



# Stratégies d'analyse spatio-temporelle de lépissage alternatif chez *Caenorhabditis elegans*

Jonathan Millet

► **To cite this version:**

Jonathan Millet. Stratégies d'analyse spatio-temporelle de lépissage alternatif chez *Caenorhabditis elegans*. Génétique. Université de Bordeaux, 2015. Français. <NNT : 2015BORD0437>. <tel-01286111>

**HAL Id: tel-01286111**

**<https://tel.archives-ouvertes.fr/tel-01286111>**

Submitted on 10 Mar 2016

**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.

THÈSE PRÉSENTÉE

POUR OBTENIR LE GRADE DE

**DOCTEUR DE**

**L'UNIVERSITÉ DE BORDEAUX**

ÉCOLE DOCTORALE DES SCIENCES DE LA VIE ET DE LA SANTÉ

SPÉCIALITÉ GÉNÉTIQUE

Par Jonathan MILLET

Stratégies d'analyse spatio-temporelle de l'épissage alternatif chez

*Caenorhabditis elegans*

Sous la direction de : Denis DUPUY

Soutenue le 18 décembre 2015

Membres du jury :

M. DOIGNON, François	Professeur - Université de Bordeaux	Président
M. MUCHARDT, Christian	Directeur de Recherche - Institut Pasteur	Rapporteur
M. POULIN, Gino	Lecturer - University of Manchester	Rapporteur
Mme MASSÉ, Karine	Maitre de conférence - Université de Bordeaux	Membre invité
M. DUPUY, Denis	Chercheur - Université de Bordeaux	Directeur de thèse



## **Stratégies d'analyse spatio-temporelle de l'épissage alternatif dans le système nerveux de *Caenorhabditis elegans***

L'épissage alternatif est un mécanisme de régulation de l'expression des gènes ayant pris une importance croissante dans l'étude du vivant. Si des méthodes existent pour déterminer les gènes qui y sont soumis, peu d'outils sont disponibles pour suivre ces événements d'épissage *in vivo* au cours du développement. Pourtant, la caractérisation des régulations sous-jacentes à ces événements et la détermination des facteurs impliqués sont dépendantes de stratégies fiables pour les visualiser dans des conditions physiologiques.

Nous avons développé un système adapté à l'étude d'événements d'épissage basé sur un rapporteur fluorescent bicolore. Nous l'avons appliqué à cinq gènes de l'organisme modèle *Caenorhabditis elegans* et avons suivi leur épissage *in vivo*.

Parmi les différents gènes suivis, deux d'entre eux suivaient un modèle d'épissage potentiellement stochastique, un autre une absence d'épissage alternatif détectable. Les deux derniers gènes présentent un profil d'épissage spécifique à certain types cellulaires mais ont un effet toxique sur l'organisme lorsque nous les avons exprimés à partir de concatémères extrachromosomiques. Pour remédier à cela, nous avons choisi de mettre en place une méthode simplifiée d'insertion en simple copie des rapporteurs utilisant le CRISPR-Cas.

Nos résultats indiquent que le système rapporteur fonctionne avec succès. Cependant, il peut encore être amélioré pour se rapprocher des taux physiologiques de transcription grâce à une insertion en simple copie dans le génome de l'organisme. Nous avons également révélé un événement sous le contrôle de régulations spatiales, temporelles et conditionnelles. De plus, nous avons créé une série de constructions capables de déterminer les éléments en *cis* impliqués dans la régulation du gène *top-1*.

*Mots clés : Caenorhabditis ; Caenorhabditis elegans ; C.elegans ; Système nerveux ; Neurones ; Épissage alternatif ; Rapporteur fluorescent ; Transgénèse ; CRISPR-Cas*

### **Strategies for spatio-temporal analysis of alternative splicing in *Caenorhabditis elegans* nervous system**

Alternative splicing is a regulatory mechanism of gene expression which is increasingly studied in Life Science. Methods exist to study this mechanism but specific tools to follow each alternative splicing event in a spatio-temporal manner are lacking. Yet, the characterization of the regulation and the elements that determines them depends on valide strategies for visualising them in physiological conditions.

We have developed a dual-fluorescent reporter-based system in order to follow alternative splicing event regulation *in vivo*. It has been applied to five different genes in the model organism *Caenorhabditis elegans*. Among the genes followed, two follow a potentially stochastic scheme, one show no visible sign of alternative splicing. The last display tissue specific splicing patterns but developed a toxic effect in the animal when expressed from a multicopy extrachromosomal array. To remediate this problem, we decided to develop a method that allows for simpler single copy insertion of fluorescent reporter using CRISPR-Cas.

Our results indicates that the dual-fluorescent reporter works well. However, this system can be upgraded by getting close to physiological rates of transcription allowed by single-copy insertion in the genome of *C.elegans*. We also discovered an alternatiove splicing event which follows a spatial, temporal and conditionnal regulation. Moreover, we constructed a set of different reporter to unravel the regulation observed in the gene *top-1*.

*Keywords : Caenorhabditis ; Caenorhabditis elegans ; C.elegans ; Nervous system ; Neurons ; Alternative splicing ; Fluorescent reporter ; Genome editing ; CRISPR-Cas*

## Unité de recherche

ARN : Régulations naturelles et artificielles - INSERM U869

Institut Européen de Chimie et de Biologie

2, rue Robert Escarpit 33067 Pessac

## Remerciements

*J'aimerais tout d'abord remercier Denis avec lequel j'ai appris durant la thèse et le Master. Merci de m'avoir fait confiance et de m'avoir guidé. Tu es un super chef et un « amigo » comme dirait le faux espagnol. Je retiendrai qu'on peut être un excellent chercheur et quelqu'un de bien tout en portant un T-shirt Captain Cavern et des espadrilles. Et n'oublie pas : « The Wheel weaves as the Wheel wills. Nom de Zeus !! »*

*J'aimerais ensuite remercier les membres du jury d'avoir accepté d'en faire partie et d'avoir pris le temps de lire ses travaux. J'aimerais remercier un peu plus particulièrement Karine Massé et François Doignon qui m'ont également formé et suivi durant plusieurs années.*

*Je ne serai jamais arriver là sans les professeurs, maîtres de conférences et enseignants de l'université Bordeaux II (oui je sais que c'est Bordeaux tout court maintenant) et plus particulièrement Didier Thoraval, qui m'a supporté bien trop longtemps, Alain blanchard, Etienne Roux et Gérard Barroso.*

*J'ai beaucoup appris au labo grâce à Karine et Esther.*

*Merci à Julia pour m'avoir prouvé qu'on avance toujours plus vite à dos de requin ... ou habillé en requin ... ou avec un requin en peluche au-dessus de vous et qui vous juge en permanence. Je garderai des Jenlains au frais.*

*Merci à Ernesto pour son soutien morale durant les soirs de manips et les journées COPAS. À Mowgli pour sa sagesse pratique et ses protocoles du fin fond de la jungle. Mais surtout, merci à Camila, avec qui j'ai partagé un labo ... en quelque sorte.*

*À « Rico » Cornes, le faux espagnol qui parle français et s'appelle Eric. Merci pour les soirées « At Boïvin's », les gri-gris ridicules et les histoires épiques – mais qui n'ont jamais aucun de sens. Ça a fini par marcher, et sans utiliser l'eau du bénitier.*

*Merci à tous les membres de l'IECB et de l'unité U869 (oui je sais que c'est plus le bon numéro) et notamment Kati, Stéphanie, Tien et Marilyn. Merci aussi à Kellian, Enzo, Emilie et Marine. Bon courage à Florian.*

*à Julie avec qui j'ai partagé un Master, un appart et un chat. Merci pour les ballades en voitures, le soutien durant la rédaction et les plats ... sans pique-assiette.*

*Merci à mes amis Thibaud et Angélique, Arnaud et Delphine, Shaan, Nina, Toff, Aurore et Bamby, à ma famille et Zohra qui doivent déjà le savoir.*

*A toute personne qui lit les remerciements d'une thèse qui traîne au labo... la pause-café est fini (oui je sais, c'est décevant comparé à d'autres).*





<b>REMERCIEMENTS</b> .....	<b>- 5 -</b>
<b>LISTE DES ABREVIATIONS</b> .....	<b>- 11 -</b>
<b>INTRODUCTION</b> .....	<b>- 13 -</b>
<b>I. DE L'INFORMATION GENOMIQUE AUX FONCTIONS BIOLOGIQUES</b> .....	<b>- 15 -</b>
1. Généralités .....	- 15 -
2. Du cellulaire au moléculaire.....	- 15 -
3. De l'importance de la régulation.....	- 16 -
<b>II. L'ARNM</b> .....	<b>- 17 -</b>
1. La transcription des gènes eucaryotes .....	- 17 -
2. Étapes de maturation des ARNs pré-messagers .....	- 17 -
3. Traduction .....	- 21 -
4. Dégradation .....	- 21 -
5. Régulation de l'expression génique .....	- 25 -
<b>III. L'ÉPISSAGE ALTERNATIF</b> .....	<b>- 27 -</b>
1. Généralités .....	- 27 -
2. Effets de l'épissage alternatif.....	- 28 -
3. Un mécanisme co-transcriptionnel .....	- 29 -
4. Les acteurs de l'épissage alternatif – « Splicing code » .....	- 35 -
5. Les différents types d'épissage alternatif.....	- 40 -
6. L'impact de l'épissage alternatif sur le vivant .....	- 43 -
<b>IV. CAENORHABDITIS ELEGANS</b> .....	<b>- 48 -</b>
1. Généralités .....	- 48 -
2. Développement et reproduction .....	- 49 -
3. Découvertes majeures et champs d'étude .....	- 50 -
4. La transgénése chez <i>C. elegans</i> .....	- 52 -
5. Les rapporteurs fluorescents .....	- 57 -
<b>OBJECTIFS</b> .....	<b>- 63 -</b>
<b>RESULTATS</b> .....	<b>- 65 -</b>
<b>I. DEVELOPPEMENT DE LA STRATEGIE</b> .....	<b>- 67 -</b>
1. Le choix des gènes.....	- 67 -
2. Méthodologie.....	- 67 -
<b>II. SUIVI DES GENES <i>UNC-86</i>, <i>CHE-1</i> ET <i>GPA-13</i></b> .....	<b>- 73 -</b>
1. <i>che-1</i> .....	- 73 -
2. <i>unc-86</i> .....	- 75 -
3. <i>gpa-13</i> .....	- 79 -
4. Conclusion .....	- 81 -
<b>III. LE CAS DES GENES <i>TOP-1</i> ET <i>HRPF-1</i></b> .....	<b>- 83 -</b>
1. Le rapporteur <i>top-1</i> .....	- 83 -
2. La recherche d'éléments en cis.....	- 93 -
3. La recherche de facteurs en trans .....	- 94 -
4. Le rapporteur <i>hrpf-1</i> .....	- 95 -
5. Co-expression du gène <i>hrpf-1</i> et du rapporteur <i>top-1</i> .....	- 101 -
<b>IV. STRATEGIES D'INSERTION EN COPIE UNIQUE</b> .....	<b>- 106 -</b>
1. <i>Mos-SCI</i> .....	- 106 -
2. <i>CRISPR-Cas</i> .....	- 106 -
3. Insertion universelle .....	- 107 -
4. Méthode.....	- 110 -
<b>DISCUSSION</b> .....	<b>- 113 -</b>
<b>I. LA NECESSITE DES RAPPORTEURS <i>IN VIVO</i></b> .....	<b>- 115 -</b>

II. UNE EVOLUTION DU SYSTEME DE PAIRES DE « MINIGENES » .....	- 115 -
III. RESULTATS OBTENUS .....	- 116 -
IV. L'INSERTION UNIVERSELLE .....	- 119 -
<b>PERSPECTIVES .....</b>	<b>- 121 -</b>
I. DIRECTES .....	- 123 -
II. A LONG TERME.....	- 123 -
1) <i>La recherche de séquence en cis par crible de mutagenèse ciblée</i> .....	- 123 -
2) <i>Modification ciblée des marques post-traductionnelle</i> .....	- 124 -
<b>MATERIEL &amp; METHODES .....</b>	<b>- 125 -</b>
<b>I. MILIEUX DE CULTURES ET SOLUTIONS .....</b>	<b>- 127 -</b>
1. <i>Solutions</i> .....	- 127 -
2. <i>Milieux de cultures</i> .....	- 128 -
<b>II. BIOLOGIE MOLECULAIRE .....</b>	<b>- 130 -</b>
1. <i>Préparation d'ADN génomique</i> .....	- 130 -
2. <i>Extraction d'ARN</i> .....	- 130 -
3. <i>Amplification par PCR</i> .....	- 131 -
4. <i>RT-PCR</i> .....	- 133 -
5. <i>Assemblages des vecteurs</i> .....	- 133 -
6. <i>Mutagenèse dirigée par PCR</i> .....	- 134 -
7. <i>Transformation</i> .....	- 135 -
8. <i>Cultures bactériennes</i> .....	- 137 -
9. <i>Extraction de plasmide</i> .....	- 137 -
10. <i>Vérification des plasmides</i> .....	- 139 -
11. <i>Stock glycérol</i> .....	- 140 -
<b>III. CULTURES ET MANIPULATION DES VERS .....</b>	<b>- 141 -</b>
1. <i>Conditions de culture</i> .....	- 141 -
2. <i>Génération de lignées transgéniques</i> .....	- 141 -
3. <i>Observation au microscope</i> .....	- 142 -
4. <i>Analyse au COPAS</i> .....	- 143 -
5. <i>Synchronisation</i> .....	- 144 -
6. <i>ARN interférence</i> .....	- 144 -
<b>ANNEXES.....</b>	<b>- 145 -</b>
BIBLIOGRAPHIE .....	- 147 -

## Liste des abréviations

ADN: Acide desoxyribonucléique ;

ARN : Acide ribonucléique ;

ARNnc : ARN non-codant ;

ARNsn : petit ARN nucléaire ;

ARNsno : petit ARN nucléolaire ;

ARNt : ARN de transfert ;

ASO : Oligonucléotide antisens ;

siARN : petit ARN interférant ;

ARNi : ARN interférence ;

CTD : Domaine carboxy-terminale

CBC : Cap-Binding Complex ;

U2AF : U2 Associated Factor ;

EJC : Exon-Junction Complex ;

NMD : Nonsense-Mediated Decay ;

NMTR : Nonsense-Mediated Translational Repression ;

NLS : Nuclear Localizing Sequence ;

hnRNP : heterogeneous nuclear RiboNucleoProtein ;

SR : Ser-Arg/Serine-Arginine rich ;

ESE : Exonic Splicing Enhancer ;

ESS : Exonic Splicing Silencer ;

ISE : Intronic Splicing Enhancer ;

ISS : Intronic Splicing Silencer ;

CE : Cassette Exon ;

MXE : Mutually exclusive exon ;

IR : Intron Retention ;

ASS : Alternative Splice Site ;

AFE : Alternative First Exon ;

ALE : Alternative Last Exon ;

PTB : Polypyrimidine Tract Binding ;

MosSCI : Mos Single-Copy Insertion ;  
TALEN : Transcription Activator-Like Effector Nuclease ;  
ZFN : Zinc-Finger Nuclease ;  
CRISPR : Clustered Regularly Interspaced Short Palindromic Repeat ;  
CRISPRi : CRISPR interference ;  
crARN : CRISPR-Related ARN ;  
tracrARN : trans-associated crARN ;  
sgARN : single-guide ARN ;  
PAM : Protospacer Adjacent Motif ;  
FSS : FrameShift Sensor ;  
GFP :Green Fluorescent Protein ;  
Kb : Kilobases ;  
RBD : RNA-Binding Domain ;  
ORF : Open Reading Frame ;

# Introduction

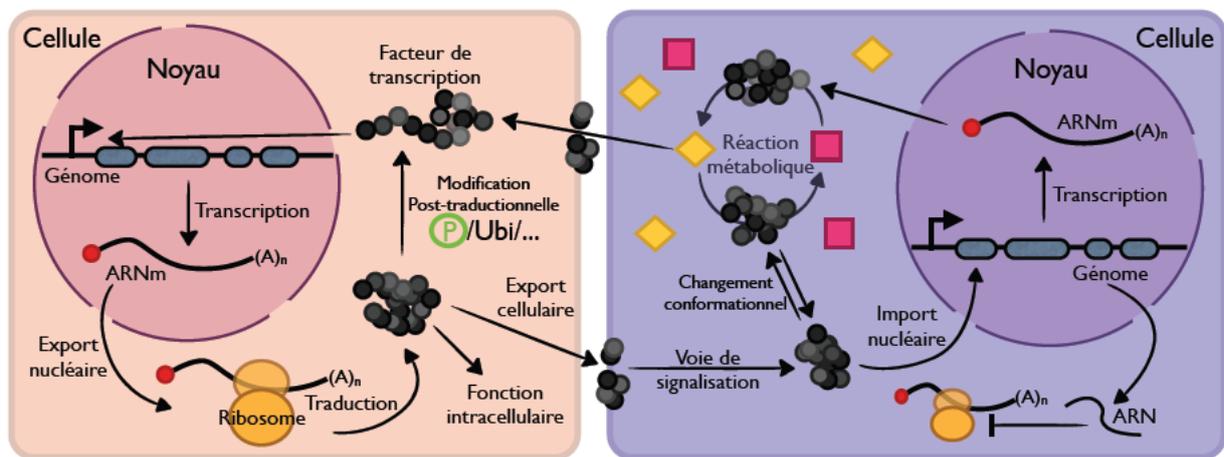


# I. De l'information génomique aux fonctions biologiques

## 1. Généralités

Le génome humain est composé de 20 000 à 25 000 gènes dispersés sur 3,4 milliards de paires de bases. Le corps humain défini par la mise en application de cette information est constitué de milliards de cellules associées sous la forme de tissus regroupés en organes pour assurer toutes les fonctions nécessaires à la survie de l'organisme.

L'unité minimale du vivant est la cellule. Capable de se reproduire de manière indépendante par la mitose, elle compose l'ensemble des diverses formes du vivant. Chez les organismes multicellulaires, elle n'est qu'un ensemble d'un tout plus grand, où la coopération et les échanges sont nécessaires à la survie de l'organisme. Ces échanges, aussi variés que nombreux (**Figure 1**), sont la base de tout système pluricellulaire.



**Figure 1 : Schéma représentant quelques mécanismes moléculaires généraux et échanges entre de deux cellules.** Les deux cellules – respectivement en rose et mauve – fonctionnent à la fois de manière indépendante (Information génétique propre ; transcription, traduction et régulations isolées) et de manière conjointe (échanges cellulaires).

## 2. Du cellulaire au moléculaire

Chez les organismes pluricellulaires, le zygote - une seule cellule- se multiplie et produit par ce biais d'autres précurseurs, qui, par migration, obtiendront leur position finale dans l'organisme et leur devenir cellulaire par un processus connu sous le nom de différenciation cellulaire.

La différenciation cellulaire est une étape cruciale du développement puisqu'il s'agit de la spécialisation fonctionnelle des cellules qui vont alors former les tissus et organes de

l'individu en devenir. Le produit de l'interaction d'un certain nombre de gènes permet cette spécification fonctionnelle des cellules.

La polyvalence du devenir cellulaire, ainsi que la capacité d'adaptation des cellules à leurs environnements, est une conséquence de la régulation des gènes, et ce, par différents mécanismes moléculaires impliquant les produits de milliers d'autres gènes.

### **3. De l'importance de la régulation**

Le séquençage du génome humain – et celui des génomes d'autres espèces – avait pour objectif de faciliter les recherches visant à analyser les relations fonctionnelles entre information génétique et fonctions cellulaires que ce soit cas par cas ou dans le cadre d'études à large spectre. Cependant, la masse de données acquise alors, additionnée des multiples études systématiques (transcriptomique, protéomique, métabolomique) générées depuis, n'a cessé de souligner l'ampleur et la complexité des interactions au sein des systèmes biologiques.

Il en découle que la fonction biologique d'un gène ne peut se définir qu'en prenant en compte l'ensemble des interactions que ses produits dérivés (ARNm/ARNnc/Protéines) entretiennent avec les produits du reste du génome.

Pour cela, il est donc nécessaire d'étudier les différentes formes de régulations révélées à ce jour, de développer des outils d'étude dans le but d'extraire des règles communes, afin de permettre la caractérisation systématique des interactions géniques. Alors seulement, l'utilisation des informations obtenues pourra permettre la modélisation de systèmes biologiques complexes capables de mimer fidèlement la réalité biologique.

L'ARN, en tant que produit intermédiaire entre information génétique et activité protéique, et en tant qu'acteur essentiel de mécanismes de régulation, est un socle important des fonctions cellulaires. De ce fait, sa régulation constitue un sujet de choix pour l'étude biologique.

## II. L'ARNm

### 1. La transcription des gènes eucaryotes

Chez les eucaryotes, les gènes sont transcrits par trois complexes protéiques connus sous le nom d'ARN polymérases. Chacun de ces complexes est constitué de plus d'une dizaine de protéines et est recruté aux régions promotrices des gènes par le biais de séquences nucléotidiques (*cis*-éléments/élément en *cis*) et de protéines (*trans*-facteur/facteur en *trans*) appelées les facteurs de transcription.

L'ARN polymérase I permet la transcription de tous les ARN ribosomiaux, à l'exception de l'ARN ribosomal 5S. Ce complexe transcrit plus de la moitié des ARNs totaux de la cellule (Drygin et al., 2010).

L'ARN polymérase II permet la transcription des ARNs pré-messagers – précurseurs des ARNs messagers traduits par les ribosomes en chaîne d'acides aminées -, la majorité des microARNs et la majorité des petits ARN nucléaires (ARNsn/snRNA – small nuclear RNA)(Tamura, 1996).

L'ARN polymérase III permet quant à elle, la transcription de l'ARN ribosomique 5S, des ARNs de transfert, l'ARNsn U6, le ARNsn 7SK, des microARNs ainsi que des ARNs nucléolaires (ARNsno/snoRNA – small nucleolar RNA)(Dieci et al., 2007). Les plantes possèdent également une ARN polymérase IV qui transcrit des petits ARNs interférents.

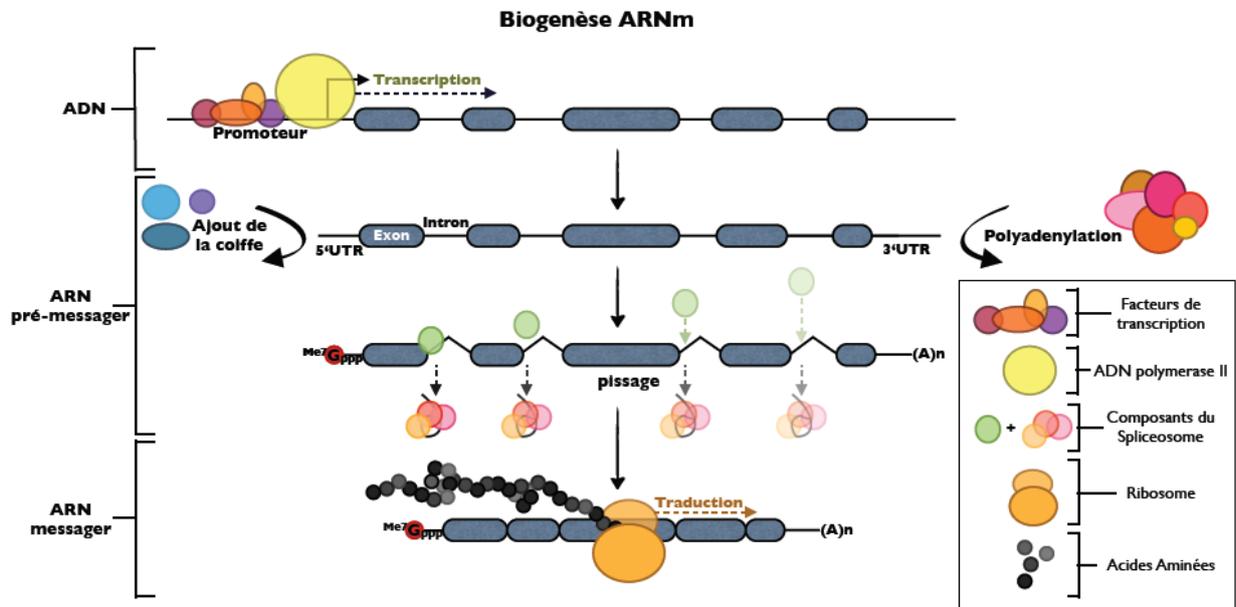
Seuls les ARNs pré-messagers transcrit par l'ARN polymérase II sont traduits en protéines par les ribosomes après avoir subi une série d'étapes de maturation.

### 2. Étapes de maturation des ARNs pré-messagers

Un ARN pré-messager synthétisé par l'ARN polymérase II va subir trois étapes essentielles de maturation avant d'être exporté vers le cytoplasme pour y être traduit :

- l'ajout de la coiffe à l'extrémité 5' de l'ARNm,
- l'épissage des introns ainsi que
- la polyadénylation de l'extrémité 3' de l'ARNm.

Ces étapes ont pour but de stabiliser l'ARN messenger final en protégeant ses extrémités de la dégradation ainsi que de révéler sa séquence codante (**Figure 2**).



**Figure 2 : Biogenèse et maturation des ARNm.** L'ARN pré-messager synthétisée va subir trois étapes de maturation : l'ajout de la coiffe, l'épissage et la polyadénylation. Une fois mature, l'ARNm est exporté dans le cytoplasme puis traduit en protéine par les ribosomes.

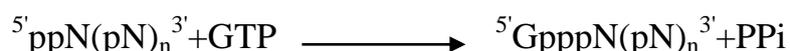
### a) L'ajout de la coiffe

L'ajout de la coiffe est un mécanisme co-transcriptionnel qui démarre sur un brin d'ARN naissant (20-25 nucléotides de long) (Shatkin and Manley, 2000) et nécessite une interaction avec un domaine de l'ARN polymérase II connu sous le nom de CTD (CarboxyTerminale Domain). L'activité de trois enzymes est nécessaire durant cette étape :

(1) Une déphosphorylation du premier nucléotide de l'ARNm (qui possède trois phosphate) par une phosphatase :



(2) L'ajout d'une guanosine monophosphate en 5' du premier nucléotide par une guanylyltransférase :



(3) La méthylation de la guanosine monophosphate par une méthyltransférase :



(Ghosh and Lima, 2010)

Une fois ces réactions terminées, le *CBC* pour *Cap-Binding Complex* va alors se fixer sur le nucléotide terminal et composer la coiffe. Ce complexe sera remplacé après le premier tour de traduction – souvent nommé *pioneer round* – par des facteurs initiateurs de traduction (Ishigaki et al., 2001). Ce premier tour de traduction est souvent considéré comme une étape de vérification des ARNm (voir plus loin).

### **b) L'épissage**

La découverte de l'épissage remonte aux années 70, lorsqu'il a été mis en évidence qu'un ARN messager mature ne s'hybride pas un seul segment d'ADN simple brin complémentaire mais à plusieurs fragments d'ADN discontinus (Berget and Sharp, 1977)(Louise T Chow, Richard E. Gelinas, Thomas R., 1977). Depuis ces travaux, une grande partie du mécanisme a été élucidé et a permis de définir les segments conservés dans l'ARNm sous le nom d'exons par opposition à ceux éliminés durant l'épissage, les introns – pour *INTRagenic RegiON* (Tonegawa et al., 1978).

Le spliceosome - le complexe ribonucléoprotéique responsable de l'épissage - est constitué de petits ARNs nucléaires non-codants riche en uridine (U-rich small nuclear RNAU - ou snRNA) et de leurs protéines associées. Ces éléments, nommés U1 à U6, se fixent séquentiellement sur les introns.

Le petit ARN nucléaire U1 sert de précurseur à la reconnaissance des séquences introniques en se fixant, via son extrémité 5', au site d'épissage 5' des introns. U2 se fixe au point de branchement tandis que la protéine U2AF (U2 Associated Factor/U2 Auxilliary Factor) reconnaîtra une séquence courte polypyrimidique suivi du dinucléotide  $5'AG3'$  situé à l'extrémité 3' de l'intron. L'association des protéines, des ARNs U1 et U2 à l'ARN pré-messager consomme de l'ATP pour former le complexe E précoce (Michaud and Reed, 1991).

Le complexe ribonucléoprotéique constitué de U4, U5 et U6 va ensuite se fixer à la place de U1, qui sera alors décroché. A ce stade, le complexe en place est nommé E tardif. De nombreux réarrangements internes s'opèrent au sein de la structure (Moore and Sharp, 1993). U4 est alors libéré du complexe. L'extrémité 5' de l'intron est détachée de l'exon en amont

puis liée à une adénine située au point de branchement pour former un lasso qui sera excisé de son extrémité 3'. Les deux exons sont joints par U5 avant que le complexe Intron-U2-U6-U5 ne soit libéré. Les ribonucléoprotéines sont alors recyclées dans une nouvelle réaction d'épissage tandis que le lasso intronique peut-être dégradé. Cependant, certains ARNs non-codants sont contenus dans les introns et sont relâchés après épissage non dégradés (Mattick and Makunin, 2006).

Une fois tous les introns éliminés, l'ARNm mature deviendra fonctionnel, la jonction des exons les uns aux autres sans interruption permettant de révéler la séquence codante finale, qui sera lue par le ribosome.

Au cours de l'élimination des introns, une fois deux exons joints, un complexe protéique, nommé l'EJC (Exon-Junction Complex) est placé sur l'ARNm naissant à 20-24 nucléotides en amont de la jonction (Le Hir et al., 2000). Doté d'un cœur hétérotetramérique, la composition de l'EJC varie de manière dynamique en incluant des composantes transitoires (Tange et al., 2005) : certains recrutés avant même que l'épissage ne commence (Reichert et al., 2002), d'autres se dissociant après l'export du noyau (Le Hir et al., 2001)(Kim et al., 2001). Parmi les composantes transitoires, deux protéines – UPF2 et UPF3 (Up-FrameShift 2 et 3) - sont recrutées sur les EJCs par le biais de la coiffe. En effet, c'est la présence du complexe CBC (Cap Binding Complexe) qui permet de joindre ces deux protéines aux EJCs. CBC est spécifique de la coiffe des ARNm « jeunes », et est remplacé après le premier tour de traduction par un facteur initiateur de traduction (Ishigaki et al., 2001)(Fortes et al., 2000) -

Agissant comme marqueur d'épissage, les EJCs déposés sur un ARNm sont détachés au cours du premier tour de traduction par le ribosome (Dostie and Dreyfuss, 2002). Plutôt que d'être détachés physiquement par le passage du ribosome, il semblerait que les EJCs soient détachés par la phosphorylation de l'un des composés du cœur de l'EJC (Hsu et al., 2005).

### **c) La polyadénylation**

La troisième étape de maturation consiste en l'ajout de la queue polyadénylée à l'extrémité 3' de l'ARN messenger. Cette étape est subdivisée en deux parties : le clivage endonucléolytique de la région 3'UTR suivi de la synthèse de la queue poly(A). Le complexe de polyadénylation est composé d'une douzaine de polypeptides et interagît avec le CTD de l'ARN polymérase II (Shatkin and Manley, 2000).

Une fois l'ARN messenger transcrit et mûri, il est exporté du noyau *via* les pores nucléaires pour rejoindre le cytoplasme où il pourra être stocké dans des granules (Parker and Sheth, 2007) ou traduit par les ribosomes.

### 3. Traduction

Les ribosomes sont des complexes macro-moléculaires ribonucléoprotéiques constitués de deux sous-unités et en charge de la synthèse protéique.

La traduction se divise en trois phases :

1. L'initiation, durant laquelle les sous-unités ribosomales sont recrutées sur l'ARNm par le biais d'un facteur d'initiation de traduction. Le ribosome va alors scanner la séquence de nucléotides jusqu'au codon initiateur AUG où le premier ARN de transfert se fixe.
2. La seconde étape est l'élongation où le ribosome scanne la séquence codante et associe à chaque codon rencontré dans le cadre de lecture, un anticodon porté par un ARN de transfert (ARNt), lui-même physiquement relié à un acide aminé particulier. Le ribosome va effectuer la liaison entre les acides aminés au fur et à mesure que ceux-ci se présentent.
3. La terminaison a lieu lorsque le ribosome s'arrête à la rencontre d'un codon stop (Ocre, Ambre, Opale).

Une fois la traduction terminée, un complexe protéique nommé SURF (SMG-1, UPF-1, eRF) est déposé par le ribosome au niveau du premier codon stop rencontré (Kashima et al., 2006). Ce complexe serait placé spécifiquement lors du premier tour de traduction, puisque seuls les ARNm porteurs du complexe CBC (Cap-Binding Complex) spécifique de la coiffe des ARNm « jeunes » (Isken and Maquat, 2008) peuvent recruter la protéine UPF-1. Le complexe SURF peut interagir rapidement avec un autre complexe protéique, si celui-ci est présent : l'EJC (Kashima et al., 2006).

Une fois la traduction terminée, les deux sous-unités ribosomales se dissocient de l'ARNm dans une quatrième étape de recyclage.

### 4. Dégradation

#### **a) Voies de dégradations « classiques »**

La coiffe à l'extrémité 5' et la queue poly(A) des ARNm les protègent de la dégradation. Chez les eucaryotes, il existe deux voies de dégradation des ARNm :

1. La voie 5' vers 3', connue sous le nom de la voie XNR1, qui nécessite la dégradation de la coiffe par l'enzyme DCP2 avant la dégradation de l'ARNm par l'exonucléase 5'-3' XNR.
2. La voie 3' vers 5', mettant en scène la déadenylation de l'extrémité 3' de l'ARNm par une déadénylase. Une fois l'extrémité poly(A) éliminée, l'ARNm peut-être décoiffé par une enzyme de décoiffage alternative (DCPS) et être soumis à la voie de dégradation 5'-3' ou être dégradé de 3' en 5' par un complexe protéique nommé l'exosome.

(Belasco, 2010)

### **b) Le NMD – Nonsense-Mediated Decay**

Le NMD – Nonsense-Mediated Decay – est un mécanisme qui dégrade sélectivement des ARNm possédant un codon stop « prématuré ». En cas de traduction de ces ARNm, les protéines synthétisées tronquées pourraient agir comme dominant-négatif ou présenter des activités gain-de-fonction délétères pour l'organisme. Le système agit en provoquant la dégradation ou la séquestration des transcrits dans des structures subcellulaires, empêchant ainsi leur traduction.

Parmi les origines possibles des codons stop prématurés se trouvent les mutations «non-sens» mais également les réarrangements génétiques programmés comme ceux qui se produisent dans les gènes des immunoglobulines G dont deux tiers des transcrits générés par ce procédé sont dégradées par le NMD (Chang et al., 2007). Les autres cas sont liés au produit de l'épissage alternatif. Le couplage avec le NMD permet de réguler le niveau d'expression du gène en éliminant sélectivement les transcrits contenant des codons “Stop” prématurés.

### **Mécanisme – Voie classique**

La voie classique du NMD, longtemps considérée comme la seule branche de cette voie de dégradation (Chang et al., 2007), passe par deux signaux distincts qui interagissent pour amorcer la dégradation de l'ARNm. Ces signaux correspondent à deux empreintes protéiques : la première – le complexe SURF - apposée par le ribosome lors de son premier tour de traduction, la seconde – l'EJC pour *Exon-Junction Complex* - étant apposée après épissage dans le noyau et conservée dans le cytoplasme (Luo and Reed, 1999)(Le Hir et al., 2000).

Une fois le premier tour de traduction terminé, si une jonction exon-exon se trouve à plus de 50-55 nucléotides en aval du codon stop rencontré par le ribosome, ce dernier sera considéré comme prématuré, destinant l'ARNm qui le porte à être pris en charge par le NMD. Ce phénomène, connu comme la « règle des 50-55 nucléotides » (Nagy and Maquat, 1998), serait liée à la présence persistante d'EJCs en aval du codon stop car non-éliminés par le ribosome lors de son premier passage (Chang et al., 2007).

Le complexe SURF, placé à la terminaison de traduction, interagit alors avec l'EJC restant pour former un super-complexe SURF-EJC qui entraîne la phosphorylation de la protéine UPF1 (Yamashita et al., 2001), entraînant le recrutement de nouvelles protéines – SMG-5 et SMG-7 – (Anders et al., 2003) qui provoque la dégradation de l'ARNm porteur. SMG-7 est considéré comme l'effecteur terminal du NMD puisque sa fixation sur un ARNm provoque la déstabilisation rapide du messager (Unterholzner and Izaurralde, 2004).

La dégradation du messager reconnu par le NMD semble s'effectuer par les voies classiques de dégradation des ARNm chez les mammifères (5' vers 3' et 3' vers 5') (Lejeune et al., 2003) bien que des voies de dégradations uniques ont été découvertes chez certaines espèces comme *Saccharomyces cerevisiae* (Takahashi et al., 2003) et *Drosophila melanogaster* (Gatfield and Izaurralde, 2004). De plus, la dégradation de l'ARNm aurait vraisemblablement lieu pendant le premier tour de traduction puisque seuls les ARNm associés au complexe CBC sont dégradés (Ishigaki et al., 2001).

L'un des aspects les moins claires de la dégradation par le mécanisme du NMD consiste au lieu de dégradation des ARNm une fois reconnus comme « aberrant ».

L'une des possibilités de lieu de dégradation est le cytoplasme, et plus particulièrement des structures sub-cellulaires connues sous le nom de corps P – ou P-bodies. Ces structures correspondent à des corps cellulaires dynamiques composées d'ARN et de protéines, dont plusieurs acteurs terminaux du NMD – SMG-7 et UPF-1 – et de facteurs de dégradation des ARNm comme les enzymes responsables de l'élimination de la coiffe et des exonucléases 5' vers 3' (Sheth and Parker, 2003). Cependant, SMG-7 – l'effecteur terminal identifié du NMD – est capable de changer de localisation, incluant notamment le noyau (Unterholzner and Izaurralde, 2004), son activité n'étant donc pas nécessairement limitée à ces corps.

Il existe également des données indiquant que la dégradation survient dans la fraction nucléaire (Urlaub et al., 1989)(Carter et al., 1996) or, la traduction – mécanisme couplé au NMD - n'a été clairement démontrée comme ayant lieu que dans le cytoplasme. Des fractions nucléaires isolées présentent une activité traductionnelle (Iborra et al., 2001)(Wilkinson and

Shyu, 2002), ce qui pourrait être expliqué par une co-précipitation cytoplasmique au cours de l'isolation. Il est également possible de voir dans ces résultats un élément indiquant une possible traduction nucléaire, comme le suggère la visualisation par immunofluorescence de synthèse de chaînes d'acides aminées dans des noyaux (David et al., 2012). Le NMD, de la reconnaissance à la dégradation pourrait donc ne pas être aussi compartimenté que ce qui a été suggéré par le passé.

Un autre élément encore obscur est le devenir de l'ARNm lui-même. En effet, bien qu'il n'ait été question jusqu'à présent que de dégradation, la voie du NMD pourrait ne provoquer qu'une répression au niveau traductionnel (You et al., 2007). Cet état des choses a conduit à l'utilisation de nouveaux termes comme celui de NMTR – Nonsense-Mediated Translational Repression – bien que le mécanisme global et qu'une partie des acteurs impliqués semblent communs (Lee et al., 2010). De plus, dans certains cas, seule une partie des ARNm est dégradée tandis que le reste peut subir une répression traductionnelle ou suivre la voie de la traduction, comme les ARNm non-reconnus par le NMD (Barberan-Soler et al., 2009).

### Autre(s) voie(s) ?

Le NMD ne provoque pas nécessairement le même devenir chez tous les ARNm – dégradation ou répression traductionnelle -, il ne nécessite pas obligatoirement les mêmes facteurs – les protéines UPF-1, UPF-3a, UPF-3b ou même le complexe EJC ne sont pas nécessaires à la voie du NMD pour certains ARNm (Gehring et al., 2005)(Bhalla et al., 2009). La mise en évidence d'une séquence inhibitrice de NMD rapproche le NMD de la vision actuelle de mécanismes de régulation tels que la transcription ou l'épissage - existence de signaux inhibiteurs de NMD et leur opposition à des signaux promoteurs (Weil and Beemon, 2006). L'enrichissement des connaissances a remis en question la vision du NMD comme mécanisme unique aux acteurs identifiés (Chang et al., 2007).

La région 3'UTR est également impliquée dans la reconnaissance des ARNm considérés comme aberrants par la machinerie cellulaire. La preuve de cette implication provient de travaux sur des ARNm à la région 3'UTR étendue suite à une mutation dans le site de polyadénylation (Pulak and Anderson, 1993)(Amrani et al., 2004). Ce mécanisme inclut des composantes communes à la voie du NMD classique (Muhlrad and Parker, 1999).

Malgré les connaissances qui s'accumulent, la recherche d'un code propre au NMD est toujours en cours, et bien qu'il soit possible d'estimer la reconnaissance probable d'un ARNm

par le NMD en se basant sur certains travaux (Cf Règle des « 50-55 nucléotides »), il reste impossible d'écarter de la dégradation ou de la répression traductionnelle un ARNm produit.

## 5. Régulation de l'expression génique

L'expression des gènes codant pour des protéines peut être régulée à quatre niveaux distincts :

1. En amont de la transcription, par la combinaison de facteurs de transcription – pouvant inhiber ou favoriser le recrutement des sous-unités de l'ARN polymérase – se liant à des signaux précis, sous la forme de séquences nucléotidiques spécifiques ou de structures secondaires de l'ADN. Cette forme de régulation agit directement sur la quantité d'ARN pré-messager produite, et peut-être modulée de manière dynamique.
2. Après la transcription, ces formes variées de régulations sont regroupées sous le nom de régulations post-transcriptionnelles. Dans ce groupe se trouve un mécanisme nommé l'épissage alternatif qui fera l'objet du prochain chapitre. Ces régulations touchent aussi bien la durée de vie de l'ARNm, sa disponibilité pour la traduction voire sa séquence elle-même.
3. Durant la traduction, par le biais des ARNt - ARN de transfert - disponibles durant la synthèse protéique. Le code génétique étant redondant – 64 triplets de codons possibles pour 20 acides aminés conventionnels et trois codons stop (Ocre, Ambre et Opale), ainsi que deux acides aminés non-conventionnels et rares - il en découle que de nombreux acides aminés correspondent à plusieurs codons. Cependant, les ARNs de transfert portant des anti-codons différents et associés au même acide aminé ne se retrouvent pas dans des proportions équivalentes, altérant la vitesse de synthèse d'une chaîne protéique et pouvant même provoquer l'arrêt de la synthèse par le décrochage du ribosome actif.
4. Après la traduction, en affectant directement l'activité, la durée de vie ou la localisation des protéines produites. Les régulations post-traductionnelles affectent les protéines cibles par le biais de réactions biochimiques visant à accrocher ou décrocher un groupement chimique à des acides aminés cibles. Ces groupements (phosphate, méthyl,

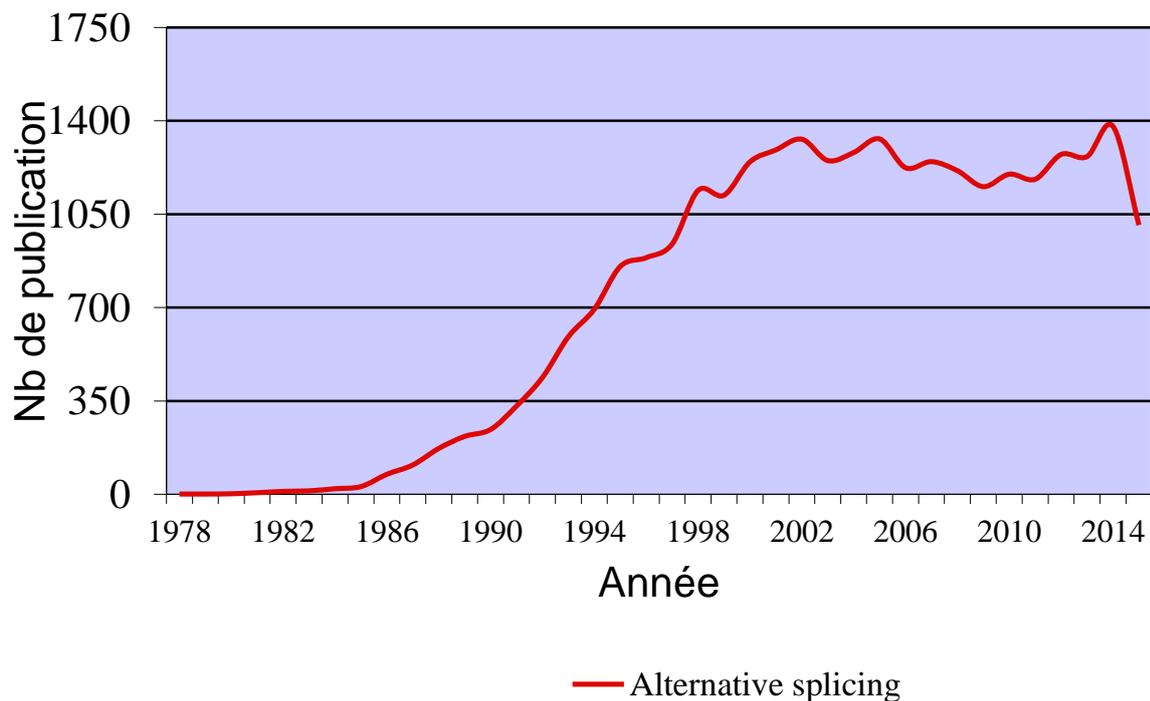
SUMO, ubiquitine, ...) peuvent permettre l'interaction avec d'autres protéines, ou un changement conformationnel de la chaîne d'acides aminés.

### III. L'épissage alternatif

#### 1. Généralités

En plus d'être une des étapes de maturation fondamentale de tout messager eucaryote, l'épissage est également un mécanisme de régulation post-transcriptionnel. Il permet, à partir d'un même ARN pré-messager, de générer plusieurs messagers différents, amenant à la synthèse de protéines distinctes en fonction du contexte cellulaire.

L'étude de l'épissage alternatif a pris une importance croissante dans les années 1990, passant d'une seule publication contenant le terme « Alternative splicing » en 1978 à plus de 1200 dans les années 2000. Depuis, le nombre est stable et compris entre 1200 et 1400 publications par an (**Figure 3**). Bien que la présence du terme « Alternative splicing » dans un article ne signifie pas qu'il s'agisse de l'intérêt principal des travaux publiés, le nombre de publication (>20 000) ne permet pas d'en faire une revue exhaustive. Le but de la partie ci-dessous consiste plus à présenter quelques exemples de modulation fonctionnelle par épissage alternatif ainsi que les liens identifiés avec d'autres mécanismes moléculaires.



**Figure 3 : Nombre d'articles scientifiques possédant le terme « Alternative splicing » au cours du temps.** La légère chute en fin de courbe correspond à l'année en cours (nombre encore non-définitif). Données obtenues à partir du site PubMed-NCBI.

L'épissage alternatif peut modifier les propriétés du produit final en altérant la séquence protéique ou en modulant le niveau d'expression du transcrit. Parmi ces événements, les modifications peuvent aller de la perte de fonction complète à des altérations dont les effets restent discrets.

## 2. Effets de l'épissage alternatif

### a) Sur les séquences protéiques

L'insertion ou l'exclusion spécifique d'exons ou d'introns peuvent éliminer des domaines abolissant les fonctions associées ou modifier les propriétés des protéines de nombreuses manières :

1) En affectant la fonction d'enzymes par la modification de leur spécificité de substrat (Christmas et al., 2001), de leurs propriétés catalytiques (Li and Koromilas, 2001) ou de la régulation de leur activité (Benedict *et al.*, 2001). Parmi ces exemples, on peut retrouver le cas d'un variant du récepteur d'hormones thyroïdiennes TSHR (Thyroid Stimulating Hormone Receptor) qui ne fixe pas le TSH, provoquant une insensibilité à l'hormone dans certaines tumeurs (Ando et al., 2001). Le récepteur à la dopamine D3 dont une isoforme est insensible à la dopamine (Nagai et al., 1993). Au-delà de la simple présence/absence de domaine et/ou de fonctions associées, les variants d'épissage peuvent engendrer des protéines aux activités similaires mais aux efficacités variables. C'est le cas de l'inclusion différentielle d'un exon dans l'ARNm du récepteur à l'insuline provoquant la division par deux de l'affinité de la protéine vis-à-vis de son ligand (Sesti et al., 2001). Ce mécanisme est tissu-spécifique et permet de contrôler la réponse à l'insuline en fonction des tissus. Une perturbation de l'abondance relatives de ces deux isoformes est caractéristique des diabètes de type II.

2) L'expression de différentes isoformes de facteurs de transcription peuvent provoquer en aval, la modification du niveau d'expression de leurs gènes cibles (Lillycrop and Latchman, 1992). Les protéines capables de se fixer à l'ADN (autres que les facteurs de transcriptions) et à l'ARN sont également concernées, bien que dans ces cas, les variants d'épissage affectent la liaison à d'autres protéines, ce qui provoque en retour une modification de l'affinité des complexes pour les acides nucléiques (Ormondroyd *et al.*, 1995).

3) La localisation cellulaire peut être modifiée *via* l'insertion de domaine de liaison aux membranes (Meshorer et al., 2004) ou par le biais d'insertion de NLS – pour *Nuclear Localization Sequence* - dans la séquence protéique finale (Kanezaki et al., 2001). Ce changement de localisation peut réguler ou compartimenter la protéine allant parfois jusqu'à

abolir sa fonction dans le cas de des Immunoglobulines  $\alpha$  et  $\beta$  chez les cellules humaines (Indraccolo et al., 2002).

4) La stabilité des protéines peut être affectée par le biais de la production d'isoformes aux durées de vie variables (Nestler et al., 1999) ou par l'apparition d'un site de clivage protéolytique (Sakurai *et al.*, 2001).

5) La stabilité, l'activité et la localisation cellulaire des isoformes protéiques peuvent également être touchées indirectement par l'ajout/suppression de sites de modifications post-traductionnelles telles que la phosphorylation (Birukov et al., 2001) ou la glycosylation (Tanahashi and Tabira, 2001).

### **b) Sur la stabilité des ARNm**

Au-delà de l'impact de l'épissage alternatif sur les protéines, les ARNm eux-mêmes sont affectés de plusieurs manières :

1) La stabilité, l'export nucléaire, la localisation cytoplasmique et l'efficacité de traduction des transcrits peuvent être affectées par l'épissage sélectif de séquences à l'intérieur des régions 5'UTR et 3'UTR. Deux exons alternatifs dans la séquence 5'UTR du VIH-1 provoquent un changement dans la stabilité de l'ARNm (Krummheuer et al., 2001).

