération des plasmides. Chaque plasmide à assembler est constitué par des fragments amplifiés indépendamment par PCR. Les fragments sont amplifiés de sorte que leurs extrémités soient chevauchantes de 20 à 30 nucléotides grâce à des queues flottantes spécifiques présentes sur certaines amorces. Les fragments à assembler sont mélangés dans un tube PCR dans de la glace, en quantité équimolaire soit 25nmol pour les fragments de plus d'1kb et 5 fois plus pour les fragments inférieurs à cette taille (conversion avec : <http://www.promega.com/techserv/tools/biomath/calc01.htm>). Le mélange final doit avoir un volume de 5µl et une quantité d'ADN en présence inférieure ou égale à 200ng pour une efficacité optimale. Toujours sur glace, 15µl de mélange de réaction (5 X tampon de réaction ISO (500 mM Tris-HCl ; 50 mM MgCl₂ ; 1 mM dNTPs ; 50 mM DTT ; 25% PEG-8000 ; 5 mM NAD), T5 exonucléase (10 U/µl), Phusion polymérase (2 U/µl), T4 Ligase (400 U/µl), H₂O) sont ajoutés dans le tube. Après 2 à 3 aspirations refoulements avec le pipetman, le tube est placé dans le thermocycleur à une température de 50°C pour 1h. Après 1h, 1µl de la réaction peut être utilisé pour transformation de bactéries compétentes.

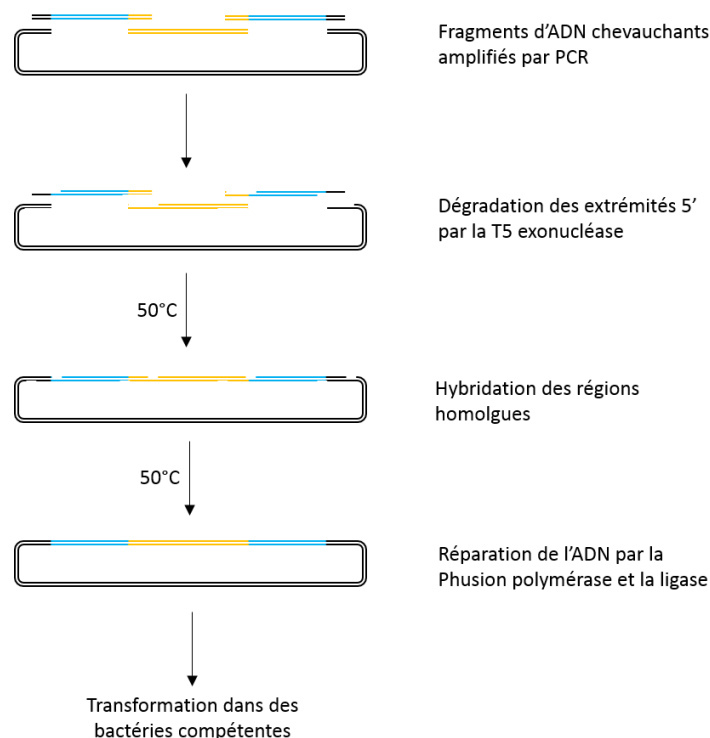


Figure 14 Principe de la technique du Chew Back Annealing. Les fragments mis en présence vont subir la dégradation de leurs extrémités 5' par la T5 exonucléase. Celle-ci est progressivement inactivée à 50°C. Les régions homologues vont ensuite s'hybrider avant que la phusion polymérase et la ligase ne viennent compléter les trous et assembler l'ensemble. Après quelques minutes de réaction l'ADN ainsi généré est inséré dans des bactéries compétentes.

5. Mutagenèse dirigée

Une partie du projet a nécessité de réaliser des mutations ciblées sur certains des plasmides utilisés comme rapporteurs. Les mutagenèses dirigées effectuées étaient exclusivement des délétions. Un protocole rapide et simple a été mis en place pour obtenir des plasmides contenant des délétions de 9 à 15 pb.

Des amorces ont été choisies de part et d'autre de la séquence à supprimer. L'ensemble du plasmide est amplifié par PCR avec ces amorces de sorte à obtenir un fragment linéaire à bord franc. L'enzyme DpnI est ajouté directement au mélange réactionnel de la PCR afin de dégrader la matrice. La réaction se poursuit dans le thermocycleur 1h à 37°C. Le fragment obtenu est ensuite purifié avec le QIAquick purification kit (QIAGEN) selon les recommandations du fournisseur. A chaque étape les produits sont vérifiés sur gel d'agarose par électrophorèse.

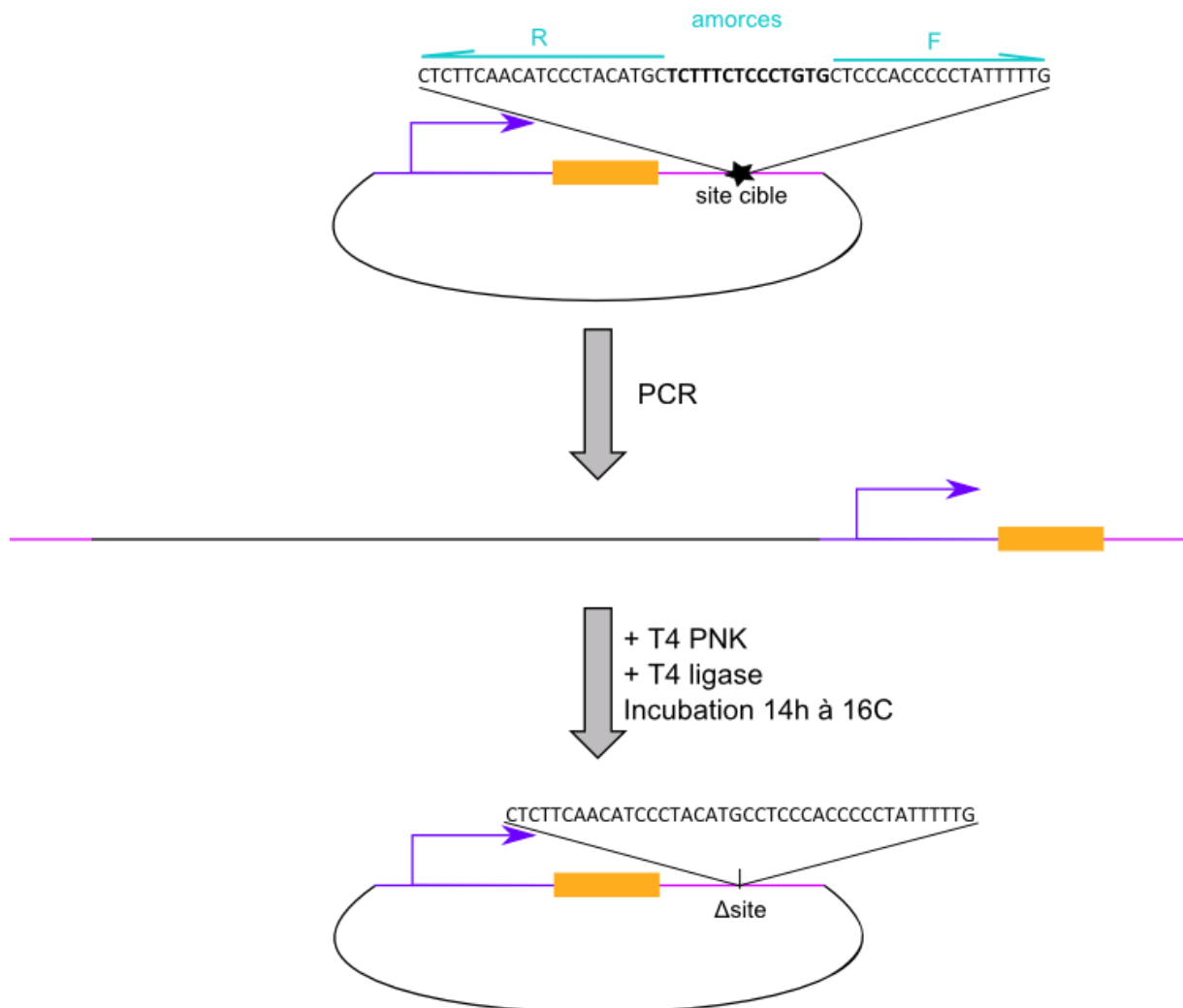


Figure 15 Mutagenèse dirigée des plasmides. Les amorces placées de part et d'autre du site cible permettent l'amplification par PCR d'un unique fragment à bord franc. Après purification le fragment est reliqué grâce à l'action conjointe de la T4 Polynucléotide Kinase et de la T4 ligase.

Après dosage au nanodrop, 200ng du fragment sont incubés avec la T4 et la PKN à 16°C sur la nuit dans un volume final de 20µl. 1,5µl de cette réaction sont utilisés directement pour transformer des bactéries compétentes.

6. Génération de plasmides

Qu'il s'agisse de la méthode du *chew back* annealing ou de la mutagenèse, les étapes pour la production des plasmides ont été réalisées de la même manière. L'ADN est introduit dans des bactéries électrocompétentes DH5α. 50µl de bactéries sont placées avec l'ADN dans des cuvettes de 2mm préalablement refroidies dans la glace. L'électroporation est réalisée à 2,3kV. Les bactéries sont ensuite transférées dans 200µl de milieu LB (Luria Broth) sans antibiotique et incubées 1h à 37°C sous agitation (180 rpm). Ces cultures sont étalées sur des boîtes LB agar additionnées d'ampicilline (100µg/ml) et incubées sur la nuit à 37°C.

Les bactéries présentes sur la boîte sont prélevées et mises en culture pour 8h à 12h dans du milieu LB liquide additionné d'ampicilline (100µg/ml). A cette étape, une PCR sur colonie peut être réalisée pour valider la présence de l'insert ou de la délétion. Les plasmides sont extraits des bactéries grâce au kit de miniprep QIAGEN utilisé selon les recommandations du fournisseur. De manière générale l'étape de l'ajout du tampon PB n'est pas réalisée, et l'élution finale a lieu dans 50µl du tampon EB fourni avec le kit et compatible avec la micro-injection dans *C. elegans*. Les plasmides sont finalement dosés et vérifiés par des digestions enzymatiques ou bien par des PCR.

Les digestions enzymatiques sont réalisées selon les recommandations du fournisseur (NEB). Environ 200ng d'ADN sont digérés dans un mélange réactionnel contenant le tampon adapté à la concentration 1X et 2U d'enzyme(s) dans un volume final de 10µl. Certains enzymes nécessitent l'ajout de BSA afin de favoriser la réaction, chaque réaction a donc été réalisée selon les conditions appropriées. Les enzymes utilisés fonctionnent à 37°C. La digestion est vérifiée par une électrophorèse sur gel d'agarose (1%).

2. Protocoles de *C. elegans*

1. Maintenance

Les animaux sont maintenus sur des boîtes de pétri sur le milieu de croissance des nématodes (Nematode Growth Media ; NGM)ensemencées avec des bactéries *E. coli* OP50 (auxotrophes pour l'uracile) comme source de nourriture.

La préparation d'1L de NGM se fait en deux étapes : 3g de NaCl, 2,5g de peptone, 12g d'agar, 1ml de cholestérol (5mg/ml) et 975ml d'eau sont autoclavés et refroidis avant ajout de 1ml de CaCl₂ (1M), 1ml de MgSO₄ (1M) et de 25ml d'une solution de potassium phosphate (1M). 1L de solution de potassium phosphate se prépare avec 132ml de K₂HPO₄ (1M) et 868ml de KH₂PO₄ (1M) stérilisés par filtration.

Une fois le milieu NGM dispensé dans les boîtes de pétri, celles-ci sont laissées au moins 24h à sécher sur la paillasse. Une fois sèches, les boîtes peuvent êtreensemencées avec une culture fraîche de bactéries OP50 et seront utilisables après 24h supplémentaires à température ambiante.

Les boîtes peuvent être stockées à 4°C plusieurs semaines avant ou après avoir étéensemencées avec les bactéries OP50.

Le laboratoire a développé la sélection de *C. elegans* sur milieu NGM additionné de G418. Les boîtes sont préparées comme décrit ci-dessus. L'antibiotique G418 (0,4g/L) est ajouté dans le milieu après l'autoclave, une fois le milieu suffisamment refroidi (65°C environ). Les boîtes sontensemencées avec une souche d'OP50neoR, une souche possédant le plasmide de résistance au G418. Les bactéries OP50neoR sont cultivées en milieu LB additionné d'antibiotique : la kanamycine (50 µg/l) ou la neomycine (50 µg/l).

La sélection par l'antibiotique G418 à une concentration de 0,4g/l empêche le développement des animaux au-delà du stade L2. Les animaux déposés sur les boîtes après avoir dépassé le stade de développement L2 peuvent survivre et poursuivre leur croissance. Ainsi chaque lot de boîte est testé en déposant une dizaine d'individus non résistants au G418 sur des boîtes. Si aucun vers ne se développe le lot peut être utilisé.

Les techniques de maintenance classiques s'appliquent également sur G418. Les lignées sont entretenues par repiquage sur les boîtes adaptées quand la source de nourriture est épuisée. Pour cela, selon les besoins, les individus sont soit repiqués individuellement à l'aide d'une anse de platine, soit déplacés sur un morceau d'agar (chunk) déposé sur la nouvelle boîte.

2. Décontamination et synchronisation des vers

Le traitement à l'hypochlorite de sodium peut être utilisé pour décontaminer ou synchroniser une population. La population est lavée dans du tampon M9 puis incubée dans une solution de soude (0,5 M) et d'hypochlorite de sodium (5 %). Le tout est agité sur un vortex à température ambiante. La solution permet de dégrader la cuticule des adultes tout en préservant les œufs, insensibles au traitement. Quand la moitié des animaux paraissent dégradés, la solution est

centrifugée 2 min à 2000g afin de précipiter les œufs. Trois lavages avec du tampon M9 sont ensuite réalisés afin d'éliminer les restes de solution d'hypochlorite de sodium et de soude. Les œufs sont ensuite déposés sur une boîte fraîche.

3. Transgénèse

La transgénèse est une technique qui consiste à introduire de l'ADN exogène dans un organisme. Pour *C. elegans*, il existe différentes méthodes conduisant à l'obtention de lignées transgéniques exprimant un transgène soit sous forme de concatémère extrachromosomique soit intégré dans l'ADN génomique. Les lignées transgéniques présentées dans cette thèse ont été générées au laboratoire par la technique de micro-injection (Mello et al. 1991) qui consiste à injecter dans les gonades du vers un mélange d'ADN.

Les différents mélanges d'injection ont été réalisés afin d'obtenir une concentration finale d'ADN de 100 ng/μl. Chaque plasmide est présent en quantité équimolaire, sauf le pRG02 contenant la résistance au G418 (gène neoR) présent à 10 ng/μl. De jeunes adultes sont immobilisés sur une lamelle recouverte d'une fine couche d'agarose 2% séchée sur laquelle une goutte d'huile est déposée. Après injection les animaux sont récupérés dans une goutte de milieu M9 et déposés sur une boîte de NGM additionnée de G418. Une boîte contrôle est réalisée avec une dizaine d'adultes de la souche utilisée pour les injections afin de contrôler l'efficacité de la sélection des boîtes. Seuls les individus transgéniques sont en mesure de se développer sur le milieu sélectif.

Lors des sessions d'injection 10 à 15 individus injectés sont déposés par boîte NGM G418. Ces individus sont les P0. Leurs descendants les F1 sont ensuite isolés au stade L4 ou jeune adulte environ 5 jours après l'injection. Les descendants des F1, les F2, sont examinés pour contrôler la présence du transgène. Si le transgène est présent, alors la nouvelle lignée ainsi générée est conservée pour analyse.

4. Microscopie

Les lignées employées pour ce travail sont des lignées fluorescentes dont l'analyse a été réalisée grâce à un microscope Axio Imager Z1 Zeiss couplé à une lampe HXP 120 et équipé d'une caméra AxioCam MRm pour l'acquisition des images.

Pour l'imagerie et l'analyse, les animaux sont placés sur une lame recouverte d'une couche d'agarose 5% dans une solution d'azoture de sodium (15mM). L'azoture de sodium est un neurotoxique permettant de paralyser les animaux sans les tuer immédiatement. L'ensemble est recouvert d'une lamelle de microscopie.

Les images ont été réalisées avec des paramètres adaptés aux lignées dont l'intensité de la fluorescence pouvait varier. La GFP a été enregistrée sur 20 à 200ms, la mCherry et la dsRed ont été enregistrées entre 200ms et 1s. Les images ont ensuite été traitées avec le logiciel ImageJ pour la réalisation des figures.

A. Coloration au DiI

Le DiI (1,1'-dioctadecyl-3,3,3'-tetramethylindocarbocyanine perchlorate) est un colorant ciblant une douzaine de neurones sensoriels (amphides, phasmides et IL sous certaines conditions) identifiés chez *C. elegans*. La coloration est révélée sous l'exposition à la lumière UV dans le rouge. L'identification des neurones est donc facilitée par l'utilisation de ce colorant qui fournit des repères spatiaux précis. Le protocole suivi est adapté de ceux disponibles sur le wormatlas. Les vers sont prélevés et incubés sur une roue pendant 2 à 3h dans une solution de DiI diluée au 1/200 dans du M9. Les animaux sont ensuite lavés deux fois dans du M9 puis remis sur une boîte NGM recouverte de bactéries OP50 pendant 1h minimum avant de les examiner au microscope.

5. Mutagenèse à l'EMS

La mutagenèse par EMS (méthanesulfonate d'éthyle) est un procédé utilisé en génétique depuis plusieurs années, y compris sur *C. elegans* (S. Brenner 1974; E. B. Lewis and Bacher 1968; Loveless 1959). L'EMS permet de générer des mutations aléatoires dans le génome en substituant les bases G en A, ce qui conduit à l'obtention de mutants le plus souvent nuls pour un ou plusieurs gènes. Sa manipulation requiert des précautions. Les déchets solides et liquides ayant été en contact avec l'EMS sont inactivés avec de la soude 1N pendant 3 jours avant d'être éliminés avec les déchets chimiques. Les manipulations doivent être effectuées sous sorbonne avec un équipement de protection adapté : 2 paires de gants, lunettes de protection et blouse.

Une boîte de diamètre 6cm contenant des vers non synchronisés est utilisée pour la mutagenèse. Les animaux sont récupérés dans du M9 et lavés 3 fois avec du M9 pour éliminer les bactéries résiduelles. Après le dernier lavage, les nématodes sont resuspendus dans une solution d'EMS pure diluée au 1/100 dans du M9 et placés sur une roue pendant 4 à 6h. A la fin de l'incubation, 3 nouveaux lavages avec du M9 sont réalisés. Les vers sont ensuite placés sur une boîte de NGM et 10 à 20 L4 sont isolés sur des nouvelles boîtes de diamètre 9cm. Après 10 à 15 jours, la descendance est examinée pour le phénotype recherché.

Les animaux présentant un phénotype mutant sont isolés manuellement et la descendance est vérifiée après plusieurs jours de croissance afin d'obtenir des lignées homozygotes pour la mutation.

Ce protocole a été réalisé sur des lignées transgéniques non intégrées résistantes au G418 sur des boîtes NGM additionnées de G418.

6. Croisements de lignées

Afin de réaliser des croisements, des mâles sont générés par choc thermique en plaçant 20 animaux de stade L4 sur une boîte NGMensemencée de bactérie OP50 pendant 6h à 30°C. Les mâles doivent apparaître dans la descendance après quelques jours de croissance.

Les mâles (3 à 5) ainsi générés sont placés sur une boîte NGMensemencée de diamètre 3.5 cm avec un hermaphrodite. Les phénotypes recherchés sont ensuite sélectionnés manuellement parmi la descendance.

IV. Résultats

1. Utilisation d'un rapporteur de régulation post-transcriptionnelle sous forme d'opéron

1. Développement du projet

De nombreuses bases de données compilent les données sur les régions 3'UTR des gènes ou les miARNs présents dans les organismes (miRBase, UTRome, microRNA.org par exemple). Des algorithmes permettant de prédire les interactions entre les 3'UTR et les miARN sont disponibles par internet et de plus en plus précis, enrichis au fur et à mesure de l'accumulation des données de séquençage et de détection de nouveaux petits ARNs non codant. L'ensemble de ces données ne permet néanmoins que de générer des prédictions, et la validation fonctionnelle de chaque interaction reste indispensable à la compréhension des mécanismes plus généraux régissant la régulation de l'expression génétique.

Plusieurs études se sont appuyées sur une différence d'expression de rapporteurs fluorescents exprimés simultanément pour observer les régulations post-transcriptionnelles (Lall et al. 2006; Merritt et al. 2008; Hsieh, Chang, and Chuang 2012). Ces études ont permis de valider expérimentalement des régulations post-transcriptionnelles s'opérant dans les régions 3'UTR des ARNm des gènes. Ces deux études se sont pour cela appuyées sur l'utilisation de rapporteurs fluorescents, indépendants ou en opéron (Figure 16). L'utilisation de rapporteurs fluorescents est aisée chez le nématode *C. elegans* en raison de sa transparence qui autorise l'observation *in vivo* de l'expression de la fluorescence.

Les travaux entrepris dans notre laboratoire par Ilyass Zniber au cours de sa thèse ont abouti à la construction d'une cassette utilisée comme rapporteur de régulation post-transcriptionnelle (RRPT). La cassette est un élément fixe composée de la séquence codante de la mCherry suivie par *unc-54* 3'UTR connectée grâce à un *operon linker* (séquence intergénique d'un opéron endogène *gpd-2/gpd-3*) à la GFP. Le promoteur du gène étudié est cloné en amont de la mCherry, et sa région 3'UTR en aval de la GFP. De cette manière, les deux protéines sont transcrites sur le même pré-ARNm. Le pré-ARNm polycistronique subit une étape de trans-épissage qui permet la maturation indépendante des différents cistrons en ARNm. Les différences d'expression entre GFP et mCherry devraient résulter d'une régulation dans les régions 3'UTR.

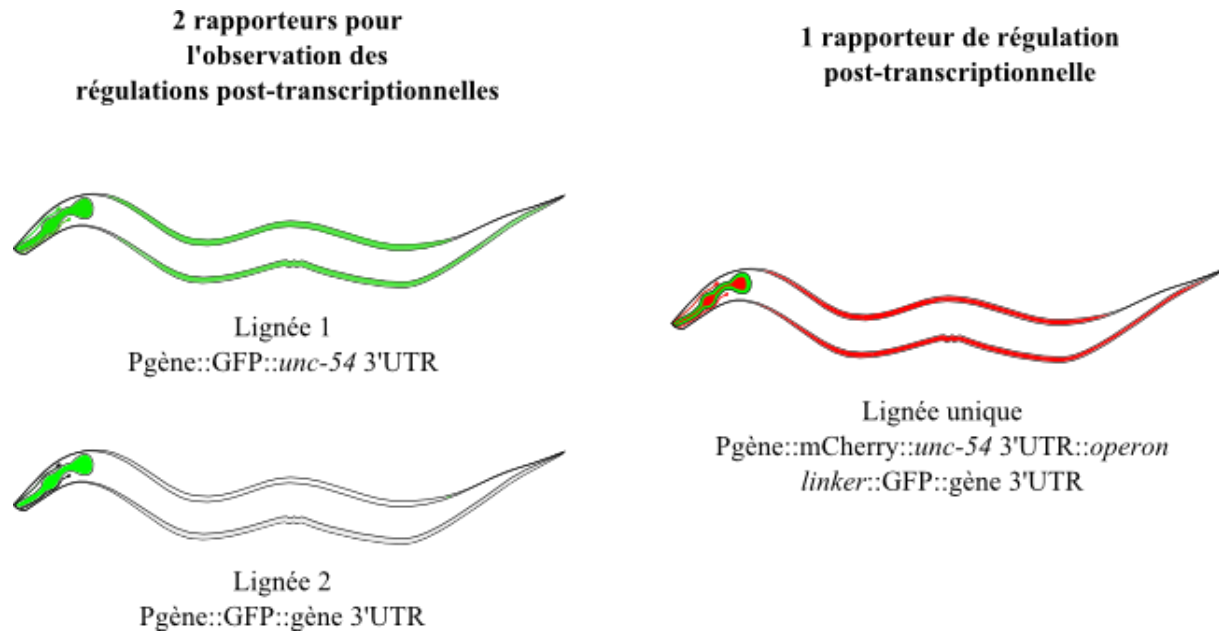


Figure 16 Les différentes stratégies d'études des régulations post-transcriptionnelles à l'aide de rapporteurs fluorescents. A gauche, la stratégie développée dans l'étude de 2006 (Lall et al. 2006) faisant appel à deux lignées exprimant le même rapporteur, à droite la stratégie employant une seule lignée exprimant deux rapporteurs.

Comme indiqué dans l'introduction, cinq gènes ont été choisis pour valider expérimentalement cette stratégie. Trois de ces gènes figuraient dans l'étude de 2006 : *lin-41*, *uba-1* et *nhr-17* (Lall et al. 2006). Deux gènes supplémentaires ont été ajoutés : *lin-28* cible du miARN *lin-4*, et *ZK757.1* cible prédictée de *let-7* dont les fonctions ne sont pas connues.

2. Analyse phénotypique des lignées transgéniques

Pour chaque gène étudié, plusieurs lignées indépendantes portant la cassette RRPT ont été générées par micro-injection. Nous avons voulu vérifier que la construction sous forme de cassette était fiable pour les études de régulation post-transcriptionnelle et qu'aucun de ses éléments n'interférait avec l'expression des rapporteurs.

En plus des plasmides contenant les rapporteurs fluorescents, le mix d'injection contenait le marqueur de sélection conférant la résistance à l'antibiotique G418 développé au laboratoire (Giordano-Santini et al. 2010).

Les lignées ont été examinées au microscope et les individus classés en trois catégories selon leur stade de développement, L1/L2, L3/L4, et Adulte. Pour chaque individu il a été noté où étaient localisées les protéines de fluorescences GFP et mCherry. Les résultats ont ensuite été compilés de manière à visualiser le nombre d'individus exprimant une fluorescence dans un tissu donné, de manière similaire à l'étude réalisée en 2006 par l'équipe de Fabio Piano (Lall et al. 2006). Cette étude était basée sur un minimum de 10 individus par stade sur 3 lignées.

Les individus observés sont prélevés au hasard dans une population d'animaux transgéniques résistants au G418 et ne sont donc pas présélectionnés sur la base de l'expression de la GFP, ce qui évite un biais de sélection en n'analysant les individus dont la fluorescence serait plus forte. Pour les stades L1/L2, quand la sélection n'a pas encore opéré, seuls des animaux exprimant au moins la GFP ou la mCherry ont été pris en compte dans l'analyse.

A. Détermination du nombre d'individus à analyser

L'analyse des populations est un travail délicat, et il faut s'assurer que l'échantillon choisi soit suffisant pour représenter le phénomène observé. Dans ce cas, les animaux sont des clones issus d'un même parent et porteurs du même fond génétique et du même transgène. Comme dans l'étude de 2006 de l'équipe de F. Piano, le transgène n'est pas intégré. L'étude initiale présente des résultats obtenus pour l'analyse de 10 individus analysés par stade, issus de 3 lignées indépendantes.

Chaque lignée générée dans cette étude exprime un concatémère extrachromosomique différent, avec un nombre de copies variable des transgènes. L'expression des transgènes peut donc varier d'une lignée à l'autre, c'est pourquoi lors d'études sur des lignées non intégrées il est essentiel d'observer plusieurs lignées. Initialement, l'étude devait porter sur 3 lignées indépendantes par transgène. Au cours de l'analyse, nous avons constaté que l'expression d'un transgène semblait ne pas varier d'une lignée à l'autre et nous avons poursuivi l'analyse sur 2 lignées en s'assurant que le transgène avait une expression similaire dans ces lignées.

Après l'analyse de deux lignées pour exprimant la GFP ou la RRPT) nous avons analysé leurs expressions respectives pour voir si l'expression du transgène était similaire entre les deux lignées. Dans le cas où l'expression des transgènes ne varie pas entre deux lignées, on peut conclure qu'elle est constante et sans doute reflète celle du gène. L'observation d'une seule lignée pourrait conduire à des conclusions erronées si les transgènes n'y sont pas exprimés de manière normale. Le phénomène de *silencing* est parfois observé dans les lignées transgéniques porteuses des transgènes en mult copie. La présence en grand nombre du même transgène empêche son expression. L'expression similaire de deux transgènes signifie que les observations conduites sur la lignée 1 ne présentent pas de différences significatives avec la lignée 2. Pour les dix transgènes différents de l'étude nous avons donc réalisé une comparaison de l'expression entre les lignées porteuses du même transgène : la GFP (Figure 17) et les fluorochromes de la cassette RRPT (Figure 18).

L'analyse a été réalisée en notant pour chaque individu la présence de la fluorescence dans le tissu observé (1) ou son absence (0). Pour chaque tissu étudié la fluorescence a été mesurée en

exprimant le nombre d'individus exprimant la fluorescence dans le tissu, ce qui a donné un score compris entre 0 et 30. Cette valeur a été convertie en pourcentage d'individus fluorescents pour chaque lignée. Ces pourcentages ont été comparés avec le test non paramétrique de Kruskal-Wallis pour l'ensemble des tissus. Le test statistique non-paramétrique de Kruskal-Wallis est une variante de l'ANOVA employée pour les analyses de population. Il permet de tester si les échantillons (2 ou plus) proviennent de la même population, ou de populations ayant des caractéristiques identiques. Le test de Krukak-Wallis a permis de valider l'absence de différence significative d'expression des transgènes ($p\text{-value} > 0.2$). Un même transgène est exprimé de la même manière entre deux lignées indépendantes entraînant un phénotype d'expression similaire. Dans le cas où le transgène induit l'expression de deux fluorophores c'est aussi le cas pour chacun. Les lignées ayant le promoteur du gène *lin-28* présentent une fluorescence dans les gonades, principalement le syncytium. C'est pourquoi une colonne « Go » figure sur les graphiques.

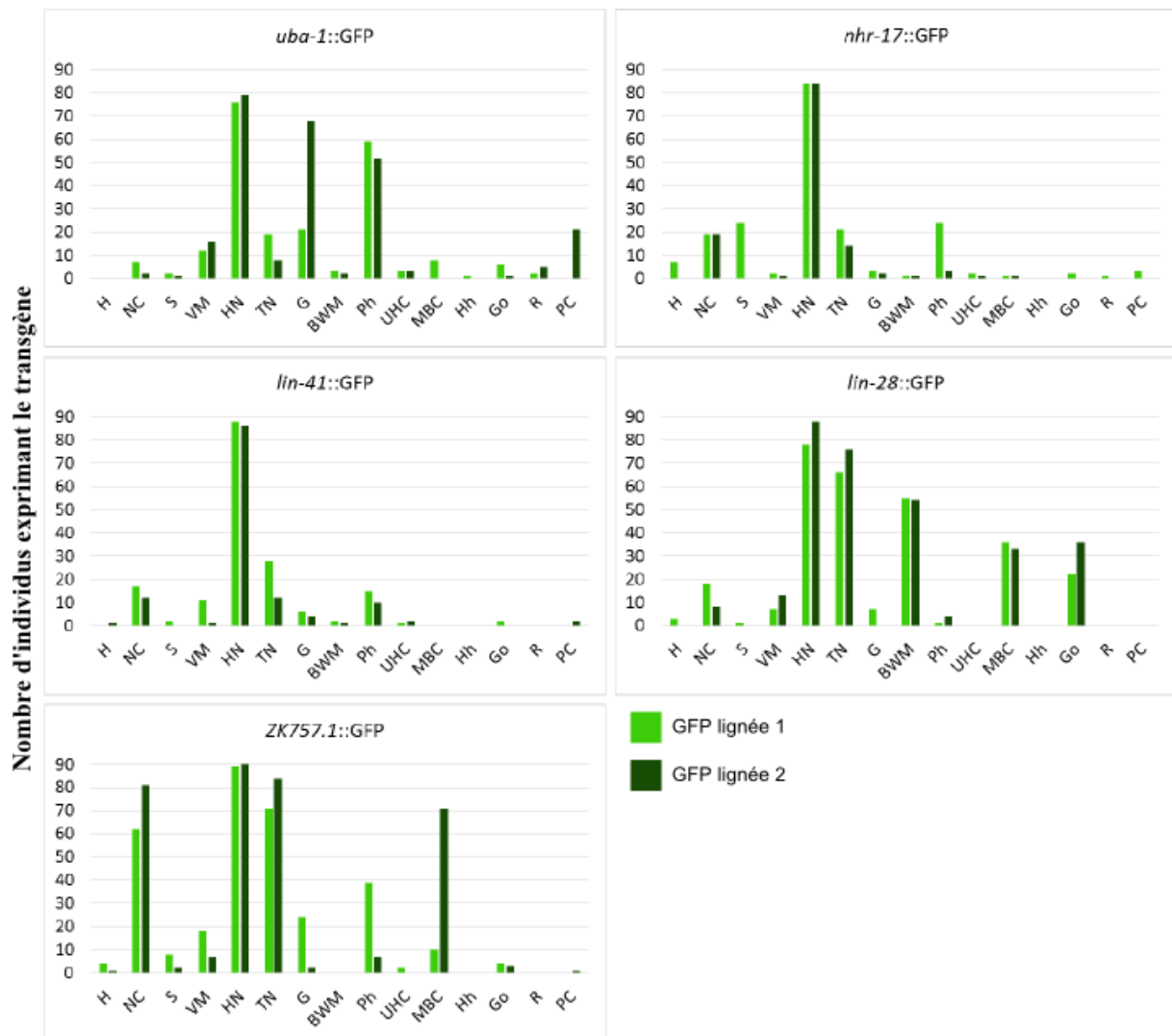


Figure 17 Expression des transgènes GFP dans les lignées indépendantes. Les transgènes ont une expression similaire entre deux lignées, les variations observées n'étant pas statistiquement significatives (test de Kruskal-Wallis $p\text{-value} > 0,2$ dans tous les cas). **Tissus :** H, hypoderme ; NC, corde nerveuse centrale ; S, cellules hypodermiques latérales ; VM, muscle de la vulve ; HN, neurones antérieurs ; TN, neurones postérieurs ; G, intestins ; BWM, muscles du corps ; Ph, pharynx ; UHC, paire de cellules musculaires de la tête ; MBC, corps cellulaires du corps ; Hh, cellules hypodermiques de la tête ; Go, gonades ; R, cellules adjacentes au rectum ; PC : cellules postérieures.

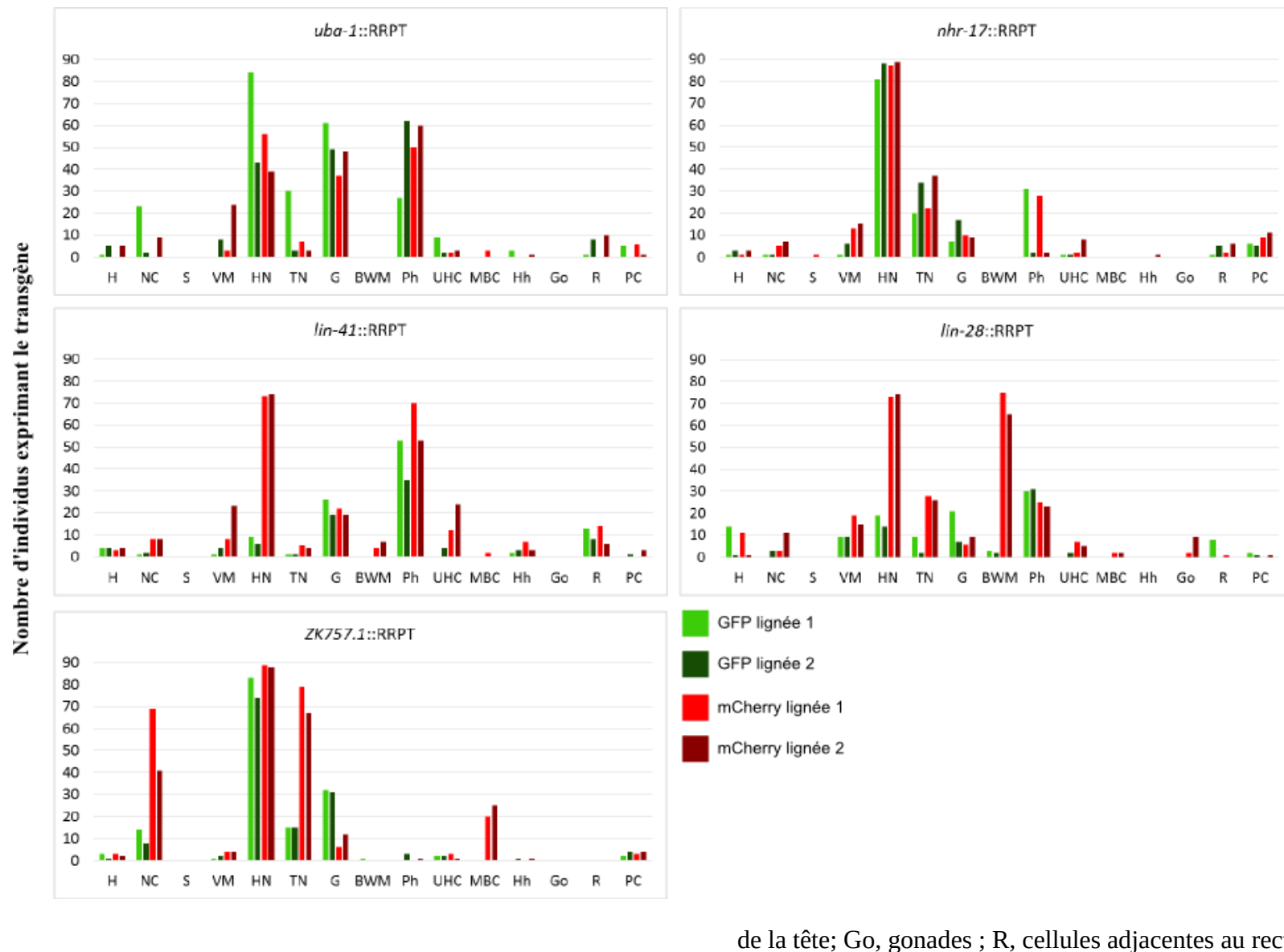


Figure 18 Expression des transgènes RRPT dans les lignées indépendantes. Sur les graphiques l'expression est donnée pour chaque transgène dans les deux lignées analysées, par tissu. Pour plus de lisibilité les GFP de chaque lignée sont présentées côte à côte suivie par les mCherry. Les transgènes ont une expression similaire entre deux lignées, les variations observées n'étant pas statistiquement significatives (test de Kruskal-Wallis p -value > 0,2 dans tous les cas). **Tissus** : H, hypoderme ; NC, corde nerveuse centrale ; S, cellules hypodermiques latérales ; VM, muscle de la vulve ; HN, neurones antérieurs ; TN, neurones postérieurs ; G, intestins ; BWB, muscles du corps ; Ph, pharynx ; UHC, paire de cellules musculaires de la tête ; MBC, corps cellulaires du corps ; Hh, cellules hypodermiques de la tête ; Go, gonades ; R, cellules adjacentes au rectum ; PC : cellules postérieures.

Dans la lignée *uba-1::RRPT* la GFP est exprimée dans des tissus en l'absence de la mCherry, ce qui indique que la région *uba-1* 3'UTR ne semble pas subir de régulation post-transcriptionnelle et qu'un autre mécanisme de régulation empêche l'expression du rapporteur contrôlé suivi par la région *unc-54* 3'UTR. Les mêmes observations avaient été faites par les auteurs de l'étude de 2006, qui avaient également émis l'hypothèse d'une autre régulation s'appliquant au rapporteur contrôlé (Lall et al. 2006; Vasudevan, Tong, and Steitz 2007).

Les lignées exprimant les mêmes constructions présentent des phénotypes d'expression des protéines de fluorescence similaires. L'étude peut donc être menée sur deux lignées par transgène. L'analyse de l'expression des transgènes a été conduite sur l'ensemble de la population, en regroupant l'ensemble des mesures réalisées sur chacun des stades.

Les animaux porteurs de transgènes non intégrés sous forme de concatémère extrachromosomique peuvent également exprimer ceux-ci de manière mosaïque. L'expression mosaïque est liée au fait que le concatémère extrachromosomique peut être perdu lors d'une mitose au cours du développement, ce qui entraînera une absence d'expression dans certains tissus de l'animal. Cet effet implique que l'expression observée du transgène peut varier entre deux individus d'une même lignée. C'est pourquoi il est essentiel d'avoir un nombre suffisant d'individus pour l'analyse.

Nous avons consolidé nos observations en menant sur plusieurs lignées une étude sur la différence des résultats qui proviendraient d'un échantillon réduit. Cette analyse a été conduite sur les lignées *uba-1::GFP* et *lin-41::GFP*. 50 individus par catégorie pour trois lignées indépendantes ont été observés. Le nombre total d'individus fluorescents a été exprimé en pourcentage d'individus fluorescents par tissus afin de refléter l'expression générale du transgène (Figure 19). Nous avons ensuite diminué le nombre d'animaux à observer par lignée. Le nombre d'individus observé par lignée est passé de 50 à 30 par catégorie. Nous avons fixé la limite d'une différence acceptable dans un intervalle de confiance calculé sur la base de l'écart type de l'échantillon le plus grand, soit 50 individus par stade et par lignée. Les résultats obtenus en observant 30 individus sont restés dans les limites fixées et ne sont pas significativement différents de ceux obtenus sur un échantillon de 50. La réduction de l'analyse à deux lignées homogènes a été permise par la vérification du comportement similaire des transgènes. Dans le cas des transgènes *uba-1::GFP* et *lin-41::GFP* nous avons aussi vérifié que cela ne modifiait pas les résultats par rapport à un échantillon plus grand (3^{ème} colonne des graphiques). Les résultats obtenus avec l'analyse de deux lignées n'est pas non plus différents de celui obtenu avec 3 lignées.

