



**HAL**  
open science

# Rôle du complexe de jonction exon-exon (EJC) dans l'épissage de l'ARN pré-messager du gène DMD et dans la différenciation myogénique.

Dylan da Cunha

## ► To cite this version:

Dylan da Cunha. Rôle du complexe de jonction exon-exon (EJC) dans l'épissage de l'ARN pré-messager du gène DMD et dans la différenciation myogénique.. Biochimie, Biologie Moléculaire. Université de Montpellier, 2022. Français. NNT : 2022UMONT077 . tel-04631561

**HAL Id: tel-04631561**

**<https://theses.hal.science/tel-04631561>**

Submitted on 2 Jul 2024

**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 POUR OBTENIR LE GRADE DE DOCTEUR DE L'UNIVERSITÉ DE MONTPELLIER

En Biologie Santé

École doctorale : Science Chimiques et Biologiques pour la Santé (CBS2)

PhyMedExp, Université de Montpellier, INSERM U1046, CNRS UMR 9214, Montpellier, France

**Rôle du complexe de jonction exon-exon (EJC)  
dans l'épissage de l'ARN pré-messager du gène  
*DMD* et dans la différenciation myogénique.**

Présentée par Dylan DA CUNHA

Le 13 décembre 2022

Sous la direction du Dr Sylvie TUFFERY-GIRAUD

Devant le jury composé de

Dr Cyril VAILLEND, Directeur de Recherche CNRS, Institut NeuroPSI, Orsay

Dr Fabrice LEJEUNE, Directeur de Recherche INSERM, Institut Pasteur, Lille

Dr Hervé LE HIR, Directeur de Recherche CNRS, IBENS, ENS, Paris

Pr Michel KOENIG, PU-PH Université de Montpellier, PhyMedExp

Dr Sylvie TUFFERY-GIRAUD, Ingénieur de Recherche Université de Montpellier, PhyMedExp

Rapporteur

Rapporteur

Examineur

Président

Directrice de thèse



UNIVERSITÉ  
DE MONTPELLIER

*“I hated every minute of training, but I said, 'Don't quit. Suffer now and live the rest of your life as a champion.’”*  
*Mohammed Ali*

# Remerciements

Je tenais tout d'abord à remercier, le Dr Fabrice Lejeune, le Dr Cyrille Vailland ainsi que le Dr Hervé le Hir d'avoir accepté d'être les membres de mon jury de thèse, d'avoir pris le temps de lire ce manuscrit et de leur implication dans mon projet.

Je remercie le laboratoire PhyMedExp de l'Université de Montpellier ainsi que le laboratoire de Génétique moléculaire du CHU de Montpellier de m'avoir accueilli et d'avoir permis de réaliser ma thèse.

Je tenais ensuite à remercier particulièrement le Dr Sylvie Tuffery-Giraud pour sa grande présence durant ces quatre années de thèse, sa pédagogie, ses conseils, sa vision de la science, ses connaissances ainsi que son implication dans mon travail qui ont permis de finaliser ce projet.

Je voudrais aussi remercier le Pr Michel Koenig, qui dirige l'équipe 5 de PhyMedExp à laquelle j'appartiens, qui possède énormément de connaissances sur la génétique et qui m'a permis d'aborder un sujet d'aspect clinique aboutissant à une publication.

J'aimerais particulièrement remercier l'AFM-téléthon pour avoir financé mon projet de thèse durant 4 années ainsi que pour son soutien à tous les projets de recherche en France et à l'international.

## Résumé

### **Titre : Rôle du complexe de jonction exon-exon (EJC) dans l'épissage de l'ARN pré-messager du gène DMD et dans la différenciation myogénique.**

L'épissage alternatif des gènes chez les eucaryotes est un mécanisme finement régulé qui permet d'augmenter la complexité du protéome et de diversifier les fonctions cellulaires. Dans les gènes humains, la reconnaissance des sites authentiques d'épissage par le spliceosome est rendue difficile en raison du nombre élevé d'exons et de la présence d'introns de grande taille comme observé dans le gène *DMD* long de 2.2 mégabases et contenant 79 exons. S'ajoutant aux séquences *cis* et facteurs *trans* déjà connus, il a récemment été décrit que le complexe de jonction exon-exon (EJC) contribue au maintien de l'intégrité du transcriptome. Le dépôt du complexe composé de quatre protéines cœur, (eIF4A3, Y14, Magoh et MLN51), 24 nucléotides en amont de chaque jonction exon-exon réprime notamment la survenue d'événements de ré-épissage indésirables. Durant ma thèse, j'ai étudié le rôle de l'EJC dans l'épissage de l'isoforme musculaire (Dp427m) du gène *DMD* qui inclut les 79 exons et de l'isoforme ubiquitaire courte (Dp71) alternativement épissée dans le cerveau, en utilisant des approches d'ARNi et de RNA-Seq ciblée dans des lignées cellulaires humaines. Nous montrons que la diminution du niveau d'expression des facteurs de l'EJC dérégule l'épissage des exons 68 à 78 de la région 3' des transcrits de manière similaire dans les deux isoformes et que les complexes associés ASAP/PSAP participent au contrôle de l'épissage des exons 71 et 78 dont l'inclusion est hautement régulée en fonction des tissus et/ou du stade de développement. Par ailleurs, nous rapportons que la diminution d'expression d'eIF4A3 empêche l'engagement de la différenciation des myoblastes en myotubes. Ainsi, nos données établissent que l'EJC contribue à la fidélité de l'épissage des exons codant pour le domaine C-terminal de la dystrophine, un domaine d'interaction protéine-protéine essentiel pour les fonctions de soutien et de signalisation de la dystrophine, et joue également un rôle dans la différenciation myogénique.

**Mots clés :** Épissage, EJC, DMD, Dp427m, Dp71, Différenciation musculaire

## Abstract

### **Title: Role of the Exon Junction Complex (EJC) in DMD gene pre-messenger RNA splicing and myogenic differentiation.**

Alternative splicing is a finely regulated mechanism in eukaryotes that allows increasing the complexity of the proteome and diversifying cellular functions. In human genes, recognition of authentic splice sites by the spliceosome is challenged by the high number of exons and the presence of large introns as observed in the 2.2 megabase long *DMD* gene containing 79 exons. In addition to the already known *cis* sequences and *trans* factors, it has been reported that the exon junction complex (EJC) contributes to the maintenance of transcriptome integrity. The deposition of the complex composed of four core proteins (eIF4A3, Y14, Magoh and MLN51) 24 nucleotides upstream of each exon-exon junction represses the occurrence of spurious re-splicing events. During my thesis, I studied the role of EJC in the splicing of the muscle isoform (Dp427m) of the *DMD* gene which includes all 79 exons and of the ubiquitous isoform (Dp71) characterized by multiple alternative splicing in the brain, using RNAi and targeted RNA-Seq approaches in human cell lines. We have shown that decreasing the expression level of EJC factors deregulates the splicing of exons 68 to 78 in the 3' region of the transcripts in a similar manner in both isoforms and that the associated ASAP/PSAP complexes are also involved in the control of the splicing of exons 71 and 78, whose inclusion level is highly regulated according to tissue and/or developmental stage. We also observed that decreased expression of eIF4A3 prevents the initiation of myoblast differentiation into myotubes. Thus, our data show that the EJC contributes to the fidelity of splicing of exons encoding the C-terminal domain of dystrophin, a protein-protein interaction domain essential for the structural and signaling functions of dystrophin and plays a role in myogenic differentiation.

**Key words:** Splicing, EJC, DMD, Dp427m, Dp71, Muscle differentiation

---

Laboratoire de Physiologie et Médecine Expérimentale du cœur et des muscles (PhyMedExp),  
Université de Montpellier, INSERM U1046, CNRS UMR 9214, 641 avenue du Doyen Gaston Giraud,  
IURC, 34093 Montpellier, France

---

# Table des matières

Remerciements.....	3
Résumé.....	4
Abstract .....	4
Table des figures.....	8
Liste des abréviations.....	10
Introduction générale .....	12
État de l'art.....	15
<b>Partie 1 - Le processus d'épissage des gènes.....</b>	<b>16</b>
I. L'épissage constitutif .....	17
A. Le mécanisme d'épissage .....	17
1. Les séquences cis de reconnaissance des introns .....	17
2. La réaction de l'épissage .....	18
3. Le spliceosome .....	18
4. Reconnaissance des sites d'épissage .....	23
II. L'épissage alternatif.....	27
A. Les mécanismes de l'épissage alternatif.....	27
1. Les différentes classes d'AS.....	27
2. Les déterminants de l'épissage alternatif .....	28
B. Autres processus de régulation de l'épissage .....	34
2. Le couplage épissage et marques épigénétiques.....	36
3. L'épissage alternatif couplé au système NMD.....	38
C. L'épissage dans le muscle .....	39
1. Les facteurs spécifiques du muscle strié.....	39
2. L'épissage alternatif du muscle squelettique.....	39
<b>Partie 2 : Le complexe de jonction d'exons (EJC) .....</b>	<b>41</b>
I. Les composants du cœur de l'EJC .....	43
A. eIF4A3 (DDX48).....	43
B. Y14/MAGOH.....	44
C. MLN51 .....	44
II. Cycle de vie de l'EJC.....	45
A. Recrutement de l'EJC dépendant de l'épissage.....	45
B. Assemblage de l'EJC.....	45
1. Recrutement d'eIF4A3 par CWC22/CWC27 .....	46
2. Fixation de l'hétérodimère Y14/MAGOH sur eIF4A3.....	47
C. EJC trimérique vs tétramérique .....	48
D. Recrutement des protéines périphériques de l'EJC .....	49
1. Les complexes ASAP/PSAP .....	49
2. Les protéines SR.....	50
3. Les protéines d'export des ARNm .....	50
4. Les facteurs à motif EBM.....	50
E. Mise en place des structures mRNP autour de l'EJC .....	50
III. Localisation de l'EJC.....	52
A. Dépôt des EJC sur le transcriptome.....	52
1. Position des EJC canoniques et non-canoniques .....	52
2. Structure secondaire de l'ARN et dépôt de l'EJC.....	54
B. Localisation de l'EJC dans la cellule .....	54
IV. Les rôles de l'EJC.....	55
A. Rôle nucléaire de l'EJC.....	55
1. Rôle de l'EJC et des protéines associées dans l'épissage.....	55
2. MLN51 controversé comme facteur d'épissage.....	57
3. Rôle des complexes ASAP/PSAP.....	57