2) Il existe également un dernier cas de figure pouvant donner lieu à la synthèse de protéines tronquées mais modifiant également le niveau d'expression du transcrit : celui de l'introduction de codon stop dit « prématuré » qui peut activer la réponse NMD décrite plus haut.

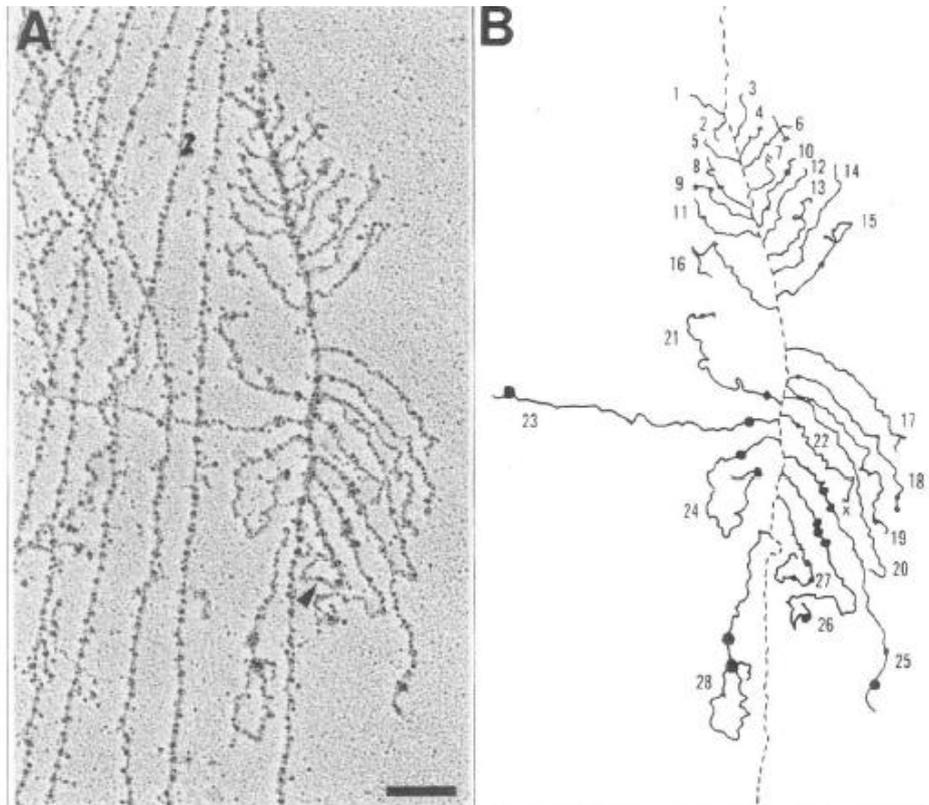
Affectant les propriétés des protéines ou des ARNm, l'épissage alternatif est un mécanisme complexe aux effets variés. Le mécanisme par lequel ces possibilités sont générées repose majoritairement sur la coopération de deux types d'éléments.

## **3. Un mécanisme co-transcriptionnel**

### **a) Couplage cinétique et CTD :**

Le premier exemple de corrélation entre transcription et épissage provient de micrographies électroniques permettant de visualiser les transcrits naissants contenant des boucles correspondant à des événements d'épissage (**Figure 4**)(Beyer and Osheim, 1988).

Depuis, la combinaison de plusieurs méthodes, dont le séquençage d'ARN naissant, a permis de confirmer ce couplage (Khodor et al., 2011)(Pandya-Jones and Black, 2009).



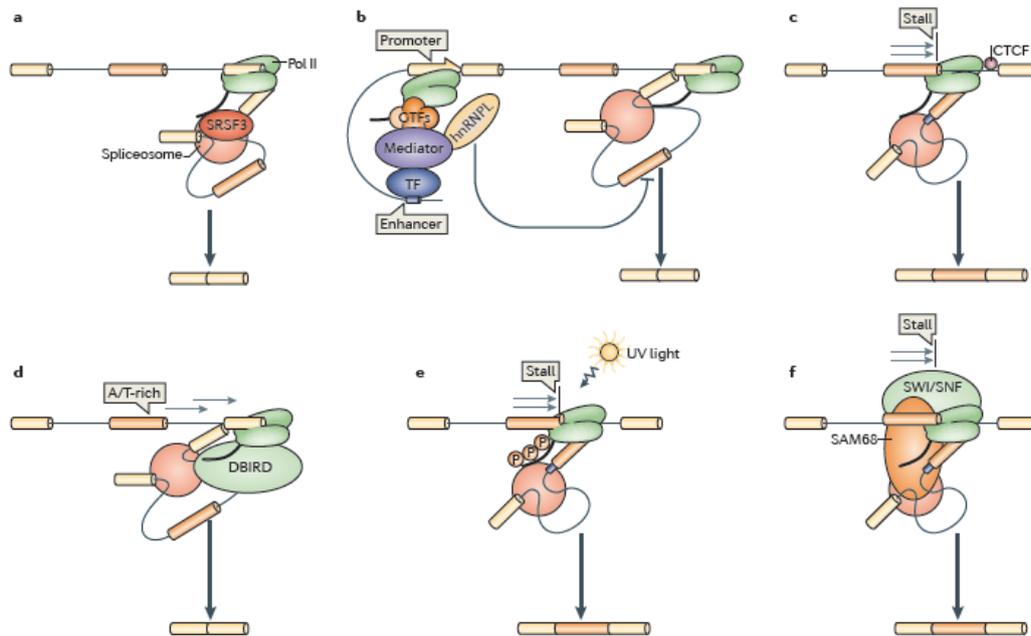
**Figure 4 : Formation de boucles le long de pré-messagers naissants ;** (A) Micrographie électronique d'unités transcriptionnelles issues d'embryons de *Drosophila* – La pointe de flèche indique un site proposé où a lieu une excision d'intron – La barre d'échelle représente 0.2  $\mu\text{m}$ ; (B) Représentation graphique de la micrographie ; (---) Matrice, (-----) fibrilles de RNP – transcrits et protéines -, (.) particules de RNP, (x) transcrits avortés présumés. Extrait de Beyer and Osheim, 1988.

Plusieurs observations expérimentales indiquent que la régulation de l'épissage alternatif peut également être dépendante de la transcription – et de toutes les composantes de ce mécanisme :

1) Des changements dans la structure des promoteurs peuvent affecter le choix des sites d'épissage, altérant le transcrit final. Ce cas a été démontré sur un exon issu du gène de la fibronectine, dont l'inclusion varie en fonction du promoteur placé en amont (Cramer et al., 1997). De plus, des composantes de la machinerie transcriptionnelle elle-même pourraient se lier physiquement avec les protéines de la famille SR impliquées dans le cas précédent (Cramer et al., 1999). Un autre article sur ce phénomène a confirmé ces résultats sur un seul

des deux gènes étudiés, démontrant que le mécanisme n'est pas applicable à tous les cas d'épissage alternatif (Pagani et al., 2003). Une étude portant sur les promoteurs dépendants de co-régulateurs – récepteurs nucléaires aux stéroïdes par exemple - démontrent que ces facteurs de transcriptions pourraient contrôler la transcription et l'épissage simultanément (Auboeuf et al., 2002).

2) Une analyse protéomique portant sur les co-purifiants de la polymérase II a mis en évidence la présence de protéines de la famille SR ainsi que des composantes de la ribonucléoparticule U1 mais pas des autres composantes du spliceosome (Das et al., 2007). SR20p, qui favorise l'exclusion de l'exon 5 du gène de la fibronectine, est recrutée spécifiquement par le CTD de la Pol II, tandis que l'inclusion du même exon favorisée par la protéine SF2/ASF est indépendante de la présence du CTD (de la Mata and Kornblihtt, 2006). L'utilisation d'une Pol II ralentie par mutation a permis de déterminer que la fonction de SR20p est indépendante de la vitesse de transcription, tandis que l'action de SF2/ASF en est dépendante (Mata et al., 2003), passant par le biais de deux séquences de 72 paires de bases qui affectent l'élongation (Kadener et al., 2002). Il existe donc plusieurs voies possibles par lesquelles la transcription affecte l'épissage (**Figure 5**) dont une voie de recrutement de régulateurs par les composantes de la polymérase, dont le CTD – lequel correspond à une plateforme pour les complexes impliqués dans la maturation des ARNm – ainsi qu'une voie cinétique, impliquant la vitesse d'élongation.



**Figure 5 : Différents mécanismes couplant transcription avec épissage alternatif.** (a) Le domaine carboxy-terminal (CTD) de l'ARN polymérase II est nécessaire au recrutement du facteur SRSF3, ce qui inhibe l'inclusion d'exons alternatifs (de la Mata and Kornblihtt, 2006). (b) Le complexe de transcription basal, connu sous le nom de complexe Mediator, s'associe à des facteurs de transcriptions au promoteur et peut recruter le facteur hnRNPL. Ce régulateur inhibe l'inclusion d'exons alternatifs durant l'élongation (Huang et al., 2012). (c) La protéine CTCF se fixe à des séquences d'ADN riches en GC et non-méthylées situées en aval d'exons alternatifs. Cette fixation gêne l'élongation, ce qui favorise l'inclusion d'exon (Shukla et al., 2011). (d) Le facteur de transcription DBIRD permet une accélération de l'élongation dans les séquences riches en A et T, ce qui inhibe l'inclusion d'exon (Close et al., 2012). (e) Les dommages à l'ADN provoqués par rayonnement UV provoquent l'hyper-phosphorylation du CTD de l'ARN polymérase II. Cette modification bloque la transcription et favorise l'inclusion d'exons alternatifs (Muñoz et al., 2009). (f) Le facteur de remodelage de la chromatine SWI/SNF recrute le régulateur d'épissage SAM68 durant la transcription, ce qui bloque l'ARN polymérase II et favorise l'inclusion d'exon (Batsché et al., 2006). Extrait de (Kornblihtt et al., 2013).

Par le biais de conditions et de mutations altérant l'élongation dans des lignées cellulaires en culture, une étude portant sur un peu moins de 6000 événements indépendants a démontré un enrichissement significatif des gènes présentant une plus forte inclusion d'exon associés aux GO term « RNA Processing » et « RNA Binding » (Ip et al., 2011). Cette étude présente un couplage Transcription-Epissage alternatif-NMD, en plus de démontrer sa conservation des cellules humaines à l'organisme modèle *Caenorhabditis elegans*.

L'élongation par l'ARN polymérase II est également affectée par la structure de la chromatine et les modifications post-traductionnelles des protéines qui la composent, les histones.

### **b) Un lien avec la structure de la chromatine :**

Un premier lien entre l'architecture d'un gène et la structure de la chromatine avait déjà été supposé dans les années 80-90 en mettant en parallèle la périodicité des nucléosomes et celle des sites d'épissage (Beckmann and Trifonov, 1991). A la fin des années 2000, deux études concomitantes relatives à la position des nucléosomes – structures constituées d'ADN génomique et d'histones associées – sont publiées en analysant des données issues de l'homme, de la souris, de la drosophile et du nématode. La première démontre un enrichissement significatif des nucléosomes au niveau des exons en comparaison des introns (Schwartz et al., 2009) ; la seconde, un enrichissement supplémentaire entre exons alternatifs et constitutifs, ainsi que l'absence de nucléosomes au niveau de pseudo-exons. Un pseudo-exon est un exon potentiel, délimité par des sites d'épissages, mais qui n'est pas reconnu par la machinerie d'épissage et donc jamais inclus dans un ARNm mature (Tilgner et al., 2009).

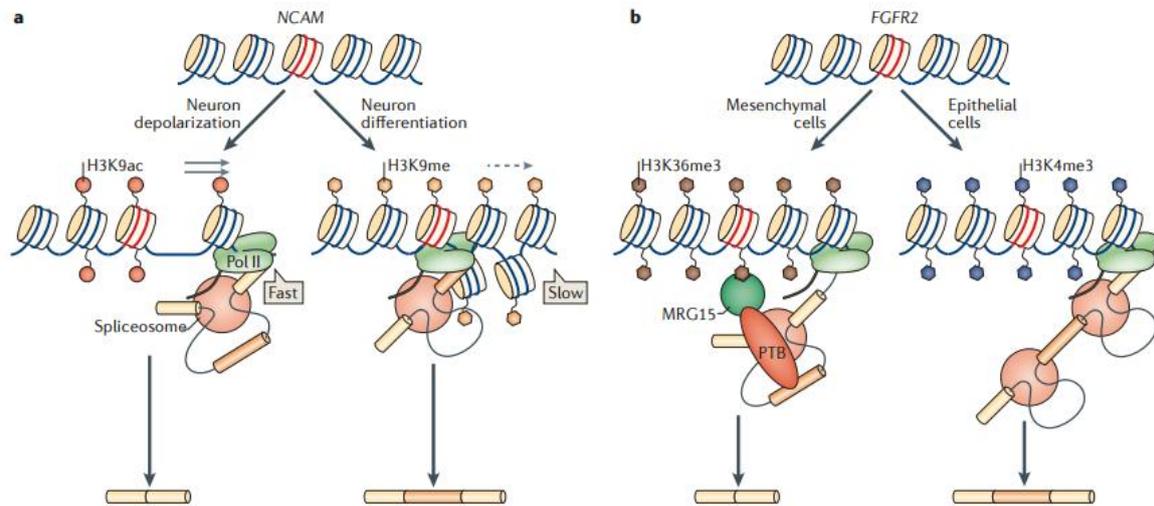
De plus, des travaux ont mis en évidence que la vitesse d'élongation de l'ARN Polymérase II n'est pas constante, qu'elle s'arrête lorsqu'elle rencontre un nucléosome et ne reprend que lorsque celui-ci s'est détaché - et potentiellement rattaché à une boucle d'ADN générée par le passage de la polymérase (Hodges et al., 2009). Ces résultats, mis en relation avec ceux sur la vitesse de transcription, permettent de déterminer que l'organisation de la chromatine peut affecter l'épissage alternatif par le biais de l'élongation du transcrit.

Tout comme pour la transcription, les modifications post-traductionnelles des histones peuvent affecter l'épissage de deux manières distinctes :

1) En affectant l'état de de la chromatine, ce qui pourrait affecter la vitesse d'élongation de la polymérase.

La dépolarisation membranaire des neurones provoque une hyperacétylation de la lysine 9 de l'histone H3 dans une région du gène NCAM, provoque un relâchement de la chromatine. Cela facilite l'élongation par l'ARN polymérase et induit l'exclusion sélective d'un exon alternatif proche (Schor et al., 2009). Durant la différenciation neuronale, le phénomène inverse se produit : les marques H3K9me2 – me2 : diméthylation ; K9 : Lysine

n°9 ; H3 : Histone 3 - et H3K27me3 – triméthylation de la lysine 27 de l’histone 3 - sont enrichies le long du gène NCAM, ralentissant l’élargissement et favorisant une l’inclusion de l’exon 18 (Schor et al., 2013). (**Figure 6**)



**Figure 6 : Comment la chromatine influence l’épissage alternatif.** (a) La méthylation ou l’acétylation de la lysine 9 de l’histone 3 affecte le couplage cinétique entre transcription et épissage alternatif. Exemple du gène NCAM dont l’épissage change en fonction de la polarisation membranaire ou de la différenciation neuronale. (b) Des modifications post-traductionnelles d’histones affectant l’épissage en recrutant des régulateurs spécifiques par le biais d’adaptateurs. Extrait de (Kornbliht et al., 2013).

2) En permettant le recrutement spécifique de régulateurs de l’épissage :

La triméthylation de la lysine 36 de l’histone 3 - H3K36me3 - est une modification post-traductionnelle fortement associée aux exons constitutifs et moins abondante aux alternatifs (Kolasinska-Zwierz et al., 2009). Elle permet le recrutement du régulateur PTB par le biais d’une autre protéine. La protéine PTB va alors se fixer sur une séquence contenue sur le pré-messager et provoquer l’exclusion de l’exon alternatif (Luco et al., 2010)(Figure 5). Cette marque permet également de recruter SRSF1 un autre régulateur de la famille des SR, par le biais d’une protéine associée aux histones (Pradeepa et al., 2012).

La marque H3K9me3 est retrouvée associée spécifiquement à plusieurs exons alternatifs, et est liée par les protéines HP1. Une de ces protéines, HP1 $\gamma$ , connue comme

répresseur transcriptionnel, est capable de freiner l'élongation de l'ARN polymérase II, favorisant ainsi l'inclusion d'un exon alternatif dans le gène *CD44* (Saint-André et al., 2011). De plus, cette protéine peut se lier physiquement à un régulateur très étudié de l'épissage, ASF/SF2 (Loomis et al., 2009). H3K9me3 semble donc affecter l'épissage à la fois par son effet sur la cinétique de transcription et par le recrutement de facteur d'épissage.

Un point intéressant de ces marques est leurs origines. Dans le cas de la modification H3K9ac liée à l'exon 18 du gène *NCAM*, l'acétylation est la résultante de la dépolarisation, éliminée après retour aux conditions standards de la cellule, tandis que pour H3K36me3, elle est apposée près des exons par la machinerie d'épissage (Kim et al., 2011), ce qui est cohérent avec une présence moindre au sein des exons alternatifs. Bien que le cas de la marque H3K36me3 soit encore un cas particulier, il se peut que les modifications des histones affectent l'épissage et que l'épissage affecte également les modifications. Les deux réponses ne sont pas nécessairement mutuellement exclusives et pourraient bien s'autoalimenter. De plus, il est probable que plusieurs mécanismes participent aux interactions entre remodelage de la chromatine et épissage.

#### 4. Les acteurs de l'épissage alternatif – « Splicing code »

L'épissage alternatif repose majoritairement sur la participation de facteurs en *trans*, le plus souvent des protéines riches en Sérine-Arginine (SR) (Fu, 1995) ou de la famille des hnRNPs - heterogeneous RiboNucleoParticle - (Krecic and Swanson, 1999). Ces facteurs interagissent physiquement avec des séquences nucléotidiques - éléments en *cis* – pour altérer l'épissage de l'ARN pré-messager naissant.

##### **a) Séquence en *cis***

L'épissage d'un ARN pré-messager nécessite trois types de séquences différentes : le point de branchement, le site d'épissage en 5' et le site d'épissage en 3'. Ces sites permettent à la machinerie moléculaire de l'épissage de déterminer la position où les ribonucléoprotéines vont se fixer, définissant ainsi les exons et les introns du pré-messager.

La reconnaissance par le spliceosome de ces séquences semble plus complexe qu'il n'y paraît. En effet, l'élimination des introns semblerait indiquer de prime abord, que le site en 5' est reconnu dans un premier temps, puis associé à un site en 3' – le spliceosome reconnaissant

les introns préférentiellement, les exons n'étant qu'une conséquence de ce procédé. Cependant, de nombreux résultats d'analyse *in vitro* et *in vivo* (Talerico and Berget, 1990)(Nakai and Sakamoto, 1994)(Kuo et al., 1991), indiqueraient une reconnaissance inverse – à savoir la reconnaissance par paire de séquences 3' et 5' délimitant un exon, ce qui impliquerait que le spliceosome délimiterait les séquences exoniques préférentiellement. Ce postulat est connu sous le nom de l' « Exon Definition » (Robberson et al., 1990).

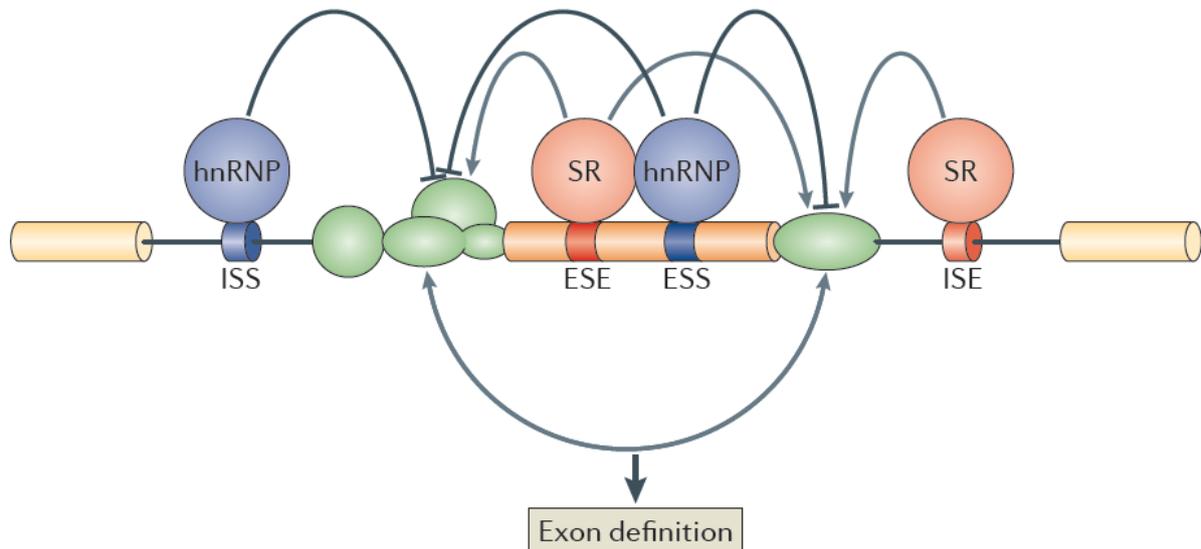
L' « Exon Definition » sous-tend – contrairement à l' « Intron Definition » - un mécanisme particulier pour le premier et le dernier exon de chaque gène. En effet, l'extrémité 5' théorique du premier exon n'est autre que la coiffe de l'ARNm et l'extrémité 3' théorique du dernier exon est située au niveau du site de clivage et de polyadénylation. Il a été montré l'importance de la coiffe dans l'épissage du premier intron en altérant le nucléotide G de la future coiffe d'un ARNm, empêchant la mise en place de la coiffe. Dans l'ARNm produit, le premier intron était moins efficacement excisé du messenger bien que le second intron ne semblait pas affecté (Ohno et al., 1987). Quelques années plus tard, des expériences ont permis de déterminer que la présence d'un intron en amont du site de polyadénylation augmentait l'efficacité de cette dernière (Niwa et al., 1990). De plus, des mutations du signal de polyadénylation inhibent l'épissage de l'intron proximal sans affecter les introns en amont (Niwa and Berget, 1991). Ces données indiquent un fort degré de coordination entre le mécanisme de l'épissage et, non plus simplement à la transcription, mais également à la présence de la coiffe (Izaurrealde et al., 1994) et à la polyadénylation.

Des données suggéreraient que la taille d'un exon serait un facteur important dans le choix des sites d'épissages (Dominski and Kole, 1991). L'apparente préférence pour une certaine longueur d'exons e (la taille moyenne d'un exon chez les vertébrés est de 137nt avec moins de 5% d'exons de plus de 300nt) semble indiquer également la nécessité pour un exon de disposer d'une taille minimale – permettant aux composantes du spliceosome de se fixer aux extrémités sans générer d'encombrement stérique – ainsi que d'une taille maximale – qui favoriserait la reconnaissance par paire – (Berget, 1995), bien que des cas d'exons de quelques nucléotides de long aient été identifiés comme l'exon N1 de 18 nucléotides du gène *c-src* chez la souris (Black, 1991). Cependant, le cas de certains mini-exons présente des particularités uniques, puisque l'épissage est dépendant d'exons en amont comme le troisième exon de 7 nucléotides du gène de la troponine I (Sternier and Berget, 1993).

La taille de l'intron en aval d'un exon affecte également l'insertion du dit exon dans le messenger final *in vitro* (Dominski and Kole, 1991), favorisant encore une fois la vue de l' « Exon Definition ».

Le cas des cassettes exons – et par extension, de toute forme d'épissage alternatif – repose sur la présence simultanée de plusieurs sites d'épissages proches. Le choix des sites utilisés par le spliceosome sera alors modulé par la présence d'autres éléments en *cis*, provoquant l'apparition de structures secondaires (Buratti and Baralle, 2004) ou la fixation de facteurs en *trans* (Ashiya and Grabowski, 1997).

Les séquences nucléotidiques présentes sur l'ARN pré-messager affectant le devenir des exons et des introns ont été baptisées ESE – Exonic Splicing Enhancer –, ESS – Exonic Splicing Silencer–, ISE – Intronic Splicing Enhancer – et ISS – Intronic Splicing Silencer – en fonction de leurs positions et de leurs effets sur l'épissage (Pagani and Baralle, 2004). Ces séquences permettent la fixation de facteur en *trans* capables, une fois fixés sur l'ARN naissant, d'affecter le choix de sites d'épissages en interagissant avec les différentes composantes du spliceosome (**Figure 7**).



**Figure 7 : Séquences en *cis* et facteurs en *trans* impliqués dans l'épissage alternatif.** Les séquences en *cis* impliquées dans l'épissage alternatif sont les séquences ESE – Exonic Splicing Enhancer –, ESS – Exonic Splicing Silencer –, ISE – Intronic Splicing Enhancer – et ISS – Intronic Splicing Silencer. Ces séquences sont liées par des facteurs en *trans* comme les protéines hnRNP (en BLEU) et SR (en ROUGE). Les éléments en VERT correspondent aux composantes du spliceosome s'associent aux sites d'épissage en 5' et 3' qui délimitent un exon. Le spliceosome est ici représenté selon le modèle de l' « Exon Definition ». Extrait de (Kornblihtt et al., 2013).

La fixation d'un régulateur dépend de son affinité vis-à-vis de la séquence. Des variations dans la séquence consensus n'empêchent pas la régulation mais modifient l'affinité de la protéine, ce qui affecte l'épissage.

Des mutations d'une séquence en *cis* affectent *in vitro* l'affinité de la protéine. En utilisant plusieurs versions d'un rapporteur fluorescent contenant différentes séquences mutantes, l'équipe du Dr Dupuy a démontré que les ratios d'isoformes visualisés *in vivo* correspondent aux affinités du régulateur mesurées *in vitro* (Amrane et al., 2014).

## **b) Facteurs en trans**

Les deux types de protéines les plus connues pour affecter l'épissage d'un ARN sont la famille des protéines riches en Sérine et Arginine – connu sous le nom de SR – et des protéines de la famille des hnRNPs – heterogeneous nuclear RiboNuclear protein Particles, ou heterogeneous nuclear RiboNucleoProtein. Il existe également des régulateurs de l'épissage alternatif qui ne font partie d'aucune de ces deux familles (Black, 2003).

Les protéines de la famille des SR fixent majoritairement les séquences ESE – *Exonic Splicing Enhancer* – tandis que celles de la famille des hnRNP sont recrutées en grande partie aux ESS – *Exonic Splicing Silencer* (Black, 2003). Bien que ces conditions ne soient pas toujours respectées, elles forment la base du « splicing code » qui permettra de déterminer, selon la nature des éléments en *cis* et des facteurs en *trans* mis en présence, la séquence du messenger final épissé.

## **La famille des protéines SR**

La famille des protéines SR (Ser-Arg) est composée de facteurs nucléaires jouant de multiples rôles dans l'épissage des pré-messagers produits par les métazoaires. Divers études ont permis de démontrer l'importance de ces facteurs dans l'épissage conventionnel, ou constitutif. Entre autres exemples, des membres de la famille des Sérine-Arginine feraient le lien entre le facteur U2AF et l'ARN U1 (Caceres & Krainer, 1997). De plus, ces protéines ont été mises en évidence dans des cas d'épissage alternatif, notamment dans le choix des sites d'épissage sélectionnés en fonction de la concentration de ces facteurs (Ge and Manley, 1990)(Krainer et al., 1990), de la répression/activation des composantes du spliceosome *via* des interactions physiques avec des séquences nucléotidiques contenues dans le pré-ARNm naissant et spécifique des membres de la famille en présence (Tacke and Manley, 1999).

Les protéines de cette famille comportent traditionnellement un à deux domaines de fixation aux ARNs – *RNA-Binding Domain* ou *RBD* – et un domaine C-terminal constitué de

répétitions d'Arginine-Sérine impliquées dans les interactions protéine-protéine (Wu and Maniatis, 1993)(Kohtz et al., 1994). Bien que de constitutions similaires, les différents membres de cette famille varient dans le choix de leurs partenaires ainsi que dans leurs affinités vis-à-vis des dits partenaires (Tacke and Manley, 1995)(Cavaloc et al., 1999). Lorsqu'un membre de cette famille dispose de deux domaines RBD, ceux-ci semblent tous deux nécessaires à la spécificité de la séquence fixée (Tacke and Manley, 1995)(Shi et al., 1997). Les séquences fixées par ces facteurs varient entre 6 à 10 nucléotides de long (Tacke and Manley, 1999), voire plus dans des cas où une structure secondaire est impliquée (Shi et al., 1997).

L'état de phosphorylation des protéines SR a également un impact sur leurs fonctions. L'utilisation d'inhibiteurs spécifiques de deux phosphatases des protéines SR a permis de démontrer la nécessité de déphosphorylation de ces facteurs afin de permettre l'activité catalytique du spliceosome dans des cellules HeLa (Mermoud et al., 1992). L'ajout de phosphatases sur des extraits nucléaires de cellules HeLa ne perturbe pas l'activité d'épissage mais empêche l'assemblage des composantes du spliceosome sur des ARNs (Mermoud et al., 1994). Il semblerait donc que l'épissage puisse être contrôlé par l'état de phosphorylation de protéines SR associées au spliceosome.

De plus, quelques exemples mis en évidence dans différentes espèces ont démontré l'impact de l'état de phosphorylation de plusieurs régulateurs de l'épissage alternatif. Le gène L1 contenu dans un adénovirus permet la production de la protéine E4-ORF4 activant la déphosphorylation de protéines SR par la phosphatase 2A dans des cellules HeLa infectées. Cette activité provoque un changement d'isoforme, favorisant la production de l'ARNm IIIa viral impliqué dans le dernier stade d'infection (Kanopka et al., 1998). Chez la drosophile, la protéine kinase LAMMER phosphoryle les régulateurs TRA, TRA2 et RBP1 impliqués dans l'épissage alternatif du gène *doublesex*. Un complexe de ces trois protéines phosphorylés permet un épissage particulier de l'ARNm de *doublesex*, l'effecteur terminal de la différenciation sexuelle des femelles (Du et al., 1998).

### La famille des hnRNP

La famille des hnRNP est composée de protéines impliquées dans de nombreux mécanismes moléculaires comme l'export nucléaire, la biogenèse des télomères, la transcription, la répression traductionnelle ou encore la polyadénylation. Bien que les

fonctions soient variées, l'une des caractéristiques les plus fréquentes chez les membres de cette famille est l'implication dans l'épissage (Krecic and Swanson, 1999).

Parmi ces protéines, une minorité a été particulièrement étudiée dans des lignées cellulaires:

1) L'un des membres les plus étudiés est la protéine hnRNP A1, qui favorise l'exclusion d'exons chez plusieurs ARNm dont son propre messenger (Del Gatto-Konczak et al., 1999)(Chabot et al., 1997). Cette protéine s'oppose à l'effet de la protéine SF2/ASF – un membre de la famille SR. Le résultat de cet antagonisme est dépendant du *ratio* de la concentration des deux protéines (Mayeda and Krainer, 1992).

2) L'insertion de l'exon A de l'ARNm du gène *c-src* est régulée par une séquence ISE sur laquelle se fixe un complexe de trois protéines, hnRNP F, hnRNP H et KSRP, et par une séquence ISS qui recrute la protéine PTB/hnRNP I (Modafferi and Black, 1999).

3) Le gène encodant la “polypyrimidine tract-binding protein” - ou PTB - régule, chez les mammifères, l'épissage de son paralogue nommé nPTB – *neural* PTB -, provoquant sa répression par le biais du NMD. La diminution de l'expression du premier permet l'expression du second ; les deux étant ainsi exclusifs et proches dans leurs fonctions de régulateurs d'épissage alternatif (Boutz et al., 2007).

Les protéines de cette famille possèdent également un à plusieurs domaines RBD et des domaines d'interactions protéine-protéine, bien qu'il y ait plus de disparités dans leurs compositions que chez les membres de la famille SR (Krecic and Swanson, 1999).

## 5. Les différents types d'épissage alternatif

Il existe, à l'heure actuelle, six grands types d'événement d'épissage alternatif (Zahler, 2012). Cette classification est basée sur la séquence finale de l'ARNm, plus que sur le mécanisme. L'un de ces six cas ne correspond pas à des isoformes dépendant pas de la machinerie d'épissage au niveau moléculaire et n'est donc pas réellement un cas d'épissage alternatif bien que le terme soit souvent utilisé abusivement pour décrire ce type d'isoformes.

1) L'exon cassette – CE pour *Cassette Exon* - correspond à une séquence identifiée comme exonique – délimitée par des sites d'épissage en 3' et en 5' - mais absente d'une ou plusieurs isoforme(s). (**Figure 8a**)

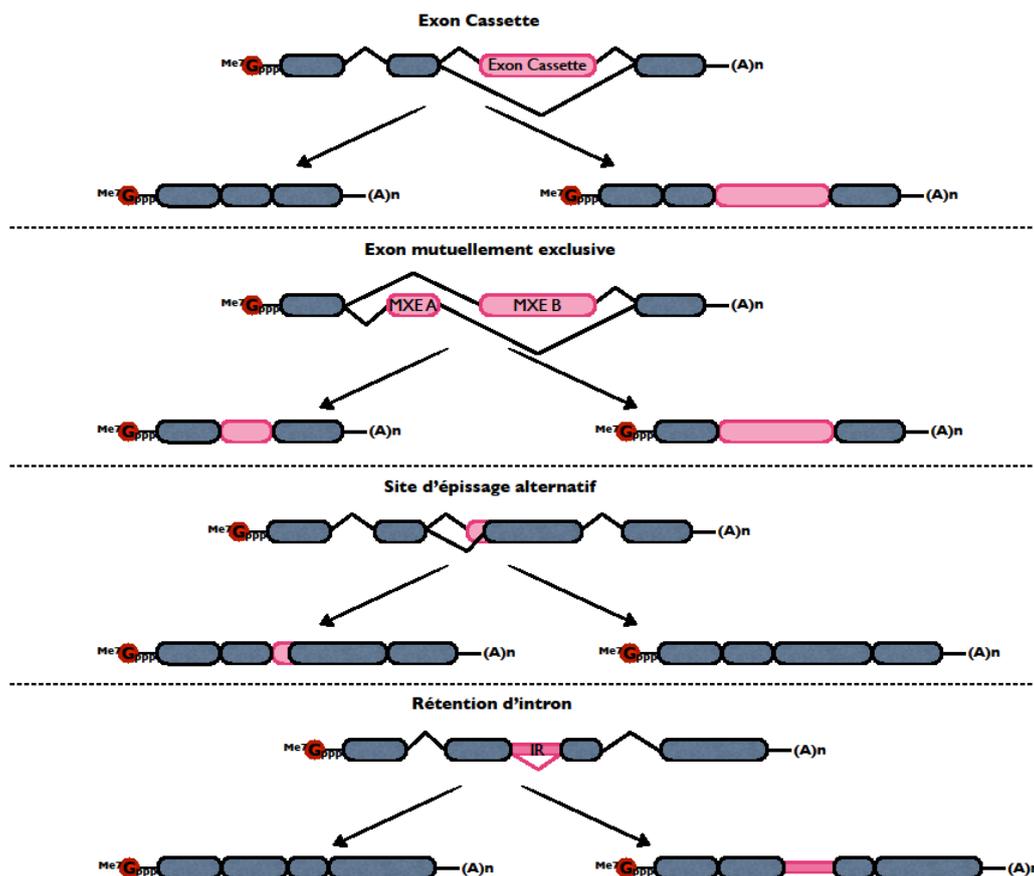
2) Le cas des exons mutuellement exclusifs – MXE pour *Mutually eXclusive Exon* – correspond à deux exons qui ne se côtoient jamais au sein d'un même transcrit ; l'inclusion de l'un entraînant l'exclusion de l'autre. (**Figure 8b**)

3) La rétention d'intron – IR pour *Intron Retention* – correspond à l'inclusion sélective d'un intron – délimité par des sites d'épissage en 5' et en 3' – au sein d'un ou plusieurs transcrit(s). (**Figure 8c**)

4) Les sites d'épissages alternatifs en 5' et 3' – 5' ASE et 3' ASE pour *5'/3' Alternative Splice Site* – correspondent à la mise en concurrence de deux ou plusieurs séquences d'ancrage des composantes du spliceosome. Ces événements provoquent l'apparition dans plusieurs isoformes de versions « tronquées » ou « allongées » des mêmes exons. (**Figure 8d**)

5) Le premier exon alternatif – AFE pour *Alternative First Exon* – correspond à un cas d'exons alternatifs – CE ou MXE – limité à la position du premier exon. Ce cas est particulier – et donc distingué des CE et MXE classiques – puisque ces événements sont plus souvent l'utilisation d'un promoteur alternatif (Ayoubi and Van De Ven, 1996).

6) Le dernier exon alternatif – ALE pour *Alternative Last Exon* – correspond à un cas d'exons alternatifs – CE ou MXE – limité à la position du dernier exon. Ce cas est particulier – et donc distingué des CE ou MXE classiques – puisque les exemples étudiés présentent des couplages entre épissage alternatif et polyadénylation (Lou et al., 1998).



**Figure 8 : Classification des évènements d'épissage alternatif.** (a) CE – Cassette Exon. Un exon différemment inclus dans le messager final. (b) MXE – Exon mutuellement exclusifs. L'inclusion d'un exon provoque l'exclusion de l'autre. (c) 5'ASS – Site d'épissage alternatif en 5'. Le site d'épissage choisi par le spliceosome varie, changeant la taille d'un exon (d) IR – Rétention d'intron. Un intron parfois conservé dans un l'ARN messager mature.

Les événements dits d'épissage alternatif ne sont vus que par la séquence des ARNm finaux produits, et lue ou vérifiée par séquençage, indépendamment de l'action du spliceosome. Cette précision est nécessaire puisqu'il existe des cas d'épissage non-conventionnel - comme celui du gène *xbp-1* - dont la classification basée sur séquençage le placerait dans la catégorie des IR alors même que la nature même de l'intron peut être remise en question. En effet, *xbp-1* présente, dans son messager final, une séquence considérée comme un intron non-conventionnel dont la présence modifie le cadre de lecture, et dont la traduction amène à la production d'une protéine tronquée supposée non-fonctionnelle (Yoshida et al., 2001). L'excision endonucléolytique de cette courte séquence, couplée à la ligature des deux parties séparées du messager, est provoquée dans des conditions particulières de stress du réticulum endoplasmique et permet de changer le cadre de lecture effectif, amenant à la production d'un facteur de transcription actif (Uemura et al., 2009). Ce

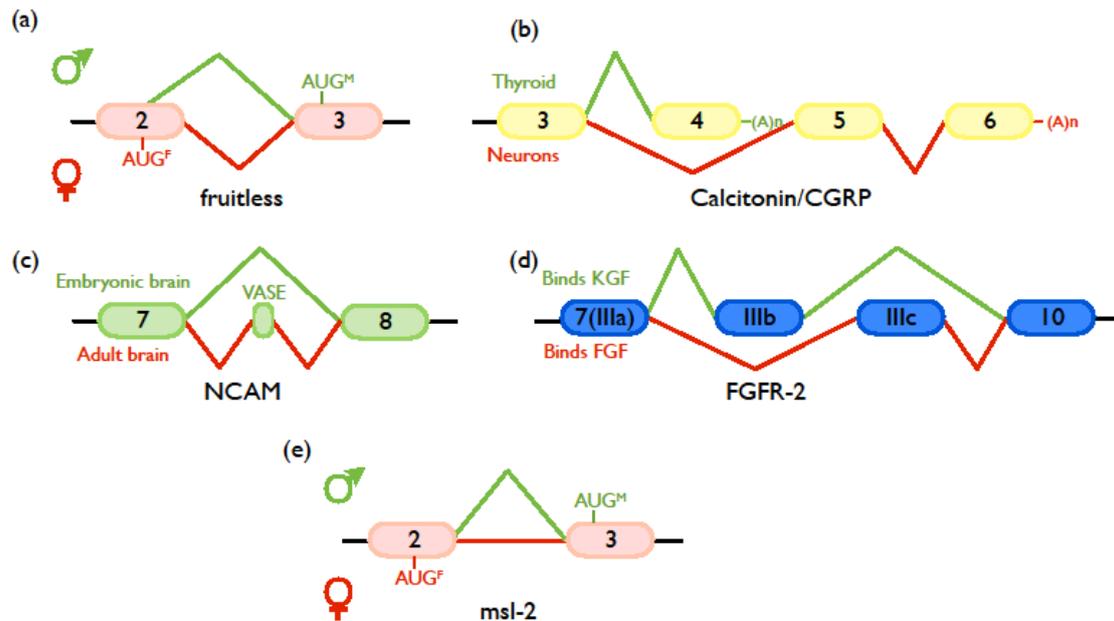
cas s'écarte donc des précédents puisqu'il s'agit d'un mécanisme indépendant de la maturation des ARNm, ou tout du moins des pré-messagers « classiques », au côté desquels, il fait figure de cas particuliers.

## 6. L'impact de l'épissage alternatif sur le vivant

L'étude de l'épissage alternatif a permis de mettre en évidence des régulations spécifiques à certains tissus, impliqués dans la mise en place de fonctions biologiques.

Chez l'organisme modèle *Drosophila melanogaster*, plusieurs exemples notables ont été mis en évidence. Chez les membres de cette espèce, l'épissage du transcrit du gène *fruitless* change l'orientation et le comportement sexuel des individus. L'expression de ce gène est limitée à environ 500 cellules du système nerveux et est l'un des premiers éléments impliqués dans la mise en place du comportement des mâles vis-à-vis des femelles – la parade nuptiale (Ryner et al., 1996). L'épissage du gène varie entre mâles et femelles, et passe par l'utilisation d'un site d'épissage 5' alternatif. Ce changement affecte le choix du codon START utilisé par la machinerie traductionnelle (**Figure 9a**). Des femelles mutantes présentant l'épissage spécifique des mâles effectuent une parade nuptiale caractéristique des mâles – comportement absent des femelles sauvages. Ces parades sont dirigées vers d'autres femelles ou vers des mâles mutants produisant des phéromones féminines (Demir and Dickson, 2005).

Le choix du sexe chez la drosophile est également lié à l'épissage. Lors du développement embryonnaire, le devenir sexuel de l'individu est dépendant du nombre de chromosomes X. La présence de deux gonosomes provoque l'expression de protéines régulatrices activant le gène *Sex-lethal* (*Sxl*). La protéine SXL active l'épissage de son propre pré-messager et maintient le mécanisme de différenciation sexuelle par une boucle de régulation positive (Salz and Erickson, 2010).



**Figure 9 : Différents évènements d'épissage alternatif et leurs conséquences biologiques.** (a) L'utilisation d'un site d'épissage 5' alternatif du gène *fruitless* de la *Drosophila* gouverne le comportement sexuel. L'épissage « mâle » est en VERT, et « femelle » en ROUGE. (b) L'épissage alternatif, en combinaison avec un site de polyadénylation spécifique à chaque isoforme, d'un gène des vertébrés encodant la calcitonine – hormone homéostatique – dans la thyroïde (en VERT) et le CGRP (Calcitonin-Gen-Related Peptide) – neuropeptide vasodilatateur – dans le système nerveux (en ROUGE). (c) L'inclusion différentielle d'un exon alternatif (VASE-Variable Alternatively Spliced Exon) du gène *NCAM* dans le cerveau embryonnaire du rat (en VERT) et le cerveau adulte (en ROUGE), contrôle la croissance axonale durant le développement. (d) Les exons mutuellement exclusif IIIb et IIIc du gène *FGFR-2* change l'affinité du récepteur produit pour des facteurs de croissance durant la progression des cancers prostatiques. La version en VERT présente une forte affinité pour le facteur KGF et en ROUGE, une forte affinité pour le FGF. (e) La présence d'un intron 'en ROUGE) dans la région 5'UTR du gène *msl-2* est spécifique des femelles drosophiles. La protéine Sex-lethal permet la rétention de l'intron dans le noyau et la répression traductionnelle de l'ARNm dans le cytoplasme, ce qui contrôle la compensation du dosage des gènes du chromosome X. Extrait de (Smith and Valcárcel, 2000).

L'exemple le plus impressionnant est celui de *Dscam*. Les produits de ce gène sont impliqués dans le système immunitaire et le câblage du système nerveux de la drosophile, dirigeant entre autre les projections neuronales du système olfactif et des neurones mécanosenseurs (Zipursky et al., 2006)(Zhu et al., 2006)(Chen et al., 2006). Le pré-messager transcrit contient 115 exons dont 95 alternatifs pour une estimation totale de 38 016 isoformes différentes (Park and Graveley, 2007). Ces messagers diffèrent notamment par leur domaines extracellulaires, capables d'homodimerisation (Wojtowicz et al., 2004). Le nombre

de ces domaines capables d'interactions isoforme-spécifique a été estimé à 19 008. Par séquençage massif des ARNs, l'expression de 18 496 isoformes de messagers a été validée *in vivo* (Sun et al., 2013). Certaines isoformes protéiques sont capables d'interagir avec *Escherichia coli*. Cela pourrait suggérer que Dscam puisse fonctionner à la manière d'un anticorps chez les insectes (Watson et al., 2005).

Quelques exemples d'événements tissu-spécifique ont également été mis en évidence dans la mise en place de fonctions biologiques supérieures.

Chez la poule (*Gallus gallus domesticus*), 576 variants d'épissage possibles d'un canal à cations sont exprimés d'après un gradient le long des 10 000 récepteurs sensoriels qui composent la cochlée. Ce gradient est impliqué dans la détection des fréquences sonores de 50 à 5000 Hz (Rosenblatt et al., 1997). Ce mécanisme pourrait être conservé puisqu'un système de régulation très similaire a été identifié dans la cochlée de la tortue de floride *Trachemys scripta elegans* (Jones et al., 1999).

La détection infrarouge d'une chauve-souris hémaphage (vampire commun - *Desmodus rotundus*) présente un événement limité à l'un de ses ganglions, provoquant la production d'une version tronquée du récepteur TRPV1. Cette variante de la protéine présente une activation thermique à 30°C, contre 43°C dans la version présente dans d'autres ganglions, et permet à l'animal de détecter les proies potentielles en les discriminants sur la base de leur température corporelle (Gracheva et al., 2011).

La dérégulation d'événements d'épissage alternatif est également la source de nombreuses maladies chez l'Homme. Les mutations affectant l'épissage pourraient participer à 15% des maladies génétiques humaines (Krawczak et al., 1992). Parmi ces dérégulations, il existe deux types de mutations qui affectent l'épissage alternatif :

1) Une mutation n'affectant pas la séquence codante du gène, peut provoquer la disparition d'éléments régulateurs en *cis*. Les maladies issues de ces mutations les mieux caractérisées sont des tauopathies – pathologies neurodégénératives liées à l'agrégation de protéines Tau – comme l'amyotrophie spinale (Lorson et al., 1999). Cette pathologie est causée par la perte du gène *smn-1* et la non-compensation du phénotype par le gène *smn-2*. La différence fondamentale entre ces deux gènes repose sur une transition C->T dans une séquence ESE favorisant l'exclusion de l'exon 7 du gène *smn-2*. Cette différence provoque la synthèse de la protéine SMN $\Delta$ 7 qui est dégradée rapidement. La perte fonctionnelle de *smn-1*

ne peut donc être compensée par le produit du gène *smn-2* et cause la maladie. Mais le nombre de copies de *smn-2* affecte la sévérité du phénotype (Gavrilina et al., 2008).

Un cas particulier est celui de la dystrophie myotonique de Steinert où les répétitions cumulées d'un triplé contenu dans un ARNm piège un régulateur de l'épissage, ce qui altère la régulation d'autres gènes (Philips et al., 1998).

2) Une mutation peut également affecter la séquence, ou la régulation, d'un facteur en *trans*. Parmi les exemples correspondants à ce type de mutation, on retrouve la dystrophie facio-scapulo-humérale. Cette pathologie neuromusculaire est la résultante de la surexpression de la protéine FRG1. La surabondance de ce régulateur change le résultat d'épissage de plusieurs ARN dont ceux des gènes *tnnt3* et *mtmr1*. TNNT3 régule la contraction musculaire et MTMR1 est impliqué dans l'atrophie musculaire (Gabellini et al., 2006).

On retrouve également des cancers présentant des altérations de l'expression de régulateurs comme FOX2. Près d'un tiers des gènes dont un exon est exclu spécifiquement dans des cancers du sein ou des ovaires possèdent des sites de fixations pour le régulateur FOX2. Parmi les produits de ces gènes, on retrouve des protéines associés à l'actine et la myosine, des kinésines et des composantes de complexes de trafic cellulaire, impliqués dans l'architecture cellulaire et la migration. De plus, le régulateur est souvent retrouvé sous-exprimé dans des cancers des ovaires et son épissage est modifié dans des tumeurs mammaires (Venables et al., 2009).

SRSF1, est un autre régulateur d'épissage, également retrouvé surexprimé dans de nombreuses tumeurs humaines. Sa surexpression entraîne la transformation maligne dans des lignées cellulaires, faisant de ce facteur une oncoprotéine (Anczuków et al., 2012). D'autres régulateurs comme les protéines hnRNPA1, hnRNPA2 et PTB affectent les proportions de deux isoformes de la pyruvate kinase, PKM1 et PKM2. Alors que PKM1 est caractéristique des cellules différenciées, PKM2 est surexprimée dans les cellules embryonnaires et cancéreuses. Cette surexpression est l'un des facteurs clés de la glycolyse élevée qui permet aux cellules cancéreuses de survivre dans des conditions anaérobiques (David et al., 2010).

La mise en évidence de l'impact de l'épissage dans les maladies permet de concevoir de nouvelles formes de traitement. Cependant, des cribles à l'aide de rapporteurs fluorescents permettent d'identifier des composants qui affectent des événements d'épissage spécifiques (Stoilov et al., 2008)(Warf et al., 2009). L'objectif est donc de mettre en évidence des molécules capables d'agir en tant qu'inhibiteur de régulateurs particuliers, comme se pourrait

être le cas pour l'amyotrophie spinale décrite plus haut. En effet, mettre en évidence une molécule capable de favoriser l'inclusion de l'exon 7 dans le produit du gène *smn-2* permettrait de contrebalancer la perte de *smn-1*. Cependant, cette approche est limitée en raison de l'implication de nombreux facteurs en *trans* dans l'épissage alternatif mais également dans l'épissage constitutif. La recherche de composés ciblant ces facteurs pourraient donc être contre-productive en affectant l'épissage de nombreux autres pré-messagers, et donc être délétère pour l'organisme.