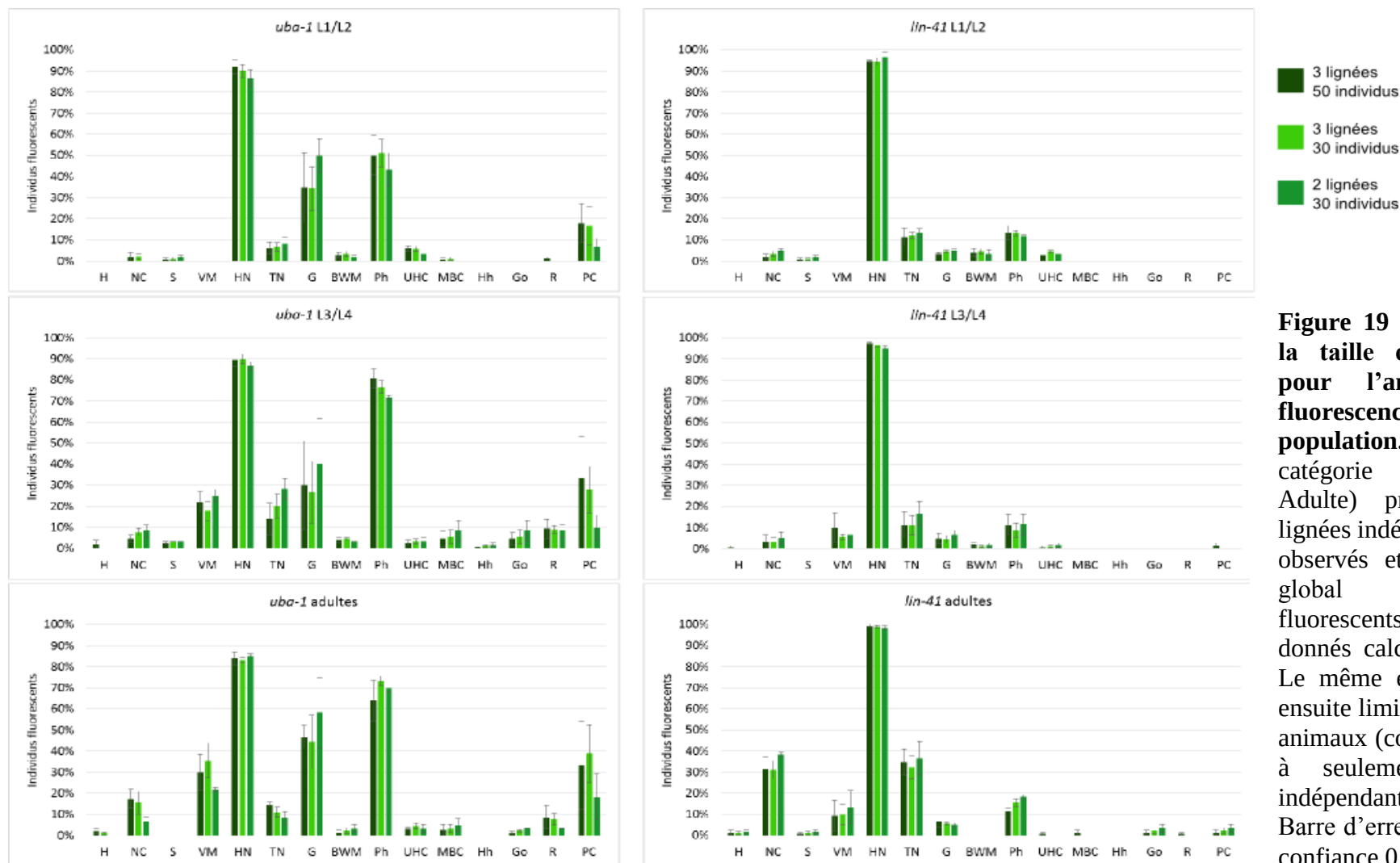


Figure 19 Modification de la taille de l'échantillon pour l'analyse de la fluorescence dans la population. 50 animaux par catégorie (L1/L2, L3/L4, Adulte) provenant de 3 lignées indépendantes ont été observés et le pourcentage global d'individus fluorescents dans les tissus donnés calculé (colonne 1). Le même échantillon a été ensuite limité au 30 premiers animaux (colonne 2) et enfin à seulement 2 lignées indépendantes (colonne 3). Barre d'erreur : intervalle de confiance 0.95

Tissus : H, hypoderme ; NC, corde nerveuse centrale ; S, cellules hypodermiques latérales ; VM, muscle de la vulve ; HN, neurones antérieurs ; TN, neurones postérieurs ; G, intestins ; BWM, muscles du corps ; Ph, pharynx ; UHC, paire de cellules musculaires de la tête ; MBC, corps cellulaires du corps ; Hh, cellules hypodermiques de la tête ; Go, gonades ; R, cellules adjacentes au rectum ; PC : cellules postérieures.

Nos analyses indiquent que l'observation de 30 animaux par stade sur 2 lignées indépendantes est suffisante pour obtenir un résultat similaire à celui obtenu avec trois lignées et 50 individus par stade.

B. Avantage du contrôle interne

L'observation des lignées transgéniques porteuses de la cassette RRPT a permis de mettre en évidence de subtiles variations dans les territoires d'expression des deux protéines de fluorescences qu'il aurait été difficile de détecter en comparant des animaux distincts porteurs de deux rapporteurs correspondant. Etant donné que ces dernières sont sous le contrôle du même promoteur, les résultats attendus étaient que les tissus expriment la mCherry et la GFP conjointement, ou uniquement la mCherry, en cas de régulation post transcriptionnelle ayant lieu sur la région 3'UTR endogène. Or, dans les neurones de plusieurs animaux de différents stades, nous avons constaté que seule la GFP était exprimée. Le phénomène s'est produit avec trois des gènes testés : *lin-28*, *uba-1* et *nhr-17* (Figure 20).

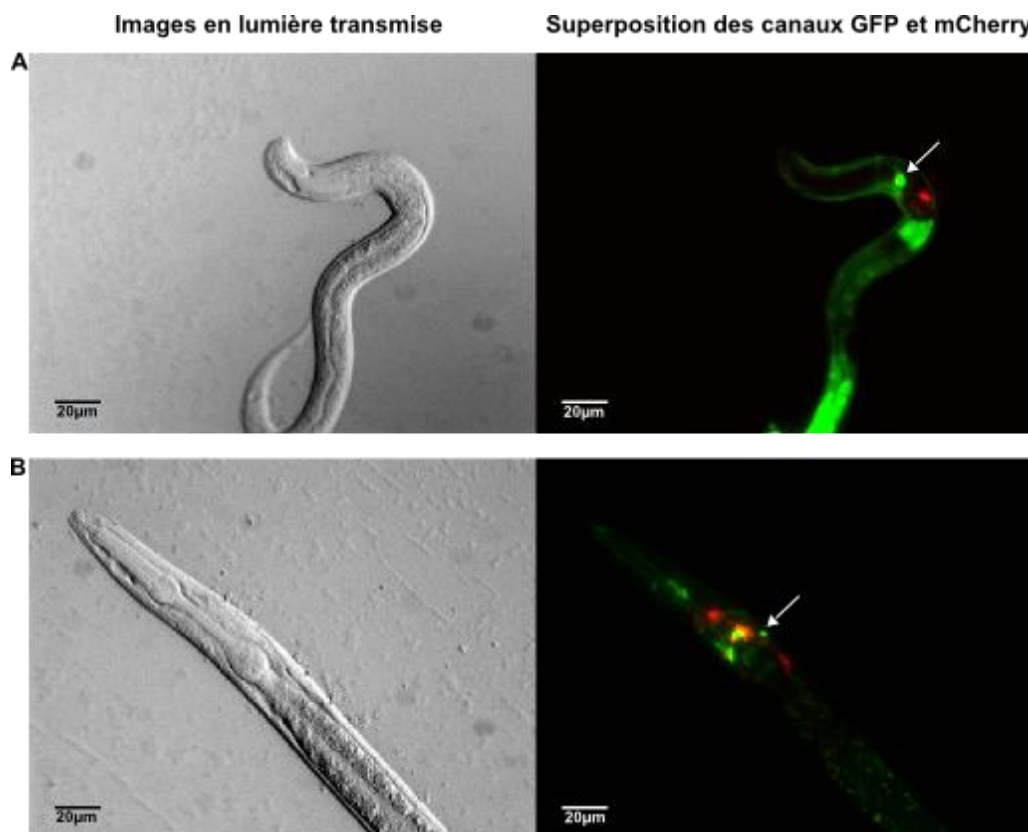


Figure 20 Expression des protéines de fluorescences de la cassette RRPT. Les images ont été prises avec un microscope couplé à une lampe UV (voir matériel et méthode). **A** Animal de stade L2 imagé au 40X exprimant la cassette contenant les éléments régulateurs du gène *uba-1*. **B** Tête d'un animal de stade L4 imagé au 40X exprimant la cassette contenant les éléments régulateurs du gène *nhr-17*.

Ces résultats soulèvent des interrogations sur les différents éléments de régulation présents dans la cassette RRPT qui conduit à l'expression de la GFP dans des neurones n'exprimant pas la

mCherry. Nous avons dans un premier temps comparé la fluorescence observée avec la cassette à celles précédemment décrites pour vérifier si les niveaux d'expression de la GFP et de la mCherry étaient en accord avec les études précédentes (Lall et al. 2006).

C. Comparaison des fluorescences

Les résultats obtenus ont été comparés avec ceux publiés en 2006 par l'équipe de Fabio Piano, pour les 3 gènes présents dans son étude (menée sur 10 individus par stade, sur un minimum de 3 lignées). Les résultats ont été mis en forme de la même manière afin d'en simplifier la lecture, et sont présentés Figure 21. De légères modifications ont été apportées au tableau original dès le début de nos observations : les colonnes comptabilisant les neurones ou paires de neurones antérieurs (HN et UAC) ont été rassemblées en une seule (HN) ; celles décrivant la fluorescence dans la corde nerveuse ventrale et ses neurones (VNC et CN) sont regroupées sous la colonne NC ; la colonne Go est ajoutée pour la fluorescence dans les gonades (présente dans les lignées transgéniques ayant le promoteur *lin-28*, non incluses dans l'étude de 2006).

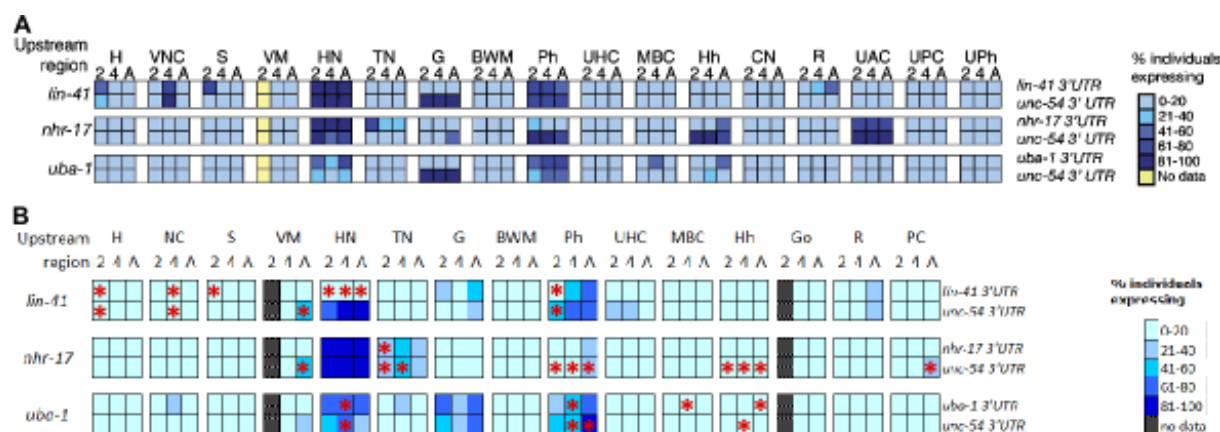


Figure 21 Comparaison des résultats obtenus à ceux publiés en 2006. A Extrait du tableau publié dans l'étude de (Lall et al. 2006) **B** Résultats de l'analyse des lignées transgéniques générées au laboratoire. Les * indiquent les différences d'expression par rapport à l'étude de 2006. **Tissus** : H, hypoderme ; NC, corde nerveuse centrale ; S, cellules hypodermiques latérales ; VM, muscle de la vulve ; HN, neurones antérieurs ; TN, neurones postérieurs ; G, intestins ; BWM, muscles du corps ; Ph, pharynx ; UHC, paire de cellules musculaires de la tête ; MBC, corps cellulaires du corps ; Hh, cellules hypodermiques de la tête ; Go, gonades ; R, cellules adjacentes au rectum ; PC : cellules postérieures.

La différence observée dans la fluorescence des intestins peut provenir d'une différence d'interprétation liée à l'autofluorescence des granules intestinaux et n'est donc pas nécessairement à considérer comme significative. Concernant le gène *lin-41*, le niveau d'expression de la GFP de la cassette n'a pas donné le même niveau d'expression que celui observé dans l'étude de 2006. La GFP de la cassette était moins exprimée au niveau des cellules nerveuses (HN et NC stade L4) et du pharynx où elle était attendue. La mCherry a donné des niveaux d'expression similaires à ceux du rapporteur contrôle de l'étude. Pour le gène *nhr-17*

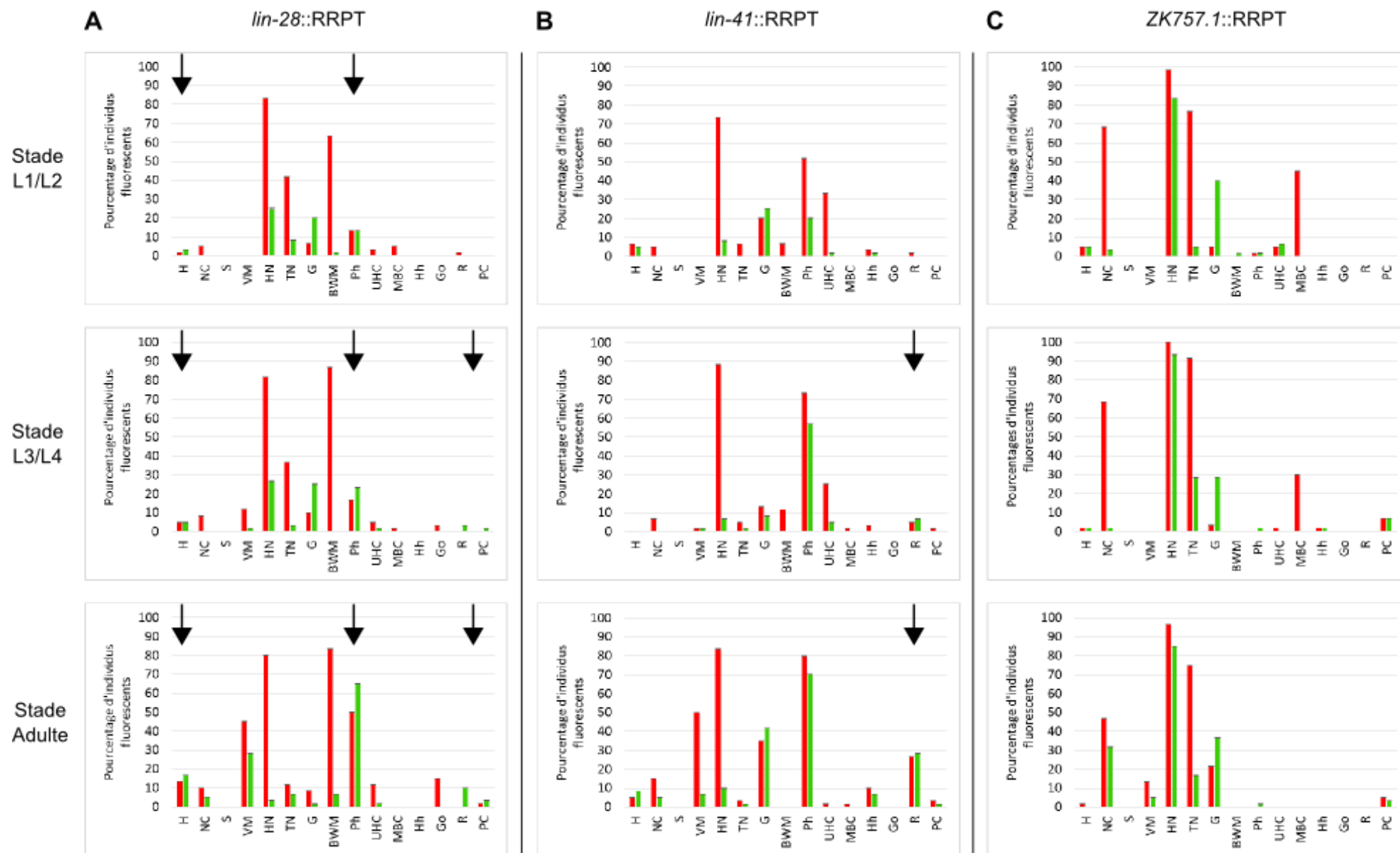
la différence d'expression avec le rapporteur suivi par la région *nhr-17* 3'UTR concerne uniquement les neurones postérieurs. D'autres différences sont observées sur l'expression du rapporteur suivi par *unc-54* 3'UTR, ce qui indique une activité différente du promoteur. Pour *uba-1* nous n'avons pas observé de variation dans l'expression des neurones et avons fait le même constat que les auteurs sur ce gène : le rapporteur suivi de la région 3'UTR endogène est plus exprimé que le contrôle.

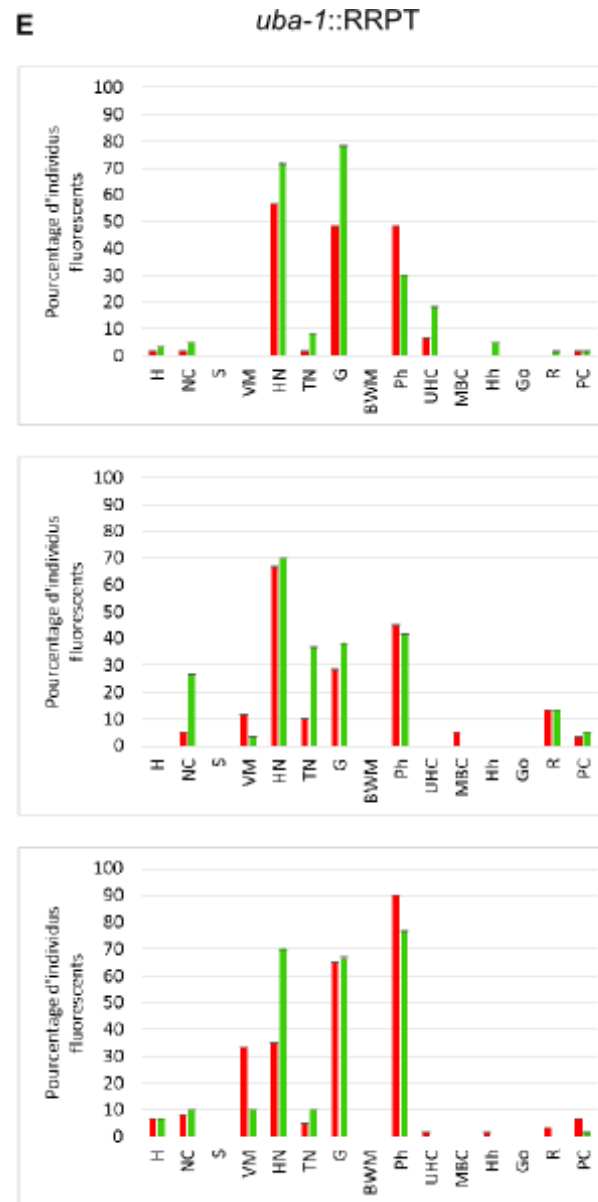
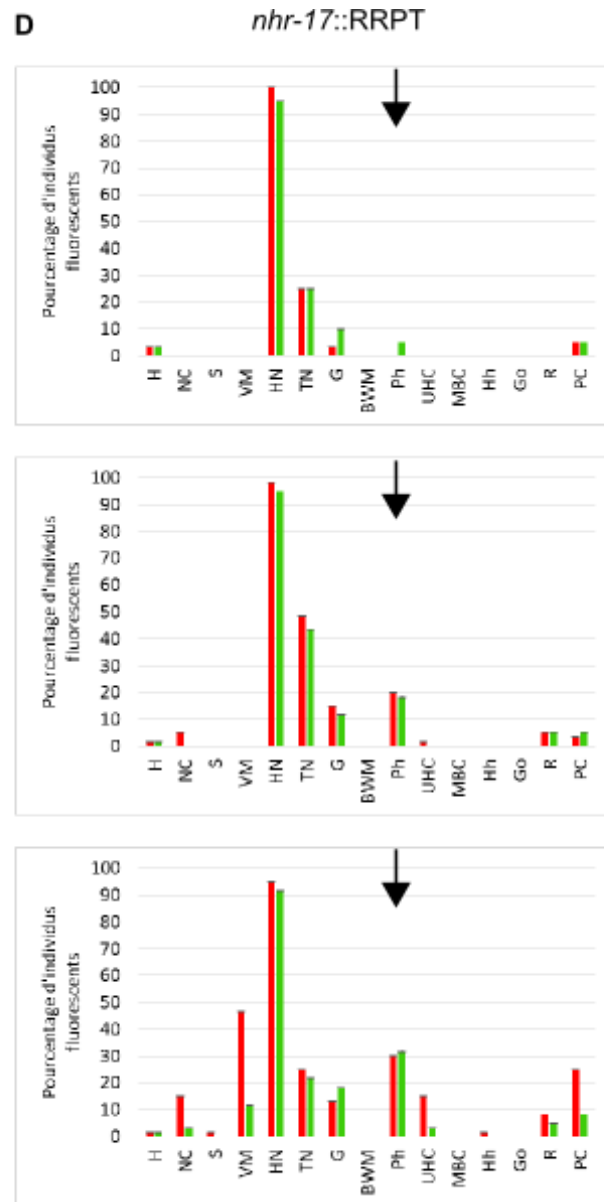
Compte tenu des différences constatées avec l'étude originale, notamment pour le gène *lin-41*, nous avons décidé de comparer la fluorescence émise par la GFP de la cassette RRPT à celle émise par la GFP des lignées contrôles. Les lignées contrôles portent la GFP directement clonée en aval du promoteur et permettent de s'assurer que les éléments de la cassette n'interfèrent pas avec l'expression de la GFP. La GFP de la cassette RRPT est située immédiatement après l'*operon linker* ce qui pourrait entraîner des défaut d'expression de celle-ci. L'*operon linker* pourrait induire une surexpression s'il agit comme un promoteur, et inversement, une diminution de l'expression si le système d'opéron n'est pas convenablement interprété par le système de transcription.

Les lignées contenant les cassettes RRPT pour les gènes étudiés, présentent globalement une expression de la mCherry égale ou supérieure à celle de la GFP, ce qui était attendu (Figure 22). Plusieurs fois, l'expression de la GFP s'est avérée plus forte que celle de la mCherry. C'est le cas dans les intestins pour plusieurs lignées, mais comme précisé plus haut, cette fluorescence peut être liée à un phénomène d'autofluorescence de la zone. Dans le cas du gène *uba-1* la GFP est exprimée de manière égale ou supérieure à la mCherry, notamment dans certains neurones antérieurs (Figure 22E). Le gène *uba-1* était prédit comme cible de *let-7* par une version précédente moins stringente de Pictar. L'étude de 2006 n'avait pas non plus mis en évidence de régulation post-transcriptionnelle de ce gène et avait fait état du même phénomène (Lall et al. 2006). La version actuelle de Pictar ne prédit aucune autre régulation par des miARN, bien que TargetScan indique plusieurs sites cibles potentiels. TargetScan indique toujours le site de régulation pour les miARN *let-7*, *miR-48*, *miR-84*, *miR-241*, *miR-795* ainsi que des sites non conservés entre les espèces (*miR-85*, *miR-4813*, *miR-85*, *miR-70*, *miR-236*, *miR-60-3p*, *miR-49/83*) ou pour des familles de miARN non conservées (*miR-1818*, *miR-792*, *miR-1018*, *miR-2209*, *miR-1829*, *miR-1832*, *miR-4809*, *miR-243*, *miR-1821*, *miR-794*, *miR-793*, *miR-356*, *miR-1020*). Le phénomène de l'expression de la GFP surpassant celle de la mCherry s'est reproduit avec les autres transgènes. Pour *lin-28* le phénomène est observé à tous les stades dans le pharynx et l'hypoderme, et dans les cellules postérieures aux stades L3/L4 et adulte (Figure 22A). Les gènes *lin-41* et *nhr-17* ont présenté ce profil d'expression où la GFP est plus exprimée

que la mCherry, respectivement dans les cellules adjacentes au rectum et le pharynx (Figure 22B et D)

La comparaison de la fluorescence des GFP entre les lignées porteuses des cassettes RRPT et les lignées ayant un plasmide simple avec la GFP mettent en évidence un biais d'expression, puisque pour 4 gènes sur 5, la GFP injectée seule est bien plus exprimée que la GFP contenue dans la cassette, y compris pour le gène *uba-1* (Figure 23). Dans le cas du gène *nhr-17*, la différence est plus ténue mais visible dans la corde ventrale et les cellules hypodermiques latérales, à tous les stades du développement de l'animal.

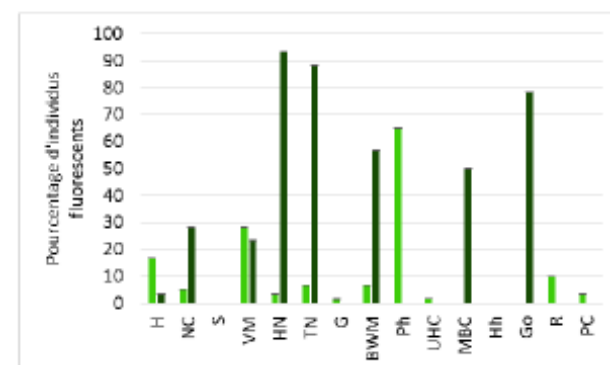
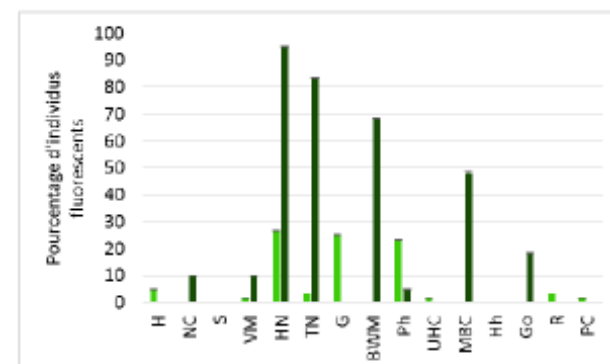
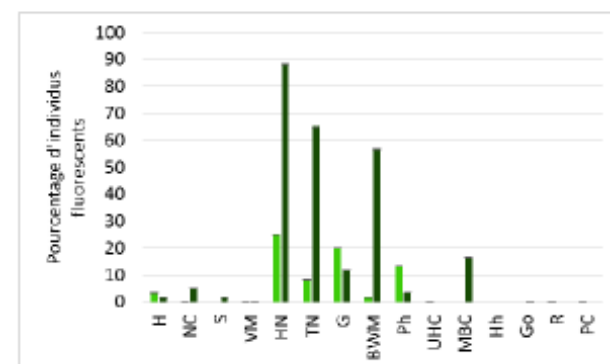
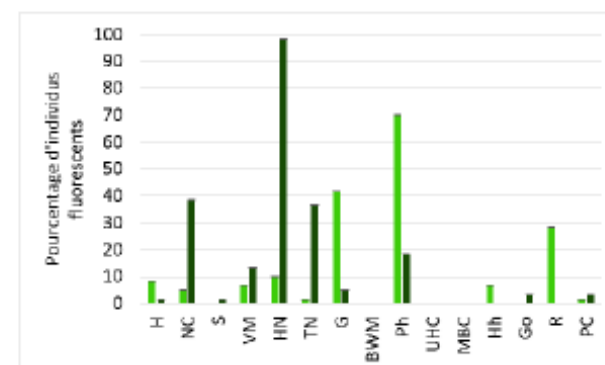
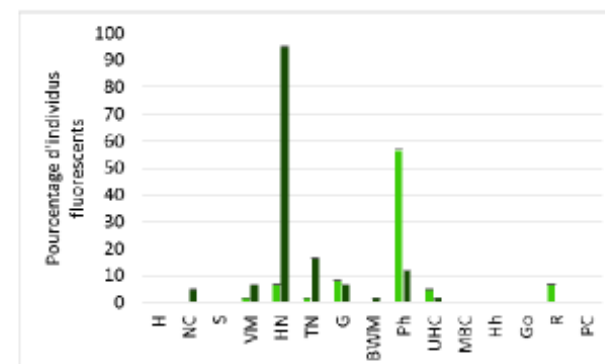
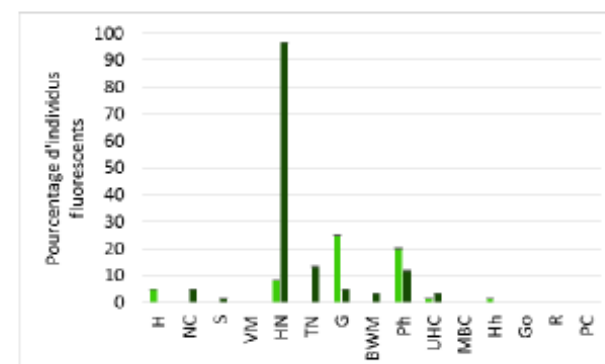
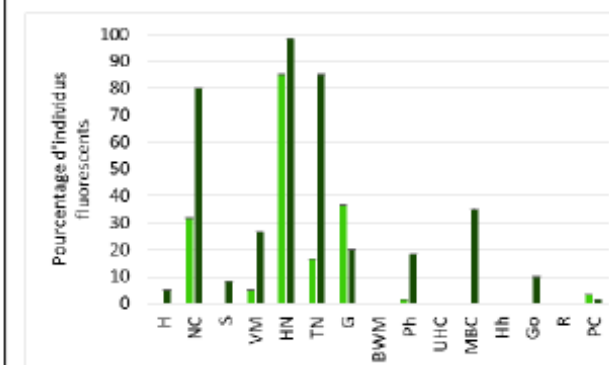
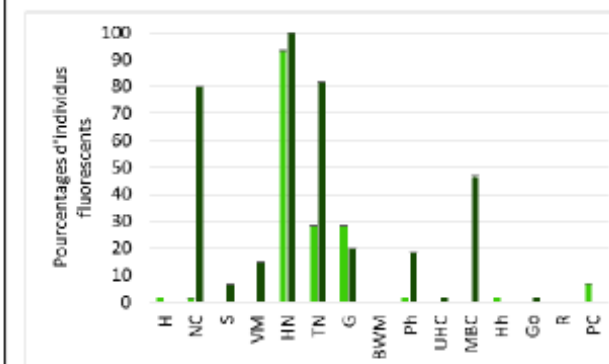
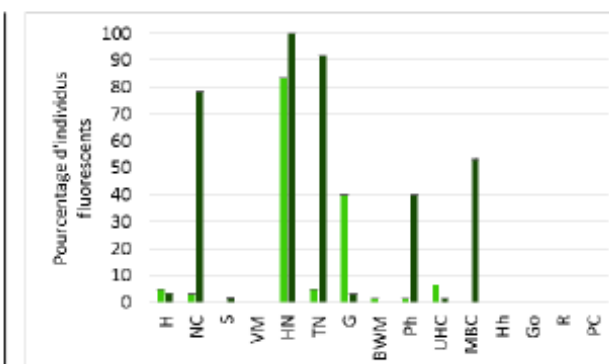




■ mCherry
de la cassette RRPT

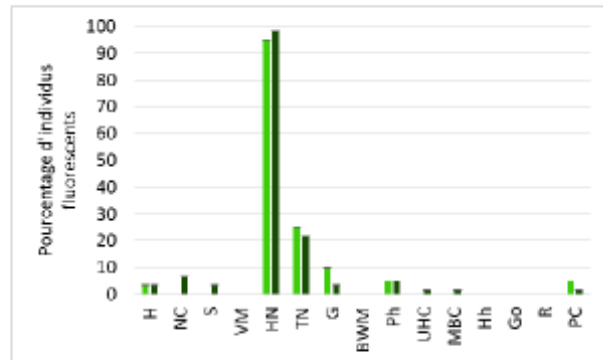
■ GFP
de la cassette RRPT

Figure 22 Comparaison de l'expression de la GFP et de la mCherry dans les lignées porteuses de la cassette RRPT. La fluorescence est exprimée en pourcentage d'individus exprimant les fluorochromes. Les flèches indiquent les zones où la GFP est plus exprimée que la mCherry. **A** Expression des différents transgènes *lin-28*. **B** Expression des différents transgènes *lin-41*. **C** Expression des différents transgènes ZK757.1. **D** Expression des différents transgènes *nhr-17*. **E** Expression des différents transgènes *uba-1*. **Tissus** : H, hypoderme ; NC, corde nerveuse centrale ; S, cellules hypodermiques latérales ; VM, muscle de la vulve ; HN, neurones antérieurs ; TN, neurones postérieurs ; G, intestins ; BWM, muscles du corps ; Ph, pharynx ; UHC, paire de cellules musculaires de la tête ; MBC, corps cellulaires du corps ; Hh, cellules hypodermiques de la tête ; Go, gonades ; R, cellules adjacentes au rectum ; PC : cellules postérieures.

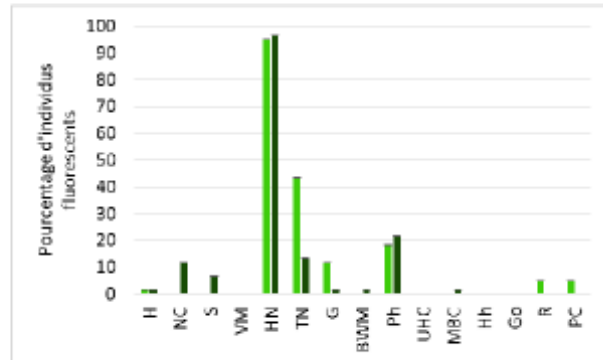
A Transgènes *lin-28***B** Transgènes *lin-41***C** Transgènes ZK757.1

D Transgènes *nhr-17*

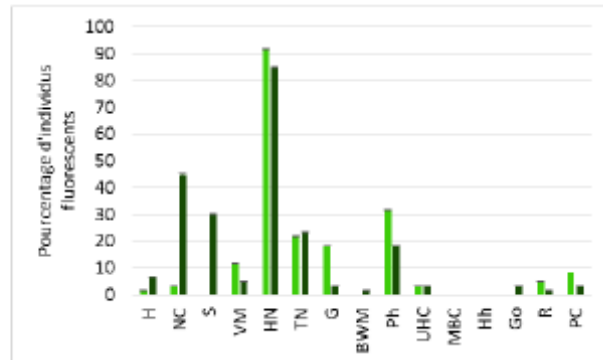
Stade
L1/L2



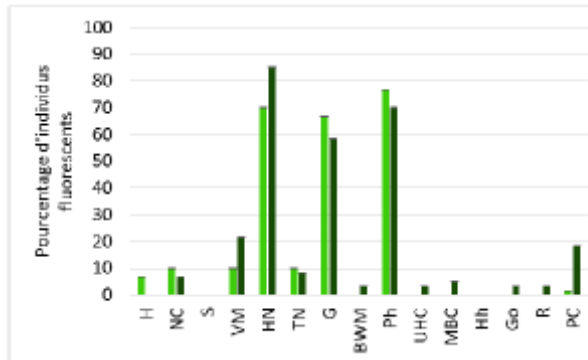
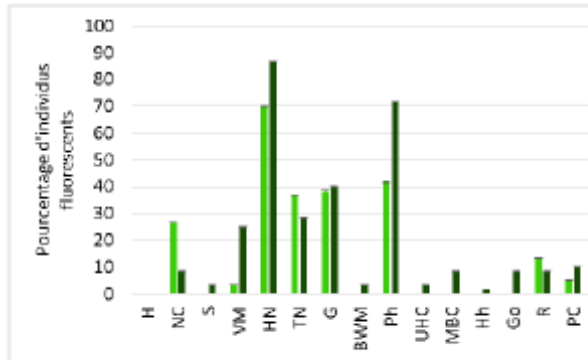
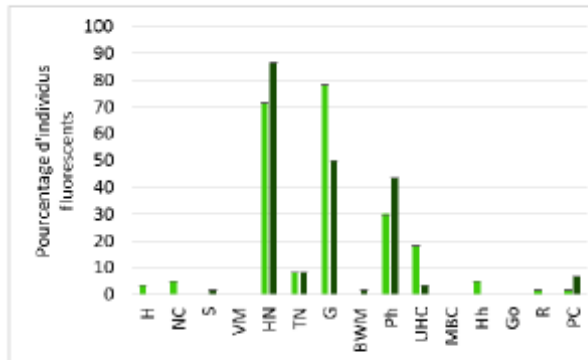
Stade
L3/L4



Stade
Adulte



E Transgènes *uba-1*



■ GFP
de la cassette RRPT

■ GFP
du vecteur contrôle
injecté seul

Figure 23 Comparaison de l'expression de la GFP selon les lignées, contenant la cassette RRPT ou le vecteur contrôle. La fluorescence est exprimée en pourcentage d'individus exprimant la GFP. **A** Expression des différents transgènes *lin-28*. **B** Expression des différents transgènes *lin-41*. **C** Expression des différents transgènes *ZK757.1*. **D** Expression des différents transgènes *nhr-17*. **E** Expression des différents transgènes *uba-1*. **Tissus** : H, hypoderme ; NC, corde nerveuse centrale ; S, cellules hypodermiques latérales ; VM, muscle de la vulve ; HN, neurones antérieurs ; TN, neurones postérieurs ; G, intestins ; BW, muscles du corps ; Ph, pharynx ; UHC, paire de cellules musculaires de la tête ; MBC, corps cellulaires du corps ; Hh, cellules hypodermiques de la tête ; Go, gonades ; R, cellules adjacentes au rectum ; PC : cellules postérieures.

L'ensemble de ces résultats nous ont amené à formuler plusieurs hypothèses sur les mécanismes qui pourraient interférer avec le fonctionnement théorique de la cassette. Premièrement, la différence entre les GFP suggère que l'*operon linker* pourrait interférer dans l'expression correcte de la GFP dans certains tissus. Deuxièmement, la différence d'expression des protéines pourrait être liée à la durée de demi-vie des protéines qui ne permettrait pas la disparition suffisamment rapide de la GFP. Troisièmement, bien qu'aucune régulation ne soit prédite par Pictar ou connue, la région *unc-54* 3'UTR pourrait être soumise à une régulation post-transcriptionnelle. En effet, celle-ci contient des sites potentiellement cibles de miARNs indiqués par Target Scan (*miR-249*, *miR-71*, *miR-85*, *miR-230*, *miR-60-3p*, *miR-2953*, *miR-1829*, *miR-4811*, *miR-1820*, *miR-1020*, *miR-4805*, *miR-2208ab*, *miR-1828*) et non pris en compte par l'algorithme Pictar car peu conservés entre les espèces de nématodes.

3. Vérification des différents éléments de la construction

A. Suppression de l'opéron linker : co-injection des rapporteurs

Une approche de co-injection de deux plasmides a été testée, l'un des plasmides porte la mCherry sous le contrôle du promoteur du gène d'intérêt et de la région *unc-54* 3'UTR, l'autre la GFP sous le contrôle du promoteur et de la région 3'UTR du gène d'intérêt. Le plasmide porteur de la GFP est donc le même que celui utilisé comme contrôle dans la première partie de ce travail.

Cette approche a été testée sur les gènes *lin-28* et T10B11.6 (Figure 24).



Figure 24 Microscopie d'animaux porteurs de plasmides indépendants exprimant deux fluorophores différents. **A** Animal de stade L2 imagé au 40X exprimant les rapporteurs du gène *T10B11.6* sur des plasmides indépendants. La flèche indique les neurones n'exprimant que la GFP. **B** Animal de stade adulte imagé au 10X exprimant les rapporteurs du gène *lin-28*. Les flèches indiquent les neurones n'exprimant que la GFP.

La suppression de l'*operon linker* a augmenté l'expression de la GFP par rapport à celle exprimée par la cassette. Cependant, elle n'a pas eu d'impact sur l'expression de la mCherry, restée au même niveau que celui de la cassette. De plus, les neurones initialement observés continuent à exprimer la GFP mais pas la mCherry.