4.	Fonction conservée de l'EJC : Répression du mécanisme de ré-épissage.....	58
5.	Rôle de l'EJC dans l'épissage co-transcriptionnel.....	61
B.	Rôle de l'EJC dans l'export des ARNm.....	63
C.	Rôle cytoplasmique de l'EJC.....	64
1.	Rôle de l'EJC dans la traduction.....	64
2.	Rôle de l'EJC dans la surveillance des ARNm.....	65
V.	Désassemblage et recyclage de l'EJC.....	72
VI.	Rôle physiologique de l'EJC en pathologies.....	74
<b>Partie 3 :</b>	<b>Le gène de la Dystrophine.....</b>	<b>75</b>
I.	Le gène DMD (Duchenne Muscular Dystrophy).....	75
II.	Les dystrophinopathies.....	75
A.	Les mutations du gène DMD.....	75
B.	Les formes cliniques de DMD et BMD.....	76
1.	La DMD.....	76
2.	La BMD.....	77
III.	Structure de la dystrophine.....	78
IV.	Le complexe dystrophine et protéines associées (DAPC).....	79
V.	Les différentes isoformes du gène DMD.....	81
A.	Les promoteurs alternatifs du gène DMD.....	81
1.	Les isoformes longues.....	81
2.	Les isoformes courtes.....	83
VI.	Spécificité tissulaire et cellulaire du DAPC.....	86
VII.	Épissages alternatifs des isoformes de dystrophine.....	89
A.	Épissage alternatif des isoformes non musculaires.....	90
1.	La Dp427c.....	90
2.	Les Dp260, Dp140, Dp116.....	90
B.	Régulation de l'épissage de la Dp427 musculaire.....	94
1.	Épissage récursif de la Dp427m.....	94
2.	Épissage non séquentiel du transcrit Dp427m.....	94
3.	Épissage alternatif de la Dp427m.....	95
4.	Épissage de la Dp427m au cours de la myogenèse.....	97
<b>Partie 4 :</b>	<b>La différenciation du muscle squelettique.....</b>	<b>99</b>
I.	Le muscle squelettique.....	99
A.	La myogenèse.....	100
B.	La différenciation musculaire.....	101
II.	Hélicases et différenciation musculaire.....	102
A.	UPF1.....	102
1.	Balance NMD/SMD.....	103
2.	Interaction d'UPF1 avec MyoD.....	105
3.	Les hélicases DDX5/17.....	106
4.	L'EJC dans la différenciation musculaire.....	107
<b>Objectifs de la thèse.....</b>	<b>108</b>	
<b>Principaux résultats obtenus durant la thèse.....</b>	<b>112</b>	
<b>Résultats.....</b>	<b>113</b>	
<b>Materials and methods.....</b>	<b>114</b>	
<b>Results.....</b>	<b>119</b>	
<b>Supplementary information.....</b>	<b>133</b>	
<b>Résultats complémentaires.....</b>	<b>155</b>	
I.	Rôle de l'EJC dans l'épissage de DMD et transcription.....	155
A.	Introduction.....	155
B.	Matériel et méthodes.....	156
C.	Résultats expérimentaux.....	157

1.	Profil d'épissage de DMD dans les cellules C25Cl48 en prolifération vs différenciation .....	157
2.	La CPT induit des variations sur les ASE .....	157
D.	Discussion et perspectives .....	158
II.	Dépôt de l'EJC et régulation de l'épissage de DMD .....	159
A.	Introduction.....	159
B.	Matériel et méthodes.....	160
1.	Analyse préliminaire des données de iCLIP et de xRIPIT-seq .....	160
2.	Expériences de RNA pull-down.....	161
C.	Résultats.....	161
1.	Visualisation des cEJC sur les exons 68 à 79 du gène DMD.....	161
2.	Hypothèse de dépôts de ncEJC dans le transcrit DMD.....	162
3.	Validation du dépôt EJC en position non canonique sur l'exon 70.....	163
4.	Validations du dépôt RNPS1 dans l'intron 77.....	164
D.	Discussion et perspective .....	165
III.	Résultats complémentaires d'UPF1 par l'étude d'UPF2 .....	166
A.	Introduction.....	166
B.	Matériel et méthodes.....	167
C.	Résultats.....	167
1.	UPF2 possède un rôle différent d'UPF1 .....	167
2.	UPF2 bloque la différenciation des C25cl48 .....	168
D.	Discussion et perspectives .....	169
	<b>Discussion générale .....</b>	<b>171</b>
	<b>Bibliographie .....</b>	<b>182</b>
	<b>ANNEXES .....</b>	<b>206</b>



## Table des figures

FIGURE 1. REPRESENTATION DES SEQUENCES DES SITES D'ÉPISSAGE DANS LES INTRONS DES PRE-ARNM DE VERTEBRES. ....	17
FIGURE 2. REPRESENTATION DES DEUX ETAPES DE TRANSESTERIFICATION DE LA REACTION D'ÉPISSAGE. ....	18
FIGURE 3. COMPOSITION DES SNRNP DU SPLICEOSOME HUMAIN. ....	19
FIGURE 4. REPRESENTATION DES COMPOSANTS ET DES INTERACTEURS DU SPLICEOSOME HUMAIN. ....	20
FIGURE 5. REPRESENTATION SCHEMATIQUE DES ETAPES DE REMODELAGE DES SNRNP AU COURS DE L'ÉPISSAGE. ....	23
FIGURE 6. REPRESENTATION DES DEUX MODELES DE DEFINITION DE L'ÉPISSAGE. ....	24
FIGURE 7. REPRESENTATION DE L'ÉPISSAGE RECURSIF DANS LES LONGS INTRONS. ....	25
FIGURE 8. MODELE DE SELECTION STOCHASTIQUE DES SITES D'ÉPISSAGE RECURSIFS. ....	26
FIGURE 9. LES DIFFERENTES CLASSES D'ÉPISSAGE ALTERNATIF. ....	28
FIGURE 10. REGULATION DE L'ÉPISSAGE DES PRE-ARNM PAR LES SEQUENCES CIS-REGULATRICES. ....	30
FIGURE 11. LISTE DES DIFFERENTES PROTEINES SR ET SR-LIKE CHEZ L'HOMME. ....	32
FIGURE 12. LISTE DES DIFFERENTES HNRNP CHEZ L'HOMME. ....	33
FIGURE 13. LE COUPLAGE DES PROCESSUS DE TRANSCRIPTION ET D'ÉPISSAGE. ....	35
FIGURE 14. LES DIFFERENTS NIVEAUX DE REGULATION DE LA TRANSCRIPTION ET DE L'ÉPISSAGE ALTERNATIF. ....	37
FIGURE 15. EXEMPLES D'ÉPISSAGE ALTERNATIF CONDUISANT A L'INTRODUCTION D'UN PTC DANS LES ARNm. ....	38
FIGURE 16. LES PRINCIPAUX REGULATEURS DE L'ÉPISSAGE ALTERNATIF DANS LE MUSCLE. ....	40
FIGURE 17. LES TROIS NIVEAUX D'ASSEMBLAGE DES FACTEURS DE L'EJC ET DES PROTEINES ASSOCIEES. ...	42
FIGURE 18. ASSEMBLAGE DE L'EJC DURANT L'ÉPISSAGE ET SES ROLES CYTOPLASMIQUES CONNUS. ....	43
FIGURE 19. REPRESENTATION SCHEMATIQUE DES DOMAINES PROTEIQUES DE EIF4A3. ....	44
FIGURE 20. FORMATION DU COMPLEXE EJC. ....	46
FIGURE 21. MECANISME DE FIXATION DE L'HELICASE A ARN EIF4A3. ....	47
FIGURE 22. COMPOSITION DES COMPLEXES ASAP ET PSAP LIES A L'EJC. ....	49
FIGURE 23. INTERACTOME DE L'EJC FAVORISANT LA FORMATION DES STRUCTURES COMPACTES DE MRNP. ....	51
FIGURE 24. DYNAMIQUE DE REGULATION DES STRUCTURES MRNP EN FONCTION DE L'INTERACTOME DE L'EJC. ....	52
FIGURE 25. REPRESENTATION SCHEMATIQUE DE L'ÉPISSAGE D'UN EXON RECURSIF (RS) REGULE PAR L'EJC ASSOCIE A RNPS1. ....	53
FIGURE 26. LOCALISATION SUBCELLULAIRE DE L'ASSEMBLAGE DE L'EJC CO-TRANSCRIPTIONNEL. ....	55
FIGURE 27. REGULATION DES SITES CRYPTIQUES D'ÉPISSAGE PAR L'EJC ASSOCIE AU COMPLEXE PSAP. ....	59
FIGURE 28. REGULATION DE L'ÉPISSAGE DES EXONS RS PAR L'EJC. ....	60
FIGURE 29. ROLE DE L'EJC DANS L'ÉPISSAGE CO-TRANSCRIPTIONNEL. ....	62
FIGURE 30. SCHEMA GLOBAL DE LA REGULATION DE L'ÉPISSAGE PAR L'EJC. ....	63
FIGURE 31. SUBSTRATS ACTIVANT LE SYSTEME NMD. ....	66
FIGURE 32. RECRUTEMENT DU SYSTEME NMD "CLASSIQUE" PAR L'EJC. ....	68
FIGURE 33. MECANISME D'ACTIVATION DU NMD EN PRESENCE DE LONGS 3'UTR. ....	69
FIGURE 34. SYSTEME NMD INDEPENDANT D'UPF3B. ....	70
FIGURE 35. SYSTEME NMD INDEPENDANT D'UPF2. ....	71
FIGURE 36. MECANISMES DE REGULATION DU SYSTEME NMD. ....	72
FIGURE 37. ROLES ET CYCLE DE VIE DE L'EJC. ....	73
FIGURE 38. DOMAINES D'INTERACTION DE L'ISOFORME MUSCULAIRE DE LA DYSTROPHINE (DP427M). ....	78
FIGURE 39. FONCTIONS ET INTERACTIONS CELLULAIRES DE LA DP427M. ....	80
FIGURE 40. LES DIFFERENTS PROMOTEURS DU GENE DMD. ....	81
FIGURE 41. SPECIFICITE TISSULAIRE DE L'EXPRESSION DES ISOFORMES DU GENE DMD. ....	85
FIGURE 42. EXPRESSION DES ISOFORMES DE DYSTROPHINE AU COURS DU DEVELOPPEMENT DU CERVEAU HUMAIN. ....	86
FIGURE 43. LES DIFFERENTS COMPLEXES DAPC EN FONCTION DES TISSUS. ....	88
FIGURE 44. MODELES DE COMPOSITION DES DAPC DANS DIFFERENTS TYPES CELLULAIRES. ....	89
FIGURE 45. ÉPISSAGES ALTERNATIFS DE LA DP427C DANS LA LIGNEE HUMAINE NEURONALE SH-SY5Y. ....	90
FIGURE 46. NOMENCLATURE DES ÉPISSAGES ALTERNATIFS DE LA DP71. ....	92
FIGURE 47. DIFFERENTES FONCTIONS DES SOUS-ISOFORMES DE LA DP71 EN FONCTION DE LEUR LOCALISATION SUBCELLULAIRE. ....	93