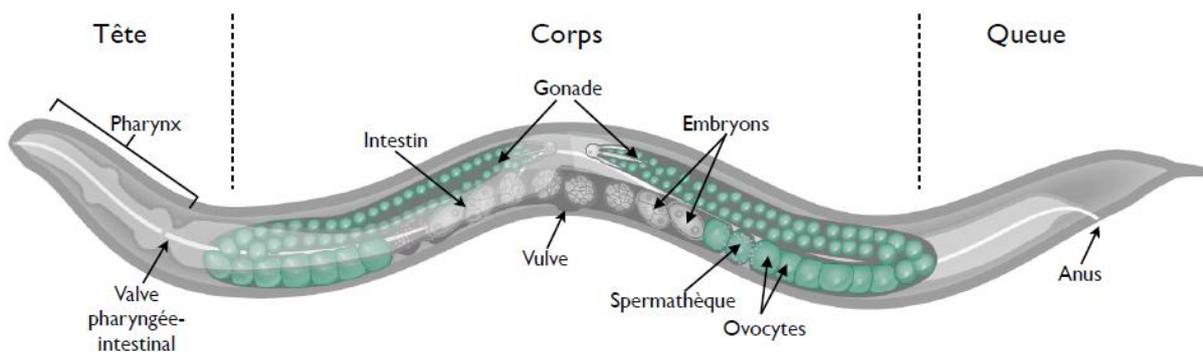
D'autres approches thérapeutiques étudiées sont basées sur l'utilisation : d'oligonucléotides antisense (ASO – AntiSens Oligonucléotides) qui ciblent des séquences de 15-20 nucléotides situées sur des pré-messagers afin d'empêcher les interactions avec le spliceosome et de rediriger l'épissage (Hua et al., 2010) ; de siRNA ciblant les isoformes aberrantes (Ryther et al., 2004) ; d'ARN U1 modifié en 5' capable de s'associer aux sites d'épissages modifiés et de rétablir l'épissage sauvage (Fernandez Alanis et al., 2012) ; d'une stratégie de *trans*-splicing permettant l'ajout d'une séquence exogène à un ARNm aberrant par le biais de la machinerie du spliceosome pour rétablir la séquence codante sauvage (Puttaraju et al., 1999).

L'épissage est un mécanisme complexe dont l'impact physiologique est important. De nombreuses questions restent à élucider pour permettre la compréhension du "splicing code" de telle sorte que l'on puisse l'exploiter à des fins thérapeutiques. Parmi le nombre d'outils utiles à l'étude de ce mécanisme de régulation, celui utilisé au cours de cette thèse consiste en un système de rapporteur fluorescent appliqué chez un organisme modèle, le nématode *Caenorhabditis elegans*.

## IV. *Caenorhabditis elegans*

### 1. Généralités

*Caenorhabditis elegans* est un ver rond de la famille des nématodes, introduit comme organisme modèle en laboratoire par le biologiste moléculaire Sydney Brenner dans les années 1970. Les avantages de cet organisme sont sa petite taille – 1mm de long pour un adulte -, la taille de la progéniture produite par un individu – près de 300 individus -, la possibilité de préserver les lignées plusieurs années à -80°C et son temps de génération faible - ~3 jours. Ces propriétés font du nématode un outil peu coûteux et facile à cultiver. *C.elegans* est également transparent à tous les stades de son développement, ce qui permet de suivre l'expression d'un gène au cours du temps en utilisant des rapporteurs fluorescents.



**Figure 10 : Schéma représentant un individu hermaphrodite adulte de l'espèce *Caenorhabditis elegans*.** L'individu est divisé en trois parties. La tête contient le pharynx, la bouche ainsi que les organes sensoriels. Dans le corps est représenté l'intestin ainsi que le système reproductif constitué des gonades et de la spermathèque. Dans la queue se termine l'appareil intestinal ainsi que d'autres organes sensoriels. Le corps est entouré de fuseaux musculaires qui permettent sa locomotion.

Son génome a été entièrement séquencé en 1998 (The *C. elegans* Sequencing Consortium, 1998) et est composé de 100 Mb, répartis en 5 autosomes, numérotés de I à V, ainsi que d'un gonosome – le chromosome X. Bien que la taille de son génome soit très inférieure à celle de celui de l'Homme (3400 Mb), le nombre de gènes identifiés est très proche de celui d'*Homo sapiens* – 20 444 gènes codant pour des protéines (WormBase release WS245, Oct. 2014) pour 19 797 chez l'Homme (Gencode 23 - March 2015 freeze, GRCh38 - Ensembl 81, 82).

En comparaison des vertébrés, les gènes de *C.elegans* sont plus petits avec une taille moyenne de 3 kilobases contre 8,5 kb chez l'homme (Jareborg et al., 2000). Cette différence

est en partie due à la taille réduite des introns du nématode – d’une taille médiane de 65pb (Spieth et al., 2013).

## 2. Développement et reproduction

*C.elegans* possède un cycle de développement très rapide. Trois jours à une température comprise entre 20 et 25°C permettent à un embryon d’atteindre le stade adulte, après avoir passé les quatre stades larvaires – L1 à L4 - nécessaires à sa maturation. Les individus sont en majorité hermaphrodites et développent, durant leur troisième stade larvaire, des spermatozoïdes amiboïdes qui seront ensuite conservés dans une structure baptisée spermathèque. C’est au cours du quatrième et dernier stade larvaire que les gonades femelles se différencient pour produire des ovocytes une fois les vers devenus adultes. Ce nématode est donc capable de se reproduire individuellement et de produire plusieurs centaines d’embryons par adulte.

Cette reproduction individuelle simplifie les conditions de culture, puisqu’un seul ver peut donner naissance à toute une population clonale poussant les populations vers l’homozygotie.

Une très faible quantité de mâles existe au sein d’une population, moins de 0.2% dans des conditions de culture en laboratoire. Ces mâles sont capables d’inséminer les individus hermaphrodites et peuvent donc être utilisés dans le cadre de croisements génétiques. Le devenir sexuel de l’individu est dépendant de la présence d’un ou deux copies du chromosome X – mâle et hermaphrodite, respectivement. La proportion de mâles peut être modifiée par augmentation de la température ou par croisement (la moitié des spermatozoïdes produits par des mâles ne portant pas de chromosome X).

Placés en conditions défavorables de culture – manque de nourriture, surpopulation, température élevée -, les larves peuvent prendre une route parallèle de développement. Cette voie provoque la mise en place d’un stade supplémentaire entre les stades larvaires L2 et L3, nommé stade *dauer*. Durant cette étape, la morphologie du ver s’adapte – pharynx clos, gonades atrophiées, cuticule renforcée - et sa durée de vie s’allonge, de quelques semaines en développement standard, à plusieurs mois sous cette forme. Cette particularité avantage fait du ver un organisme facile à maintenir en laboratoire ainsi qu’un outil de choix pour l’étude du vieillissement.

### 3. Découvertes majeures et champs d'étude

La transparence de l'animal a permis de suivre l'ensemble des divisions cellulaires qui permettent à un embryon de devenir un adulte. Ces travaux ont permis de déterminer que le nombre de cellules chez ce nématode est quasiment invariable et d'obtenir le premier lignage cellulaire complet d'un métazoaire (Sulston and Horvitz, 1977)(Kimble and Hirsh, 1979)(Sulston et al., 1983). En parallèle de ces travaux, un effort important fut apporté à la description de l'anatomie de système nerveux et des connexions neuronales (Ward et al., 1975)(White et al., 1976)(White et al., 1986).

L'utilisation de *Caenorhabditis elegans* a permis de découvrir et caractériser des voies de signalisation impliquées dans un grand nombre de fonctions biologiques variées, comme par exemple celle de l'apoptose (Hedgecock et al., 1983)(Ellis and Horvitz, 1986) et la voie de signalisation *Notch* dans le développement embryonnaire (Priess, 2005). L'analyse des phénotypes hétérochroniques - c'est à dire présentant des altérations de la durée ou de la vitesse du développement -, ainsi que leur causes génétiques, a permis la découverte du premier microARN, *lin-4* (Ambros and Horvitz, 1984)(Lee et al., 1993).

Année	Découverte(s)	Référence(s)
1974	Mutations affectant le comportement	(Brenner, 1974)
1977	Premier clonage et séquençage d'un gène	(MacLeod et al., 1977)
1983	Premier lignage cellulaire complet	(Sulston and Horvitz, 1977) (Kimble and Hirsh, 1979) (Sulston et al., 1983)
1983	Identification de gènes de mort cellulaire	(Hedgecock et al., 1983)
1984	Découverte des gènes hétérochroniques	(Ambros and Horvitz, 1984)
1986	Diagramme complet des connexions neuronales	(White et al., 1986)

1988	Identification des premiers facteurs de transcription à homéodomaines (POU, LIM)	(Way and Chalfie, 1988) (Finney et al., 1988)
1993	Mutations allongeant la durée de vie	(Kenyon et al., 1993)
1993	Premier microARN <i>lin-4</i> et sa cible <i>lin-14</i> identifiés	(Lee et al., 1993) (Wightman et al., 1993)
1993	Identification des gènes <i>smg</i> et de leur rôle dans le NMD	(Pulak and Anderson, 1993)
1994	Utilisation de la GFP pour suivre l'expression d'un gène <i>in vivo</i>	(Chalfie et al., 1994)
1998	Premier génome métazoaire séquencé	(The <i>C. elegans</i> Sequencing Consortium, 1998)
1998	Découverte de l'ARN interférence	(Fire et al., 1998)
2000	Mis en place du premier crible par ARNi	(Fraser et al., 2000)
2005	Utilisation de l'optogénétique	(Nagel et al., 2005)

**Figure 11 : Découvertes et travaux majeurs effectués grâce à l'organisme modèle *Caenorhabditis elegans*.** (Ce tableau n'est en aucun cas exhaustif et les travaux associés à ces découvertes ne sont pas tous référencés par souci de lisibilité).

Sydney Brenner, Robert Horvitz and John Sulston ont été récompensés en 2002 par le Prix Nobel de Physiologie ou Médecine pour leurs découvertes concernant la régulation génétique du développement d'un organisme et la programmation de la mort cellulaire. Andrew Fire et Craig Mello ont obtenu le Prix Nobel de Physiologie ou Médecine en 2006 pour leurs découvertes de l'extinction de l'expression des gènes par des ARN interférents double-brins. Osamu Shimomura, Martin Chalfie et Roger Tsien ont été récompensé du Prix Nobel de Chimie en 2008 pour la découverte de la GFP et son utilisation pour suivre l'expression des gènes *in vivo*.

## 4. La transgénèse chez *C. elegans*

La transgénèse est l'un des outils essentiels à l'étude de la fonction biologique et de l'expression des gènes, ainsi qu'à la mise en place de nombreux cribles génétiques. Il existe deux techniques permettant la transgénèse chez *Caenorhabditis elegans* : le bombardement de micro-particules et la micro-injection. Seule la seconde sera abordée puisqu'elle a été utilisée durant ce travail de thèse.

### **a) Introduction d'ADN exogène**

La micro-injection consiste en l'injection de matériel génétique (plasmides ou produits de PCR) dans le bras distal de la gonade. Cette partie de l'organisme est constituée d'un syncytium où tous les noyaux des ovocytes en devenir partagent le même cytoplasme. L'ADN exogène circule donc librement dans une zone limitée de la gonade où il peut pénétrer les noyaux avant que ceux-ci ne soient compartimentés dans des cytoplasmes d'ovocytes individuels.

Cet ADN peut-être répliqué et transmis lors des divisions cellulaires subséquentes sous la forme d'un concatémère extrachromosomique. Cette longue molécule d'ADN est formée des monomères présents dans le mélange injecté (plasmides, produits de PCR) par le biais de recombinaisons homologue et hétérologue. Ces concatémères peuvent être transmis correctement lors des mitoses car les chromosomes du ver ne possèdent pas de centromères, ils sont donc holocentriques. L'attachement du fuseau mitotique s'opère à plusieurs endroits des chromosomes et ne nécessitent pas de séquence spécifique. Ceci permet à un multimère d'ADN exogène d'être considéré comme un chromosome par la machinerie cellulaire si celui possède une taille suffisante (estimée à centaines de copies de plasmides). La composition du concatémère suit la composition du mélange d'injection de manière stœchiométrique.

(Mello et al., 1991)

La recombinaison homologue au *locus* endogène peut arriver à très faible fréquence (Broverman, 1993)(Berezikov et al., 2004).

La transmission des concatémères n'étant pas aussi efficace que celle des chromosomes, on utilise un marqueur de co-injection afin de permettre la sélection des individus transgéniques.

## **b) La sélection**

Afin de maintenir les lignées transgéniques, un marqueur de sélection est nécessaire pour suivre la présence des mélanges d'injections. Parmi les marqueurs les plus utilisés on retrouve:

- des allèles mutants provoquant des phénotypes dominant. Les gènes les plus communément utilisés sont de la famille des *dpy* (phénotype : dumpy), *rol* (phénotype : roller) et *unc* (phénotype : uncoordinated) (Mello et al., 1991). Ces marqueurs provoquent une modification du phénotype des individus transgéniques. La sélection se fait grâce à la physionomie des vers ou sur leurs locomotions. Cependant, les phénotypes mutants engendrés peuvent gêner le croisement génétique, perturber la visualisation de rapporteurs fluorescents et engendrer un plus faible nombre de descendants en comparaison d'individus sauvages. De plus, le contexte génétique mutant peut affecter les régulations génétiques étudiées et rendre les conclusions d'expérience moins claires.

- des marqueurs de type *rescue* injectés chez une lignée de vers mutants afin de rétablir un phénotype sauvage. Le marqueur de référence est le gène *unc-119* injecté dans une lignée mutante pour ce gène. L'avantage de ce système est que les mutants, en plus d'être identifiables de par leur locomotion particulière, ne sont pas capables de passer en stade *dauer*, ce qui permet de donner un avantage selectif individus transgéniques dans un contexte de culture défavorables. Le premier marqueur de ce type fut celui du gène *pha-1*. A une température de 25°C, les embryons mutants ne peuvent survivre. L'ajout de copies de gènes *pha-1* sauvage corrige cette particularité (Granato et al., 1994).

- des marqueurs fluorescents. L'utilisation de ces marqueurs pose problème en combinaison de rapporteurs fluorescents.

Le désavantage majeur des systèmes présentés plus haut repose sur la sélection manuelle.

- finalement, un nouveau type de marqueur basé sur la résistance aux antibiotiques s'est développé. En 2010, un article présente les travaux menés au sein du laboratoire du Dr Denis Dupuy sur la résistance à un antibiotique eucaryote nommée G418 ou *Geneticin*<sup>TM</sup>. À une concentration de 0,40 mg/mL de G418 dans le milieu de culture, les individus sauvages ne se développent pas jusqu'au stade adulte qui leur permet de se reproduire, contrairement aux individus transgéniques porteurs du marqueur de résistance – le gène *Neo* (Giordano-Santini et al., 2010). Soumis à l'action de l'antibiotique, les populations présentent un enrichissement suffisant des individus transgéniques pour maintenir une lignée, et ce, sans

sélection manuelle. Deux autres antibiotiques ont également été appliqués chez le ver : la puromycine et l'hygromycine (Semple et al., 2010)(Radman et al., 2013).

### **c) Intégration en simple copie**

L'expression de constructions génétiques sous la forme de concatémères extrachromosomiques pose plusieurs problèmes.

Premièrement, il existe chez l'organisme *Caenorhabditis elegans* un mécanisme moléculaire de répression des éléments génomiques répétés au sein des cellules de la lignée germinale (Kelly et al., 1997). Ce mécanisme se met en place sur plusieurs générations et empêche l'expression prolongée de rapporteurs dans les gonades et les embryons au sein d'une lignée transgénique.

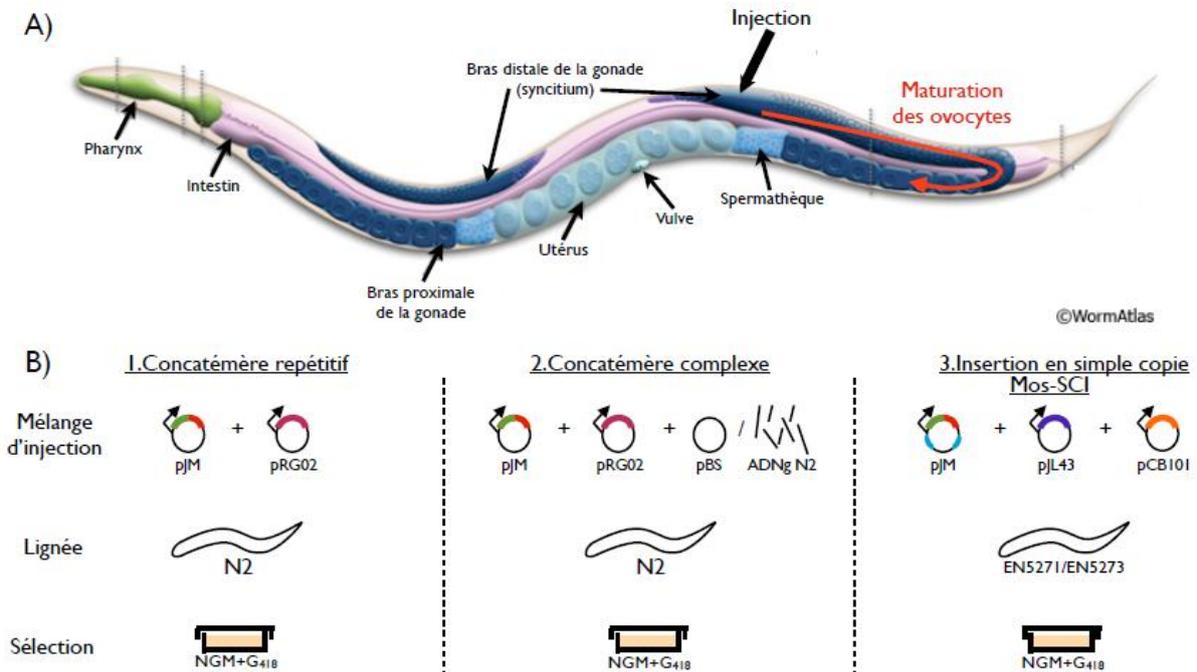
Deuxièmement, les constructions génétiques exprimées en grand nombre de copies peuvent présenter, chez *C.elegans*, une expression partiellement stochastique en fonction des éléments de régulation choisis (Qun Xu *et al.*, Fire Lab Vector Supplement, 1997). Au sein d'une même lignée, il peut donc y avoir des variations d'expression d'un individu à l'autre. Cette particularité est généralement contrecarrée en étudiant plusieurs individus issus de deux à trois lignées indépendantes. Deux lignées sont indépendantes lorsque les individus qui les ont générées sont issus de deux concatémérisation indépendantes. Ainsi, deux individus issus de la progéniture d'un hermaphrodite injecté peuvent donner naissance à deux lignées indépendantes.

Un rapporteur en très grand nombre de copies peut également provoquer des problèmes de létalité. En effet, une construction générée par la fusion d'une séquence codante partielle peut entraîner la production d'un dominant négatif aux effets délétères pour l'organisme. De plus, l'utilisation de séquences de régulation peut entraîner la titration des facteurs impliqués dans cette régulation et entraîner une modification de l'expression endogène d'autres gènes.

Un moyen pour lutter efficacement contre tous ces problèmes est la diminution du nombre de copies de rapporteur. Ceci peut être réalisé en modifiant la composition du mélange d'injection. Les proportions entre les éléments du mélange et la composition du concatémère étant stœchiométriques, il est possible de rajouter des séquences d'ADN génomiques digéré afin de diminuer le nombre de copies de la construction d'intérêt dans les lignées générés pour l'étude (Kelly et al., 1997). Les concatémères extra-chromosomiques sont dits « complexes ».

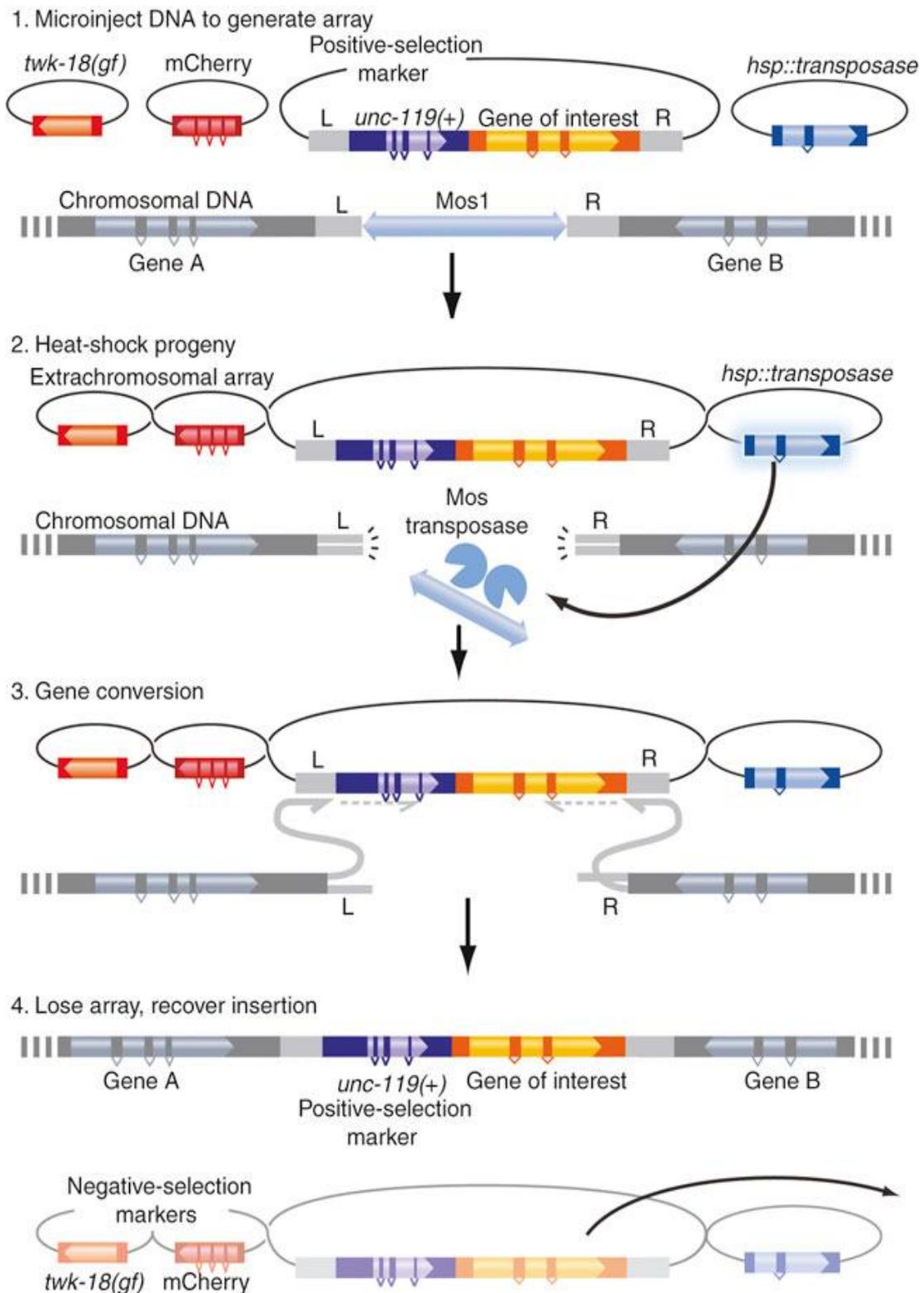
Une manière plus radicale de contourner ces problèmes est de passer par l'intégration des transgènes dans le génome. Il existe une méthode d'intégration aléatoire en faible nombre

de copies mais elle est dépendante d'une autre technique de transgénèse, le bombardement à microparticules (Praitis et al., 2001) que je ne développerai pas ici car elle a été supplantée par les méthodes d'intégration ciblée en simple copies comme la technique du MosSCI – Mos Single Copy Insertion (Frøkjær-jensen et al., 2009).



**Figure 12 : Micro-injection et expression de transgènes chez *Caenorhabditis elegans*.** (A) L'aiguille de micro-injection permet d'inoculer la matériel génétique désiré dans le syncytium de la gonade. Ce sont les noyaux des ovocytes en maturation qui sont perméables aux plasmides injectés. (B) La micro-(injection permet de jouer sur le nombre de copie de transgènes en fonction de la composition du mélange mais également de la lignée sélectionnée. L'expression sous la forme de concatémères répétitifs se fait dans une lignée classique (N2) avec seulement le rapporteur (pJM ici) et le marqueur de sélection (pRG02). La formation de concatémères complexes se fait en rajoutant des plasmides vides ou de l'ADN génomique digéré afin de diminuer la quantité de rapporteur. L'insertion en simple copie par la méthode du MosSCI nécessite l'utilisation de lignées particulières ainsi que des vecteurs spécifiques (pJM, pJL43) présentant les séquences d'homologies et la transposase. Image du ver récupérée de WormAtlas.

La technique du MosSCI repose sur l'insertion dans le génome de *Caenorhabditis elegans* d'un transposon issu de la drosophile, le transposon *Mos*. Inséré à un endroit précis du génome, il peut être éliminé par le biais de l'expression d'une transposase spécifique, provoquant ainsi une cassure double brin. Il est alors possible d'insérer une séquence d'intérêt flanquée des séquences homologues aux zones proches de la cassure, favorisant ainsi l'insertion par recombinaison homologue (**Figure 13**).



**Figure 13: Représentation schématique de l'insertion en simple copie par MosSCL.** Un transposon *Mos1* est situé à un locus non-codant. Cet élément peut être éliminé du génome par expression d'une transposase provoquant une coupure double brins. Grâce à des séquences d'homologie, l'insertion d'une séquence d'intérêt est favorisée par le biais de la recombinaison homologue. Dans l'exemple présenté, deux marqueurs négatifs sont présents dans le mélange d'injection. Ces rapporteurs ne peuvent être insérés et servent à discriminer les événements d'insertion. Le premier est un rapporteur fluorescent. Le second – *twk-18(gf)* – provoque la paralysie des individus le portant à 25°C mais pas à 15°C. Extrait de (Frøkjær-jensen et al., 2009).

D'autres méthodes basées sur les coupures double-brin et la recombinaison homologue utilisées chez de nombreuses espèces ont également été appliquées chez le nématode comme l'utilisation des nucléases à doigts de zinc – ZFN (Morton et al., 2006) – et des TALENs (Wood et al., 2012). Depuis 2013, la méthode de CRISPR-Cas9 s'est développée chez *Caenorhabditis elegans* d'abord en août 2013 (Friedland et al., 2013) puis lors de la publication de cinq articles traitant de l'application de cet outil à cet organisme modèle dans la revue Genetics de novembre 2013 (Waaijers et al., 2013)(Tzur et al., 2013). Cette méthode pourrait devenir celle de référence et sera développer dans la partie Résultats.

## 5. Les rapporteurs fluorescents

L'un des avantages majeurs de l'organisme modèle *Caenorhabditis elegans* réside dans le fait que les individus sont transparents à tous les stades de leur développement, permettant de suivre l'expression d'un gène *in vivo*. De plus, la largeur du ver (75 µm) permet de se passer de microscopie à haute résolution.

Cet avantage a permis initialement le développement de rapporteurs de l'expression des gènes basés sur le gène *LacZ* (Fire et al., 1990). Dans un second temps, une nouvelles séries de rapporteurs basés sur la séquence codante de la GFP a pu voir le jour (Chalfie et al., 1994). Ces travaux, comprenant la découverte de la protéine fluorescente, ont été récompensés par le Prix Nobel de Chimie en 2008.

L'utilisation de ces rapporteurs a permis l'étude de l'expression génique, de l'étude des séquences en *cis* et l'identification cellulaire. Depuis, ce type d'outils est appliqués à l'étude de nombreux mécanismes de régulation qu'il serait trop long de citer ici.

Les prochaines parties ont pour but de présenter les différents types de rapporteurs qui ont pu voir le jour chez *Caenorhabditis elegans* dans le cadre de l'étude de l'épissage alternatif en particulier.

### **a) Approche « classique »**

Ce type de système rapporteur est appelé ici « classique » majoritairement en raison de l'absence de nom particulier. Il s'agit dans ce cas de créer deux constructions pour étudier un seul événement d'épissage. Une première construction consiste à l'addition de la GFP à l'extrémité 3' de la séquence codante du gène étudié afin de visualiser l'expression de toutes les isoformes – dans la limite imposé par les séquences de régulation transcriptionnelle utilisées. Une seconde construction repose sur l'addition de la GFP directement au sein de la séquence différenciellement inclus dans le messenger final (exons ou introns). Puisque les deux

protéines hybrides présentent les mêmes caractéristiques de fluorescence, l'analyse repose sur la comparaison des deux types de lignées générées (Malone et al., 2003).

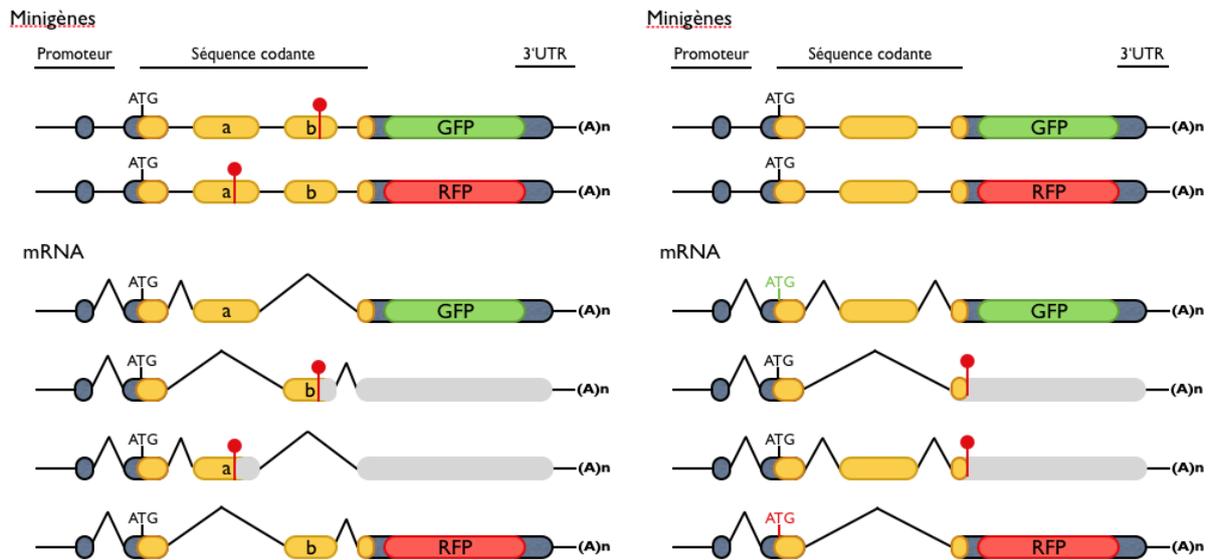
Ce système possède des limitations. La nécessité des deux lignées présente des difficultés de comparaison. De plus, la présence dans une cellule d'une forme n'exclue pas la présence de l'autre et limite donc l'analyse de la régulation. Cependant malgré cela, des études récentes continuent d'utiliser ce système (Craig et al., 2013).

### **b) Approche “paire de minigènes”**

Les minigènes correspondent à des constructions minimales composées des éléments de régulations – Promoteurs, régions 3'UTR –, des séquences codantes partielles, comprenant la séquence d'intérêt et son environnement génomique immédiat – les deux introns et exons proximaux -, couplées à des séquences de fluorophores. Deux rapporteurs sont générés pour étudier un événement d'épissage alternatif et injecté ensemble permettant la comparaison des profils d'expression des deux protéines fluorescentes au sein d'un même individu.

Dans le cas d'exons mutuellement exclusifs, les constructions nécessitent la présence de codons stop afin de différencier les deux isoformes (Figure 14-Gauche). Pour le cas d'un exon cassette, il est nécessaire de disposer d'un changement de cadre de lecture qui permet de visualiser la présence ou l'absence de la séquence (Figure 14-Droite).

Les séquences choisies afin d'assembler les rapporteurs ne contiennent pas nécessairement tous les éléments de régulations des gènes étudiés. Les promoteurs utilisés correspondent aux 2 kb en amont du codon START – suivant ainsi les règles établies par le projet du ‘Promoterome » (Dupuy et al., 2004) - et les régions 3'UTR à une séquence de 1 kb en aval du codon STOP – suivant les règles dégagées par l'étude du projet « UTRome » (Mangone et al., 2008). De plus, la séquence contenue entre le START et le STOP est partielle. De ces faits, le système des minigènes n'est pas utilisé pour suivre avec précision l'expression des isoformes mais plutôt pour servir d' « appât à régulateurs », et ce, afin de mettre en évidence les éléments qui gouverne la régulation.



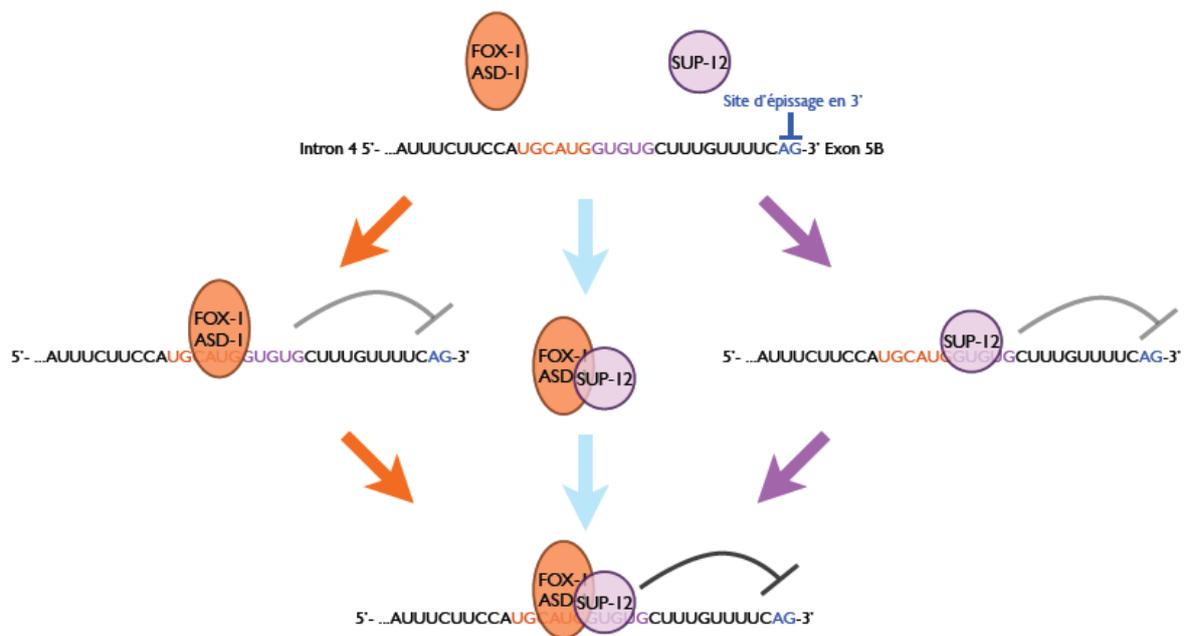
**Figure 14 : Représentation schématique des constructions minigènes.** (À gauche) Minigènes ajustés à l'étude de deux exons mutuellement exclusifs. Deux constructions similaires sont réalisées en changeant la séquence codante du fluorophore. Pour chaque minigènes, un codon stop est rajouté dans la séquence codante d'un des exons alternatifs. Une fois transcrits et épissés, l'inclusion de chaque exon est suivie par la fluorescence des protéines hybrides générées. (À droite) Minigènes adaptés à l'étude d'un exon cassette. Les constructions sont réalisées de telle sorte que l'inclusion de l'exon alternatif change le cadre de lecture - que cela soit naturel ou par mutagenèse dirigée. L'inclusion de l'exon est suivie par la comparaison des deux fluorescences. Modifié de (Kuroyanagi et al., 2010).

Les paires de « minigènes » se sont développées pour suivre l'épissage par RT-PCR (Lorson et al., 1999) ou par le biais de rapporteur luciférase (Weg-Remers et al., 2001) avant d'être couplées à la GFP pour suivre l'épissage des gènes SMN1 et SMN2 humains dans des lignées cellulaires en culture afin d'étudier les éléments en *cis* impliqués dans leur régulation (Zhang et al., 2001).

Ce type de rapporteur a notamment été utilisé chez *Caenorhabditis elegans* par le Dr Hidehito Kuroyanagi. Dans un premier temps le système a été appliqué sur l'épissage du gène *let-2* dans le cadre d'un crible de génétique classique afin d'isoler des mutants affectant le profil de fluorescence (Ohno et al., 2008). Ces travaux ont permis de mettre en évidence le facteur ASD-2 ainsi que des éléments en *cis* nécessaire à la régulation. L'équipe en charge de ces travaux publie deux ans plus tard une méthodologie à appliquer à l'étude systématique de l'épissage alternatif. Cette identification est basée sur la mise en évidence des séquences en *cis* par approche bio-informatique puis par identification par le biais de mutagenèse dirigée sur les rapporteurs. La mise évidence des facteurs en *trans* se fait par le biais de l'utilisation des rapporteurs, de cribles de génétiques classiques et de séquençage (Kuroyanagi et al., 2010). Cette méthode est ensuite appliquée à l'étude du gène *unc-60*, mettant en évidence une

régulation conjointe par ASD-2 et SUP-12 (Ohno et al., 2012), et à l'étude du gène *unc-32* permettant l'identification du facteur UNC-75 dans une régulation de l'épissage alternatif spécifique aux neurones (Kuroyanagi et al., 2013).

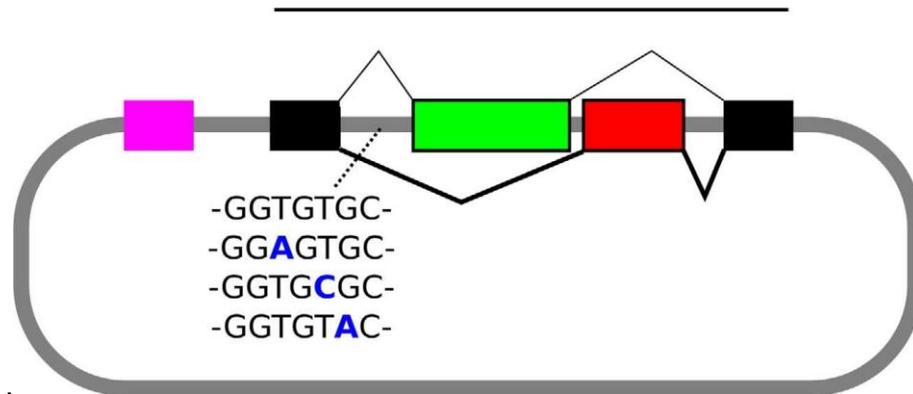
La même équipe a développé, précédemment aux travaux présentés plus haut, un autre rapporteur qui reprend le même principe que les paires de « minigènes » mais basé sur une seule construction et applicable à l'étude des exons mutuellement exclusifs. Il s'agit d'un minigène n'utilisant que les exons différenciellement inclus et leurs séquences proches – introns et exon à proximité – mais en intégrant les séquences des fluorophores directement dans les exons concernés. Ce rapporteur a été utilisé pour l'étude du gène *egl-15* et a permis la mise en évidence du gène *asd-1*, le premier gène de la famille des "alternative splicing deficient". Ce facteur, ainsi que son paralogue FOX-1, affectent l'épissage de l'ARNm d'*egl-15* (Kuroyanagi et al., 2006). Un autre régulateur, SUP-12, est découvert l'année suivante grâce au même rapporteur, et coopère avec ASD-1/FOX-1 (**Figure 15**) (Kuroyanagi et al., 2007).



**Figure 15 : Coopération des facteurs ASD-1/FOX-1 et SUP-12 dans l'épissage alternatif.** L'utilisation des minigènes a permis l'étude des paralogues ASD-1 et FOX-1 et du facteur SUP-12 ainsi que des séquences en *cis* que ces facteurs fixent. La fixation des deux facteurs à leurs séquences respectives et proches permet d'agir sur l'épissage. Extrait de (Kuroyanagi et al., 2007).

Ces travaux mettent en évidence les éléments en *cis* fixés par les facteurs de régulation ainsi que le phénotype des mutants et doubles mutants des gènes *asd-1*, *fox-1* et *sup-12* (Kuroyanagi et al., 2007).

La construction utilisée pour l'étude d'*egl-15* ont également été utilisés dans un étude quantitative qui a permis de démontrer que l'affinité *in vitro* de SUP-12 pour une séquence affecte le ratio des isoformes d'épissage *in vivo* (**Figure 16**)(Amrane et al., 2014). Cette démonstration a permis de pousser un peu plus loin l'étude des régulateurs de l'épissage alternatif qui se limitait à la présence des facteurs.



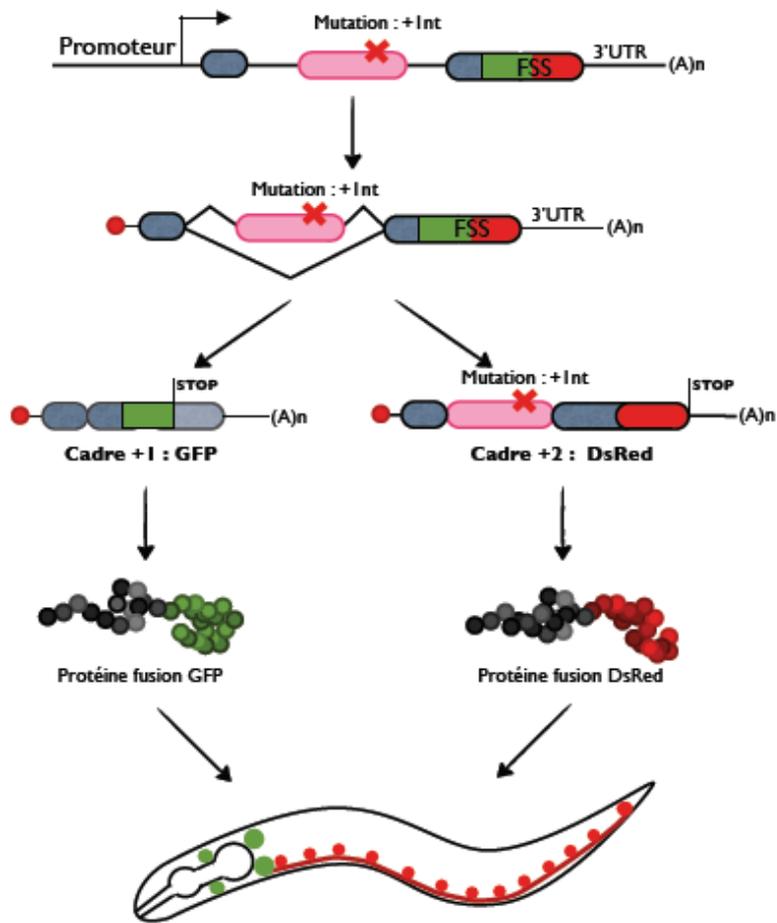
**Figure 16 : Système rapporteur utilisé pour étudier l'impact de l'affinité d'un régulateur pour une séquence en *cis* sur le résultat d'épissage alternatif.** Extrait de (Amrane et al., 2014).

Ces rapporteurs permettent une analyse qualitative, mais qui pourrait également devenir quantitative, et systématique de l'épissage alternatif et de ses composantes. La seule limitation d'une approche à haut débit réside dans l'assemblage et le *design* des constructions rapportrices.

Récemment, un nouveau type de construction assez similaire a vu le jour.

### **c) Approche "FrameShift Sensor"**

Le FrameShift Sensor – FSS – est une cassette composée de la séquence de la GFP suivie de la séquence de la mCherry. L'avantage de cette stratégie par rapport à celle des minigènes, est qu'elle ne nécessite le clonage que d'un seul plasmide d'expression. Les deux fluorophores ne sont pas codés dans le même cadre de lecture. L'inclusion d'un exon en amont de la cassette FSS causera donc la traduction d'un fluorophore diffère que son exclusion. Si l'exon alternatif endogène ne provoque pas de changement de cadre de lecture, il est alors nécessaire d'ajouter un nucléotide dans sa séquence par mutagenèse dirigée (**Figure 17**).



**Figure 17 : Schéma représentant le principe du rapporteur FrameShift Sensor.** La construction est composée de la fusion entre une partie de la séquence du gène étudié à la cassette GFP-mCherry. L'exon alternatif présente un changement de cadre de lecture si inclus sous la forme d'un nucléotide additionnel. Après épissage, l'ARNm obtenu permet la synthèse d'une protéine hybride fluorescente différente en fonction de la présence ou de l'absence de l'exon alternatif. Ainsi, à partir d'une seule construction génétique et au sein d'un seul individu, on peut observer deux isoformes différentes.

Ce système rapporteur, a été développé par le Dr John Calarco et depuis été appliqué chez le nématode pour découvrir deux régulateurs du système nerveux (Norris et al., 2014).

Cette évolution de l'idée des minigènes simplifie les étapes de clonage *in vitro* mais permet également d'exprimer les deux fluorophores à partir d'une même séquence génétique évitant ainsi les potentiels artefacts liés à la stoechiométrie des deux vecteurs dans la stratégie paire de « minigènes ». C'est sur ce système que les travaux effectués durant ces trois années sont basés.

## Objectifs

L'épissage alternatif prend une importance croissante dans les résultats de séquençage ARNs à haut débit, pouvant toucher jusqu'à 97% des gènes humains. Ces données s'accumulent mais les stratégies pour déterminer la pertinence des résultats obtenues ainsi que pour étudier en détail les régulations mis à jour sont limitées. Nous souhaitons mettre en place une stratégie simple pour suivre l'épissage alternatif *in vivo* grâce au rapporteur « FrameShift Sensor ». Une telle stratégie pourra permettre de suivre deux isoformes et étudier les éléments impliqués dans cette régulation. Ces données pourront compléter les éléments connus du « splicing code ».

Le but central de cette thèse était d'établir un système rapporteur permettant de visualiser les deux isoformes d'un gène pour pouvoir identifier ensuite des régulateurs impliqués dans la régulation observée. L'utilisation d'un cytomètre en flux adapté aux larges objets COPAS-Profiler – dont le nématode *Caenorhabditis elegans* – aurait servi à isoler des individus mutants issus de mutagenèse aléatoire (Doitsidou et al., 2008). Il s'agit d'une méthodologie similaire à celle publiée sur le système des paires de « minigènes » (Kuroyanagi et al., 2010) mais plus rapide et simple dans l'assemblage des constructions ainsi que dans la mise en place du crible – le cytomètre en flux n'ayant pas été utilisé jusque-là dans le contexte de l'épissage alternatif.

La méthodologie de clonage a été mise en place avant d'être appliquée à plusieurs gènes. Trois d'entre eux ne disposait que de données de séquençage.



# Résultats



# I. Développement de la stratégie

## 1. Le choix des gènes

Dans un premier temps, une liste de gènes répertoriés comme pouvant générer plusieurs isoformes ont été récupérés par le biais du portail BioMart ([central.biomart.org](http://central.biomart.org)). En utilisant cet outil, la base de données en ligne WormBase (WS220) a été criblée pour les gènes présentant des événements d'épissage alternatif des types CE, MXE, IR, 5'ASS, 3'ASS, AFE et ALE. Cette liste contient 2534 gènes.

Il a été décidé de limiter le choix possible aux gènes exprimés dans le système nerveux. En effet, de nombreuses régulations d'épissage ont été mises en évidence dans le système nerveux, faisant de ce tissu, un système de choix (Grabowski and Black, 2001). La liste a ensuite été criblée pour ressortir les gènes associés à une expression neuronale (WormMine, Base de donnée : Wormbase-WS220). Sur les 191 gènes extraits, seulement 29 était associés à un événement d'épissage alternatif unique de type exon cassette - CE. Cette dernière contrainte a été fixée pour limiter les cas étudiés à seulement deux isoformes pour simplifier l'étude. De plus, notre but étant l'étude de l'épissage alternatif, les événements de type AFE – Alternative First Exon – sont à mettre de côté puisqu'étant le fruit de promoteurs alternatifs. Enfin, les événements de type ALE – Alternative Last Exon – ne sont pas adaptés pour le système utilisé en raison des codons stop et séquences 3'UTR associé à chaque isoforme. C'est dans cette dernière liste que les trois premiers gènes ont été extraits : *che-1*, *unc-86* et *gpa-13*.

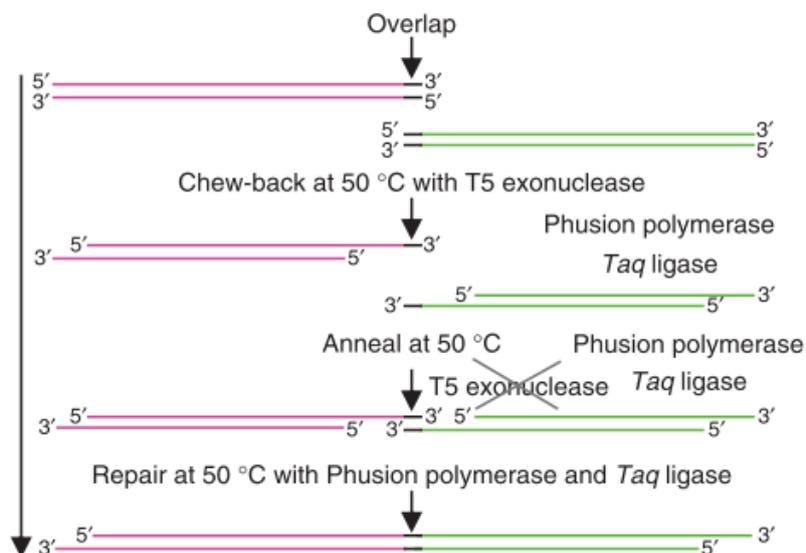
Deux autres gènes, *top-1* et *hrpf-1*, ont été extraits d'une étude publiée en 2011 (Ramani et al., 2011). Par une combinaison de séquençage des ARNs et de microarrays, cette étude a permis de mettre en évidence des gènes dont l'épissage alternatif est régulé durant le développement. Ces deux gènes font partie des 2534 gènes récupérés plus haut. Ces deux gènes ont été choisis puisque l'on dispose pour ces gènes d'informations supérieures. Les premiers gènes choisis ne sont reconnus comme tel que par la détection de plusieurs isoformes par le biais du séquençage massif des ARN totaux extraits à partir de populations mixtes – embryons, larves et adultes. Ces événements d'épissage ne sont donc pas vérifiés et peuvent provenir d'une détection précise au point de visualiser des produits aberrants ou sous le contrôle d'aucune régulation active. Les gènes *top-1* et *hrpf-1* ont été plus étudiés et des régulateurs sont connus, ce qui permet de vérifier ces données grâce au système rapporteur et d'en découvrir d'autres.