Si l'*operon linker* avait provoqué une fuite de transcription permettant l'expression de la GFP indépendamment de la mCherry, sa suppression aurait dû corriger le problème, les deux fluorophores étant sous le contrôle direct d'un promoteur identique. Les territoires d'expression des deux protéines de fluorescence infirment donc cette hypothèse. Une partie des cellules expriment les deux rapporteurs conjointement, certaines n'expriment que la mCherry et d'autres n'expriment que la GFP. Il semblerait donc que ce résultat indique deux mécanismes de régulation post-transcriptionnelle, l'un s'opérant sur la région 3'UTR d'intérêt comme attendu et l'autre concernant *unc-54* 3'UTR. Une autre possibilité est qu'il existe une différence dans la demi-vie des protéines fluorescentes et que l'on observe des protéines produites plus tôt au cours du développement qui n'ont pas été dégradées.

B. Les protéines fluorescentes présentent le même profil d'expression placées sous le contrôle des mêmes séquences régulatrices

Deux plasmides contenant respectivement la GFP et la mCherry ont été construits avec comme promoteur et 3'UTR les régions du gène *lin-28*. La co-injection de ces plasmides a permis de restaurer la colocalisation parfaite des protéines fluorescentes, particulièrement dans les neurones ayant montré précédemment des divergences d'expression entre la GFP et la mCherry (Figure 25).

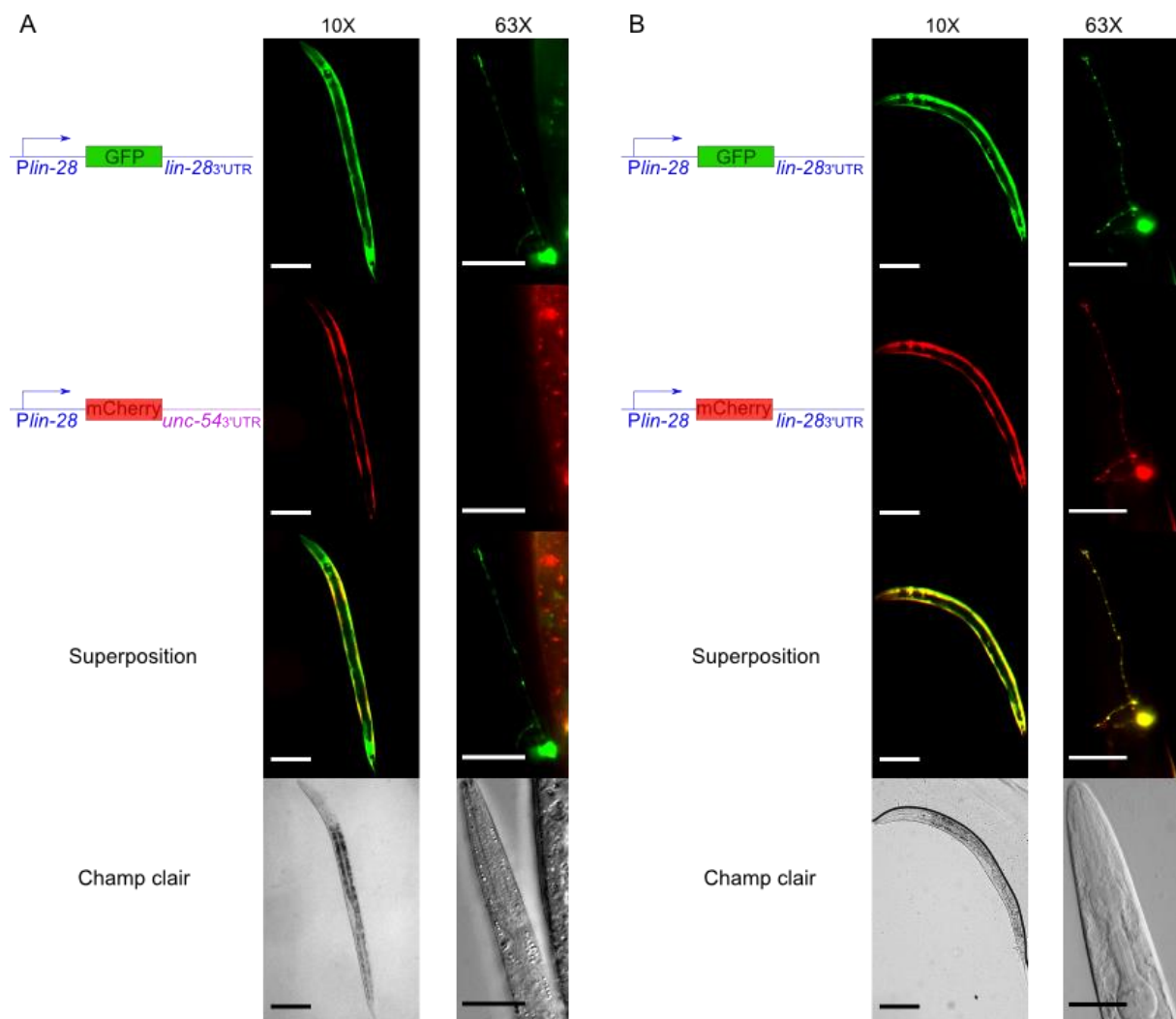


Figure 25 **A** Images de microscopie d'un vers entier (10X, barre d'échelle 100 μ m) et des neurones de la tête (63X, barre d'échelle 25 μ m) d'une lignée portant les transgènes *Plin-28::GFP::lin-28* 3'UTR et *Plin-28::mCherry::unc-54* 3'UTR. **B** Images de microscopie d'un vers entier (10X, barre d'échelle 100 μ m) et des neurones de la tête (63X, barre d'échelle 25 μ m) d'une lignée portant les transgènes *Plin-28::GFP::lin-28* 3'UTR et *Plin-28::mCherry::lin-28* 3'UTR.

Les lignées obtenues présentent un profil d'expression des deux protéines de fluorescence parfait. Ce phénotype visible a été confirmé par l'analyse en microscopie. La Figure 26 montre l'expression des protéines dans les différents tissus.

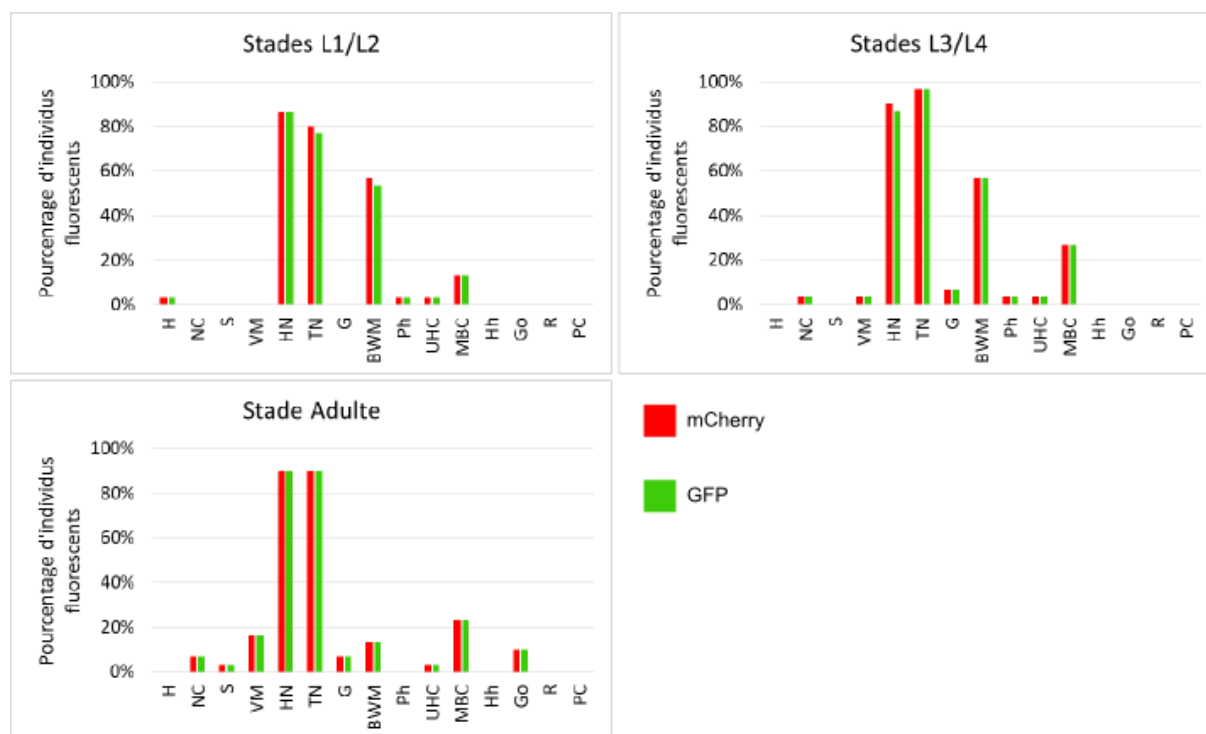


Figure 26 30 individus ont été observés par stade. **Tissus** : H, hypoderme ; NC, corde nerveuse centrale ; S, cellules hypodermiques latérales ; VM, muscle de la vulve ; HN, neurones antérieurs ; TN, neurones postérieurs ; G, intestins ; BWM, muscles du corps ; Ph, pharynx ; UHC, paire de cellules musculaires de la tête ; MBC, corps cellulaires du corps ; Hh, cellules hypodermiques de la tête ; Go, gonades ; R, cellules adjacentes au rectum ; PC : cellules postérieures.

Cette expérience confirme que placées sous les mêmes éléments de régulation, les deux protéines mCherry et GFP présentent un profil d'expression identique. Cela signifie donc que les profils observés sont liés à l'intervention de deux mécanismes de régulation post-transcriptionnelle opérant sur les régions 3'UTR. Dans les cellules exprimant les deux protéines, aucune régulation n'a lieu. Celles exprimant uniquement la mCherry soulignent l'évènement de régulation lié à la région 3'UTR du gène *lin-28*. Inversement, les cellules n'exprimant que la GFP mettent en évidence un évènement de régulation sur la région 3'UTR du gène *unc-54*.

2. Caractérisation de la régulation post-transcriptionnelle du gène *unc-54*

1. Identification des neurones présentant une inhibition de la séquence *unc-54* 3'UTR

A. Détermination de la morphologie des neurones

Le Wormatlas (Altun et al.) regroupe l'ensemble des informations disponibles sur l'anatomie de *C. elegans* et les 302 neurones de l'animal y sont schématiquement représentés selon leur emplacement précis. Nous avons utilisé ces représentations pour déterminer quels neurones sont marqués par la GFP sous le contrôle du promoteur du gène *lin-28* mais n'expriment pas la mCherry sous le contrôle de la région 3'UTR du gène *unc-54*.

Image de microscopie :



Neurones candidats :

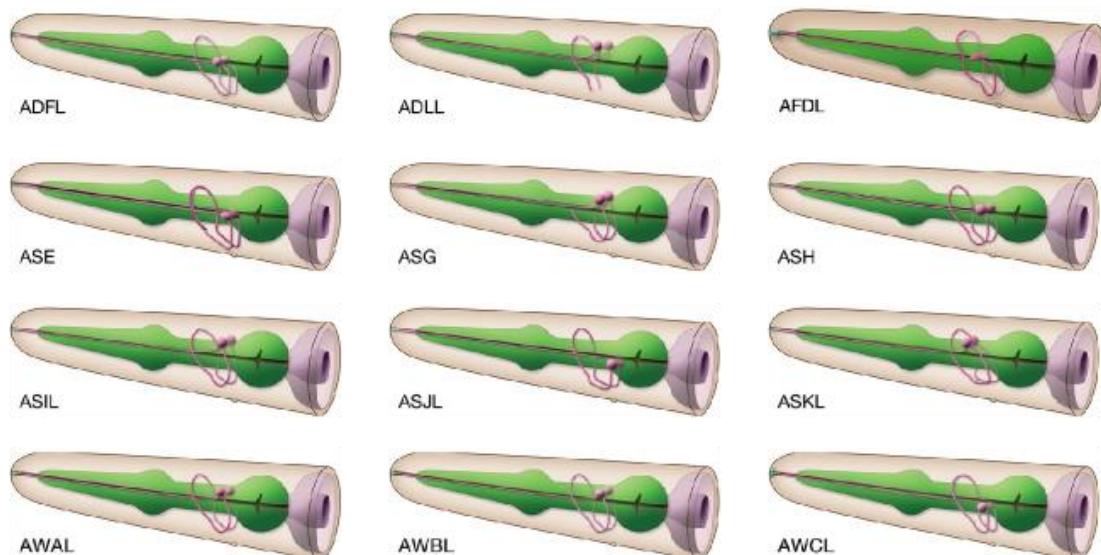


Figure 27 Identification des neurones exprimant la GFP sous le contrôle du promoteur *lin-28* et où *unc-54* 3'UTR est régulé. **A** Superposition des canaux GFP et champ clair, et GFP seule afin de localiser l'emplacement des neurones. **B** Vue au 63X de la paire de neurones, l'animal étant sur le ventre.

De plus, l'extrémité des dendrites a pu être visualisée par microscopie et permet ainsi d'identifier les neurones (Figure 28B). L'extrémité de la dendrite se termine en fourche bifide, ce qui permet de réduire le nombre de candidats à trois neurones : AWB, ADF et ADL (Figure 28).

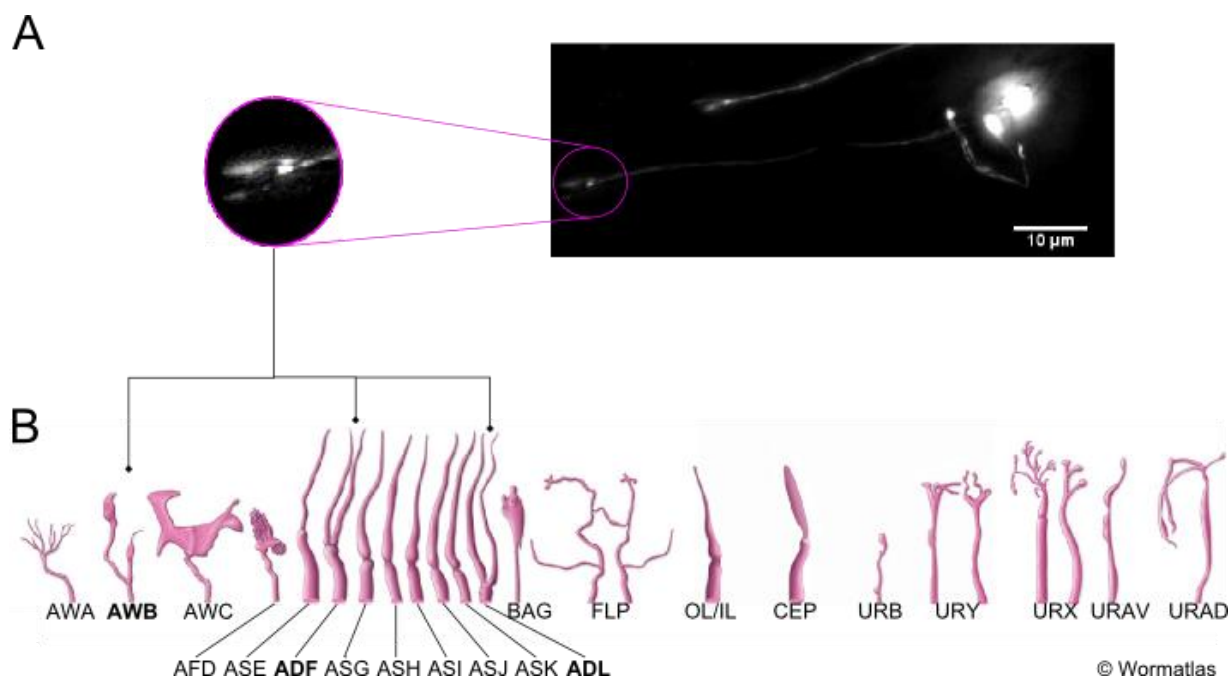


Figure 28 Identification des neurones présentant une régulation de la région *unc-54* 3'UTR. **A** Image en fluorescence de la tête d'un ver transgénique portant le transgène *Plin-28::GFP::lin-28* 3'UTR. La GFP est exprimée dans une paire de neurones de la tête. L'extrémité de la dendrite se divise en deux parties (zoom). **B** Les différentes formes des extrémités dendritiques des neurones présents dans la tête de *C. elegans*, adapté du Wormatlas. La forme de l'extrémité de la dendrite permet de restreindre les candidats à 3 neurones.

B. Confirmation de l'identification par coloration au DiI

Pour identifier précisément la paire de neurones où se déroule la régulation post-transcriptionnelle de *unc-54* 3'UTR, une coloration au DiI a été réalisée. Le DiI permet de marquer chez *C. elegans* les neurones amphides ADL, ASH, ASI, ASJ, ASK, AWB, phasmides PHA et PHB ainsi que les neurones IL1 et IL2. Parmi les candidats identifiés, seuls les paires AWB et ADL sont marquées par ce colorant. Les résultats possibles étaient donc un double marquage rouge (DiI) et vert (GFP du transgène) pour AWB et ADL, ou un simple marquage vert pour ADF.

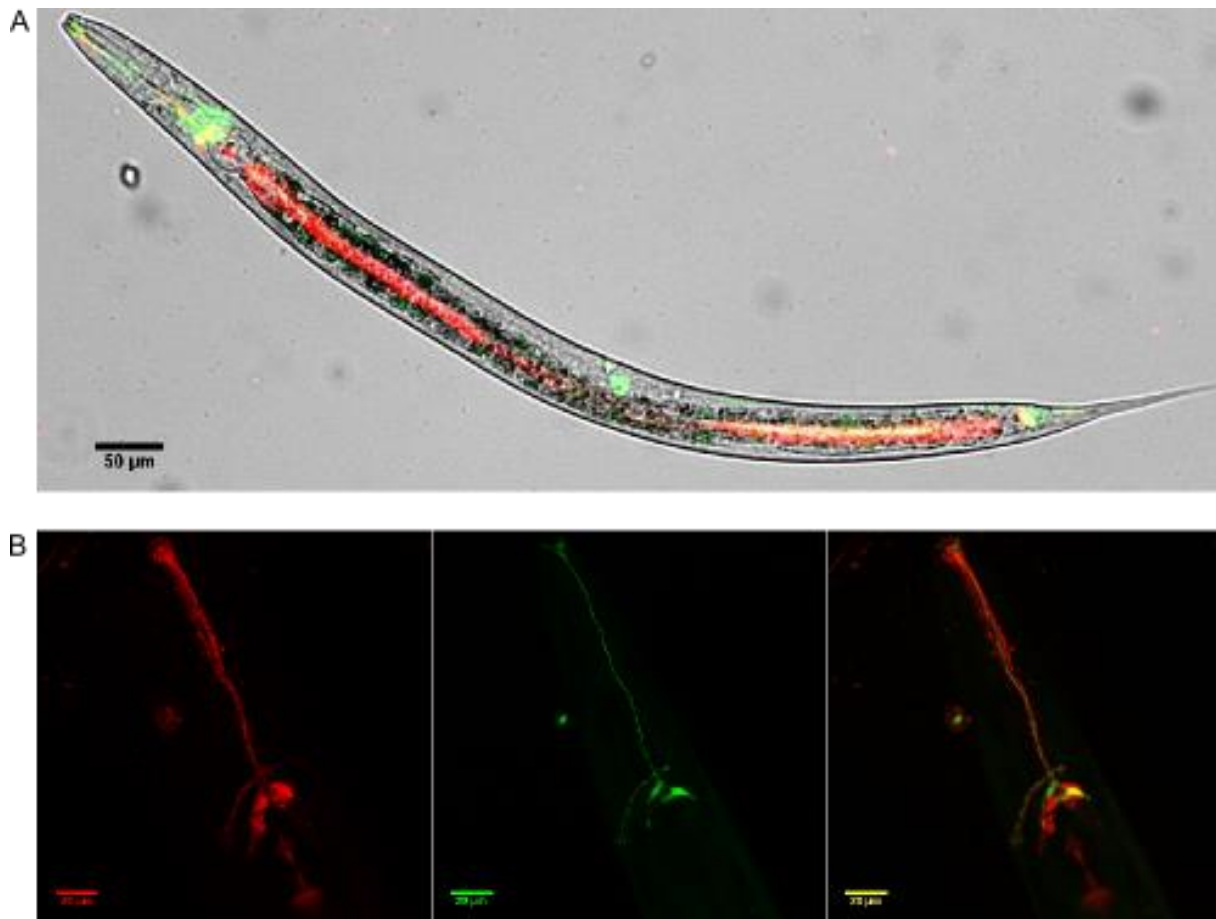


Figure 29 Marquage au DiI des neurones de la lignée transgénique portant le transgène *Plin-28::GFP::lin-28* 3'UTR *Plin-28::mCherry::unc-54* 3'UTR. A Superposition des canaux vert, rouge et champ clair sur animal entier de stade L4 marqué au DiI (rouge). B Images de microscopie confocale mettant en évidence l'absence de coloration au DiI des neurones exprimant la GFP seule (barre d'échelle 20µm). La couleur jaune sur l'image de droite est liée à la projection sur un même plan de l'ensemble des fluorescences, le second neurone de la paire apparaissant en décalé est d'ailleurs parfaitement vert. L'identification des neurones ADF est confirmée.

La forme de la dendrite ainsi que la coloration au DiI ont permis de déterminer avec certitude que les neurones dans lesquels la séquence *unc-54* 3'UTR subit une régulation post-transcriptionnelle sont ceux de la paire ADF. Les neurones ADF sont les seuls neurones sérotoninergiques sensoriels présents chez *C. elegans*. Ils associeraient d'ailleurs la neurotransmission de sérotonine à la détection de nourriture dans l'environnement (Jafari et al. 2011).

2. Mutagenèse dirigée sur la séquence *unc-54* 3'UTR

Les résultats précédents semblent clairement indiquer une régulation post-transcriptionnelle tissu-spécifique de *unc-54* 3'UTR. Afin de déterminer avec précision quelle partie de la région est responsable de cette régulation, des expériences de mutagenèse dirigée ont été menées.

A. Analyse *in silico* de la séquence et choix des sites cibles

L'analyse de la séquence avec le logiciel TargetScan a permis de mettre en évidence que la région 3'UTR contient des sites potentiels de fixation des miARNs (Figure 30).

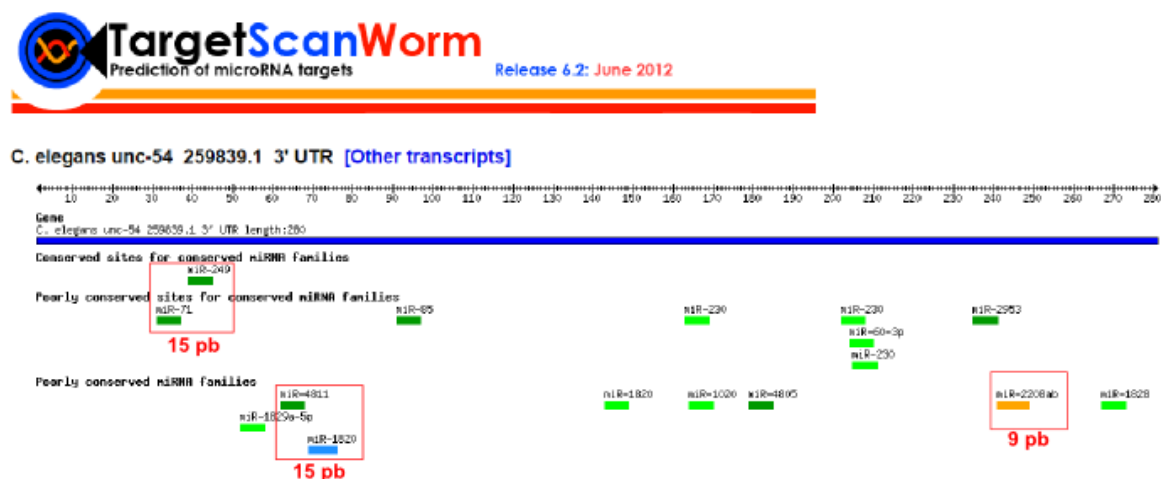


Figure 30 Capture d'écran des résultats de l'analyse avec TargetScan. En rouge, les sites choisis pour la mutagenèse dirigée. Les sites ayant une barre vert foncé indiquent une correspondance exacte des positions 2-7 du miARN mature (site graine), vert clair une correspondance des positions 2-6 suivie par un A. La barre bleue claire indique une correspondance exacte des positions 2-7 du miARN mature (site graine) suivie par un A et la barre orange indique une correspondance exacte des positions 2-8 du miARN mature (site graine et position 8) suivie par un U.

Trois sites ont été choisis pour des délétions spécifiques. La première sélection des candidats a été basée sur leur conservation entre les espèces *C. elegans*, *C. remanei* et *C. briggsae*, nous avons donc éliminé les sites représentés en vert clair sur la Figure 30. Le site 1 est une délétion de 15 paires de bases contenant les sites cibles des miARNs *mir-71* et *mir-249*. Le site *mir-249* étant conservé au sein des différentes familles de nématodes constitue un bon candidat potentiel, même si, à ce jour, aucune donnée sur le profil d'expression de ce miARN n'existe. Le *mir-71* est quant à lui exprimé tout au long du développement, même si sa fonction reste inconnue. Le site 2 est une délétion de 15 paires de bases contenant les sites cibles des miARNs *mir-1820* et *mir-4811* pour lesquels aucune donnée d'expression n'est disponible. Le dernier site candidat, le site 3, ne contient qu'un seul site cible pour le *mir-2208ab*. Ce site est long de 8 pb ce qui en faisait un bon candidat, malgré l'absence de donnée d'expression concernant ce miARN. Dans le cas où ces délétions ne donneraient pas de résultats, les autres sites seraient aussi testés.

B. Résultats des délétions ciblées

La suppression des sites 1, 2 et 3, respectivement ci-après $\Delta S1$, $\Delta S2$ et $\Delta S3$, a été réalisée sur le plasmide *Plin-28::mCherry::unc-54 3'UTR* par PCR (Mutagenèse dirigée). Les délétions ont

toutes été confirmées par séquençage. Des animaux N2 ont alors été injectés avec les différents plasmides, toujours avec *Plin-28::GFP::lin-28 3'UTR* comme contrôle interne (Figure 31).

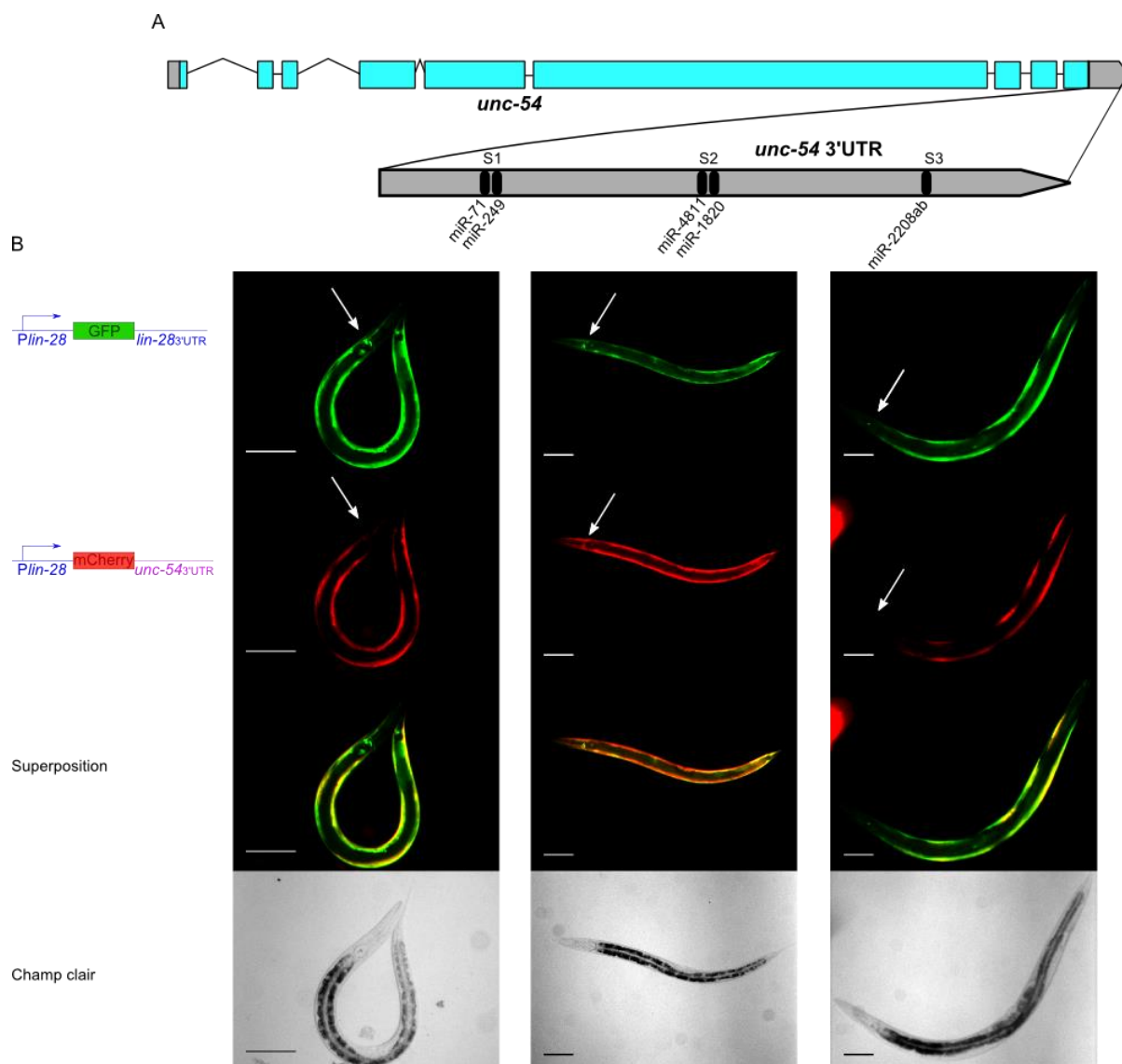


Figure 31 Mutagenèse dirigée sur la séquence *unc-54 3'UTR*. **A** Représentation du gène *unc-54* et de la région 3'UTR avec les sites S1, S2 et S3 délétés spécifiquement. **B** Animaux de stade L4 issus des lignées exprimant *Plin-28::GFP::lin-28 3'UTR* et *Plin-28::mCherry::unc-54 3'UTR* porteurs des délétions $\Delta S1$, $\Delta S2$ et $\Delta S3$ respectivement. Les mutations $\Delta S1$ et $\Delta S3$ ne modifient pas le phénotype alors que $\Delta S2$ entraîne l'expression de la mCherry dans les neurones ADF (flèches) et l'augmente également dans le muscle.

Seule la délétion $\Delta S2$ a entraîné une variation phénotypique observable. En effet, la mutation a permis d'obtenir l'expression de la mCherry dans les neurones ADF, mais aussi de l'augmenter dans le muscle.

La superposition des canaux rouge (mCherry) et vert (GFP) a révélé un phénotype d'expression des rapporteurs variable au cours du temps. Aux stades L1 et L2, l'animal semble vert dans les muscles bien qu'il exprime les deux protéines de fluorescence. Aux stades L3 et L4, l'animal devient jaune ce qui signifie que les 2 rapporteurs sont exprimés de manière comparable. Enfin,

au stade adulte, l'animal est globalement rouge, même si les neurones conservent une fluorescence jaune (Figure 32). Ces résultats ont été vérifiés en procédant aux injections avec inversion des rapporteurs. Deux lignées indépendantes exprimant les transgènes *Plin-28::mCherry::lin-28* 3'UTR et *Plin-28::GFP::unc-54* 3'UTR Δ S2 ont été générées et donnent le même profil d'expression, avec un changement dans les ratios d'expression de la GFP et la mCherry stade spécifique (Figure 32). L'augmentation de l'expression des protéines fluorescentes dans les muscles semblent indiquer que la région *unc-54* 3'UTR serait soumise à une régulation post-transcriptionnelle dans les muscles.

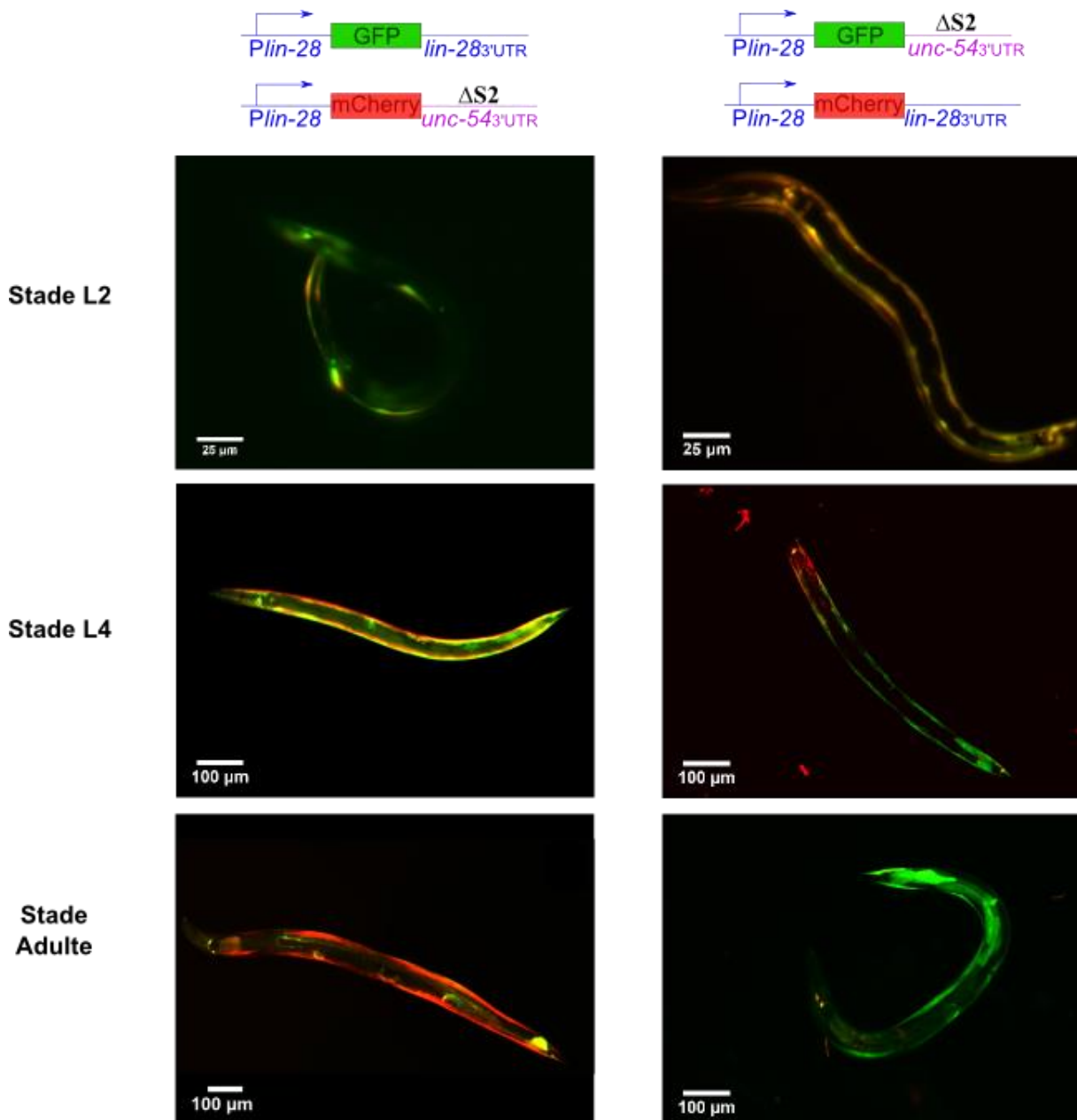


Figure 32 A gauche une lignée portant *Plin-28::mCherry::unc-54* 3'UTR Δ S2 et *Plin-28::GFP::lin-28* ; à droite une lignée portant *Plin-28::GFP::unc-54* 3'UTR Δ S2 et *Plin-28::mCherry::lin-28*. Superposition d'images de microscopie acquises au 40X pour le stade L2, 10X pour les stades ultérieurs, dans les canaux GFP et RFP. Les neurones expriment les deux fluorochromes à tous les stades dans les neurones ADF.

Dans les deux cas, les observations mettent en évidence que l'expression dans les neurones antérieurs ADF est constante lorsque la région *unc-54* 3'UTR est mutée par la délétion du site 2. Les neurones conservent une double fluorescence constante au cours du temps.

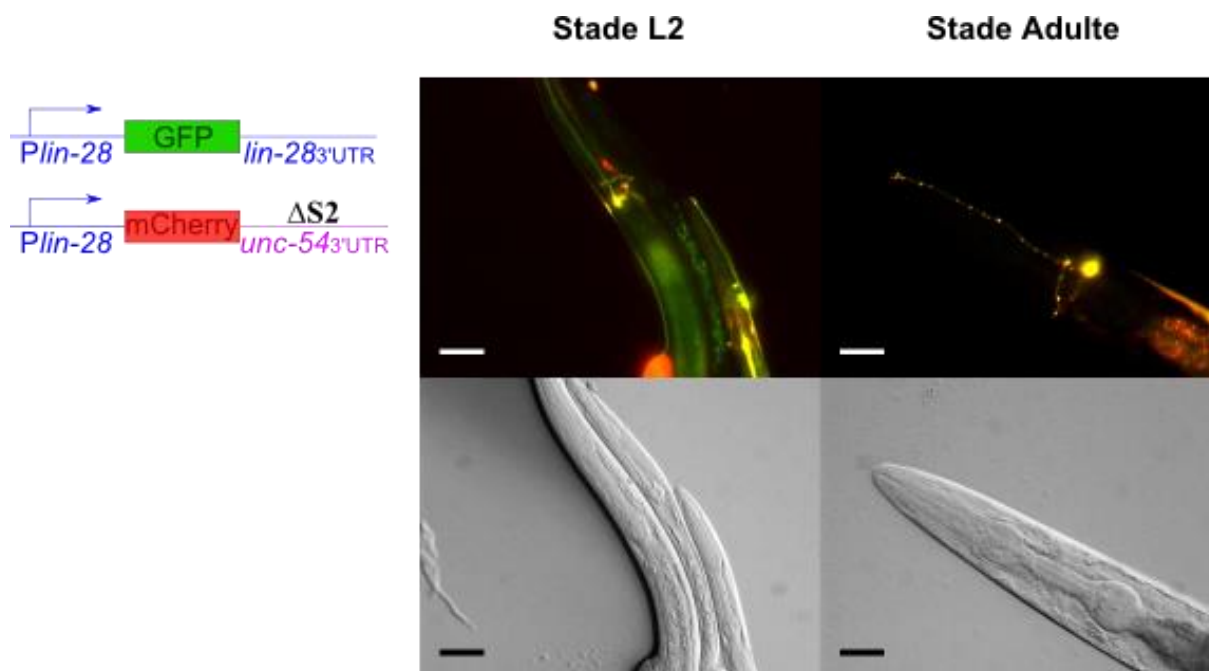


Figure 33 Expression neuronale des rapporteurs fluorescents dans les lignées portant *Plin-28::mCherry::unc-54* 3'UTR Δ S2 et *Plin-28::GFP::lin-28*. Barre d'échelle 25 μ m.

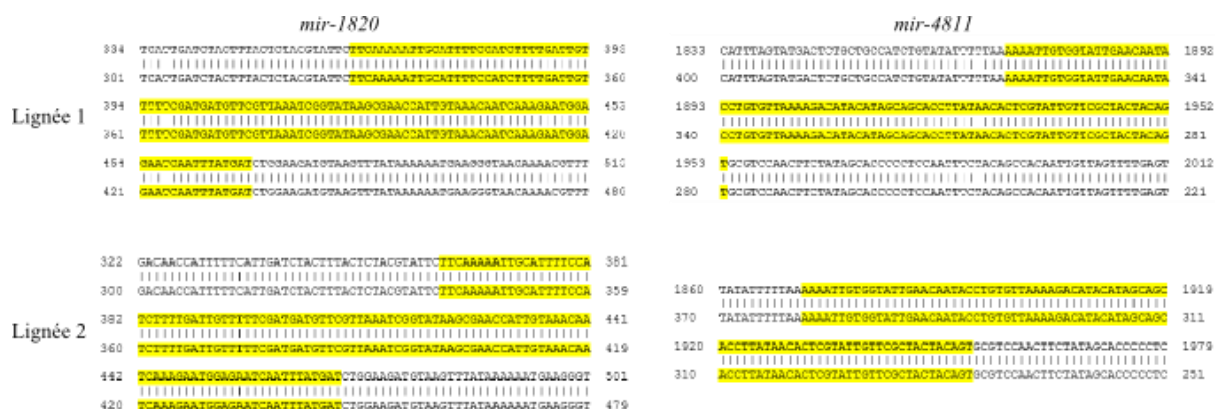
Les neurones ADF sont donc le site d'une régulation post-transcriptionnelle qui s'applique également à *unc-54* 3'UTR. Nous avons décidé d'entreprendre un crible génétique par mutagenèse aléatoire pour identifier les régulateurs impliqués dans la régulation de la région *unc-54* 3'UTR.

3. Crible génétique à l'EMS

La délétion du site 2 de la séquence *unc-54* 3'UTR a conduit à un changement drastique de l'expression de la mCherry. Afin d'identifier la ou les voie(s) de régulation impliquée(s) dans ce phénomène, nous avons opté pour un crible basé sur la mutagenèse aléatoire à l'EMS (Ethyl methane sulfonate). L'EMS est un agent hautement mutagène permettant la substitution de bases G en bases A dans l'ADN. C'est un outil puissant dont l'utilisation chez *C. elegans* a débuté dès 1974 et dont le protocole a été mis au point par S. Brenner (S. Brenner 1974 ; 5. Mutagenèse à l'EMS). Le taux de mutation par l'EMS peut atteindre $2,5 \times 10^{-3}$ mutation par gène par génération (Gengyo-Ando and Mitani 2000) et génère fréquemment des codons stop (Flibotte et al. 2010).