FIGURE 48. ÉPISSAGE SEQUENTIEL ET NON SEQUENTIEL DES INTRONS DE LA DP427M. ....	95
FIGURE 49. ÉPISSAGE ALTERNATIF PHYSIOLOGIQUE DANS LE MUSCLE SQUELETTIQUE. ....	96
FIGURE 50. STRUCTURE TISSULAIRE ET CELLULAIRE DU MUSCLE SQUELETTIQUE. ....	100
FIGURE 51. MYOGENESE ET DIFFERENCIATION DU MUSCLE SQUELETTIQUE. ....	102
FIGURE 52. INTERACTION D'UPF1 ET STAUFEN-1 DANS LE SYSTEME SMD. ....	103
FIGURE 53. BALANCE NMD/SMD DEPENDANT DE UPF1. ....	104
FIGURE 54. BALANCE DES VOIES DE NMD/SMD DANS LA DIFFERENCIATION MUSCULAIRE. ....	104
FIGURE 55. MECANISME D'ACTION DE UPF1 SUR L'EXPRESSION DE MYOD DURANT LA MYOGENESE.....	105
FIGURE 56. ROLE DE DDX5/17 DANS LA DIFFERENCIATION MUSCULAIRE. ....	107
FIGURE 57. ANALYSE DE L'EXPRESSION DE LA DP427M DANS LES CELLULES C25CL48 AU COURS DE LA DIFFERENCIATION MYOGENIQUE. ....	109
FIGURE 58. VALIDATION DU MODELE CELLULAIRE C25CL48 COMPARE AU MUSCLE SQUELETTIQUE. ....	110
FIGURE 59. METHODOLOGIE DU PROJET DE THESE. ....	111
FIGURE 60. SPLICING PROFILE OF THE DP427M ISOFORM IN EIF4A3- OR Y14-DEPLETED HUMAN C25CL48 MUSCLE CELLS. ....	120
FIGURE 61. VALIDATION OF ALTERNATIVE SPLICING EVENTS IN THE DMD GENE UPON DEPLETION OF EJC COMPONENTS. ....	123
FIGURE 62. EFFECT OF UPF1 KNOCKDOWN ON EJC-DEPENDENT SPLICING EVENTS.....	125
FIGURE 63. VALIDATION OF THE HELa CELL LINE AS A MODEL TO STUDY EJC-DEPENDENT SPLICING EVENTS IN THE DMD GENE. ....	127
FIGURE 64. EIF4A3 AND Y14 INDIVIDUALLY CONTRIBUTE TO THE INCLUSION OF DMD EXONS BUT COMPLEX FORMATION IS REQUIRED FOR CORRECT DMD SPLICING. ....	128
FIGURE 65. THE EJC-PERIPHERAL PROTEINS REGULATE SPLICING OF DMD EXONS 71 AND 78. ....	129
FIGURE 66. DEPLETION OF EJC COMPONENTS IN MUSCULAR C25CL48 CELLS DECREASES DYSTROPHIN EXPRESSION AND IMPAIRS MYOGENIC DIFFERENTIATION.....	131
FIGURE 67. ROLE DE L'EJC DANS L'EPISSAGE COUPLE A LA TRANSCRIPTION. ....	156
FIGURE 68. GEL D'AGAROSE MONTRANT L'EFFET DE LA CPT SUR L'EPISSAGE ALTERNATIF DE DMD. ....	157
FIGURE 69. SEQUENCES DES SONDAS DES EXONS 70 ET 78 UTILISEES DANS LES EXPERIENCES DE RNA PULL- DOWN.....	161
FIGURE 70. DISTRIBUTION DES DEPOTS EJC ET DE RNPS1 SUR LA PARTIE DISTALE DU GENE DMD (EXONS 68 A 79). ....	162
FIGURE 71. ANALYSES DES DONNEES DE ICLIP ET xRIPIT SUR L'EXON 70 ET L'INTRON 77 DE DMD. ....	163
FIGURE 72. WESTERN BLOT REVELANT LA FIXATION D'EIF4A3 SUR LES SONDAS DE L'EXON 70 ANALYSEES EN RNA PULL-DOWN. ....	164
FIGURE 73. ÉTUDE DU SITE DE FIXATION DE RNPS1.....	165
FIGURE 74. EFFET DU SI-UPF2 SUR L'EPISSAGE ET L'EXPRESSION DE DMD. ....	168
FIGURE 75. QUANTIFICATION DE L'EXPRESSION DES FACTEURS DE LA MYOGENESE ET DE LA DP427M. ....	169
FIGURE 76. RESULTATS PRINCIPAUX DU PROJET DE THESE. ....	181

## Liste des abréviations

3'UTR : Région 3' non traduite  
5'UTR : Région 5' non traduite  
ACINUS : Apoptotic Chromatin Condensation Inducer 1  
ARNi/siRNA : ARN interférents  
ARNm : Acide RiboNucléique messenger mature  
AS : épissage alternatif  
ASAP : apoptosis and splicing-associated protein  
ATP : Adénosine triphosphate  
BMD : Becker Muscular Dystrophy  
pb : paire de base  
BS: site de branchement / Branch Site  
BS-A : Adénosine du site de branchement  
CBC : Complexe de liaison à la coiffe  
CDE : Complexe de définition d'exon  
CLIP-Seq : crosslinking and immunoprecipitation of RNA-protein complexes  
CPT : Camptothécine  
Cryo-EM : Cryo-microscopie électronique  
CTD : domaine carboxy- terminale  
CWC22/27 : Spliceosome Associated Protein Homolog  
DAPC : Dystrophin-Associated Proteins Complex  
DDX : protéines à motif DEAD  
DMD : Dystrophie musculaire de Duchenne  
DMD : gène de la dystrophie musculaire de Duchenne  
EBM : motif de liaison à l'EJC  
eIF4A3 : eucaryotic initiation factor 4A3  
EJC : Complexe de jonction exon-exon  
ESE : activateur d'épissage exonique  
ESS : répresseur d'épissage exonique  
FUBP1 : Far Upstream element-Binding Protein 1  
hnRNP : ribonucléoprotéines nucléaires hétérogènes  
IBC : Complexe de liaison à l'intron  
ILS : Spliceosome contenant le lariat d'intron  
ISE : activateur d'épissage intronique  
ISS : répresseur d'épissage intronique  
m<sup>1</sup>G : mono-méthyl-guanosine  
m<sup>3</sup>G : tri-méthyl-guanosine  
MAGOH : mago-nashi homolog

MBNL1 : muscleblind-like 1  
MIF4G : Middle of Initiation Factor 4G  
MLN51 Metastatic Lymph Node 51  
mRNP : particules de ribonucléoprotéines messagers  
NLS : Nuclear export signal  
NMD : Système de dégradation des ARNm portant un PTC  
nt : nucléotides/nucléotides  
NTC : Complexe lié à Prp19  
NTR : Complexe lié au NTC  
ORF : Open Reading Frame  
PABP : protéine se liant à la queue polyA  
PAP : Polymérase PolyA  
PE : exon poison  
PI : intron poison  
PININ : Desmosome Associated Protein  
Pol II/III : ARN polymérase II ou III  
PPT : Tract de poly-pyrimidine  
Pré-ARNm : Acide Ribonucléique pré-messager  
PSAP : Pinin and splicing-associated protein  
PTC : Codon stop prématuré  
pTEFb : positive transcription elongation factor  
RAF : facteur associés aux RBP  
RBP : Protéines de liaison à l'ARN  
RNA-Seq : Séquence de l'ARN  
RNPS1 : RNA-binding protein with serine-rich domain 1  
RRM : motif de reconnaissance de l'ARN  
RS : épissage récursif  
SAP18 : Sin3A Associated Protein 18  
SELOR SpEckle LOcalizer and RNA-binding module  
SMD : Staufen1 (STAU1)-mediated mRNA decay (SMD)  
SF : facteur d'épissage  
SNC : système nerveux central  
snRNA : petits ARN nucléaires  
snRNP : petites ribonucléoprotéines nucléaires  
SR : serine/arginine-rich domain  
TTN : gène de la titine  
UPF : Up Framshift proteins  
xRIPit-seq : RIP en tandem couplée à l'utilisation de formaldéhyde pour l'étape de cross-link  
Y14/RBM8A : RNA Binding Motif Protein 8

## Introduction générale

Mes travaux de thèse s'inscrivent dans la thématique développée au sein de l'équipe 5 de PhyMedExp concernant l'étude de la régulation de l'épissage du gène *DMD* (*Duchenne Muscular Dystrophy*) dont les mutations sont responsables de dystrophies musculaires de Duchenne et de Becker. Le gène *DMD* code pour la dystrophine, une grande protéine du cytosquelette qui joue un rôle majeur dans le maintien de l'intégrité des fibres musculaires lors de la contraction.

La régulation spatio-temporelle de l'expression de ce gène immense (2.2 mégabases) est particulièrement complexe. Sept promoteurs et des épissages alternatifs génèrent un grand nombre d'isoformes, principalement dans les muscles squelettiques, cardiaques, le système nerveux central et la rétine. Les profils d'épissage diffèrent selon les isoformes et le tissu d'expression. Ainsi, dans le muscle squelettique adulte, les 79 exons du gène sont inclus dans le transcrit mature pour produire une protéine pleine longueur (Dp427m), au contraire des isoformes exprimées dans le cerveau comme la Dp427c ou la Dp71, caractérisées par de très nombreux épissages alternatifs.

La compréhension de l'importance fonctionnelle et biologique des isoformes produites dans les différents tissus reste incomplète à ce jour. De même, les mécanismes moléculaires qui déterminent les différences d'épissage alternatif entre les tissus ont été très peu étudiés.

L'élimination des introns du pré-ARNm lors de la réaction d'épissage est catalysée par le spliceosome, un large complexe composé de cinq petites particules nucléaires ribonucléoprotéiques (snRNP) et plus de 150 de protéines. Chez les eucaryotes supérieurs, de multiples paramètres modulent la sélection des sites d'épissage (la force des sites, la présence ou le niveau d'expression de protéines régulatrices activatrices ou inhibitrices, la structure secondaire de l'ARN, le couplage fonctionnel avec la transcription, ...) et confèrent la flexibilité nécessaire à la réalisation de l'épissage alternatif qui diversifie les transcrits produits à partir d'un seul gène en fonction des besoins cellulaires et tissulaires.

Les protéines de liaison à l'ARN (RBP) qui se lient aux ARNm épissés au sein des particules de ribonucléoprotéines (mRNP) interviennent dans leur stabilité, leur localisation ou l'efficacité de leur traduction de ces ARNm. Le complexe de jonction exon-exon (EJC) est un constituant majeur des mRNP. Il est déposé comme une marque de l'épissage en amont des jonctions exon-exon et est bien connu pour son rôle dans l'export de l'ARNm du noyau vers le cytoplasme où

il est impliqué dans l'initiation de la traduction et la dégradation des ARNm porteurs d'un codon stop prématuré par le système du NMD (*Nonsense-Mediated mRNA Decay*).

Des études récentes ont montré que l'EJC intervient également en amont lors de l'épissage de l'ARN pré-messager. Son dépôt a des effets immédiats sur l'intégrité et la maturation correcte des ARNm, particulièrement dans les gènes possédant de longs introns. Les EJC présents aux jonctions exon-exon nouvellement formées permet notamment de réprimer les événements de ré-épissage ou l'utilisation de sites cryptiques dans le pré-ARNm en cours d'épissage (Schlautmann and Gehring, 2020).

Nous avons émis l'hypothèse que l'EJC pourrait être un déterminant important de la fidélité de l'épissage du gène *DMD* compte tenu de sa taille, de sa structure complexe et de la durée exceptionnelle de sa transcription (environ 10 heures). L'épissage de ce très grand pré-ARNm est majoritairement non séquentiel. Il nécessite un épissage récursif pour éliminer les introns de très grande taille et très probablement la contribution de nombreuses protéines régulatrices pour guider et positionner correctement le spliceosome sur les sites d'épissage authentiques des exons dispersés au sein de cette immense séquence.