## 2. Méthodologie

## a) Méthode d'assemblage

Nous voulons utiliser la cassette FSS pour construire des rapporteurs contenant : la séquence promotrice du gène étudié afin de permettre l'expression ; les exons et introns flanquant l'exon alternatif ; la cassette FSS – FrameShift Sensor ; la région 3'UTR du gène étudié ; le plasmide. En raison du nombre d'éléments à insérer et de la complexité des clonages, la méthode choisie pour les constructions est la méthode d'assemblage par “chew-back/annealing” (Gibson et al., 2009).

Cette méthode repose sur l'action de trois enzymes : l'exonucléase T5, la polymérase Phusion et la Taq ligase. Les séquences à cloner sont amplifiées par des oligonucléotides possédant chacun une queue flottante de 20 à 30 nucléotides. Ces queues sont homologues à l'extrémité de la séquence à assembler. Les amplifiats sont placés dans un même mélange réactionnel en présence des trois enzymes à 4°C avant d'être incubés une heure à 50°C. Ce changement de température permet l'activation temporaire de l'exonucléase – température optimale : 37°C – qui hydrolyse de 5' en 3', générant des séquences à bouts cohésifs. Une fois à 50°C et l'exonucléase inactivé, les fragments vont pouvoir s'assembler par leurs extrémités analogues. La phusion et la ligase vont combler les trous et lier les séquences. (Figure 18)



**Figure 18 : Représentation de la méthode d'assemblage utilisée pour les constructions.** Les fragments en rose et vert possèdent une séquence en commun (en noir). En passant de 4°C à 50°C, l'exonucléase T5 dégrade les séquences de 5' vers 3'. Les séquences homologues peuvent alors s'hybrider. A 50°C, l'exonucléase n'est plus active et la phusion et la ligase peuvent combler et lier le produit de la fusion. Extrait de (Gibson et al., 2009).

C'est grâce à cette méthode que tous les rapporteurs présentés ont pu être réalisés.

## **b) Les constructions**

Pour chaque gène étudié, un rapporteur contenant une partie de la séquence du gène couvrant l'événement d'épissage alternatif, les éléments de régulation du gène et la cassette « FrameShift Sensor » a été assemblé. La méthodologie utilisée pour suivre l'épissage par paire de « minigènes » n'inclut pas les promoteurs endogènes (Kuroyanagi et al., 2006). Nous avons fait le choix dans notre stratégie de les utiliser afin de s'approcher le plus possible des régulations endogènes et de prendre en compte l'impact que peut avoir le promoteur sur l'épissage (**Introduction.III.3.a**). Pour cela, des règles ont été établies. En l'absence de séquence clairement identifiée pour le promoteur du gène étudié, j'ai suivi les règles introduites par le projet du « Promoterome » (Dupuy et al., 2004) en récupérant deux kilobases en amont du codon initiateur pour servir de promoteur. Pour la région 3'UTR, un kilobase en aval du codon stop du gène avait été sélectionnée. Les assemblages nécessitaient donc quatre composantes : Promoteur, séquence – exons et introns -, cassette FSS et 3'UTR. Ces composantes se découpent en 3 à 4 séquences, en fonction de la distance entre le promoteur et l'événement d'épissage – amplifiés ensemble lorsque la distance était minimale.

De plus, le rapporteur nous permet de visualiser les deux isoformes à travers deux marqueurs fluorescents différents lorsqu'un changement de cadre de lecture est induit par la présence ou l'absence de l'exon. Dans la plupart des gènes étudiés, ce n'était pas le cas. Il a donc été nécessaire de provoquer une mutation dans la séquence différentiellement incluse afin de provoquer ce changement. Ces mutations ont été introduites lors du processus d'assemblage en amplifiant le gène non pas en une mais en deux parties, rajoutant la mutation sur les séquences d'homologies nécessaires à l'assemblage par la méthode de Gibson.

Cette première méthodologie – basée sur le protocole établi au laboratoire - a été mise en œuvre mais les résultats n'ont pas suivis après des dizaines de tentatives. Il a donc été nécessaire d'apporter des modifications vis-à-vis du protocole en place. Les constructions sont complexes et la taille des rapporteurs finaux – au-delà de 6kb – peuvent jouer comme facteur limitant. Il a donc été décidé de limiter le nombre de séquences à assembler. Pour simplifier les assemblages, la cassette FSS a été clonée dans un plasmide standard, le pBS et associée à un 3'UTR générique – discuté plus loin. Cette modification a permis de diminuer le nombre de fragments jusqu'à 2-3. Cependant, cette modification n'a pas suffi à obtenir des résultats positifs.

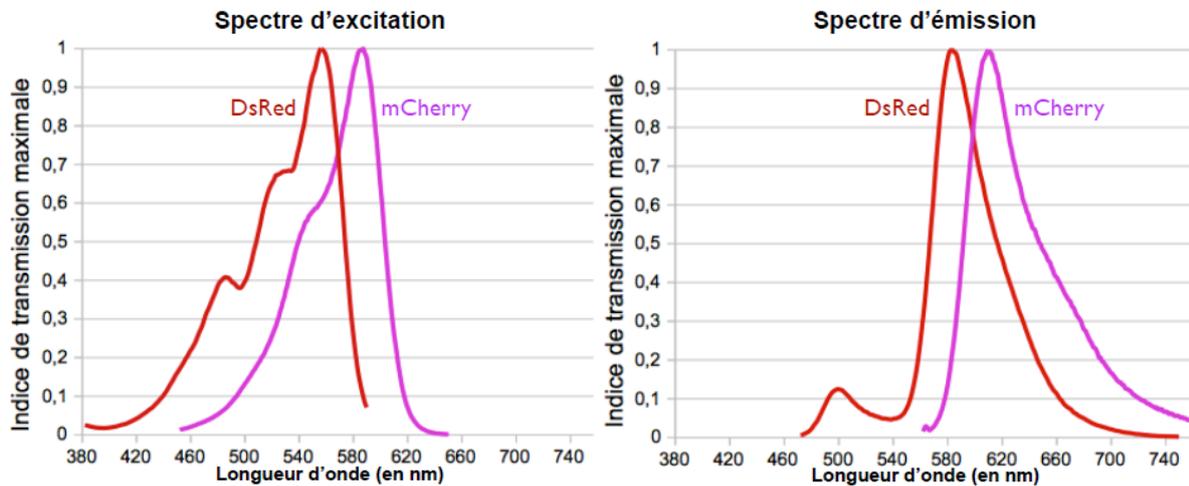
Le protocole final a été obtenu par le biais de modifications établis sur l'assemblage des rapporteurs *top-1*, *gpa-13* et *unc-86*. Plusieurs changements ont été effectués sur le protocole publié (Gibson et al., 2009). Parmi ces modifications, deux changements majeurs : la quantité des séquences lors de l'assemblage, établis à 25 Fmol et augmenté deux à trois fois ; le choix des séquences d'homologies – établis à 20nt au laboratoire – afin d'adapter la taille des séquences à des Tm proches de 55-60°C.

Le protocole finale est une combinaison de ces modifications et a depuis été utilisé au laboratoire pour six autres constructions et ont permis d'obtenir des résultats allant de 30 à 100% de clones positif après la première transformation. Le nombre de positif ne semble pas varier en fonction du nombre de séquences à assembler mais en fonction des séquences d'homologie utilisés.

La mise en place de cette méthode permet aujourd'hui d'assembler des rapporteurs simplement, en moins de deux semaines - de la construction théorique au résultat de séquençage.

### **c) La séquence du second fluorophore**

La seconde séquence du FSS a été modifiée. La séquence de la mCherry était placée après celle de la GFP, or, le laboratoire est équipé d'un appareil de cytométrie en flux adapté à l'analyse de nématodes (COPAS-Profiler Union Biometrica) calibré pour la visualisation de la DsRed. Les deux protéines, relativement proche en terme de propriétés, présente des différences au niveau des longueurs d'ondes d'absorption et d'émission. Nous avons donc remplacé la séquence codante de la mCherry par la séquence codante de la DsRed de façon à pouvoir utiliser notre cytomètre pour des analyses quantitatives à haut débit.



**Figure 19 :** Spectres comparés d'excitation et d'émission de la DsRed et de la mCherry. La courbe rouge correspond aux propriétés de la DsRed et en rose à celles de la mCherry.

La substitution a également eu un effet bénéfique sur notre efficacité de clonage. En effet, la séquence de la mCherry contenu dans le vecteur initial était précédée d'une séquence d'une dizaine de nucléotides identiques au début de la séquence de la GFP. La méthode d'assemblage de Gibson étant basée sur des homologies de séquences, il n'était pas rare de retrouver parmi les clones sélectionnés, des plasmides sans GFP, où la séquence du gène était directement associée à celle de la mCherry. Grâce au changement effectué, ce problème a été complètement éliminé.

#### **d) La région 3'UTR**

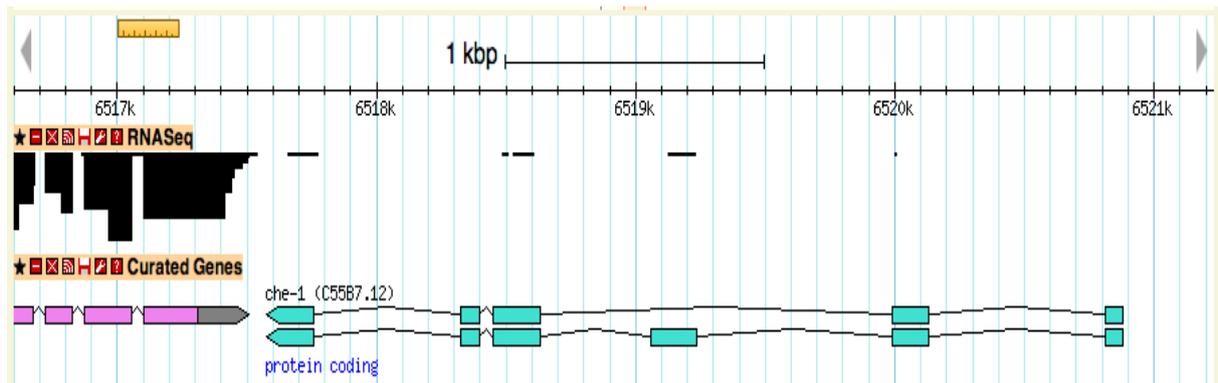
Afin de faciliter l'assemblage des constructions durant le projet, il a également été décidé de limiter le nombre d'éléments à associer, et ce, en accolant au FSS une région 3'UTR générique. La région 3'UTR d'un gène est cruciale à la stabilité de l'ARNm produit mais est également source de régulation - notamment par les microARNs -, il est nécessaire d'utiliser une séquence qui n'interférera pas avec la régulation que l'on souhaite observer. La région 3'UTR générique utilisée chez *Caenorhabditis elegans* est traditionnellement celle du gène *unc-54* - codant pour une protéine de la chaîne lourde de la myosine - considérée comme permissif dans tous les tissus, et ce, à tous les stades de développement depuis 1995 (Boulin et al., 2006). Cependant, des travaux conduits au sein de l'équipe ont démontré que cette séquence était soumise à régulation dans certains neurones (Cécile Quéré et Denis Dupuy - Travaux non publiés). En l'absence de région 3'UTR synthétique minimale ou de région démontrée comme permissive dans tous les tissus, la séquence du gène *unc-119* a été choisie. *unc-119* est un gène à l'expression pan-neuronale (Maduro and Pilgrim, 1995). Sa séquence

3'UTR ne connaît pour le moment aucun régulateur identifié et aucun microARN identifié à ce jour n'est prédit par l'algorithme de TargetScan comme potentiel régulateur (Lewis et al., 2003).

## II. Suivi des gènes *unc-86*, *che-1* et *gpa-13*

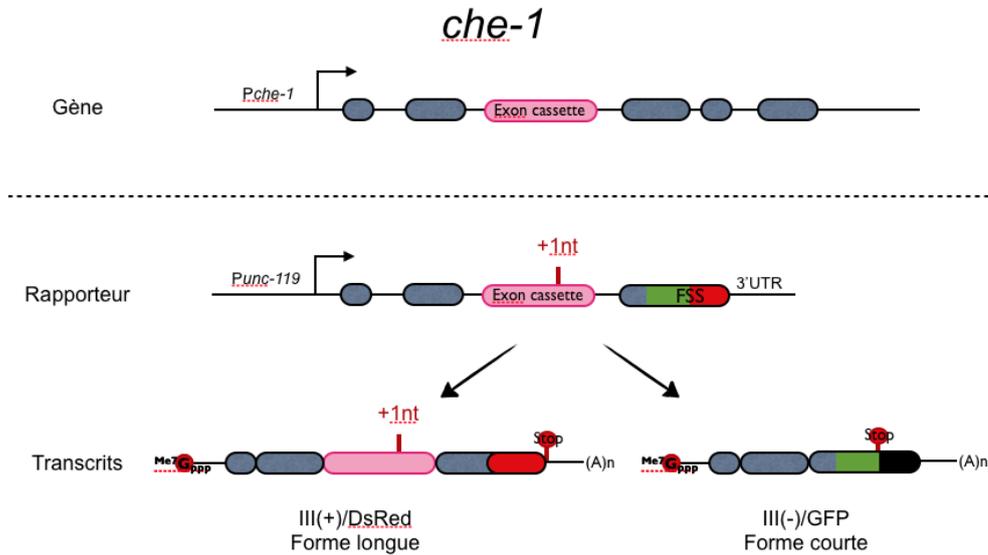
### 1. *che-1*

CHE-1 est un facteur de transcription à doigt de zinc responsable de la mise en place de l'identité cellulaire et de la fonction des neurones ASE, une paire de neurones chimio-senseurs (Uchida et al., 2003)(Chang et al., 2003)(O'Meara et al., 2009). Un rapporteur *che-1::GFP* est exprimé dans les neurones ASE (Sarin et al., 2009). Situé sur le chromosome I, le gène *che-1* est composé de cinq exons constitutifs et d'un exon alternatif – le III (Figure 20).



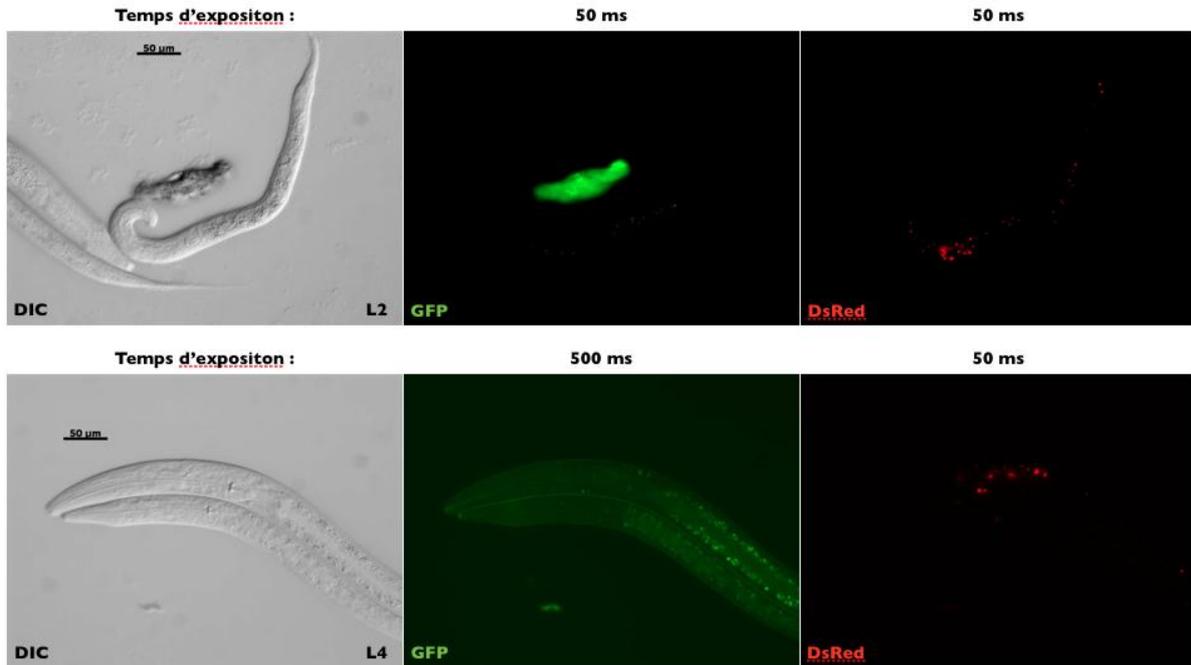
**Figure 20 : Modèle du gène *che-1*.** Les exons bleus représentent les deux isoformes du gène *che-1*. Les données de séquençage des ARNs disponibles sont représentées au-dessus en noir. Extrait de WormBase.

Les données de séquençage disponible sont largement insuffisantes pour s'assurer du modèle proposée (Figure 20). La construction contenant le promoteur et les quatre premiers exons n'a pas pu être assemblée, et ce, après plusieurs tentatives. Une construction alternative a été réalisée en prenant le promoteur du gène pan-neuronale *unc-119* puisque l'expression de *che-1* est limité aux neurones (Sarin et al., 2009)(Figure 21).



**Figure 21 : Représentation du gène *che-1* et de la construction réalisée.** Associé au promoteur du gène *unc-119*, la séquence du gène *che-1* va de l'ATG du premier exon jusqu'au 85ème nucléotide de l'exon IV. Un nucléotide supplémentaire a été inséré à la position 132 de l'exon alternatif. La forme longue – III(+) – est dans le cadre de lecture de la DsRed et la forme courte – III(-) – dans le cadre de la GFP.

Deux lignées transgéniques ont été isolées et présentent le même profil. L'expression est neuronale et nucléaire. L'isoforme longue – III(+)/DsRed – est fortement visible à la loupe à fluorescence. Cependant, la GFP est visible très faiblement et co-localise parfaitement avec la DsRed (Figure 22).



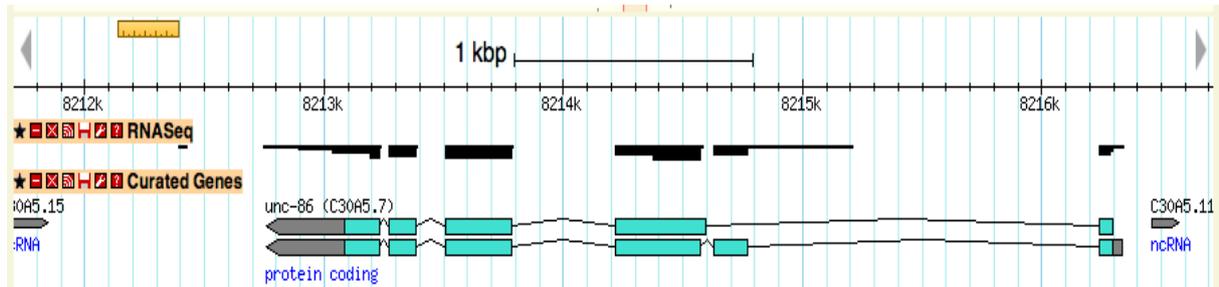
**Figure 22 : Micrographie d'individus porteurs du rapporteur *che-1::FSS*.** La première colonne correspond au canal GFP, la seconde correspond à la Dsred et la troisième à une superposition des deux. La barre noire située dans la colonne DIC représente 20µm. La forme longue – en ROUGE – est exprimé majoritairement mais la forme courte – en VERT - est légèrement visible et les deux formes co-localisent. (Le signal GFP localisé faible dans le corps des individus de la troisième ligne est un cas d'auto-fluorescence connu chez le nématode).

Le rapporteur ne permet pas de visualiser une différence spatiale ou temporelle dans l'expression des deux isoformes. L'épissage de l'exon III peut ne pas être contrôlé activement, les deux isoformes coexistant sans impact de régulateurs dans des conditions standard de culture. La construction peut manquer des éléments de régulations dans la séquence du gène. La transcription et les facteurs de transcription peuvent être couplés à l'épissage (**Introduction III.3**). Le promoteur utilisé pourrait donc affecter la régulation.

## 2. *unc-86*

UNC-86 est un facteur de transcription de type POU – qui tire son nom d'un domaine conservé par les protéines PIT-1, OCT-1, OCT-2 et UNC-86. Ce facteur est un régulateur majeur de la différenciation neuronale, impliqué notamment dans la croissance des dendrites des neurones sérotoninergiques, la différenciation des neurones mécano-senseurs et des inter-neurones chimio-senseurs (Horvitz and Sulston, 1980)(Sze et al., 2002)(Sze and Ruvkun, 2003). La protéine UNC-86 est détectée par immuno-histo-chimie dans les noyaux d'une cinquantaine de neurones (Finney and Ruvkun, 1990).

Situé sur le chromosome III, le gène *unc-86* est composé de cinq exons constitutifs et d'un alternatif. Le cas de ce gène est un peu particulier puisque deux ARNm ont été identifiés sur la base de données WormBase qui diffèrent par un exon cassette – le second exon – mais également par la taille du troisième exon. Lorsque le second exon est présent, le troisième exon est composé de 352 nucléotides. L'isoforme la plus courte manque le second exon et présente un exon III de 375 nucléotides.



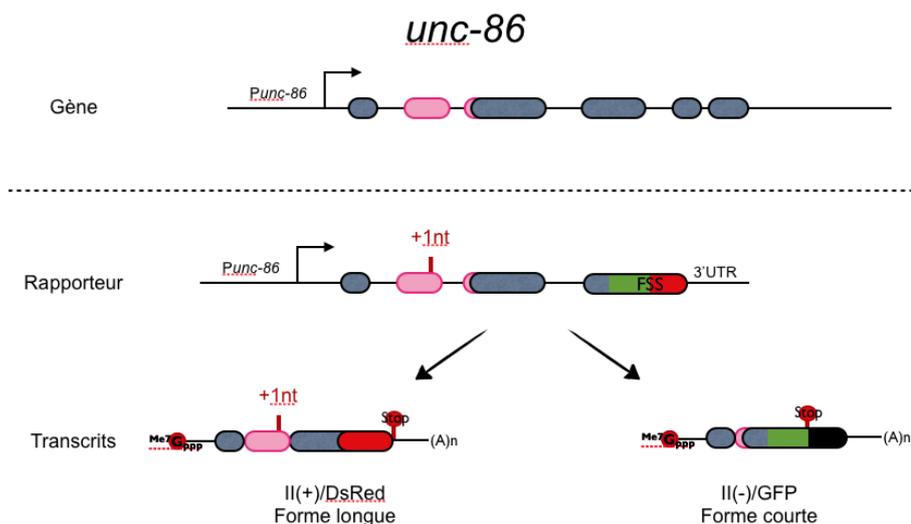
**Figure 23 : Modèle du gène *unc-86*.** Les exons bleus représentent les deux isoformes du gène *unc-86*. La forme incluant l'exon alternatif possède également un troisième exon plus court. Les données de séquençage des ARNs disponibles sont représentées au-dessus en noir. Extrait de WormBase.

L'exon cassette mesure 137 nucléotides et changerait le cadre de lecture s'il était intégré avec l'exon III le plus long. Il y a deux possibilités théoriques pour expliquer ce cas de figure :

(1) Il est possible que la présence de l'exon II et la taille de l'exon III soient deux événements d'épissage alternatif indépendants – CE et 5'ASS. Dans ce cas, les deux autres possibilités d'ARNm seraient dégradées par le NMD. En effet, les deux autres possibilités théoriques issues de l'épissage possèdent chacun un codon stop dans l'exon III – qu'il soit court ou long. La présence d'un STOP en amont d'introns, et donc de complexes EJC, permettrait la reconnaissance par la voie principale du NMD (**Introduction II.4.b**).

(2) Il est aussi possible qu'il ne s'agisse que d'un seul événement ; l'épissage du second exon affectant le site d'épissage suivant par le biais d'un couplage. En effet, l'intron situé entre l'exon II et III(long) mesure 36nt de long, ce qui pourrait constituer une limite physique dans le cas d'un épissage suivant le concept de l'« Intron Definition » - qui n'est pas la vision de référence mais n'a pas été réfutée définitivement - et provoquer une préférence pour le site d'épissage suivant (**Introduction III.4.a**).

Dans la construction réalisée, la version II(+)/III(court) permet la synthèse de la DsRed (isoforme longue du gène) et la version II(-)/III(long) permet la synthèse de la GFP (isoforme courte du gène) (Figure 24).

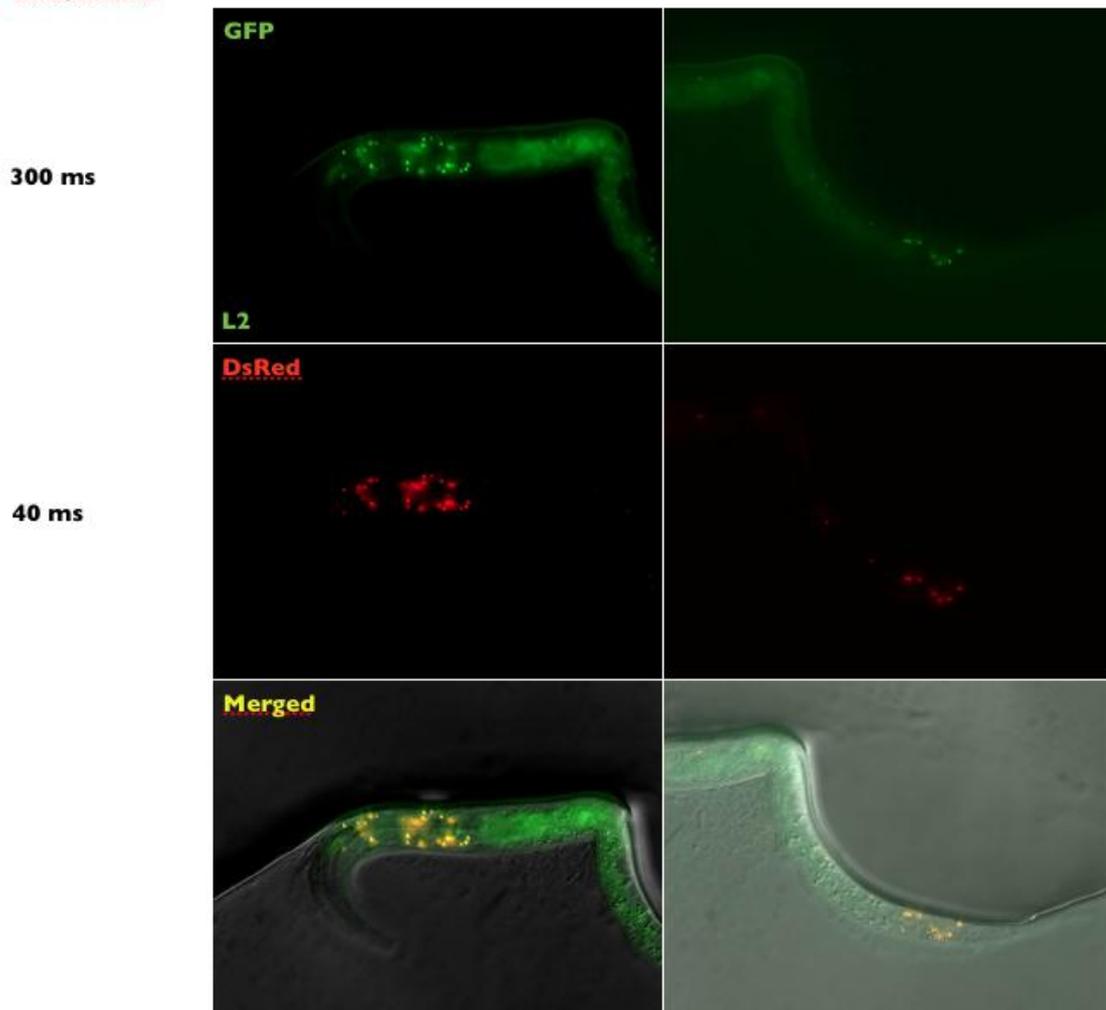


**Figure 24 : Représentation du gène *unc-86* et de la construction réalisée.** La construction assemblée couvre les 2kb en amont du codon START, jusqu'à la moitié de l'exon III (147nt version courte/160nt version longue). Un nucléotide a été rajouté à la position 131 de l'exon II pour décaler le cadre de lecture. La forme longue est associée à la séquence codante de la DsRed et la forme courte à celle de la GFP.

Cette construction permet donc de visualiser deux versions représentant les ARNm identifiés. Les deux autres possibilités théoriques ne sont associées à aucune ORF fluorescentes. La version II(-)/III(court) présente une ORF de 60nt de l'ATG au premier STOP, et la version II(+)/III(long) à une ORF de 312nt.

La construction générée a été injectée et a permis d'obtenir deux lignées fluorescentes aux profils identiques. L'expression est visible majoritairement dans les noyaux de cellules de la tête, proche de la base du pharynx, le long du corps de l'individu et dans la queue. Cette expression correspond à des cellules du système nerveux, ce qui coïncide avec les résultats de rapporteur GFP publié (Feng, Craig and Hope, 2012). Cette expression ne semble pas varier au cours du développement du premier stade larvaire au stade adulte. A la loupe à fluorescence, le seul signal le plus fort est celui de la DsRed qui correspond à la version II(+)/III(court) – l'isoforme la plus longue. Cependant, la seconde isoforme est faiblement visible et co-localise parfaitement avec celle de la DsRed (Figure 25).

Temps d'exposition :



**Figure 25 : Micrographie d'un individu porteur du rapporteur *unc-86::FSS*.** La deuxième ligne correspond au canal GFP, la troisième correspond à la DsRed et la quatrième à une superposition des canaux. La première colonne correspond à l'objectif 10x ; la barre noire représente 20µm. Les secondes et troisièmes colonnes utilisent un objectif plus élevée ; la barre noire représente 50µm. La forme longue – en ROUGE – est exprimé majoritairement mais la forme courte – en VERT - est légèrement visible et les deux formes co-localisent (Le temps d'exposition est dix fois supérieur dans le canal GFP). (Le signal diffus GFP situé dans le corps de l'individu est un cas d'auto-fluorescence connu chez le nématode).

Durant l'isolation d'individus transgéniques issus d'injection, ils ont également été observés à la loupe à fluorescence. Un individu unique présentant un profil de fluorescence GFP a pu être observé. Cet individu est mort durant le premier stade larvaire. Cet événement n'a jamais été à nouveau observé.

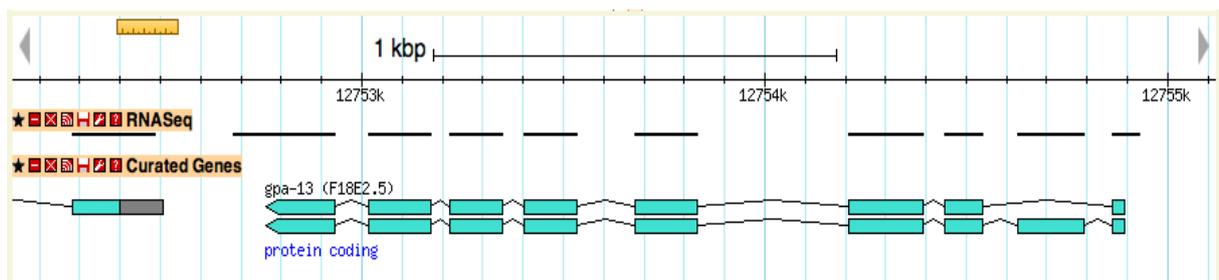
L'expression de cette construction ressemble à celle de la construction réalisée pour le gène *che-1* et pose donc les mêmes problèmes. La visualisation des deux fluorescences pourrait être le fruit d'une « fuite » du système engendrée par une fuite de transcription. Elle

pourrait également être le reflet d'une réalité plus complexe que ce qu'il n'y paraît, à savoir l'impact des sites d'épissages eux-mêmes sur l'efficacité d'inclusion d'un exon.

Des sites de fixation potentiels à des facteurs d'épissage comme SUP-12 ou PTB-1 ont été identifiés dans le premier intron. Cependant, l'analyse de ce gène n'a pas été poussée plus en avant durant cette étude en raison de son profil d'épissage. La présence massive d'une isoforme par rapport à l'autre ainsi que la co-localisation rend difficile les cribles génétiques et l'analyse de mutants.

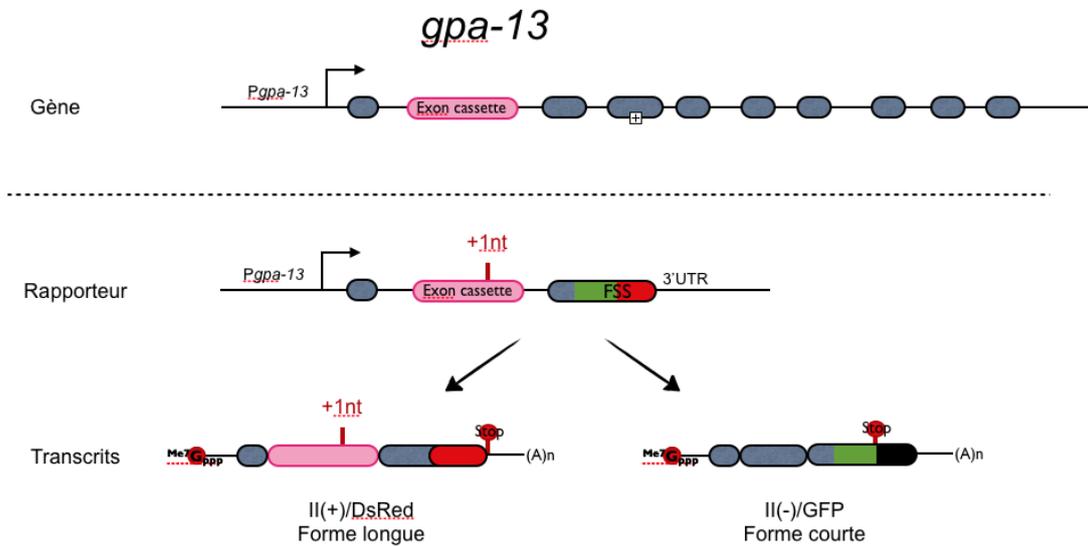
### 3. *gpa-13*

GPA-13 est une GTPase exprimée dans les neurones sensoriels du nématode (Jansen et al., 1999). Situé sur le chromosome V, le gène *gpa-13* est constitué de 8 exons constitutifs et d'un exon alternatif – le II. Le gène est référencé avec deux isoformes bien que les données de séquençage ARN disponibles sont limitées par la faiblesse de l'expression du gène.



**Figure 26 : Modèle du gène *gpa-13*.** Les exons bleus représentent les deux isoformes du gène *gpa-13*. Les données de séquençage des ARNs disponibles sont représentées au-dessus en noir. Extrait de WormBase.

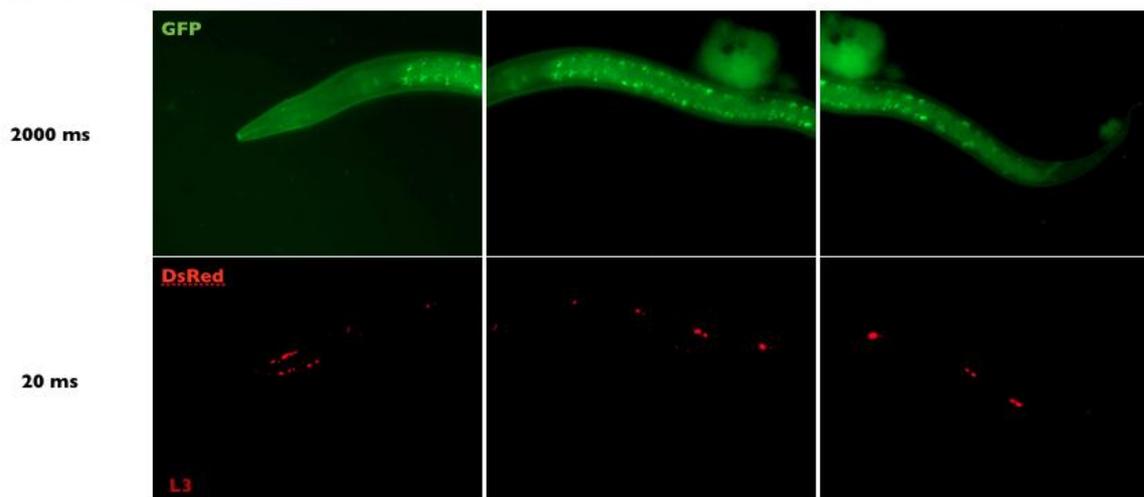
La construction réalisée est constituée de telle manière que la forme courte – II(-) – est dans le cadre de lecture de la GFP, et la forme longue – II(+) – est couplée à la DsRed (Figure 27).



**Figure 27 : Représentation du gène *gpa-13* et de la construction réalisée.** La construction réalisée comprend le promoteur, et les trois premiers exons. Un nucléotide supplémentaire a été inséré à la position 132 de l'exon II. La forme courte – II(-) – est dans le cadre de lecture de la GFP, et la forme longue – II(+)- est couplée à la DsRed.

À la suite de plusieurs séries d'injection, trois lignées transgéniques indépendantes ont été obtenues. Les individus observés de ces trois lignées présentaient des profils identiques. L'expression ne semble pas varier au cours du temps et se limite à des cellules de la tête et de la queue. Seul la DsRed est visible, et ce, même à très forte exposition du signal GFP (Figure 28).

**Temps d'exposition :**



**Figure 28 : Micrographie d'un individu porteur du rapporteur *gpa-13::FSS*.** La première ligne correspond au canal GFP, la deuxième ligne correspond à la DsRed et la troisième à une superposition des deux. La forme longue – en ROUGE – est exprimée le long du corps ainsi que dans la tête. Le ver

est visible dans le canal GFP en raison de la surexposition ; l'amas visible proche du ver est issu du tapis bactérien des milieux de culture (Le temps d'exposition est cent fois supérieur dans le canal GFP comparé à celui de la DsRed).

Cette construction, contrairement aux deux précédentes, ne présente pas de signal GFP. Seule l'isoforme longue est visible. L'absence totale de GFP visualisée grâce à cette construction permet d'éliminer l'hypothèse de la « fuite » des constructions précédentes. Ce résultat conforte l'idée que les deux formes visualisées grâce aux rapporteurs de *unc-86* et *che-1* sont issues d'un épissage rare.

La présence d'une seule isoforme visible grâce à notre rapporteur pourrait remettre en question le modèle du gène. La visualisation des deux isoformes par séquençage des ARNs peut être liée à l'expression faible du gène (Figure 26).

#### 4. Conclusion

Les constructions réalisées pour ces trois gènes n'ont pas permis de mettre en évidence une régulation spatio-temporelle visible par le biais du rapporteur. Si le système peut être remis en question, il se peut également que l'épissage alternatif ne soit pas un mécanisme aussi tissu-spécifique que ce qui a été suggéré. Dans un article publié en 2013, l'équipe du Dr Ian Hope a appliqué le système de rapporteurs GFP « classiques » (**Introduction IV.5.a**) à 29 gènes présentant plusieurs isoformes et encodant des facteurs de transcriptions. La majorité des rapporteurs construits ne présentaient aucune différence dans l'expression des transcrits alternatifs (Craig et al., 2013).

L'épissage des rapporteurs *unc-86* et *che-1* est différent de celui du rapporteur *gpa-13*. En effet, dans ces deux premiers cas la forme longue est majoritaire mais la forme courte est néanmoins présente, ce qui n'est pas le cas du troisième. Cette absence peut être due à l'absence dans notre rapporteur des éléments des régulations ou à la surexpression du rapporteur. Les données obtenues par séquençage des ARNs ont également pu mettre en avant un artefact, l'expression du gène *gpa-13* étant très faible.

Pour *che-1* et *unc-86*, la présence aussi faible de la forme courte pourrait mettre en évidence la « force » relative des sites d'épissages, permettant la production de différentes isoformes sans l'action de facteurs extérieurs au spliceosome. Une question se pose alors : la « force » de ces sites est-elle une aberration d'épissage sans conséquence pour l'organisme, ou bien un mécanisme de régulation en lui-même ?

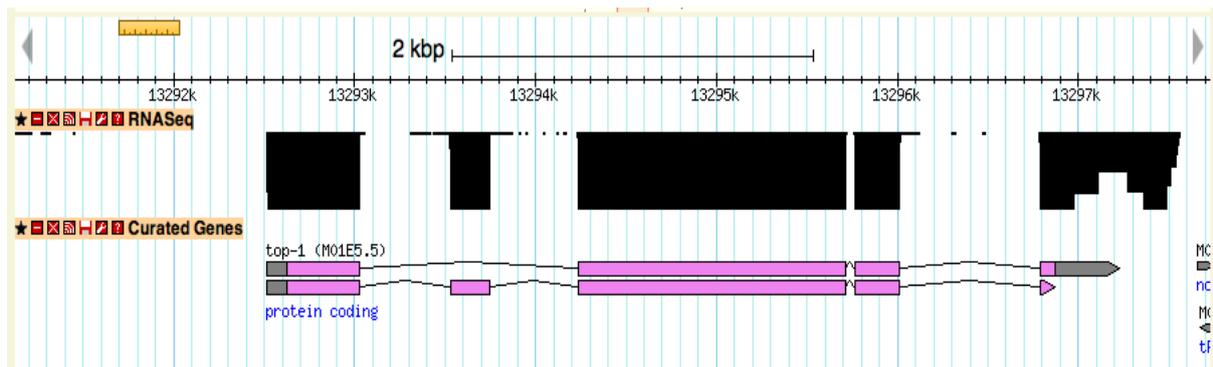
Pour répondre à cette question, il serait d'abord nécessaire d'étudier l'impact des sites d'épissage sur la production de plusieurs isoformes.

### III. Le cas des gènes *top-1* et *hrpf-1*

#### 1. Le rapporteur *top-1*

Le gène *top-1* encode la protéine TOP-1, une topo-isomérase de type I. Cette protéine est nécessaire au développement embryonnaire et à la prolifération cellulaire (Lee et al., 1993)(Morham et al., 1996). Les topo-isomérases de type I se fixent à l'ADN, provoquent des coupures de la chaîne nucléotidique et changent la structure topologique de l'ADN en catalysant le passage des segments d'ADN à travers ces coupures. Elles permettent notamment d'ajouter et d'enlever des super-tours dans les molécules d'ADN (Wang, 2002). La protéine TOP-1 humaine, dont l'homologie avec la protéine de *C.elegans* est de 95,5%, affecte la transcription de plusieurs gènes, parfois en tant que répresseur et parfois en tant qu'activateur - dépendant du promoteur (Merino et al., 1993).

Situé sur le chromosome I, le gène *top-1* est composé de quatre exons constitutifs et d'un exon alternatif – le II (Figure 29).

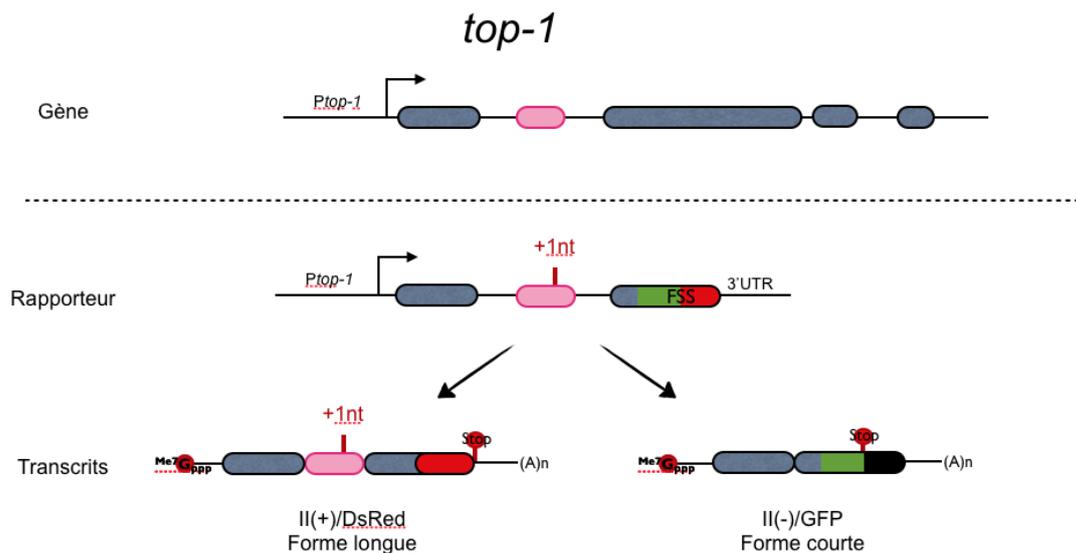


**Figure 29 : Modèle du gène *top-1*.** Les exons roses représentent les deux isoformes du gène *top-1*. Les données de séquençage des ARNs disponibles sont représentées au-dessus en noir. Extrait de WormBase.

Les résultats de séquençage ARN disponibles n'indiquent pas la présence de la forme courte (Figure 29). L'épissage alternatif étant probablement masqué par la régulation de celui-ci durant le développement. Le séquençage ARN des différents stades de développement et l'utilisation de micro-arrays ont démontré que les deux isoformes du gène sont régulées au cours du développement (Lee et al., 1998)(Barberan-Soler and Zahler, 2008)(Ramani et al., 2011). Une diminution de l'expression du gène *hrpf-1* par ARN interférence a pu démontrer que le facteur HRPF-1 est un régulateur de l'épissage alternatif du pré-messager de *top-1* (Barberan-Soler and Zahler, 2008). Par immunohistochimie, l'utilisation d'anticorps isoforme-spécifique a démontré que la forme courte de la protéine est présente dans les neurones tandis que la forme longue est présente dans de nombreuses cellules de l'organisme

(Cha et al., 2012). Un rapporteur GFP permet l'expression du fluorochrome dans les cellules du système nerveux seulement et ne se recoupe pas complètement avec la présence de la forme longue de la protéine visualisé (Cha et al., 2012). Cette différence peut s'expliquer par : des éléments de régulations insuffisants pour mimer l'expression endogène dans le rapporteur GFP ; l'absence d'expression du rapporteur durant l'embryogenèse en raison d'une répression de l'expression des transgènes dans les cellules de la lignée germinale propre à *Caenorhabditis elegans* ; la présence d'artefacts ou de reconnaissance peu spécifique des anticorps utilisées.

La construction réalisée est présentée dans la Figure 30.



**Figure 30 : Représentation du gène *top-1* et de la construction réalisée.** La construction réalisée utilise un promoteur de 2kb ainsi que les trois premiers exons. La fusion avec la cassette se situe après le nucléotide 21 de l'exon III. Un nucléotide additionnel a été ajouté à la position 213 de l'exon II. La forme longue de l'ARNm est couplée à la DsRed – II(+)/DsRed - et la forme courte à la GFP – II(-)/GFP.

Un premier mélange d'injection a été réalisé pour permettre la formation d'un concatémère répétitif avec la construction *top-1* à une concentration de 90ng/μL (**Introduction.IV.4.a**). Après de nombreuses séries d'injection, aucune lignée n'a pu être obtenue.

Un second mélange d'injection a été réalisé pour permettre la formation d'un concatémère « complexe » avec la construction *top-1* à une concentration de 10ng/μL (**Introduction.IV.4.a**). Après de nombreuses séries d'injection, une seule lignée a pu être obtenue. En plus des difficultés d'obtenir des individus transgéniques, il se trouve que la lignée elle-même est instable. L'utilisation du G418 comme moyen de sélection n'est pas

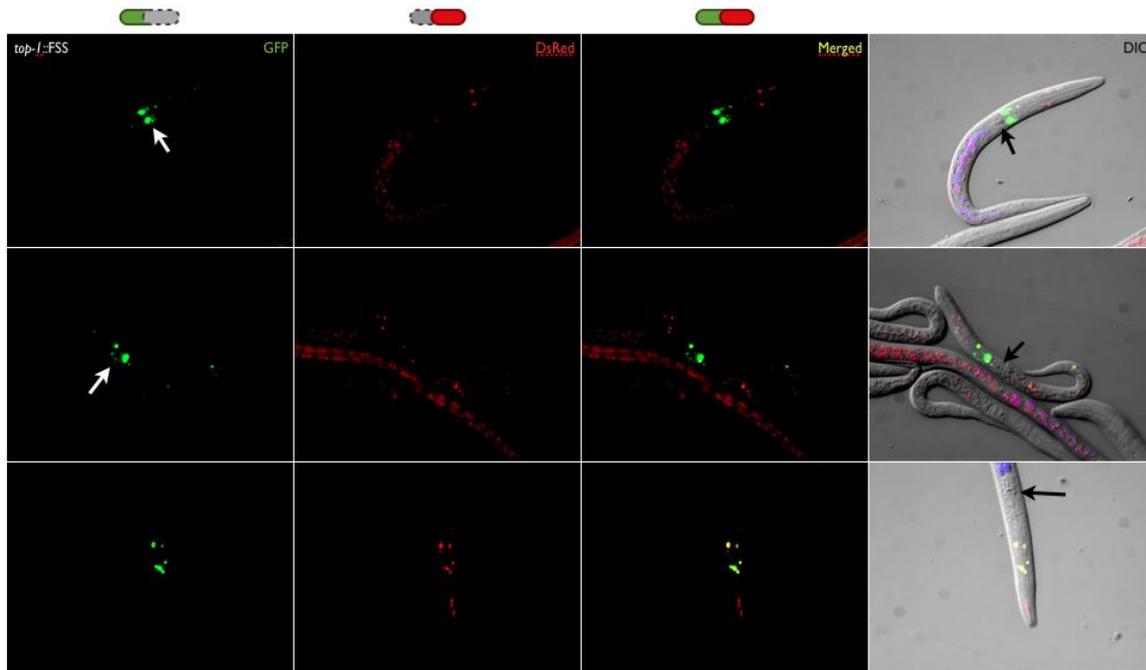
suffisant. La lignée isolée peut se maintenir durant quelques repiquages, mais ne peut être maintenu à long terme que par une sélection manuelle supplémentaire. De nombreux individus fluorescents ne se développent pas jusqu'au stade adulte. La construction et son expression en multiples copies nuisent aux individus qui les portent. Un troisième mélange d'injection a été réalisé pour permettre la formation d'un concatémère « complexe » avec la construction *top-1* à une concentration de 1ng/μL mais n'a donné aucune amélioration. Les observations présentées ci-dessous se sont donc portées sur cette unique lignée.