Cinq cribles ont été conduits successivement sur ces lignées, permettant de couvrir environ 11 000 génomes haploïdes. Pour chaque crible, 20 animaux incubés dans l'EMS ont été isolés sur des boîtes NGM contenant du G418 afin de maintenir la pression de sélection sur le transgène. Après 10 à 15 jours les boîtes sont examinées pour détecter la présence de phénotypes mutants dans le profil d'expression des protéines fluorescentes.

L'ADN génomique de deux de ces lignées (lignées 1 et 2) a été extrait et ne présente pas de mutation dans les régions des miARN *mir-4811* et *mir-1820* (Figure 34, voir annexe pour les séquences 5' et 3'). Il est donc possible que le crible ait atteint des gènes impliqués à différents niveaux dans la régulation post-transcriptionnelle observée sur *unc-54* 3'UTR. Les lignées 3 et 4 n'ayant été obtenues que récemment, le séquençage de ces gènes est en cours.



90

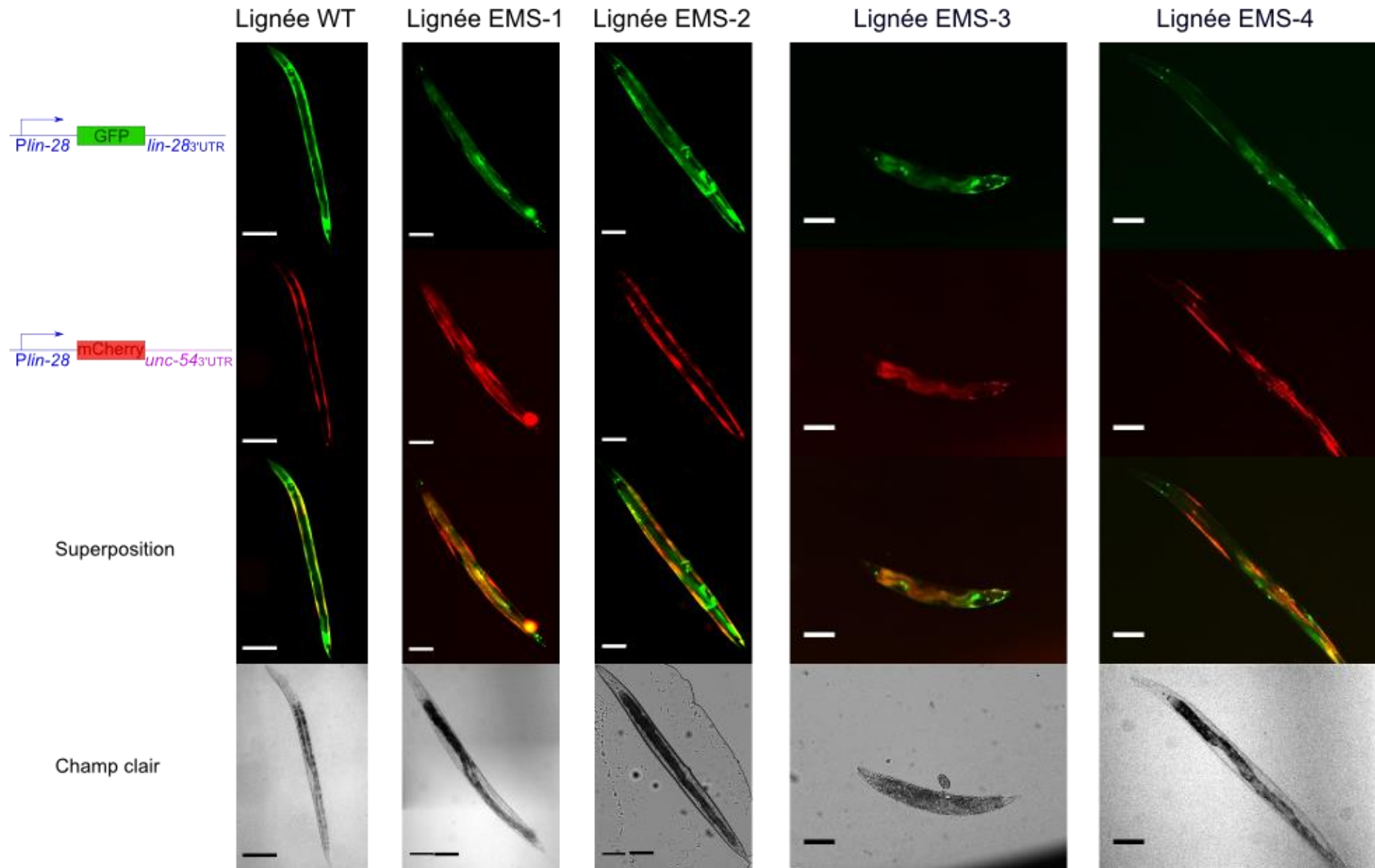


Figure 35 Comparaison des différents phénotypes obtenus après l'exposition de la lignée WT à l'EMS. Images de microscopie prises au 10X, barre d'échelle 100µm. Dans les lignées 1, 2 et 4 l'expression de la mCherry est augmenté dans les muscles.

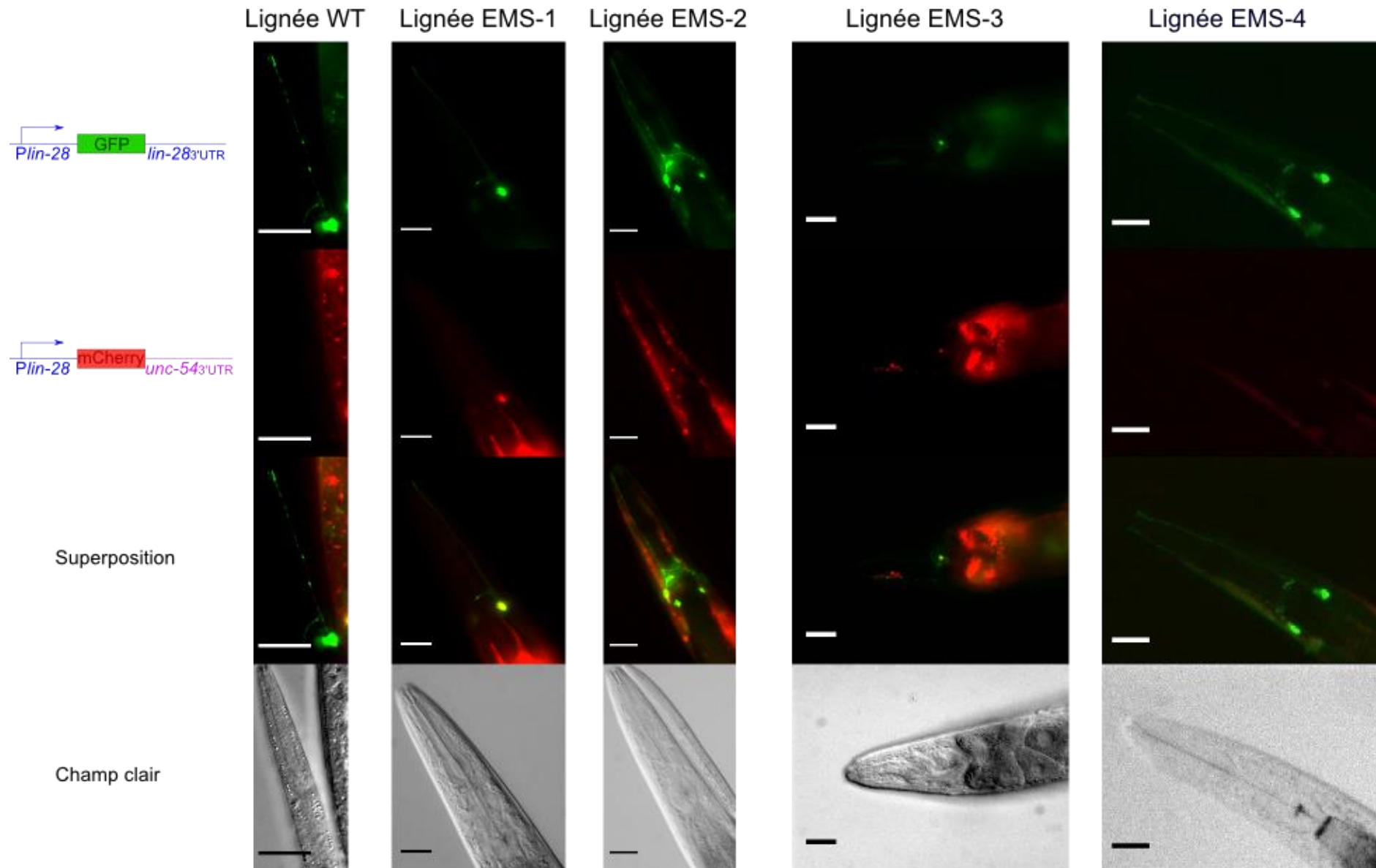


Figure 36 Comparaison des différents phénotypes obtenus après l'exposition de la lignée WT à l'EMS. Images de microscopie prises au 40X (sauf WT au 63X), barre d'échelle 25µm. Dans les lignées 1 et 3 l'expression de la mCherry est restaurée dans les neurones.

Ces lignées mutantes pourront être utilisées pour essayer d'identifier des régulateurs impliqués dans la régulation musculaire et neuronale observée, grâce aux méthodes de séquençage. Tout d'abord les gènes *mir-1820* et *mir-4811* des lignées 3 et 4 seront séquencés. Si comme dans le cas des lignées 1 et 2, les miARN ne sont pas mutés, c'est que la différence de phénotype vient d'un autre régulateur.

Avant de séquencer le génome, une série de croisements avec la souche N2 sera nécessaire afin d'éliminer les mutations silencieuses présentes à cause de l'EMS (Zuryn and Jarriault 2013). Pour mesurer l'effet des mutations de ces lignées, elles pourront être débarrassées du transgène sous forme de concatémère extrachromosomique et croisées avec une des souches portant *unc-54* 3'UTR Δ S2 afin de voir si la mutation modifie le profil d'expression de ce transgène.

4. Analyse de l'expression du gène *unc-54* in vivo

Le gène *unc-54* est exprimé dans les tissus musculaires localisés le long de l'organisme, répartis sur quatre cadrans. Le gène et sa protéine UNC-54 (myosine B) ont été parmi les premiers à être caractérisés chez *C. elegans* (Dibb et al. 1989; A. Fire and Waterston 1989; MacLeod, Karn, and Brenner 1981). L'expression du gène est constante au cours du temps et du développement de l'animal. La comparaison des ratios d'ARNm et des protéines entre le gène de la paramyosine codée par *myo-3* et *unc-54* a révélé que ces ratios étaient différents. Le ratio des ARNm est de 1 pour 20 quand le ratio des protéines est de 1 pour 1.7 (S. Honda and Epstein 1990). Les auteurs ont à l'époque suggéré un mécanisme de régulation post-transcriptionnel ou post-traductionnel pour expliquer ces différences. En 2003, l'étude de G. White (G. E. White, Petry, and Schachat 2003) permet de mettre en évidence que l'accumulation de la protéine UNC-54 est directement liée à la présence de paramyosine encodée par le gène *unc-15*. En cas d'absence de paramyosine (UNC-15) fonctionnelle, la synthèse et l'accumulation de UNC-54 sont diminuées. Les mutations nulles du gène *unc-54* n'ont en revanche pas d'impact sur la synthèse et l'accumulation de la paramyosine et de la myosine A (*myo-3*). L'explication la plus plausible est l'existence d'un phénomène d'épistasie qui reflète l'assemblage des fibres musculaires débutant par l'assemblage de la myosine A et la paramyosine. UNC-54, la myosine B, se fixe sur la structure ainsi formée pour permettre l'élongation des muscles.

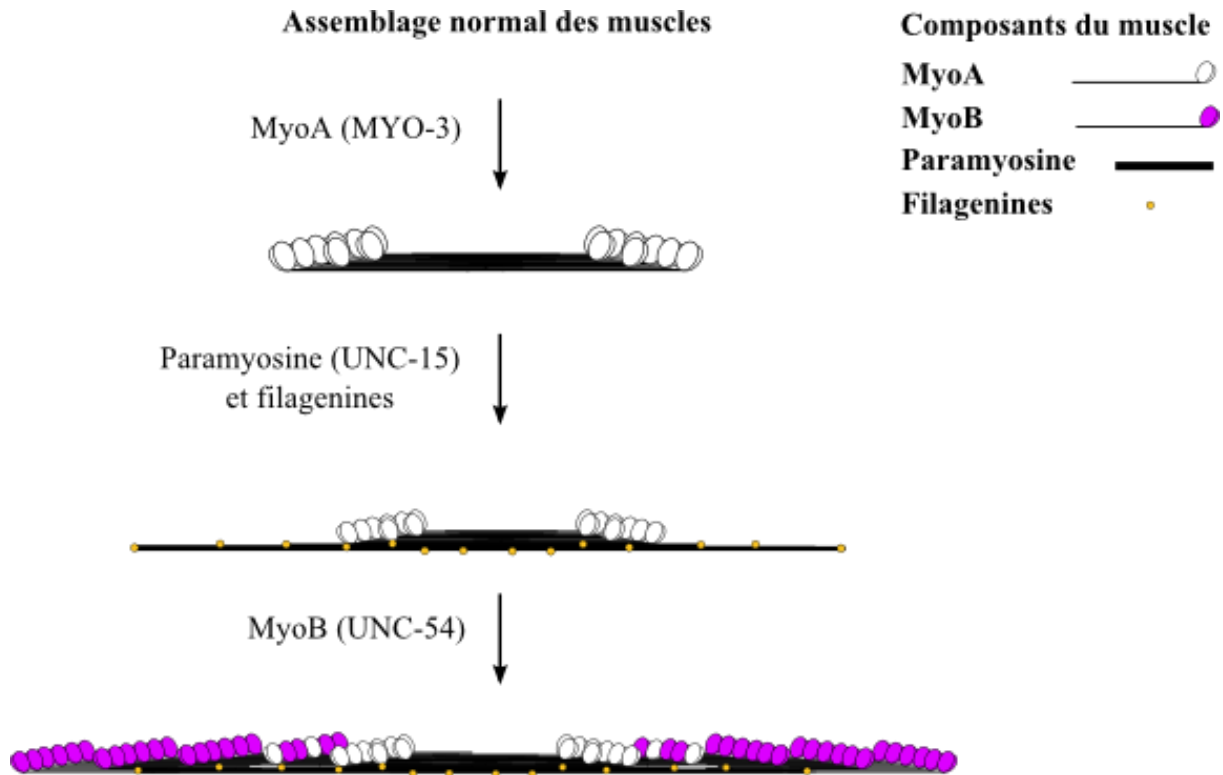


Figure 37 Rôle de UNC-54 dans l'assemblage des fuseaux musculaires (G. E. White, Petry, and Schachat 2003).

La régulation post-transcriptionnelle observée sur *unc-54* 3'UTR avec le promoteur du gène *lin-28* a également été observée dans les muscles.

Des lignées porteuses d'un double rapporteur avec le promoteur du gène *unc-54* ont été obtenues par la co-injection des plasmides *Punc-54::GFP::unc-54* 3'UTR et *Punc-54::mCherry::unc-54* 3'UTR, $\Delta S1$, $\Delta S2$, $\Delta S3$. Contrairement à ce qui a été observé avec le promoteur du gène *lin-28* aucune des lignées obtenues avec les différentes co-injections de plasmides n'ont pas révélé de différence d'expression entre la GFP et la mCherry malgré les mutations aux sites cibles (Figure 38).

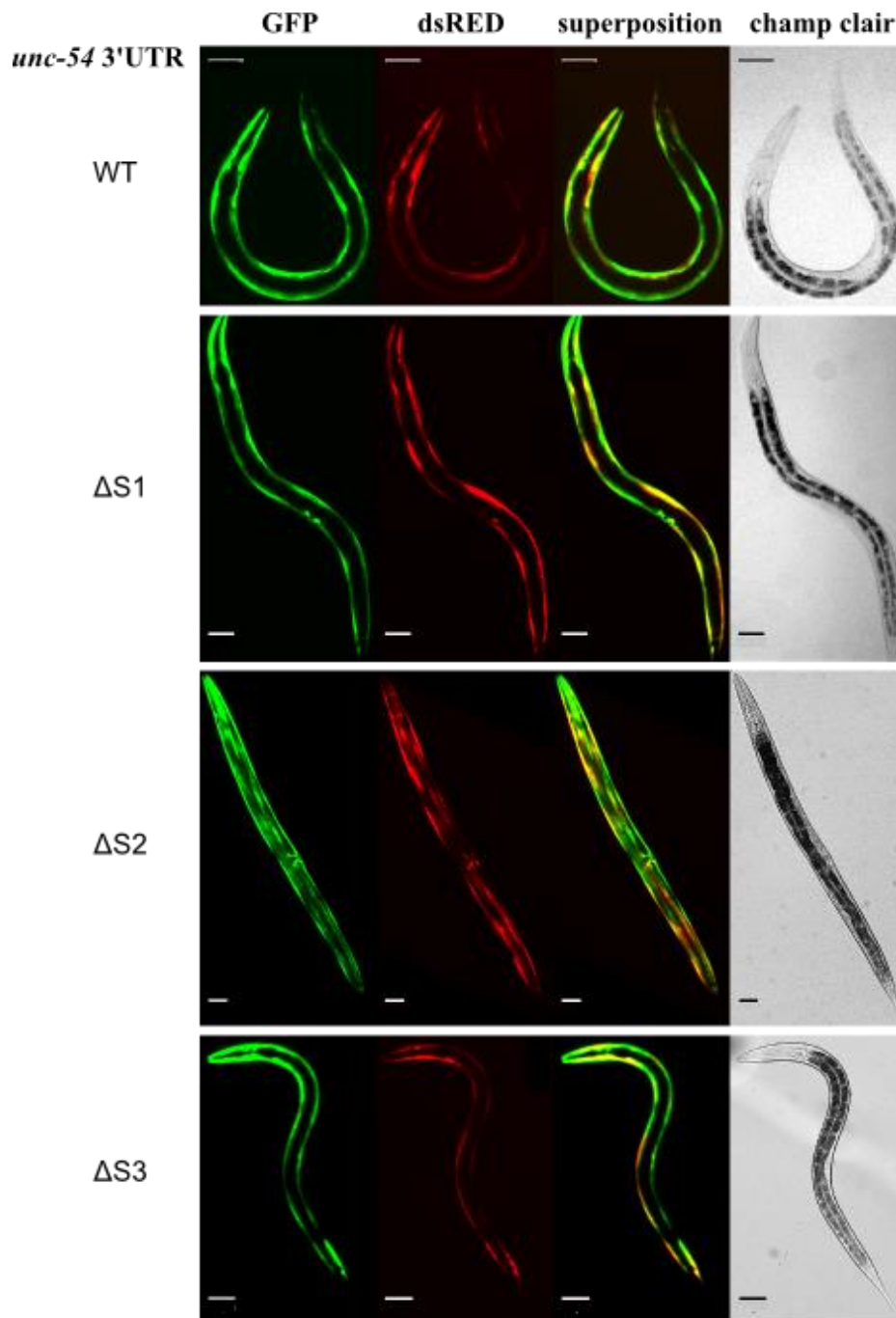


Figure 38 Expression des rapporteurs fluorescents sous le contrôle du promoteur *unc-54*. La GFP est exprimé sous les séquences régulatrices du promoteur et 3'UTR du gène *unc-54*. La dsRed est exprimé sous le contrôle du promoteur *unc-54* et sous les régions *unc-54* 3'UTR endogène (WT) ou délétées pour les sites 1 ($\Delta S1$), 2 ($\Delta S2$), 3 ($\Delta S3$). Image de microscopie au 10X, barre d'échelle 50 μm .

Les rapporteurs fluorescents exprimés sous forme de concatémère extrachromosomique n'ont pas permis de voir une régulation post-transcriptionnelle sur ce gène. Par opposition à l'expression constante du gène *unc-54*, le gène *lin-28* est régulé au cours du temps. Le niveau d'ARNm généré avec les deux promoteurs peut donc être très variable. Le niveau d'ARN généré par les promoteurs est en cours de vérification par RT-PCR semi-quantitative. Les ARNs de quatre lignées ont été extraits aux stades L2, L4 et Adultes afin de mesurer l'expression de

4 ARN : la GFP et la mCherry (ou la dsRed) pour comparer le niveau d'expression des deux rapporteurs, les gènes *unc-54* et *lin-28* au niveau d'un exon. Nous souhaitons comparer l'expression de ces ARN dans les lignées portant les plasmides où *unc-54* 3'UTR non muté et celles où le site 2 a été supprimé.

5. Potentiels gènes cibles de la régulation opérée sur la région *unc-54* 3'UTR

A. Sélection des gènes candidats

Les résultats obtenus ont permis de déterminer que la régulation post-transcriptionnelle s'opérant sur la séquence *unc-54* 3'UTR est exercée en partie dans les neurones, et causée par *mir-4811* ou *mir-1820*.

Le logiciel TargetScan a ensuite été utilisé pour établir une liste des gènes présentant un site potentiel de fixation pour *mir-4811* et *mir-1820* avec une stringence minimale afin d'obtenir une vaste liste de candidats. L'analyse a donné respectivement 2 493 et 4 354 gènes cibles potentiels.

Le laboratoire de Shawn Lockery a publié une liste des promoteurs spécifiques des neurones (http://lockerylab.uoregon.edu/groups/main/wiki/d55a4/Neuronspecific_promoters.html) correspondant aux gènes spécifiquement exprimés dans ceux-ci.

En comparant ces listes de cibles potentielles à celle des gènes exprimés dans les neurones, nous avons obtenu :

- 73 gènes exprimés dans les neurones et cible potentielles de *mir-1820*
- 28 gènes exprimés dans les neurones et cibles potentielles de *mir-4811*
- 17 gènes possédant les deux sites cibles des deux miARNs.

Le choix a été fait de se concentrer d'abord sur les gènes pouvant être régulés par les deux miARNs. Parmi les 17 gènes présentant une séquence 3'UTR possédant les mêmes sites de régulation celle du gène *unc-54*, seuls 5 (*daf-1*, *daf-4*, *eat-16*, *egl-10*, *osm-9*) sont potentiellement exprimés dans la paire de neurone ADF (Tableau 1).

Gènes	Profil d'expression neuronal
acr-5	Exprimé dans les moto-neurones de type B
daf-1	Exprimé dans les amphides
daf-4	Exprimé dans les amphides
daf-7	Limité aux ASI
dbl-1	Neurones dorsaux et rectaux
eat-16	Tous les neurones et pharynx

egl-10	Tous les neurones
egl-19	Neurones de l'anneau nerveux et rectaux
F59B2.13	DVA DVB DVC RMED RMEV
glt-1	Neurones de la queue mâle
her-1	Neurones spécifiques du mâle
mig-13	Neurones de la corde nerveuse ventrale
osm-9	Quelques neurones dont ADF
sek-1	Neurones de la corde nerveuse ventrale
unc-115	Tous les neurones
unc-64	Tous les neurones
unc-97	Neurones sensoriels

Tableau 1 Liste des gènes dont la région 3'UTR présente des sites de fixation aux *mir-1820* et *mir-4811* ainsi que leur profil d'expression dans le système nerveux. En gris, les neurones exprimés dans les neurones ADF.

Les cinq gènes ainsi répertoriés semblent être de bons candidats. Néanmoins, les gènes *eat-16* et *eat-10* sont exprimés dans l'ensemble du système nerveux, ce qui rend difficile la visibilité de régulations fines se jouant à l'échelle d'une paire de neurones. Les gènes *daf-1* et *daf-4* sont également des gènes dont l'expression est étendue à bon nombre d'autres tissus du corps. Le gène *osm-9* dont l'expression est spécifiquement cantonnée à quelques neurones, dont les neurones ADF est donc le candidat le plus intéressant pour vérifier l'existence d'une régulation par *mir-1820* et/ou *mir-4811*.

B. Expression du gène *osm-9* et analyse

Le gène *osm-9* est impliqué dans les réponses neuronales aux odeurs (évitement du benzaldéhyde) et aux stimuli sensoriels (changement de direction lorsque le nez de l'animal est touché)(Colbert and Bargmann 1995). Le gène est exprimé dans différents neurones de l'animal. Son rôle a été mis en évidence dans les neurones AWA et ASH, où la protéine est nécessaire au fonctionnement, mais pas au développement (Colbert and Bargmann 1995; Colbert, Smith, and Bargmann 1997).

Dans un premier temps, nous avons exprimé le transgène *Posm-9::GFP::osm-9* 3'UTR afin d'identifier les neurones dans lesquels il est présent de manière endogène. La séquence du promoteur a été commencée 2kb en amont du codon ATG définissant le début de l'exon 1. La séquence génomique du gène *osm-9* indique également la présence de nombreux introns. Les introns peuvent également contenir des informations d'expression, comme c'est le cas pour le gène *pie-1* (Reese et al. 2000) et donc influencer la localisation des rapporteurs. Dans un premier temps nous avons réalisé la construction *Posm-9::GFP::osm-9* 3'UTR sans y ajouter d'intron. La séquence 3'UTR choisie a été raccourcie par rapport à celle employée dans l'étude de 1997 qui couvrait la région intergénique en aval de la séquence codante du gène *osm-9*. A

ce jour, seules des séquences prédites pour la région *osm-9* 3'UTR sont disponibles, mais aucune expérience de séquençage de son ARN ne permet d'établir s'il existe plusieurs isoformes de la région. La séquence *osm-9* 3'UTR du gène a donc été limitée à 675pb soit un peu plus que les séquences théoriques indiquées par TargetScan (+100pb) et Wormbase (+500pb). La lignée injectée a été marquée au DiI afin de localiser les amphides et l'anneau nerveux.

Les premières observations ont montré que le gène *osm-9* est exprimé dans 4 neurones de la tête, à proximité du pharynx, au-dessus du nerve ring (Figure 39).

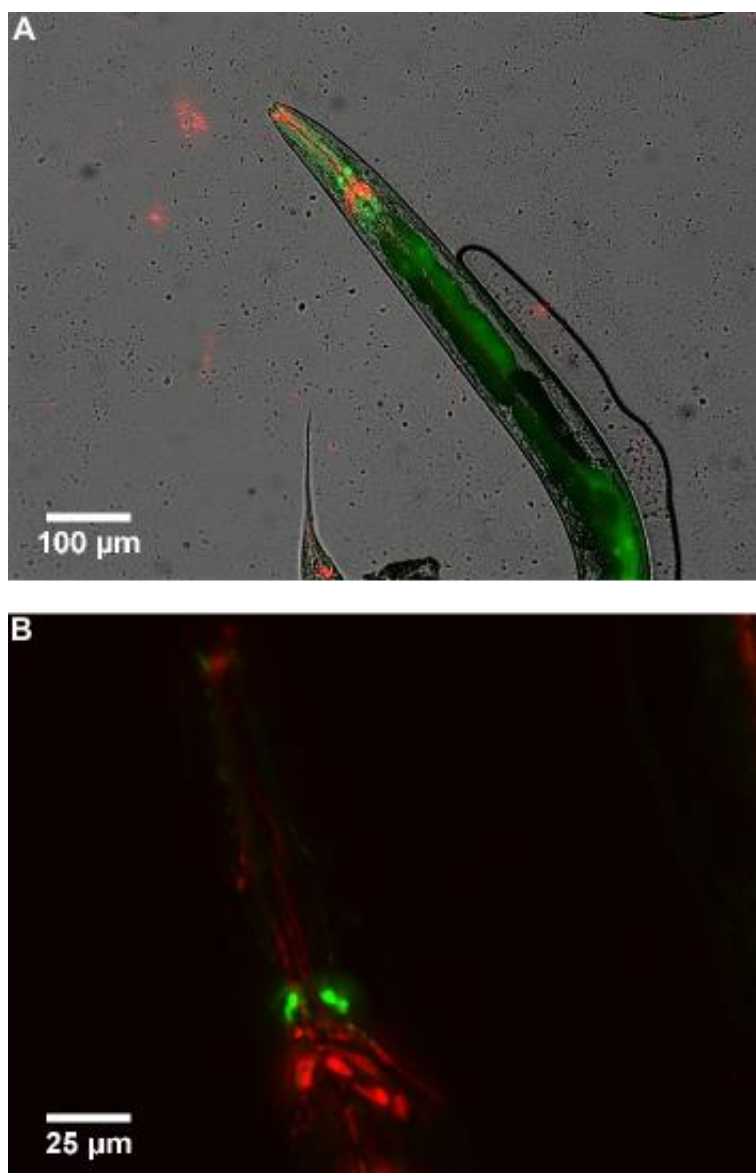


Figure 39 Images de microscopie de la lignée exprimant *Posm-9::GFP::osm-9* 3'UTR après coloration au DiI. A Animal de stade L4 imagé au 10X. Superposition des canaux vert, rouge et champ clair. La GFP est exprimé dans des neurones au milieu du pharynx, à distance des amphides colorés au DiI. **B** Tête d'un animal de stade L4 imagé au 40X. Les canaux vert et rouge sont superposés mettant en évidence 4 noyaux exprimant la GFP au-dessus des amphides colorés en rouge par le DiI.

Afin d'identifier les neurones marqués en GFP, des images à plus fort grossissement ont été réalisées, permettant de déterminer la forme de la dendrite et des axones.

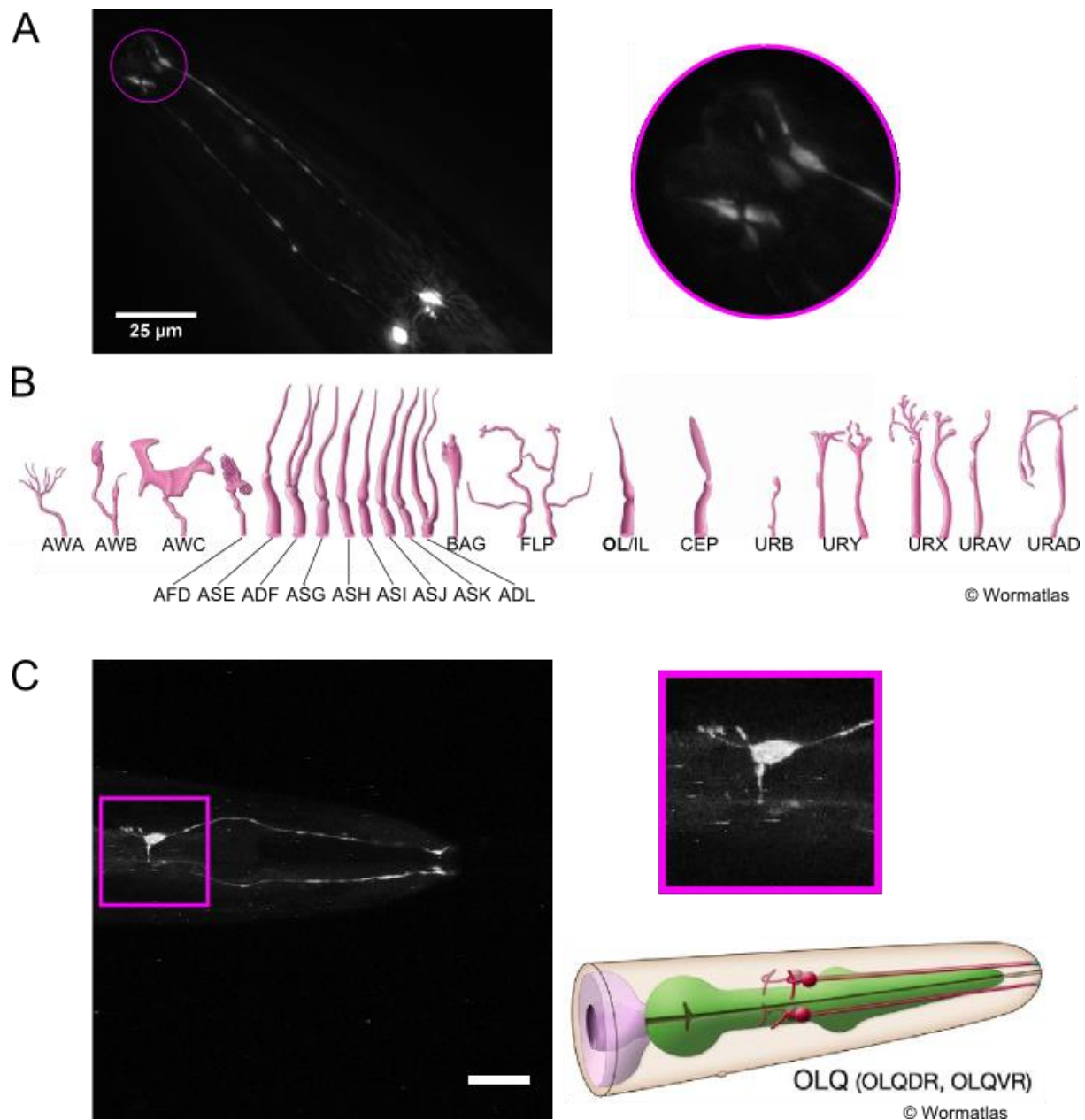


Figure 40 Identification des neurones exprimant le transgène *Posm-9::GFP::osm-9* 3'UTR. **A** Projection sur l'axe Z d'une pile d'image de la tête d'un vers stade L4 imagé au 63X en immersion. L'extrémité des dendrites est agrandie en médaillon. **B** Schéma représentant les extrémités dendritiques des différents types de neurones chez *C. elegans*. Les dendrites de l'image A sont semblables à celle des neurones OLQ. **C** Projection sur l'axe Z d'une pile d'image d'un des 4 neurones marqués en GFP imagé au 63X en immersion avec un microscope confocal. Le noyau et l'axone sont agrandis dans le médaillon. La forme de l'axone confirme l'identification des neurones OLQ.

La simple fusion *Posm-9::GFP::osm-9* 3'UTR s'exprime dans les lignées transgéniques dans les neurones OLQ. L'étude de 1997 décrit que la fusion de la GFP avec le promoteur entraîne l'expression dans les neurones ADL et OLQ uniquement. Pour permettre l'expression dans AWA il faut les 80 premiers nucléotides de la séquence codante. L'expression dans les autres

neurones IL2, AWC, ASE, ADF, ASG, ASH et ASI n'a été observée qu'en ajoutant la séquence codante de la protéine ainsi que la région 3'UTR (Colbert, Smith, and Bargmann 1997).

6. Analyse du site 2 muté responsable du phénotype observé

La délétion $\Delta S2$ dans la séquence *unc-54* 3'UTR correspond à la suppression de deux sites cibles de miARN, *mir-4811* présent une seule fois dans la séquence, et *mir-1820* présentant un second site en aval. Pour préciser lequel des deux est responsable, de nouvelles délétions affectant les deux sites séparément ont été entreprises.

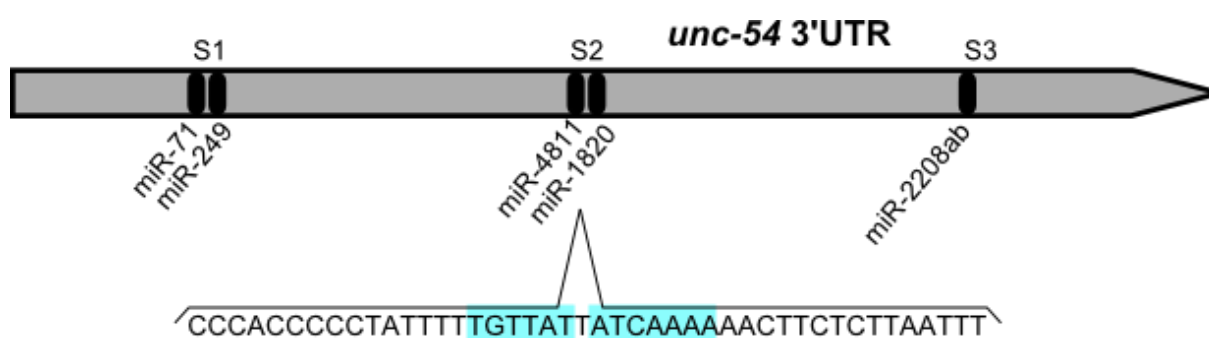


Figure 41 Séquences des sites *mir-4811* (6pb) et *mir-1820* (7pb). L'ensemble des deux constitue la délétion originale $\Delta S2$ (15pb).

Un autre plasmide a été généré par mutagenèse dirigée sur le plasmide par PCR et séquencé pour s'assurer de la délétion du site cible de *mir-4811*. La co-injection des plasmides porteurs de *unc-54* 3'UTR $\Delta S4811$ a permis d'obtenir plusieurs lignées transgéniques, 2 sous le contrôle de chaque promoteur *lin-28* et *unc-54*. Ces lignées n'ont pas présenté de différence d'expression des rapporteurs fluorescents par rapport aux lignées portant *unc-54* 3'UTR non muté (Figure 42).

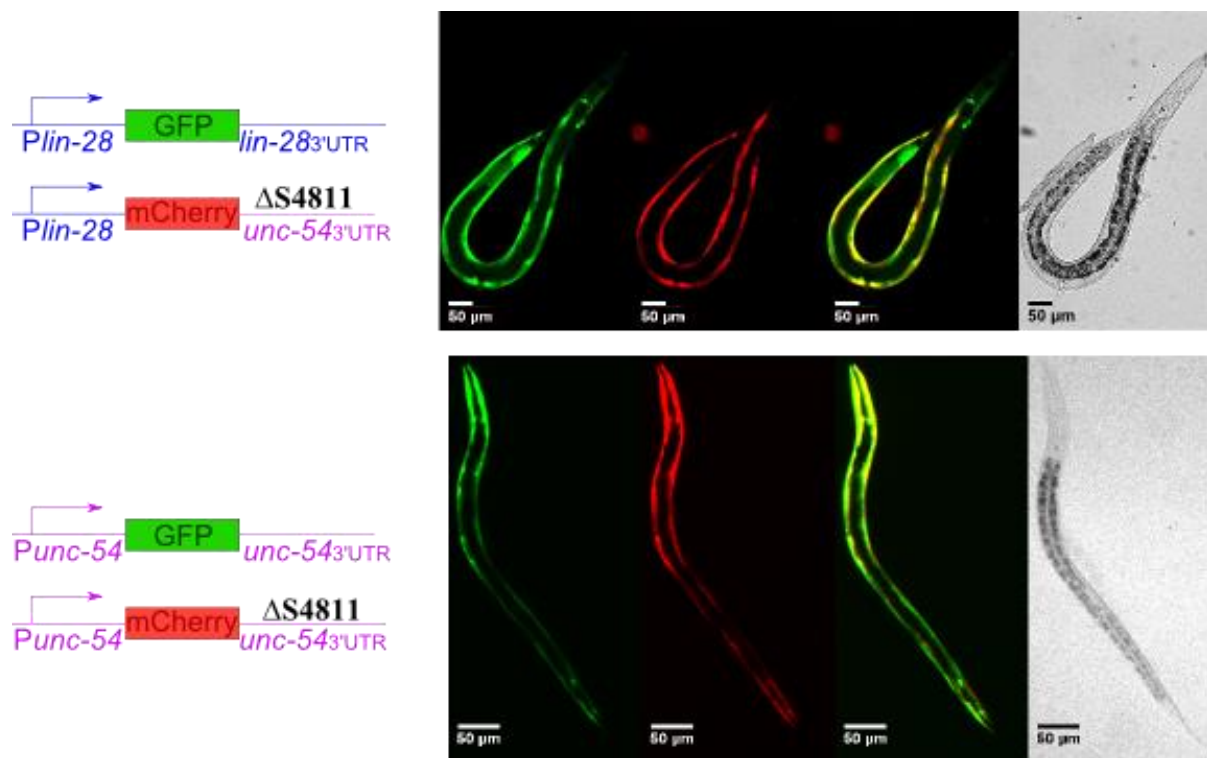


Figure 42 Impact de la délétion du site cible de *mir-4811* dans la région 3'UTR de *unc-54* 3'UTR. La délétion ne modifie pas le profil d'expression des rapporteurs fluorescents.

La délétion du site cible de *mir-4811* ne modifie pas l'expression de la mCherry dans les lignées, ce qui semble indiquer que *mir-1820* serait suffisant pour exercer la régulation. Les plasmides porteurs des délétions $\Delta S1820$ n'ont pas encore été obtenus pour confirmer son rôle *in vivo*.

Nous nous sommes donc intéressés au profil d'expression de *mir-1820*. La fusion *Pmir-1820::GFP::mir-1820* 3'UTR a permis d'obtenir des animaux exprimant la GFP. Le plasmide a été co-injecté avec *Plin-28::mCherry::lin-28* 3'UTR afin de voir s'il y avait une colocalisation dans les muscles et les neurones ADF (Figure 43). Pour l'instant une seule lignée a été obtenue avec ce transgène, mais il semble que *mir-1820* soit exprimé faiblement dans les muscles latéraux et fortement dans le muscle stomatointestinal et les neurones CAN et ADF (Figure 44).