Au cours de ma thèse, j'ai étudié le rôle de l'EJC dans l'épissage des 79 exons de l'isoforme musculaire (Dp427m) du gène *DMD* par des approches d'ARN interférence ciblant les composants de l'EJC dans des cellules musculaires et en analysant le profil d'épissage du transcrit *DMD* par RNA-Seq ciblé. J'ai mis en évidence le rôle de l'EJC et ses protéines associées dans l'épissage de la partie distale du gène *DMD*. Cette région de la dystrophine est fonctionnellement importante. Elle est impliquée dans de nombreuses interactions protéine-protéine essentielles à son ancrage au sarcolemme ainsi que pour ses fonctions dans la signalisation cellulaire. De plus, nous avons aussi observé un rôle potentiel de l'EJC dans la différenciation musculaire.

Le manuscrit de thèse comporte trois parties. Une première grande partie est consacrée à une revue de la littérature divisée en quatre sous-chapitres qui décrivent : (1) l'épissage constitutif, alternatif et la régulation de l'épissage dans le tissu musculaire, (2) le complexe EJC, ses composants et fonctions en développant plus particulièrement son rôle dans la régulation l'épissage des gènes, (3) le gène *DMD*, ses isoformes et la régulation de son expression dans les différents tissus et plus précisément dans le muscle, et enfin (4) le processus de différenciation musculaire.

La deuxième partie inclut en premier lieu les objectifs de la thèse et la présentation de la méthodologie utilisée en faisant référence à un article figurant en Annexe 1 dont je suis co-auteur et qui a contribué à la mise en place de mon projet. Les résultats de mes travaux de thèse faisant partie d'un article en préparation dont je suis premier auteur sont ensuite présentés, suivis par des résultats complémentaires. Les données obtenues sont ensuite discutées et mises en perspective des connaissances actuelles concernant le rôle de l'EJC dans l'épissage des gènes humains et la régulation de l'expression des isoformes du gène *DMD*.

La dernière partie correspond aux Annexes comprenant l'article dont je suis co-auteur discuté dans la présentation de la méthodologie (Annexe 1) et un deuxième article paru en 2022 dans lequel je suis co-premier auteur auquel j'ai contribué en analysant la stabilité des transcrits mutés du gène *RFC1* chez des patients atteints du syndrome de CANVAS (Annexe 2). Ce dernier article ne sera pas discuté dans le manuscrit de thèse.

# État de l'art



## Partie 1 - Le processus d'épissage des gènes

L'expression des gènes chez les eucaryotes est un processus hautement régulé impliquant la transcription, la maturation de l'ARN pré-messager (pré-ARNm) nucléaire, l'export, la traduction et la dégradation du transcrit mature dans le cytoplasme. L'étape de maturation des ARN pré-messagers transcrits par la polymérase II (Pol II) inclut la mise en place de la coiffe 7-méthylguanosine à l'extrémité 5' des transcrits nouvellement synthétisés. Celle-ci se lie au complexe de liaison à la coiffe (CBC) qui sert de plateforme d'interaction pour de nombreux facteurs (Rambout and Maquat, 2020). Vient ensuite l'épissage des introns puis l'ajout d'une séquence poly-Adénosine (poly(A)) dans la région 3' UTR des transcrits par la polyA polymérase (PAP) (Xu et al., 2022).

L'épissage des pré-ARNm est essentiellement co-transcriptionnel chez la levure comme dans les cellules humaines car les transcrits en cours de synthèse sont accessibles à la machinerie d'épissage. Les deux processus sont intimement coordonnés et capables de s'influencer mutuellement. L'épissage permet d'éliminer les introns (séquences non codantes) des pré-ARNm et de lier les exons (séquences codantes) entre eux pour former un ARNm mature qui sera traduit en protéine après son export dans le cytoplasme (Borao et al., 2021). L'épissage est catalysé par le spliceosome, un large complexe ribonucléoprotéique dont le remodelage est très actif au cours des deux étapes de transestérification successives de la réaction d'épissage (Wilkinson et al., 2020). Les études récentes en Cryo-microscopie électronique (Cryo-EM) des structures du spliceosome chez la levure et chez l'homme ont permis de clarifier comment les composants du spliceosome stabilisent et remodelent le site actif autour de l'ARN pour effectuer la catalyse de l'épissage. Ces données ont également révélé une plus grande complexité du spliceosome chez les mammifères comparées aux levures et le rôle de protéines humaines dans la modulation de l'activation du spliceosome ou de la dynamique du site actif médiée par l'ATPase (Fica, 2020; Wilkinson et al., 2020).

# I. L'épissage constitutif

## A. Le mécanisme d'épissage

### 1. Les séquences *cis* de reconnaissance des introns

Les séquences présentes aux jonctions intron/exon sont partiellement dégénérées chez les eucaryotes à l'exception de deux paires de dinucléotides (GT en 5' de l'intron et AG en 3' de l'intron) hautement conservés (Sibley et al., 2016). Les séquences consensus des sites donneurs et accepteurs d'épissage ont été établies par l'analyse systématique des séquences aux jonctions intron/exon dans les gènes humains codants pour des protéines (Senapathy et al., 1990; Shapiro and Senapathy, 1987) (Figure 1). Le site donneur d'épissage (5'ss) est défini par une courte séquence de 10 nucléotides tandis que le site accepteur (3'ss) comprend en amont une région riche en pyrimidines, appelée tract de poly-pyrimidines (PPT) et le site de branchement situé à une distance de 18 à 40 nt (Wilkinson et al., 2020). Les sites proches de la séquence consensus (dits sites forts) sont reconnus avec une forte affinité par les facteurs d'épissage et responsables de l'épissage constitutif (inclusion totale de l'exon). Des sites faibles ne seront utilisés que dans certaines conditions physiologiques. Ils permettent, en lien avec d'autres protéines de liaison à l'ARN (RBP) activatrices ou inhibitrices de l'épissage, une régulation plus fine de l'inclusion ou de l'exclusion des exons dans les transcrits matures (épissage alternatif) (Srebrow and Kornblihtt, 2006; Will and Lührmann, 2011; Yeo and Burge, 2004).

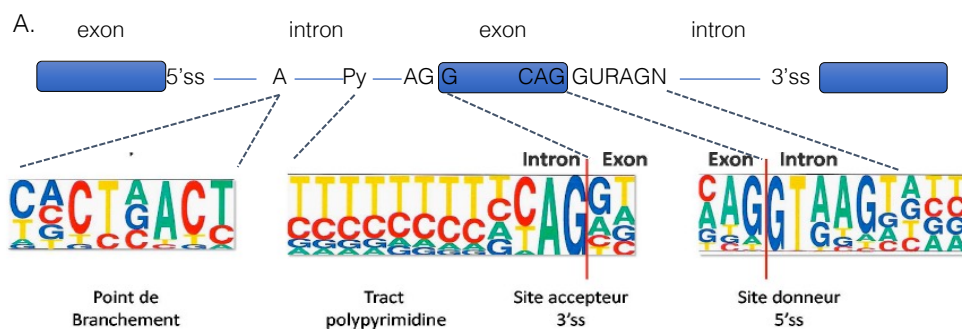
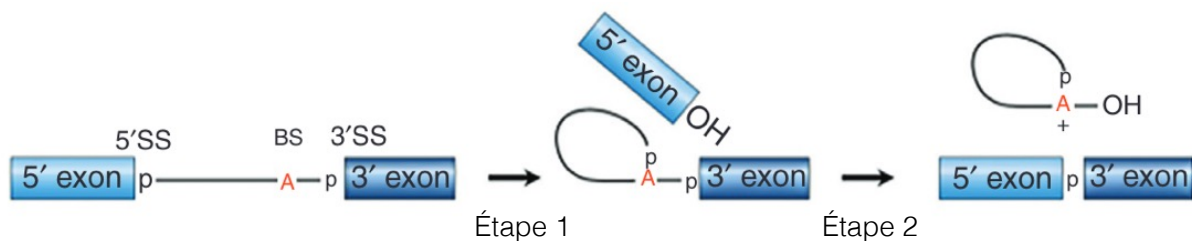


Figure 1. Représentation des séquences des sites d'épissage dans les introns des pré-ARNm de vertébrés. Les séquences sous forme de matrice de poids des sites d'épissage donneurs (5') et accepteurs (3') définissant les limites de l'intron, ainsi que du point de branchement sont présentées pour les introns de type AG/GT reconnus par le spliceosome majeur (dépendant de U2).

## 2. La réaction de l'épissage

La réaction d'épissage consiste en deux étapes successives de transestérification afin d'éliminer les introns des pré-ARNm sous forme de lasso (ou lariat). L'étape 1 de transestérification, communément appelée branchement, consiste en une attaque nucléophile du groupe 2'-OH de l'adénosine du site de branchement (BS-A) au niveau de la liaison phosphodiester du 5'ss. Cette étape aboutit à un exon 5' libre et à un lariat-exon 3' intermédiaire (Figure 2). Lors de l'étape 2, le groupe 3'-OH de l'exon 5' attaque le 3'ss, ce qui entraîne la ligation de l'exon et l'excision de l'intron (Kastner et al., 2019). L'intron est alors linéarisé et dégradé (Jurica et al., 2002).



Adaptée de Kastner et al., 2019

Figure 2. Représentation des deux étapes de transestérification de la réaction d'épissage.

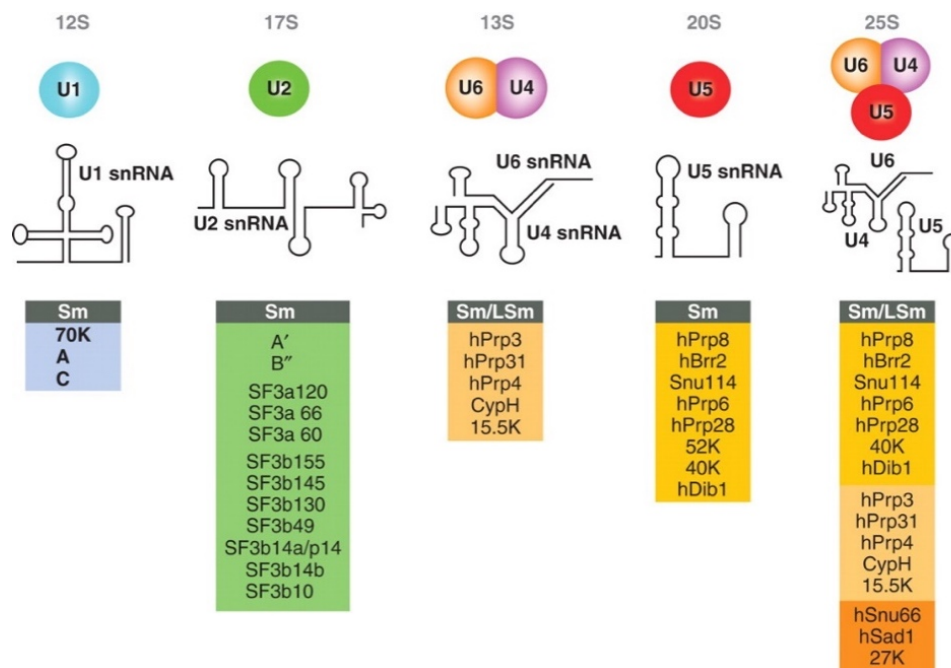
Lors de l'étape 1, l'attaque du groupe 2'-OH de l'Adénosine du site de branchement (BS-A) sur le phosphate (p) du résidu à l'extrémité 5' de l'intron (site 5'SS) conduit à un exon 5' libre et à un exon 3' lié à l'intron structuré en lasso. Lors de l'étape 2, l'attaque du 3' OH de l'exon 5' sur le phosphate à l'extrémité 3' de l'intron (3'ss) permet la libération de l'intron et la ligation des deux exons.