### **a) Au cours du développement**

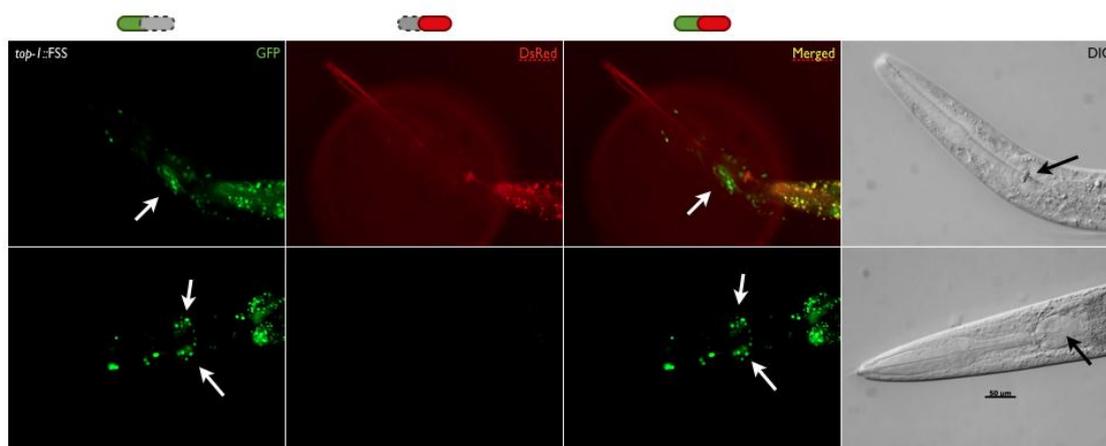
L'expression du rapporteur se limite à des neurones situés dans la région la tête et de la queue de l'organisme. Cette expression correspond à celle observée avec un rapporteur GFP contrôlé par le même promoteur (Cha et al., 2012).

Le rapporteur est exprimé chez les larves et les adultes. Malgré quelques variations, l'expression de la forme courte II(-)/GFP est maintenue dans des cellules de la tête situées au niveau de l'anneau nerveux – qui contient la majorité des *soma* des neurones de la tête (Figure 31 ; Figure 32). Les variations adviennent majoritairement sur la quantité de ces cellules fluorescentes. Cette forme est également visualisée dans des cellules de la queue et dans deux cellules proches de la vulve qui, aux vues de leurs positions et de l'expression du rapporteur GFP, correspondent aux neurones HSN – Hermaphrodite-Specific Neuron.

La forme longue II(+)/DsRed est présente durant les premiers stades larvaires, dans des cellules à proximité du premier bulbe du pharynx – qui pourraient correspondre aux neurones Ils identifiés précédemment (Cha et al., 2012). On retrouve cette forme co-exprimée avec la forme courte II(-)/GFP dans quelques cellules de la tête et de la queue chez les adultes (Figure 31 ; Figure 32).



**Figure 31 : Micrographie d'individus porteurs du rapporteur *top-1::FSS* aux stades larvaires.** La première colonne correspond au canal GFP, la deuxième colonne correspond à la DsRed et la troisième à une superposition des deux. La forme longue – en ROUGE – et la forme courte – en VERT – sont exprimées dans des cellules différentes de la tête chez les jeunes larves et se co-localise partiellement à des stades plus tardifs. Les flèches blanches représentent la position de l'anneau nerveux. Les flèches noires représentent la position de la base du pharynx. La forme courte est exprimée fortement dans l'anneau nerveux dans les premiers stades larvaires et disparaît dans les stades larvaires suivants.



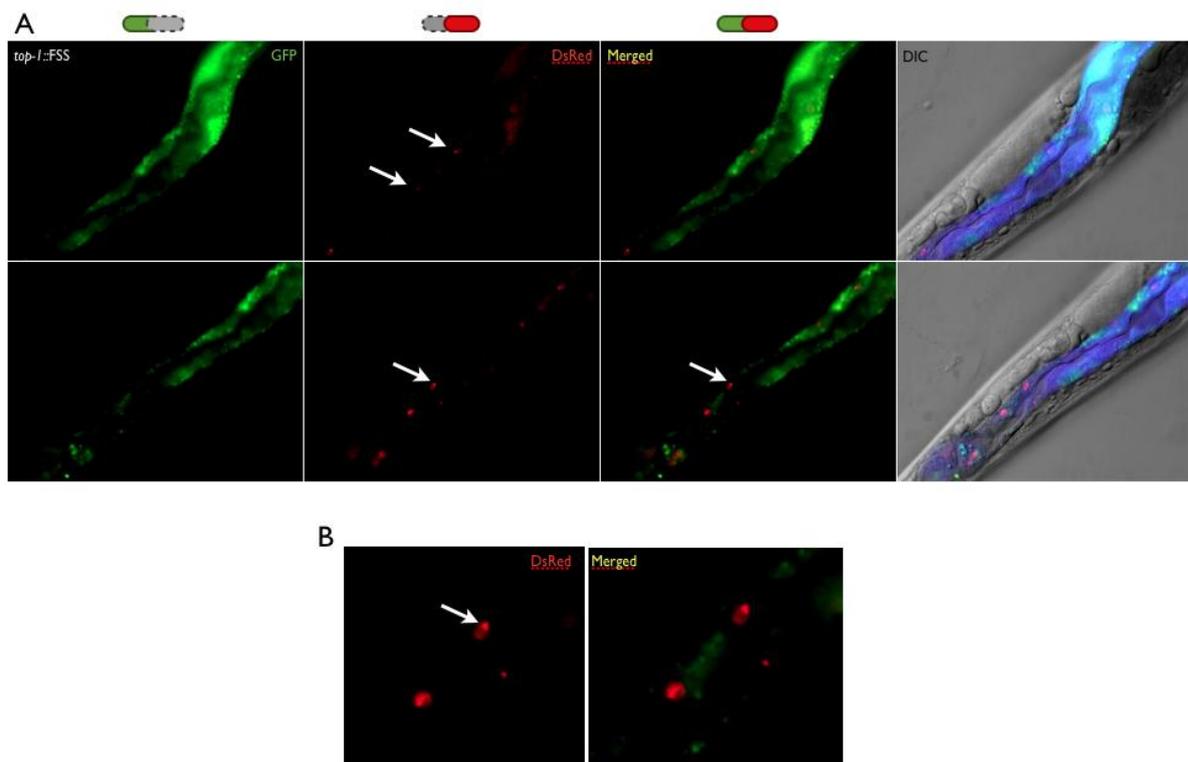
**Figure 32 : Micrographie d'individus porteurs du rapporteur *top-1::FSS* aux stades adultes.** La première colonne correspond au canal GFP, la deuxième colonne correspond à la DsRed et la troisième à une superposition des deux. Les flèches blanches représentent la position de l'anneau nerveux. Les flèches noires représentent la position de la base du pharynx. La forme courte – en VERT – est exprimée fortement dans l'anneau nerveux chez les adultes. La forme longue est très peu présente chez les adultes et co-localise souvent avec la forme longue.

Cette lignée a permis de visualiser pour la première fois au laboratoire une différence de localisation entre deux isoformes en utilisant ma version de la cassette FSS modifiée. Cette régulation varie au cours du temps. Cependant, l'absence d'autres lignées, la relative instabilité de celle-ci – qui limite fortement le nombre d'individus visualisables – et la stochasticité des transgènes exprimés en copies multiples, rendent difficile l'identification formelle des cellules en question et l'utilisation de cette lignée pour l'identification des déterminants moléculaires de cette régulation.

Malgré cela, la lignée a permis de réaliser des observations intéressantes.

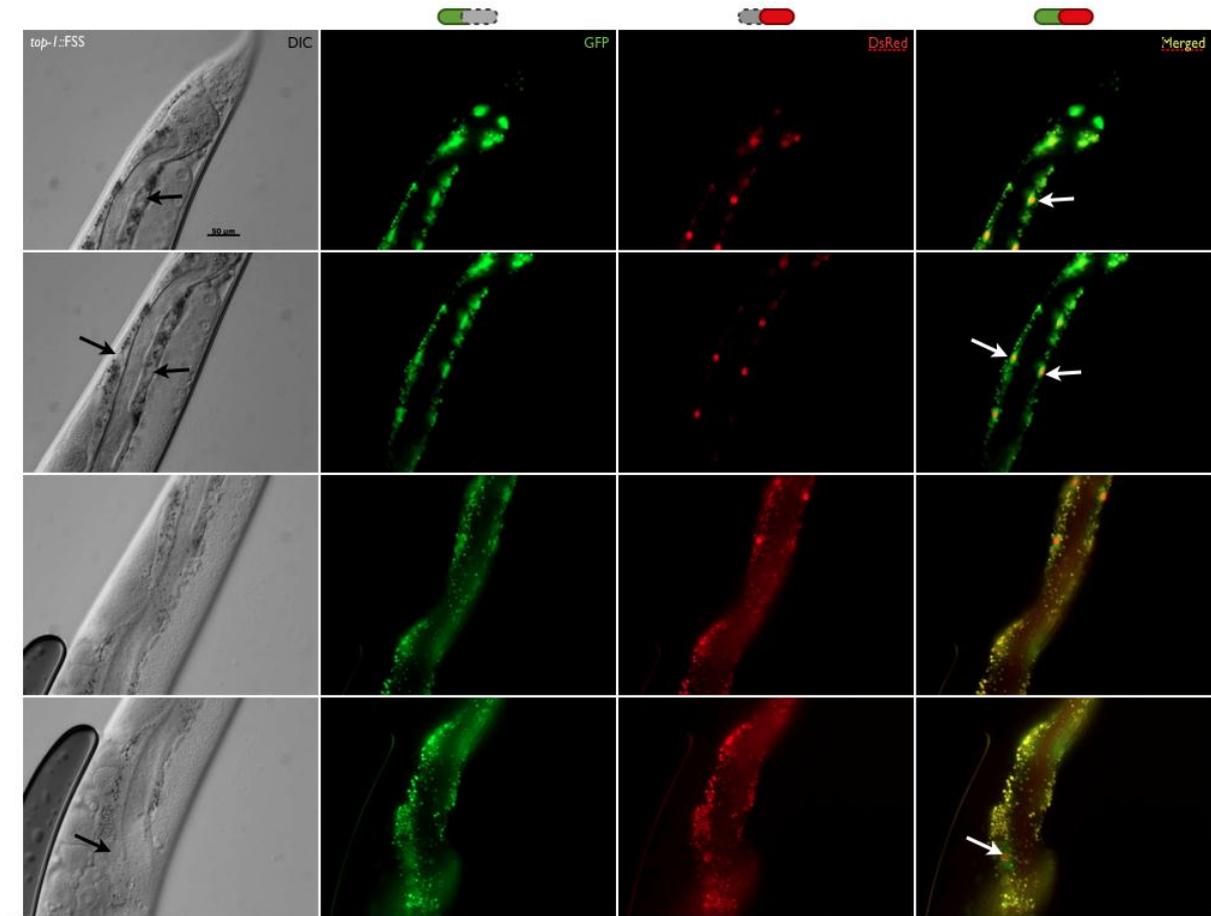
### **b) Expression dans les noyaux des cellules intestinales**

Les individus adultes présentent une expression dans les intestins qui semblent varier avec l'âge. La forme longue II(+)/DsRed se retrouve exprimée dans les noyaux des cellules intestinales d'individus adultes de la section postérieure de l'intestin (Figure 33).



**Figure 33 : Micrographie d'individus porteurs du rapporteur *top-1::FSS* aux stades adultes. (A)** Corps d'un adulte. La première colonne correspond au canal GFP, la deuxième colonne correspond à la DsRed et la troisième à une superposition des deux. La forme longue – en **ROUGE** – est présente dans des noyaux des cellules intestinales. Les flèches blanches désignent la position de signal de DsRed. **(B)** Agrandissement des micrographies présentées en (A). La flèche blanche désigne un sous-localisation du signal de la DsRed. Le signal GFP diffus est une caractéristique de l'intestin du nématode.

La forme courte II(-)/GFP apparait avec la forme longue chez des individus plus âgés (Figure 34).

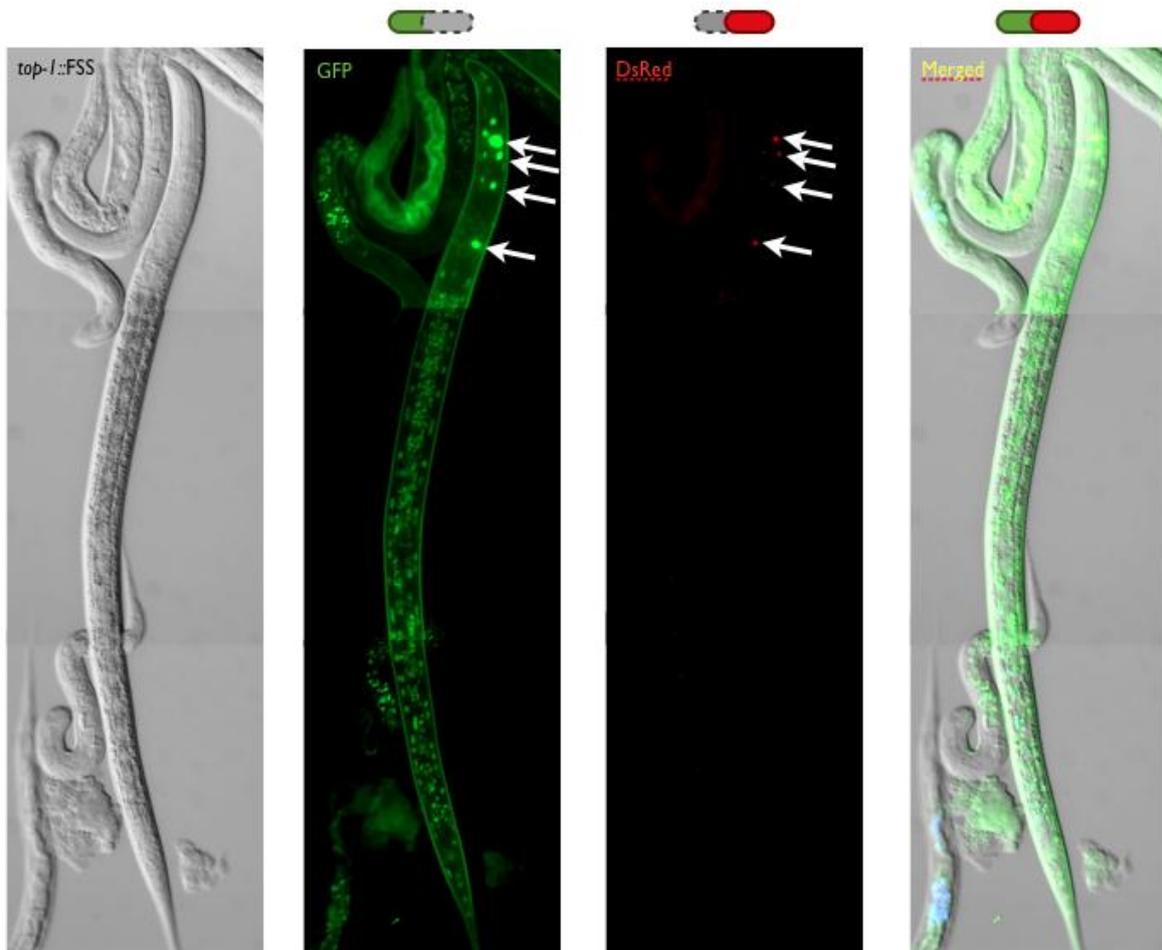


**Figure 34 : Micrographie d'individus porteurs du rapporteur *top-1::FSS* aux stades adultes.** La seconde colonne correspond au canal GFP, la troisième colonne correspond à la DsRed et la quatrième à une superposition des deux. La forme longue – en ROUGE – et la forme courte – en VERT – est présente dans des noyaux des cellules intestinales. Les flèches noires et blanches désignent la position de noyaux de cellules intestinales. Les signaux fluorescents sont plus forts à l'extrémité postérieure.

Cette expression des deux formes semble ne toucher que les individus âgés et suivre un gradient antéro-postérieur.

### **c) Les individus *dauer***

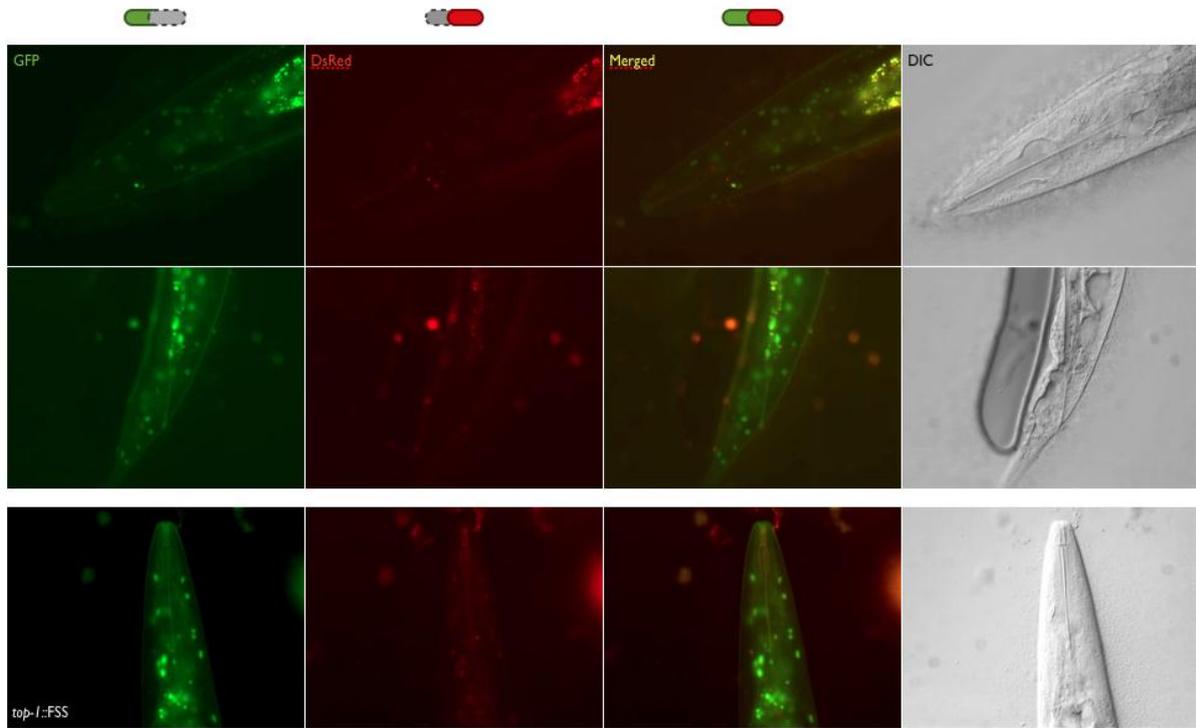
Cette lignée a également permis de mettre en évidence une autre particularité de l'expression du rapporteur. En cas de stress de « crowding » (surpopulation des vers), des individus *dauer* exprimant le rapporteur ont pu être observés. Ces individus présentent une expression très forte de la GFP qui ne se cantonne plus à quelques cellules (Figure 35).



**Figure 35 : Micrographie d'un individu de stade *dauer* porteurs du rapporteur *top-1::FSS*.** La seconde colonne correspond au canal GFP, la troisième colonne correspond à la DsRed et la quatrième à une superposition des deux. La forme longue – en **ROUGE** – est désignée par les flèches blanches, et co-localise avec la forme courte – en **VERT**. La forme courte est également exprimée le long du corps, longeant l'auto-fluorescence intestinale sous la forme de cercles qui pourraient correspondre à des noyaux de cellules musculaires. La GFP est également présente faiblement dans de nombreuses cellules de la tête.

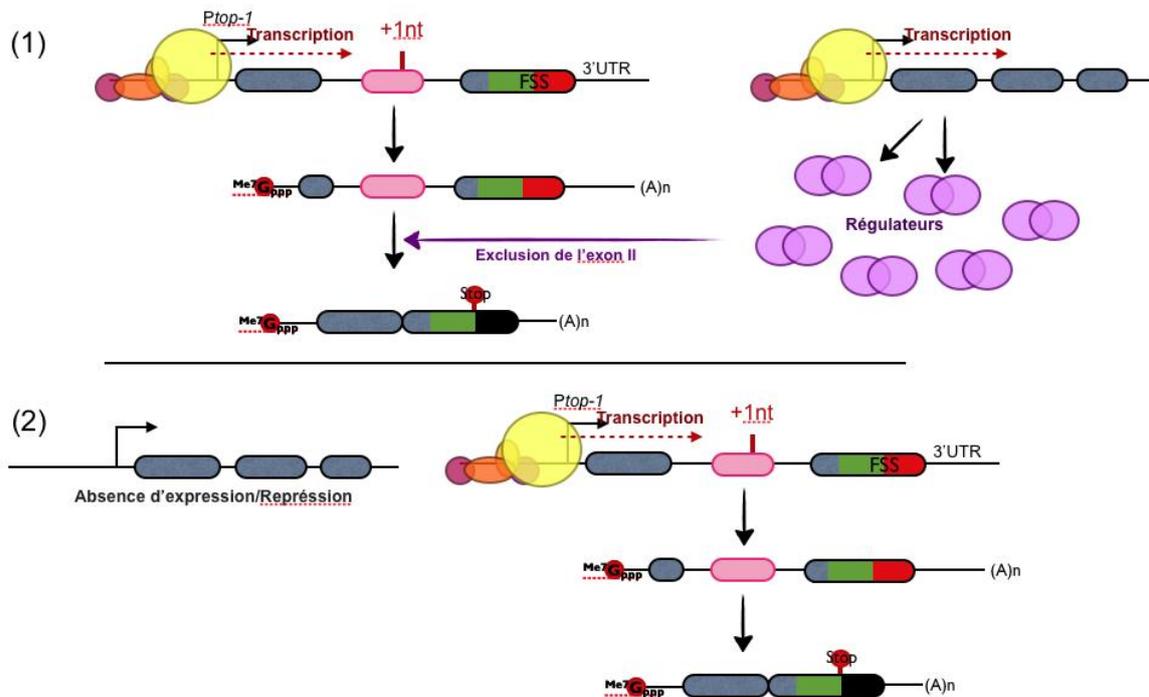
Chez ces individus, la forme longue II(+)/DsRed est toujours visible mais se limite à quelques cellules de la tête, ce qui correspond à l'expression de cette forme dans les individus du premier stade larvaire – le stade *dauer* étant un stade larvaire L2 alternatif.

L'expression de la GFP est également retrouvée plus faiblement chez des adultes soumis aux mêmes conditions (Figure 36).



**Figure 36 : Micrographie d'individus porteurs du rapporteur *top-1::FSS* aux stades adultes et en conditions de surpopulation.** La première colonne correspond au canal GFP, la seconde colonne correspond à la DsRed et la troisième à une superposition des deux. Les deux premières lignes sont issues du même organisme. La forme courte – en **VERT** - est présente dans des noyaux des cellules du corps en conditions de stress. La forme longue – en **ROUGE** - est également faiblement présente.

Cette expression est liée à un changement d'expression du rapporteur et n'est donc pas le fruit d'un changement dans la régulation de l'épissage alternatif. Cependant, la présence d'une seule isoforme – dans le cas des *dauer* - est intéressante puisque l'épissage est dans ce cas unanime. Deux possibilités simples sont à envisager : (1) Une régulation forte favorisant l'exclusion de l'exon II ou (2) l'absence totale de régulation favorisant l'inclusion de l'exon II (Figure 37).

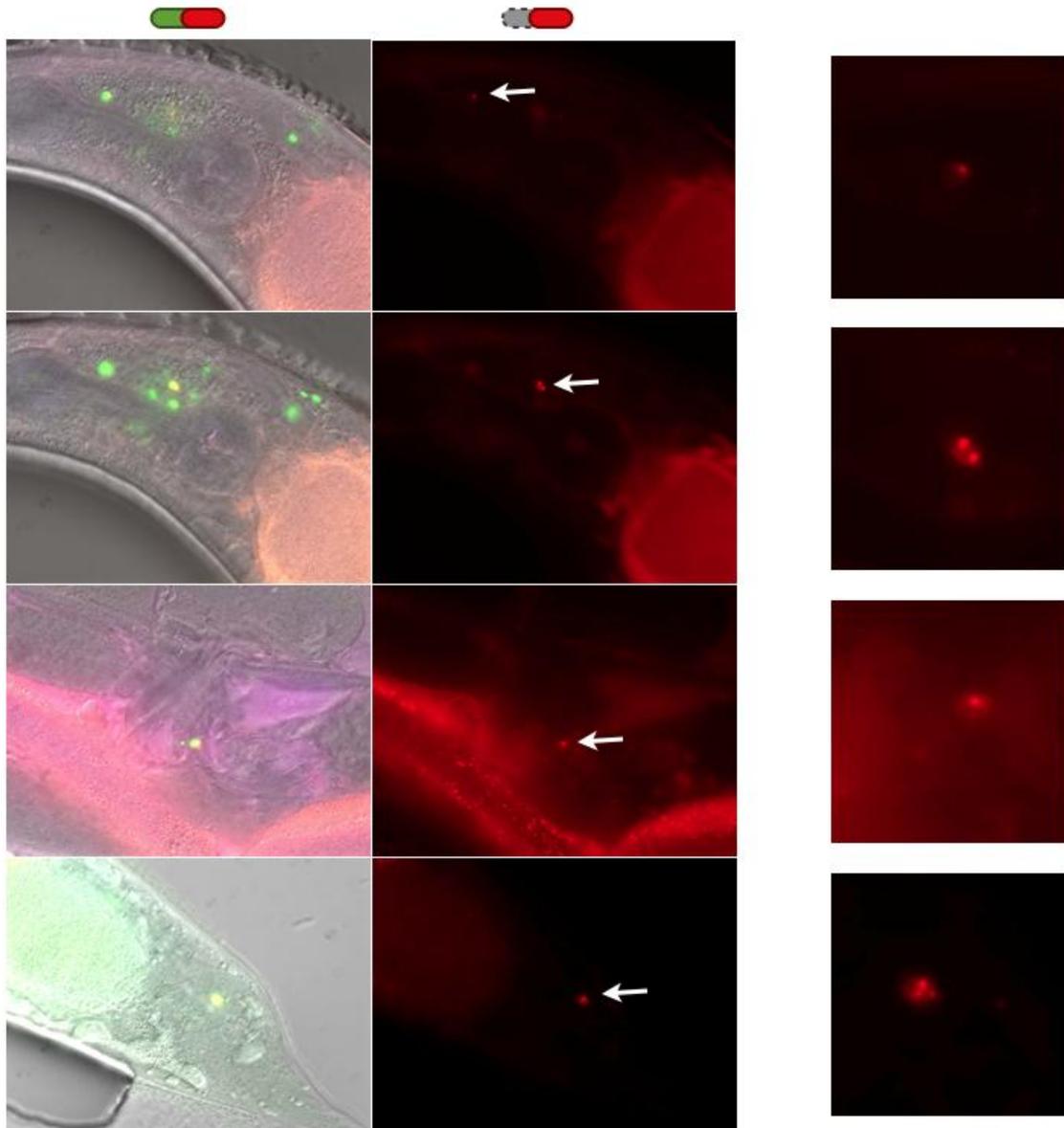


**Figure 37 : Représentation des hypothèses relatives à l'origine de l'expression de la forme courte dans des conditions de stress. (1)** Une régulation forte favorisant l'exclusion de l'exon II. Le rapporteur et un régulateur favorisant l'exclusion de l'exon alternatif – en VIOLET - sont exprimés. La présence de ce régulateur rend l'épissage unanime ; **(2)** l'absence totale de régulation favorisant l'inclusion de l'exon II. Dans ce cas de figure, le ou les régulateurs de l'épissage de *top-1* favorisant l'inclusion de l'exon alternatif ne sont pas exprimés dans les cellules où le rapporteur est visualisé comme exprimé dans les conditions de stress.

En l'absence de détails sur les éléments impliqués dans la régulation de *top-1*, il est impossible de trancher entre les deux hypothèses.

#### **d) Localisation nucléaire de la forme longue**

La forme longue – II(+)/DsRed - a pu être visualisée concentrée dans des *foci* nucléaires chez plusieurs individus différents, mais également dans des cellules différentes. Cette sous-localisation particulière semble aléatoire, sans être particulière d'un stade de développement, d'une condition de culture ou même de cellules spécifiques. La majorité des individus présentaient une DsRed nucléaire similaire à la GFP. Cette sous-localisation n'a jamais été observée avec la forme courte. Ces observations ont été réalisées à l'objectif 63x à immersion (Figure 38).



**Figure 38 : Micrographie d'un individu adulte porteur du rapporteur *top-1::FSS*.** Les deux premières lignes correspondent à deux plans différents de la tête de l'animal. La troisième au niveau de la vulve et la dernière dans la queue. La localisation de la DsRed est pointée par les flèches blanches. La troisième colonne consiste en un agrandissement de la cellule désignée par les flèches blanches. La forme longue est localisée dans un à plusieurs *foci* nucléaires.

Cette localisation a également pu être observée dans les noyaux des cellules intestinales (**Résultats.III.1.B** ; Figure 33).

La protéine TOP-1 est une topo-isomérase qui se fixe sur l'ADN. Cette sous-localisation de la forme longue issue de notre construction pourrait donc être le fruit d'une activité partielle de la protéine hybride générée, s'accumulant sur certaines séquences. La létalité du rapporteur pourrait donc être liée à un effet « dominant négatif » d'une protéine capable de se fixer à l'ADN comme TOP-1 sans déployer l'activité complète de ce facteur.

## e) Conclusion

Le rapporteur utilisé n'a pas permis de générer de lignée suffisamment stable pour permettre une analyse génétique de la régulation de l'épissage de *top-1*. Cet effet peut être la conséquence d'un effet délétère de la construction et de sa surexpression au sein des individus. Cet effet nocif peut prendre sa source de plusieurs manières. La surreprésentation de séquences régulatrices provoquant un phénotype en titrant des facteurs a été mis en évidence chez *Caenorhabditis elegans* avec l'exemple du promoteur de *hlh-17* (Stout Jr et al., 2013). Cependant, la titration de facteurs vitaux par la surreprésentation de la séquence promotrice est peu probable en raison du rapporteur GFP déjà exprimé, apparemment sans problèmes majeurs (Cha et al., 2012). La nocivité de la construction peut également être due à un effet « dominant négatif » qui pourrait concerner les deux isoformes, mais plus particulièrement la forme longue – II(+)/DsRed – en raison de sa localisation particulière. Les concatémères affectent parfois l'expression des gènes endogènes d'une manière similaire à l'ARN interférence (Dernburg et al., 2000). Cette hypothèse est d'autant plus valable puisqu'une diminution de l'expression du gène *top-1* par ARNi est létal pour les embryons et pour les larves (Fernandez et al., 2005)(Simmer et al., 2003).

Malgré ces problèmes, la lignée a permis de mettre en avant plusieurs choses. La première est de pouvoir visualiser une différence dans la localisation cellulaire des deux isoformes ce qui confirme la capacité de notre stratégie à rendre visible des événements d'épissage alternatif *in vivo*. De plus, la localisation dans les cellules intestinale des adultes ainsi que l'expression de la construction durant le stade *dauer* n'avaient jamais été rapportées auparavant. Ces observations sont à prendre avec précaution puisque ces résultats n'ont pu être confirmés par l'observation d'autres lignées et que l'expression du rapporteur en multicopies extrachromosomiques peut présenter des différences avec l'expression endogène. Cependant, ces éléments sont intéressants à suivre et pointent vers un lien entre l'expression du rapporteur *top-1* et la réponse physiologique à des conditions de stress.

## 2. La recherche d'éléments en *cis*

La recherche d'éléments en *cis* impliqués dans la régulation a commencé avec le seul régulateur connu de *top-1*, HRPF-1. L'implication directe du facteur dans l'épissage de *top-1* n'a pas été déterminée. En effet, l'implication du facteur HRPF-1 sur l'épissage de *top-1* a été identifiée par un Knock-Down du gène *hrpf-1* par ARN interférence. Ce KD change le ratio des ARNm de *top-1* (Barberan-Soler and Zahler, 2008). Il se peut que cette régulation soit directe, par le biais de la fixation du facteur HRPF-1 sur le pré-messager de *top-1*. Or, la

séquence à laquelle se fixe le facteur HRPF-1 n'est pas connue. Cependant, les séquences sur lesquelles les domaines RBDs de ses orthologues humains hnRNP F et hnRNP H se fixent sont connues. Des travaux réalisés *in vivo* et portant sur l'affinité de ces protéines vis-à-vis de différentes séquences ont pu mettre en évidence une séquence optimale : AGGGAU (Dominguez et al., 2010). La modification des nucléotide A ou U de la séquence n'a que des effets modérés sur l'affinité de liaison tandis que la modification d'un G entraîne une diminution de l'affinité cent à mille fois inférieure en fonction du G touché (Dominguez et al., 2010). En raison de ces données, le rapporteur *top-1* a été scanné pour la présence de triplets de G. Trois séquences ont été identifiées : à la position 50-52 de l'intron I, à la position 104-106 de l'exon II alternatif et à la position 205-207 de l'intron II. Pour déterminer si ces sites sont impliqués dans la régulation, des rapporteurs mutés ont été générés : mut1, mut2 et mut3 qui perdent le triplet aux positions respectives des sites 1, 2 et 3. Des combinaisons ont également été réalisées : mut1mut2 ; mut2mut3 ; mut1mut3. La dernière combinaison – mut1mut2mut3 – n'a pas été réalisée par manque de temps.

En se basant sur les éléments en *cis* connus chez *Caenorhabditis elegans*, deux séquences types ont pu être identifiées : la séquence GCAUG reconnue par ASD-1/FOX-1 à la position 30-34 de l'intron II ; la séquence CUAAC reconnue par ASD-2 aux positions 438-442 de l'intron I, 362-366 et 421-425 de l'intron II. L'expression connue des gènes *asd-1* et *fox-1* correspond à celle du rapporteur *top-1* observé (Kuroyanagi et al., 2007), ce qui soutient l'idée d'une participation dans la régulation. La séquence GCATG a donc été modifiée par mutagenèse dirigée pour obtenir un nouveau rapporteur - mut4.

Si l'expression concorde pour ASD-1/FOX-1, ce n'est pas le cas de l'expression du facteur ASD-2, qui est musculaire (Ohno et al., 2008). Pour cette raison, ces séquences ont été mises de côté pour le moment.

Toutes les constructions ont été injectées. Aucune n'a permis d'obtenir une lignée, ce qui rejoint les difficultés inhérentes au rapporteur sauvage.

### 3. La recherche de facteurs en *trans*

La relative instabilité de la lignée obtenue avec le rapporteur *top-1* empêche la mise en place d'un crible par génétique classique. En effet, un crible par mutagenèse aléatoire nécessite une très grande quantité d'individus à cribler, ce qui rentre en conflit avec la faible transmission de notre rapporteur.

Pour identifier des facteurs de régulations, une autre voie existe par le biais d'un crible de génétique inverse. Pour cela, des banques de souches bactériennes exprimant des ARN double brin existent.

Une sélection a été réalisée sur la banque ORFeome afin de créer une mini-banque plus facile et rapide d'utilisation qui contient les souches capables de diminuer l'expression de toutes les protéines possédant un domaine de liaison à l'ARN soit 176 clones sélectionnés. La présence de ce type de domaine étant la seule constante chez les régulateurs de l'épissage. De plus, toutes les souches capables de diminuer l'expression des gènes *smg* disponibles ont également été incluses dans cette banque. Les gènes *smg* sont impliqués dans le NMD et leur expression peut donc affecter certaines isoformes produites.

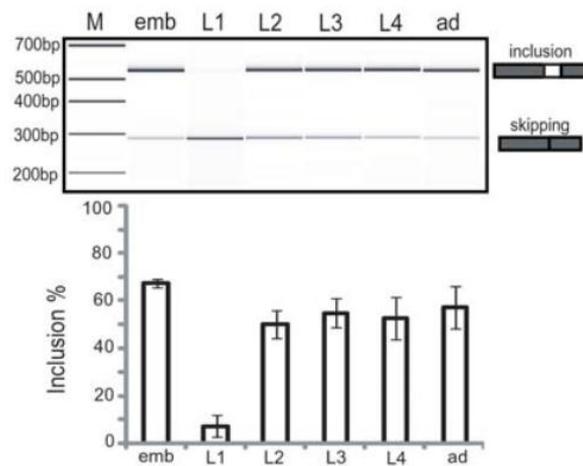
Cette banque peut donc être utilisée pour mettre en place de cribles de génétique inverse à la recherche de facteurs d'épissage. Cependant, cette banque n'a pu être appliquée à la lignée du rapporteur *top-1*. En effet, des expériences préliminaires d'ARNi réalisées sur quelques cibles – dont *hrpf-1* – ont démontrées l'impossibilité technique de cette méthode sur une lignée aussi instable – le nombre de vers fluorescents étant trop faible.

Un régulateur de l'épissage de *top-1* est déjà connu, HRPF-1. Puisque ce régulateur régule lui-même son épissage, nous nous sommes tournés vers lui.

#### 4. Le rapporteur *hrpf-1*

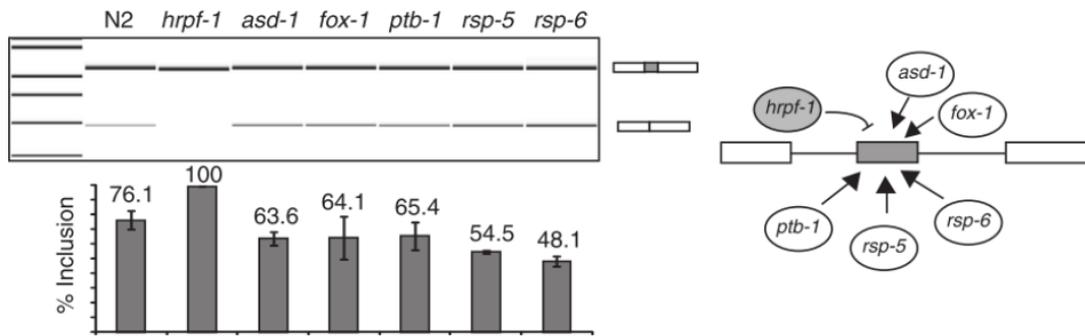
HRPF-1 est un régulateur de l'épissage alternatif (Barberan-Soler and Zahler, 2008). Situé sur le chromosome I, le gène *hrpf-1* possède deux ARNm identifiés. La première isoforme est constitué de 7 exons encodant trois domaines RBD – RNA-Binding Domain. L'isoforme la plus courte ne contient pas l'exon V, ce qui change le cadre de lecture, amenant à l'apparition d'un codon stop dans l'exon VI. La protéine synthétisée avec cette seconde forme ne contiendrait que deux RBD. Cependant, cette seconde protéine pourrait ne pas être synthétisée. En effet, l'ARNm court – *hrpf-1b* ou V(-) – est la cible du NMD qui provoque la dégradation de cette isoforme (Barberan-Soler and Zahler, 2008). Une faible quantité de cette isoforme n'est pas dégradée mais sa traduction est supposée être réprimée (Barberan-Soler et al., 2009). L'inclusion de l'exon V est modulée au cours du développement (Figure 39).

### *hrpf-1* alternative splicing during development



**Figure 39 : Régulation de l'épissage alternatif de *hrpf-1* au cours du développement.** RT-PCR réalisé sur des extractions d'ARN totaux des différents stades de développement du nématode (emb : embryon ; L1-4 : stades larvaires ; ad : adultes). La quantité relative des deux isoformes change fortement durant le premier stade larvaire avant de se rétablir au cours du développement à un ratio proche de son niveau embryonnaire. Le pourcentage d'inclusion au cours du développement est représenté en bas. Extrait de (Barberan-Soler and Zahler, 2008).

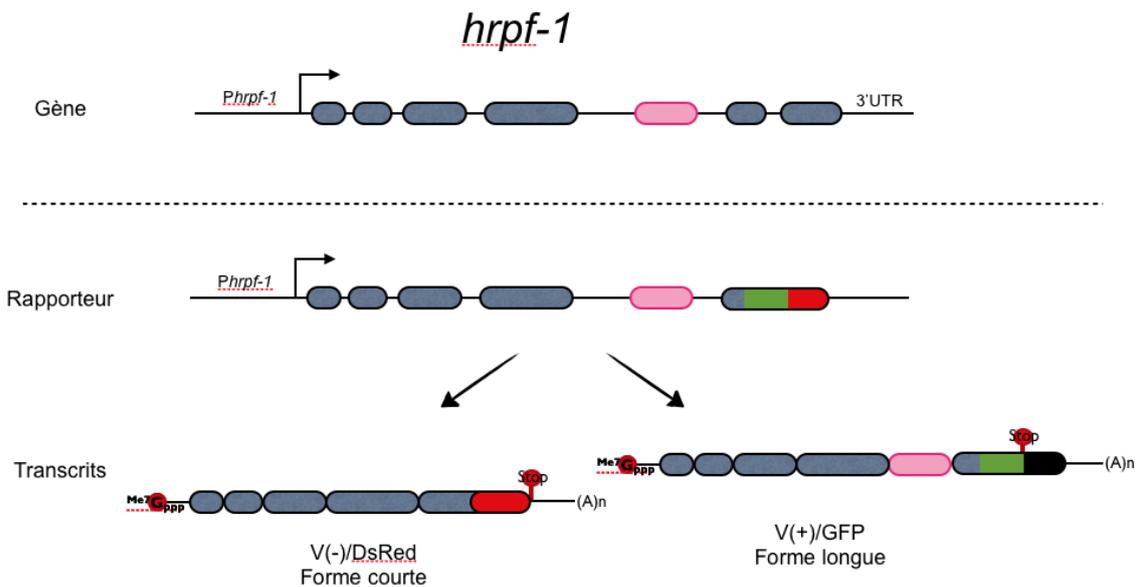
HRPF-1 favorise l'exclusion de l'exon V de l'ARNm Hrfp-1 (Figure 40). Cette action est contrebalancée par l'action des facteurs ASD-1, FOX-1, PTB-1, RSP-5 et RSP-6 qui favorisent l'inclusion de l'exon V (Barberan-Soler et al., 2011). Bien que de nombreux facteurs en *trans* impliqués dans la régulation soient connus, les éléments en *cis* ne le sont pas. Les séquences consensus que fixent les facteurs ASD-1, FOX-1 et PTB-1 sont connus et des séquences correspondantes sont présentes dans les introns proches de l'exon alternatif. L'étude de ce gène avec notre système rapporteur pourrait donc permettre de compléter les informations déjà connues ainsi que de donner une explication mécanistique au changement drastique de l'épissage observé durant le stade L1 (Figure 39).



**Figure 40 : RT-PCR réalisé sur des ARNs totaux issus de lignées N2 et mutantes.** Ces résultats ont été générés sur BioAnalyzer (Agilent). Les barres d'erreur représentent les variations obtenues pour au moins deux réactions de RT-PCR différentes. Le pourcentage d'inclusion au cours du développement est représenté en bas. Extrait de (Barberan-Soler and Zahler, 2011).

### a) La première construction

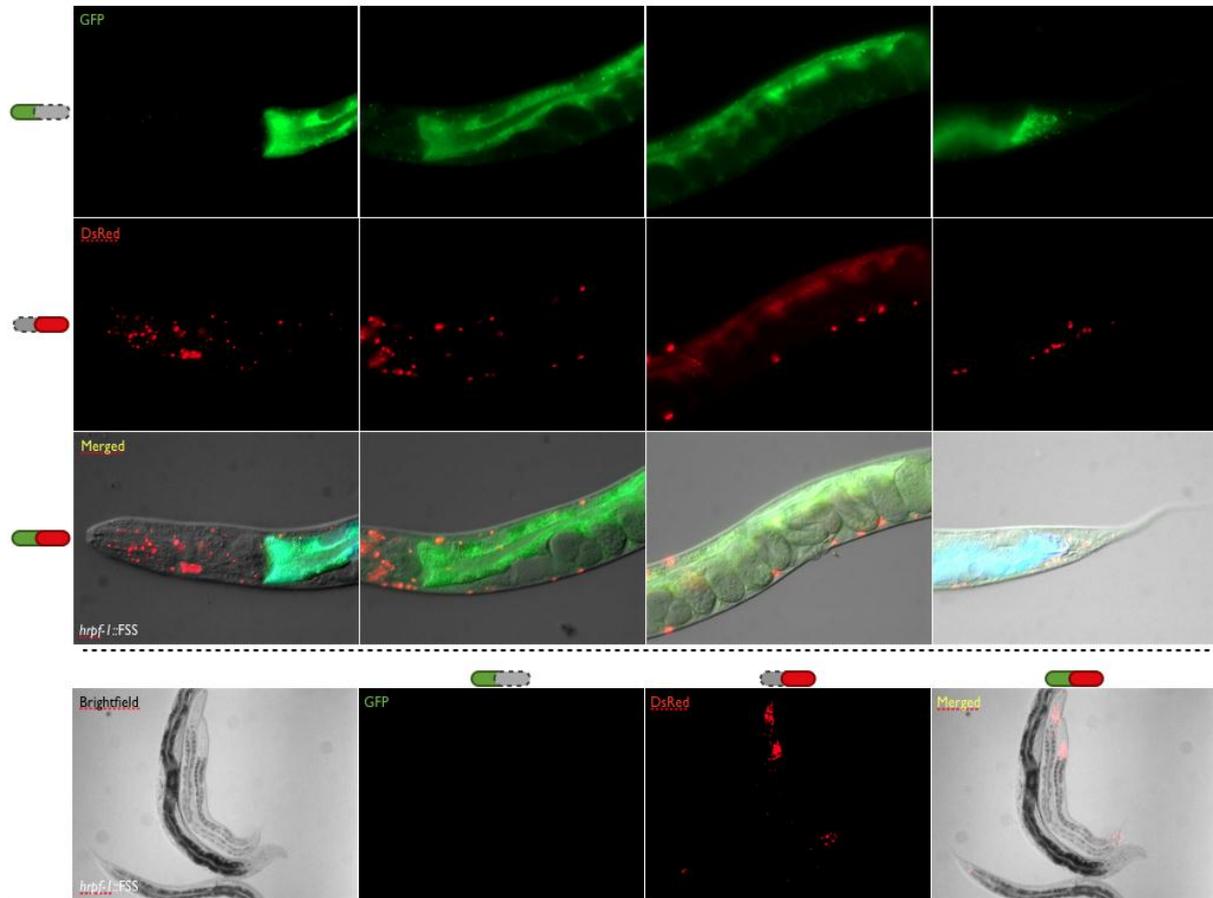
La première construction réalisée pour suivre ce gène tire parti du changement de cadre de lecture naturel liée à l'absence de l'exon V (Figure 41).



**Figure 41 : Représentation du gène *hrpf-1* et de la première construction réalisée.** Le promoteur et la séquence du gène de l'exon I au début de l'exon VI sont associés à la cassette. La forme longue de l'isoforme – V(+) – est associée au cadre de lecture de la GFP et la forme courte –V(-) – à celui de la DsRed.

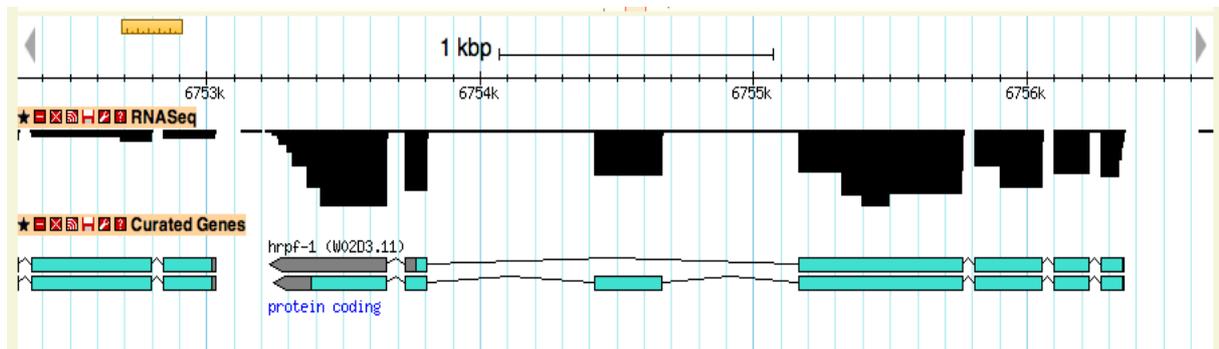
Deux lignées transgéniques portant des concatémères répétitifs ont été obtenues. Les individus expriment la forme courte V(-)/DsRed dans la tête et la queue, ainsi que le long de

l'organisme. Cette forme est prédominante mais la seconde – V(+)/GFP – est exprimée faiblement et co-localise avec la forme courte (Figure 42). Un grand nombre de cellules expriment les protéines fluorescentes et il n'y a aucune variation flagrante de l'expression au cours du développement.



**Figure 42 : Micrographie d'individus porteurs du rapporteur *top-1::FSS*.** Les trois premières lignes correspondent aux canaux GFP, DsRed et superposition d'un individu adulte à l'objectif 40x. La quatrième ligne correspond aux différents canaux utilisés sur deux individus fluorescents. La forme courte – en ROUGE – est présente massivement dans la tête de l'individu, mais aussi dans la queue et le long du corps. La forme longue – en VERT – est faiblement visible à l'objectif 40x et co-localise avec la forme longue (Les temps d'expositions sont du même ordre de grandeur).

Les données de séquençage des ARNs semblent aller dans le sens des travaux publiés, à savoir une plus forte expression de la forme longue. Cependant, les données accessibles montrent des résultats ambigus (Figure 43).



**Figure 43 : Modèle du gène *hrpf-1*.** Les exons bleus représentent les deux isoformes du gène *hrpf-1*. L'exclusion de l'exon alternatif change le cadre de lecture – les deux isoformes n'ont pas le même codon stop. Les données de séquençage des ARNs disponibles sont représentées au-dessus en noir. Extrait de WormBase.

Les résultats obtenus vont dans le sens inverse de ce qui est connu :

(a) Tout d'abord, la présence forte de la forme courte vont à l'encontre de la reconnaissance de cette forme par le NMD (Barberan-Soler and Zahler, 2008). Cependant, la voie classique du NMD passe par la reconnaissance de codon stop en amont de jonction exon-exon par le biais du complexe EJC (**Introduction.II.2.b**). Or, la cassette FSS, placée avant le codon stop de la forme courte, ne contient pas d'intron. Le rapporteur ne possède donc pas de signaux pour être reconnu par la voie classique du NMD – ce qui est cohérent avec les résultats.

(b) De plus, la surreprésentation d'une forme vis-à-vis de l'autre et l'absence de régulation visible au cours du développement va à l'encontre de ce qui est publié (Barberan-Soler and Zahler, 2008). Cependant, les deux isoformes hybrides produites par la construction portent la majeure partie de la séquence codante dont deux domaines de fixation aux ARNs. Ces deux domaines sont responsables de l'exclusion de l'exon V. En effet, un mutant du gène *hrpf-1* ne possédant pas ces domaines aboutit à l'abolition de la régulation. Ce mutant a une délétion de 426pb qui change le cadre de lecture et provoque l'apparition d'un codon stop prématuré (Barberan-Soler et al., 2011). Il se pourrait donc que la régulation soit perturbée puisque nous sur-exprimons le régulateur et l'élément régulé avec la même construction et donc que notre protéine de fusion régule son propre épissage expliquant ainsi le profil observé.