Le profil d'expression fourni par le transgène *Pmir-1820::GFP::mir-1820* 3'UTR englobe les muscles et les neurones ADF où la régulation sur *unc-54* 3'UTR a été observée. Il est donc probable que l'inhibition observée sur *unc-54* 3'UTR soit due à une interaction avec ce miARN.

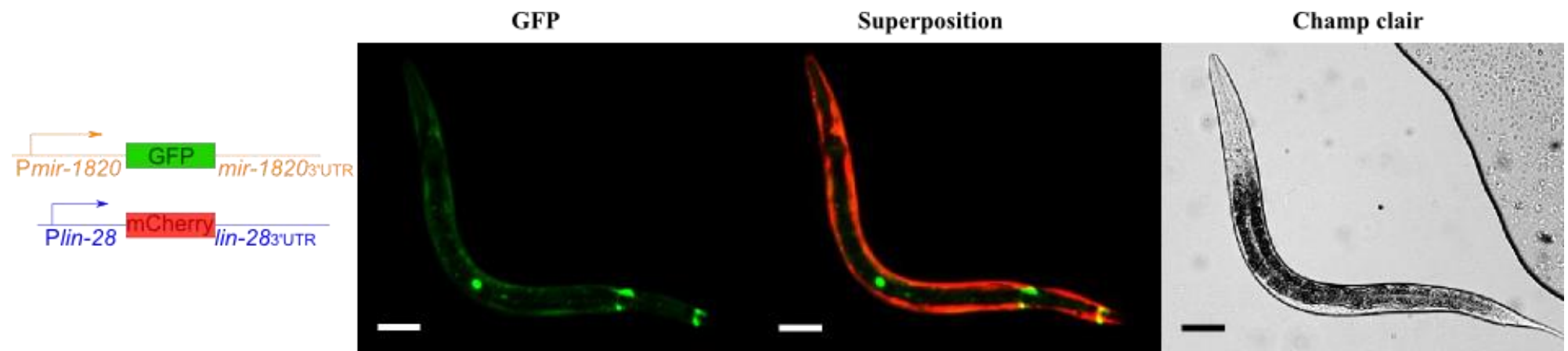


Figure 43 Activité du promoteur de la région *mir-1820* comparée avec le rapporteur *P_{lin-28}::mCherry::lin-28 3'UTR*. Images de microscopie prises au 10X, barre d'échelle 50µm.

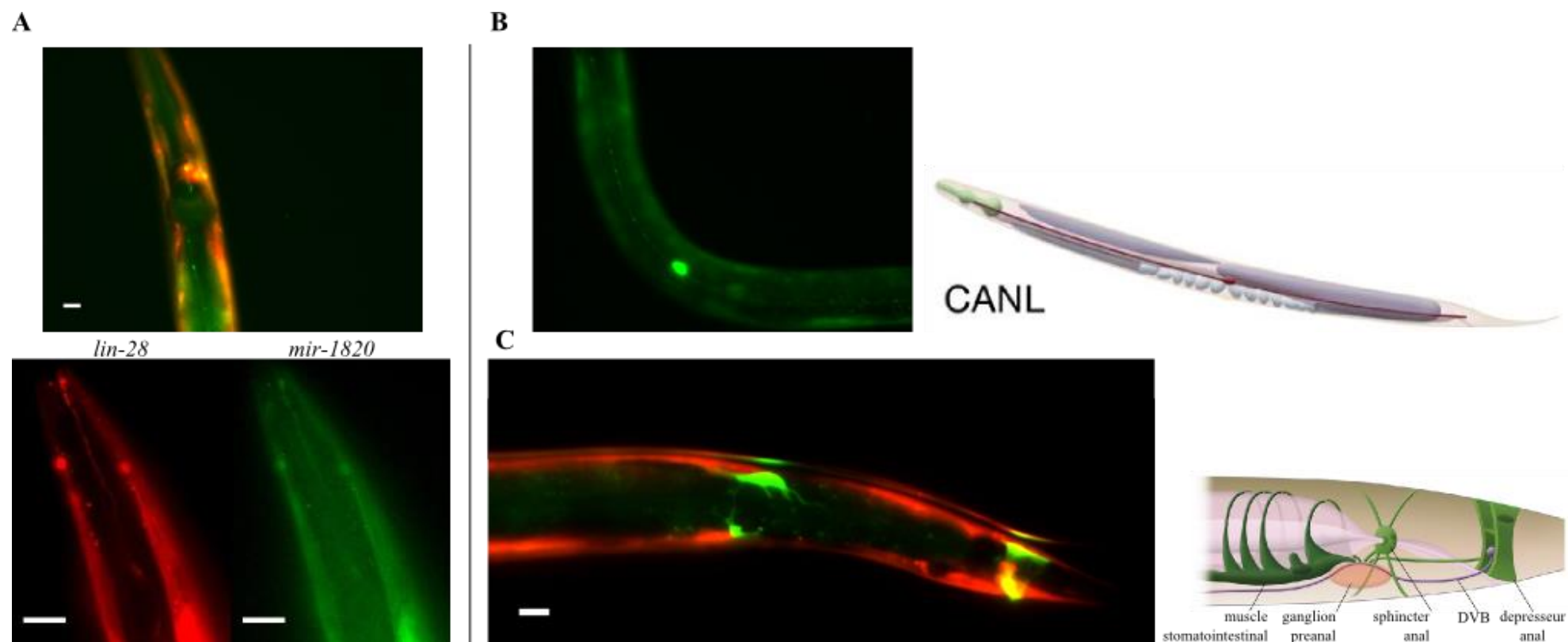


Figure 44 Identification des zones d'expression de *Pmir-1820::GFP::mir-1820* 3'UTR et *Plin-28::mCherry::lin-28* 3'UTR. **A** Expression dans les neurones de la tête. La GFP et la mCherry sont colocalisées dans les neurones ADF. **B** Expression de la GFP au milieu de l'animal correspondant aux neurones CAN. **C** Expression de la GFP dans la partie postérieure de l'animal. Barres d'échelle 10µm.

3. Biologie moléculaire : clonage par *Chew Back Annealing*

Annealing

1. Le choix d'une méthode novatrice

Au cours du travail présenté précédemment, bon nombre des lignées transgéniques utilisées furent générées par la micro-injection d'un mélange de plasmides. Au départ, il était prévu de mettre au point un rapporteur unique contenant deux fluorophores sous forme d'une cassette. Par la suite, nous nous sommes dirigés vers la co-injection de plasmides ce qui impliquait deux fois plus de clonages à effectuer pour chaque gène de l'étude.

L'analyse d'un grand nombre de gènes nécessite une méthode permettant de cloner rapidement les séquences régulatrices : le promoteur et la région 3'UTR. Situé en amont et en aval des rapporteurs, il fallait une méthode permettant d'assembler au moins 3 fragments dans un vecteur, le plus simplement possible. C'est pourquoi, la construction de ces plasmides a été réalisée grâce à la méthode du *Chew Back Annealing* ou *Gibson assembly* (Gibson 2011; Gibson et al. 2009).

Le *Chew Back Annealing* est une méthode permettant l'assemblage de plusieurs fragments d'ADN en une seule réaction d'une heure. Le plus souvent, les fragments sont des amplifiats obtenus par PCR, y compris le vecteur de destination. Il est néanmoins possible d'utiliser des fragments issus de digestions enzymatiques et purifiés sur gel. Les fragments d'ADN sont mis en présence en quantité équimolaire dans un mélange réactionnel contenant les enzymes T5 exonucléase, polymérase phusion et ligase avec du polyéthylène glycol pour favoriser la recombinaison (Figure 45).

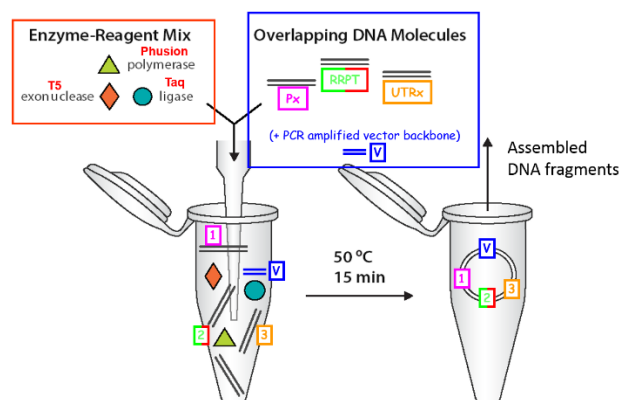


Figure 45 Réalisation de la réaction de *chew back annealing* (adaptation d'une communication personnelle de C. Lartigue). Les fragments d'ADN et les réactifs sont mis en présence dans un microtube et laissés de 15 à 50 min à une température de 50°C dans le thermocycleur. À l'issue de la réaction, l'ADN circulaire généré est inséré dans des bactéries compétentes.

La majorité des plasmides de ce travail de thèse a nécessité le clonage de 3 inserts indépendants dans un même plasmide. Les méthodes conventionnelles de clonage (par digestion et ligation) implique de cloner les inserts successivement, en plusieurs fois. En comparaison, la méthode *chew back assembly* permet le clonage de tous les fragments en une étape. Le temps d'obtention des plasmides est ainsi réduit (Figure 46).

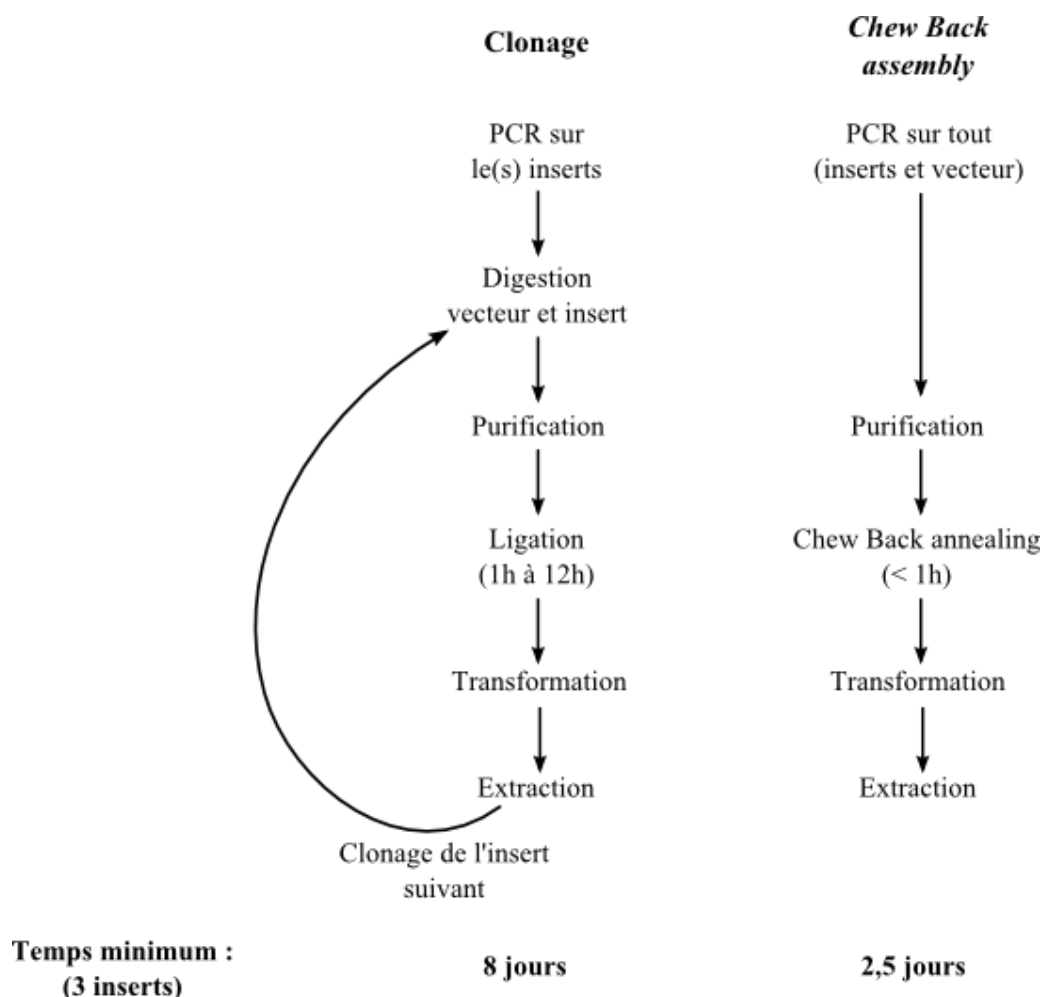


Figure 46 Comparaison des méthodes de clonage conventionnelle et du *Chew Back assembly*. Le temps donné est calculé pour cloner 3 inserts dans un vecteur, dans les conditions optimales de fonctionnement, en supposant que chaque étape fonctionne.

A posteriori, il a été possible d'analyser l'efficacité de la méthode selon les différents assemblages réalisés.

2. Test de l'efficacité de la méthode

Au cours de ces travaux, 58 essais d'assemblages de plasmides ont eu lieu avec cette méthode. Il a donc été possible d'analyser le rendement de la technique. Le nombre de tentatives à faire pour obtenir le plasmide attendu, ainsi que le nombre de colonies positives ont donc été comptabilisés.

Tout d'abord, le nombre de tentatives nécessaires pour que l'assemblage fonctionne. Un essai correspond aux étapes successives de la réaction dans le thermocycleur (1h à 50°C) suivi par la transformation (électroporation), la culture des colonies puis l'extraction et l'analyse des plasmides par restriction ou PCR. Un essai se déroule sur 2,5 jours pour confirmer le bon assemblage du plasmide.

Sur les 58 plasmides, 14 n'ont pas été assemblés et 44 ont été assemblés correctement (76%). Parmi ces 44 plasmides, 32 ont été assemblés dès le premier essai (Figure 47) soit 73% des assemblages. Seul le plasmide *Punc-119::mCherry::unc-54* 3'UTR a demandé 5 essais pour obtenir l'assemblage.

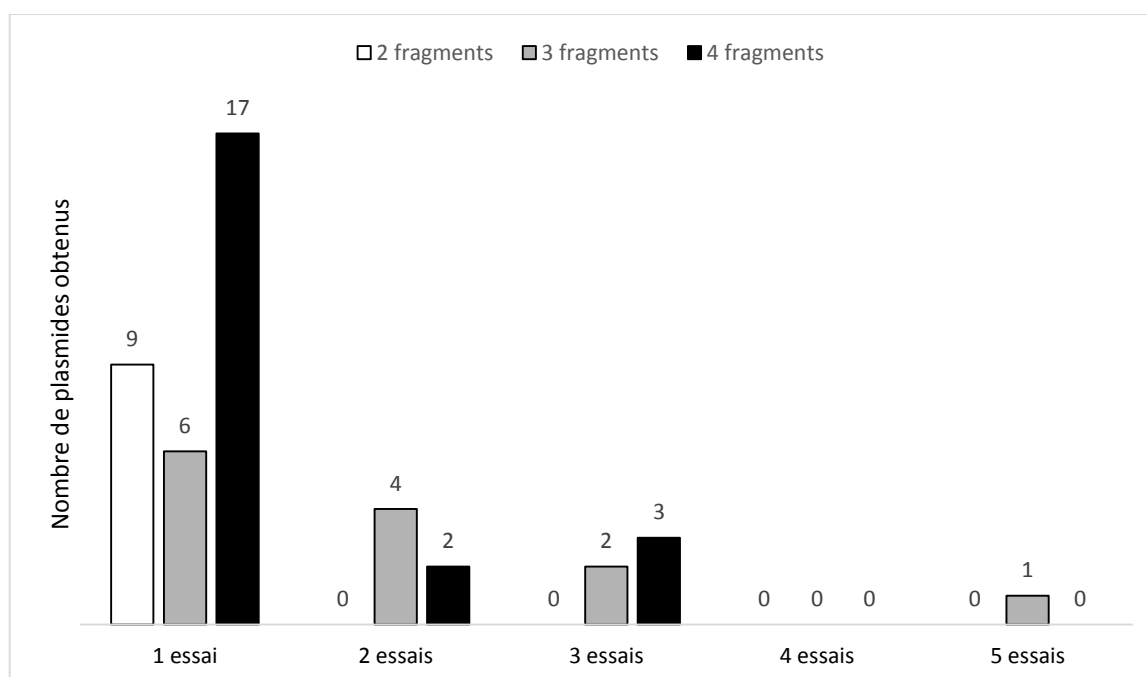


Figure 47 Répartition des assemblages réalisés en fonction du nombre de fragments mis en présence et du nombre d'essais entrepris.

Les résultats mettent aussi en évidence que le nombre de fragments n'est pas une source de difficultés supplémentaires pour l'utilisation de cette technique, déjà utilisée avec succès pour réaliser des assemblages de 4 à 5 fragments de 60 pb à plusieurs kb (Gibson et al. 2010). En effet, sur 22 assemblages de 4 fragments (soit le vecteur et 3 inserts), 17 ont été assemblés au premier essai. L'assemblage ayant demandé le plus d'essais ne comptait que 3 fragments à assembler.

Les transformations ont abouti à un nombre de colonies allant de 20 à un tapis bactérien sur la boîte. Selon les tentatives, un nombre plus ou moins important de ces colonies a été testé à partir des boîtes. Le nombre de colonies à tester par réaction d'assemblage pour avoir le plasmide assemblé correctement ne semble pas non plus dépendant du nombre de fragments (Figure 48). En moyenne, pour les assemblages réussis, 3 colonies testées sur 8 contiennent le plasmide

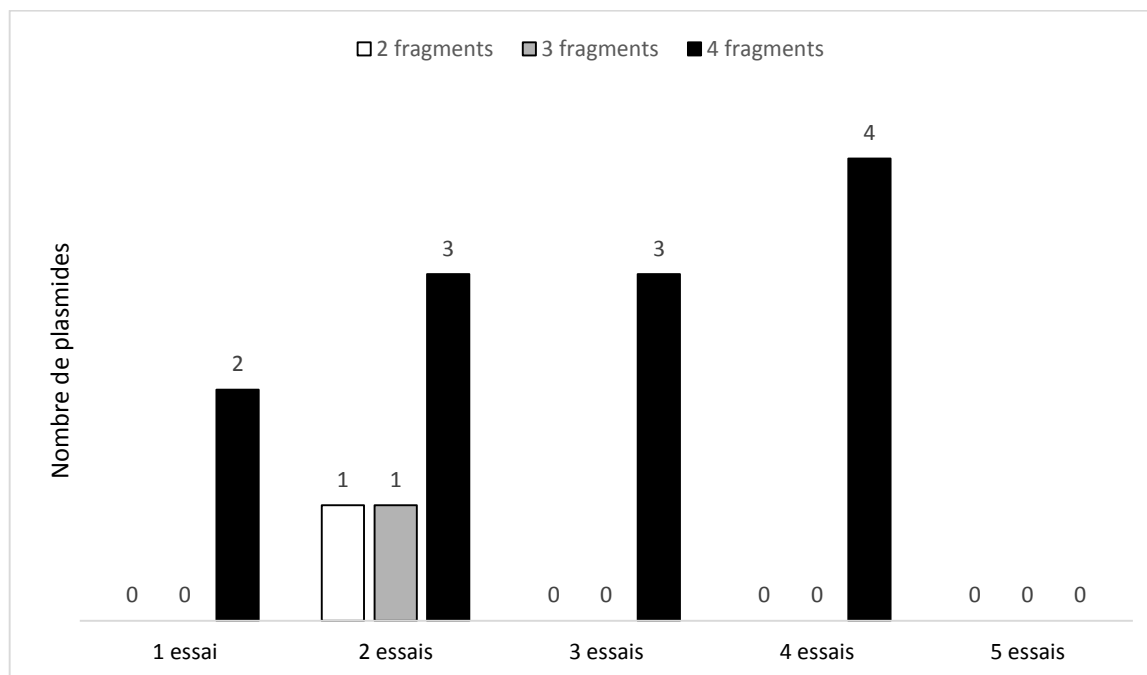


Figure 49 Répartition des assemblages non réalisés en fonction du nombre de fragments mis en présence et du nombre d'essais entrepris. Le seul assemblage ayant été testé 5 fois a fonctionné (Figure 47), les autres assemblages n'ont pas été testé une 5^{ème} fois.

Aucun assemblage n'a demandé plus de 5 essais pour obtenir la construction désirée. Les assemblages à 2 et 3 fragments ont fonctionné pour 90% des constructions envisagées. Les assemblages à 4 fragments donnent des résultats plus variables puisque 12 plasmides sur 34 n'ont pas été obtenus mais 17 assemblages sur 34 ont fonctionné dès le premier essai. Cette différence dans l'efficacité semble pouvoir être imputée aux fragments d'ADN mis en présence plutôt qu'à la méthode elle-même. Le nombre de fragments mis en présence dans la réaction ne serait donc pas prédictif de son efficacité.

En conclusion, la méthode permet de construire des plasmides à partir de 2 à 4 fragments de PCR, de manière simple et efficace, en une seule étape, dans un délai raisonnable.

V. Discussion

1. Les constructions sous forme d'opéron

1. La cassette RRPT

Un rapporteur de régulation post-transcriptionnelle sous la forme d'un opéron constituerait un outil extrêmement pratique pour les études sur les régulations opérées dans les régions 3'UTR des ARNm. L'assemblage d'une construction sous la forme d'un opéron est grandement facilité par les techniques de biologie moléculaire moderne, comme le *chew back annealing*. L'absence de cicatrices laissées par la méthode la rend idéale, surtout pour la réalisation de fusion dans des régions précises où l'insertion de sites de restriction peut être délicate.

Ilyass Zniber avait déjà obtenu plusieurs constructions contenant la cassette lorsque j'ai débuté mon travail. La cassette permet l'expression de la GFP et de la mCherry sous le contrôle du même promoteur. La mCherry joue le rôle du contrôle interne et est suivie d'une région 3'UTR permissive. La GFP reflète l'expression du gène en étant en amont de la région 3'UTR étudiée. Le profil d'expression attendu était donc l'observation de cellules exprimant les deux protéines et une partie n'exprimant que la mCherry mettant en évidence une régulation post-transcriptionnelle.

L'obtention des lignées analysées au cours de cette étude a suivi de près mon arrivée au laboratoire. L'analyse des lignées portant la cassette RRPT a rapidement mis en évidence un problème d'absence d'expression de la mCherry dans certaines cellules exprimant la GFP, dans le cas du gène *lin-28* dans les neurones ADF en particulier. La comparaison avec les lignées qui portent un transgène de type promoteur::GFP::3'UTR a également souligné une différence d'expression entre la GFP de la cassette et la GFP sous le contrôle direct du promoteur.

Initialement, nous avons suspecté un problème de régulation lié à l'*operon linker* inséré entre la GFP et la mCherry de la cassette. La séparation des éléments sur deux plasmides a apporté des éléments de réponse concernant cette hypothèse. Le niveau d'expression de la GFP dans les lignées obtenues est devenu identique à celui observé dans les lignées contrôle qui exprimait la GFP sous le contrôle direct du promoteur. Cependant, l'expression de la GFP dans des cellules n'exprimant pas la mCherry a perduré.

L'*operon linker* n'est donc visiblement pas au centre du problème de l'expression des protéines. Néanmoins, le niveau d'expression de la GFP était plus bas dans les cassettes que directement sous le contrôle du promoteur. La séquence utilisée comme *operon linker* provenait de la région intergénique *gpd-2/gpd-3*. Le trans-épissage de cette région implique l'utilisation d'un brin SL2 (Spieth et al. 1993). Hors, le trans-épissage subit par le pré-ARNm dépend de la maturation du

3'UTR du gène précédent (Kuersten et al. 1997) et pourrait influencer le traitement de la région 3'UTR du gène (M. Mangone et al. 2010). Il existe environ 200 gènes pour lesquels le trans-épissage se fait avec SL1, mais l'organisation intergénique est telle que le site PAS du premier gène est très proche du site de trans-épissage du second (Allen et al. 2011; C. Williams, Xu, and Blumenthal 1999). Il faudrait donc une région 3'UTR contrôle non seulement permissive mais satisfaisante aux contraintes de tailles et de signaux permettant que les deux gènes de l'opéron soient trans-épissés en SL1.

2. Les applications des constructions en opéron

L'expression de gènes en opéron constitue un avantage majeur non seulement pour simplifier la transgénèse de l'organisme étudié, mais aussi pour disposer d'un contrôle interne à l'expression du transgène.

L'une des applications possibles d'un transgène sous forme d'opéron est de ne pas recourir à une protéine de fusion dont l'expression peut être légèrement modifiée par l'ajout de la protéine fluorescente. Plusieurs études chez *C. elegans* ont ainsi construits des rapporteurs basés sur un modèle du type *Pgène::gène::operon linker::GFP* afin de déterminer le profil d'expression du gène (Cheung et al. 2004; Coates and de Bono 2002; Palaima et al. 2010; Tullet et al. 2008). Ces expériences ont souvent confirmé les profils d'expression obtenus avec des protéines de fusion ou des hybridations *in situ*.

Les opérons possèdent aussi un avantage indéniable pour bénéficier d'un contrôle interne dans le cas d'analyses de régulations post-transcriptionnelles. Cette technique a été testée sur des embryons de souris porteurs d'un rapporteur LacZ exprimé constitutivement et dont la région 3'UTR portait des sites cibles spécifiques d'un miARN. Cette méthode a permis de définir les lieux d'expression des miARN (Mansfield et al. 2004). Plus tard, des constructions en opérons ont été utilisées pour étudier l'impact des régions 3'UTR sur l'expression des gènes de *C. elegans*. La GFP et la mCherry étaient exprimées sous le contrôle du promoteur *pie-1* dans les gonades de l'animal et suivie de région 3'UTR différentes. Les différences d'expression ainsi repérées ont permis de confirmer les zones où les gènes étaient régulés de manière post-transcriptionnelle (Merritt et al. 2008).

Une application originale de l'usage des opérons a été développée récemment chez *C. elegans*. L'une des méthodes classiquement employée avec cet organisme modèle est l'ARN interférence (ARNi) afin d'étudier des gènes dont les délétions sont létales. L'ARNi peut être réalisée de manière ciblée à différents stades de développement afin de contourner la létalité qu'une mutation induirait. L'ARNi est aussi souvent utilisée pour réaliser des cribles génétiques

afin d'identifier les gènes d'une voie de régulation, dans ce cas le crible est généralement basé sur le sauvetage d'un phénotype, mais il peut aussi être basé sur une modification phénotypique comme la fluorescence d'un transgène. Les neurones de *C. elegans* sont moins sensibles à l'ARNi, ce qui a conduit au développement d'une souche mutante augmentant la sensibilité des neurones (D. Wang et al. 2005). L'ajout d'une structure en opéron des séquences codantes des gènes *sid-1* et *rde-1* connectées par un site de trans-épissage de type SL2 et placées sous le contrôle d'un promoteur spécifique a permis de délimiter l'action des ARNi à un type précis de neurones (Firnhaber and Hammarlund 2013). La construction a été testée avec l'ARNi du gène *ama-1* codant la RNA Pol II entraînant normalement une forte létalité qui a été dans ce cas complètement abolie.

Cependant, ces études étaient très ciblées sur certaines parties de l'organisme (neurones spécifiques, gonades) et les niveaux d'expression du gène situé en aval de la séquence de trans-épissage SL2 n'ont jamais été vérifiés. L'utilisation de cette séquence ne donne donc aucune garantie sur le niveau d'expression des gènes en aval et peut donner lieu à une certaine variabilité. Une autre approche avec une séquence IRES (*i*nternal *r*ibosomal *e*ntry *s*ite) (Pfleger et al. 2006) a également été essayée chez *C. elegans* sous le contrôle du promoteur *myo-3* et a conduit à la co-expression des deux protéines RFP et GFP mais sans études quantitatives de leur niveau respectif (D. Li and Wang 2012).

Le problème de l'expression des protéines sous un même promoteur semble toutefois avoir trouvé une réponse dans l'utilisation des séquences 2A de peptides viraux. Ces séquences permettent l'expression stœchiométrique de protéines sous le contrôle d'un seul promoteur. Un seul ARNm est généré, et c'est à l'étape de traduction que différents peptides seront formés à partir du transcrit unique (Ahier and Jarriault 2014).

A ce jour les transgènes sous forme d'opéron ont permis d'étudier avec succès les régulations génétiques chez les eucaryotes. Cependant, la question de la quantification du niveau d'expression des différents transcrits générés n'a été que partiellement traitée. Le système d'expression des protéines en proportion stœchiométriques, basé sur l'inclusion de séquences 2A, vient d'être développé et semble très prometteur pour la conduite d'études à venir.

L'évolution des méthodes de clonages, notamment le *chew back annealing*, facilite l'obtention rapide de transgène. Cette méthode de clonage présente deux gros avantages qui sont de pouvoir obtenir en quelques jours des assemblages à partir de plusieurs fragments et en une seule étape, sans cicatrices laissées par l'ajout de sites de restriction. Au cours de ce travail nous avons substitué deux plasmides co-injectés à un seul, et avons doublé le nombre de construction sans difficultés. L'expression des plasmides présents en multicopie peut parfois être compromise par

le phénomène de *silencing*, ce qui n'a pas été le cas dans cette étude. La co-injection des plasmides n'a dans ce cas pas posé de problèmes pour l'observation de régulations post-transcriptionnelles. Nous avons pu observer des différences d'expression entre les plasmides liées à la région 3'UTR placée en aval du rapporteur fluorescent. L'analyse des régulations post-transcriptionnelles peut donc être faite avec deux plasmides co-injectés.

2. La régulation du gène *unc-54*

1. La région *unc-54* 3'UTR subit une régulation post-transcriptionnelle

Au cours de notre étude nous avons pu mettre en évidence une régulation s'opérant sur la région *unc-54* 3'UTR. Nous avons observé cette régulation dans les neurones ADF ainsi que dans les muscles latéraux. Cette régulation a été remarquée grâce à l'absence d'expression du rapporteur mCherry suivie par *unc-54* 3'UTR tant dans les lignées porteuses des constructions contenant la cassette RRPT que dans les lignées porteuses de transgènes indépendants co-injectés.

La délétion d'une région de 15 pb a suffi à modifier dans l'ensemble de l'organisme l'expression de la mCherry suivie par *unc-54* 3'UTR sous le contrôle du promoteur de *lin-28*. Cette délétion a provoqué l'expression de la mCherry dans les neurones ADF et augmenté son expression dans les muscles. Le même système de co-injection de plasmides contenant des rapporteurs différents a été employé et a permis d'obtenir des lignées exprimant la GFP et la dsRed sous le contrôle du promoteur du gène *unc-54*. Dans les lignées exprimant les deux rapporteurs sous le contrôle de *unc-54* 3'UTR les protéines ont été observées dans le muscle, comme attendu, avec une colocalisation parfaite. D'autres lignées portant les plasmides avec les délétions $\Delta S1$, $\Delta S2$ ou $\Delta S3$ n'ont en revanche pas présenté de phénotypes de fluorescence différents, y compris pour $\Delta S2$.

Les 15 pb du site 2 comprennent deux sites cibles pour les miARN *mir-1820* et *mir-4811*. Nous nous sommes d'abord intéressés à *mir-4811* car *unc-54* 3'UTR ne présente qu'un seul site cible pour ce miARN alors qu'il en existe deux pour *mir-1820*. Le site supprimé ayant conduit au phénotype variant observé avec le promoteur du gène *lin-28* ne retirait qu'un seul des sites cibles de *mir-1820*. La suppression du site *mir-4811* seulement n'a pas suffi à modifier le profil d'expression de la mCherry, ce qui semble indiquer que le site cible de *mir-1820* serait suffisant. L'absence de variation observée avec le promoteur du gène *unc-54* semble indiquer que les deux sites cibles de *mir-1820* sont nécessaires à la régulation.

La régulation potentiellement exercée par le miARN *mir-1820* a été examinée. Le miARN *mir-1820* est fortement exprimé dans les neurones ADF, CAN, le muscle stomato-intestinal et plus faiblement dans les muscles latéraux. La régulation que nous avons observée sur la construction *Plin-28::mCherry::unc-54* 3'UTR correspond à ces zones d'expression. Le fait que ces régulations ne soient pas observées avec le promoteur du gène *unc-54* peut provenir de la différence d'activité entre les promoteurs ou du fait que les deux sites cibles de *mir-1820* sont nécessaires à la régulation. Une expérience de RT-PCR semi-quantitative ou quantitative permettrait de vérifier cette hypothèse.

Une autre approche que nous n'avons pas essayée serait de cibler le *mir-1820* avec le système CRISPR CAS-9 (Selma Waaijers and Boxem 2014; S. Waaijers et al. 2013). Il serait possible de générer des souches mutantes délétées pour ce miARN et de les croiser avec les lignées transgéniques exprimant un transgène de type *mCherry::unc-54* 3'UTR afin de voir si le phénotype obtenu avec la mutation $\Delta S2$ est reproductible.

2. Vers l'élucidation de la voix de régulation

Les lignées transgéniques obtenues par mutagenèse aléatoire avec l'EMS présentent des phénotypes d'expression variables. Certaines lignées ont montré un changement dans l'expression des rapporteurs uniquement au niveau du muscle ou uniquement au niveau des neurones ADF. Deux des quatre lignées obtenues ne sont pas mutées dans les régions de *mir-1820* ce qui indique que la mutagenèse a atteint d'autres protéines de la voie de régulation. Trois des lignées ont un phénotype et une croissance normale à 20°C ce qui permet de supposer que ces mutations concernent une voie de régulation spécifique. La quatrième lignée présente un phénotype *dumpy* qui pourrait être lié à la différence d'expression des rapporteurs. Les croisements avec les N2 permettront de vérifier s'il existe un lien entre les deux phénotypes. Il est possible que nous ayons atteint des protéines de liaison à l'ARN ou des facteurs de transcription.

Le génome de *C. elegans* comporte plus de 500 gènes qui produisent des protéines possédant un domaine de fixation à l'ARN (M.-H. Lee 2006; Tamburino, Ryder, and Walhout 2013). Ces protéines jouent plusieurs rôles essentiels dans la régulation de l'expression génétique en permettant l'épissage, la poly-adénylation, la stabilité, la localisation et enfin la traduction des ARNm. Néanmoins il reste difficile de déterminer avec précision le rôle de ces RBP qui ont de nombreuses cibles et dont les délétions entraînent l'apparition de phénotype qui découlent de la disfonction de l'ensemble d'un réseau de régulation.

La protéine LIN-28 de *C. elegans* est un gène hétérochronique représentatif de la complexité de ces régulations. La protéine est localisée dans le cytoplasme et possède deux domaines de liaison à l'ARN (Moss, Lee, and Ambros 1997). Elle cible sans doute plusieurs ARNm et ne permet leur expression qu'au bon moment car son ARNm est lui-même régulé par un –ou plusieurs- miARN.

3. Etude du gène *osm-9*

La régulation observée dans les neurones ADF pouvait être une régulation exogène liée à l'activité de régulateurs ciblant d'autres gènes. Nous avons donc cherché si d'autres gènes étaient les cibles de ces deux miARNs et pouvaient donc subir les mêmes régulations que celles observées sur *unc-54* 3'UTR. Nous avons choisi de commencer l'étude du gène *osm-9* exprimé dans quelques neurones dont la paire ADF. La fusion *Posm-9::GFP* a entraîné l'expression du gène dans les neurones OLQ uniquement. Pour l'expression dans les neurones ADF, l'étude de 1997 (Colbert, Smith, and Bargmann 1997) indique qu'elle n'est observée qu'en présence de l'ensemble des séquences de la protéine et du 3'UTR. Les premières observations réalisées au laboratoire indiquent que, seules, les régions promotrices et *osm-9* 3'UTR ne sont pas suffisante pour permettre l'expression dans les neurones ADF. Ces observations sont pour l'instant en accord avec une régulation possible par *mir-1820*, la mutagenèse de *osm-9* 3'UTR permettra de déterminer si le promoteur est actif dans ces neurones et si une régulation post-transcriptionnelle a lieu sur le 3'UTR. Néanmoins, l'expression du rapporteur est limitée aux neurones OLQ et ne s'étend pas aux neurones ADL comme décrit dans l'étude de 1997 (Colbert, Smith, and Bargmann 1997). L'ajout des séquences de certains introns est parfois nécessaire pour permettre l'expression d'un transgène, comme c'est le cas pour le gène *pie-1* (Reese et al. 2000). Si la mutagenèse du site cible de *mir-1820* ne suffit pas son expression dans les neurones ADF, il faudra sans doute ajouter des séquences régulatrices présentes dans la séquence codante.

3. Le choix de la séquence 3'UTR contrôle

Pendant des années la séquence *unc-54* 3'UTR a été utilisée comme région 3'UTR permissive pour l'expression de transgènes. Les données accumulées au cours de ce travail mettent en évidence que cette région peut subir des régulations post-transcriptionnelles même si jusqu'ici le cas ne s'était visiblement pas présenté.

Ces résultats soulignent la difficulté de déterminer une région 3'UTR contrôle exempte de régulation post-transcriptionnelle. La manière dont les cibles des miARN sont déterminées

n'est pas connue et la seule présence d'un site cible dans la région 3'UTR ne se traduit pas forcément par la réalité d'une régulation.

Actuellement, de plus en plus d'équipe remplace la séquence *unc-54* 3'UTR par celle de *let-858* 3'UTR (Maher et al. 2013; Reese et al. 2000; Kai et al. 2013; Grossman, Giurumescu, and Chisholm 2013; Delahaye et al. 2014; Ahier and Jarriault 2014). Le gène *let-858* code pour la nucampholine et présente des similarités avec la protéine eIF-4 gamma (Kelly et al. 1997). La région *let-858* 3'UTR est très courte (122 nucléotides) mais présente aussi 5 sites potentiellement cibles des miARN (*mir-792*, *mir-2218b*, *mir-85*, *mir-4805*). De la même manière que pour le gène *unc-54* exprimé de manière constante au cours du temps, il se pourrait qu'une régulation s'exerce aussi sur cette région dans des cellules particulières.

Une stratégie pour valider que l'UTR contrôle est adapté à l'étude réalisée serait de comparer son comportement par rapport à des versions dont les sites cibles présents dans la séquence seraient mutés.

4. Conclusion

Les travaux réalisés au cours de ma thèse n'ont pas permis de développer un rapporteur unique pour l'étude des régulations post-transcriptionnelle et semble indiquer que ce n'est pas indispensable pour l'étude des régulations post-transcriptionnelles. La principale difficulté réside dans la sélection d'une séquence 3'UTR sur laquelle aucune régulation ne s'exercera. La liaison des différents rapporteurs en une seule construction a tout de même permis de mettre en évidence des phénomènes de régulations post-transcriptionnelles. La découverte de la régulation de la région *unc-54* 3'UTR a été permise grâce à l'usage de deux rapporteurs, ce qui démontre clairement l'intérêt d'étudier les régulations potentielles dans une seule lignée, plutôt que d'en comparer deux. La caractérisation de cette régulation nous a amené à étudier deux miARN, dont *mir-1820* dont l'existence a été vérifiée en 2011 (Jan et al. 2011) mais dont l'expression et les rôles restaient à ce jour non déterminés. Les résultats préliminaires semblent indiquer qu'il joue un rôle dans certains neurones ainsi que dans une partie des muscles. Le miARN *mir-1820* cible potentiellement plus de 4 000 transcrits chez *C. elegans* parmi lesquels 177 sont conservés chez *C. briggsae* (TargetScan).

L'évolution rapide des techniques de séquençages permettent de valider de plus en plus de petits ARN régulateurs dont les rôles et places au sein des réseaux de régulations restent à déterminer. Le développement des techniques de clonage, comme le *chew back annealing*, permet d'obtenir rapidement les plasmides pour générer des lignées transgéniques. La mise au point de méthodes efficace d'édition du génome comme le *Mos*, les TALENs ou le système CRISPR, déjà

optimisés pour *C. elegans* augmente aussi grandement les possibilités d'étudier de manière spécifique le rôle d'un gène ou d'un miARN.

Le développement de ce projet devrait inclure systématiquement deux contrôles, qui ont été faits à posteriori dans cette étude mais ont permis de comprendre le problème sous-jacent avec notre construction de départ, puis la région 3'UTR contrôle. Le premier contrôle à faire est de réaliser la transgénèse conjointe des différents rapporteurs sous le contrôle des mêmes éléments régulateurs afin de valider le dispositif. Par exemple, la cassette RRPT aurait pu être modifiée en remplaçant *unc-54* 3'UTR par un des autres 3'UTR. Ensuite la région contrôle ou témoin devrait être validée en réalisant des expériences de mutagenèse pour voir si le profil d'expression s'en trouve modifié.

L'étude des régulations dans un organisme modèle permet d'avoir un aperçu de la complexité des réseaux de régulation qui permettent le développement d'un organisme, même avec une organisation simple comme *C. elegans*. Les mécanismes biologiques à l'œuvre dans un nématode simple comme *C. elegans* ont souvent été retrouvés chez d'autres organismes et permettent de mieux appréhender le fonctionnement global d'organismes plus complexes.