## 3. Le spliceosome

L'épissage d'une grande majorité des pré-ARNm est réalisée par le spliceosome majeur (également appelé spliceosome U2-dépendant). A chaque réaction d'épissage, le spliceosome s'assemble de façon dynamique autour d'un intron, réalise la réaction d'épissage puis doit se désassembler pour pouvoir accomplir sa fonction sur un nouvel intron. Le spliceosome est un macro-complexe composé des 5 snRNP (*small nuclear ribonucleoproteins*) U1, U2, U4-U6 et U5 et d'un grand nombre de protéines qui engagent des interactions ARN-protéines ou protéine-protéine pour accomplir la réaction d'épissage. Le mécanisme d'épissage des pré-ARNm est hautement conservé chez les eucaryotes mais le spliceosome humain a dû évoluer et s'adapter pour prendre en charge l'épissage de pré-ARNm contenant de multiples introns. Il a intégré des facteurs protéiques supplémentaires pour permettre une plasticité accrue de la reconnaissance du site d'épissage et potentiellement pour réguler l'épissage alternatif (Fica, 2020; Kastner et al., 2019).

### a. Les snRNP

Les principaux composants du spliceosome sont les cinq petites ribonucléoprotéines nucléaires (snRNP). Elles sont chacune composées d'un petit ARN nucléaire (snRNA) qui est assemblé avec différentes protéines Sm ou LSm (Figure 3). Les quatre snRNA U1, U2, U4, U5 sont synthétisés par l'ARN polymérase II (Pol II). Ils possèdent une coiffe tri-méthyl-guanosine (m<sup>3</sup>G) à leur extrémité 5' lors de leur maturation dans le noyau. La formation des snRNP se produit grâce à l'interaction de sept protéines Sm homologues dans le cytoplasme. Celles-ci s'assemblent en un anneau autour de la séquence riche en U (connue sous le nom de site Sm), située à l'extrémité 3' du snRNA replié en structure secondaire, pour former le domaine central des snRNP (Jurica and Moore, 2003; Yang et al., 2021). Contrairement aux autres snRNA, le snRNA U6 est synthétisé par l'ARN polymérase III (Pol III) et possède une coiffe mono-méthyl-guanosine (m<sup>1</sup>G). La snRNP U6 est formée grâce à un complexe paralogue de sept protéines LSm (Zhou et al., 2014). Une fois les snRNP formées, elles sont réimportées dans le noyau par des importines afin de remplir leur rôle au sein du spliceosome (Gruss et al., 2017).



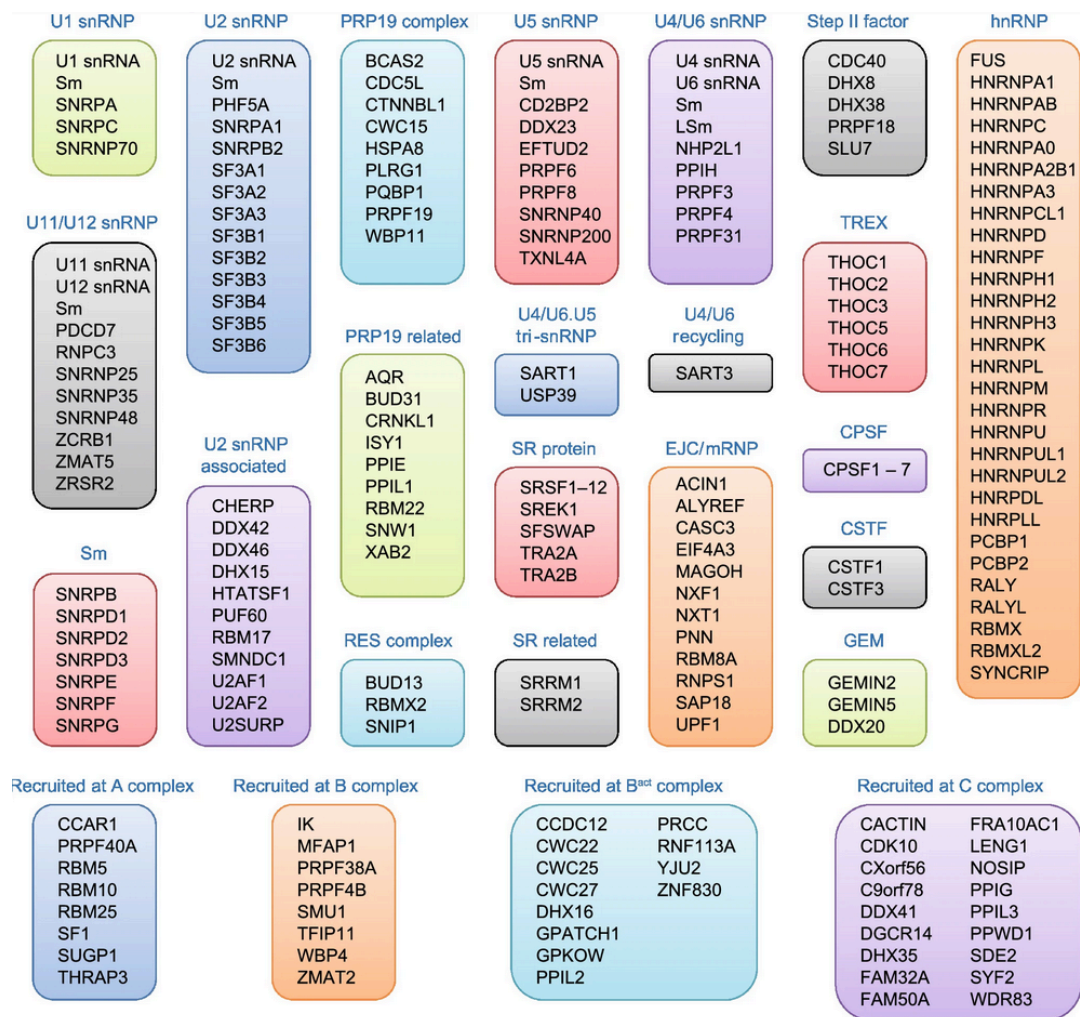
Will and Lührmann, 2011

Figure 3. Composition des snRNP du spliceosome humain.

Chacune des cinq snRNP (U1, U2, U4, U5 et U6) sont composées de snRNA dont la structure secondaire est représentée, associées à des protéines Sm ou LSm (Like Sm) spécifiques listées pour chacune des snRNP.

### b. Complexité de la composition du spliceosome

Le spliceosome est un macro-complexe ribonucléoprotéique activement remodelé au cours de la réaction d'épissage. Il implique 150 à 300 protéines associées ou non aux snRNP, qui contribuent au bon déroulement de l'épissage des pré-ARNm (Figure 4) (Jurica and Moore, 2003; Yang et al., 2021).



Yang et al., 2021

Figure 4. Représentation des composants et des interacteurs du spliceosome humain.

Les facteurs recrutés en fonction des différents complexes formés lors des étapes de l'épissage sont représentés.

Abréviations : RES, rétention et épissage ; SR protein, protéines riches en sérine et arginine ; EJC, complexe de jonction exon-exon ; mRNP, particule de ribonucléoprotéines messagers ; TREX, transcription-exportation ; CPSF, facteur de spécificité de clivage et de polyadénylation ; CSTF, facteur de stimulation du clivage ; GEM, gémini des corps enroulés (ou corps de Cajal) ; hnRNP, ribonucléoprotéines nucléaires hétérogènes.

### c. Dynamique de l'assemblage du spliceosome

L'assemblage du spliceosome s'effectue par différentes étapes catalytiques successives grâce à de multiples interactions dynamiques entre les facteurs d'épissage et l'ARN. Les différents complexes intermédiaires (E, A, B, C, P et ILS) formés indépendamment ou non de l'ATP, assurent l'élimination des introns du pré-ARNm (Figure 5).

#### i. Le complexe E, le complexe d'engagement

La formation du complexe E est la première étape d'assemblage du spliceosome. Cette étape permet la reconnaissance des sites d'épissage 5' et 3'. La snRNP U1 entre en interaction avec le site d'épissage 5' par appariement de base en reconnaissant le GU qui borde l'intron (Wan et al., 2020). En parallèle, la séquence du site de branchement du pré-ARNm est reconnue par

le facteur d'épissage SF1. L'hétérodimère U2AF65/35 se lie à la séquence poly-pyrimidine et sur le site d'épissage 3' en reconnaissant le dinucléotide AG (Wilkinson et al., 2020). La formation du complexe E ne nécessite pas l'hydrolyse de l'ATP.

#### *ii. Le complexe A*

La formation du complexe A est dépendante de l'hydrolyse de l'ATP qui permet l'ouverture des structures secondaires de l'ARN double brin. La transition du complexe E vers le complexe A est médiée par Prp5 et UAP56, deux ATPases/hélicases, qui libèrent SF1 du BS et U2AF65/35 du 3'ss afin de recruter la snRNP U2 au niveau de l'adénosine du site de branchement (Fica, 2020). Cette liaison est stabilisée par les sous-complexes SF3a et SF3b en interaction avec Prp9/11. L'ensemble de ces interactions induit une modification de la structure de l'ARN et le dégagement de l'adénosine du site de branchement qui permet alors le déclenchement de la première attaque nucléophile du site d'épissage en 5' par le groupement hydroxyle 2'OH de l'adénosine du BS (Wan et al., 2020; Wilkinson et al., 2020).

#### *iii. Le pré-complexe B*

La tri-snRNP U4/U6 et U5 préalablement assemblée rejoint le pré-complexe B, premier complexe dans lequel les cinq snRNP sont présentes (Wan et al., 2020). La formation du pré-complexe B réunit les substrats de la première réaction de branchement que sont le 5'ss, le BS et les snRNP U2, U5 et U6 nécessaires à la formation du site actif du spliceosome. Cependant, la snRNP U6 reste dans sa conformation inactive et le 5'ss est toujours lié à la snRNP U1 (Wilkinson et al., 2020).

#### *iv. Les complexes B, B<sup>act</sup> et B<sup>\*</sup>*

Le complexe B est initié par Prp28. La fermeture ATP-dépendante de ses domaines RecA libère le site d'épissage 5' de la snRNP U1. U6 se lie ensuite au 5'ss. L'appariement de la snRNP U5 sur l'exon 5' ainsi que Prp4 qui phosphoryle les composants de la tri-snRNP permettent la stabilisation de ce trimère (Wilkinson et al., 2020). L'ATPase/hélicase Brr2 détache U4 de U6, ce qui permet la formation du complexe B<sup>act</sup> dans lequel la snRNP U6 libérée va s'apparier à U2. Dans une deuxième étape, le complexe associé à Prp19 (NTC), le complexe associé au NTC (NTR) et le complexe de liaison à l'intron (IBC) sont recrutés, induisant l'activation du spliceosome (De et al., 2015; Hogg et al., 2010). Ces complexes stabilisent le pré-ARNm sur le site actif qui se situe au niveau de la protéine spliceosomale centrale Prp8. Pendant cette transition, l'exon 5' et le 5'ss sont guidés par le duplex U5/SRm300 afin d'être stabilisés sur le site actif (Fica, 2020; Kastner et al., 2019; Wan et al., 2020; Wilkinson et al., 2020). Le complexe B<sup>act</sup> est le premier intermédiaire possédant un noyau catalytique entièrement formé.