## **b) La seconde construction**

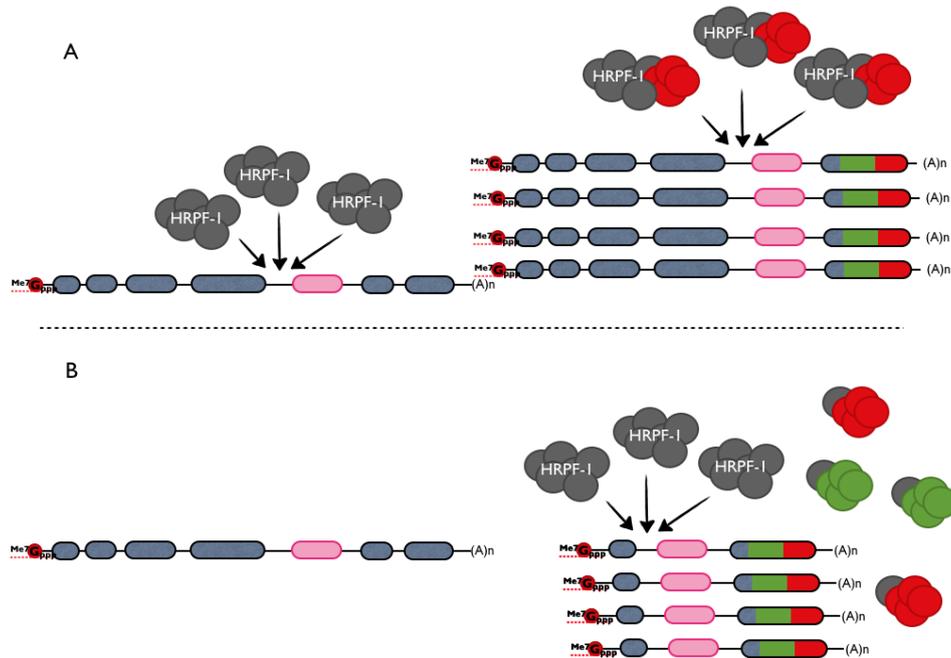
Une seconde construction a été réalisée avec la même séquence de promoteur suivi d'un ATG fusionné à la moitié de l'exon IV. Les isoformes sont couplées aux mêmes fluorophores – court/V(-)/DsRed et long/V(+)/GFP.

Malgré les injections répétées, aucune lignée transgénique n'a pu être générée avec ce rapporteur. Il semblerait que cette construction ait également un effet délétère chez les individus la portant en multiples copies extra-chromosomiques. En effet, après les injections, quelques individus fluorescents ont pu être observés. Ceux-ci présentaient une expression de la GFP et de la DsRed visible à la loupe à fluorescence et présent dans des cellules distinctes. Cependant, ces larves présentaient également des difformités sévères et ne se sont pas développées au-delà du premier stade larvaire.

Si aucune lignée n'a pu être obtenue avec un ce second rapporteur, les quelques individus fluorescents vont dans le sens de l'autorégulation de la protéine hybride. Ce problème de létalité fait écho aux problèmes rencontrés avec le gène *top-1* (**Résultats III.1**) dont *hrpf-1* est le seul régulateur connu.

Le cas de ce gène reste intéressant bien que problématique en raison de notre incapacité à obtenir une lignée avec le second rapporteur, alors que le premier n'avait posé aucun problème.

Il est possible qu'il s'agisse d'un phénomène de titration où HRPF-1 endogène est saturé par l'expression du rapporteur en multiples copies. Dans les lignées porteuses de la première construction, une protéine hybride dotée de la grande partie de la séquence de *hrpf-1* – dont deux RBD - était surexprimée alors que ces domaines manquaient à la seconde construction (Figure 44).



**Figure 44 : Représentation de l’hypothèse relative à l’expression des deux rapporteurs *hrpf-1*.** (A) Le premier rapporteur pourrait entraîner la surexpression de protéines hybrides fonctionnelles, créant ainsi le régulateur en même temps que l’élément régulé – l’ARNm ; (B) Le second rapporteur pourrait engendrer l’expression d’une protéine hybride sans la fonction de l’épissage du facteur HRPF-1 et titrer le facteur endogène pour son propre épissage.

En raison de ces résultats, il a été décidé de co-injecter le gène *hrpf-1* avec le rapporteur *top-1* afin de voir si, comme avec la forme longue du rapporteur *hrpf-1*, l’expression du régulateur pouvait lever l’effet délétère du rapporteur.

## 5. Co-expression du gène *hrpf-1* et du rapporteur *top-1*

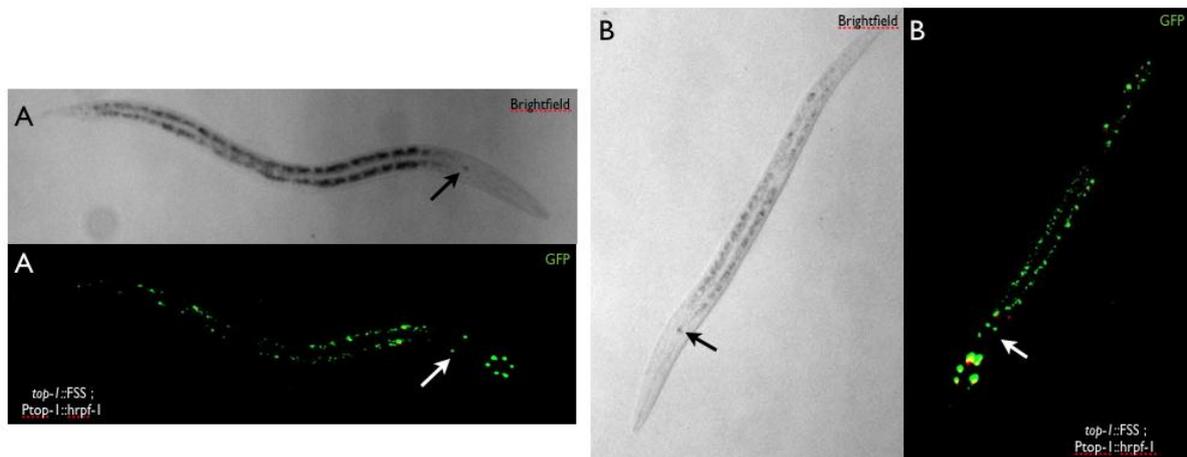
Pour confirmer l’impact du facteur HRPF-1 sur l’épissage *top-1*, une nouvelle construction a été réalisée. La séquence du gène *hrpf-1*, comprenant exons et introns, a été clonée sous le contrôle du promoteur de *top-1* et la région 3’UTR du gène *unc-119* afin de mimer l’expression du rapporteur.

Un mélange d’injection a été réalisé pour permettre la formation d’un concatémère répétitif avec le rapporteur *top-1* à une concentration de 45ng/μL et la construction *Ptop-1::hrpf-1::3’UTRunc-119* à une concentration de 45ng/μL (**Introduction.IV.4.a**). Ce choix de réaliser à nouveau des injections avec le rapporteur de *top-1* peut paraître incongru aux vues des difficultés posées lors des précédentes injections. Cependant, la première construction du gène *hrpf-1* n’a posé aucun problème contrairement à la seconde. Cette différence pourrait provenir d’un effet « dominant négatif » de la seconde construction tronquée mais également

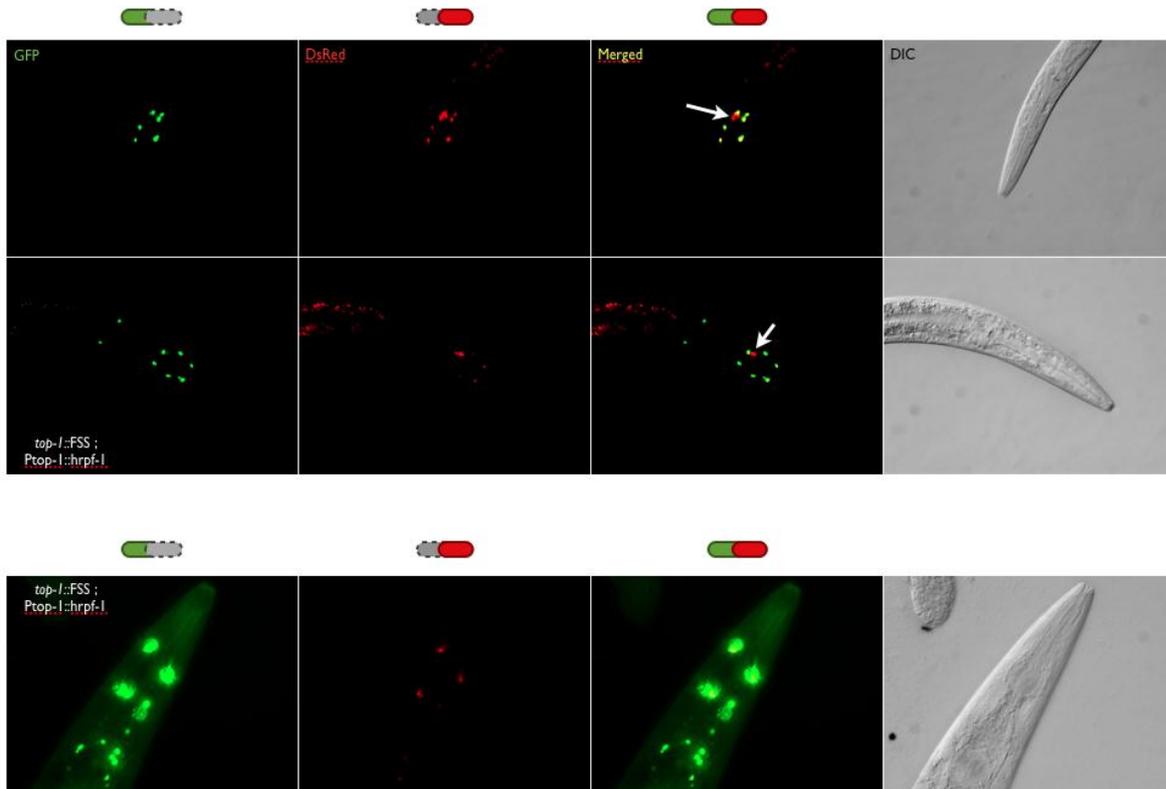
d'un phénomène de titration du facteur HRP-1 endogène non-compensé par le transgène (**Résultats.II.4**). Il a donc été décidé de déterminer si la surexpression du facteur HRP-1 pourrait lever l'effet délétère du rapporteur de *top-1* d'une part et d'autre part permettre de visualiser l'effet direct de la présence d'HRPF-1 sur le choix de l'isoforme de *top-1* exprimée.

Trois lignées indépendantes ont pu être obtenues en injectant ce mélange. En plus de la facilité avec laquelle ces lignées ont générées, celles-ci ne présente aucune instabilité et se maintiennent très bien à l'aide du G418. L'expression de cette construction *hrpf-1* semble donc avoir permis de corriger l'instabilité du rapporteur *top-1*.

Chaque lignée présente des variations de l'expression des deux fluorochromes entre les individus. Cependant, ces variations d'expression sont très similaires entre les trois lignées. L'expression du rapporteur se limite à des cellules de la tête et de la queue de l'organisme (Figure 45 ; Figure 46).



**Figure 45 : Micrographie d'individus porteurs du rapporteur *top-1::FSS* et du régulateur *Ptop-1::hrpf-1*.** (A) et (B) individus larvaires observés à l'objectif 10x. Les flèches noires pointent la base du pharynx et les flèches blanches, la position de l'anneau nerveux. Les individus présentent un signal GFP dans deux cellules de l'anneau nerveux et dans six cellules proches du premier bulbe du pharynx. Les larves présentent majoritairement un signal DsRed co-localisée avec la GFP. De nombreux présentent une seule cellule de la tête présentant un signal DsRed sans GFP.



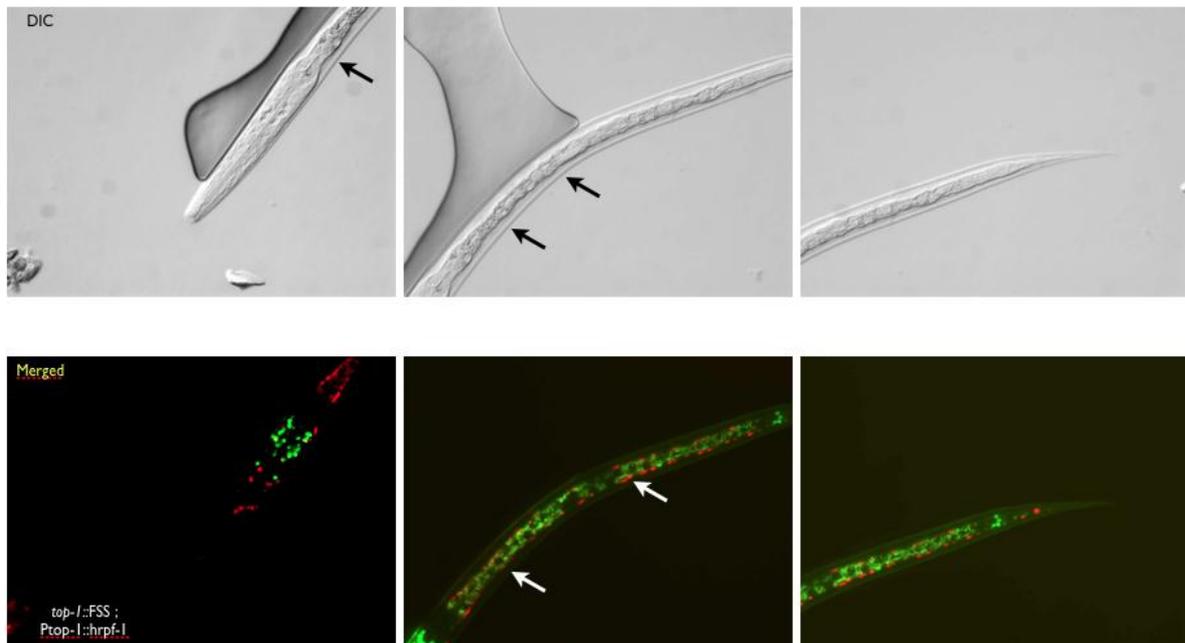
**Figure 46 : Micrographie d'individus porteurs du rapporteur *top-1::FSS* et du régulateur *Ptop-1::hrpf-1*.** Les deux premières lignes présentent des larves. Les flèches blanches pointent l'unique cellule présentant un signal DsRed sans signal GFP visualisée. La troisième ligne correspond à la tête d'un adulte. La forme courte – en **VERT** – est présente dans des cellules de l'anneau nerveux ainsi que des cellules proche du premier bulbe. La forme longue – en **ROUGE** – est présente dans ces cellules et co-localisée chez les adultes avec la forme courte.

Les cellules fluorescentes de la tête de l'organisme ne sont pas identiques à celles retrouvée dans la lignée exprimant le rapporteur *top-1* seul. En effet, dans cette lignée la majorité des cellules GFP+ se situait autour de la base du pharynx où se situe l'anneau nerveux. Dans les trois lignées possédant la construction *hrpf-1*, les cellules de l'anneau nerveux présentent peu voir pas de fluorescences contrairement à des cellules situés autour du premier bulbe du pharynx. Ces cellules correspondent aux neurones IL – *Inner labia* – identifiés dans l'expression du rapporteur GFP (Cha et al., 2012). Le profil d'expression des cellules intestinales de la lignée porteuse du rapporteur de *top-1* seul n'a pas été observé chez les individus issus de ces trois lignées.

Il est difficile de statuer quant à ce changement d'expression. En effet, l'effet délétère du rapporteur seul n'a permis l'obtention que d'une seule lignée ce qui empêche de confirmer les résultats. De plus, les variations au sein de chaque lignée sont probablement le fruit d'une expression partiellement stochastique, connue des rapporteurs exprimés en multiples connus (Mello and Fire, 1995).

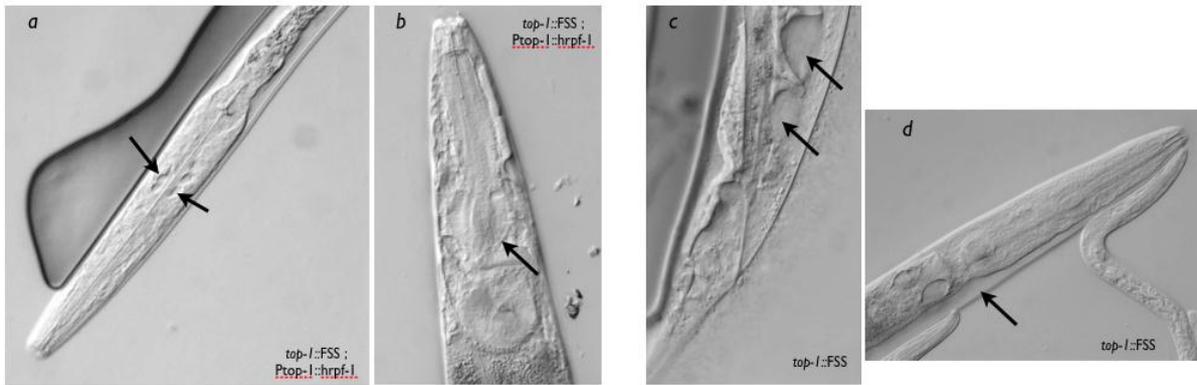
La majorité des cellules observées présentent la forme courte, soit seul comme dans l'anneau nerveux, soit co-exprimée avec la forme longue, ce qui va dans le sens d'une exclusion de l'exon alternatif causée par HRPF-1. Chez de nombreuses larves, une cellule présentant un signal DsRed uniquement a pu être observé. HRPF-1 pourrait ne pas y être stable ou bien son action peut être contrée par des facteurs qui favorisent l'inclusion.

En cas de stress de « crowding », des individus ont pu être observés avec un profil particulier. La forme courte II(-)/GFP est retrouvée dans de nombreuses cellules de l'anneau nerveux. La forme longue I(+)/DsRed est retrouvée dans des cellules de la tête et le long du corps qui n'expriment pas la GFP (Figure 47).



**Figure 47 : Micrographie d'un individu adulte porteur du rapporteur *top-1::FSS* et du régulateur *Ptop-1::hrpf-1* en conditions de surpopulation.** Les flèches noires indiquent une anomalie du ver qui ressemble à un creux. La forme courte – en VERT – est exprimée dans les cellules situées au niveau de l'anneau nerveux. La forme longue – en ROUGE – est retrouvée dans des cellules situées au niveau du premier bulbe du pharynx et dans la queue.

Les cellules le long du corps sont difficiles à identifier en raison du phénotype des vers. Ils présentent des trous le long du corps, en dessous de la cuticule (Figure 47). De telles anomalies ont pu être observées dans ces lignées, sans conditions de stress, ainsi que dans la lignée du rapporteur *top-1* (Figure 48).



**Figure 48 : Micrographie d'individus porteurs du rapporteur *top-1::FSS* et d'individus porteurs du rapporteur *top-1::FSS* et du régulateur P *top-1::hrpf-1*. (a) et (b) Individus porteurs du rapporteur *top-1::FSS* et du régulateur P*top-1::hrpf-1* ; (c) et (d) Individus porteurs du rapporteur *top-1::FSS*. Les flèches noires pointent des anomalies des vers présentés. (a) Des trous au niveau du premier bulbe du pharynx ; (b) La pharynx est anormalement tordu ; (c) Des trous dans la queue ; (d) Un creux en-dessous de la cuticule.**

La présence de ces anomalies signifie que la levée de nocivité due à l'expression du régulateur est seulement partielle. Les vers présentent un phénotype anormal et inconstant. De plus, l'effet du facteur HRPF-1 est sujet à précaution puisque l'on exprime un facteur et son régulateur par la même construction. Sa surexpression doit donc être modérée par rapport à celle du rapporteur *top-1*. De plus, le facteur HRPF-1 peut se comporter différemment dans un contexte cellulaire qui ne correspond pas forcément à celui dans lequel il est normalement présent – stabilité de la protéine, absence/présence de facteurs associés – et ne pas avoir le même effet sur l'épissage de *top-1*. Cependant, ces trois dernières lignées ont permis de mettre en avant plusieurs choses comme l'effet du facteur sur la survie des individus porteur du rapporteur.

Malgré tout, le problème probablement associé à la présence d'un trop grand nombre de copies de notre rapporteur pour *top-1* n'a pas pu être résolu par la co-transformation avec *hrpf-1*. C'est pourquoi nous nous sommes tournés vers l'intégration des rapporteurs en simple copie. Cette méthode permet de limiter la surexpression et la surreprésentation des constructions qui pourraient être responsable de ces problèmes et permette l'obtention de lignées stables exprimant le transgène à des niveaux plus physiologiques.

## IV. Stratégies d'insertion en copie unique

### 1. Mos-SCI

La stratégie d'insertion en MosSCI nécessite de transférer les constructions dans un nouveau plasmide, le pRG5271Neo et d'injecter dans une nouvelle lignée porteuse du transposon *Mos*, EN5271 (Giordano-Santini et al., 2010). Le pRG5271 est constitué d'une base de pBlueScript, de deux séquences d'homologie L et R – identiques aux séquences qui flanquent le transposon – et du gène de résistance *Neo* permettant la sélection de individus transgéniques.

Pour assembler la construction *top-1* à l'intérieur de ce vecteur, le rapporteur a été amplifié en une seule séquence. Deux paires d'amorces ont été utilisées : (1) possédant des queues flottantes de 30nt, homologues des extrémités du vecteur, pour permettre le clonage par « *chew-back annealing* », et (2) possédant des queues flottantes contenant un site de restriction *ApaI* pour permettre l'insertion par clonage de type « restriction-ligation ». Aucune de ces méthodes n'a permis d'obtenir la construction insérée dans ce vecteur.

Le vecteur de destination mesurant 8kb, l'insertion du rapporteur le porterait à 14kb. Cette taille pourrait constituer un obstacle au clonage.

Depuis, une nouvelle technique d'édition de génome utilisant le CRISPR-Cas a été introduite dans la communauté *C. elegans*.

### 2. CRISPR-Cas

#### **a) Principe**

CRISPR-Cas est un mécanisme de défense commun aux *archae* et des bactéries (Horvath and Barrangou, 2010) qui permet la reconnaissance d'ADN étranger et l'immunisation par la production d'ARN CRISPR (Clustered Regularly Interspaced Short Palindromic Repeat) – crARN - qui dirige la coupure séquence-spécifique grâce aux endonucléases Cas (CRISPR associated). Les crARNs s'hybrident à un autre type d'ARN, le tracrARN – trans-activating crARN – qui sert de lien avec les endonucléases. Ce mécanisme permet ainsi de reconnaître les séquences étrangères et de les éliminer. Les séquences cibles sont portées par les crARNs et dirigent une coupure double-brins des séquences homologues des cibles.

L'expression de la Cas9 et d'un ARN hybride, le sgARN - fusion du crARN et du tracrARN -, suffit pour causer une coupure génomique chez un organisme qui ne possède pas naturellement ce mécanisme (Friedland et al., 2013)(Feng et al., 2013). La séquence cible du sgARN n'est constituée que de 20 nucléotides et peut donc être facilement adaptée pour cibler

une coupure double-brin, faisant du système un outil d'édition du génome à la fois très versatile et très spécifique.

La technologie du CRISPR-Cas s'est développée rapidement et a remplacé d'autres méthodes telles que l'utilisation des nucléases à doigts de zinc et de TALENs, dont l'inconvénient majeur par rapport au CRISPR-Cas repose sur la complexité des clonages à réaliser en amont et sur les contraintes de cibles (Gaj et al., 2013). La technologie du CRISPR-Cas a été appliquée aux souris et rats (Li et al., 2013), aux plantes (Feng et al., 2013), poissons-zèbres (Hwang et al., 2013) ainsi que chez le nématode (Friedland et al., 2013).

### **b) Méthodologie appliquée à *Caenorhabditis elegans***

La méthodologie a été appliquée à l'organisme modèle *Caenorhabditis elegans* au cours de l'année 2013 (Friedland et al., 2013). L'injection de constructions permettant l'expression de la Cas9 et du sgARN suffisent à provoquer des délétions aux sites désirés dans la progéniture des individus injectés. Différentes combinaisons d'éléments injectés ont données des résultats : constructions plasmidiques (Friedland et al., 2013)(Waijers et al., 2013) ARNs (Chiu et al., 2013) et complexe ribonucléoprotéique (Cho et al., 2013). Ces travaux ayant été réalisés en parallèle sur des cibles différentes, il était encore impossible de déterminer laquelle de ces combinaisons était la plus efficace jusqu'à récemment (Paix et al., 2015). Les résultats obtenus à ce jour sont peu nombreux et présentent des efficacités variables en fonction du site de coupure sélectionné (Tzur et al., 2013)(Farboud and Meyer, 2015).

### **c) Séquence PAM et contraintes**

La seule contrainte à la cible de la séquence repose sur la présence du motif NGG directement placé en 3' de la séquence cible. Ce motif est connu sous le nom de séquence PAM – Protospacer Adjacent Spacer – et est nécessaire à la reconnaissance de la cible par la Cas9 (Gasiunas et al., 2012). Cependant, un article publié cette année met en avant une amélioration de l'efficacité de coupure lorsque la cible se termine par un dinucléotide GG (en amont de la séquence PAM)(Farboud and Meyer, 2015).

## **3. Insertion universelle**

En raison du nombre de constructions à intégrer – rapporteur *top-1* et ses versions mutées, rapporteur *hrpf-1* -, la méthode actuelle pose trois problèmes.

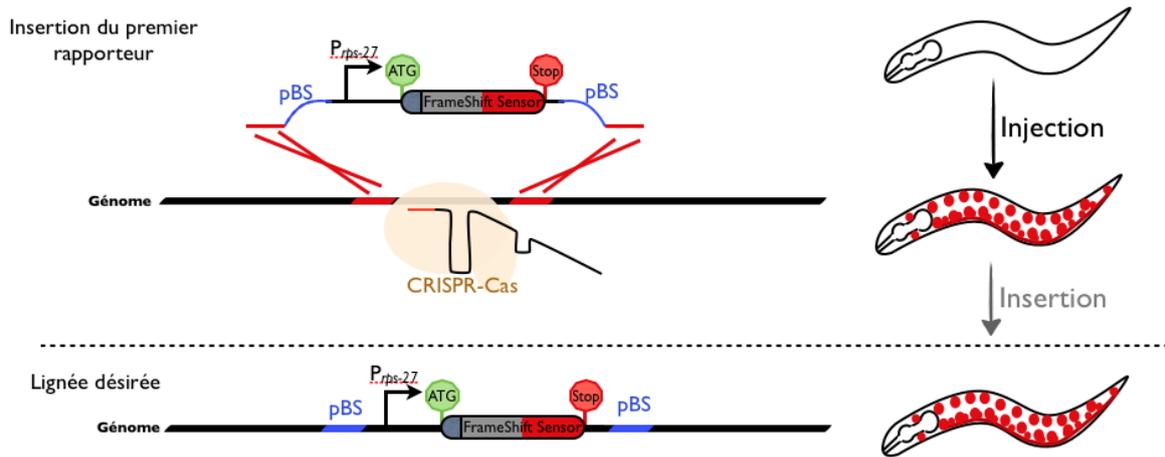
Le premier est relatif à la variation dans les résultats observés jusqu'à présent. Ce problème pourrait provenir en partie du changement dans les séquences cibles et celles utilisées pour permettre l'insertion par recombinaison homologue. Pour stabiliser l'efficacité de l'insertion, il serait donc nécessaire de fixer les constantes qui peuvent l'être – cible de coupure et séquences d'homologie pour la recombinaison.

Le second problème est relatif au système de sélection. Notre objectif étant d'insérer des rapporteurs fonctionnels en extra-chromosomique, la fluorescence ne peut servir de marqueur de sélection. La méthode actuelle n'utilise aucun système de sélection, préférant isoler des centaines d'individus pour les cribler par PCR (Waijers and Boxem, 2014). Ce système prendrait un temps considérable et rend difficile le nombre d'insertion qui nous intéressent.

Le troisième problème est relatif aux clonages nécessaires. En effet, pour permettre l'insertion, il est nécessaire de flanquer les constructions de séquences d'homologies – traditionnellement 1 à 1,5 kb (Tzur et al., 2013). L'utilisation de séquences de micro-homologie de seulement 30 nucléotides est possible chez le nématode (Paix et al., 2014). Cependant, l'efficacité de ces micro-homologies reste à démontrer sur un plus grand nombre de cibles.

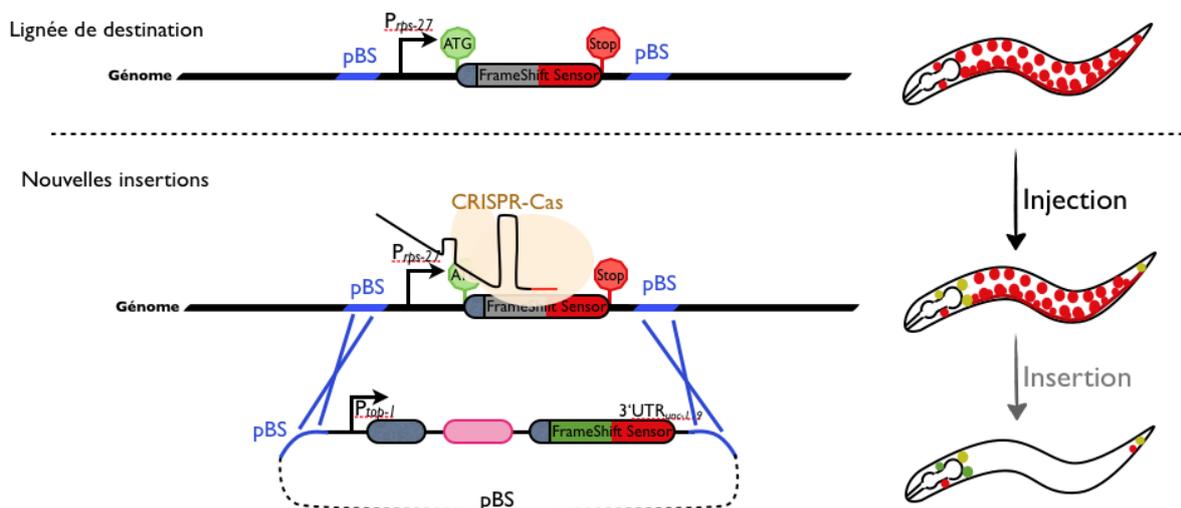
De plus, le problème posé par *top-1* et *hrpf-1* liée à la surexpression constitue un obstacle potentiel pour l'étude de nombreux autres gènes, ce qui limiterait grandement l'intérêt du système développé. Il est donc intéressant de mettre au point une méthode permettant une insertion facilitée tout en limitant la quantité de clonage et d'étapes de biologie moléculaire nécessaire pour passer d'une expression en multiples copie extra-chromosomique à une intégration génomique en simple copie.

Pour permettre cela, nous avons décidé d'insérer dans le génome de *C. elegans*, un rapporteur fluorescent flanqué de deux séquences d'un kilobase issues du plasmide utilisé au laboratoire pour porter les constructions assemblées durant ma thèse, le pBluscript(SK+) (ou pBS) (Figure 49).



**Figure 49 : Représentation de l’insertion du premier rapporteur par CRISPR-Cas.** L’insertion utilise des séquences de micro-homologie comme décrite dans Paix et al., 2014. Les séquences de 1kb issues du pBS sont représentées en BLEU. La coupure double-brins va être engendrée par CRISPR-Cas. La coupure favorise l’insertion du rapporteur et des séquences du pBS par recombinaison homologue. La lignée désirée est fluorescente.

Le rapporteur inséré permet l’expression d’une DsRed nucléaire à partir d’un promoteur fort – celui du gène *rps-27*. Une fois cette lignée obtenue, l’utilisation d’un sgARN dirigeant la coupure à l’intérieur du rapporteur inséré peut être réalisée pour le remplacer (Figure 50).



**Figure 50 : Représentation de l’insertion des rapporteurs dans la lignée de destination par CRISPR-Cas.** Le rapporteur inséré est ciblé par CRISPR-Cas, provoquant une coupure double-brins dans le rapporteur. Les séquences du plasmide – en BLEU – servent de séquences d’homologie pour l’insertion des rapporteurs désirés. La perte du premier signal sert à identifier et isoler des candidats potentiels. La transmission mendélienne du second signal et la vérification par PCR serviront à valider les intégrations.

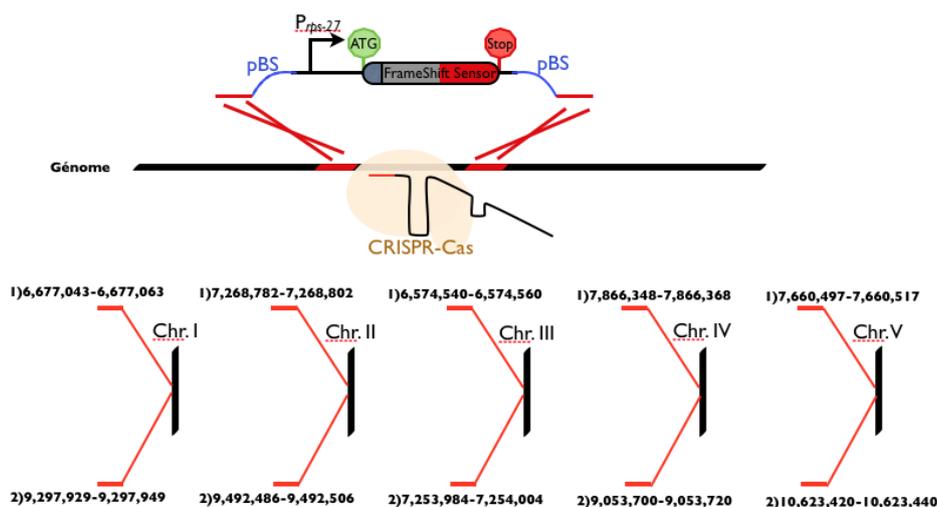
Ce système est en réalité très similaire à celui du MosSCI – le rapporteur inséré en premier lieu remplace le transposon – mais sans nécessiter d'étape de clonage dans un nouveau plasmide adapté.

L'intérêt d'une lignée porteuse de ces éléments est multiple. D'une part, la lignée de départ étant fluorescente, les insertions de rapporteur peuvent être suivies. La perte de l'expression initiale issue du rapporteur inséré permettra d'isoler des individus candidats. La transmission mendélienne du second rapporteur pourra servir à sélectionner des candidats et l'insertion vérifiée par PCR. De plus, les séquences issues du pBS deviennent des séquences d'homologies, ce qui engendre la possibilité de cloner des constructions dans un vecteur unique, permettant l'expression en multiples copies et en simple copie. Enfin, grâce à ce système, la cible de coupure et les séquences d'homologie sont identiques à chaque insertion, ce qui pourrait permettre de lever une partie de la variabilité observée dans l'efficacité de coupure et d'insertion (Tzur et al., 2013)(Farboud and Meyer, 2015). Ce dernier point reste à être confirmé ou infirmé grâce à l'insertion de nos constructions. Cette stratégie ayant été développée dans les tous derniers mois de ma thèse elle n'a pas encore été validée expérimentalement.

## 4. Méthode

### a) Sélection des sites de coupures dans le génome – Première insertion

Pour mettre au point ce système, le premier rapporteur doit être incorporé au génome. Pour cela, dix sites de d'insertion candidats ont été sélectionnés situés sur les cinq autosomes de l'organisme, à raison de deux sites par chromosome (Figure 51).

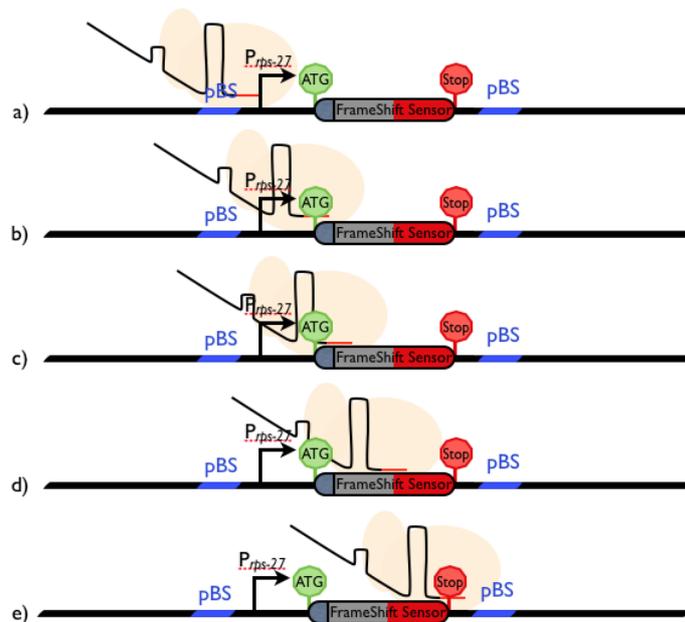


**Figure 51 : Représentation des sites d'insertions.** Dix sites d'insertions ont été sélectionnés à raison de deux par autosomes. Ces sites sont séparés de 4 à 5kb des gènes les plus proches.

Le chromosome X a été laissé de côté en raison de son comportement particulier – les mâles n’en possèdent qu’un seul. Les sites sélectionnés sont situés au centre des chromosomes pour permettre l’expression dans les cellules de la lignée germinale (Froakjensen *et al.*, - à paraître). Les sites de coupures sont séparés d’au moins 4kb des gènes adjacents afin de limiter la perturbation des éléments de régulations endogènes. La séquence à insérer est amplifiée avec des oligonucléotides portant des séquences d’homologie de 30nt. La première insertion se fait donc avec des séquences de micro-homologies (Paix et al., 2014).

### **b) Sélection des sites de coupures au sein du rapporteur – Seconde insertion**

Pour cibler spécifiquement le rapporteur préexistant et éviter de recouper la construction finale désirée les cibles choisies sont situées à la jonction des différents éléments utilisés afin de ne pas causer de coupures double-brins non-voulues (Figure 52).



**Figure 52 : Représentation des cibles sélectionnées pour permettre l’insertion dans la lignée de destination.** Ces cinq sites sont situés le long du premier rapporteur intégré. La première cible (a) est située à cheval entre la séquence issue du pBS - en BLEU – et le promoteur du gène *rps-27*. La seconde et la troisième cible (b et c) sont situées entre le promoteur et la séquence de la GFP issue du FSS. La quatrième cible (d) se situe entre la séquence de la GFP et celle de la DsRed. La dernière cible (e) est située entre la séquence de la DsRed et la région 3’UTR du gène *unc-119*.

### **c) Système de sélection des progénitures « jackpot »**

Lors de l'insertion du premier rapporteur, nous ne disposons pas d'outil pour permettre de visualiser l'insertion, puisque le rapporteur s'exprimera en extra-chromosomique. Pour cela, nous nous sommes tournés vers un système de co-conversion, reposant sur l'utilisation – en parallèle d'un sgARN d'intérêt – d'une seconde cible faisant office de marqueur. La coupure à ce second site engendre un phénotype dominant – *dumpy* – et l'intégration à ce second site permet l'obtention d'un autre phénotype dominant – *roller*. Ces deux phénotypes étant dominant – déclenché malgré le caractère hétérozygote de la modification – ils permettent de visualiser facilement les individus touchés à ce locus. La progéniture d'individus injectés présente donc deux types de vers : sauvages et mutants. Les mutants *dumpy* et *roller* présentent plus de chances de présenter des modifications au second locus – 0,4%-7% des individus sauvages contre 40%-60% des individus mutants (Arribere et al., 2014). L'effet des mutations au locus marqueur étant dominant et les modifications du génome dans les premières générations étant majoritairement hétérozygotes, il est facile d'éliminer la mutation de suivi par une simple sélection supplémentaire d'individus sauvages à la génération suivante (Paix et al., 2014)(Farboud and Meyer, 2015). Ce système a donc été utilisé pour faciliter l'obtention des premières insertions.

### **d) Résultats**

Toutes les constructions nécessaires ont été assemblées et les dix mélanges d'injection ont été réalisés. Les injections ont commencées et sont toujours en cours à l'heure où j'écris ces lignes.

# Discussion



## I. La nécessité des rapporteurs *in vivo*

L'utilisation croissante des méthodes séquençages des ARNs à permis de détecter des événements d'épissage alternatif en masse. Cependant, les données issues d'extraction d'ARN d'organismes entiers ne peuvent permettre de caractériser les régulations spatiales et temporelles. Il est nécessaire de développer des outils et des méthodes capables d'étudier ces événements et leurs complexités *in vivo*.

Les gènes récupérés dans ma première liste couvrent plus 10% des gènes de l'organisme *Caenorhabditis elegans*. Cette liste est probablement incomplète puisque des études mettent en évidence près de 25% des gènes (Ramani et al., 2011) comparés au 10% estimé précédemment (Brett et al., 2002). Chez l'Homme, cette estimation monte à plus de 90%. Il est donc nécessaire d'évaluer si la totalité de ces transcrits est régulé, et si ce n'est pas le cas, d'identifier les éléments qui permettent de discriminer les événements qui relèvent de la stochasticité de la machinerie d'épissage de ceux qui ont une régulation précise chez l'organisme.

Les données accumulées par le séquençage des ARNs peuvent permettre, grâce à des traitements bio-informatiques, d'identifier des éléments en *cis* enrichies près des séquences différentiellement incluses. Cependant, seules les séquences massivement représentées peuvent être identifiés et leurs fonctions dans l'inclusion et l'exclusion d'exons et d'introns ne peuvent être déterminées sans validation expérimentale.

L'utilisation d'un système rapporteur *in vivo* ne peut se substituer aux méthodes à haut débit, comme ces méthodes ne peuvent se substituer à l'étude du cas par cas. Au contraire, ces méthodologies s'alimentent mutuellement pour affiner notre compréhension du système dans son ensemble. L'étude de l'épissage alternatif, comme tout autre mécanisme de régulation est complexe et nécessite une combinaison d'approches différentes, qui, une fois appliquées de concert, pourront permettre de révéler la réalité moléculaire du vivant.

## II. Une évolution du système de paires de « minigènes »

La méthodologie développée ici est très proche de celle appliquée grâce aux paires de « minigènes ». Cependant, notre méthodologie utilise les promoteurs endogènes et nous permet de créer moins de constructions et de lignées pour arriver au même résultat – la méthodologie des « minigènes » utilisent plusieurs promoteurs tissu-spécifiques pour révéler une régulation.

Le second intérêt de la cassette FSS est qu'elle permet de suivre deux isoformes à partir d'un seul pré-messager. La transcription d'un gène étant couplée avec l'épissage, l'utilisation de deux constructions portées par des gènes indépendant peut perturber le résultat

d'épissage. Le ratio des deux isoformes est également probablement plus fiable car deux lignées indépendantes ne contiennent pas nécessairement le même nombre de rapporteur, ce qui peut affecter la quantité relative des deux isoformes produites avec deux rapporteurs distincts. Ce n'est pas le cas lorsque les deux isoformes sont produites par un seul gène. De plus, une seule construction est plus facilement intégrée en simple copie dans le génome, ce qui facilite l'étude des cas problématiques comme *top-1*.

La combinaison de ces éléments fait du système FSS une évolution méthodologique de la paire de « minigènes ». Ce système est appliqué chez le nématode puisque cet organisme nous permet de suivre la régulation des rapporteurs au cours du temps. Il est toutefois possible d'appliquer ce système chez des cellules en cultures par exemple, afin de suivre des événements dépendants de conditions ou de l'expression de régulateur particulier. Un outil adapté à l'étude de l'épissage alternatif et adaptable à différentes études et organismes est une avancée nécessaire à une meilleure compréhension de l'épissage alternatif de manière globale à long terme, et de manière ciblée dans le cas de l'étude d'un réseau de régulation particulier.

### III. Résultats obtenus

Le travail que j'ai effectué ces dernières années a permis de mettre au point au laboratoire une méthodologie adaptée à l'étude de l'épissage alternatif et de l'appliquer à plusieurs cas. Si après trois ans de travail, une étude poussée débouchant sur une publication scientifique n'a pas pu être finalisée complètement, c'est en raison de la nature de la régulation des gènes initialement choisis. En effet, la grande partie des travaux qui touchent aux transcrits alternatifs mettent en avant l'épissage alternatif comme un mécanisme tissu-spécifique simple : la présence d'un régulateur change l'isoforme produite par le gène. Cette vision simpliste est renforcée par le biais de publication. En effet, les cas les plus simples sont les plus facilement résolus et donc les plus rapidement publiables. Il se trouve donc que les résultats obtenus jusqu'à aujourd'hui vont dans le sens d'un contrôle fort des événements d'épissage alternatif mais cela pourrait ne pas refléter l'étendue de la complexité de la réalité biologique. Une publication récente issue du laboratoire du Dr Ian Hope a permis de mettre en doute cette vision de la régulation (Craig et al., 2013). Travaillant avec des rapporteurs « classiques » (**Introduction.IV.5.a**), ils ont dû créer plusieurs centaines de rapporteurs pour suivre tous les transcrits alternatifs produits par les 29 gènes qu'ils avaient sélectionnés. En effet, ce type de rapporteur ne permet de suivre les isoformes qu'une par une. Parmi les 29 gènes, de nombreux rapporteurs n'ont démontré aucune variation dans l'expression des transcrits alternatifs de chaque gène. Il est dommageable que le seul gène que nous avons en commun, *unc-86*, n'a pu être étudié en raison de données erronées sur le modèle du gène

utilisé dans leur étude. En effet, l'étude publiée à pris le modèle du gène *unc-86* présenté dans une précédente version de la base de donnée WormBase, modifiée depuis. Les rapporteurs qu'ils ont générés sur la base de ce modèle obsolète pour suivre ce gène n'ont permis de détecter aucune fluorescence.

D'autres rapporteurs ont pu être assemblés durant ma thèse pour les gènes *tax-6*, *unc-36* et *mlk-1* choisis de la même manière que *gpa-13*, *unc-86* et *che-1*, mais le manque de temps, lié à l'échéance, n'a pas permis d'obtenir des lignées pour étudier leurs profils d'expression. Ils pourront néanmoins servir à décrire de nouveaux événements.

Sur le long terme, le suivi de l'épissage alternatif permettra de résoudre la proportion d'événements d'épissage alternatif régulés et celle des événements stochastiques, les erreurs de prédiction bio-informatique des modèles d'épissage proposés ainsi que les régulations tissus-spécifique et leurs facteurs.

Les profils avec une isoforme majoritaire et une isoforme largement minoritaire exprimée dans les mêmes cellules (observés pour *unc-86* et *che-1*) pourraient ne pas correspondre à une situation physiologique. Il est possible que la forme correspondant à la dsRed dans ces deux cas soient promue par un système activateur en quantité limitante qui est saturé par le nombre de copies de transgène présents sur le concatémère extra-chromosomique.

A l'inverse, il est possible que la forme minoritaire représente un épissage stochastique qui ne se produit que dans une fraction des événements d'épissage et qui est trop rare pour être significatif mais qui est rendu visible en raison de la large surexpression de notre rapporteur.

Dans ces deux cas, notre lignée ne représente pas fidèlement la situation endogène, et il sera important de confirmer ces résultats avec notre stratégie d'insertion du transgène en copie unique.

Un événement d'épissage stochastique - absence de régulation active -, peut être lié à l'affinité des composantes du spliceosome pour les sites d'épissage en présence. Les séquences des sites d'épissage bordant les exons présentent de nombreuses variations qui peuvent influencer sur l'épissage (Farrer et al., 2002). Il est donc légitime de se demander quel impact ces variations ont sur l'inclusion de séquences dans le messager final, et si le changement de ses séquences peut modifier l'inclusion d'un exon constitutif. L'avantage du système rapporteur développé au laboratoire est qu'il permet de tester un certain nombre d'hypothèses sur les règles régissant le "splicing code".

Un rapporteur construit sur le même principe que ceux présentés dans les résultats mais constitué d'exons constitutifs serait nécessaire. Par mutagenèse dirigée, les sites d'épissages placés en bordure de l'exon constitutif pourraient être modifiés pour générer différentes constructions. Il serait alors possible de créer des lignées porteuses de rapporteurs qui ne diffèrent que par leurs sites d'épissage. Le cytomètre en flux du laboratoire peut analyser et quantifier les deux fluorescences sur un très grand nombre d'individus et pourrait permettre de comparer le ratio GFP/DsRed observé avec les différents rapporteurs (Amrane et al., 2014). Ce système permettrait de connaître la capacité de différentes séquences des sites d'épissages à permettre l'exclusion d'un exon constitutif. Les sites d'épissage des gènes *unc-86* et *che-1* pourraient alors être évalués au vu de ces données et permettre de déterminer si les résultats observés sont l'effet d'un événement stochastique.

Le cas de *gpa-13* est particulier puisque nous n'observons qu'une seule forme. Nous sommes face à un cas où il devient difficile de répondre sans vérifier la version endogène. Cependant, la forme courte théorique pourrait n'être qu'un artefact de prédiction ou être exprimée si faiblement qu'il serait virtuellement impossible de le confirmer à l'heure actuelle.

Un autre problème se pose alors : comment prouver la non-existence d'un transcrit alternatif théorisé ? Notre système ne nous permet pas de répondre définitivement à la question, puisqu'il est toujours possible qu'une isoforme très faiblement exprimé soit indétectable. Dans ce cas, la présence de la seconde isoforme doit être validée ou infirmée par une autre méthode. Seule une combinaison d'approches et la vérification des résultats par différentes méthodes peuvent permettre d'affiner les modèles d'épissage.

Il est cependant à nuancer l'impact d'une isoforme rare au point de passer inaperçue par certaines méthodes de détection. La représentation très faible d'un ARNm vis-à-vis d'un autre se ressentirait dans la quantité de protéines synthétisées. Cela n'aurait impact – autre qu'une diminution faible de l'expression - si l'isoforme rare ne différait de la forme majoritaire que par une simple perte de fonction. L'absence d'une activité de manière faible serait compensée par la forme majoritaire, ce qui ne serait pas le fait du cas inverse. Le gain, même faible, d'une activité pourrait avoir un rôle important dans l'organisme en apportant une fonction biologique. Un événement aussi rare pourrait donc n'avoir aucun impact sur l'organisme à l'échelle moléculaire.