VI. Bibliographie

- Abrahante, Juan E., Aric L. Daul, Ming Li, Mandy L. Volk, Jason M. Tennessen, Eric A. Miller, and Ann E. Rougvie. 2003. "The *Caenorhabditis Elegans* Hunchback-like Gene *Lin-57/hbl-1* Controls Developmental Time and Is Regulated by MicroRNAs." *Developmental Cell* 4 (5): 625–37. doi:10.1016/S1534-5807(03)00127-8.
- Adenle, A.A., B. Johnsen, and N.J. Szewczyk. 2009. "Review of the Results from the International *C. Elegans* First Experiment (ICE-FIRST)." *Advances in Space Research* 44 (2): 210–16. doi:10.1016/j.asr.2009.04.008.
- Ahier, Arnaud, and Sophie Jarriault. 2014. "Simultaneous Expression of Multiple Proteins Under a Single Promoter in *Caenorhabditis Elegans* via a Versatile 2A-Based Toolkit." *Genetics* 196 (3): 605–13. doi:10.1534/genetics.113.160846.
- Allen, Mary Ann, LaDeana W. Hillier, Robert H. Waterston, and Thomas Blumenthal. 2011. "A Global Analysis of *C. Elegans* Trans-Splicing." *Genome Research* 21 (2): 255–64. doi:10.1101/gr.113811.110.
- Altun, Z.F., L.A. Herndon, C. Crocker, R. Lints, and D.H. Hall. "WormAtlas." <http://www.wormatlas.org>.
- Ambros, Victor. 1989. "A Hierarchy of Regulatory Genes Controls a Larva-to-Adult Developmental Switch in *C. Elegans*." *Cell* 57 (1): 49–57. doi:10.1016/0092-8674(89)90171-2.
- . 2011. "MicroRNAs and Developmental Timing." *Current Opinion in Genetics & Development*, Differentiation and gene regulation, 21 (4): 511–17. doi:10.1016/j.gde.2011.04.003.
- Ambros, Victor, Rosalind C. Lee, Ann Lavanway, Peter T. Williams, and David Jewell. 2003. "MicroRNAs and Other Tiny Endogenous RNAs in *C. Elegans*." *Current Biology* 13 (10): 807–18.
- Antoshechkin, Igor, and Paul W. Sternberg. 2007. "The Versatile Worm: Genetic and Genomic Resources for *Caenorhabditis Elegans* Research." *Nature Reviews Genetics* 8 (7): 518–32. doi:10.1038/nrg2105.
- Avery, O. T., C. M. Macleod, and M. McCarty. 1944. "STUDIES ON THE CHEMICAL NATURE OF THE SUBSTANCE INDUCING TRANSFORMATION OF PNEUMOCOCCAL TYPES: INDUCTION OF TRANSFORMATION BY A DESOXYRIBONUCLEIC ACID FRACTION ISOLATED FROM PNEUMOCOCCUS TYPE III." *The Journal of Experimental Medicine* 79 (2): 137–58.
- Banerjee, A. K. 1980. "5'-Terminal Cap Structure in Eucaryotic Messenger Ribonucleic Acids." *Microbiological Reviews* 44 (2): 175–205.
- Barberis, A., and L. Gaudreau. 1998. "Recruitment of the RNA Polymerase II Holoenzyme and Its Implications in Gene Regulation." *Biological Chemistry* 379 (12): 1397–1405.
- Bartel, David P. 2004. "MicroRNAs: Genomics, Biogenesis, Mechanism, and Function." *Cell* 116 (2): 281–97. doi:10.1016/S0092-8674(04)00045-5.
- Berglund, A.-C., E. Sjolund, G. Ostlund, and E. L. L. Sonnhammer. 2007. "InParanoid 6: Eukaryotic Ortholog Clusters with Inparalogs." *Nucleic Acids Research* 36 (Database): D263–66. doi:10.1093/nar/gkm1020.
- Bernstein, Emily, Amy A. Caudy, Scott M. Hammond, and Gregory J. Hannon. 2001. "Role for a Bidentate Ribonuclease in the Initiation Step of RNA Interference." *Nature* 409 (6818): 363–66. doi:10.1038/35053110.
- Blumenthal, Thomas, Donald Evans, Christopher D. Link, Alessandro Guffanti, Daniel Lawson, Jean Thierry-Mieg, Danielle Thierry-Mieg, et al. 2002. "A Global Analysis of *Caenorhabditis Elegans* Operons." *Nature* 417 (6891): 851–54. doi:10.1038/nature00831.
- Brenner, John L., Benedict J. Kemp, and Allison L. Abbott. 2012. "The Mir-51 Family of microRNAs Functions in Diverse Regulatory Pathways in *Caenorhabditis Elegans*." *PLoS ONE* 7 (5): e37185. doi:10.1371/journal.pone.0037185.
- Brenner, S. 1974. "The Genetics of *Caenorhabditis Elegans*." *Genetics* 77 (1): 71–94.

- Brenner, S., F. Jacob, and M. Meselson. 1961. "An Unstable Intermediate Carrying Information from Genes to Ribosomes for Protein Synthesis." *Nature* 190 (May): 576–81.
- Brenner, Sydney. 1973. "The Genetics of Behaviour." *British Medical Bulletin* 29 (3): 269–71.
- Brody, E., and J. Abelson. 1985. "The 'Spliceosome': Yeast Pre-Messenger RNA Associates with a 40S Complex in a Splicing-Dependent Reaction." *Science (New York, N.Y.)* 228 (4702): 963–67.
- Calin, George A., and Carlo M. Croce. 2006. "MicroRNA Signatures in Human Cancers." *Nature Reviews Cancer* 6 (11): 857–66. doi:10.1038/nrc1997.
- Carthew, Richard W., and Erik J. Sontheimer. 2009. "Origins and Mechanisms of miRNAs and siRNAs." *Cell* 136 (4): 642–55. doi:10.1016/j.cell.2009.01.035.
- Cassada, Randall C., and Richard L. Russell. 1975. "The Dauerlarva, a Post-Embryonic Developmental Variant of the Nematode *Caenorhabditis Elegans*." *Developmental Biology* 46 (2): 326–42. doi:10.1016/0012-1606(75)90109-8.
- Caswell-Chen, E. P., J. Chen, E. E. Lewis, G. W. Douhan, S. A. Nadler, and J. R. Carey. 2005. "Revising the Standard Wisdom of *C. Elegans* Natural History: Ecology of Longevity." *Science of Aging Knowledge Environment: SAGE KE* 2005 (40): pe30. doi:10.1126/sageke.2005.40.pe30.
- Cevec, M., C. Thibaudeau, and J. Plavec. 2010. "NMR Structure of the Let-7 miRNA Interacting with the Site LCS1 of Lin-41 mRNA from *Caenorhabditis Elegans*." *Nucleic Acids Research* 38 (21): 7814–21. doi:10.1093/nar/gkq640.
- Chalfie, M., Y. Tu, G. Euskirchen, W. W. Ward, and D. C. Prasher. 1994. "Green Fluorescent Protein as a Marker for Gene Expression." *Science (New York, N.Y.)* 263 (5148): 802–5.
- Cheung, Benny H. H, Fausto Arellano-Carbajal, Irene Rybicki, and Mario de Bono. 2004. "Soluble Guanylate Cyclases Act in Neurons Exposed to the Body Fluid to Promote *C. Elegans* Aggregation Behavior." *Current Biology* 14 (12): 1105–11. doi:10.1016/j.cub.2004.06.027.
- Clark, S. G., X. Lu, and H. R. Horvitz. 1994. "The *Caenorhabditis Elegans* Locus Lin-15, a Negative Regulator of a Tyrosine Kinase Signaling Pathway, Encodes Two Different Proteins." *Genetics* 137 (4): 987–97.
- Coates, Juliet C., and Mario de Bono. 2002. "Antagonistic Pathways in Neurons Exposed to Body Fluid Regulate Social Feeding in *Caenorhabditis Elegans*." *Nature* 419 (6910): 925–29. doi:10.1038/nature01170.
- Cochella, Luisa, and Oliver Hobert. 2012. "Embryonic Priming of a miRNA Locus Predetermines Postmitotic Neuronal Left/Right Asymmetry in *C. Elegans*." *Cell* 151 (6): 1229–42. doi:10.1016/j.cell.2012.10.049.
- Colbert, Heather A, and Cornelia I Bargmann. 1995. "Odorant-Specific Adaptation Pathways Generate Olfactory Plasticity in *C. Elegans*." *Neuron* 14 (4): 803–12. doi:10.1016/0896-6273(95)90224-4.
- Colbert, Heather A., Tracy L. Smith, and Cornelia I. Bargmann. 1997. "OSM-9, A Novel Protein with Structural Similarity to Channels, Is Required for Olfaction, Mechanosensation, and Olfactory Adaptation in *Caenorhabditis Elegans*." *The Journal of Neuroscience* 17 (21): 8259–69.
- Consortium, The *C. elegans* Sequencing. 1998. "Genome Sequence of the Nematode *C. Elegans*: A Platform for Investigating Biology." *Science* 282 (5396): 2012–18. doi:10.1126/science.282.5396.2012.
- Cornes, Eric, Cécile A. L. Quéré, Rosina Giordano-Santini, and Denis Dupuy. 2014. "Applying Antibiotic Selection Markers for Nematode Genetics." *Methods, C. Elegans Methods*, 68 (3): 403–8. doi:10.1016/j.ymeth.2014.04.016.
- Cosma, Maria Pia, Tomoyuki Tanaka, and Kim Nasmyth. 1999. "Ordered Recruitment of Transcription and Chromatin Remodeling Factors to a Cell Cycle- and Developmentally Regulated Promoter." *Cell* 97 (3): 299–311. doi:10.1016/S0092-8674(00)80740-0.

- Darnell, J. E., R. Wall, and R. J. Tushinski. 1971. "An Adenylic Acid-Rich Sequence in Messenger RNA of HeLa Cells and Its Possible Relationship to Reiterated Sites in DNA." *Proceedings of the National Academy of Sciences of the United States of America* 68 (6): 1321.
- Delahaye, Jared L., Olivia K. Foster, Annalise Vine, Daniel S. Saxton, Thomas P. Curtin, Hannah Somhegyi, Rebecca Salesky, and Greg J. Hermann. 2014. "Caenorhabditis Elegans HOPS and CCZ-1 Mediate Trafficking to Lysosome-Related Organelles Independently of RAB-7 and SAND-1." *Molecular Biology of the Cell* 25 (7): 1073–96. doi:10.1091/mbc.E13-09-0521.
- Denli, Ahmet M., Bastiaan B. J. Tops, Ronald H. A. Plasterk, René F. Ketting, and Gregory J. Hannon. 2004. "Processing of Primary microRNAs by the Microprocessor Complex." *Nature* 432 (7014): 231–35. doi:10.1038/nature03049.
- Dibb, Nick J., Ichiro N. Maruyama, Michael Krause, and Jonathan Karn. 1989. "Sequence Analysis of the Complete Caenorhabditis Elegans Myosin Heavy Chain Gene Family." *Journal of Molecular Biology* 205 (3): 603–13. doi:10.1016/0022-2836(89)90229-5.
- Doench, John G., Christian P. Petersen, and Phillip A. Sharp. 2003. "siRNAs Can Function as miRNAs." *Genes & Development* 17 (4): 438–42. doi:10.1101/gad.1064703.
- Dougherty, E. C., and H. G. Calhoun. 1948. "Possible Significance of Free-Living Nematodes in Genetic Research." *Nature* 161 (4079): 29.
- Dougherty, E. C., and E. L. Hansen. 1956. "Axenic Cultivation of Caenorhabditis Briggsae (Nematoda: Rhabditidae). V. Maturation on Synthetic Media." *Proceedings of the Society for Experimental Biology and Medicine. Society for Experimental Biology and Medicine (New York, N.Y.)* 93 (2): 223–27.
- Duursma, Anja M., Martijn Kedde, Mariette Schrier, Carlos le Sage, and Reuven Agami. 2008. "miR-148 Targets Human DNMT3b Protein Coding Region." *RNA (New York, N.Y.)* 14 (5): 872–77. doi:10.1261/rna.972008.
- Edmonds, Mary. 2002. "A History of Poly A Sequences: From Formation to Factors to Function." In *Progress in Nucleic Acid Research and Molecular Biology*, Volume 71:285–389. Academic Press. <http://www.sciencedirect.com/science/article/pii/S0079660302710465>.
- Edmonds, Mary, Maurice H. Vaughan, Jr, and Hiroshi Nakazato. 1971. "Polyadenylic Acid Sequences in the Heterogeneous Nuclear RNA and Rapidly-Labeled Polyribosomal RNA of HeLa Cells: Possible Evidence for a Precursor Relationship." *Proceedings of the National Academy of Sciences of the United States of America* 68 (6): 1336.
- Ellis, Hilary M., and H. Robert Horvitz. 1986. "Genetic Control of Programmed Cell Death in the Nematode C. Elegans." *Cell* 44 (6): 817–29. doi:10.1016/0092-8674(86)90004-8.
- Epstein, H. F., R. H. Waterston, and S. Brenner. 1974. "A Mutant Affecting the Heavy Chain of Myosin in Caenorhabditis Elegans." *Journal of Molecular Biology* 90 (2): 291–300.
- Evans, Thomas. 2006. "Transformation and Microinjection." *WormBook*. doi:10.1895/wormbook.1.108.1.
- Fatt, H. V., and E. C. Dougherty. 1963. "Genetic Control of Differential Heat Tolerance in Two Strains of the Nematode Caenorhabditis Elegans." *Science (New York, N.Y.)* 141 (3577): 266–67. doi:10.1126/science.141.3577.266.
- Fire, A., D. Albertson, S. W. Harrison, and D. G. Moerman. 1991. "Production of Antisense RNA Leads to Effective and Specific Inhibition of Gene Expression in C. Elegans Muscle." *Development (Cambridge, England)* 113 (2): 503–14.
- Fire, A., S. W. Harrison, and D. Dixon. 1990. "A Modular Set of lacZ Fusion Vectors for Studying Gene Expression in Caenorhabditis Elegans." *Gene* 93 (2): 189–98.
- Fire, Andrew. 1986. "Integrative Transformation of Caenorhabditis Elegans." *The EMBO Journal* 5 (10): 2673–80.
- Fire, Andrew, SiQun Xu, Mary K. Montgomery, Steven A. Kostas, Samuel E. Driver, and Craig C. Mello. 1998. "Potent and Specific Genetic Interference by Double-Stranded RNA in Caenorhabditis Elegans." *Nature* 391 (6669): 806–11. doi:10.1038/35888.

- Fire, A., and R. H. Waterston. 1989. "Proper Expression of Myosin Genes in Transgenic Nematodes." *The EMBO Journal* 8 (11): 3419.
- Firnhaber, Christopher, and Marc Hammarlund. 2013. "Neuron-Specific Feeding RNAi in *C. Elegans* and Its Use in a Screen for Essential Genes Required for GABA Neuron Function." *PLoS Genetics* 9 (11). doi:10.1371/journal.pgen.1003921.
- Flibotte, Stephane, Mark L. Edgley, Iasha Chaudhry, Jon Taylor, Sarah E. Neil, Aleksandra Rogula, Rick Zapf, et al. 2010. "Whole-Genome Profiling of Mutagenesis in *Caenorhabditis Elegans*." *Genetics* 185 (2): 431–41. doi:10.1534/genetics.110.116616.
- Forman, Joshua J., Aster Legesse-Miller, and Hilary A. Collier. 2008. "A Search for Conserved Sequences in Coding Regions Reveals That the Let-7 microRNA Targets Dicer within Its Coding Sequence." *Proceedings of the National Academy of Sciences of the United States of America* 105 (39): 14879–84. doi:10.1073/pnas.0803230105.
- Friedland, Ari E, Yonatan B Tzur, Kevin M Esvelt, Monica P Colaiácovo, George M Church, and John A Calarco. 2013. "Heritable Genome Editing in *C. Elegans* via a CRISPR-Cas9 System." *Nature Methods* 10 (8): 741–43. doi:10.1038/nmeth.2532.
- Gengyo-Ando, Keiko, and Shohei Mitani. 2000. "Characterization of Mutations Induced by Ethyl Methanesulfonate, UV, and Trimethylpsoralen in the Nematode *Caenorhabditis Elegans*." *Biochemical and Biophysical Research Communications* 269 (1): 64–69. doi:10.1006/bbrc.2000.2260.
- Gibson, Daniel G. 2011. "Enzymatic Assembly of Overlapping DNA Fragments." In *Methods in Enzymology*, 498:349–61. Elsevier. <http://linkinghub.elsevier.com/retrieve/pii/B9780123851208000152>.
- Gibson, Daniel G., Hamilton O. Smith, Clyde A. Hutchison Iii, J. Craig Venter, and Chuck Merryman. 2010. "Chemical Synthesis of the Mouse Mitochondrial Genome." *Nature Methods* 7 (11): 901–3. doi:10.1038/nmeth.1515.
- Gibson, Daniel G., Lei Young, Ray-Yuan Chuang, J. Craig Venter, Clyde A. Hutchison, and Hamilton O. Smith. 2009. "Enzymatic Assembly of DNA Molecules up to Several Hundred Kilobases." *Nature Methods* 6 (5): 343–45. doi:10.1038/nmeth.1318.
- Gilbert, W. 1978. "Why Genes in Pieces?" *Nature* 271 (5645): 501.
- Giordano-Santini, Rosina, Stuart Milstein, Nenad Svrzikapa, Domena Tu, Robert Johnsen, David Baillie, Marc Vidal, and Denis Dupuy. 2010. "An Antibiotic Selection Marker for Nematode Transgenesis." *Nature Methods* 7 (9): 721–23. doi:10.1038/nmeth.1494.
- Granato, M., H. Schnabel, and R. Schnabel. 1994. "Pha-1, a Selectable Marker for Gene Transfer in *C. Elegans*." *Nucleic Acids Research* 22 (9): 1762.
- Greiss, Sebastian, and Jason W. Chin. 2011. "Expanding the Genetic Code of an Animal." *Journal of the American Chemical Society* 133 (36): 14196–99. doi:10.1021/ja2054034.
- Grey, Finn, Rebecca Tirabassi, Heather Meyers, Guanming Wu, Shannon McWeeney, Lauren Hook, and Jay A. Nelson. 2010. "A Viral microRNA Down-Regulates Multiple Cell Cycle Genes through mRNA 5'UTRs." *PLoS Pathog* 6 (6): e1000967. doi:10.1371/journal.ppat.1000967.
- Großhans, Helge, Ted Johnson, Kristy L. Reinert, Mark Gerstein, and Frank J. Slack. 2005. "The Temporal Patterning MicroRNA Let-7 Regulates Several Transcription Factors at the Larval to Adult Transition in *C. Elegans*." *Developmental Cell* 8 (3): 321–30. doi:10.1016/j.devcel.2004.12.019.
- Grossman, Emily N., Claudiu A. Giurumescu, and Andrew D. Chisholm. 2013. "Mechanisms of Ephrin Receptor Protein Kinase-Independent Signaling in Amphid Axon Guidance in *Caenorhabditis Elegans*." *Genetics* 195 (3): 899–913. doi:10.1534/genetics.113.154393.
- Grosswendt, Stefanie, Andrei Filipchuk, Mark Manzano, Filippos Klironomos, Marcel Schilling, Margareta Herzog, Eva Gottwein, and Nikolaus Rajewsky. 2014. "Unambiguous Identification of miRNA:Target Site Interactions by Different Types of Ligation Reactions." *Molecular Cell* 54 (6): 1042–54. doi:10.1016/j.molcel.2014.03.049.

- Gwizdek, Carole, Batool Ossareh-Nazari, Amy M. Brownawell, Alain Doglio, Edouard Bertrand, Ian G. Macara, and Catherine Dargemont. 2003. "Exportin-5 Mediates Nuclear Export of Minihelix-Containing RNAs." *The Journal of Biological Chemistry* 278 (8): 5505–8. doi:10.1074/jbc.C200668200.
- Hamilton, Andrew J., and David C. Baulcombe. 1999. "A Species of Small Antisense RNA in Posttranscriptional Gene Silencing in Plants." *Science* 286 (5441): 950–52. doi:10.1126/science.286.5441.950.
- Han, Jinju, Yoontae Lee, Kyu-Hyun Yeom, Young-Kook Kim, Hua Jin, and V. Narry Kim. 2004. "The Drosha-DGCR8 Complex in Primary microRNA Processing." *Genes & Development* 18 (24): 3016–27. doi:10.1101/gad.1262504.
- He, Lin, and Gregory J. Hannon. 2004. "MicroRNAs: Small RNAs with a Big Role in Gene Regulation." *Nature Reviews Genetics* 5 (7): 522–31. doi:10.1038/nrg1379.
- Hirsh, D., K. J. Kemphues, D. T. Stinchcomb, and R. Jefferson. 1985. "Genes Affecting Early Development in *Caenorhabditis Elegans*." *Cold Spring Harbor Symposia on Quantitative Biology* 50: 69–78.
- Hodgkin, Jonathan A., and Sydney Brenner. 1977. "Mutations Causing Transformation of Sexual Phenotype in the Nematode *Caenorhabditis Elegans*." *Genetics* 86 (2): 275–87.
- Honda, S., and H. F. Epstein. 1990. "Modulation of Muscle Gene Expression in *Caenorhabditis Elegans*: Differential Levels of Transcripts, mRNAs, and Polypeptides for Thick Filament Proteins during Nematode Development." *Proceedings of the National Academy of Sciences* 87 (3): 876–80. doi:10.1073/pnas.87.3.876.
- Honda, Yoko, Shuji Honda, Marco Narici, and Nathaniel J. Szewczyk. 2014. "Spaceflight and Ageing: Reflecting on *Caenorhabditis Elegans* in Space." *Gerontology* 60 (2): 138–42. doi:10.1159/000354772.
- House, Amy E., and Kristen W. Lynch. 2008. "Regulation of Alternative Splicing: More than Just the ABCs." *Journal of Biological Chemistry* 283 (3): 1217–21. doi:10.1074/jbc.R700031200.
- Hsieh, Yi-Wen, Chieh Chang, and Chiou-Fen Chuang. 2012. "The MicroRNA Mir-71 Inhibits Calcium Signaling by Targeting the TIR-1/Sarm1 Adaptor Protein to Control Stochastic L/R Neuronal Asymmetry in *C. Elegans*." *PLoS Genet* 8 (8): e1002864. doi:10.1371/journal.pgen.1002864.
- Huang, X. Y., and D. Hirsh. 1989. "A Second Trans-Spliced RNA Leader Sequence in the Nematode *Caenorhabditis Elegans*." *Proceedings of the National Academy of Sciences* 86 (22): 8640–44.
- Hunt-Newbury, Rebecca, Ryan Viveiros, Robert Johnsen, Allan Mah, Dina Anastas, Lily Fang, Erin Halfnight, et al. 2007. "High-Throughput In Vivo Analysis of Gene Expression in *Caenorhabditis Elegans*." *PLoS Biology* 5 (9): e237. doi:10.1371/journal.pbio.0050237.
- Hwang, Hun-Way, Erik A. Wentzel, and Joshua T. Mendell. 2007. "A Hexanucleotide Element Directs MicroRNA Nuclear Import." *Science* 315 (5808): 97–100. doi:10.1126/science.1136235.
- Jafari, Gholamali, Yusu Xie, Andrey Kullyev, Bin Liang, and Ji Ying Sze. 2011. "Regulation of Extrasynaptic 5-HT by Serotonin Reuptake Transporter Function in 5-HT-Absorbing Neurons Underscores Adaptation Behavior in *Caenorhabditis Elegans*." *The Journal of Neuroscience* 31 (24): 8948–57. doi:10.1523/JNEUROSCI.1692-11.2011.
- Jan, Calvin H., Robin C. Friedman, J. Graham Ruby, and David P. Bartel. 2011. "Formation, Regulation and Evolution of *Caenorhabditis Elegans* 3'UTRs." *Nature* 469 (7328): 97–101. doi:10.1038/nature09616.
- Johnston, Robert J., and Oliver Hobert. 2003. "A microRNA Controlling Left/right Neuronal Asymmetry in *Caenorhabditis Elegans*." *Nature* 426 (6968): 845–49. doi:10.1038/nature02255.
- Jones, Donald, Emily Crowe, Tracy A. Stevens, and E. Peter M. Candido. 2002. "Functional and Phylogenetic Analysis of the Ubiquitylation System in *Caenorhabditis Elegans*:

- Ubiquitin-Conjugating Enzymes, Ubiquitin-Activating Enzymes, and Ubiquitin-like Proteins.” *Genome Biology* 3 (1): RESEARCH0002.
- Jorgensen, Erik M., and Susan E. Mango. 2002. “The Art and Design of Genetic Screens: *Caenorhabditis Elegans*.” *Nature Reviews Genetics* 3 (5): 356–69. doi:10.1038/nrg794.
- Kai, Zoya S., Emily F. Finnegan, Stacey Huang, and Amy E. Pasquinelli. 2013. “Multiple Cis-Elements and Trans-Acting Factors Regulate Dynamic Spatio-Temporal Transcription of *Let-7* in *C. Elegans*.” *Developmental Biology* 374 (1): 223–33. doi:10.1016/j.ydbio.2012.11.021.
- Kelly, William G., SiQun Xu, Mary K. Montgomery, and Andrew Fire. 1997. “Distinct Requirements for Somatic and Germline Expression of a Generally Expressed *Caenorhabditis Elegans* Gene.” *Genetics* 146 (1): 227–38.
- Kim, V. Narry, Jinju Han, and Mikiko C. Siomi. 2009. “Biogenesis of Small RNAs in Animals.” *Nature Reviews Molecular Cell Biology* 10 (2): 126–39. doi:10.1038/nrm2632.
- Kornberg, A., J. F. Scott, and L. L. Bertsch. 1978. “ATP Utilization by Rep Protein in the Catalytic Separation of DNA Strands at a Replicating Fork.” *Journal of Biological Chemistry* 253 (9): 3298–3304.
- Korswagen, H. C., J. H. Park, Y. Ohshima, and R. H. Plasterk. 1997. “An Activating Mutation in a *Caenorhabditis Elegans* Gs Protein Induces Neural Degeneration.” *Genes & Development* 11 (12): 1493–1503. doi:10.1101/gad.11.12.1493.
- Kramer, J. M., R. P. French, E. C. Park, and J. J. Johnson. 1990. “The *Caenorhabditis Elegans* *Rol-6* Gene, Which Interacts with the *Sqt-1* Collagen Gene to Determine Organismal Morphology, Encodes a Collagen.” *Molecular and Cellular Biology* 10 (5): 2081–89. doi:10.1128/MCB.10.5.2081.
- Krek, Azra, Dominic Grün, Matthew N. Poy, Rachel Wolf, Lauren Rosenberg, Eric J. Epstein, Philip MacMenamin, et al. 2005. “Combinatorial microRNA Target Predictions.” *Nature Genetics* 37 (5): 495–500. doi:10.1038/ng1536.
- Kuersten, S., K. Lea, M. MacMorris, J. Spieth, and T. Blumenthal. 1997. “Relationship between 3’ End Formation and SL2-Specific Trans-Splicing in Polycistronic *Caenorhabditis Elegans* Pre-mRNA Processing.” *RNA* 3 (3): 269–78.
- Kulkarni, Madhura, and Harold E. Smith. 2008. “E1 Ubiquitin-Activating Enzyme UBA-1 Plays Multiple Roles throughout *C. Elegans* Development.” *PLoS Genet* 4 (7): e1000131. doi:10.1371/journal.pgen.1000131.
- Kumar, Pankaj, Jordan Anaya, Suresh B. Mudunuri, and Anindya Dutta. 2014. “Meta-Analysis of tRNA Derived RNA Fragments Reveals That They Are Evolutionarily Conserved and Associate with AGO Proteins to Recognize Specific RNA Targets.” *BMC Biology* 12 (1): 78. doi:10.1186/PREACCEPT-5867533061403216.
- Lall, Sabbi, Dominic Grün, Azra Krek, Kevin Chen, Yi-Lu Wang, Colin N. Dewey, Pranidhi Sood, et al. 2006. “A Genome-Wide Map of Conserved MicroRNA Targets in *C. Elegans*.” *Current Biology* 16 (5): 460–71. doi:10.1016/j.cub.2006.01.050.
- Lee, I., S. S. Ajay, J. I. Yook, H. S. Kim, S. H. Hong, N. H. Kim, S. M. Dhanasekaran, A. M. Chinnaiyan, and B. D. Athey. 2009. “New Class of microRNA Targets Containing Simultaneous 5’-UTR and 3’-UTR Interaction Sites.” *Genome Research* 19 (7): 1175–83. doi:10.1101/gr.089367.108.
- Lee, Min-Ho. 2006. “RNA-Binding Proteins.” *WormBook*. doi:10.1895/wormbook.1.79.1.
- Lee, Rosalind C., Rhonda L. Feinbaum, and Victor Ambros. 1993. “The *C. Elegans* Heterochronic Gene *Lin-4* Encodes Small RNAs with Antisense Complementarity to *Lin-14*.” *Cell* 75 (5): 843–54. doi:10.1016/0092-8674(93)90529-Y.
- Lee, Tong Ihn, and Richard A. Young. 2000. “Transcription of Eukaryotic Protein-Coding Genes.” *Annual Review of Genetics* 34 (1): 77–137. doi:10.1146/annurev.genet.34.1.77.
- Lee, Yoontae, Chiyoung Ahn, Jinju Han, Hyounjeong Choi, Jaekwang Kim, Jeongbin Yim, Junho Lee, et al. 2003. “The Nuclear RNase III Droscha Initiates microRNA Processing.” *Nature* 425 (6956): 415–19. doi:10.1038/nature01957.

- Lewis, Benjamin P., Christopher B. Burge, and David P. Bartel. 2005. "Conserved Seed Pairing, Often Flanked by Adenosines, Indicates That Thousands of Human Genes Are MicroRNA Targets." *Cell* 120 (1): 15–20. doi:10.1016/j.cell.2004.12.035.
- Lewis, B. P., R. E. Green, and S. E. Brenner. 2003. "Evidence for the Widespread Coupling of Alternative Splicing and Nonsense-Mediated mRNA Decay in Humans." *Proceedings of the National Academy of Sciences* 100 (1): 189–92. doi:10.1073/pnas.0136770100.
- Lewis, E. B., and F. Bacher. 1968. "Method of Feeding Ethyl Methane Sulfonate (EMS) to *Drosophila* Males." *Dros. Inf. Serv* 43: 193.
- Liao, Jian-You, Li-Ming Ma, Yan-Hua Guo, Yu-Chan Zhang, Hui Zhou, Peng Shao, Yue-Qin Chen, and Liang-Hu Qu. 2010. "Deep Sequencing of Human Nuclear and Cytoplasmic Small RNAs Reveals an Unexpectedly Complex Subcellular Distribution of miRNAs and tRNA 3' Trailers." Edited by Shuang-yong Xu. *PLoS ONE* 5 (5): e10563. doi:10.1371/journal.pone.0010563.
- Li, Dirui, and Ming Wang. 2012. "Construction of a Bicistronic Vector for the Co-Expression of Two Genes in *Caenorhabditis Elegans* Using a Newly Identified IRES." *BioTechniques* 52 (3): 173–76. doi:10.2144/000113821.
- Lindbo, J. A., L. Silva-Rosales, W. M. Proebsting, and W. G. Dougherty. 1993. "Induction of a Highly Specific Antiviral State in Transgenic Plants: Implications for Regulation of Gene Expression and Virus Resistance." *The Plant Cell Online* 5 (12): 1749–59. doi:10.1105/tpc.5.12.1749.
- Lints, R., and D.H. Hall. 2009. "Reproductive system, germ line." In *WormAtlas*. <http://www.wormatlas.org>.
- Liu, Pengpeng, Lijiang Long, Kai Xiong, Bo Yu, Nannan Chang, Jing-Wei Xiong, Zuoyan Zhu, and Dong Liu. 2014. "Heritable/conditional Genome Editing in *C. Elegans* Using a CRISPR-Cas9 Feeding System." *Cell Research* 24 (7): 886–89. doi:10.1038/cr.2014.73.
- Li, Zhou Fang, Yi Min Liang, Pui Ngan Lau, Wei Shen, Dai Kui Wang, Wing Tai Cheung, Chun Jason Xue, Lit Man Poon, and Yun Wah Lam. 2013. "Dynamic Localisation of Mature MicroRNAs in Human Nucleoli Is Influenced by Exogenous Genetic Materials." Edited by Sebastien Pfeffer. *PLoS ONE* 8 (8): e70869. doi:10.1371/journal.pone.0070869.
- Lohman, Timothy M., Eric J. Tomko, and Colin G. Wu. 2008. "Non-Hexameric DNA Helicases and Translocases: Mechanisms and Regulation." *Nature Reviews Molecular Cell Biology* 9 (5): 391–401. doi:10.1038/nrm2394.
- Loveless, A. 1959. "The Influence of Radiomimetic Substances on Deoxyribonucleic Acid Synthesis and Function Studied in *Escherichia coli*/Phage Systems. III. Mutation of T2 Bacteriophage as a Consequence of Alkylation in Vitro: The Uniqueness of Ethylation." *Proceedings of the Royal Society of London. Series B - Biological Sciences* 150 (941): 497–508. doi:10.1098/rspb.1959.0038.
- Lund, Elsebet, Stephan Güttinger, Angelo Calado, James E. Dahlberg, and Ulrike Kutay. 2004. "Nuclear Export of MicroRNA Precursors." *Science* 303 (5654): 95–98. doi:10.1126/science.1090599.
- MacLeod, Alexander R., Jonathan Karn, and Sydney Brenner. 1981. "Molecular Analysis of the Unc-54 Myosin Heavy-Chain Gene of *Caenorhabditis Elegans*." *Nature* 291 (5814): 386–90. doi:10.1038/291386a0.
- Maduro, M., and D. Pilgrim. 1995. "Identification and Cloning of Unc-119, a Gene Expressed in the *Caenorhabditis Elegans* Nervous System." *Genetics* 141 (3): 977.
- Maher, Kathryn N., Aishwarya Swaminathan, Parth Patel, and Daniel L. Chase. 2013. "A Novel Strategy for Cell-Autonomous Gene Knockdown in *Caenorhabditis Elegans* Defines a Cell-Specific Function for the G-Protein Subunit GOA-1." *Genetics* 194 (2): 363–73. doi:10.1534/genetics.113.149724.

- Mangone, Marco, Philip MacMenamin, Charles Zegar, Fabio Piano, and Kristin C. Gunsalus. 2007. "UTRome.org: A Platform for 3'UTR Biology in *C. Elegans*." *Nucleic Acids Research* 36 (suppl 1): D57–62. doi:10.1093/nar/gkm946.
- Mangone, M., A. P. Manoharan, D. Thierry-Mieg, J. Thierry-Mieg, T. Han, S. D. Mackowiak, E. Mis, et al. 2010. "The Landscape of *C. Elegans* 3'UTRs." *Science* 329 (5990): 432–35. doi:10.1126/science.1191244.
- Mansfield, Jennifer H., Brian D. Harfe, Robert Nissen, John Obenauer, Jalagani Srineel, Aadel Chaudhuri, Raphael Farzan-Kashani, et al. 2004. "MicroRNA-Responsive 'Sensor' Transgenes Uncover Hox-like and Other Developmentally Regulated Patterns of Vertebrate microRNA Expression." *Nature Genetics* 36 (10): 1079–83. doi:10.1038/ng1421.
- Maragkakis, Manolis, Panagiotis Alexiou, Giorgio L. Papadopoulos, Martin Reczko, Theodore Dalamagas, George Giannopoulos, George Goumas, et al. 2009. "Accurate microRNA Target Prediction Correlates with Protein Repression Levels." *BMC Bioinformatics* 10 (1): 295. doi:10.1186/1471-2105-10-295.
- Martinez, Natalia J., Maria C. Ow, M. Inmaculada Barrasa, Molly Hammell, Reynaldo Sequerra, Lynn Doucette-Stamm, Frederick P. Roth, Victor R. Ambros, and Albertha J. M. Walhout. 2008. "A *C. Elegans* Genome-Scale microRNA Network Contains Composite Feedback Motifs with High Flux Capacity." *Genes & Development* 22 (18): 2535–49. doi:10.1101/gad.1678608.
- Maupas, E. 1901. "Modes et formes de reproduction des nématodes." In *Archives de zoologie expérimentale et générale.*, 463–623. Centre national de la recherche scientifique (Paris).
- Maxam, A. M., and W. Gilbert. 1977. "A New Method for Sequencing DNA." *Proceedings of the National Academy of Sciences* 74 (2): 560–64.
- Mello, Craig C., James M. Kramer, Dan Stinchcomb, and Victor Ambros. 1991. "Efficient Gene Transfer in *C. Elegans*: Extrachromosomal Maintenance and Integration of Transforming Sequences." *The EMBO Journal* 10 (12): 3959.
- Melters, Daniël P., Leocadia V. Paliulis, Ian F. Korf, and Simon W. L. Chan. 2012. "Holocentric Chromosomes: Convergent Evolution, Meiotic Adaptations, and Genomic Analysis." *Chromosome Research: An International Journal on the Molecular, Supramolecular and Evolutionary Aspects of Chromosome Biology* 20 (5): 579–93. doi:10.1007/s10577-012-9292-1.
- Merritt, Christopher, Dominique Rasoloson, Darae Ko, and Geraldine Seydoux. 2008. "3' UTRs Are the Primary Regulators of Gene Expression in the *C. Elegans* Germline." *Current Biology* 18 (19): 1476–82. doi:10.1016/j.cub.2008.08.013.
- Miska, Eric A, Ezequiel Alvarez-Saavedra, Allison L Abbott, Nelson C Lau, Andrew B Hellman, Shannon M McGonagle, David P Bartel, Victor R Ambros, and H. Robert Horvitz. 2007. "Most *Caenorhabditis Elegans* microRNAs Are Individually Not Essential for Development or Viability." *PLoS Genetics* 3 (12). doi:10.1371/journal.pgen.0030215.
- Moore, Melissa J. 2005. "From Birth to Death: The Complex Lives of Eukaryotic mRNAs." *Science* 309 (5740): 1514–18. doi:10.1126/science.1111443.
- Morand, S., M. J. Wilson, and D. M. Glen. 2004. "Nematodes (Nematoda) Parasitic in Terrestrial Gastropods." In *Natural Enemies of Terrestrial Molluscs*, edited by G. M. Barker, 525–57. Wallingford: CABI. <http://www.cabi.org/cabebooks/ebook/20043115152>.
- Moss, Eric G., Rosalind C. Lee, and Victor Ambros. 1997. "The Cold Shock Domain Protein LIN-28 Controls Developmental Timing in *C. Elegans* and Is Regulated by the Lin-4 RNA." *Cell* 88 (5): 637–46.
- Nigon, Victor. 1949. "Les modalités de la reproduction et le déterminisme du sexe chez quelques nématodes libres." In *Annales de Sciences Naturelles*, 1–132.

- Orphanides, George, and Danny Reinberg. 2002. "A Unified Theory of Gene Expression." *Cell* 108 (4): 439–51. doi:10.1016/S0092-8674(02)00655-4.
- Palaima, Elizabeth, Nancy Leymarie, Dave Stroud, Rahman M. Mizanur, Jonathan Hodgkin, Maria J. Gravato-Nobre, Catherine E. Costello, and John F. Cipollo. 2010. "The *Caenorhabditis Elegans* Bus-2 Mutant Reveals a New Class of O-Glycans Affecting Bacterial Resistance." *The Journal of Biological Chemistry* 285 (23): 17662–72. doi:10.1074/jbc.M109.065433.
- Park, Chang Won, Yan Zeng, Xiaoxiao Zhang, Subbaya Subramanian, and Clifford J. Steer. 2010. "Mature microRNAs Identified in Highly Purified Nuclei from HCT116 Colon Cancer Cells." *RNA Biology* 7 (5): 606–14. doi:10.4161/rna.7.5.13215.
- Pasquinelli, Amy E., Brenda J. Reinhart, Frank Slack, Mark Q. Martindale, Mitzi I. Kuroda, Betsy Maller, David C. Hayward, et al. 2000. "Conservation of the Sequence and Temporal Expression of Let-7 Heterochronic Regulatory RNA." *Nature* 408 (6808): 86–89. doi:10.1038/35040556.
- Paulsen, J. E., E. E. Capowski, and S. Strome. 1995. "Phenotypic and Molecular Analysis of Mes-3, a Maternal-Effect Gene Required for Proliferation and Viability of the Germ Line in *C. Elegans*." *Genetics* 141 (4): 1383–98.
- Pfleger, Brian F., Douglas J. Pitera, Christina D. Smolke, and Jay D. Keasling. 2006. "Combinatorial Engineering of Intergenic Regions in Operons Tunes Expression of Multiple Genes." *Nature Biotechnology* 24 (8): 1027–32. doi:10.1038/nbt1226.
- Pillai, Ramesh S., Suvendra N. Bhattacharyya, Caroline G. Artus, Tabea Zoller, Nicolas Cougot, Eugenia Basyuk, Edouard Bertrand, and Witold Filipowicz. 2005. "Inhibition of Translational Initiation by Let-7 MicroRNA in Human Cells." *Science* 309 (5740): 1573–76. doi:10.1126/science.1115079.
- Praitis, Vida, Elizabeth Casey, David Collar, and Judith Austin. 2001. "Creation of Low-Copy Integrated Transgenic Lines in *Caenorhabditis Elegans*." *Genetics* 157 (3): 1217–26.
- Radman, Inja, Sebastian Greiss, and Jason W. Chin. 2013. "Efficient and Rapid *C. Elegans* Transgenesis by Bombardment and Hygromycin B Selection." *PLoS ONE* 8 (10): e76019. doi:10.1371/journal.pone.0076019.
- Reece-Hoyes, John S., Jane Shingles, Denis Dupuy, Christian A. Grove, Albertha J. M. Walhout, Marc Vidal, and Ian A. Hope. 2007. "Insight into Transcription Factor Gene Duplication from *Caenorhabditis Elegans* Promoterome-Driven Expression Patterns." *BMC Genomics* 8: 27. doi:10.1186/1471-2164-8-27.
- Reese, Kimberly J., Melanie A. Dunn, James A. Waddle, and Geraldine Seydoux. 2000. "Asymmetric Segregation of PIE-1 in *C. Elegans* Is Mediated by Two Complementary Mechanisms That Act through Separate PIE-1 Protein Domains." *Molecular Cell* 6 (2): 445–55. doi:10.1016/S1097-2765(00)00043-5.
- Rehmsmeier, Marc, Peter Steffen, Matthias Höchsmann, and Robert Giegerich. 2004. "Fast and Effective Prediction of microRNA/target Duplexes." *RNA* 10 (10): 1507–17. doi:10.1261/rna.5248604.
- Rigoutsos, Isidore. 2009. "New Tricks for Animal MicroRNAs: Targeting of Amino Acid Coding Regions at Conserved and Nonconserved Sites." *Cancer Research* 69 (8): 3245–48. doi:10.1158/0008-5472.CAN-09-0352.
- Robert, Valerie, and Jean-Louis Bessereau. 2007. "Targeted Engineering of the *Caenorhabditis Elegans* Genome Following Mos1-Triggered Chromosomal Breaks." *The EMBO Journal* 26 (1): 170–83. doi:10.1038/sj.emboj.7601463.
- Roeder, R. G., and W. J. Rutter. 1969. "Multiple Forms of DNA-Dependent RNA Polymerase in Eukaryotic Organisms." *Nature* 224 (5216): 234–37.
- Ruby, J. Graham, Calvin Jan, Christopher Player, Michael J. Axtell, William Lee, Chad Nusbaum, Hui Ge, and David P. Bartel. 2006. "Large-Scale Sequencing Reveals 21U-RNAs and Additional MicroRNAs and Endogenous siRNAs in *C. Elegans*." *Cell* 127 (6): 1193–1207. doi:10.1016/j.cell.2006.10.040.