L'ATPase/hélicase Prp2 permet de remodeler le complexe B<sup>act</sup> en B\* (Wilkinson et al., 2020). Des facteurs spécifiques du site branchement sont ensuite associés avec le duplex 5'ss/U6 dans le site actif. C'est à cette étape que se produit la première réaction de transestérification, entraînant la libération de l'exon 5' et la pseudo-formation d'un lariat d'intron.

#### v. *Les complexes C et C activé (C\*)*

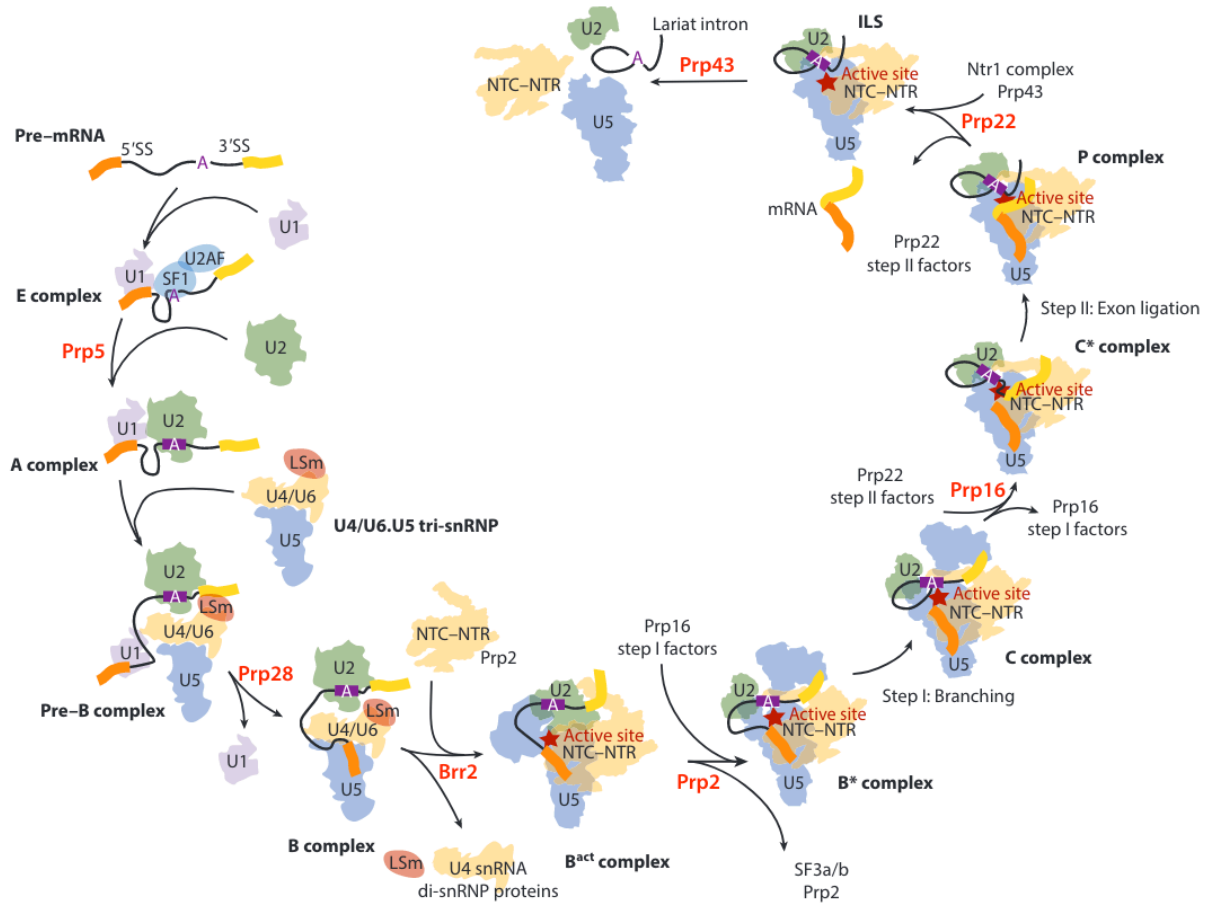
Les complexes B\* et C sont pratiquement équivalents. Ils diffèrent seulement par le fait que le groupe phosphate du premier nucléotide de l'intron (G) est encore lié à l'exon 5' dans le complexe B\*, alors que cette guanine est liée à l'adénosine du site de branchement dans le complexe C lors du branchement (Wilkinson et al., 2020). Le remodelage du complexe C par l'ATPase/hélicase Prp16, engendre la formation du complexe C\*. Prp16 se lie à l'intron en aval du site de branchement et entraîne le déplacement des facteurs spécifiques de cette séquence (CCDC49/94). Pour stabiliser le branchement, Prp17, un facteur de ligation d'exon lié à U6 entre en interaction avec U2. Ceci permet de libérer un espace dans le site actif pour l'arrimage du site 3'ss. FAM32A se lie à l'exon 5' et permet au complexe C\* d'effectuer la jonction de l'exon 5' et du 3'ss qui engendre la ligation des exons (Fica et al., 2019; Wilkinson et al., 2020).

#### vi. *Le complexe P, ILS*

Une fois la ligation effectuée, le complexe P est formé dans lequel la liaison entre le 3'ss et l'exon 3' est remplacée par la liaison avec l'exon 5'. CCDC49/94 et le NTC sont dissociés du site de branchement et de la snRNP U2 (Zhang et al., 2019). FAM32A stabilise la ligation des exons. L'ATPase Prp22 induit la libération des exons qui ont été ligués en interagissant avec l'exon 3' (Fica et al., 2019; Shi, 2017; Wan et al., 2020). L'hydrolyse de l'ATP de Prp22 lui permet de transloquer le long de l'ARNm dans le sens 3'→5'. Ceci conduit à la perte des facteurs de ligation d'exon FAM32A, et à la libération de l'ARNm du spliceosome. En parallèle, l'intron excisé sous forme de lariat est lié à Cwf19L2, un facteur permettant d'initier la dégradation de l'intron (Nguyen et al., 2016).

#### d. *Dissociation du spliceosome*

L'ATPase/hélicase Prp43 se lie au spliceosome contenant le lariat d'intron (ILS) grâce au complexe NTR afin de dissocier le spliceosome. La dissociation des trois snRNP U2, U5, U6 de l'ARNm entraîne la dissociation du complexe NTC induisant ainsi la libération et le recyclage de ces trois snRNP dans de nouveaux spliceosomes (Fica, 2020; Fica et al., 2019; Wan et al., 2020; Wilkinson et al., 2020).



Wilkinson et al., 2020

Figure 5. Représentation schématique des étapes de remodelage des snRNP au cours de l'épissage.

Le spliceosome s'assemble de manière séquentielle sur l'intron à épisser en différents complexes (E, A, Pré-B, B, B<sup>act</sup>, B\*, C, C\* P) nécessaires à la formation du site actif et à la réalisation des étapes de l'épissage (branchement, ligation des exons) conduisant à la libération de l'ARNm. Les ATPases (en rouge) sont impliquées dans le remodelage des différents complexes. BP, point de branchement ; ILS, spliceosome intron-lariat ; mRNA, ARN messenger ; NTC, complexe associé à Prp19 ; NTR, complexe lié au NTC ; snRNA, petit ARN nucléaire ; snRNP, petite ribonucléoprotéine nucléaire ; SS, site d'épissage.

## 4. Reconnaissance des sites d'épissage

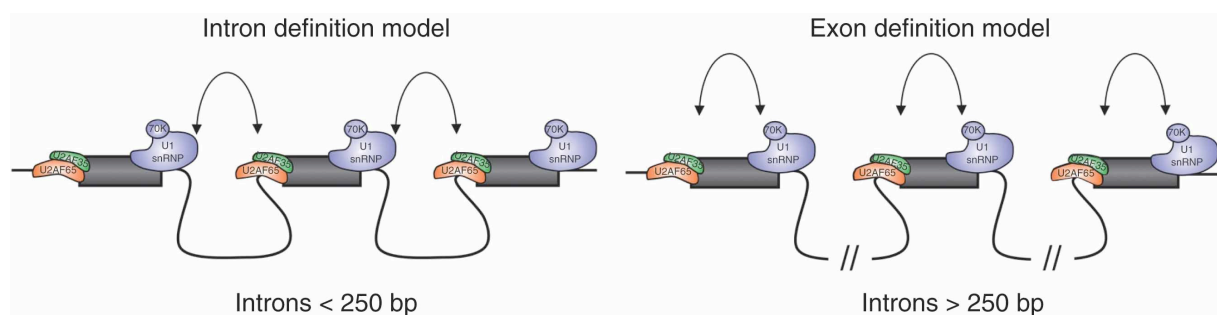
### a. Le modèle de définition d'exon

L'architecture des gènes influence la manière dont les introns et les exons sont reconnus par le spliceosome au cours de l'épissage. Chez les eucaryotes inférieurs où les gènes sont principalement constitués de grands exons entourés de petits introns (< 250 pb), les sites d'épissage sont essentiellement appariés au travers de l'intron (modèle de définition de l'intron). A l'opposé, lorsque les exons sont petits (en moyenne 170 pb) et entourés par des introns considérablement plus longs (> 250 pb à plusieurs kb) comme dans les gènes humains et chez la plupart des vertébrés, la communication entre les sites d'épissage a plus de chance de s'établir au travers de l'exon que de l'intron (Ke and Chasin, 2011; Zlotorynski, 2019). Ce processus est appelé « **définition d'exon** ». Dans ce modèle, la fixation de la snRNP U1 sur le 5'ss en aval favorise la reconnaissance du site 3'ss en amont par U2AF35/65 et la fixation de la snRNP U2



sur le site de branchement (Figure 6) (De Conti et al., 2013; Ke and Chasin, 2011). La tri-snRNP U4/U6 et U5 est aussi présente dans ces complexes de définition d'exon (CDE). La transition du CDE vers le complexe B peut donc se faire sans la formation au préalable d'un complexe A (Schneider et al., 2010). Sur le plan mécanistique, il n'y a pas de différence dans l'assemblage du spliceosome par définition d'exon ou intron. Chaque type de recrutement peut se produire dans différentes parties du même pré-ARNm.

L'épissage basé sur la reconnaissance des exons pourrait servir de contrôle qualité de l'épissage en guidant le dépôt du spliceosome aux bordures d'exons véritables, empêchant ainsi l'utilisation de sites d'épissage isolés (Ke and Chasin, 2011).



De Conti et al., 2013

Figure 6. Représentation des deux modèles de définition de l'épissage.

A gauche, le modèle de définition de l'intron s'applique à l'épissage des introns courts (<250pb) chez les eucaryotes inférieurs. A droite, le modèle de définition de l'exon utilisé pour l'épissage des longs introns (>250pb) chez les vertébrés.