Le rapporteur *top-1* présente bien des variations à la fois spatiales, temporelles et conditionnelles. Mais cette fois, un problème de toxicité a été observé avec notre rapporteur en multiples copies. Ce nouvel artefact potentiel m'a poussé à consacrer la fin de ma thèse à mettre en place un système d'insertion des constructions rapportrices en simple copie dans un

locus invariant de façon à éliminer ce type d'artefacts. Cela permettrait non-seulement d'étudier *top-1* grâce aux différentes constructions réalisées, mais également de proposer une solution à d'autres gènes et diverse études présentant un problème similaire.

Une fois cela réalisé, l'insertion des constructions *top-1* et *hrpf-1* pourra permettre de mettre en évidence leurs régulations et les éléments en *cis* impliqués. Nous disposons également de lignées mutantes pour plusieurs facteurs d'épissage ainsi qu'une banque de RNAi dédiée aux régulateurs potentiels. Ces outils permettront, avec des lignées stables, de caractériser les partenaires génétiques participant à la régulation de ces gènes.

Pour la séquence des triplets de G, l'observation des différentes constructions est nécessaire avant de déterminer leurs implications. Si ces séquences étaient validées, il serait alors possible de se pencher sur le lien entre HRPF-1 et ces motifs. Le laboratoire étant en collaboration active avec l'équipe du Dr Cameron Mackereth, il sera alors possible de déterminer *in vitro* l'affinité du facteur HRPF-1 pour les triplets de G.

Déterminer les facteurs impliqués dans la régulation de l'épissage alternatif est une étape essentielle dans l'étude des gènes. En effet, grâce à la connaissance des séquences en *cis* impliquées et les nouvelles techniques d'édition du génome, il est désormais possible de créer non plus des KO de gènes entiers mais des KO spécifiques d'isoformes, par l'élimination au locus des éléments impliqués dans l'exclusion ou l'inclusion d'exons. Ces KO pourraient permettre de déterminer les fonctions biologiques de chaque isoforme indépendamment des autres, et constituerait un pas en avant dans l'étude de la complexité du vivant. Le paradigme actuel repose sur l'étude de gènes comme unité des fonctions biologiques. L'étude des différentes isoformes n'étant pas systématique à l'heure actuelle. Cependant, la généralisation de l'utilisation d'outils adaptés pourrait modifier cet état des choses à terme.

Avant d'en arriver là, l'étude des séquences en *cis* est fondamentale. Celle de *top-1* est en cours et sera résolue lorsque les insertions génomiques de rapporteur auront été obtenues.

#### **IV. L'insertion universelle**

Le dernier obstacle à l'étude approfondie de *top-1* réside dans le problème posé par les lignées porteuses des différentes constructions. Les différentes hypothèses envisagées – effet dominant négatif, titration des régulateurs ou répression des copies endogènes – à mener à une solution unique : l'insertion en simple copie. Pour cela, nous avons imaginé un système simple d'utilisation basée sur la technologie CRISPR-Cas. Toujours en cours de réalisation expérimentale, il est indispensable à l'étude des différentes constructions réalisées mais son utilisation pourrait s'étendre à l'intégration de nombreux autres rapporteurs ne nécessitant pas d'intégration au locus d'intérêt.

Au-delà de la résolution du problème rencontré, l'utilisation du CRISPR-Cas et plus particulièrement l'intégration de nos rapporteurs ouvre la voie vers de nouvelles formes d'étude irréalisables jusqu'à ce jour.

# Perspectives



## I. Directes

Les perspectives à court terme correspondent à l'intégration des rapporteurs *top-1* et *hrpf-1*. L'observation de lignées stables permettra l'identification complète des cellules exprimant le rapporteur, l'impact des diverses séquences identifiées et l'identification des facteurs – si HRPF-1 et ASD-1/FOX-1 ne sont pas les seuls impliqués. La résolution de cette régulation pourra donner lieu à une publication similaire à celles issues des laboratoires du Dr Hidehito Kuroyanagi et du Dr John Calarco.

Le système de l'insertion universelle est une solution aux problèmes rencontrés mais également une méthode d'intérêt pour la communauté de *C. elegans*.

Des rapporteurs pour d'autres gènes ont également été assemblés et sont en attente d'injection. Les résultats obtenus permettront de dévoiler d'autres cas potentiels d'étude mais également d'autres cas similaires à *unc-86* et *che-1*. L'impact des séquences de sites d'épissage est cependant nécessaire à évaluer pour déterminer si ces cas peuvent être sous le contrôle du spliceosome seul. Ce projet repose sur des méthodologies mises en place au laboratoire, ce qui devrait faciliter cette étude.

## II. A long terme

L'intégration en simple copie des rapporteurs développés ici rend possible plusieurs nouvelles approches d'intérêt.

### 1) La recherche de séquence en cis par crible de mutagenèse ciblée

Parmi elles se trouvent une méthode potentielle pour identifier efficacement toutes séquences en *cis* impliquées dans une régulation. En effet, la recherche de ces éléments passe par deux méthodes : l'analyse de séquences conservées communes à plusieurs espèces et la recherche de motifs connus. La première est limitée à une conservation de la régulation, la seconde aux connaissances préalables dans le domaine. Lorsque ces analyses ne sont pas réalisables, seule reste la méthode de l'*intron bashing*, qui consiste à éliminer des portions entières des introns, puis à affiner l'identification par de nouvelles délétions dans le rapporteurs. Cette méthode repose donc sur un grand nombre de constructions et de lignées, rendues particulièrement difficiles dans des cas similaires à celui de *top-1*.

L'intégration du rapporteur permet, grâce aux technologies d'édition du génome, de cibler les introns du rapporteur directement. Cela aura pour effet de causer des coupures double-brins dans le rapporteur et le gène natif, qui seront réparées, apportant délétions ou additions à la séquence. L'avantage du rapporteur fluorescent est qu'il permettra alors de

cribler des individus présentant un profil de fluorescence différent. L'isolation de ces individus et le séquençage du rapporteur qu'ils portent permettra alors de récolter des données qui pourront permettre l'identification de sites particuliers en comparant les séquences mutées.

Les cribles de génétique classique – mutagenèse aléatoire du génome - et inverse – grâce aux banques d'ARN interférence – sont déjà en place au laboratoire. A ces cribles, pourront s'ajouter une nouvelle méthode qui, en combinaison des deux autres, pourra permettre d'étudier la régulation d'un gène sans avoir à se reposer sur les informations déjà connues.

## 2) Modification ciblée des marques post-traductionnelle

L'intégration en simple copie permettra également de développer un nouveau type d'outil, pour pouvoir étudier l'impact des modifications post-traductionnelles des histones sur l'épissage alternatif d'un rapporteur. En effet, la technologie du CRISPR-Cas repose sur le ciblage d'une protéine à un locus déterminé par un ARN guide. Il est donc envisageable qu'à terme, la Cas9 utilisée soit modifiée pour ne plus induire de coupure et porte d'autres domaines d'activités comme cela a déjà pu être réalisé Deux système ont vu le jour : la répression de transcription grâce à une *nuclease-dead cas9* débarrassée de son activité endonucléolytique, nommée le CRISPRi pour CRISPR interférence (Qi et al., 2013) ; l'activation transcriptionnelle grâce à une *nuclease-dead cas9* également débarrassée de son activité endonucléolytique mais fusionnée à un domaine d'activation de transcription, nommée le CRISPR-on (Cheng et al., 2013).

A partir de là, il est possible d'envisager l'utilisation d'un domaine de méthylation ou d'acétylation dans la construction de Cas9 chimériques. Il serait donc possible, en théorie, de cibler des modifications particulières aux histones environnant le locus, et ainsi de visualiser l'impact de ces modifications sur l'épissage observé grâce au système rapporteur mis en place.

L'édition du génome de manière ciblée, en couplage avec le rapporteur développé ici, présentent de nombreuses voies d'études futures de l'épissage alternatif et de ses divers composantes.

L'ensemble des méthodes mises en place et des méthodes théorisées devrait à long terme faciliter l'étude de l'épissage alternatif au cas par cas. En combinaison avec les d'autres méthodes de détection, les données accumulées permettront de découvrir la nature détaillée du « splicing code ».

# Matériel & Méthodes



# I. Milieux de cultures et solutions

## 1. Solutions

Tampon de réaction ISO 5x : 3mL de solution de Tris-HCl à 1mol/L, 150µL de solution de MgCl<sub>2</sub> à 2mol/L, 60µL de solution de dNTP (NEB) à 100mmol/L, 300µL de solution de DTT à 1mol/L, 1.5g de PEG-8000, 300µL de solution de NAD à 100mmol/L, H<sub>2</sub>O qsp 6mL.

Le tampon de réaction ISO 5x est utilisé dans le protocole d'assemblage par la méthode de Gibson.

Mélange Enzyme-Réactifs : 320µL de Tampon de réaction ISO 5x, 0.64 d'exonucléase T5 (Epicentre) à 10U/µL, 20µL de polymérase Phusion (NEB) à 2U/µL, 160µL de Taq ligase (NEB) à 40U/µL, H<sub>2</sub>O qsp 1.2mL.

Le mélange Enzyme-Réactifs est utilisé dans le protocole d'assemblage par la méthode de Gibson.

M9 solution : 3g de KH<sub>2</sub>PO<sub>4</sub>, 6g de Na<sub>2</sub>HPO<sub>4</sub>, 5g de NaCl, de 1ml MgSO<sub>4</sub> à 1mol/L, H<sub>2</sub>O qsp 1L.

Le milieu M9 est une solution permettant la survie des vers et utilisée pour manipuler des populations (Préparation d'ADN génomique, régénération après injection, préparation de stock glycérol).

Solution filtré à l'aide d'unité de filtration.

Tampon Potassium Phosphate : 132mL de solution de K<sub>2</sub>HPO<sub>4</sub> à 1mol/L, 868mL de solution de KH<sub>2</sub>PO<sub>4</sub> à 1mol/L.

Solution filtré à l'aide d'une unité de filtration.

Solution de congélation : 0.58g de NaCl, 0.68g de KH<sub>2</sub>PO<sub>4</sub>, 30g de glycérol, H<sub>2</sub>O qsp 50mL.

La solution de congélation permet la conservation longue durée de lignées de *Caenorhabditis elegans* à une température de -80°C.

Solution de synchronisation : 2.5mL de NaOH à 2M, 2mL de NaClO à 4%, H<sub>2</sub>O qsp 5mL.

Réalisée extemporanément.

Solution TB (Transformation Broth) : 3.02g de Pipes, 2.21g de CaCl<sub>2</sub>, 16.64g de KCl, H<sub>2</sub>O qsp 900mL. pH ajusté à 6.7. 10.89g de MnCl<sub>2</sub> est ajouté, et le volume ajusté à 1L.

Solution filtré à l'aide d'une unité de filtration.

## 2. Milieux de cultures

Milieu LB (Luria Broth) : 10g Tryptone, 5g d'extrait de levure, 10g NaCl et H<sub>2</sub>O qsp - quantité suffisante pour 1 litre - 1L. Pour obtenir du milieu LB solide, 12g d'Agar sont rajoutés pour 1L de milieu.

Milieu autoclavé avant utilisation.

Pour rendre le milieu de culture sélectif, de l'ampicilline ou de la kanamycine est rajouté à hauteur de 100 µg/mL et 50 µg/mL respectivement. Dans le cadre des expériences de ARNi, le LB est supplémenté d'ampicilline (100 µg/mL) et de tétracycline (15 µg/mL). 1mM d'IPTG (isopropyl β-D-1-thiogalactopyranoside) est ajouté.

Le milieu LB est un milieu de culture liquide ou solide riche utilisé pour les cultures bactériennes.

Milieu SOB (Super Optimal Broth) : 20g de Bactotryptone, 5g d'extrait de levure, 2mL de solution de NaCl à 5mol/L, 2.5mL de solution de KCl à 1mol/L, 10mL de solution de MgCl<sub>2</sub> à 1mol/L, 10mL de solution de MgSO<sub>4</sub> à 1mol/L, H<sub>2</sub>O qsp 1L.

Milieu autoclavé avant utilisation.

Le milieu SOB est un milieu riche utilisé lors du protocole d'obtention de bactéries thermocompétentes.

Milieu NGM (Nematoda Growth Medium) : 3g NaCl, 2.5g Peptone, 12g Agar, 5mg Cholestérol et H<sub>2</sub>O qsp 975 µL. Après passage à l'autoclave, 1mL CaCl<sub>2</sub> à 1mol/L filtré, 1mL MgSO<sub>4</sub> à 1mol/L filtré, 25mL Tampon Potassium Phosphate à 1mol/L filtrés sont rajoutés au milieu. Pour rendre le milieu sélectif, du G418 est ajouté à hauteur de 0.40 g/L (Ref Rosina).

Après 24h, une solution d'*Escherichia coli* (souche OP50 ou OP50-neoR) mis en culture dans du LB liquide est déposé sur les boîtes de NGM. Les boîtes sont incubées à température ambiante pendant 24h.

Le milieu NGM est un milieu de culture permettant le maintien de lignées de *Caenorhabditis elegans*.

Les produits suivants proviennent du fournisseur BD<sup>TM</sup> : Agar, Bactotryptone, Extrait de levure, Peptone.

Les produits suivants proviennent du fournisseur FischerScientific<sup>TM</sup> : Ampicilline, Kanamycine, Tetracycline, CaCl<sub>2</sub>, KCl, KH<sub>2</sub>PO<sub>4</sub>, K<sub>2</sub>HPO<sub>4</sub>, MgSO<sub>4</sub>, MnCl<sub>2</sub>, NaCl, Na<sub>2</sub>HPO<sub>4</sub>, Nalgene<sup>TM</sup> Rapid-Flow<sup>TM</sup> Sterile Disposable Filter.

Les produits suivants proviennent du fournisseur AcrossOrganics<sup>TM</sup> : Agarose, Cholestérol, IPTG, NaOH, NaClO.

Les produits suivants ont été empruntés à d'autres équipes de recherche : Pipes (Équipe McCusker), NAD, PEG-8000, Tris-HCl et DTT (Équipe Teichmann).

## II. Biologie moléculaire

### 1. Préparation d'ADN génomique

Une fois la source de nourriture épuisée sur une boîte recouverte de vers de la souche N2 (Bristol) - Boîte de Pétri de 9cm de diamètre -, les vers sont récupérés dans 15mL d'eau distillée puis centrifugés 2 minutes à 400g. Deux autres lavages sont réalisés dans les mêmes conditions (15mL de dH<sub>2</sub>O ; Centrifugation 2' - 400g). Le surnageant est éliminé. L'ADN génomique des vers est extrait à l'aide du kit NucleoSpin Tissue (Macherey-Nagel). Les vers sont resuspendus dans du tampon de lyse (Tampon T1) puis congelés à -20°C pendant une heure. L'échantillon est ensuite décongelé et 25µL de protéinase K à 22.5mg/mL sont ajoutés. L'échantillon est incubé à 56°C pendant la nuit. Une fois les vers lysés, la récupération de l'ADN génomique est effectuée grâce à une purification sur colonne de silice en suivant les instructions du fournisseur. L'élution finale est réalisée dans 100µL d'eau milliQ (pré-chauffé à 70°C pour améliorer le rendement de la purification). La quantité d'ADN est mesurée par spectrophotométrie à l'aide d'un NanoDrop (ThermoScientific) et stockés à 4°C. La concentration est ajustée par ajout d'H<sub>2</sub>O milliQ jusqu'à une concentration de 50ng/µL.

### 2. Extraction d'ARN

Une fois la source de nourriture épuisée sur une boîte recouverte de vers de la souche N2 (Bristol) - Boîte de Pétri de 9cm de diamètre -, les vers sont récupérés dans 15mL d'eau distillée puis centrifugés 2 minutes à 400g. Deux autres lavages sont réalisés dans les mêmes conditions (15mL de dH<sub>2</sub>O ; Centrifugation 2' - 400g). Le surnageant est éliminé.

Les vers sont resuspendus dans 1mL de Trizol (Sigma-Aldrich) et le mélange est homogénéisé par aspiration-refoulement à la pipette. Le tube est ensuite congelé dans de l'azote liquide et décongelé à 37°C à deux reprises. 200µL de chloroforme (Sigma-Aldrich) et 25µL de glycogène à 20mg/mL (ThermoScientific) sont ajoutés au mélange, puis homogénéisé vigoureusement par agitation manuelle pendant 15 secondes. Le mélange est incubé à 4°C pendant 15 minutes avant d'être centrifugé à 10000g pendant 10 minutes à 4°C.

La phase aqueuse est récupérée. Un volume de trizol et 1/10 de volume de chloroforme sont ajoutés. Le mélange est agité vigoureusement pendant 15 secondes avant d'être incubé à 4°C pendant 10 minutes et centrifugé à 10000g pendant 10 minutes à 4°C.

La phase aqueuse est récupérée et un volume de chloroforme est ajouté. Le mélange est homogénéisé vigoureusement pendant 15 secondes et incubé 5 minutes à 4°C avant d'être centrifugé à 10000g pendant 5 minutes à 4°C.

La phase aqueuse est récupérée et un volume d'isopropanol est ajouté. Le mélange est homogénéisé vigoureusement pendant 15 secondes et incubé 10 minutes à température ambiante avant d'être placé à -20°C pour la nuit.

Le mélange est centrifugé à 13000g pendant 30 minutes à 4°C. Le surnageant est éliminé, le culot est rincé à l'éthanol 75% et centrifugé à 7500g pendant 10 minutes à 4°C. Le surnageant est éliminé et le culot séché à l'air libre. Le culot est repris dans 50µL d'eau distillée.

L'extrait d'ARN totaux est traité à la DNase (4 Unités de Turbo DNase - Ambion), en présence de tampon DNase, pendant 25 minutes à 37°C. La DNase est inactivée par réactif d'inactivation pendant 2 minutes à température ambiante. Le mélange est centrifugé à 10000g pendant 5 minutes à 4°C. Le surnageant est récupéré, dosé par spectrophotométrie (NanoDrop - ThermoScientific) et la qualité des ARNs est analysée par puce Agilent (Agilent RNA 6000 Nano kit). Les échantillons sont conservés lorsque le RIN (RNA Integrity Number) est supérieur ou égale à 8. L'étape de reverse transcription est réalisée immédiatement et les extraits d'ARNs totaux sont conservés à -80°C.

### **3. Amplification par PCR**

#### **a) Conditions de PCR**

Les fragments d'ADN nécessaires aux assemblages sont amplifiés par PCR en utilisant comme matrice 50ng d'ADN génomique (*C.elegans* - N2 bristol) ou 10pg de plasmide, 200µmol/L de dNTPs (NEB), 0,5µmol/L de chaque amorces, 1 Unité de polymérase Phusion (NEB) ou de Q5 (NEB), 10µL de tampon associé à la polymérase utilisée, dilué dans de l'eau milliQ pour un volume réactionnel final de 50µL. Les réactions d'amplification sont réalisées dans un thermocycleur avec les étapes suivantes :

Cycle	Étape	Température	Durée
1x	Dénaturation initiale	98°C	30''
35x	Dénaturation	98°C	10''
	Hybridation	Fonction des amorces	20''
	Élongation	72°C	30''/kb
1x	Élongation finale	72°C	5'

Chaque amplification est réalisée en parallèle d'un contrôle négatif dans les mêmes conditions en remplaçant la matrice par 1µL d'eau milliQ.

### **b) Électrophorèse**

5µL de chaque réaction sont additionnés de bleu de charge 6x (Promega) et chargés dans un gel de TAE0.5x-Agarose 1% contenant du bromure d'éthidium à une concentration final de 0.5µg/mL. L'électrophorèse est conduite à 100V pendant 30 à 40 minutes dans une cuve de TAE0.5x et le résultat est visualisé grâce à un Trans-illuminateur UV. La taille des produits de digestion est estimée en comparaison d'un marqueur de taille 1kb Ladder (Promega). Les réactions sont conservées lorsque le contrôle négatif ne présente aucune bande et l'amplification présente une seule bande à la taille attendue.

### **c) Purification des produits de PCR**

Les réactions positives sont purifiées à l'aide du kit PCR Purification (QIAGEN) en suivant les instructions données par le fournisseur. L'élution finale est réalisée dans 30µL de tampon EB. La concentration en ADN des amplifications est dosée par spectrophotométrie (NanoDrop-ThermoScientific). Les échantillons sont considérés comme suffisamment purs et stockés à 4°C lorsque le résultat du ratio A260/A280 se situe entre 1.8 et 2.0, le résultat du ratio A260/A230 entre 2.0 et 2.4 et la concentration supérieure à 20ng/µL.

## 4. RT-PCR

La synthèse des ADNs complémentaires est réalisée avec 1 µg d'extrait d'ARN totaux en présence de 100pmol d'amorce oligo-dT<sub>23</sub>, de 20 Unités de Murine RNase Inhibitor, 10nmol de dNTP, 200 Unités de M-MuLV Reverse Transcriptase dans 20µL d'eau (nuclease-free) en présence de tampon de RT. Les amorces et ARNs sont dénaturées à 70°C pendant 5 minutes. La réaction s'opère à 42°C pendant 1 heure. Les ADNc sont conservés à -20°C.

Les amplifications sont réalisées dans les conditions suivantes :

100ng d'ADN complémentaire (issus des extractions d'ARN), 200µmol/L de dNTPs (NEB), 0,5µmol/L de chaque amorces, 1 Unité de polymérase Taq Thermopol (NEB), 10µL de tampon associé à la polymérase utilisée, le tout dilué dans de l'eau milliQ pour un volume réactionnel final de 50µL. Les réactions d'amplification sont réalisées dans un thermocycleur avec les étapes suivantes :

Cycle	Étape	Température	Durée
1x	Dénaturation initiale	95°C	2'
30x	Dénaturation	95°C	20''
	Hybridation	Fonction des amorces	30''
	Élongation	68°C	1'/kb
1x	Élongation finale	72°C	5'

Chaque amplification est réalisée en parallèle d'un contrôle négatif dans les mêmes conditions en remplaçant la matrice par 1µL d'eau milliQ. Le contrôle positif de RT-PCR est une amplification de 327pb sur un extrait d'ADNc issu de la RT d'un extrait d'ARN totaux de foie de Rat (NEB). Le contrôle négatif est réalisé à partir d'un extrait d'ARN non-traité à la Reverse-Transcriptase.

## 5. Assemblages des vecteurs

### **a) Assemblage par la méthode de Gibson**

Les fragments à assembler sont mélangés à hauteur de 25Fmol (vecteur linéarisé) et 75 Fmol (inserts), à 4°C, et complétés par de l'H<sub>2</sub>O pour un volume réactionnel final de 5µL. La conversion mol-gramme est effectué en prenant en compte le nombre exacte de nucléotide de chaque fragment et en utilisant la formule mathématique suivante :

$$\text{Quantité ADN (Fmol)} \times 660 \text{ (pg/mol)} \times 10^{-6} \times n = \text{Quantité ADN (ng)}$$

(n étant le le nombre de paires de nucléotides du fragment et 660pg/mol étant le poids moléculaire moyen d'une paire de nucléotides). 15µL de mélange Enzyme-Réactifs sont ajoutés au 5µL de volume réactionnel. Le tube est incubé à 50°C pendant une heure (Gibson et al., 2009).

### **b) Assemblage des vecteurs sgARN**

Un mélange contenant 150ng de pMB70, 500nmol/L d'amorces, 10 Unités de BsaI, 20 Unités de Quick ligase sont dilués dans 20µL d'eau milliQ additionnée de tampon de T4 ligase est placé dans un thermocycleur dans les conditions suivantes :

Cycle(s)	Température	Temps
<b>10x</b>	37°C	5'
	16°C	10'
<b>1x</b>	50°C	15'
	80°C	5'

## **6. Mutagenèse dirigée par PCR**

Ing de matrice est amplifié avec les amorces de mutagénèse dirigée, et l'amplification est vérifiée, en utilisant les conditions données dans la section B.3.1 du Matériel & Méthode.

Les amplifications vérifiées sont ensuite purifiées (B.3.2) dans 30µL de tampon EB puis digérées par DpnI et NdeI en ajoutant 3.5µL de tampon NEB4 pendant 2h à 37°C. 3µL de tampon T4 ligase et 1000 unités de Quick Ligase sont ajoutées. Le mélange est incubé à 25°C pendant 5 minutes.

## 7. Transformation

### **a) Obtention de bactéries électro-compétentes**

Une pré-culture de 10mL de LB liquide est inoculée de bactéries Escherichia coli - souche DH5α - est incubée sur la nuit à 37°C. Une culture d'un litre de LB liquide estensemencée avec la pré-culture de la veille. La densité optique (D.O.) est mesuré toutes les demi-heures jusqu'à obtention d'une valeur de 0.35 à 0.4. La culture est ensuite placé à 4°C pendant 30 minutes. La culture est séparé dans des pots de centrifugation de 250mL de contenance.

Toutes les centrifugations sont opérées à 1000g pendant 20 minutes à 4°C. La culture est centrifugée à deux reprises, les surnageants sont éliminés et les culots repris dans de l'eau distillée à 4°C (200mL puis 100mL par pot).

Les cultures sont rassemblées en deux pots de 200mL et centrifugé une troisième fois. Les culots sont resuspendus dans 40mL d'eau distillée à 4°C et placés dans des tubes à fond conique de 50mL avant d'être à nouveau centrifugés. Les culots sont repris dans 1mL de glycérol 10% à 4°C et aliquotés. Les aliquotes de 50µL sont conservées à -80°C.

L'absence de contamination à été vérifiée en transformant 1µL d'eau distillée, les colonies sont dénombrés après incubation 24h à 37°C et le lot est considéré comme non-contaminé si aucune colonie ne s'est développée sur LB+Ampicilline. L'efficacité de transformation est estimée par aliquote en transformant 10pg de plasmide pBS (pBlueScript). Après incubation 24h à 37°C, le nombre de colonies présent sur boîte LB+Ampicilline est dénombré, et la formule suivante est appliquée :

**Efficacité de transformation = Nb de colonies/Qte d'ADN(ng) x 1000(ng/ug)**

### **b) Transformation par électroporation**

1µL de produits d'assemblage est utilisé pour transformer 50µL de bactéries électro-compétentes de l'espèce Escherichia coli - souche DH5α ou DH10B - par électroporation, à l'aide d'un électroporateur GenePulser (BioRad). Les aliquotes de bactéries électro-compétentes utilisées provenaient de différents lots et présentaient tous une efficacité de transformation de l'ordre de 10<sup>-6</sup>.

Les mélanges bactérie-ADN sont ensuite étalées sur boîtes LB+antibiotique(s). Les boîtes sont incubées sur la nuit à 37°C.

### **c) Obtention de bactéries thermo-compétentes**

Une pré-culture de 12mL de SOB liquide estensemencée de bactéries Escherichia coli - souche DH5α - et incubée sur la nuit à 37°C. Une culture d'un litre de milieu SOB liquide estensemencée avec la pré-culture de la veille et est incubée sur la nuit à 18°C. L'absorbance est mesurée toutes les heures le lendemain jusqu'à obtention d'une valeur de 0.6. La culture est placée à 4°C pendant 10 minutes et centrifugée à 2500g pendant 10 minutes à 4°C. Le surnageant est éliminé et le culot est resuspendu dans 25mL de tampon TB. 295mL sont ajoutés après homogénéisation. La culture est placée à 4°C pendant 5minutes avant une nouvelle centrifugation à 2500g pendant 10 minutes à 4°C. Le surnageant est à nouveau éliminé et le culot est resuspendu dans 40mL de tampon TB. 3mL de DMSO est ajouté (Cf dMSO : 7%). Les bactéries sont aliquotés par 100µL dans des tubes qui sont ensuite immédiatement congelés en les trempant dans de l'azote liquide avant d'être stockés à -80°C.

L'absence de contamination à été vérifiée en transformant 5µL d'eau distillée, les colonies sont dénombrés après incubation 24h à 37°C et le lot est considéré comme non-contaminé si aucune colonie ne s'est développée sur LB+Ampicilline, LB+Kanamycine et LB+Tetracycline. L'efficacité de transformation est estimée de la même manière que pour les bactéries électro-compétentes.

#### **d) Transformation par choc thermique**

5µL de produits d'assemblage, ou de dH<sub>2</sub>O - dans le cadre des contrôles négatifs -, est utilisé pour transformer 20µL de bactéries thermo-compétentes de l'espèce *Escherichia coli* - souche DH5α - par choc thermique.

Les aliquotes de bactéries thermo-compétentes utilisées provenaient du même lot - réalisé en mai 2014 - et présentaient une efficacité de transformation de l'ordre de 10<sup>-7</sup>.

Le choc thermique est effectué en plaçant les mélanges bactérie-ADN pendant 30 minutes à 4°C, puis placés à 42°C pendant 1 minute suivi par un passage à 4°C pendant dans 5 minutes, avant d'être étalés au râteau sur une boîte LB+antibiotique(s) solide.

Les boîtes sont ensuite incubées à 37°C sur la nuit.

### **8. Cultures bactériennes**

Après incubation, le nombre de colonies sur boîte est dénombré et 18 colonies - ou toutes les colonies si le nombre de colonies était inférieur à 18 - sont mises en culture dans 2mL (Miniprep), ou 25mL (Midiprep), de LB+antibiotique(s) chacune et incubées pendant 24h.

### **9. Extraction de plasmide**

#### **a) Extraction de plasmide par Miniprep**

L'extraction de plasmide est réalisé par lyse alcaline en utilisant le kit QIAprep Spin Miniprep (QIAGEN) en suivant les instructions du fournisseur.

1,5mL de culture bactérienne est centrifugé à 6800G pendant 3 minutes. Le surnageant est éliminé et le culot est resuspendu dans 250µL de tampon P1. 250µL de tampon P2 sont ajoutés, mélangé par retournement, et 350µL de tampon N3 sont rajoutés avant homogénéisation par retournement. Le mélange est centrifugé 10 minutes à 18 000G, et le surnageant est placé dans une colonne du kit, puis centrifugé 1 minute à 18 000G.

Un premier lavage est réalisé avec 500µL de tampon PB, et un second lavage est réalisé avec 750µL tampon PE. Une nouvelle centrifugation est réalisé pour éliminer les restes de tampon. La colonne est alors placé dans un nouveau tube, et l'élution finale est réalisé avec 50µL de tampon EB.

(Les lavages et l'élution nécessitent une étape de centrifugation d'une minute à 18 000G)

### **b) Extraction de plasmide par Midiprep**

Les souches portant les plasmides pBS, pBS(FSS::3'UTR<sub>unc-119</sub>), pJM001, pJM002, pJM003, pJM004, pRG02, pJL43, pBS(P<sub>eft-3</sub>::Cas9::SV40-NLS::3'UTR<sub>tbb-2</sub>), pMB67, pMB70, pMB78, pMB79, pRG5271Neo et pJJR50 sont cultivées dans 25mL de LB+antibiotique(s).

L'extraction de plasmide est réalisé par lyse alcaline en utilisant le kit Plasmid Plus Purification (QIAGEN) en suivant les instructions du fournisseur.

25mL de culture bactérienne sont centrifugés à 6800G pendant 3 minutes. Le surnageant est éliminé et le culot est resuspendu dans 2mL de tampon P1. 2mL de tampon P2 sont ajoutés, l'ensemble mélangé par retournement, et 2mL de tampon S3 sont rajoutés avant homogénéisation par retournement. Le mélange est incubé sur cartouche filtrante QIAfiltrer pendant 10 minutes. Le lysat est filtré et 2mL de tampon BB sont ajoutés. Le mélange est homogénéisé par inversion et filtré dans une colonne fournie avec le kit à l'aide d'une pompe à vide.

Un premier lavage est réalisé avec 700µL de tampon ETR, et un second lavage est réalisé avec 700µL tampon PE. Une nouvelle centrifugation est réalisé pour éliminer les restes de tampon. La colonne est alors placé dans un nouveau tube, et l'élution finale est réalisé avec 200µL de tampon EB.

(Les lavages et l'élution nécessitent une étape de centrifugation d'une minute à 18 000G)

### **c) Dosage des plasmides**

Les plasmides sont alors dosés par spectrophotométrie (NanoDrop-ThermoScientific) en suivant les règles décrites dans la partie Amplification par PCR et conservés à 4°C.

## 10. Vérification des plasmides

### **a) Vérification par restriction**

Les plasmides sont vérifiés par digestion en utilisant les enzymes de restriction suivantes : ApaI, BamHI-HF, EcoRI-HF, EcoRV, HindIII et PvuII-HF (NEB) en suivant les indications fournies par le fournisseur. 250ng sont digérés avec 1U d'enzyme dans le tampon conseillé pendant une heure à 37°C.

Les produits de digestion sont visualisés après électrophorèse (30minutes - 100V) en gel d'agarose 1% (TAE 0.5x, Bromure d'éthidium) grâce à un Trans-illuminateur UV. La taille des produits de digestion est estimée en comparaison d'une échelle 1kb (Promega).

Les plasmides présentant un bon profil de restriction sont conservés pour être vérifiés par séquençage.

### **b) Vérification par amplification**

Les plasmides pMB70<sub>(cible)</sub> sont vérifiés par amplification via l'amorce M13 Forward et l'amorce sens utilisée durant l'assemblage en utilisant 1ng de plasmide à vérifier et en suivant les conditions données dans la section B.3.1 du Matériel & Méthode.

### **c) Vérification par séquençage**

Le séquençage par la société GATC-Biotech dans les conditions suivantes : 40 à 50ng/μL de plasmide et 2.5μmol/L d'amorce dans un volume de 10μL. Les amorces utilisées recouvrent chacune 800pb et sont choisies pour couvrir l'intégralité des constructions assemblées.

Les chromatogrammes sont lus avec le logiciel 4Peaks, et les séquences sont comparées via le logiciel ApE (A plasmid Editor).

Les plasmides dont les constructions ont été vérifiées sont stockés à -20°C.

## **11. Stock glycérol**

Les souches bactériennes porteuses des plasmides vérifiées sont stockées à -80°C dans du LB+Glycérol (16%).

### III. Cultures et manipulation des vers

#### 1. Conditions de culture

Les vers sont cultivés sur boîtes de milieu NGM ou NGM+G418 recouvert d'un tapis bactérien de *Escherichia coli* souche OP50 ou OP50 neoR, dans un incubateur à 20°C ou 15°C. Les vers sont transférés vers de nouvelles boîtes tous les 3 à 4 jours.

#### 2. Génération de lignées transgéniques

##### a) Mélanges d'injection

Les mélanges d'injection de type concatémère répétitifs se composent de 10ng/μL de pRG02 et 90ng/μL de plasmide rapporteur.

Les mélanges d'injection de type concatémère complexes se composent de 10ng/μL de pRG02, 10ng/μL de plasmide rapporteur et 80ng/μL d'ADN génomique (*C.elegans*-N2 bristol) digéré par PvuII-HF (NEB).

Les mélanges d'injection permettant l'insertion en copie unique par MosSCI se composent de 45ng/μL de plasmide rapporteur, de 45ng/μL de pJL43 (Mos Transposase) et de pCB101 (marqueur de co-injection).

La première version des mélanges d'injection permettant l'insertion par CRISPR-Cas9 se composent de 50ng/μL de vecteur pMB70<sub>(cible)</sub> (sgARN), de 50ng/μL de pBS(P<sub>eft-3</sub>::Cas9::SV40-NLS::3'UTR tbb-2), de 40 ng/μL de plasmide rapporteur.

La seconde version des mélanges d'injection permettant l'insertion par CRISPR-Cas9 se composent de 50ng/μL de vecteur pMB70<sub>(cible)</sub> (sgARN), de 50ng/μL de pBS(P<sub>eft-3</sub>::Cas9::SV40-NLS::3'UTR tbb-2), de 50 ng/μL de séquence à insérer, 25ng/μL de pMB70(dpy-10) ainsi que de 20ng/μL d'oligonucléotide (ssODN) dpy-10(cn64).

Les mélanges d'injection permettant la délétion par CRISPR-Cas9 se composent de 50ng/μL de vecteur pMB70<sub>(cible)</sub> (sgARN) et de 50ng/μL de pBS(P<sub>eft-3</sub>::Cas9::SV40-NLS::3'UTR tbb-2).

## **b) Injection**

Les mélanges d'injection sont centrifugés à 10 000g pendant 30 minutes avant chaque séance d'injection pour culotter d'éventuelles particules qui pourraient boucher l'aiguille. 5  $\mu$ L de mélange d'injection sont chargés dans un capillaire de micro-injection (Eppendorf) branché à un injecteur FemtoJet (Eppendorf). La micro-injection est visualisée grâce à un microscope (Nikon Eclipse Ti). L'injection est réalisée dans la partie syncytiale d'une des deux gonades d'individu jeune adulte (individu porteur d'1 à 6 embryons) immobilisé dans une goutte d'huile sur une lame recouverte d'un tapis d'agarose à 5%. Cinq adultes P<sub>0</sub> injectés sont déposés par boîtes NGM+G418. La sélectivité du milieu NGM+G418 est vérifiée en déposant cinq individus n'ayant pas subi d'injection sur une boîte provenant du même lot.

## **c) Sélection F1 et F2**

A partir des boîtes P<sub>0</sub>, tous les individus F<sub>1</sub> - descendants direct des P<sub>0</sub> injectés - sont isolés sur de nouvelles boîtes, à raison de 10 vers par boîte.

A partir des boîtes F<sub>1</sub>, les individus F<sub>2</sub> - descendants direct des F<sub>1</sub> - sont isolés individuellement sur de nouvelles boîtes.

A partir de chaque boîte de F<sub>1</sub>, une seule lignée indépendante peut être générée et est conservée.

## **d) Conservation des lignées**

Les vers sont récupérés avec 1mL de solution M9. 1mL de Solution de congélation est ajouté. Les 2mL sont séparés en quatre tubes de 500 $\mu$ L, qui sont ensuite placés à -80°C.

## **3. Observation au microscope**

Les vers sont déposés sur des lames recouvertes d'un tapis d'agarose à 5%, dans une goutte d'azoture de sodium (AcrossOrganics) à 10mmol/L ou du levamisole (AcrossOrganics)

à 1mmol/L, afin de les anesthésier, et recouverts d'une lamelle. Les lames sont observés à l'objectif 40x ou à l'objectif à immersion 63x avec une goutte d'huile par DIC (Differential Interference Contrast Microscopy - microscopie basée sur l'interférence de deux faisceaux lumineux), puis dans les canaux d'observation de la GFP (excitation à 395nm) et de la DsRed (excitation à 558nm).

#### 4. Analyse au COPAS

Les boîtes sont analysées au COPAS lorsqu'elles contiennent une population dense mais avant que la source de nourriture ne soit épuisée. Les vers sont resuspendus dans de l'eau distillée, et la suspension est versée dans la coupe à échantillon du cytomètre. La suspension est diluée lorsque le nombre de vers analysés par seconde est égale à 300 (limite technique). Les paramètres suivants sont utilisés :

The image shows a software interface for COPAS with the following settings:

Gains		Thresholds	
	Signal	Signal	TOF Minimum
EXT	1.6	30	20
Green	3		
Yellow	3		
Red	3		
Trigger	EXT	Close	

PMT Control

Channel	Wavelength (nm)
Green	700
Yellow	700
Red	900

Les données générées sont ensuite traitées par un logiciel mis au point au laboratoire.

## 5. Synchronisation

Les boîtes contenant une population dense (un jour avant que la source de nourriture ne soit épuisée) sont utilisées pour la synchronisation. Les vers sont récupérés dans 15mL de solution M9 et centrifugés 2 minutes à 400g. Trois étapes de lavage sont effectuées en reprenant les vers dans 15mL de M9 avant d'être centrifugés 2 minutes à 400g. Après les surnageant du dernier lavage est éliminé, les vers sont repris dans 1mL d'une solution de synchronisation - réalisée extemporanément. Les tubes sont agités énergiquement pendant 5 minutes. La réaction est arrêtée par ajout de 14mL de M9 et les tubes centrifugés 1 minute à 400g. Le surnageant est éliminé et trois nouvelles étapes de lavage sont effectuées en reprenant les vers dans 15mL de M9 avant d'être centrifugés 2 minutes à 400g. Le dernier surnageant est éliminé et les vers sont repris dans un volume final de 1mL de M9.

Les vers peuvent être dénombrés par dépôt de 10 $\mu$ L sur boîte.

Les vers sont ensuite, soit, incubés à 20°C avec agitation rotative (pour éviter l'agglutination et la mort des vers) pendant 24h et déposés sur milieu - meilleure synchronisation -, soit, déposés immédiatement sur milieu.

**(Porta-de-la-Riva *et al.* 2012, JoVE)**

## 6. ARN interférence

Les souches bactériennes utilisées lors des expériences d'ARN interférence proviennent toutes de la banque ORFeome (Ref ORFeome). La souche exprimant l'ARNi dirigé contre le gène *dpy-1* est utilisée comme contrôle d'expérience.

Les souches bactériennes sont incubées pendant à 24h à 37°C dans 2mL de LB+Amp+Tet. 20 $\mu$ L des cultures sont alorsensemencés dans 2mL de LB+Amp+Tet+IPTG et incubées à 37°C pendant 24h.

Les cultures sont alors déposées sur boîte NGM+Amp+Tet+IPTG, et les boîtes de NGM sont incubées à 20°C pendant 24h.

Dix vers hermaphrodite au stade L4 sont alors placés sur boîte. Une semaine après les vers sont visualisés à la loupe à fluorescence ou au microscope.

# Annexes



## Bibliographie

- Ambros, V., and Horvitz, H.R. (1984). Heterochronic mutants of the nematode *Caenorhabditis elegans*. *Science* 226, 409–416.
- Amrane, S., Reborá, K., Zniber, I., Dupuy, D., and Mackereth, C.D. (2014). Backbone-independent nucleic acid binding by splicing factor SUP-12 reveals key aspects of molecular recognition. *Nat. Commun.* 5, 4595.
- Amrani, N., Ganesan, R., Kervestin, S., Mangus, D. a, Ghosh, S., and Jacobson, A. (2004). A faux 3'-UTR promotes aberrant termination and triggers nonsense-mediated mRNA decay. *Nature* 432, 112–118.
- Anczuków, O., Rosenberg, A.Z., Akerman, M., Das, S., Zhan, L., Karni, R., Muthuswamy, S.K., and Krainer, A.R. (2012). The splicing factor SRSF1 regulates apoptosis and proliferation to promote mammary epithelial cell transformation. *Nat. Struct. Mol. Biol.* 19, 220–228.
- Anders, K.R., Grimson, A., and Anderson, P. (2003). SMG-5, required for *C.elegans* nonsense-mediated mRNA decay, associates with SMG-2 and protein phosphatase 2A. *EMBO J.* 22, 641–650.
- Ando, S., Sarlis, N.J., Krishnan, J., Feng, X., Refetoff, S., Zhang, M.Q., Oldfield, E.H., and Yen, P.M. (2001). Aberrant alternative splicing of thyroid hormone receptor in a TSH-secreting pituitary tumor is a mechanism for hormone resistance. *Mol. Endocrinol.* 15, 1529–1538.
- Arribere, J. a., Bell, R.T., Fu, B.X.H., Artiles, K.L., Hartman, P.S., and Fire, a. Z. (2014). Efficient Marker-Free Recovery of Custom Genetic Modifications with CRISPR/Cas9 in *Caenorhabditis elegans*. *Genetics* 198, 837–846.
- Ashiya, M., and Grabowski, P.J. (1997). A neuron-specific splicing switch mediated by an array of pre-mRNA repressor sites: evidence of a regulatory role for the polypyrimidine tract binding protein and a brain-specific PTB counterpart. *RNA* 3, 996–1015.
- Auboeuf, D., Höning, A., Berget, S.M., and O'Malley, B.W. (2002). Coordinate regulation of transcription and splicing by steroid receptor coregulators. *Science* 298, 416–419.
- Ayoubi, T. a, and Van De Ven, W.J. (1996). Regulation of gene expression by alternative promoters. *FASEB J.* 10, 453–460.
- Barberan-Soler, S., and Zahler, A.M. (2008). Alternative splicing regulation during *C. elegans* development: splicing factors as regulated targets. *PLoS Genet.* 4, e1000001.
- Barberan-Soler, S., Lambert, N.J., and Zahler, A.M. (2009). Global analysis of alternative splicing uncovers developmental regulation of nonsense-mediated decay in *C. elegans*. *RNA* 15, 1652–1660.
- Barberan-Soler, S., Medina, P., Estella, J., Williams, J., and Zahler, A.M. (2011). Co-regulation of alternative splicing by diverse splicing factors in *Caenorhabditis elegans*. *Nucleic Acids Res.* 39, 666–674.

- Batsché, E., Yaniv, M., and Muchardt, C. (2006). The human SWI/SNF subunit Brm is a regulator of alternative splicing. *Nat. Struct. Mol. Biol.* *13*, 22–29.
- Beckmann, J.S., and Trifonov, E.N. (1991). Splice junctions follow a 205-base ladder. *Proc. Natl. Acad. Sci. U. S. A.* *88*, 2380–2383.
- Belasco, J.G. (2010). All things must pass: contrasts and commonalities in eukaryotic and bacterial mRNA decay. *Nat. Rev. Mol. Cell Biol.* *11*, 467–478.
- Berezikov, E., Bargmann, C.I., and Plasterk, R.H.A. (2004). Homologous gene targeting in *Caenorhabditis elegans* by biolistic transformation. *Nucleic Acids Res.* *32*, 40e – 40.
- Berget, S.M. (1995). Exon recognition in vertebrate splicing. *J. Biol. Chem.* *270*, 2411–2414.
- Berget, S.M., and Sharp, P. a (1977). A spliced sequence at the 5'-terminus of adenovirus late mRNA. *Proc. Natl. Acad. Sci. U. S. A.* *74*, 3171–3175.
- Beyer, a. L., and Osheim, Y.N. (1988). Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. *Genes Dev.* *2*, 754–765.
- Bhalla, a. D., Gudikote, J.P., Wang, J., Chan, W.-K., Chang, Y.-F., Olivas, O.R., and Wilkinson, M.F. (2009). Nonsense Codons Trigger an RNA Partitioning Shift. *J. Biol. Chem.* *284*, 4062–4072.
- Birukov, K.G., Csontos, C., Marzilli, L., Dudek, S., Ma, S.F., Bresnick, a R., Verin, a D., Cotter, R.J., and Garcia, J.G. (2001). Differential regulation of alternatively spliced endothelial cell myosin light chain kinase isoforms by p60(Src). *J. Biol. Chem.* *276*, 8567–8573.
- Black, D.L. (1991). Does steric interference between splice sites block the splicing of a short c-src neuron-specific exon in non-neuronal cells? *Genes Dev.* *5*, 389–402.
- Black, D.L. (2003). Mechanisms of alternative pre-messenger RNA splicing. *Annu. Rev. Biochem.* *72*, 291–336.
- Boulin, T., Etchberger, J.F., and Hobert, O. (2006). Reporter gene fusions. *WormBook* 1–23.
- Boutz, P., Stoilov, P., Li, Q., and Lin, C. (2007). A post-transcriptional regulatory switch in polypyrimidine tract-binding proteins reprograms alternative splicing in developing neurons. *Genes ...* 1636–1652.
- Brenner, S. (1974). The genetics of *Caenorhabditis elegans*. *Genetics* *77*, 71–94.
- Brett, D., Pospisil, H., Valcárcel, J., Reich, J., and Bork, P. (2002). Alternative splicing and genome complexity. *Nat. Genet.* *30*, 29–30.
- Broverman, S. (1993). Alteration of *Caenorhabditis elegans* gene expression by targeted transformation. *Proc. ...* *90*, 4359–4363.
- Buratti, E., and Baralle, F.E. (2004). Influence of RNA Secondary Structure on the Pre-mRNA Splicing Process. *Mol. Cell. Biol.* *24*, 10505–10514.

- Carter, M.S., Li, S., and Wilkinson, M.F. (1996). A splicing-dependent regulatory mechanism that detects translation signals. *EMBO J.* *15*, 5965–5975.
- Cavaloc, Y., Bourgeois, C.F., Kister, L., and Stévenin, J. (1999). The splicing factors 9G8 and SRp20 transactivate splicing through different and specific enhancers. *RNA* *5*, 468–483.
- Cha, D.S., Hollis, S.E., Datla, U.S., Lee, S., Ryu, J., Jung, H.R., Kim, E., Kim, K., Lee, M., Li, C., et al. (2012). Differential subcellular localization of DNA topoisomerase-1 isoforms and their roles during *Caenorhabditis elegans* development. *Gene Expr. Patterns* *12*, 189–195.
- Chabot, B., Blanchette, M., Lapierre, I., and La Branche, H. (1997). An intron element modulating 5' splice site selection in the hnRNP A1 pre-mRNA interacts with hnRNP A1. *Mol. Cell. Biol.* *17*, 1776–1786.
- Chalfie, M., Tu, Y., Euskirchen, G., Ward, W.W., and Prasher, D.C. (1994). Green fluorescent protein as a marker gene expression. *Science* *263*, 802–805.
- Chang, S., Johnston, R.J., and Hobert, O. (2003). A transcriptional regulatory cascade that controls left/right asymmetry in chemosensory neurons of *C. elegans*. *Genes Dev.* *17*, 2123–2137.
- Chang, Y.-F., Imam, J.S., and Wilkinson, M.F. (2007). The nonsense-mediated decay RNA surveillance pathway. *Annu. Rev. Biochem.* *76*, 51–74.
- Chen, B.E., Kondo, M., Garnier, A., Watson, F.L., Püettmann-Holgado, R., Lamar, D.R., and Schmucker, D. (2006). The Molecular Diversity of Dscam Is Functionally Required for Neuronal Wiring Specificity in *Drosophila*. *Cell* *125*, 607–620.
- Cheng, A.W., Wang, H., Yang, H., Shi, L., Katz, Y., Theunissen, T.W., Rangarajan, S., Shivalila, C.S., Dadon, D.B., and Jaenisch, R. (2013). Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. *Cell Res.* *23*, 1163–1171.
- Chiu, H., Schwartz, H.T., Antoshechkin, I., and Sternberg, P.W. (2013). Transgene-free genome editing in *Caenorhabditis elegans* using CRISPR-Cas. *Genetics* *195*, 1167–1171.
- Cho, S.W., Lee, J., Carroll, D., Kim, J.-S., and Lee, J. (2013). Heritable gene knockout in *Caenorhabditis elegans* by direct injection of Cas9-sgRNA ribonucleoproteins. *Genetics* *195*, 1177–1180.
- Christmas, P., Jones, J.P., Patten, C.J., Rock, D. a, Zheng, Y., Cheng, S.M., Weber, B.M., Carlesso, N., Scadden, D.T., Rettie, a E., et al. (2001). Alternative splicing determines the function of CYP4F3 by switching substrate specificity. *J. Biol. Chem.* *276*, 38166–38172.
- Close, P., East, P., Dirac-Svejstrup, a. B., Hartmann, H., Heron, M., Maslen, S., Chariot, A., Söding, J., Skehel, M., and Svejstrup, J.Q. (2012). DBIRD complex integrates alternative mRNA splicing with RNA polymerase II transcript elongation. *Nature* *484*, 386–389.
- Craig, H.L., Wirtz, J., Bamps, S., Dolphin, C.T., and Hope, I. a (2013). The significance of alternative transcripts for *Caenorhabditis elegans* transcription factor genes, based on expression pattern analysis. *BMC Genomics* *14*, 249.