- Sanger, F., S. Nicklen, and A. R. Coulson. 1977. "DNA Sequencing with Chain-Terminating Inhibitors." *Proceedings of the National Academy of Sciences* 74 (12): 5463–67.
- Seggerson, Kathy, Lingjuan Tang, and Eric G. Moss. 2002. "Two Genetic Circuits Repress the *Caenorhabditis Elegans* Heterochronic Gene *Lin-28* after Translation Initiation." *Developmental Biology* 243 (2): 215–25. doi:10.1006/dbio.2001.0563.
- Semple, Jennifer I., Laura Biondini, and Ben Lehner. 2012. "Generating Transgenic Nematodes by Bombardment and Antibiotic Selection." *Nature Methods* 9 (2): 118–19. doi:10.1038/nmeth.1864.
- Semple, Jennifer I., Rosa Garcia-Verdugo, and Ben Lehner. 2010. "Rapid Selection of Transgenic *C. Elegans* Using Antibiotic Resistance." *Nature Methods* 7 (9): 725–27. doi:10.1038/nmeth.1495.
- Shen, Wei-Fang, Yu-Long Hu, Lalita Uttarwar, Emmanuelle Passegue, and Corey Largman. 2008. "MicroRNA-126 Regulates *HOXA9* by Binding to the Homeobox." *Molecular and Cellular Biology* 28 (14): 4609–19. doi:10.1128/MCB.01652-07.
- Shiu, Philip K., Jimmy J. Zhuang, and Craig P. Hunter. 2014. "Assays for Direct and Indirect Effects of *C. Elegans* Endo-siRNAs." *Methods in Molecular Biology (Clifton, N.J.)* 1173: 71–87. doi:10.1007/978-1-4939-0931-5_7.
- Shuman, S. 2001. "Structure, Mechanism, and Evolution of the mRNA Capping Apparatus." *Progress in Nucleic Acid Research and Molecular Biology* 66: 1–40.
- Simon, David J., Jon M. Madison, Annie L. Conery, Katherine L. Thompson-Peer, Michael Soskis, Gary B. Ruvkun, Joshua M. Kaplan, and John K. Kim. 2008. "The MicroRNA *miR-1* Regulates a MEF-2-Dependent Retrograde Signal at Neuromuscular Junctions." *Cell* 133 (5): 903–15. doi:10.1016/j.cell.2008.04.035.
- Singh, Jarnail, and Richard A. Padgett. 2009. "Rates of in Situ Transcription and Splicing in Large Human Genes." *Nature Structural & Molecular Biology* 16 (11): 1128–33. doi:10.1038/nsmb.1666.
- Slack, Frank J., Michael Basson, Zhongchi Liu, Victor Ambros, H. Robert Horvitz, and Gary Ruvkun. 2000. "The *Lin-41* RBCC Gene Acts in the *C. Elegans* Heterochronic Pathway between the *Let-7* Regulatory RNA and the *LIN-29* Transcription Factor." *Molecular Cell* 5 (4): 659–69. doi:10.1016/S1097-2765(00)80245-2.
- Sluder, Ann E., Siuyien Wong Mathews, David Hough, Viravuth P. Yin, and Claude V. Maina. 1999. "The Nuclear Receptor Superfamily Has Undergone Extensive Proliferation and Diversification in Nematodes." *Genome Research* 9 (2): 103–20. doi:10.1101/gr.9.2.103.
- Spieth, John, Glenn Brooke, Scott Kuersten, Kristi Lea, and Thomas Blumenthal. 1993. "Operons in *C. Elegans*: Polycistronic mRNA Precursors Are Processed by Trans-Splicing of *SL2* to Downstream Coding Regions." *Cell* 73 (3): 521–32. doi:10.1016/0092-8674(93)90139-H.
- Stinchcomb, D. T., J. E. Shaw, S. H. Carr, and D. Hirsh. 1985. "Extrachromosomal DNA Transformation of *Caenorhabditis Elegans*." *Molecular and Cellular Biology* 5 (12): 3484.
- Sulston, J. E. 1976. "Post-Embryonic Development in the Ventral Cord of *Caenorhabditis Elegans*." *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 275 (938): 287–97.
- Sulston, J. E., and S. Brenner. 1974. "The Dna of *Caenorhabditis Elegans*." *Genetics* 77 (1): 95–104.
- Sulston, J. E., and H. R. Horvitz. 1977. "Post-Embryonic Cell Lineages of the Nematode, *Caenorhabditis Elegans*." *Developmental Biology* 56 (1): 110–56. doi:10.1016/0012-1606(77)90158-0.
- Sulston, John E., H. Robert Horvitz, and Judith Kimble. 1988. "APPENDIX 3 Cell Lineage." In *The nematode Caenorhabditis elegans*, 457–89.
- Tamburino, Alex M., Sean P. Ryder, and Albertha J. M. Walhout. 2013. "A Compendium of *Caenorhabditis Elegans* RNA Binding Proteins Predicts Extensive Regulation at

- Multiple Levels.” *G3: Genes|Genomes|Genetics* 3 (2): 297–304. doi:10.1534/g3.112.004390.
- Tamura, Taka-Aki, Yoshiyuki Konishi, Yasutaka Makino, and Katsuhiko Mikoshiba. 1996. “Mechanisms of Transcriptional Regulation and Neural Gene Expression.” *Neurochemistry International* 29 (6): 573–81. doi:10.1016/S0197-0186(96)00048-4.
- The C. elegans Deletion Mutant Consortium. 2012. “Large-Scale Screening for Targeted Knockouts in the Caenorhabditis Elegans Genome.” *G3: Genes|Genomes|Genetics* 2 (11): 1415–25. doi:10.1534/g3.112.003830.
- Toscano-Garibay, Julia D., and Guillermo Aquino-Jarquin. 2014. “Transcriptional Regulation Mechanism Mediated by miRNA–DNA•DNA Triplex Structure Stabilized by Argonaute.” *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* 1839 (11): 1079–83. doi:10.1016/j.bbagr.2014.07.016.
- Tullet, Jennifer M. A., Maren Hertweck, Jae Hyung An, Joseph Baker, Ji Yun Hwang, Shu Liu, Riva P. Oliveira, Ralf Baumeister, and T. Keith Blackwell. 2008. “Direct Inhibition of the Longevity-Promoting Factor SKN-1 by Insulin-like Signaling in C. Elegans.” *Cell* 132 (6): 1025–38. doi:10.1016/j.cell.2008.01.030.
- Tzur, Yonatan B., Ari E. Friedland, Saravanapriah Nadarajan, George M. Church, John A. Calarco, and Monica P. Colaiácovo. 2013. “Heritable Custom Genomic Modifications in Caenorhabditis Elegans via a CRISPR–Cas9 System.” *Genetics* 195 (3): 1181–85. doi:10.1534/genetics.113.156075.
- Vallin, Elodie, Joseph Gallagher, Laure Granger, Edwige Martin, Jérôme Belougne, Julien Maurizio, Yohann Duverger, et al. 2012. “A Genome-Wide Collection of Mos1 Transposon Insertion Mutants for the C. Elegans Research Community.” *PLoS ONE* 7 (2): e30482. doi:10.1371/journal.pone.0030482.
- Vanderslice, Rebecca, and David Hirsh. 1976. “Temperature-Sensitive Zygote Defective Mutants of Caenorhabditis Elegans.” *Developmental Biology* 49 (1): 236–49. doi:10.1016/0012-1606(76)90269-4.
- Vasudevan, Shobha, Yingchun Tong, and Joan A. Steitz. 2007. “Switching from Repression to Activation: MicroRNAs Can Up-Regulate Translation.” *Science* 318 (5858): 1931–34. doi:10.1126/science.1149460.
- Voinnet, Olivier, and David C. Baulcombe. 1997. “Systemic Signalling in Gene Silencing.” *Nature* 389 (6651): 553–553. doi:10.1038/39215.
- Waaijers, Selma, and Mike Boxem. 2014. “Engineering the Caenorhabditis Elegans Genome with CRISPR/Cas9.” *Methods, C. Elegans Methods*, 68 (3): 381–88. doi:10.1016/j.ymeth.2014.03.024.
- Waaijers, S., V. Portegijs, J. Kerver, B. B. L. G. Lemmens, M. Tijsterman, S. van den Heuvel, and M. Boxem. 2013. “CRISPR/Cas9-Targeted Mutagenesis in Caenorhabditis Elegans.” *Genetics* 195 (3): 1187–91. doi:10.1534/genetics.113.156299.
- Wahl, Markus C., Cindy L. Will, and Reinhard Lührmann. 2009. “The Spliceosome: Design Principles of a Dynamic RNP Machine.” *Cell* 136 (4): 701–18. doi:10.1016/j.cell.2009.02.009.
- Wang, Duo, Scott Kennedy, Darryl Conte, John K. Kim, Harrison W. Gabel, Ravi S. Kamath, Craig C. Mello, and Gary Ruvkun. 2005. “Somatic Misexpression of Germline P Granules and Enhanced RNA Interference in Retinoblastoma Pathway Mutants.” *Nature* 436 (7050): 593–97. doi:10.1038/nature04010.
- Wang, Juan, and Maureen M. Barr. 2005. “RNA Interference in Caenorhabditis Elegans.” In *Methods in Enzymology*, edited by and John J. Rossi David R. Engelke, Volume 392:36–55. RNA Interference. Academic Press. <http://www.sciencedirect.com/science/article/pii/S0076687904920034>.
- Wang, Luwen, Ning Jiang, Lin Wang, Ou Fang, Lindsey J. Leach, Xiaohua Hu, and Zewei Luo. 2014. “3′ Untranslated Regions Mediate Transcriptional Interference between Convergent Genes Both Locally and Ectopically in Saccharomyces Cerevisiae.” *PLoS Genet* 10 (1): e1004021. doi:10.1371/journal.pgen.1004021.

- Watson, J. D., and F. H. Crick. 1953. "Molecular Structure of Nucleic Acids; a Structure for Deoxyribose Nucleic Acid." *Nature* 171 (4356): 737–38.
- White, Glenn E., Christine M. Petry, and Fred Schachat. 2003. "The Pathway of Myofibrillogenesis Determines the Interrelationship between Myosin and Paramyosin Synthesis in *Caenorhabditis Elegans*." *Journal of Experimental Biology* 206 (11): 1899–1906. doi:10.1242/jeb.00377.
- White, J. G., E. Southgate, J. N. Thomson, and S. Brenner. 1986. "The Structure of the Nervous System of the Nematode *Caenorhabditis Elegans*." *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 314 (1165): 1–340.
- Wightman, B., T. R. Bürglin, J. Gatto, P. Arasu, and G. Ruvkun. 1991. "Negative Regulatory Sequences in the Lin-14 3'-Untranslated Region Are Necessary to Generate a Temporal Switch during *Caenorhabditis Elegans* Development." *Genes & Development* 5 (10): 1813–24.
- Wightman, Bruce, Ilho Ha, and Gary Ruvkun. 1993. "Posttranscriptional Regulation of the Heterochronic Gene Lin-14 by Lin-4 Mediates Temporal Pattern Formation in *C. Elegans*." *Cell*.
- Williams, Andrew H, Ning Liu, Eva van Rooij, and Eric N Olson. 2009. "MicroRNA Control of Muscle Development and Disease." *Current Opinion in Cell Biology, Nucleus and gene expression*, 21 (3): 461–69. doi:10.1016/j.ceb.2009.01.029.
- Williams, Carol, Lei Xu, and Thomas Blumenthal. 1999. "SL1 Trans Splicing and 3'-End Formation in a Novel Class of *Caenorhabditis elegans* Operon." *Molecular and Cellular Biology* 19 (1): 376–83.
- Wilm, T., P. Demel, H. U. Koop, H. Schnabel, and R. Schnabel. 1999. "Ballistic Transformation of *Caenorhabditis Elegans*." *Gene* 229 (1-2): 31–35.
- Wolkow, C.A., and D.H. Hall. "The Dauer Pharynx." In *WormAtlas*.
- Xu, Wenlong, Anthony San Lucas, Zixing Wang, and Yin Liu. 2014. "Identifying microRNA Targets in Different Gene Regions." *BMC Bioinformatics* 15 Suppl 7: S4. doi:10.1186/1471-2105-15-S7-S4.
- Yan, Dong, Zilu Wu, Andrew D. Chisholm, and Yishi Jin. 2009. "The DLK-1 Kinase Promotes mRNA Stability and Local Translation in *C. Elegans* Synapses and Axon Regeneration." *Cell* 138 (5): 1005–18. doi:10.1016/j.cell.2009.06.023.
- Yeom, Kyu-Hyeon, Yoontae Lee, Jinju Han, Mi Ra Suh, and V. Narry Kim. 2006. "Characterization of DGCR8/Pasha, the Essential Cofactor for Drosha in Primary miRNA Processing." *Nucleic Acids Research* 34 (16): 4622–29. doi:10.1093/nar/gkl458.
- Zorio, Diego A. R., Niansheng Nick Cheng, Thomas Blumenthal, and John Spieth. 1994. "Operons as a Common Form of Chromosomal Organization in *C. Elegans*." *Nature* 372 (6503): 270–72. doi:10.1038/372270a0.
- Zou, Yan, Hui Chiu, Dorothée Domenger, Chiou-Fen Chuang, and Chieh Chang. 2012. "The Lin-4 MicroRNA Targets the LIN-14 Transcription Factor to Inhibit Netrin-Mediated Axon Attraction." *Science Signaling* 5 (228): ra43–ra43. doi:10.1126/scisignal.2002437.
- Zuryn, Steven, and Sophie Jarriault. 2013. "Deep Sequencing Strategies for Mapping and Identifying Mutations from Genetic Screens." *Worm* 2 (3): e25081. doi:10.4161/worm.25081.

VII. Annexes

1. Liste des amorces

Numéro	Nom d'amorce	Séquence
51	GFP_F	ATGAGTAAAGGAGAAGAAGT
52	GFP_R	CTATTTGTATAGTTCATCCA
53	RFP_F	ATGAGTAAAGTCTCAAAGGGTGAAGAAG
64	mCherry_R (NotI)	AGCGGCCGCTTATACAATTCATCCATGC
85	M13 F	CGACGTTGTAAAACGACGGCCAGT
86	M13 R	CAGGAAACAGCTATGACCATG
322	Fos29a_U_R_CB20	CTAAAGGGAACAAAAGCTGGCGCCCGCAATAGCGGAAGCTT
323	pBS_Fos29a_F_CB20	AAGTTCCGCTATTGCGGGCGCCAGCTTTTGTTCCTTTAG
324	Fos29a_P_F_CB20	GGATCCACTAGTTCTAGAGCATGTGAAATTGAGAGTGTGA
325	pBS_Fos29a_R_CB20	TCACACTCTCAATTTACATGCTCTAGAACTAGTGGATCC
326	Fos29a_U_R_CB40	AAGCTCGAAATTAACCCTCACTAAAGGGAACAAAAGCTGGC GCCCCGAATAGCGGAAGCTT
327	pBS_F_CB40	CCAGCTTTTGTTCCTTTAG
328	Fos29a_P_F_CB40	TGGAATTCCTGCAGCCCGGGGATCCACTAGTTCTAGAGCAT GTGAAATTGAGAGTGTGA
329	pBS_R_CB40	GCTCTAGAACTAGTGGATCC
330	Fos30_U_R_CB20	CTAAAGGGAACAAAAGCTGGCTGTATGACTATTTTCGTAAT
331	pBS_Fos30_F_CB20	ATTACGAAATAGTCATACAGCCAGCTTTTGTTCCTTTAG
332	Fos30_P_F_CB20	GGATCCACTAGTTCTAGAGCTATACGTAAAGTTTAAAATA
333	pBS_Fos30_R_CB20	TATTTTAAACTTTACGTATAGCTCTAGAACTAGTGGATCC
334	Fos30_U_R_CB40	AAGCTCGAAATTAACCCTCACTAAAGGGAACAAAAGCTGGCT GTATGACTATTTTCGTAAT
335	Fos30_P_F_CB40	TGGAATTCCTGCAGCCCGGGGATCCACTAGTTCTAGAGCTA TACGTAAAGTTTAAAATA
336	nhr17_P_F_CB20	GGATCCACTAGTTCTAGAGCTCATTTACTATATATAAAGC
337	nhr17_P_R_gfp	AGTTCTTCTCCTTTACTCATTACTGAAATGAAATAATGAA
338	nhr17_U_R_CB20	CTAAAGGGAACAAAAGCTGGCTTGCTGACAGCTCAAATG
339	uba1_P_F_CB20	GGATCCACTAGTTCTAGAGCCTAGTACGCCCTGGAAGCCG
340	uba1_P_R_gfp	AGTTCTTCTCCTTTACTCATTAATCACGAAAGATCAGAAC
341	uba1_U_R_CB20	CTAAAGGGAACAAAAGCTGGATTATTTGTTCAACTTG
342	lin41_P_F_CB20	GGATCCACTAGTTCTAGAGCGTGTGTAGAGAGTGGGAACG
343	lin41_P_R_gfp	AGTTCTTCTCCTTTACTCATTTCACTTTTTTCCAAGTCTG
344	lin41_U_R_CB20	CTAAAGGGAACAAAAGCTGGGCACAGGGAAGGGGCGGAG
345	lin28_P_R_gfp	AGTTCTTCTCCTTTACTCATCTGAAAAAGATTTTTAAAA
346	ZK7571_P_R_gfp	AGTTCTTCTCCTTTACTCATGAGCTCTTTGAGCATTTAAACG
394	T10B11.6_P_F_CB20	GGATCCACTAGTTCTAGAGCGATATTGAGGTCAGGAATTG
395	T10B11.6_P_Rr_CB20	TCTTCACCCTTTGAGACCATCTGAAAAATATTAATTTGATGG
396	T10B11.6_P_Rg_CB20	AGTTCTTCTCCTTTACTCATCTGAAAAATATTAATTTGATGG
397	T10B11.6_U_F_CB20	CATGGATGAACTATACAAATAGCGTATAAATGAAGAAATGA GG
398	T10B11.6_U_R_CB20	CTAAAGGGAACAAAAGCTGGCAAAATCTCAAAAGTACACAT G
399	mCh_F	ATGGTCTCAAAGGGTGAAGA
400	unc-54_U_R_CB20	CTAAAGGGAACAAAAGCTGGAAACAGTTATGTTTGGTATATT GGGA
425	unc-54_F_CB20	CATGGATGAACTATACAAATAGCGGCCGCTCCAATTACTCT T
426	nhr-17_U_F_CB	ATGGATGAATTGTATAAGCGGCCGCTAAATTTACATATTTAT GATGCAACTTAC
427	lin-28_U_F_CB	ATGGATGAATTGTATAAGCGGCCGCTAATCATCTAGACACTG AGAATATTG

452	ubc-1_U_F_CB20	ATGGATGAATTGTATAAGCGGCCGCTTAAATTTTCAGGAATT TTTCTG
453	ubc-1_U_R_CB20	CTAAAGGGAACAAAAGCTGGGCCATCAAAATTTTTTATTC
454	act-1_U_F_CB20	ATGGATGAATTGTATAAGCGGCCGCTTAAATGCACAACTTCG TCAA
455	act-1_U_R_CB20	CTAAAGGGAACAAAAGCTGGATCAATTTTTAAATTTTTATTC AC
464	Punc-54_F	GGATCCACTAGTTCTAGAGCAAACCTGCGACAGCCACTAGT
465	Punc-54_R_GFP	AGTTCTTCTCCTTTACTCATGATTCTCGCTTCTTCAAATGG
466	Punc-54_R_mCh	TCTTCACCCTTTGAGACCATGATTTCTCGCTTCTTCAAATGG
537	let-858_UTR_R_pBS	CTAAAGGGAACAAAAGCTGGGATTGAATTAATTTTTTAGA
538	let-858_UTR_F_mCh	ATGGATGAATTGTATAAGCGGCCGCTATTTTCAAATTTTAAAT ACTG
539	let-858_UTR_F_gfp	CATGGATGAACTATACAAATAGATTTTCAAATTTTAAATACT G
555	DsRed_F	ATGGCCTCCTCCGAGAACG
556	DsRed_R	CTACAGGAACAGGTGGTGGC
557	unc-54_F_dsRedtail	GCCACCACCTGTTCTGTAGCGGCCGCGTCCAATTACTCTT
572	Munc-54_S1_F	TCCCACCCCTATTTTTG
573	Munc-54_S1_R	GCATGTAGGGATGTTGAAGAG
574	Munc-54_S1_F2	TCTTCAACATCCCTACATGCTCCCACCCCTATTTTTG
575	Munc-54_S2_F	AACTTCTCTTAATTTCTTTGTTTTTAG
576	Munc-54_S2_R	AAATAGGGGGTGGGAGC
577	Munc-54_S2_R2	CAAAGAAATTAAGAGAAGTTAAATAGGGGGTGGGAGC
578	Munc-54_S3_F	CTTATGTTTTTTACTTCTGATAAA
579	Munc-54_S3_R	CAGAACATTGTGTAGATTTTG
580	Munc-54_S3_R2	CAAATCTACACAATGTTCTGCTTATGTTTTTTACTTCTGA
581	Munc-54_S2_F_Verif	CATGCTCTTTCTCCCTGTGC
582	Munc-54_S2_R_Verif	GTGACTTAAAAGAAGCTAAAAAACAAAG
583	Munc-54_S3_F_Verif	ATAATTCTATCCCAAATCTAC
584	Munc-54_S3_R_Verif	CGGTTTTTTCTATGATGTTTC
598	3'let-858_F_dsRed	GCCACCACCTGTTCTGTAGATTTTCAAATTTTAAATACTG
640	Pmir-1820_F_pBS	GGATCCACTAGTTCTAGAGCATTATTTCAATTCAAATTGAC
641	Pmir-1820_R_GFP	AGTTCTTCTCCTTTACTCATGAATACGTAGAGTAAAGTAG
642	mir-1820_F	TTCAAAAATTGCATTTTCCATCT
643	mir-1820_R	CTAAAGGGAACAAAAGCTGGTTCTCAAAACCTTGACTGAAC
644	Pmir-4811_F_pBS	GGATCCACTAGTTCTAGAGCATTCTTACCGCGACTATATTC
645	Pmir-4811_R_GFP	AGTTCTTCTCCTTTACTCATTTAAAAATATACAGATGGCAGC
646	mir-4811_F	AAAATTGTGGTATTGAACAATAC
647	mir-4811_R	CTAAAGGGAACAAAAGCTGGCATTACTCAGTGTAGACCTG
648	Munc-54_MS4811	ATCAAAAAACTTCTCTTAATTTT
649	Punc-54_R_dsRed	ACGTTCTCGGAGGAGGCCATCATGATTCTCGCTTCTTTCAA TGG
658	Pmir-1820_R_dsRed	ACGTTCTCGGAGGAGGCCATGAATACGTAGAGTAAAGTAG
659	Pmir-4811_R_dsRed	ACGTTCTCGGAGGAGGCCATTTAAAAATATACAGATGGCAGC
660	3'mir-4811_F_GFP	CATGGATGAACTATACAAATAGGCGTCCAACCTTCTATAGC
661	3'mir-1820_F_GFP	CATGGATGAACTATACAAATAGCTGGAAGATGTAAGTTTATA AAA
670	Punc-54_R_mir4811	TTGTTCAATACCACAATTTTGATTTCTCGCTTCTTTCAAATGG
671	Pmir-4811_R_ok	AAAATATACAGATGGCAGCAG
689	Posm-9_F	GGATCCACTAGTTCTAGAGCCCTTAGAGGCACATTTTAAGG
690	Posm-9_R_GFP	AGTTCTTCTCCTTTACTCATGTTTGGTTTCTGAAAAAATTGG
691	3'osm-9_F_GFP	CATGGATGAACTATACAAATAGGAACTTTTTTCTTCTAATTTT TTG

692	3'osm-9_R	CTAAAGGGAACAAAAGCTGGAACAAGACACACCTCTCATC
693	Posm-9_R_dsRed	ACGTTCTCGGAGGAGGCCATGTTTGGTTTCTGAAAAAATTGG
714	3'lin28_dsRed	GCCACCACCTGTTCTGTAGAATCATCTAGACACTGAGAATA TTG
715	Posm-9_f_30	GCAGCCCGGGGGATCCACTAGTTCTAGAGCCCTTAGAGGGCC ACATTTTAAGG
716	Posm-9_R_GFP_30	CTCCAGTGAAAAGTTCTTCTCCTTTACTCATGTTTGGTTTCTG AAAAAATTGG

2. Liste des plasmides

Plasmides utilisés dans ces travaux	Description
pBS (SK-)	vecteur utilisé comme base des plasmides et les clonages
<i>Plin-28::GFP::lin-28 3'UTR</i>	rapporteur de l'expression du gène <i>lin-28</i>
<i>Plin-28::GFP::unc-54 3'UTR</i>	rapporteur témoin de l'activité du promoteur du gène <i>lin-28</i>
<i>Plin-28::mCherry::act-1 3'UTR</i>	rapporteur témoin de l'activité du promoteur du gène <i>lin-28</i>
<i>Plin-28::mCherry::lin-28 3'UTR</i>	rapporteur de l'expression du gène <i>lin-28</i>
<i>Plin-28::mCherry::unc-1 3'UTR</i>	rapporteur témoin de l'activité du promoteur du gène <i>lin-28</i>
<i>Plin-28::mCherry::unc-54 3'UTR</i>	rapporteur témoin de l'activité du promoteur du gène <i>lin-28</i>
<i>Plin-28::mCherry::unc-54 3'UTR ΔS1</i>	rapporteur portant la délétion du site 1 dans <i>unc-54 3'UTR</i>
<i>Plin-28::mCherry::unc-54 3'UTR ΔS2</i>	rapporteur portant la délétion du site 2 dans <i>unc-54 3'UTR</i>
<i>Plin-28::mCherry::unc-54 3'UTR ΔS3</i>	rapporteur portant la délétion du site 3 dans <i>unc-54 3'UTR</i>
<i>Plin-28::mCherry::unc-54 3'UTR ΔS4811</i>	rapporteur portant la délétion du site 4811 dans <i>unc-54 3'UTR</i>
<i>Plin-28::RRPT::lin-28 3'UTR</i>	contient la RRPT pour le gène <i>lin-28</i>
<i>Plin-41::GFP::lin-41 3'UTR</i>	rapporteur de l'expression du gène <i>lin-41</i>
<i>Plin-41::RRPT::lin-41 3'UTR</i>	contient la RRPT pour le gène <i>lin-41</i>
<i>Pmir-1820::GFP::mir-1820 3'UTR</i>	rapporteur de l'expression du <i>mir-1820</i>
<i>Pmir-4811::GFP::mir-4811 3'UTR</i>	rapporteur de l'expression du <i>mir-4811</i>
<i>Pnhr-17::GFP::nhr-17 3'UTR</i>	rapporteur de l'expression du gène <i>nhr-17</i>
<i>Pnhr-17::mCherry::nhr-17 3'UTR</i>	rapporteur de l'expression du gène <i>nhr-17</i>
<i>Pnhr-17::RRPT::nhr-17 3'UTR</i>	contient la RRPT pour le gène <i>nhr-17</i>
<i>Posm-9::GFP::osm-9 3'UTR</i>	rapporteur de l'activité du gène <i>osm-9</i>

<i>pRG02</i>	porteur du gène de résistance au G418
<i>PT10B11.6::GFP::T10B11.6 3'UTR</i>	rapporteur de l'activité du gène <i>T10B11.6</i>
<i>PT10B11.6::mCherry::unc-54 3'UTR</i>	rapporteur témoin de l'activité du promoteur du gène <i>T10B11.6</i>
<i>Puba-1::GFP::uba-1 3'UTR</i>	rapporteur de l'expression du gène <i>uba-1</i>
<i>Puba-1::RRPT::uba-1 3'UTR</i>	contient la RRPT pour le gène <i>uba-1</i>
<i>Punc-54::dsRed::unc-54 3'UTR</i>	rapporteur de l'expression du gène <i>unc-54</i>
<i>Punc-54::dsRed::unc-54 3'UTR ΔS1</i>	rapporteur portant la délétion du site 1 dans <i>unc-54 3'UTR</i>
<i>Punc-54::dsRed::unc-54 3'UTR ΔS2</i>	rapporteur portant la délétion du site 2 dans <i>unc-54 3'UTR</i>
<i>Punc-54::dsRed::unc-54 3'UTR ΔS3</i>	rapporteur portant la délétion du site 3 dans <i>unc-54 3'UTR</i>
<i>Punc-54::dsRed::unc-54 3'UTR ΔS4811</i>	rapporteur portant la délétion du site 4811 dans <i>unc-54 3'UTR</i>
<i>Punc-54::GFP::unc-54 3'UTR</i>	rapporteur de l'expression du gène <i>unc-54</i>
<i>PZK757.1::GFP::ZK757.1 3'UTR</i>	rapporteur de l'expression du gène <i>ZK757.1</i>
<i>PZK757.1::RRPT::ZK757.1 3'UTR</i>	contient la RRPT pour le gène <i>ZK757.1</i>

3. Liste des lignées

<i>C. elegans</i>	Génotype
DUD61	N2+/+ ; dudEx48[pdestDD04neo; pBS ; Pnhr-17::mCherry::3'UTR unc-54::OperonLinker::gfp::3'UTR nhr-17]
DUD062	N2+/+ ; dudEx49[pdestDD04neo; pBS ; Pnhr-17::mCherry::3'UTR unc-54::OperonLinker::gfp::3'UTR nhr-17]
DUD063	N2+/+ ; dudEx50[pdestDD04neo; pBS ; Pnhr-17::mCherry::3'UTR unc-54::OperonLinker::gfp::3'UTR nhr-17]
DUD064	N2+/+ ; dudEx51[pdestDD04neo; pBS ; Pnhr-17::mCherry::3'UTR unc-54::OperonLinker::gfp::3'UTR nhr-17]
DUD065	N2+/+ ; dudEx52[pdestDD04neo; pBS ; Pnhr-17::mCherry::3'UTR unc-54::OperonLinker::gfp::3'UTR nhr-17]
DUD069	N2+/+ ; dudEx56[pdestDD04neo; pBS ; Plin-41::mCherry::3'UTR unc-54::OperonLinker::gfp::3'UTR lin-41]
DUD070	N2+/+ ; dudEx57[pdestDD04neo; pBS ; Plin-41::mCherry::3'UTR unc-54::OperonLinker::gfp::3'UTR lin-41]
DUD071	N2+/+ ; dudEx58[pdestDD04neo; pBS ; Plin-41::mCherry::3'UTR unc-54::OperonLinker::gfp::3'UTR lin-41]
DUD072	N2+/+ ; dudEx59[pdestDD04neo; pBS ; Plin-41::mCherry::3'UTR unc-54::OperonLinker::gfp::3'UTR lin-41]
DUD073	N2+/+ ; dudEx60[pdestDD04neo; pBS ; Plin-41::mCherry::3'UTR unc-54::OperonLinker::gfp::3'UTR lin-41]
DUD086	N2+/+ ; dudEx73[pdestDD04neo; pBS ; Puba-1::mCherry::3'UTR unc-54::OperonLinker::gfp::3'UTR unba-1]
DUD087	N2+/+ ; dudEx74[pdestDD04neo; pBS ; Puba-1::mCherry::3'UTR unc-54::OperonLinker::gfp::3'UTR unba-1]
DUD088	N2+/+ ; dudEx75[pdestDD04neo; pBS ; Puba-1::mCherry::3'UTR unc-54::OperonLinker::gfp::3'UTR unba-1]
DUD089	N2+/+ ; dudEx76[pdestDD04neo; pBS ; Puba-1::mCherry::3'UTR unc-54::OperonLinker::gfp::3'UTR unba-1]
DUD090	N2+/+ ; dudEx77[pdestDD04neo; pBS ; Puba-1::mCherry::3'UTR unc-54::OperonLinker::gfp::3'UTR unba-1]
DUD091	N2+/+ ; dudEx78[pdestDD04neo; pBS ; Puba-1::mCherry::3'UTR unc-54::OperonLinker::gfp::3'UTR unba-1]
DUD092	N2+/+ ; dudEx79[pdestDD04neo; pBS ; Puba-1::mCherry::3'UTR unc-54::OperonLinker::gfp::3'UTR unba-1]
DUD093	N2+/+ ; dudEx80[plin-28::mCherry::3'UTR unc-54::OperonLinker::gfp::3'UTR lin-28 ; prps-27::neo::3'UTR unc-54]
DUD094	N2+/+ ; dudEx81[plin-28::mCherry::3'UTR unc-54::OperonLinker::gfp::3'UTR lin-28 ; prps-27::neo::3'UTR unc-54]
DUD095	N2+/+ ; dudEx82[plin-28::mCherry::3'UTR unc-54::OperonLinker::gfp::3'UTR lin-28 ; prps-27::neo::3'UTR unc-54]
DUD096	N2+/+ ; dudEx83[plin-28::mCherry::3'UTR unc-54::OperonLinker::gfp::3'UTR lin-28 ; prps-27::neo::3'UTR unc-54]
DUD097	N2+/+ ; dudEx84[plin-28::mCherry::3'UTR unc-54::OperonLinker::gfp::3'UTR lin-28 ; prps-27::neo::3'UTR unc-54]
DUD098	N2+/+ ; dudEx85[plin-28::mCherry::3'UTR unc-54::OperonLinker::gfp::3'UTR lin-28 ; prps-27::neo::3'UTR unc-54]
DUD099	N2+/+ ; dudEx86[Pnhr-17::gfp::3'UTR nhr-17 ; Prps-27::neo::3'UTR unc-54]
DUD100	N2+/+ ; dudEx87[Pnhr-17::gfp::3'UTR nhr-17 ; Prps-27::neo::3'UTR unc-54]
DUD101	N2+/+ ; dudEx88[Pnhr-17::gfp::3'UTR nhr-17 ; Prps-27::neo::3'UTR unc-54]
DUD102	N2+/+ ; dudEx89[Pnhr-17::gfp::3'UTR nhr-17 ; Prps-27::neo::3'UTR unc-54]
DUD103	N2+/+ ; dudEx90[Puba-1::gfp::3'UTR uba-1 ; Prps-27::neo::3'UTR unc-54]
DUD104	N2+/+ ; dudEx91[Puba-1::gfp::3'UTR uba-1 ; Prps-27::neo::3'UTR unc-54]
DUD105	N2+/+ ; dudEx92[Puba-1::gfp::3'UTR uba-1 ; Prps-27::neo::3'UTR unc-54]
DUD106	N2+/+ ; dudEx93[Puba-1::gfp::3'UTR uba-1 ; Prps-27::neo::3'UTR unc-54]
DUD107	N2+/+ ; dudEx94[Puba-1::gfp::3'UTR uba-1 ; Prps-27::neo::3'UTR unc-54]
DUD108	N2+/+ ; dudEx95[Puba-1::gfp::3'UTR uba-1 ; Prps-27::neo::3'UTR unc-54]

DUD109 N2+/+ ; dudEx96[Pzk757.1::gfp::3'UTR zk757.1 ; Prps-27::neo::3'UTR unc-54]

DUD110 N2+/+ ; dudEx97[Pzk757.1::gfp::3'UTR zk757.1 ; Prps-27::neo::3'UTR unc-54]

DUD111 N2+/+ ; dudEx98[Pzk757.1::gfp::3'UTR zk757.1 ; Prps-27::neo::3'UTR unc-54]

DUD112 N2+/+ ; dudEx99[Pzk757.1::gfp::3'UTR zk757.1 ; Prps-27::neo::3'UTR unc-54]

DUD113 N2+/+ ; dudEx100[Pzk757.1::gfp::3'UTR zk757.1 ; Prps-27::neo::3'UTR unc-54]

DUD114 N2+/+ ; dudEx101[Pzk757.1::gfp::3'UTR zk757.1 ; Prps-27::neo::3'UTR unc-54]

DUD115 N2+/+ ; dudEx102[Pzk757.1::gfp::3'UTR zk757.1 ; Prps-27::neo::3'UTR unc-54]

DUD116 N2+/+ ; dudEx103[Plin-41::gfp::3'UTR lin-41 ; Prps-27::neo::3'UTR unc-54]

DUD117 N2+/+ ; dudEx104[Plin-41::gfp::3'UTR lin-41 ; Prps-27::neo::3'UTR unc-54]

DUD118 N2+/+ ; dudEx105[Plin-41::gfp::3'UTR lin-41 ; Prps-27::neo::3'UTR unc-54]

DUD119 N2+/+ ; dudEx106[Plin-41::gfp::3'UTR lin-41 ; Prps-27::neo::3'UTR unc-54]

DUD120 N2+/+ ; dudEx107[Plin-41::gfp::3'UTR lin-41 ; Prps-27::neo::3'UTR unc-54]

DUD121 N2+/+ ; dudEx108[Plin-41::gfp::3'UTR lin-41 ; Prps-27::neo::3'UTR unc-54]

DUD122 N2+/+ ; dudEx109[Plin-41::gfp::3'UTR lin-41 ; Prps-27::neo::3'UTR unc-54]

DUD123 N2+/+ ; dudEx110[Plin-41::gfp::3'UTR lin-41 ; Prps-27::neo::3'UTR unc-54]

DUD124 N2+/+ ; dudEx111[Plin-28::gfp::3'UTR lin-28 ; Prps-27::neo::3'UTR unc-54]

DUD125 N2+/+ ; dudEx112[Plin-28::gfp::3'UTR lin-28 ; Prps-27::neo::3'UTR unc-54]

DUD126 N2+/+ ; dudEx113[Plin-28::gfp::3'UTR lin-28 ; Prps-27::neo::3'UTR unc-54]

DUD127 N2+/+ ; dudEx114[Plin-28::gfp::3'UTR lin-28 ; Prps-27::neo::3'UTR unc-54]

DUD129 N2+/+ ; dudEx116[Plin-28::gfp::3'UTR lin-28 ; Prps-27::neo::3'UTR unc-54]

DUD130 N2+/+ ; dudEx117[Plin-28::gfp::3'UTR lin-28 ; Prps-27::neo::3'UTR unc-54]

DUD131 N2+/+ ; dudEx118[Plin-28::gfp::3'UTR lin-28 ; Prps-27::neo::3'UTR unc-54]

DUD132 N2+/+ ; dudEx119[PZK757.1::mCherry::3'UTR unc-54::OperonLinker::gfp::3'UTR ZK757.1 ; prps-27::neo::3'UTR unc-54]

DUD133 N2+/+ ; dudEx120[pZK757.1::mCherry::3'UTR unc-54::OperonLinker::gfp::3'UTR ZK757.1 ; prps-27::neo::3'UTR unc-54]

DUD134 N2+/+ ; dudEx121[pZK757.1::mCherry::3'UTR unc-54::OperonLinker::gfp::3'UTR ZK757.1 ; prps-27::neo::3'UTR unc-54]

DUD135 N2+/+ ; dudEx122[pZK757.1::mCherry::3'UTR unc-54::OperonLinker::gfp::3'UTR ZK757.1 ; prps-27::neo::3'UTR unc-54]

DUD136 N2+/+ ; dudEx123[pZK757.1::mCherry::3'UTR unc-54::OperonLinker::gfp::3'UTR ZK757.1 ; prps-27::neo::3'UTR unc-54]

DUD153 N2+/+ ; dudEx140 [pT10B11.6::mCherry::3'UTR unc-54 ; pT10B11.6::gfp::3'UTR T10B11.6 ; prps-27::neo::3'UTR unc-54]