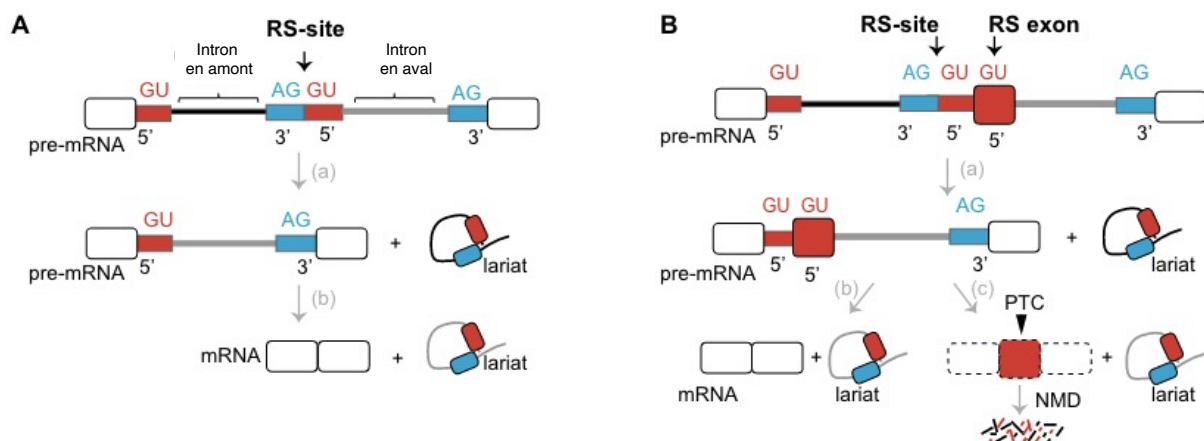
De petites séquences cis-régulatrices dites auxiliaires, situées en dehors des sites consensus d'épissage contribuent au contrôle de l'épissage en favorisant ou en inhibant selon le contexte cellulaire, la reconnaissance des exons via leur interaction avec des protéines spécifiques de liaison à l'ARN. En particulier, les protéines SR recrutées sur les séquences exoniques entre les deux sites, vont aider à la communication entre les extrémités de l'exon et à la connexion entre les snRNP U1 et U2 en stabilisant le réseau d'interaction protéine-protéine à travers l'exon (Ke and Chasin, 2011). La coexistence des snRNP U1 et U2 ainsi que de nombreuses autres protéines dans le CDE a pu être démontrée dans des expériences d'épissage *in vitro* (Schneider et al., 2010). Le rôle de ces séquences auxiliaires d'épissage et des facteurs associés est déterminant dans la régulation de l'épissage alternatif (voir paragraphe II.2).

Dans ce modèle, la taille des exons peut influencer la reconnaissance des sites d'épissage du fait notamment de l'encombrement stérique. Ainsi, un exon constitutif dont la taille a été réduite à moins de 50 pb est exclu *in vivo* dans les cellules HeLa (Dominski and Kole, 1991). Inversement, un exon exclu de moins de 20 pb dont la taille est augmentée (> 100 pb) devient constitutif chez la souris (Black, 1991). L'activation de sites d'épissage cryptiques à l'intérieur

de longs exons (>300 pb) est également en faveur d'un épissage basé sur le modèle de définition de l'exon chez les vertébrés (De Conti et al., 2013; Sakharkar et al., 2005).

### b. Cas particulier : Épissage des longs introns par épissage récursif

Le modèle canonique de l'épissage du pré-ARNm selon lequel l'intron est éliminé au cours d'une seule réaction sous la forme d'une entité unique est difficilement applicable à l'épissage des introns de grande taille. Le processus d'"épissage récursif" (RS) initialement décrit chez *D. melanogaster*, permet l'élimination des grands introns (>10 kb) en plusieurs étapes. Ce mécanisme repose sur la présence dans les introns de sites récursifs d'épissage (sites RS) correspondant aux séquences accolées (AG/GU) de sites canoniques 3' et 5' d'épissage (Duff et al., 2015; Kumari et al., 2021). Le site RS fonctionne initialement comme un site 3'ss, s'appariant avec le site 5'ss qui borde l'exon en amont pour éliminer le premier segment intronique. Le site 5'ss reconstitué interagit avec le site d'épissage 3'ss de l'exon en aval pour l'élimination de la région intronique restante (Figure 7A). Bien que prévalents dans les très longs introns (>40 kb), tous les longs introns ne sont pas éliminés par RS chez la drosophile. Les déterminants de l'activation de ce processus restent encore mal compris.



Adaptée de Kumari et al., 2021

Figure 7. Représentation de l'épissage récursif dans les longs introns.

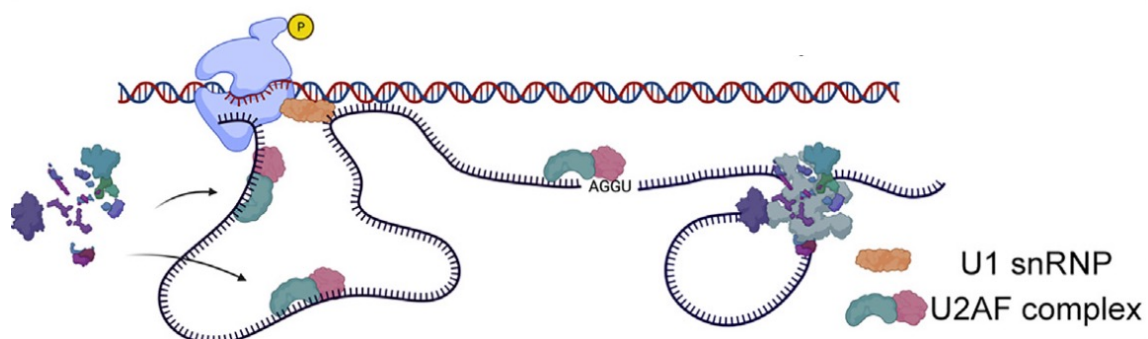
(A) Chez la drosophile, chaque site RS contient un motif de séquence AG/GU comprenant une paire de sites d'épissage 3' et 5' juxtaposés. Ces sites d'épissage divisent un long intron en deux ou plusieurs segments qui seront épissés séquentiellement : (a) intron en amont puis (b) intron en aval. (B) Chez les vertébrés, le processus nécessite un exon RS en aval du site RS. Les sites d'épissage 3' des sites RS doivent être suffisamment forts pour être reconnus par le spliceosome à la première étape de l'épissage récursif (a) avant l'excision finale de l'intron (b). L'exon RS peut être inclus dans certaines isoformes (c) conduisant à une diminution de la stabilité de l'ARNm via le système NMD lorsqu'il contient un codon stop prématuré (PTC).

Le mécanisme de RS est conservé au cours de l'évolution et présent chez les vertébrés. Chez l'homme, la moitié des gènes qui codent pour des protéines ont au moins un intron de plus de 10 kb, pouvant parfois atteindre une taille exceptionnelle comme dans le gène *DMD* (>200 kb). Néanmoins, le séquençage d'ARN total provenant de vingt tissus humains n'a permis de mettre en évidence l'implication du RS que dans un nombre restreint d'introns de très grande taille,

faisant partie principalement de gènes impliqués dans le développement du cerveau (Duff et al., 2015; Sibley et al., 2015). Le mécanisme de RS chez les vertébrés est différent. Il requiert la présence d'un exon supplémentaire appelé exon RS en aval du site RS intronique (AG/GU). La définition de cet exon aide à la reconnaissance du site 3'ss du site RS, qui doit être suffisamment fort pour être reconnu par le spliceosome à la première étape de l'épissage récursif (Joseph et al., 2018; Pai et al., 2018; Sibley et al., 2015; Zhang et al., 2018). Le site 5'ss de l'exon RS va alors entrer en compétition avec le site 5'ss du site RS. Le résultat de cette compétition détermine si l'exon RS sera correctement éliminé, ou inclus dans le transcrit mature. Dans certaines isoformes, l'exon RS contient un codon stop prématuré (PTC) et son inclusion diminue la stabilité de l'ARNm *via* le mécanisme du NMD (Figure 7B). Leur inclusion, généralement à des niveaux faibles, pourrait contribuer à réguler le niveau d'expression des gènes (Sibley et al., 2016). Les exons RS pourraient avoir co-évolué avec les sites RS de manière à maintenir ce modèle de définition d'exon requis chez les vertébrés. La moitié des événements d'épissage récursifs sont post-transcriptionnels et les deux types d'épissage (RS et canonique) peuvent intervenir dans l'élimination des introns contenant des sites RS (Zhang et al., 2018). L'épissage récursif compléterait l'épissage canonique pour éliminer les longs introns.

Bien qu'identifié initialement dans un nombre restreint d'introns, une étude récente indique que l'épissage récursif serait utilisé pour l'élimination de la majorité des introns des gènes humains, selon un modèle de sélection stochastique (Wan et al., 2021). Les sites RS seraient utilisés fréquemment mais de manière aléatoire, non séquentielle, par les spliceosomes à l'échelle du transcriptome (Figure 8). Des transcrits incomplètement épissés sont donc présents dans la cellule, contenant des introns complets ou seulement des parties d'introns correspondant aux étapes intermédiaires du RS.

### Modèle de sélection stochastique des sites d'épissage



Adaptée de Wan et al., 2021

Figure 8. Modèle de sélection stochastique des sites d'épissage récursifs.

L'empreinte du complexe U2AF sur le transcriptome au niveau des sites d'épissage récursifs des introns permet d'établir que la réaction d'épissage se produit de manière aléatoire sur les différents sites RS et non de manière séquentielle.

## II. L'épissage alternatif

L'épissage alternatif (AS) est un processus essentiel qui permet de produire des ARNm messagers qui possèdent des structures et des fonctions différentes à partir d'un même promoteur de gène (Liu et al., 2022). Il est établi que 95 à 100 % des gènes humains subissent un épissage alternatif et que 100% des gènes produisent au moins deux isoformes (Bao et al., 2019; Lee and Rio, 2015; Pan et al., 2008; Reixachs-Solé and Eyrales, 2021). Cette diversité d'expression des transcriptomes contribue à la complexité du protéome (Titus et al., 2021). Les données du séquençage à haut débit de l'ARN dans le cadre du projet génome humain ont permis d'établir que les 20352 gènes produisent 266331 transcrits codant pour des protéines (Pertea et al., 2018). Tous les événements d'épissage alternatifs n'aboutissent pas cependant à la production de protéines fonctionnelles mais peuvent contribuer par d'autres mécanismes à la régulation de l'expression des gènes *via* la production d'ARN non codants ou de transcrits éligibles au mécanisme du NMD (Baralle and Giudice, 2017). L'épissage alternatif joue un rôle primordial dans divers processus de transition chez l'homme. Il contribue à la mise en place coordonnée de programmes d'expression génique requis pour le développement des tissus et des organes notamment le cerveau, dans lequel la régulation spatio-temporelle de l'épissage alternatif a été largement décrite, le cœur, le muscle squelettique ou le foie. Il joue un rôle également dans l'acquisition et le maintien de l'identité tissulaire et cellulaire ou la différenciation cellulaire (Baralle and Giudice, 2017). La dérégulation de ce mécanisme fondamental est impliquée dans de nombreuses pathologies.