- Cramer, P., Pesce, C.G., Baralle, F.E., and Kornblihtt, a R. (1997). Functional association between promoter structure and transcript alternative splicing. *Proc. Natl. Acad. Sci. U. S. A.* *94*, 11456–11460.
- Cramer, P., Cáceres, J.F., Cazalla, D., Kadener, S., Muro, A.F., Baralle, F.E., and Kornblihtt, A.R. (1999). Coupling of Transcription with Alternative Splicing. *Mol. Cell* *4*, 251–258.
- Das, R., Yu, J., Zhang, Z., Gygi, M.P., Krainer, A.R., Gygi, S.P., and Reed, R. (2007). SR Proteins Function in Coupling RNAP II Transcription to Pre-mRNA Splicing. *Mol. Cell* *26*, 867–881.
- David, a., Dolan, B.P., Hickman, H.D., Knowlton, J.J., Clavarino, G., Pierre, P., Bennink, J.R., and Yewdell, J.W. (2012). Nuclear translation visualized by ribosome-bound nascent chain puromycylation. *J. Cell Biol.* *197*, 45–57.
- David, C.J., Chen, M., Assanah, M., Canoll, P., and Manley, J.L. (2010). HnRNP proteins controlled by c-Myc deregulate pyruvate kinase mRNA splicing in cancer. *Nature* *463*, 364–368.
- Demir, E., and Dickson, B.J. (2005). fruitless splicing specifies male courtship behavior in *Drosophila*. *Cell* *121*, 785–794.
- Dernburg, A.F., Zalevsky, J., and Colaiácovo, M.P. (2000). Transgene-mediated cosuppression in the *C. elegans* germ line. *Genes Dev.* *14*, 1578–1583.
- Dieci, G., Fiorino, G., Castelnovo, M., Teichmann, M., and Pagano, A. (2007). The expanding RNA polymerase III transcriptome. *Trends Genet.* *23*, 614–622.
- Doitsidou, M., Flames, N., Lee, A., Boyanov, A., and Hobert, O. (2008). Automated screening for mutants affecting dopaminergic-neuron specification in *C. elegans*. *Nat. Methods* *5*, 869–872.
- Dominguez, C., Fiset, J.-F., Chabot, B., and Allain, F.H.-T. (2010). Structural basis of G-tract recognition and encaging by hnRNP F quasi-RRMs. *Nat. Struct. Mol. Biol.* *17*, 853–861.
- Dominski, Z., and Kole, R. (1991). Selection of splice sites in pre-mRNAs with short internal exons. *Mol. Cell. Biol.* *11*, 6075–6083.
- Dostie, J., and Dreyfuss, G. (2002). Translation is required to remove Y14 from mRNAs in the cytoplasm. *Curr. Biol.* *12*, 1060–1067.
- Drygin, D., Rice, W.G., and Grummt, I. (2010). The RNA polymerase I transcription machinery: an emerging target for the treatment of cancer. *Annu. Rev. Pharmacol. Toxicol.* *50*, 131–156.
- Du, C., McGuffin, M.E., Dauwalder, B., Rabinow, L., and Mattox, W. (1998). Protein phosphorylation plays an essential role in the regulation of alternative splicing and sex determination in *Drosophila*. *Mol Cell* *2*, 741–750.
- Dupuy, D., Li, Q.-R., Deplancke, B., Boxem, M., Hao, T., Lamesch, P., Sequerra, R., Bosak, S., Doucette-Stamm, L., Hope, I. a., et al. (2004). A first version of the *Caenorhabditis elegans* Promoterome. *Genome Res.* *14*, 2169–2175.

- Ellis, M.H., and Horvitz, H.R. (1986). Genetic Control of Programmed Cell Death in the Nematode *C. elegans*. *Cell* 44, 817–829.
- Farboud, B., and Meyer, B.J. (2015). Dramatic enhancement of genome editing by CRISPR / Cas9 through improved guide RNA design. *Genetics* 199, 959–971.
- Farrer, T., Roller, a B., Kent, W.J., and Zahler, A.M. (2002). Analysis of the role of *Caenorhabditis elegans* GC-AG introns in regulated splicing. *Nucleic Acids Res.* 30, 3360–3367.
- Feng, Z., Zhang, B., Ding, W., Liu, X., Yang, D., Wei, P., Cao, F., Zhu, S., Zhang, F., Mao, Y., et al. (2013). Efficient genome editing in plants using a CRISPR-Cas system. *Nat. Biotechnol.* 31, 227–229.
- Fernandez, A.G., Gunsalus, K.C., Huang, J., Chuang, L., Ying, N., Liang, H., Tang, C., Schetter, A.J., Zegar, C., Rual, J., et al. (2005). New genes with roles in the *C. elegans* embryo revealed using RNAi of ovary-enriched ORFeome clones. *Genome Res.* 15, 250–259.
- Fernandez Alanis, E., Pinotti, M., Dal Mas, a., Balestra, D., Cavallari, N., Rogalska, M.E., Bernardi, F., and Pagani, F. (2012). An exon-specific U1 small nuclear RNA (snRNA) strategy to correct splicing defects. *Hum. Mol. Genet.* 21, 2389–2398.
- Finney, M., and Ruvkun, G. (1990). The *unc-86* gene product couples cell lineage and cell identity in *C. elegans*. *Cell* 63, 895–905.
- Finney, M., Ruvkun, G., and Horvitz, H.R. (1988). The *C. elegans* cell lineage and differentiation gene *unc-86* encodes a protein with a homeodomain and extended similarity to transcription factors. *Cell* 55, 757–769.
- Fire, a, Xu, S., Montgomery, M.K., Kostas, S. a, Driver, S.E., and Mello, C.C. (1998). Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 391, 806–811.
- Fire, A., Harrison, S.W., and Dixon, D. (1990). A modular set of lacZ fusion vectors for studying gene expression in *Caenorhabditis elegans*. *Gene* 93, 189–198.
- Fortes, P., Inada, T., Preiss, T., Hentze, M.W., Mattaj, I.W., and Sachs, A.B. (2000). The Yeast Nuclear Cap Binding Complex Can Interact with Translation Factor eIF4G and Mediate Translation Initiation. *Mol. Cell* 6, 191–196.
- Fraser, a G., Kamath, R.S., Zipperlen, P., Martinez-Campos, M., Sohrmann, M., and Ahringer, J. (2000). Functional genomic analysis of *C. elegans* chromosome I by systematic RNA interference. *Nature* 408, 325–330.
- Friedland, A.E., Tzur, Y.B., Esvelt, K.M., Colaiácovo, M.P., Church, G.M., and Calarco, J. a (2013). Heritable genome editing in *C. elegans* via a CRISPR-Cas9 system. *Nat. Methods* 10, 741–743.
- Frøkjær-jensen, C., Davis, M.W., Hopkins, C.E., Newman, B., Thummel, J.M., Olesen, S., Grunnet, M., and Jørgensen, E.M. (2009). Single-Copy Insertion of transgenes in *C.elegans*. *Nature* 40, 1375–1383.
- Fu, X.D. (1995). The superfamily of arginine/serine-rich splicing factors. *RNA* 1, 663–680.

Gabellini, D., D'Antona, G., Moggio, M., Prelle, A., Zecca, C., Adami, R., Angeletti, B., Ciscato, P., Pellegrino, M.A., Bottinelli, R., et al. (2006). Facioscapulohumeral muscular dystrophy in mice overexpressing FRG1. *Nature* 439, 973–977.

Gaj, T., Gersbach, C.A., and Barbas III, C.F. (2013). ZFN, TALEN and CRISPR/Cas based methods for genome engineering. *Trends Biotechnol.* 31, 397–405.

Gasiunas, G., Barrangou, R., Horvath, P., and Siksnys, V. (2012). Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl. Acad. Sci. U. S. A.* 109, E2579–E2586.

Gatfield, D., and Izaurralde, E. (2004). Nonsense-mediated messenger RNA decay is initiated by endonucleolytic cleavage in *Drosophila*. *Nature* 429, 575–578.

Del Gatto-Konczak, F., Olive, M., Gesnel, M.C., and Breathnach, R. (1999). hnRNP A1 recruited to an exon in vivo can function as an exon splicing silencer. *Mol. Cell. Biol.* 19, 251–260.

Gavrulina, T.O., McGovern, V.L., Workman, E., Crawford, T.O., Gogliotti, R.G., DiDonato, C.J., Monani, U.R., Morris, G.E., and Burghes, A.H.M. (2008). Neuronal SMN expression corrects spinal muscular atrophy in severe SMA mice while muscle-specific SMN expression has no phenotypic effect. *Hum. Mol. Genet.* 17, 1063–1075.

Ge, H., and Manley, J.L. (1990). A protein factor, ASF, controls cell-specific alternative splicing of SV40 early pre-mRNA in vitro. *Cell* 62, 25–34.

Gehring, N.H., Kunz, J.B., Neu-Yilik, G., Breit, S., Viegas, M.H., Hentze, M.W., and Kulozik, A.E. (2005). Exon-Junction Complex Components Specify Distinct Routes of Nonsense-Mediated mRNA Decay with Differential Cofactor Requirements. *Mol. Cell* 20, 65–75.

Ghosh, A., and Lima, C.D. (2010). Enzymology of RNA cap synthesis. *Wiley Interdiscip. Rev. RNA* 1, 152–172.

Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison, C.A., and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* 6, 343–345.

Giordano-Santini, R., Milstein, S., Svrzikapa, N., Tu, D., Johnsen, R., Baillie, D., Vidal, M., and Dupuy, D. (2010). An antibiotic selection marker for nematode transgenesis. *Nat. Methods* 7, 721–723.

Grabowski, P.J., and Black, D.L. (2001). Alternative RNA splicing in the nervous system. *Prog. Neurobiol.* 65, 289–308.

Gracheva, E.O., Cordero-Morales, J.F., González-Carcacía, J. a, Ingolia, N.T., Manno, C., Aranguren, C.I., Weissman, J.S., and Julius, D. (2011). Ganglion-specific splicing of TRPV1 underlies infrared sensation in vampire bats. *Nature* 476, 88–91.

Granato, M., Schnabe, H., and Schnabe, R. (1994). pha-1, a selectable marker for gene-transfer in *C. elegans*. *Nucleic Acids Res.* 22, 1762–1763.

Hedgecock, E.M., Sulston, J.E., and Thomson, J.N. (1983). Mutations affecting programmed cell deaths in the nematode *Caenorhabditis elegans*. *Science* (80- ). 220, 1277–1279.

Le Hir, H., Izaurralde, E., Maquat, L.E., and Moore, M.J. (2000). The spliceosome deposits multiple proteins 20-24 nucleotides upstream of mRNA exon-exon junctions. *EMBO J.* 19, 6860–6869.

Le Hir, H., Gatfield, D., Izaurralde, E., and Moore, M.J. (2001). The exon-exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay. *EMBO J.* 20, 4987–4997.

Hodges, C., Bintu, L., Lubkowska, L., Kashlev, M., and Bustamante, C. (2009). Nucleosomal fluctuations govern the transcription dynamics of RNA Polymerase II. *Science* (80- ). 325, 626–628.

Horvath, P., and Barrangou, R. (2010). CRISPR/Cas, the immune system of bacteria and archaea. *Science* 327, 167–170.

Horvitz, H.R., and Sulston, J.E. (1980). Isolation and genetic characterization of cell-lineage mutants of the nematode *Caenorhabditis elegans*. *Genetics* 96, 435–454.

Hsu, I.-W., Hsu, M., Li, C., Chuang, T.-W., Lin, R.-I., and Tarn, W.-Y. (2005). Phosphorylation of Y14 modulates its interaction with proteins involved in mRNA metabolism and influences its methylation. *J. Biol. Chem.* 280, 34507–34512.

Hua, Y., Sahashi, K., Hung, G., Rigo, F., Passini, M. a., Bennett, C.F., and Krainer, A.R. (2010). Antisense correction of SMN2 splicing in the CNS rescues necrosis in a type III SMA mouse model. *Genes Dev.* 24, 1634–1644.

Huang, Y., Li, W., Yao, X., Lin, Q.J., Yin, J.W., Liang, Y., Heiner, M., Tian, B., Hui, J., and Wang, G. (2012). Mediator Complex Regulates Alternative mRNA Processing via the MED23 Subunit. *Mol. Cell* 45, 459–469.

Hwang, W.Y., Fu, Y., Reyon, D., Maeder, M.L., Shengdar, Q., Sander, J.D., Peterson, R.T., Yeh, J.J., and Keith, J. (2013). Efficient In Vivo Genome Editing Using RNA-Guided Nucleases. *Nat. Biotechnol.* 31, 227–229.

Iborra, F.J., Jackson, D.A., and Cook, P.R. (2001). Coupled Transcription and Translation Within Nuclei of Mammalian Cells. *Science* (80- ). 293, 1139–1142.

Indraccolo, S., Minuzzo, S., Noonan, D.M., and Superiore, I. (2002). Differential effects of angiostatin, endostatin and interferon-alpha1 gene transfer on in vivo growth of human breast cancer cells. *Gene Ther.* 9, 867–878.

Ip, J.Y., Schmidt, D., Pan, Q., Ramani, A.K., Fraser, A.G., Odom, D.T., and Blencowe, B.J. (2011). Global impact of RNA polymerase II elongation inhibition on alternative splicing regulation. *Genome Res.* 21, 390–401.

Ishigaki, Y., Li, X., Serin, G., and Maquat, L.E. (2001). Evidence for a pioneer round of mRNA translation: mRNAs subject to nonsense-mediated decay in mammalian cells are bound by CBP80 and CBP20. *Cell* 106, 607–617.

- Isken, O., and Maquat, L.E. (2008). The multiple lives of NMD factors: balancing roles in gene and genome regulation. *Nat. Rev. Genet.* *9*, 699–712.
- Izaurrealde, E., Lewis, J., McGuigan, C., Jankowska, M., Darzynkiewicz, E., and Mattaj, L.W. (1994). A nuclear cap binding protein complex involved in pre-mRNA splicing. *Cell* *78*, 657–668.
- Jansen, Thijssen, K.L., Werner, P., van der Horst, M., Hazendonk, E., and Plasterk, R.H. (1999). The complete family of genes encoding G proteins of *Caenorhabditis elegans*. *Nat. Genet.* *21*, 414–419.
- Jareborg, N., Birney, E., and Durbin, R. (2000). Mouse and Human Gene Pairs Comparative Analysis of Noncoding Regions of 77 Orthologous Mouse and Human Gene Pairs. *Genome Res.* *9*, 815–824.
- Jones, E.M., Gray-Keller, M., and Fettiplace, R. (1999). The role of Ca<sup>2+</sup>-activated K<sup>+</sup> channel spliced variants in the tonotopic organization of the turtle cochlea. *J. Physiol.* *518* ( Pt 3), 653–665.
- Kadener, S., Fededa, J.P., Rosbash, M., and Kornblihtt, A.R. (2002). Regulation of alternative splicing by a transcriptional enhancer through RNA pol II elongation. *Proc. Natl. Acad. Sci. U. S. A.* *99*, 8185–8190.
- Kanezaki, R., Toki, T., Yokoyama, M., Yomogida, K., Sugiyama, K., Yamamoto, M., Igarashi, K., and Ito, E. (2001). Transcription factor BACH1 is recruited to the nucleus by its novel alternative spliced isoform. *J. Biol. Chem.* *276*, 7278–7284.
- Kanopka, A., Mühlemann, O., Petersen-Mahrt, S., Estmer, C., Ohrmalm, C., and Akusjärvi, G. (1998). Regulation of adenovirus alternative RNA splicing by dephosphorylation of SR proteins. *Nature* *393*, 185–187.
- Kashima, I., Yamashita, A., Izumi, N., Kataoka, N., Morishita, R., Hoshino, S., Ohno, M., Dreyfuss, G., and Ohno, S. (2006). Nonsense-mediated mRNA decay (NMD) is a surveillance mechanism that degrades mRNA containing premature termination codons (PTCs). In mammalian cells, recognition of PTCs requires translation and depends on the presence on the mRNA with the splicing-depend. *Genes Dev.* *20*, 355–367.
- Kelly, W.G., Xu, S., Montgomery, M.K., and Fire, a (1997). Distinct requirements for somatic and germline expression of a generally expressed *Caenorhabditis elegans* gene. *Genetics* *146*, 227–238.
- Kenyon, C., Chang, J., Gensch, E., Rudner, a, and Tabtiang, R. (1993). A *C. elegans* mutant that lives twice as long as wild type. *Nature* *366*, 461–464.
- Khodor, Y.L., Rodriguez, J., Abruzzi, K.C., Tang, C.H.A., Marr, M.T., and Rosbash, M. (2011). Nascent-seq indicates widespread cotranscriptional pre-mRNA splicing in *Drosophila*. *Genes Dev.* *25*, 2502–2512.
- Kim, S., Kim, H., Fong, N., Erickson, B., and Bentley, D.L. (2011). Pre-mRNA splicing is a determinant of histone H3K36 methylation. *Proc. Natl. Acad. Sci. U. S. A.* *108*, 13564–13569.

- Kim, V.N., Kataoka, N., and Dreyfuss, G. (2001). Role of the nonsense-mediated decay factor hUpf3 in the splicing-dependent exon-exon junction complex. *Science* 293, 1832–1836.
- Kimble, J., and Hirsh, D. (1979). The postembryonic cell lineages of the hermaphrodite and male gonads in *Caenorhabditis elegans*. *Dev. Biol.* 70, 396–417.
- Kohtz, J.D., Jamison, S.F., Will, C.L., Zuo, P., Lührmann, R., Garcia-Blanco, M. a., and Manley, J.L. (1994). Protein-protein interactions and 5'-splice-site recognition in mammalian mRNA precursors. *Nature* 368, 119–124.
- Kolasinska-Zwierz, P., Down, T., Latorre, I., Liu, T., Liu, X.S., and Ahringer, J. (2009). Differential chromatin marking of introns and expressed exons by H3K36me3. *Nat. Genet.* 41, 376–381.
- Kornblihtt, A.R., Schor, I.E., Alló, M., Dujardin, G., Petrillo, E., and Muñoz, M.J. (2013). Alternative splicing: a pivotal step between eukaryotic transcription and translation. *Nat. Rev. Mol. Cell Biol.* 14, 153–165.
- Krainer, a. R., Conway, G.C., and Kozak, D. (1990). The essential pre-mRNA splicing factor SF2 influences 5' splice site selection by activating proximal sites. *Cell* 62, 35–42.
- Krawczak, M., Reiss, J., and Cooper, D.N. (1992). The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: Causes and consequences. *Hum. Genet.* 90, 41–54.
- Krecic, A.M., and Swanson, M.S. (1999). hnRNP complexes: composition, structure, and function. *Curr. Opin. Cell Biol.* 11, 363–371.
- Krummheuer, J., Lenz, C., Kammler, S., Scheid, a, and Schaal, H. (2001). Influence of the small leader exons 2 and 3 on human immunodeficiency virus type 1 gene expression. *Virology* 286, 276–289.
- Kuo, H., Nasim, F.H., and Grabowski, P.J. (1991). Control of alternative splicing by the differential binding of U1 small nuclear ribonucleoprotein particle. *Science* (80-. ). 251, 1045–1050.
- Kuroyanagi, H., Kobayashi, T., Mitani, S., and Hagiwara, M. (2006). Transgenic alternative-splicing reporters reveal tissue-specific expression profiles and regulation mechanisms in vivo. *Nat. Methods* 3.
- Kuroyanagi, H., Ohno, G., Mitani, S., and Hagiwara, M. (2007). The Fox-1 family and SUP-12 coordinately regulate tissue-specific alternative splicing in vivo. *Mol. Cell. Biol.* 27, 8612–8621.
- Kuroyanagi, H., Ohno, G., Sakane, H., Maruoka, H., and Hagiwara, M. (2010). Visualization and genetic analysis of alternative splicing regulation in vivo using fluorescence reporters in transgenic *Caenorhabditis elegans*. *Nat. Protoc.* 5, 1495–1517.
- Kuroyanagi, H., Watanabe, Y., and Hagiwara, M. (2013). CELF Family RNA-Binding Protein UNC-75 Regulates Two Sets of Mutually Exclusive Exons of the unc-32 Gene in Neuron-Specific Manners in *Caenorhabditis elegans*. *PLoS Genet.* 9, e1003337.

- De la Mata, M., and Kornblihtt, A.R. (2006). RNA polymerase II C-terminal domain mediates regulation of alternative splicing by SRp20. *Nat. Struct. Mol. Biol.* *13*, 973–980.
- Lee, H.C., Oh, N., Cho, H., Choe, J., and Kim, Y.K. (2010). Nonsense-mediated translational repression involves exon junction complex downstream of premature translation termination codon. *FEBS Lett.* *584*, 795–800.
- Lee, M.H., Jang, Y.J., and Koo, H.S. (1998). Alternative splicing in the *Caenorhabditis elegans* DNA topoisomerase I gene. *Biochim. Biophys. Acta* *1396*, 207–214.
- Lee, R.C., Feinbaum, R.L., and Ambros, V. (1993). The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* *75*, 843–854.
- Lejeune, F., Li, X., and Maquat, L.E. (2003). Nonsense-mediated mRNA decay in mammalian cells involves decapping, deadenylation, and exonucleolytic activities. *Mol. Cell* *12*, 675–687.
- Lewis, B.P., Shih, I.H., Jones-Rhoades, M.W., Bartel, D.P., and Burge, C.B. (2003). Prediction of Mammalian MicroRNA Targets. *Cell* *115*, 787–798.
- Li, S., and Koromilas, A.E. (2001). Dominant Negative Function by an Alternatively Spliced Form of the Interferon-inducible Protein Kinase PKR. *J. Biol. Chem.* *276*, 13881–13890.
- Li, D., Qiu, Z., Shao, Y., Chen, Y., Guan, Y., Liu, M., Li, Y., Gao, N., Wang, L., Lu, X., et al. (2013). Heritable gene targeting in the mouse and rat using a CRISPR-Cas system. *Nat. Biotechnol.* *31*, 681–683.
- Lillycrop, K. a, and Latchman, D.S. (1992). Alternative splicing of the Oct-2 transcription factor RNA is differentially regulated in neuronal cells and B cells and results in protein isoforms with opposite effects on the activity of octamer/TAATGARAT-containing promoters. *J. Biol. Chem.* *267*, 24960–24965.
- Loomis, R.J., Naoe, Y., Parker, J.B., Savic, V., Bozovsky, M.R., Macfarlan, T., Manley, J.L., and Chakravarti, D. (2009). Chromatin Binding of SRp20 and ASF/SF2 and Dissociation from Mitotic Chromosomes Is Modulated by Histone H3 Serine 10 Phosphorylation. *Mol. Cell* *33*, 450–461.
- Lorson, C.L., Hahnen, E., Androphy, E.J., and Wirth, B. (1999). A single nucleotide in the SMN gene regulates splicing and is responsible for spinal muscular atrophy. *Proc. Natl. Acad. Sci. U. S. A.* *96*, 6307–6311.
- Lou, H., Neugebauer, K.M., Gagel, R.F., and Berget, S.M. (1998). Regulation of alternative polyadenylation by U1 snRNPs and SRp20. *Mol. Cell. Biol.* *18*, 4977–4985.
- Louise T Chow, Richard E. Gelinias, Thomas R., B. and R.J.R. (1977). An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell* *12*, 1–8.
- Luco, R., Pan, Q., and Tominaga, K. (2010). Regulation of alternative splicing by histone modifications. *Science* (80-. ). *327*, 996–1000.
- Luo, M.J., and Reed, R. (1999). Splicing is required for rapid and efficient mRNA export in metazoans. *Proc. Natl. Acad. Sci. U. S. A.* *96*, 14937–14942.

- MacLeod, a R., Waterston, R.H., Fishpool, R.M., and Brenner, S. (1977). Identification of the structural gene for a myosin heavy-chain in *Caenorhabditis elegans*. *J. Mol. Biol.* *114*, 133–140.
- Maduro, M., and Pilgrim, D. (1995). Identification and cloning of *unc-119*, a gene expressed in the *Caenorhabditis elegans* nervous system. *Genetics* *141*, 977–988.
- Malone, C.J., Misner, L., Le Bot, N., Tsai, M.C., Campbell, J.M., Ahringer, J., and White, J.G. (2003). The *C. elegans* Hook Protein, ZYG-12, Mediates the Essential Attachment between the Centrosome and Nucleus. *Cell* *115*, 825–836.
- Mangone, M., Macmenamin, P., Zegar, C., Piano, F., and Gunsalus, K.C. (2008). UTRome.org: a platform for 3'UTR biology in *C. elegans*. *Nucleic Acids Res* *36*, D57–D62.
- Mata, M. De, Alonso, C.R., Fededa, J.P., Pelisch, F., Cramer, P., Bentley, D., and Kornblihtt, A.R. (2003). A slow RNA Polymerase II affects alternative splicing in vivo. *Mol. Cell* *12*, 525–532.
- Mattick, J.S., and Makunin, I. V. (2006). Non-coding RNA. *Hum. Mol. Genet.* *15*, R17–R29.
- Mayeda, a, and Krainer, a R. (1992). Regulation of alternative pre-mRNA splicing by hnRNP A1 and splicing factor SF2. *Cell* *68*, 365–375.
- Mello, C.C., Kramer, J.M., Stinchcomb, D., and Ambros, V. (1991). Efficient gene transfer in *C.elegans*: extrachromosomal maintenance and integration of transforming sequences. *EMBO J.* *10*, 3959–3970.
- Merino, A., Madden, K.R., Lane, W.S., Champoux, J.J., and Reinberg, D. (1993). DNA topoisomerase I is involved in both repression and activation of transcription. *Nature* *365*, 227–232.
- Mermoud, J.E., Cohen, P., and Lamond, a I. (1992). Ser/Thr-specific protein phosphatases are required for both catalytic steps of pre-mRNA splicing. *Nucleic Acids Res.* *20*, 5263–5269.
- Mermoud, J.E., Cohen, P.T., and Lamond, a I. (1994). Regulation of mammalian spliceosome assembly by a protein phosphorylation mechanism. *EMBO J.* *13*, 5679–5688.
- Meshorer, E., Toiber, D., Zurel, D., Sahly, I., Dori, A., Cagnano, E., Schreiber, L., Grisaru, D., Tronche, F., and Soreq, H. (2004). Combinatorial complexity of 5' alternative acetylcholinesterase transcripts and protein products. *J. Biol. Chem.* *279*, 29740–29751.
- Michaud, S., and Reed, R. (1991). An ATP-independent complex commits pre-mRNA to the mammalian spliceosome assembly pathway. *Genes Dev.* *5*, 2534–2546.
- Modafferi, E.F., and Black, D.L. (1999). Combinatorial control of a neuron-specific exon. *RNA* *5*, 687–706.
- Moore, M.J., and Sharp, P. a (1993). Evidence for two active sites in the spliceosome provided by stereochemistry of pre-mRNA splicing. *Nature* *365*, 364–368.
- Morham, S.G., Kluckman, K.D., Voulomanos, N., and Smithies, O. (1996). Targeted disruption of the mouse topoisomerase I gene by camptothecin selection. *Mol. Cell. Biol.* *16*, 6804–6809.

- Morton, J., Davis, M.W., Jorgensen, E.M., and Carroll, D. (2006). Induction and repair of zinc-finger nuclease-targeted double-strand breaks in *Caenorhabditis elegans* somatic cells. *Proc. Natl. Acad. Sci. U. S. A.* *103*, 16370–16375.
- Muhrad, D., and Parker, R. (1999). Aberrant mRNAs with extended 3' UTRs are substrates for rapid degradation by mRNA surveillance. *RNA* *5*, 1299–1307.
- Muñoz, M.J., Santangelo, M.S.P., Paronetto, M.P., de la Mata, M., Pelisch, F., Boireau, S., Glover-Cutter, K., Ben-Dov, C., Blaustein, M., Lozano, J.J., et al. (2009). DNA Damage Regulates Alternative Splicing through Inhibition of RNA Polymerase II Elongation. *Cell* *137*, 708–720.
- Nagai, Y., Ueno, S., Saeki, Y., Soga, F., and Yanagihara, T. (1993). Expression of the D3 dopamine receptor gene and a novel variant transcript generated by alternative splicing in human peripheral blood lymphocytes. *Biochem. Biophys. Res. Commun.* *194*, 368–374.
- Nagel, G., Brauner, M., Liewald, J.F., Adeishvili, N., Bamberg, E., and Gottschalk, A. (2005). Light Activation of Channelrhodopsin-2 in Excitable Cells of *Caenorhabditis elegans* Triggers Rapid Behavioral Responses. *Curr. Biol.* *15*, 2279–2284.
- Nagy, E., and Maquat, L.E. (1998). A rule for termination-codon position within intron-containing genes: When nonsense affects RNA abundance. *Trends Biochem. Sci.* *23*, 198–199.
- Nakai, K., and Sakamoto, H. (1994). Construction of a novel database containing aberrant splicing mutations of mammalian genes. *Gene* *141*, 171–177.
- Nestler, E.J., Kelz, M.B., and Chen, J. (1999). DeltaFosB: a molecular mediator of long-term neural and behavioral plasticity. *Brain Res.* *835*, 10–17.
- Niwa, M., and Berget, S.M. (1991). Mutation of the AAUAAA polyadenylation signal depresses in vitro splicing of proximal but not distal introns. *Genes Dev.* *5*, 2086–2095.
- Niwa, M., Rose, S.D., and Berget, S.M. (1990). In vitro polyadenylation is stimulated by the presence of an upstream intron. *Genes Dev.* *4*, 1552–1559.
- Norris, A.D., Gao, S., Norris, M.L., Ray, D., Ramani, A.K., Fraser, A.G., Morris, Q., Hughes, T.R., Zhen, M., and Calarco, J. a. (2014). A Pair of RNA-binding proteins controls networks of splicing events contributing to specialization of neural cell types. *Mol. Cell* *54*, 946–959.
- O'Meara, M.M., Bigelow, H., Flibotte, S., Etchberger, J.F., Moerman, D.G., and Hobert, O. (2009). Cis-regulatory mutations in the *Caenorhabditis elegans* homeobox gene locus *cog-1* affect neuronal development. *Genetics* *181*, 1679–1686.
- Ohno, G., Hagiwara, M., and Kuroyanagi, H. (2008). STAR family RNA-binding protein ASD-2 regulates developmental switching of mutually exclusive alternative splicing in vivo. *Genes Dev.* *22*, 360–374.
- Ohno, G., Ono, K., Togo, M., Watanabe, Y., Ono, S., Hagiwara, M., and Kuroyanagi, H. (2012). Muscle-Specific Splicing Factors ASD-2 and SUP-12 Cooperatively Switch Alternative Pre-mRNA Processing Patterns of the ADF/Cofilin Gene in *Caenorhabditis elegans*. *PLoS Genet.* *8*, e1002991.

- Ohno, M., Sakamoto, H., and Shimura, Y. (1987). Preferential excision of the 5' proximal intron from mRNA precursors with two introns as mediated by the cap structure. *Proc. Natl. Acad. Sci. U. S. A.* *84*, 5187–5191.
- Pagani, F., and Baralle, F.E. (2004). Genomic variants in exons and introns: identifying the splicing spoilers. *Nat. Rev. Genet.* *5*, 389–396.
- Pagani, F., Stuani, C., Zuccato, E., Kornblihtt, A.R., and Baralle, F.E. (2003). Promoter architecture modulates CFTR exon 9 skipping. *J. Biol. Chem.* *278*, 1511–1517.
- Paix, a., Folkmann, a., Rasoloson, D., and Seydoux, G. (2015). High Efficiency, Homology-Directed Genome Editing in *Caenorhabditis elegans* Using CRISPR-Cas9 Ribonucleoprotein Complexes. *Genetics* *201*, 47–54.
- Paix, A., Wang, Y., Smith, H.E., Lee, C.-Y.S., Calidas, D., Lu, T., Smith, J., Schmidt, H., Krause, M.W., and Seydoux, G. (2014). Scalable and Versatile Genome Editing Using Linear DNAs with Micro-Homology to Cas9 Sites in *Caenorhabditis elegans*. *Genetics* *198*, 1347–1356.
- Pandya-Jones, A., and Black, D.L. (2009). Co-transcriptional splicing of constitutive and alternative exons. *RNA* *15*, 1896–1908.
- Park, J.W., and Graveley, B.R. (2007). Complex alternative splicing. *Adv. Exp. Med. Biol.* *623*, 50–63.
- Parker, R., and Sheth, U. (2007). P bodies and the control of mRNA translation and degradation. *Mol. Cell* *25*, 635–646.
- Philips, a V, Timchenko, L.T., and Cooper, T. a (1998). Disruption of splicing regulated by a CUG-binding protein in myotonic dystrophy. *Science* *280*, 737–741.
- Pradeepa, M.M., Sutherland, H.G., Ule, J., Grimes, G.R., and Bickmore, W. a. (2012). Psp1/Ledgf p52 binds methylated histone H3K36 and splicing factors and contributes to the regulation of alternative splicing. *PLoS Genet.* *8*.
- Praitis, V., Casey, E., Collar, D., and Austin, J. (2001). Creation of low-copy integrated transgenic lines in *Caenorhabditis elegans*. *Genetics* *157*, 1217–1226.
- Priess, J.R. (2005). Notch signaling in the *C. elegans* embryo. *WormBook* 1–16.
- Pulak, R., and Anderson, P. (1993). mRNA surveillance by the *Caenorhabditis elegans* stag genes. *Genes Dev.* *7*, 1885–1897.
- Puttaraju, M., Jamison, S.F., Mansfield, S.G., Garcia-Blanco, M. a, and Mitchell, L.G. (1999). Spliceosome-mediated RNA trans-splicing as a tool for gene therapy. *Nat. Biotechnol.* *17*, 246–252.
- Qi, L.S., Larson, M.H., Gilbert, L. a, Doudna, J. a, Weissman, J.S., Arkin, A.P., and Lim, W. a (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* *152*, 1173–1183.
- Radman, I., Greiss, S., and Chin, J.W. (2013). Efficient and rapid *C. elegans* transgenesis by bombardment and hygromycin B selection. *PLoS One* *8*, e76019.

- Ramani, A.K., Calarco, J. a, Pan, Q., Mavandadi, S., Wang, Y., Nelson, A.C., Lee, L.J., Morris, Q., Blencowe, B.J., Zhen, M., et al. (2011). Genome-wide analysis of alternative splicing in *Caenorhabditis elegans*. *Genome Res.* *21*, 342–348.
- Reichert, V.L., Le Hir, H., Jurica, M.S., and Moore, M.J. (2002). 5' Exon interactions within the human spliceosome establish a framework for exon junction complex structure and assembly. *Genes Dev.* *16*, 2778–2791.
- Robberson, B.L., Cote, G.J., and Berget, S.M. (1990). Exon definition may facilitate splice site selection in RNAs with multiple exons. *Mol. Cell. Biol.* *10*, 84–94.
- Rosenblatt, K.P., Sun, Z.P., Heller, S., and Hudspeth, a J. (1997). Distribution of Ca<sup>2+</sup>-activated K<sup>+</sup> channel isoforms along the tonotopic gradient of the chicken's cochlea. *Neuron* *19*, 1061–1075.
- Ryner, L.C., Goodwin, S.F., Castrillon, D.H., Anand, A., Villella, A., Baker, B.S., Hall, J.C., Taylor, B.J., and Wasserman, S. a. (1996). Control of male sexual behavior and sexual orientation in *Drosophila* by the fruitless gene. *Cell* *87*, 1079–1089.
- Ryther, R.C., Flynt, a S., Harris, B.D., Phillips III, J. a, and Patton, J.G. (2004). GH1 splicing is regulated by multiple enhancers whose mutation produces a dominant-negative GH isoform that can be degraded by allele-specific small interfering RNA (siRNA). *Endocrinology* *145*, 2988–2996.
- Saint-André, V., Batsché, E., Rachez, C., and Muchardt, C. (2011). Histone H3 lysine 9 trimethylation and HP1 $\gamma$  favor inclusion of alternative exons. *Nat. Struct. Mol. Biol.* *18*, 337–344.
- Salz, H.K., and Erickson, J.W. (2010). Sex determination in *Drosophila* : The view from the top. *Fly (Austin)*. *4*, 60–70.
- Sarin, S., Antonio, C., Tursun, B., and Hobert, O. (2009). The *C. elegans* Tailless/TLX transcription factor nhr-67 controls neuronal identity and left/right asymmetric fate diversification. *Development* *136*, 2933–2944.
- Schor, I.E., Rascovan, N., Pelisch, F., Alló, M., and Kornblihtt, A.R. (2009). Neuronal cell depolarization induces intragenic chromatin modifications affecting NCAM alternative splicing. *Proc. Natl. Acad. Sci. U. S. A.* *106*, 4325–4330.
- Schor, I.E., Fiszbein, A., Petrillo, E., and Kornblihtt, A.R. (2013). Intragenic epigenetic changes modulate NCAM alternative splicing in neuronal differentiation. *EMBO J.* *32*, 2264–2274.
- Schwartz, S., Meshorer, E., and Ast, G. (2009). Chromatin organization marks exon-intron structure. *Nat. Struct. Mol. Biol.* *16*, 990–995.
- Semple, J.I., Garcia-Verdugo, R., and Lehner, B. (2010). Rapid selection of transgenic *C. elegans* using antibiotic resistance. *Nat. Methods* *7*, 725–727.
- Sesti, G., Federici, M., Lauro, D., Sbraccia, P., and Lauro, R. (2001). Molecular mechanism of insulin resistance in type 2 diabetes mellitus: role of the insulin receptor variant forms. *Diabetes. Metab. Res. Rev.* *17*, 363–373.

- Shatkin, a J., and Manley, J.L. (2000). The ends of the affair: capping and polyadenylation. *Nat. Struct. Biol.* 7, 838–842.
- Sheth, U., and Parker, R. (2003). Decapping and decay of messenger RNA occur in cytoplasmic processing bodies. *Science* 300, 805–808.
- Shi, H., Hoffman, B.E., and Lis, J.T. (1997). A specific RNA hairpin loop structure binds the RNA recognition motifs of the *Drosophila* SR protein B52. *Mol. Cell. Biol.* 17, 2649–2657.
- Shukla, S., Kavak, E., Gregory, M., Imashimizu, M., Shutinoski, B., Kashlev, M., Oberdoerffer, P., Sandberg, R., and Oberdoerffer, S. (2011). CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature* 479, 74–79.
- Simmer, F., Moorman, C., van der Linden, A.M., Kuijk, E., van den Berghe, P.V.E., Kamath, R.S., Fraser, A.G., Ahringer, J., and Plasterk, R.H. a (2003). Genome-wide RNAi of *C. elegans* using the hypersensitive *rrf-3* strain reveals novel gene functions. *PLoS Biol.* 1, E12.
- Smith, C.W., and Valcárcel, J. (2000). Alternative pre-mRNA splicing: the logic of combinatorial control. *Trends Biochem. Sci.* 25, 381–388.
- Spieth, J., Lawson, D., Davis, P., Williams, W., and Howe, K. (2013). Overview of gene structure in *C. elegans*. *WormBook* 1–30.
- Sterner, D. a, and Berget, S.M. (1993). In vivo recognition of a vertebrate mini-exon as an exon-intron-exon unit. *Mol. Cell. Biol.* 13, 2677–2687.
- Stoilov, P., Lin, C.-H., Damoiseaux, R., Nikolic, J., and Black, D.L. (2008). A high-throughput screening strategy identifies cardiotonic steroids as alternative splicing modulators. *Proc. Natl. Acad. Sci.* 105, 11218–11223.
- Stout Jr, R.F., Grubišić, V., and Parpura, V. (2013). A *Caenorhabditis elegans* Locomotion Phenotype Caused by Transgenic Repeats of the *hlh-17* Promoter Sequence. *PLoS One* 8, e81771.
- Sulston, J.E., and Horvitz, H.R. (1977). Post-embryonic cell lineages of the nematode, *Caenorhabditis elegans*. *Dev. Biol.* 56, 110–156.
- Sulston, J.E., Schierenberg, E., White, J.G., and Thomson, J.N. (1983). The embryonic cell lineage of the nematode *Caenorhabditis elegans*. *Dev. Biol.* 100, 64–119.
- Sun, W., You, X., Gogol-Döring, A., He, H., Kise, Y., Sohn, M., Chen, T., Klebes, A., Schmucker, D., and Chen, W. (2013). Ultra-deep profiling of alternatively spliced *Drosophila* Dscam isoforms by circularization-assisted multi-segment sequencing. *EMBO J.* 32, 2029–2038.
- Sze, J.Y., and Ruvkun, G. (2003). Activity of the *Caenorhabditis elegans* UNC-86 POU transcription factor modulates olfactory sensitivity. *Proc. Natl. Acad. Sci. U. S. A.* 100, 9560–9565.
- Sze, J.Y., Zhang, S., Li, J., and Ruvkun, G. (2002). The *C. elegans* POU-domain transcription factor UNC-86 regulates the *tph-1* tryptophan hydroxylase gene and neurite outgrowth in specific serotonergic neurons. *Development* 129, 3901–3911.

- Tacke, R., and Manley, J.L. (1995). The human splicing factors ASF/SF2 and SC35 possess distinct, functionally significant RNA binding specificities. *Embo J* 14, 3540–3551.
- Tacke, R., and Manley, J.L. (1999). Determinants of SR protein specificity. *Curr Opin Cell Biol* 11, 358–362.
- Takahashi, S., Araki, Y., Sakuno, T., and Katada, T. (2003). Interaction between Ski7p and Upf1p is required for nonsense-mediated 3'-to-5' mRNA decay in yeast. *EMBO J.* 22, 3951–3959.
- Talerico, M., and Berget, S.M. (1990). Effect of 5' splice site mutations on splicing of the preceding intron. *Mol. Cell. Biol.* 10, 6299–6305.
- Tamura, T. (1996). Mechanisms of Transcriptional Regulation and Neural Gene Expression. *Neurochem. Int.* 29, 573–581.
- Tanahashi, H., and Tabira, T. (2001). Three novel alternatively spliced isoforms of the human beta-site amyloid precursor protein cleaving enzyme (BACE) and their effect on amyloid beta-peptide production. *Neurosci. Lett.* 307, 9–12.
- Tange, T.Ø., Shibuya, T., Jurica, M.S., and Moore, M.J. (2005). Biochemical analysis of the EJC reveals two new factors and a stable tetrameric protein core. *RNA* 11, 1869–1883.
- The *C. elegans* Sequencing Consortium (1998). Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* 282, 2012–2018.
- Tilgner, H., Nikolaou, C., Althammer, S., Sammeth, M., Beato, M., Valcárcel, J., and Guigó, R. (2009). Nucleosome positioning as a determinant of exon recognition. *Nat. Struct. Mol. Biol.* 16, 996–1001.
- Tonegawa, S., Maxam, a M., Tizard, R., Bernard, O., and Gilbert, W. (1978). Sequence of a mouse germ-line gene for a variable region of an immunoglobulin light chain. *Proc. Natl. Acad. Sci. U. S. A.* 75, 1485–1489.
- Tzur, Y.B., Friedland, A.E., Nadarajan, S., Church, G.M., Calarco, J. a, and Colaiácovo, M.P. (2013). Heritable custom genomic modifications in *Caenorhabditis elegans* via a CRISPR-Cas9 system. *Genetics* 195, 1181–1185.
- Uchida, O., Nakano, H., Koga, M., and Ohshima, Y. (2003). The *C. elegans che-1* gene encodes a zinc finger transcription factor required for specification of the ASE chemosensory neurons. *Development* 130, 1215–1224.
- Uemura, A., Oku, M., Mori, K., and Yoshida, H. (2009). Unconventional splicing of XBP1 mRNA occurs in the cytoplasm during the mammalian unfolded protein response. *J. Cell Sci.* 122, 2877–2886.
- Unterholzner, L., and Izaurralde, E. (2004). SMG7 acts as a molecular link between mRNA surveillance and mRNA decay. *Mol. Cell* 16, 587–596.
- Urlaub, G., Mitchell, P.J., Ciudad, C.J., and Chasin, L. a (1989). Nonsense mutations in the dihydrofolate reductase gene affect RNA processing. *Mol Cell Biol* 9, 2868–2880.

- Venables, J.P., Klinck, R., Koh, C., Gervais-Bird, J., Bramard, A., Inkel, L., Durand, M., Couture, S., Froehlich, U., Lapointe, E., et al. (2009). Cancer-associated regulation of alternative splicing. *Nat. Struct. Mol. Biol.* *16*, 670–676.
- Waaaijers, S., and Boxem, M. (2014). Engineering the *Caenorhabditis elegans* genome with CRISPR/Cas9. *Methods*.
- Waaaijers, S., Portegijs, V., Kerver, J., Lemmens, B.B.L.G., Tijsterman, M., van den Heuvel, S., and Boxem, M. (2013). CRISPR/Cas9-targeted mutagenesis in *Caenorhabditis elegans*. *Genetics* *195*, 1187–1191.
- Wang, J.C. (2002). Cellular roles of DNA topoisomerases: a molecular perspective. *Nat. Rev. Mol. Cell Biol.* *3*, 430–440.
- Ward, S., Thomson, N., White, J.G., and Brenner, S. (1975). Electron microscopical reconstruction of the anterior sensory anatomy of the nematode *Caenorhabditis elegans*. *J. Comp. Neurol.* *160*, 313–337.
- Warf, M.B., Nakamori, M., Matthys, C.M., Thornton, C. a, and Berglund, J.A. (2009). Pentamidine reverses the splicing defects associated with myotonic dystrophy. *Proc. Natl. Acad. Sci. U. S. A.* *106*, 18551–18556.
- Watson, F.L., Puttmann-Holgado, R., Thomas, F., Lamar, D.L., Hugues, M., Kondo, M., Rebel, V.I., and Schmucker, D. (2005). Extensive Diversity of Ig-Superfamily Proteins in the Immune System of Insects. *Science* (80-. ). *309*, 1874–1878.
- Way, J.C., and Chalfie, M. (1988). *mec-3*, a homeobox-containing gene that specifies differentiation of the touch receptor neurons in *C. elegans*. *Cell* *54*, 5–16.
- Weg-Remers, S., Ponta, H., Herrlich, P., and König, H. (2001). Regulation of alternative pre-mRNA splicing by the ERK MAP-kinase pathway. *EMBO J.* *20*, 4194–4203.
- Weil, J.E., and Beemon, K.L. (2006). A 3' UTR sequence stabilizes termination codons in the unspliced RNA of Rous sarcoma virus. *RNA* *12*, 102–110.
- White, J.G., Southgate, E., Thomson, J.N., and Brenner, S. (1976). The structure of the ventral nerve cord of *Caenorhabditis elegans*. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* *275*, 327–348.
- White, J.G., Southgate, E., Thomson, J.N., and Brenner, S. (1986). The Mind of a Worm. *Philos. Trans. R. Soc. B Biol. Sci.* *314*, 1–340.
- Wightman, B., Ha, I., and Ruvkun, G. (1993). Posttranscriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal pattern formation in *C. elegans*. *Cell* *75*, 855–862.
- Wilkinson, M.F., and Shyu, A.-B. (2002). RNA surveillance by nuclear scanning? *Nat. Cell Biol.* *4*, E144–E147.
- Wojtowicz, W.M., Flanagan, J.J., Millard, S.S., Zipursky, S.L., and Clemens, J.C. (2004). Alternative splicing of *Drosophila Dscam* generates axon guidance receptors that exhibit isoform-specific homophilic binding. *Cell* *118*, 619–633.

Wood, A.J., Lo, T., Zeitler, B., Pickle, C.S., Edward, J., Lee, A.H., Amora, R., Miller, J.C., Leung, E., Zhang, L., et al. (2012). Targeted Genome Editing Across Species Using ZFNs and TALENs. *Science* (80-. ). 333, 3–6.

Wu, J.Y., and Maniatis, T. (1993). Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. *Cell* 75, 1061–1070.

Yamashita, A., Ohnishi, T., Kashima, I., Taya, Y., and Ohno, S. (2001). Human SMG-1, a novel phosphatidylinositol 3-kinase-related protein kinase, associates with components of the mRNA surveillance complex and is involved in the regulation of nonsense-mediated mRNA decay. *Genes Dev.* 15, 2215–2228.

Yoshida, H., Matsui, T., Yamamoto, A., Okada, T., and Mori, K. (2001). XBP1 mRNA Is Induced by ATF6 and Spliced by IRE1 in Response to ER Stress to Produce a Highly Active Transcription Factor. *Cell* 107, 881–891.

You, K.T., Li, L.S., Kim, N.-G., Kang, H.J., Koh, K.H., Chwae, Y.-J., Kim, K.M., Kim, Y.K., Park, S.M., Jang, S.K., et al. (2007). Selective translational repression of truncated proteins from frameshift mutation-derived mRNAs in tumors. *PLoS Biol.* 5, e109.

Zahler, A.M. (2012). Pre-mRNA splicing and its regulation in *Caenorhabditis elegans*. *WormBook* 1–21.

Zhang, M.L., Lorson, C.L., Androphy, E.J., and Zhou, J. (2001). An in vivo reporter system for measuring increased inclusion of exon 7 in SMN2 mRNA: potential therapy of SMA. *Gene Ther.* 8, 1532–1538.

Zhu, H., Hummel, T., Clemens, J.C., Berdnik, D., Zipursky, S.L., and Luo, L. (2006). Dendritic patterning by Dscam and synaptic partner matching in the *Drosophila* antennal lobe. *Nat. Neurosci.* 9, 349–355.

Zipursky, S.L., Wojtowicz, W.M., and Hattori, D. (2006). Got diversity? Wiring the fly brain with Dscam. *Trends Biochem. Sci.* 31, 581–588.

Cáceres, J. F., & Krainer, A. R. (1997). Mammalian pre-mRNA splicing factors. *Eukaryotic mRNA processing*, 174-212.

Mello, C., & Fire, A. (1995). DNA transformation. *Methods Cell Biol.* 48(48), 451-482.

Feng, H., Craig, H. L., & Hope, I. A. (2012). Expression pattern analysis of regulatory transcription factors in *Caenorhabditis elegans*. In *Gene Regulatory Networks* (pp. 21-50). Humana Press.