DUD154 N2+/+ ; dudEx141 [Plin-28::mCherry::3'UTR act-1 ; Plin-28::gfp::3'UTR lin-28 ; prps-27::neo::3'UTR unc-54]

DUD155 N2+/+ ; dudEx142 [Plin-28::mCherry::3'UTR act-1 ; Plin-28::gfp::3'UTR lin-28 ; prps-27::neo::3'UTR unc-54]

DUD156 N2+/+ ; dudEx143 [pT10B11.6::mCherry::3'UTR unc-54 ; pT10B11.6::gfp::3'UTR T10B11.6 ; prps-27::neo::3'UTR unc-54]

DUD158 N2+/+ ; dudEx145 [plin-28::mCherry::3'UTR ubc-1 ; plin-28::gfp::3'UTR lin-28 ; prps-27::neo::3'UTR unc-54]

DUD159 N2+/+ ; dudEx146 [plin-28::mCherry::3'UTR ubc-1 ; plin-28::gfp::3'UTR lin-28 ; prps-27::neo::3'UTR unc-54]

DUD160 N2+/+ ; dudEx147 [plin-28::mCherry::3'UTR unc-54 ; plin-28::gfp::3'UTR lin-28 ; prps-27::neo::3'UTR unc-54]

DUD161 N2+/+ ; dudEx148 [plin-28::mCherry::3'UTR unc-54 ; plin-28::gfp::3'UTR lin-28 ; prps-27::neo::3'UTR unc-54]

DUD162 N2+/+ ; dudEx149 [plin-28::mCherry::3'UTR unc-54_ΔS1 ; plin-28::gfp::3'UTR lin-28 ; prps-27::neo::3'UTR unc-54]

[illegible]

DUD200	N2+/+ ; dudEx185 [plin-28::mCherry::3'UTR lin-28 ; plin-28::gfp::3'UTR unc-54_ΔS2 ; prps-27::neo::3'UTR unc-54]
DUD201	N2+/+ ; dudEx186 [plin-28::mCherry::3'UTR lin-28 ; plin-28::gfp::3'UTR unc-54_ΔS2 ; prps-27::neo::3'UTR unc-54]
DUD202	N2+/+ ; dudEx187 [plin-28::mCherry::3'UTR lin-28 ; plin-28::gfp::3'UTR unc-54_ΔS2 ; prps-27::neo::3'UTR unc-54]
DUD203	N2+/+ ; dudEx188 [plin-28::mCherry::3'UTR lin-28 ; plin-28::gfp::3'UTR unc-54_ΔS2 ; prps-27::neo::3'UTR unc-54]
DUD204	N2+/+ ; dudEx189 [plin-28::mCherry::3'UTR lin-28 ; plin-28::gfp::3'UTR unc-54_ΔS2 ; prps-27::neo::3'UTR unc-54]
DUD205	N2+/+ ; EMS ; dudEx158 [plin-28::mCherry::3'UTR unc-54 ; plin-28::gfp::3'UTR lin-28 ; prps-27::neo::3'UTR unc-54]
DUD206	N2+/+ ; EMS ; dudEx158 [plin-28::mCherry::3'UTR unc-54 ; plin-28::gfp::3'UTR lin-28 ; prps-27::neo::3'UTR unc-54]
DUD207	N2+/+ ; EMS ; dudEx158 [plin-28::mCherry::3'UTR unc-54 ; plin-28::gfp::3'UTR lin-28 ; prps-27::neo::3'UTR unc-54]
DUD208	N2+/+ ; EMS ; dudEx158 [plin-28::mCherry::3'UTR unc-54 ; plin-28::gfp::3'UTR lin-28 ; prps-27::neo::3'UTR unc-54]
DUD209	N2+/+ ; dudEx190 [plin-28::mCherry::3'UTR lin-28 ; pmir-1820::gfp::3'UTR mir-1820 ; prps-27::neo::3'UTR unc-54]
DUD210	N2+/+ ; dudEx191 [plin-28::mCherry::3'UTR lin-28 ; pmir-1820::gfp::3'UTR mir-1820 ; prps-27::neo::3'UTR unc-54]
DUD211	N2+/+ ; dudEx192 [posm-9::gfp::3'UTR osm-9 ; prps-27::neo::3'UTR unc-54]
DUD212	N2+/+ ; dudEx193 [posm-9::gfp::3'UTR osm-9 ; prps-27::neo::3'UTR unc-54]
DUD213	N2+/+ ; dudEx194 [posm-9::gfp::3'UTR osm-9 ; prps-27::neo::3'UTR unc-54]

4. Alignements de séquence

Lignée 1 mir-1820 (Depuis 5')

Query	34	TCTGTTGATTTCGGCCAATTATCCTTTGTGACCACTCCCACTGTGTTGCAGATGCGGAGC	93
Sbjct	1	TCTGTTGANTTCGGCCANTTATCCTTTGTGACCACTCCCACTGTGTTGCAGATGCGGAGC	60
Query	94	ACACATATGGTTTCGAGTATTCTCGACACGTGCACAATTCTTTGTTTGTAAATTGAGCT	153
Sbjct	61	ACACATATGGTTTCGAGTATTCTCGACACGTGCACAATTCTTTGTTTGTAAATTGAGCT	120
Query	154	GGCAAGATATTATTTTATTCAAAATGTTTAAACTGATATAAGCTATTTAATTTGGAATA	213
Sbjct	121	GGCAAGATATTATTTTATTCAAAATGTTTAAACTGATATAAGCTATTTAATTTGGAATA	180
Query	214	TTATTTAATGACTAACAAAGTTTGGTTTGACATTTTATTCGCCTCTCACAACTTGAAAAT	273
Sbjct	181	TTATTTAATGACTAACAAAGTTTGGTTTGACATTTTATTCGCCTCTCACAACTTGAAAAT	240
Query	274	TCTCAAAGCGAACCCCTAATTACTACTGTTTACAACGTTTGCATAATGACAACCATTTT	333
Sbjct	241	TCTCAAAGCGAACCCCTAATTACTACTGTTTACAACGTTTGCATAATGACAACCATTTT	300
Query	334	TCATTGATCTACTTTACTCTACGTATTC	393
Sbjct	301	TCATTGATCTACTTTACTCTACGTATTC	360
Query	394	TTTTCGATGATGTTTCGTTAAATCGGTATAAGCGAACCATTTGTAACAATCAAAGAATGGA	453
Sbjct	361	TTTTCGATGATGTTTCGTTAAATCGGTATAAGCGAACCATTTGTAACAATCAAAGAATGGA	420
Query	454	GAATCAATTTATGATCTGGAAGATGTAAGTTTATAAAAAATGAAGGGTAACAAAACGTTT	513
Sbjct	421	GAATCAATTTATGATCTGGAAGATGTAAGTTTATAAAAAATGAAGGGTAACAAAACGTTT	480
Query	514	TCCATTATATGAATTCCTCAAAAAGATTAAATCAAAGTTTTTGTGCGCGGAGGAGTATG	573
Sbjct	481	TCCATTATATGAATTCCTNN-AAAAGATTAAATCAAAGTTTTTGTGCGCGGAGGAGTATG	539
Query	574	AAAAGAGCAAACGCTTAAGTAAATACATGTCTAGCAGTGTGCGATTGCTCTACAACAATA	633
Sbjct	540	AAAAGAGCAAACGCTTAAGTAAATACATGTCTAGCAGTGTGCGATTGCTCTACNACAATA	599
Query	634	aaaaaaaaaacagtttttaaaattattttcaaaagtccaatttattttcggttttaaaaaa	693
Sbjct	600	AAAAAANCAGTTTTAAAAATTATTTTCNNAAGTTCAATTTATTTTCGTTTTAAAAAA	659
Query	694	aacctaaaaga 704	
Sbjct	660	AACCTAAAAGA 670	

Lignée 1 mir-1820 (Depuis 5')

Query	1	ATTATTTCAATTCAAATTGACTATCTTATTCTCTCTGTTGATTTCGGCCAATTATCCTTT	60
Sbjct	712	ATTATTTCAATTCAAATTGACTATCTTATTCTCTCTGTTGATTTCGGCCAATTATCCTTT	653
Query	61	GTGACCACTCCCACTGTGTTGCAGATGCGGAGCACACATATGGTTTCGAGTATTCTCGAC	120
Sbjct	652	GTGACCACTCCCACTGTGTTGCAGATGCGGAGCACACATATGGTTTCGAGTATTCTCGAC	593
Query	121	ACGTGCACAATTCTTTGTTTGTAAATTGAGCTGGCAAGATATTATTTTATTCAAAATG	180
Sbjct	592	ACGTGCACAATTCTTTGTTTGTAAATTGAGCTGGCAAGATATTATTTTATTCAAAATG	533
Query	181	TTTAAACTGATATAAGCTATTTAATTTGGAATATTATTTAATGACTAACAAAGTTTGGTTT	240
Sbjct	532	TTTAAACTGATATAAGCTATTTAATTTGGAATATTATTTAATGACTAACAAAGTTTGGTTT	473
Query	241	GACATTTTATTCGCCTCTCACAACTTGAAAATCTCAAAGCGAACCCCTAATTACTAC	300
Sbjct	472	GACATTTTATTCGCCTCTCACAACTTGAAAATCTCAAAGCGAACCCCTAATTACTAC	413
Query	301	TGTTTACAACGTTTGCATAATGACAACCATTTTTCATTGATCTACTTTACTCTACGTATT	360
Sbjct	412	TGTTTACAACGTTTGCATAATGACAACCATTTTTCATTGATCTACTTTACTCTACGTATT	353
Query	361	CTTCAAAAATTGCATTTTCCATCTTTTGATTGTTTTTCGATGATGTTTCGTTAAATCGGTA	420
Sbjct	352	CTTCAAAAATTGCATTTTCCATCTTTTGATTGTTTTTCGATGATGTTTCGTTAAATCGGTA	293
Query	421	TAAGCGAACCATTTGTAACAATCAAAGAATGGAGAATCAATTTATGATCTGGAAGATGTA	480
Sbjct	292	TAAGCGAACCATTTGTAACAATCAAAGAATGGAGAATCAATTTATGATCTGGAAGATGTA	233
Query	481	AGTTTATAAAAAATGAAGGGTAACAAAACGTTTTCCATTATATGAATTCCTCAAAAAGA	540
Sbjct	232	AGTTTATAAAAAATGAAGGGTAACAAAACGTTTTCCATTATATGAATTCCTCAAAAAGA	173
Query	541	TTAAATCAAAGTTTTTGTGCGCGGAGGAGTATGAAAAGAGCAAACGCTTAAGTAAATAC	600
Sbjct	172	TTAAATCAAAGTTTTTGTGCGCGGAGGAGTATGAAAAGAGCAAACGCTTAAGTAAATAC	113
Query	601	ATGTCTAGCAGTGTGCGATTGCTCTACAACAATAaaaaaaacagtttttaaaattat	660
Sbjct	112	ATGTCTAGCAGTGTGCGATTGCTCTACAACAATAAAAAAACAGTTTTAAAAATTATTT	53
Query	661	ttcaaaagtccaatttattttcggttttaaaaaaa-cctaaagaaaaa 708	
Sbjct	52	TTCAAAAGTTCAATTTATTTTCGTTTNAAAAAAACCTAAANGNAAAA 4	

Lignée 1 mir-4811 (Depuis 5')

Query	27	TGttttttttAGAATAATTAT-A-TATTTATTCTAACAACTTACTTTCCTCTTTTTCCT	84
Sbjct	7	TGTTTTTTTAG-AT-ATTTATAANTATTTATTCINA-NACTTACTTTCCTCTTTTTCCT	63
Query	85	GTCAACATGTATTTTCTAAAAGTATCGCAATCAATTTAACTTTCAATCGTTTTCAAGTTC	144
Sbjct	64	GTCAACATGTATTTTCTAAAAGTATCGCAATCAATTTAACTTTCAATCGTTTTCAAGTTC	123
Query	145	ATACTGATTTGACTTTTGAGTTGATCCTTCATGCGACATCCCAATCAAAGTTATAACTCT	204
Sbjct	124	ATACTGATTTGACTTTTGAGTTGATCCTTCATGCGACATCCCAATCAAAGTTATAACTCT	183
Query	205	GAGCTTTATTGTGTTTGATGCACCTGTATTTCGATTACACATTTCTAAAATATGACAtttt	264
Sbjct	184	GAGCTTTATTGTGTTTGATGCACCTGTATTTCGATTACACATTTCTAAAATATGACATTTT	243
Query	265	tttCCCCTTAACAATTTCAATTTTGCAAGCCTAAAAATAAGCCTAAAAATAAGATAGTTTT	324
Sbjct	244	TTTCCCCTTAACAATTTCAATTTTGCAAGCCTAAAAATAAGCCTAAAAATAAGATAGTTTT	303
Query	325	AAATGAACATTCACGTGAGAATTAAATTCGAATAATTCAATGAGTGGAATATTCTTTTAA	384
Sbjct	304	AAATGAACATTCACGTGAGAATTAAATTCGAATAATTCAATGAGTGGAATATTCTTTTAA	363
Query	385	AAATGCGGGATAAAAAAGGAAAAGTACCTAAATACCTTATAGTAAACTAGCTAGATACAC	444
Sbjct	364	AAATGCGGGATAAAAAAGGAAAAGTACCTAAATACCTTATAGTAAACTAGCTAGATACAC	423
Query	445	TTGAAAAATTAGGCTGCAGAGCACATGCTGTGAAAAATCGATTTTATGAGGAATAAATCTG	504
Sbjct	424	TTGAAAAATTAGGCTGCAGAGCACATGCTGTGAAAAATCGATTTTATGAGGAATAAATCTG	483
Query	505	TCAACTTTTTTCGCTTTACCTGAAATTTAATTCTGAATTTGACAGGTTAAGTGCAAAAGC	564
Sbjct	484	TCAACTTTTTTCGCTTTACCTGAAATTTAATTCTGAATTTGACAGGTTAAGTGCAAAAGC	543
Query	565	AAGAAAAGTACAATTCGAAAAAGACAGCAAAGCTAAATTTTACTTTGAGTTTGAAGTA	624
Sbjct	544	AAGAAAAGTACAATTCGAAAAAGACAGCAAAGCTAAATTTTACTTTGAGTTTGAAGTA	603
Query	625	CGATTCAATGTGAAATTTTCCAAATCTCACTTTGAAATACAATATTTTTTGaaaaaaagt	684
Sbjct	604	CGATTCAATGTGAAATTTTCCAAATCTCACTTTGAAATACAATATTTTTTGAAAAAAGT	663
Query	685	ttaaaacataatgtggattgaaatcaattaaaaacaaaaaagaaGTAACAGTATATTTA	744
Sbjct	664	TTAAACATAATGTGGATTGAAATCAATTAAACAAAAAAGAAAGTAACAGTATATTTA	723
Query	745	CAAGTCTGTTTCGATTTCATCGAAGGAACATACGTGCTTTAATATAGGGCGGAGCTAATCGG	804
Sbjct	724	CAAGTCTGTTTCGATTTCATCGAAGGAACATACGTGCTTTAATATAGGGCGGAGCTAATCGG	783
Query	805	GATTGACCGAATGCTACCACCGAATATGCGTGACTTACGGTAGTTGTA	852
Sbjct	784	GATTGACCGAATGCTACCACCGAATATGCGTGACTTACGGTAGTTGTA	831

Lignée 1 *mir-4811* (Depuis 3')

Query	1233	TTTTGTCGATTGCTGGCTGaaaaaaaattGCCAGATTGTGCGCTCGCAGAACCTGACCTT	1292
Sbjct	1000	TTTTGTCGATTGCTGGCNNAAAAAAATTCGCGATTGNNGCTNNGCAGAACCTGACCTT	941
Query	1293	TATCGATAACTCTATTTTGTTCACCTTTGATGACTTCCATACTCAATTTTGTATAAATGTA	1352
Sbjct	940	NNTNGATAACTCTANTNTGTTCACTTTGATGACTTCCATACTCAATTTTGTATAAATGTA	881
Query	1353	TTGCTTTCCAATTAACATATCATCTAAAGATGACATCTGAAGTTTTTATAAAGAGAAATGT	1412
Sbjct	880	TTGCTTTCCAATTAACATATCATCTAAAGATGACATCTGAAGTTTTTATAAAGAGAAATGT	821
Query	1413	GAAATGTTCAAACATGCAAGCCGAATGCCACACACTTCTTGGGAACATGTTCTGGAATG	1472
Sbjct	820	GAAATGTTCAAACATGCAAGCCGAATGCCACACACTTCTTGGGAACATGTTCTGGAATG	761
Query	1473	GTGTTAAATTTTATGAAAGCTAATGTAAAAACCATTATATTTAAACAAATTATTGTCGA	1532
Sbjct	760	GTGTTAAATTTTATGAAAGCTAATGTAAAAACCATTATATTTAAACAAATTATTGTCGA	701
Query	1533	TTCCATTTTGTGGAAGTTATGCCACCTTCAAATATCCAAAACGTTTATGATGCTCACATC	1592
Sbjct	700	TTCCATTTTGTGGAAGTTATGCCACCTTCAAATATCCAAAACGTTTATGATGCTCACATC	641
Query	1593	CAAGGTGAACATTAATCATCGAACAAAACAGTTTATGATAAACTtttttttAAAAACAA	1652
Sbjct	640	CAAGGTGAACATTAATCATCGAACAAAACAGTTTATGATAAACTTTTTTTTAAAAACAA	581
Query	1653	CATTCAGCTTCAATTTTGTGTTTGAGAGAGAACGAAAGCTTCAAAAACATTAttttttcc	1712
Sbjct	580	CATTCAGCTTCAATTTTGTGTTTGAGAGAGAACGAAAGCTTCAAAAACATTATTTTTTCC	521
Query	1713	attttttCCATGCTCAACGGGGTTTCATAATAGACAGAAAAACAATGCAGACTAAATGC	1772
Sbjct	520	ATTTTTTCCATGCTCAACGGGGTTTCATAATAGACAGAAAAACAATGCAGACTAAATGC	461
Query	1773	TGCTCCATAAGTATACGTTTCTTACGGTAGCCATTGGTATCAACACTTCGGAGAAAAC	1832
Sbjct	460	TGCTCCATAAGTATACGTTTCTTACGGTAGCCATTGGTATCAACACTTCGGAGAAAAC	401
Query	1833	CATTTAGTATGACTCTGCTGCCATCTGTATATTTTAA ~~AAAATTGTGGTATTGAACAATA~~	1892
Sbjct	400	CATTTAGTATGACTCTGCTGCCATCTGTATATTTTAA ~~AAAATTGTGGTATTGAACAATA~~	341
Query	1893	~~CCTGTGTTAAAAAGACATACATAGCAGCACCTTATAACACTCGTATTGTTTCGCTACTACAG~~	1952
Sbjct	340	~~CCTGTGTTAAAAAGACATACATAGCAGCACCTTATAACACTCGTATTGTTTCGCTACTACAG~~	281
Query	1953	~~TGCGTCCAACCTCTATAGCACCCCTCCAATTTCTACAGCCACAATTGTTAGTTTGGAGT~~	2012
Sbjct	280	~~TGCGTCCAACCTCTATAGCACCCCTCCAATTTCTACAGCCACAATTGTTAGTTTGGAGT~~	221
Query	2013	GAAACTGTCAGACTACAGTGAAGAATGGAATAAAAAATTTAAAAAGCAGTCTGGATACTAAA	2072
Sbjct	220	GAAACTGTCAGACTACAGTGAAGAATGGAATAAAAAATTTAAAAAGCAGTCTGGATACTAAA	161

Query	2073	GTTTTTTAATAAGGAAAGTTCAGAAAAACGACCACAGGTTGCGTTCTCGATGCTGTAAAC	2132
Sbjct	160	GTTTTTTAATAAGGAAAGTTCAGAAAAACGACCACAGGTTGCGTTCTCGATGCTGTAAAC	101
Query	2133	GGTAAAAAGGTAGAGCACTCCattttttgcacattttgtaaaaatatcaaattaaatgta	2192
Sbjct	100	GGTAAAAAGGTAGAGCACTCCattttttgcacattttgtaaaaatatcaaattaaatgta	41
Query	2193	atttttattttgttt-ca-ggtotacatattttttccett	2230
Sbjct	40	ATTTTATTTTGTTCATGGTCTACANANTTTTCCCTT	1

Lignée 2 mir-1820 (Depuis 5')

Query	22	TATCTTATTCTCTCTGTTGATTTTCGGCCAATTATCCTTTGTGACCACTCCCACTGTGTTG	81
Sbjct	1	TATCNTATTCTCTCTGTTGATTTTCGG-NNATTATCCTTTGTGACCACTCCCACTGTGTTG	59
Query	82	CAGATGCGGAGCACACATATGGTTTCGAGTATTCTCGACACGTGCACAATTCTTTGTTTG	141
Sbjct	60	CAGATGCGGAGCACACATATGGTTTCGAGTATTCTCGACACGTGCACAATTCTTTGTTTG	119
Query	142	TTAAATTGAGCTGGCAAGATATTATTTTTATTCAAAATGTTTAACTGATATAAGCTATT	201
Sbjct	120	TTAAATTGAGCTGGCAAGATATTATTTTTATTCAAAATGTTTAACTGATATAAGCTATT	179
Query	202	TAATTTGGAATATTATTTAATGACTAACAAGTTTGGTTTGACATTTTATTCGCCTCTCAC	261
Sbjct	180	TAATTTGGAATATTATTTAATGACTAACAAGTTTGGTTTGACATTTTATTCGCCTCTCAC	239
Query	262	AAACTTGAAAATTCTCAAAGCGAACCCCTAATTACTACTGTTTACAACGTTTGACATAAT	321
Sbjct	240	AAACTTGAAAATTCTCAAAGCGAACCCCTAATTACTACTGTTTACAACGTTTGACATAAT	299
Query	322	GACAACCATTTTTCATTGATCTACTTTACTCTACGTATTCTTCAAAAATTGCATTTTCCA	381
Sbjct	300	GACAACCATTTTTCATTGATCTACTTTACTCTACGTATTCTTCAAAAATTGCATTTTCCA	359
Query	382	TCTTTTGATTGTTTTTCGATGATGTTTCGTTAAATCGGTATAAGCGAACCATTTGTAAACAA	441
Sbjct	360	TCTTTTGATTGTTTTTCGATGATGTTTCGTTAAATCGGTATAAGCGAACCATTTGTAAACAA	419
Query	442	TCAAAGAATGGAGAATCAATTTATGATCTGGAAGATGTAAGTTTATAAAAAATGAAGGGT	501
Sbjct	420	TCAAAGAATGGAGAATCAATTTATGATCTGGAAGATGTAAGTTTATAAAAAATGAAGGGT	479
Query	502	AACAAAACGTTTTCCATTATATGAATTCTTCAAAAAAGATTAAATCAAAGTTTTGTGCGG	561
Sbjct	480	AACAAAACGTTTTCCATTATATGAATTCTTCAAAAAAGATTAAATCAAAGTTTTGTGCGG	539
Query	562	CGGAGGAGTATGAAAAGAGCAAACGCTTAACTGAAATACATGCTAGCAGTGTGCGATTGC	621
Sbjct	540	CGGAGGAGTATGAAAAGAGCAAACGCTTAACTGAAATACATGCTAGCAGTGTGCGATTGC	599
Query	622	TCTACAACAATaaaaaaacagttttaaaaaattattttcaaaagttcaatttatttt	681
Sbjct	600	TCTACAACAATAAAAAAAACAGTTTTAAAAATTATTTTTCAAAAGTTCAATTTATTTT	659
Query	682	cgttttaaaaaaaacctaagaaaaaTGTTTCAGTCAAGGTTTTGAGAA	730
Sbjct	660	CGTTTTAAAAAAACCTAAAGAAAAATGTTTCAGTCAAGGTTTTGAGAA	708

Lignée 2 mir-1820 (Depuis 3')

Query	1	ATTATTTCAATTCAAATTGACTATCTTATTCTCTCTGTTGATTTTCGGCCAATTATCCTTT	60
Sbjct	703	ATTATTTCAATTCAAATTGACTATCTTATTCTCTCTGTTGATTTTCGGCCAATTATCCTTT	644
Query	61	GTGACCACTCCCACTGTGTTGCAGATGCGGAGCACACATATGGTTTCGAGTATTCTCGAC	120
Sbjct	643	GTGACCACTCCCACTGTGTTGCAGATGCGGAGCACACATATGGTTTCGAGTATTCTCGAC	584
Query	121	ACGTGCACAATTCTTTGTTTGTAAATTGAGCTGGCAAGATATTATTTTTATTCAAAATG	180
Sbjct	583	ACGTGCACAATTCTTTGTTTGTAAATTGAGCTGGCAAGATATTATTTTTATTCAAAATG	524
Query	181	TTTAAACTGATATAAGCTATTAAATTGGAATATTATTTAATGACTAACAAGTTTGGTTT	240
Sbjct	523	TTTAAACTGATATAAGCTATTAAATTGGAATATTATTTAATGACTAACAAGTTTGGTTT	464
Query	241	GACATTTTATTTCGCCTCTCACAACCTTGAAAATTCTCAAAGCGAACCCCTAATTACTAC	300
Sbjct	463	GACATTTTATTTCGCCTCTCACAACCTTGAAAATTCTCAAAGCGAACCCCTAATTACTAC	404
Query	301	TGTTTACAACGTTTGCATAATGACAACCATTTTTCATTGATCTACTTTACTCTACGTATT	360
Sbjct	403	TGTTTACAACGTTTGCATAATGACAACCATTTTTCATTGATCTACTTTACTCTACGTATT	344
Query	361	TTCAAAAATTGCATTTTCCATCTTTTGATTGTTTTTCGATGATGTTTCGTTAAATCGGTA	420
Sbjct	343	TTCAAAAATTGCATTTTCCATCTTTTGATTGTTTTTCGATGATGTTTCGTTAAATCGGTA	284
Query	421	TAAAGCGAACCATTTGTAAACAATCAAAGAATGGAGAATCAATTTATGATCTGGAAGATGTA	480
Sbjct	283	TAAAGCGAACCATTTGTAAACAATCAAAGAATGGAGAATCAATTTATGATCTGGAAGATGTA	224
Query	481	AGTTTATAAAAAATGAAGGGTAACAAAACGTTTCCATTATATGAATTCTTCAAAAAAGA	540
Sbjct	223	AGTTTATAAAAAATGAAGGGTAACAAAACGTTTCCATTATATGAATTCTTCAAAAAAGA	164
Query	541	TTAAATCAAAGTTTTTGTGCGCGGAGGAGTATGAAAAGAGCAAACGCTTAACTGAAATAC	600
Sbjct	163	TTAAATCAAAGTTTTTGTGCGCGGAGGAGTATGAAAAGAGCAAACGCTTAACTGAAATAC	104
Query	601	ATGTCTAGCAGTGTGCGATTGCTCTACAACAATaaaaaaacagttttaaaaaattatttt	660
Sbjct	103	ATGTCTAGCAGTGTGCGATTGCTCTACAACAATAAAAAAAACAGTTTTAAAAATTATTT	44
Query	661	ttcaaaagtccaattttattttcgtttttaaaaaaacctaaaa	702
Sbjct	43	TTCAAAAGTTCAATTTATTTTCGTTT-AAAAAANCCTAAAA	3

Lignée 2 mir-4811 (Depuis 5')

Query	26	ATGttt-tttttAGAATAATTTATATATTTATTTCTAACAACCTTACTTTCCTTCTTTTCTCT	84
Sbjct	1	ATGTTGTTTTTAGNATAANTTATANATTTATTTCTAACAACCTTACTTTCCTTCTTTTCTCT	60
Query	85	GTCAACATGTATTTTCTAAAAGTATCGCAATCAATTTAACTTTCAATCGTTTTCAAGTTC	144
Sbjct	61	GTCAACATGTATTTTCTAAAAGTATCGCAATCAATTTAACTTTCAATCGTTTTCAAGTTC	120
Query	145	ATACTGATTTGACTTTTGAGTTGATCCTTCATGCGACATCCCAATCAAAGTTATAACTCT	204
Sbjct	121	ATACTGATTTGACTTTTGAGTTGATCCTTCATGCGACATCCCAATCAAAGTTATAACTCT	180
Query	205	GAGCTTTATTGTGTTTGATGCACCTGTATTCGATTACACATTTCTAAAATATGACAtttt	264
Sbjct	181	GAGCTTTATTGTGTTTGATGCACCTGTATTCGATTACACATTTCTAAAATATGACATTTT	240
Query	265	tttCCCCCTTAACAATTTTCATTTTGCAAGCCTAAAAATAAGCCTAAAAATAAGATAGTTTT	324
Sbjct	241	TTTCCCCCTTAACAATTTTCATTTTGCAAGCCTAAAAATAAGCCTAAAAATAAGATAGTTTT	300
Query	325	AAATGAACATTCACGTGAGAATTAAATTCGAATAATTCATGAGTGGAATATTCTTTTAA	384
Sbjct	301	AAATGAACATTCACGTGAGAATTAAATTCGAATAATTCATGAGTGGAATATTCTTTTAA	360
Query	385	AAATGCGGGATAAAAAAGGAAAAGTACCTAAATACCTTATAGTAACTAGCTAGATACAC	444
Sbjct	361	AAATGCGGGATAAAAAAGGAAAAGTACCTAAATACCTTATAGTAACTAGCTAGATACAC	420
Query	445	TTGAAAAATTAGGCTGCAGAGCACATGCTGTGAAAATCGATTTTATGAGGAATAAATCTG	504
Sbjct	421	TTGAAAAATTAGGCTGCAGAGCACATGCTGTGAAAATCGATTTTATGAGGAATAAATCTG	480
Query	505	TCAACTTTTTTCGCTTTACCTGAAATTTAATTCTGAATTTGACAGGTTAAGTGCAAAAGC	564
Sbjct	481	TCAACTTTTTTCGCTTTACCTGAAATTTAATTCTGAATTTGACAGGTTAAGTGCAAAAGC	540
Query	565	AAGAAAAGTACAATTCGAAAAAGACAGCAAAGCTAAATTTTACTTTGAGTTTGAAGTA	624
Sbjct	541	AAGAAAAGTACAATTCGAAAAAGACAGCAAAGCTAAATTTTACTTTGAGTTTGAAGTA	600
Query	625	CGATTCAATGTGAAATTTTCCAAATCTCACTTTGAAATACAATATTTTTTGaaaaaagt	684
Sbjct	601	CGATTCAATGTGAAATTTTCCAAATCTCACTTTGAAATACAATATTTTTTGAAAAAAGT	660
Query	685	ttaaaaacataatgtggattgaaatcaattaaaaacaaaaaagaaaGTAAACAGTATATTTA	744
Sbjct	661	TTAAACATAATGTGGATTGAAATCAATTAAACAAAAAAGAAAGTAACAGTATATTTA	720
Query	745	CAAGTCTGTTTCGATTTCATCGAAGGAACATACGTGCTTTAATATAGGGCGGAGCTAATCGG	804
Sbjct	721	CAAGTCTGTTTCGATTTCATCGAAGGAACATACGTGCTTTAATATAGGGCGGAGCTAATCGG	780
Query	805	GATTGACCGAATGCTACCACCGAATATGCGTGACTTACGGTAGTTGTAGGCCCCACCATT	864
Sbjct	781	GATTGACCGAATGCTACCACCGAATATGCGTGACTTACGGTAGTTGTAGGCCCCACCATT	840

Query	865	AGGCCCCACGATCAACATATTACGTTTTTATTCTACTTCTAATTGTATTTTCTACTCATG	924
Sbjct	841	AGGCCCCACGATCAACATATTACGTTTTTATTCTACTTCTAATTGTATTTTCTACTCATG	900
Query	925	CAAGTTCGTGGTACTGACCAATGTTAGTGTTTGaaaaaaaGTATTTTCGTTGCACGCTC	984
Sbjct	901	CAAGTTCGTGGTACTGACCAATGTTAGTGTTTGAAAAAANAGTATTTTCGTTGCACGCTC	960
Query	985	ATTGTTCACTACTTTTGAATTTGATTGTTCAAGAAAATTGTACGAAAGTTaaaaaa-aa	1043
Sbjct	961	ATTGTTCACTACTTTTGAATTTGATTGTTCAAGAAAATTGTACGAAAGTTAAANAAANAA	1020
Query	1044	TTGGC-TTTTGTGCGCATGTGC-TGGACGTCGTAT-CGTTCTTAt-ttttttGA	1093
Sbjct	1021	TTGGCTTTTTGTGCGCATGTGCNTGGACGTCGTANNCGNTCTTANNTTTTTGA	1074

Lignée 2 mir-4811 (Depuis 3')

Query	1140	ATAAAGTTATTCAATAAAGATGCAATCAGTAAAGTTCAGCATTTTGTGCACAAAAGTACAC	1199
Sbjct	1090	ATAAAGTTNTTCANTAAAGATGCAATCAGTAAAGTTCAGCATTTTGTGCACAAAAGTNCAC	1031
Query	1200	GCTTTACAATATACAGAAATTCAAAACTATATTTTGTGCGATTGCTGGCTGAAAAAAA	1259
Sbjct	1030	GCTTTACAATATACAGAAATTCAAAACTATATNTTGTGCGATTGCTGGCTGAAAAAAA	971
Query	1260	TTGCCAGATTGTGCGCTCGCAGAACCTGACCTTTATCGATAACTCTATTTTGTTCACCTT	1319
Sbjct	970	TTGCCAGATTGTGCGCTCGCAGAACCTGACCTTTATCGATAACTCTATTTTGTTCACCTT	911
Query	1320	GATGACTTCCATACTCAATTTTGTATAAATGTATTGCTTTCCAATTAACATATCATCTAAA	1379
Sbjct	910	GATGACTTCCATACTCAATTTTGTATAAATGTATTGCTTTCCAATTAACATATCATCTAAA	851
Query	1380	GATGACATCTGAAGTTTTTATAAAGAGAAATGTGAAAAATGTTCAACATGCAAGCCGAAT	1439
Sbjct	850	GATGACATCTGAAGTTTTTATAAAGAGAAATGTGAAAAATGTTCAACATGCAAGCCGAAT	791
Query	1440	GCCACACACTTCTTGGGAACATGTTCTGGAATGGTGTAAATTTTATGAAAGCTAATGTA	1499
Sbjct	790	GCCACACACTTCTTGGGAACATGTTCTGGAATGGTGTAAATTTTATGAAAGCTAATGTA	731
Query	1500	AAAACCATTATATTTAAAACAAATTATTGTGCGATTCCATTTTGTGGAAGTTATGCCACCT	1559
Sbjct	730	AAAACCATTATATTTAAAACAAATTATTGTGCGATTCCATTTTGTGGAAGTTATGCCACCT	671
Query	1560	TCAAAATATCCAAAACGTTTATGATGCTCACATCCAAGGTGAACATTAATCATCGAACAAA	1619
Sbjct	670	TCAAAATATCCAAAACGTTTATGATGCTCACATCCAAGGTGAACATTAATCATCGAACAAA	611
Query	1620	ACAGTTTATGATAAACTtttttttAAAAATACAACATTCAGCTTCAATTTTGTGTTTGAGA	1679
Sbjct	610	ACAGTTTATGATAAACTTTTTTTAAAAATACAACATTCAGCTTCAATTTTGTGTTTGAGA	551
Query	1680	GAGAACGAAAGCTTCAAAAACATTAttttttccattttttCCATGCTCAACGGGGTTTCA	1739
Sbjct	550	GAGAACGAAAGCTTCAAAAACATTATTTTTTCCATTTTTCATGCTCAACGGGGTTTCA	491
Query	1740	TAATAGACAGAAAAACAATGCAGACTAAATGCTGCCTCCATAAGTATACGTTTCTTACG	1799
Sbjct	490	TAATAGACAGAAAAACAATGCAGACTAAATGCTGCCTCCATAAGTATACGTTTCTTACG	431
Query	1800	GTAGCCATTGGTATCAACACTTCGGAGAAAACTCATTTAGTATGACTCTGCTGCCATCTG	1859
Sbjct	430	GTAGCCATTGGTATCAACACTTCGGAGAAAACTCATTTAGTATGACTCTGCTGCCATCTG	371
Query	1860	TATATTTTAA ~~AAAAATTGTGGTATTGAACAATACCTGTGTTAAAAGACATACATAGCAGC~~	1919
Sbjct	370	TATATTTTAA ~~AAAAATTGTGGTATTGAACAATACCTGTGTTAAAAGACATACATAGCAGC~~	311
Query	1920	~~ACCTTATAACACTCGTATTGTTTCGCTACTACAGT~~ GCGTCCAACCTCTATAGCACCCCTC	1979
Sbjct	310	~~ACCTTATAACACTCGTATTGTTTCGCTACTACAGT~~ GCGTCCAACCTCTATAGCACCCCTC	251

Query	1980	CAATTTCTACAGCCACAATTGTTAGTTTTGAGTGAACTGTGAGACTACAGTGAAGAATG	2039
Sbjct	250	CAATTTCTACAGCCACAATTGTTAGTTTTGAGTGAACTGTGAGACTACAGTGAAGAATG	191
Query	2040	GAATAAAAAATTTAAAAGCAGTCTGGATACTAAAGTTTTTTAATAAGGAAAGTTCAGAAAA	2099
Sbjct	190	GAATAAAAAATTTAAAAGCAGTCTGGATACTAAAGTTTTTTAATAAGGAAAGTTCAGAAAA	131
Query	2100	ACGACCACAGGTTGCGTTCTCGATGCTGTAAACGGTAAATGGTAGAGCACTCCattttt	2159
Sbjct	130	ACGACCACAGGTTGCGTTCTCGATGCTGTAAACGGTAAATGGTAGAGCACTCCATTTTT	71
Query	2160	tgcacattttgtaaaaatatcaaattaaatgtaatttttattttgtttca-ggtctacat	2218
Sbjct	70	TGCACATTTGTAAAAATATCAAATTAATGTAATTTTATTNGNNTCNTGGTCTACAT	11
Query	2219	attttttccct	2229
Sbjct	10	ANTTTT-CCCT	1

Régulation post-transcriptionnelle du gène *unc-54* de *Caenorhabditis elegans* identifiée *in vivo* par un système de double rapporteurs fluorescents.

Le développement de *Caenorhabditis elegans* est finement régulé et aboutit au nombre constant de 959 cellules par individu. Une partie de ces régulations repose sur les microARN, ARN simple brin d'environ 22 nucléotides. Ils peuvent diminuer l'expression des gènes en ciblant des séquences homologues dans les régions 3' non transcrites (3'UTR) des ARN messagers. Pour déterminer l'impact de cette régulation, nous utilisons une méthode permettant de visualiser *in vivo* une différence d'expression induite post-transcriptionnellement. Cette méthode se base sur l'utilisation de deux protéines fluorescentes : la GFP (verte) et la mCherry (rouge) exprimées sous le contrôle du promoteur du gène d'intérêt. La mCherry est suivie par un 3'UTR contrôle et reflète l'activité du promoteur. La GFP est suivie par le 3'UTR du gène d'intérêt et subit les mêmes régulations que le gène endogène. La différence d'expression entre les deux protéines devrait donc refléter la régulation s'opérant sur le 3'UTR du gène. La transparence de *C. elegans* permet de localiser les protéines rapportrices par microscopie en fluorescence à tous les stades du développement dans l'organisme entier. Initialement, le 3'UTR du gène *unc-54* (myosine de type II) a été choisi comme contrôle. Le rapporteur bicolore a révélé une régulation s'opérant sur *unc-54* 3'UTR jusqu'ici considérée comme permissif. La régulation observée s'opère dans les muscles et les neurones ADF. La caractérisation de cette régulation a permis de mettre en évidence le rôle potentiel de *mir-1820*. Le profil d'expression de *mir-1820* a pu être établi grâce à une fusion avec la GFP et correspond au profil des régulations observées sur *unc-54* 3'UTR.

Mots clés : *Caenorhabditis elegans*, microARN, régulation, *unc-54*, transgénèse, *mir-1820*.sélection

Post-transcriptional regulation of *unc-54* *Caenorhabditis elegans* gene has been identified *in vivo* through a fluorescing double reporters system.

Caenorhabditis elegans development is very tightly regulated, leading to the same number of cells in each individual. Part of this regulation network relies on small single strand RNAs (miRNAs), which can target homologous sequences in the 3' untranslated regions (3'UTR) of messenger RNAs. We want to investigate the contribution of miRNAs during neurons differentiation. In order to study the miRNA contribution to gene regulation we use double fluorescent reporters that allow us to visualize the posttranscriptional contribution to regulation throughout development. The GFP and the mCherry are expressed under the control of the gene promoter, but followed by either the 3'UTR of interest, or a control 3'UTR. We first chose as a control 3'UTR the *unc-54* 3'UTR (myosin class II). The gene *unc-54* is expressed through all larval stages and in the adult worms. The two colors reporter system showed that *unc-54* 3'UTR undergoes a regulation in the ADF pair of neurons and partially in the body wall muscle. The characterization of this regulation pointed out a potential role for *mir-1820*. The GFP was cloned between *mir1820* 5' and 3' sequences and the construction displayed an expression profile overlapping with the regulation pattern observed on *unc-54* 3'UTR.

Keywords: *Caenorhabditis elegans*, microRNA, regulation, *unc-54*, transgenesis, *mir-1820*.