### A. Les mécanismes de l'épissage alternatif

#### 1. Les différentes classes d'AS

On recense sept types principaux d'épissage alternatifs : (i) le saut d'exon (exon cassette) est le plus fréquent (30%), (ii) les exons mutuellement exclusifs qui génèrent des isoformes alternatives ne conservant qu'un seul exon d'un groupe d'exons voisins dans le transcrit mature, (iii) la rétention d'intron, (iv-v) l'utilisation de sites d'épissage alternatifs donneurs ou accepteurs (respectivement 8% et 18% des transcrits alternatifs chez les eucaryotes supérieurs) présents souvent à proximité des sites d'épissage constitutifs, (vi) l'utilisation de promoteurs alternatifs et enfin (vii) l'utilisation de sites PolyA alternatifs lors de la maturation des pré-ARNm (Figure 9) (Borao et al., 2021; Liu et al., 2022).

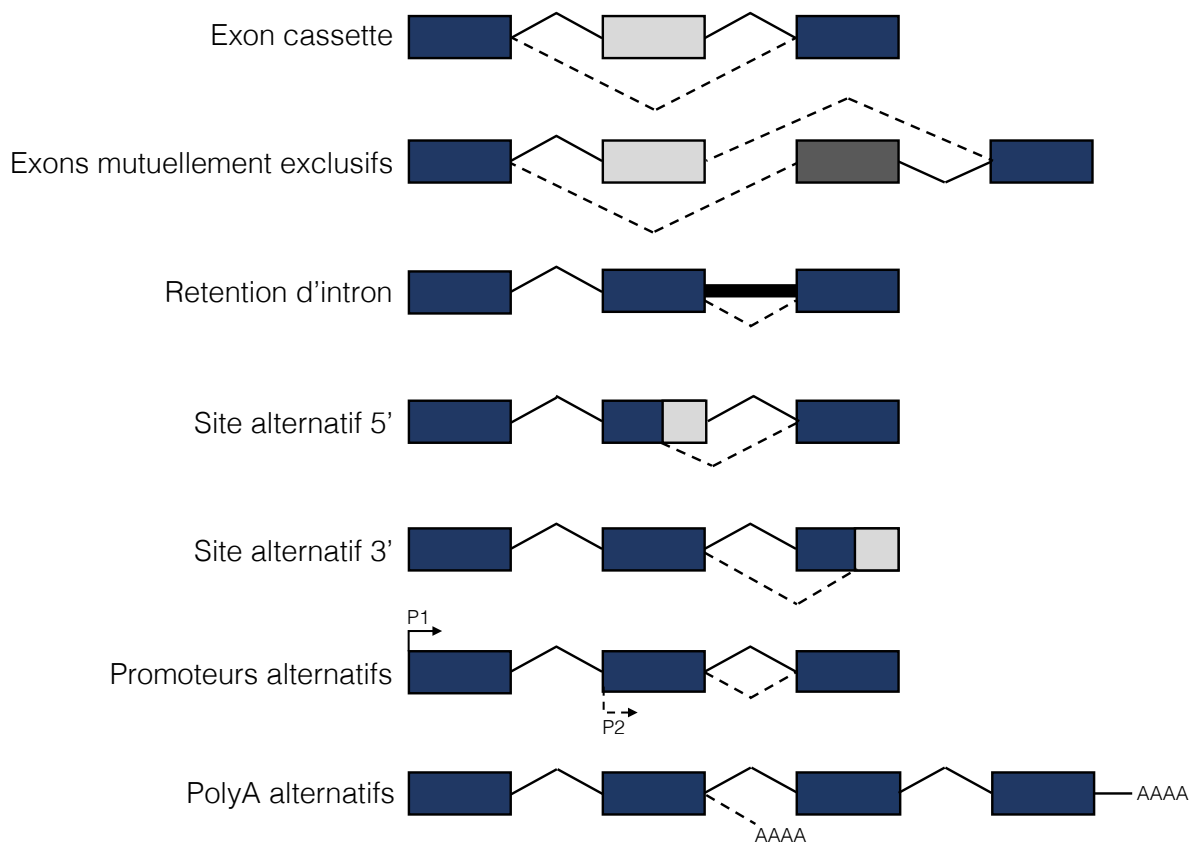


Figure 9. Les différentes classes d'épissage alternatif.

Les boîtes bleues représentent des exons ou des portions d'exons constitutifs. Les boîtes grises représentent les exons alternatifs. Les différents épissages constitutifs (lignes pleines) et les différents épissages alternatifs (lignes en pointillés) sont indiqués. La ligne noire épaisse représente une rétention d'intron dans le transcrit mature final.

## 2. Les déterminants de l'épissage alternatif

Différents facteurs influencent l'épissage alternatif. En premier lieu, les séquences *cis*-régulatrices et les facteurs agissant en *trans* favorisent ou empêchent la reconnaissance du site d'épissage et l'assemblage du spliceosome pour l'inclusion de l'exon dans l'ARNm. D'autres processus fonctionnellement liés à l'épissage comme la transcription ou la régulation épigénétique peuvent également intervenir dans ce mécanisme (Baralle and Giudice, 2017; Dvine, 2018). L'environnement transcriptionnel d'un gène et son contexte cellulaire déterminent son profil d'épissage (abondance relative des différentes isoformes de transcrits).

### a. Les séquences agissant en *cis*

Parmi les séquences agissant en *cis*, on peut distinguer les sites d'épissage canoniques nécessaires à l'assemblage et à la catalyse du spliceosome, les séquences régulatrices auxiliaires reconnues par des activateurs ou des répresseurs d'épissage, et enfin les séquences qui influencent l'épissage par la formation de structures secondaires de l'ARN.

### *i) Les sites consensus et les séquences cis-régulatrices*

L'assemblage du spliceosome dépend d'un appariement entre les régions situées aux jonctions intron/exon et les petits ARN non codants (snRNA) contenus dans les particules snRNP, un appariement qui dépend du niveau de complémentarité. Le niveau de dégénérescence des sites d'épissage est corrélé à la survenue d'évènements d'épissage alternatif. Les exons ayant des sites d'épissage forts (forte homologie de séquence avec les snRNA) sont majoritairement constitutifs.

Bien que la dégénérescence des sites d'épissage et l'émergence de l'épissage alternatif découlent d'un relâchement de la pression de sélection au niveau de ces régions régulatrices, il est noté que les séquences des sites faibles, bien qu'elles diffèrent des sites forts, sont fréquemment conservées entre la souris et l'homme suggérant leur importance fonctionnelle pour la régulation de l'épissage (Dvinge, 2018).

Cette dégénérescence des sites d'épissage s'est accompagnée de l'émergence de séquences *cis*-régulatrices capables de recruter des RBP jouant un rôle déterminant dans la reconnaissance des exons alternatifs. Ces séquences sont constituées de courts motifs situés dans les exons ou les introns. Celles-ci vont moduler l'épissage en favorisant ou inhibant l'assemblage du spliceosome l'épissage (Liu et al., 2022; Reixachs-Solé and Eyras, 2021). Elles sont décrites selon leur localisation et l'impact positif ou négatif sur l'épissage observé dans un contexte donné. On distingue ainsi les activateurs d'épissage exoniques (ESE) ou introniques (ISE) et les répresseurs d'épissage exoniques (ESS) ou introniques (ISS). Ces séquences sont des plateformes de recrutement de facteurs d'épissage généraux que sont les protéines riches en sérine/arginine (SR) et les ribonucléoprotéines nucléaires hétérogènes (hnRNP). Les séquences ESE et ISE recrutent principalement les protéines SR tandis que les hnRNP sont le plus souvent présentes sur les séquences inhibitrices ESS et ISS (Figure 10) (Lee and Rio, 2015). Néanmoins, cette classification entre régulateurs positifs et négatifs n'est pas figée. Les approches de HTS-CLIP/CLIP-Seq montrent que chacune de ces classes de protéines peut jouer les deux rôles selon leur dynamique de recrutement (position exonique ou intronique sur le pré-ARNm et en amont ou en aval des exons) (Lee and Rio, 2015; Liu et al., 2022).

L'ensemble de ces séquences forme une combinaison spécifique d'éléments régulateurs agissant de façon synergique ou antagoniste sur l'inclusion d'un exon. Des études de génomique comparative indiquent qu'il y a eu une évolution compensatrice entre les sites d'épissage et les motifs ESE ou ESS d'un même exon. L'exon dans son ensemble est conservé comme unité d'épissage (Ke and Chasin, 2011).

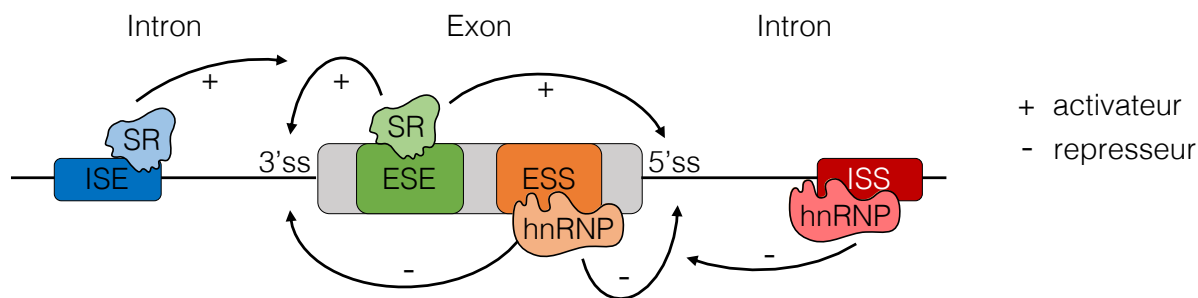


Figure 10. Régulation de l'épissage des pré-ARNm par les séquences cis-régulatrices.

Les protéines SR sont activatrices de l'épissage en se fixant sur les séquences activatrices exoniques (ESE, vert) ou introniques (ISE, bleu). Les hnRNP ont un rôle d'inhibition en se fixant sur les séquences inhibitrices exoniques (ESS, orange) ou introniques (ISS, rouge). La balance de régulation du niveau d'inclusion de l'exon se fait en fonction des besoins cellulaires.

## ii) La structure secondaire des ARN

Les pré-ARNm simples brins synthétisés peuvent adopter différentes structures secondaires en fonction de leurs séquences qui modulent l'accessibilité des sites d'épissage ou des motifs de liaison aux RBP (Bartys et al., 2019). Ces structures peuvent être des structures en tige boucle repliant l'ARN en double brin, des G-quadruplexes (4 brins) ou des interactions ARN-ARN longue distance. Le blocage d'un site d'épissage ou d'une séquence activatrice dans une structure fermée en épingle à cheveux empêche sa reconnaissance par le spliceosome ou une RBP et conduit à une exclusion de l'exon. A l'opposé, ces structures peuvent agir positivement sur l'épissage en rapprochant des éléments régulateurs les uns des autres ou en exposant dans des boucles des motifs activateurs ou un exon entier pour les rendre plus accessibles. La formation de duplex d'ARN peut s'établir à distance entre de petits motifs structuraux de séquences conservées (appelées box) dans les introns permettant d'éliminer une large séquence du transcrit mature comme observé dans le cas des exons mutuellement exclusifs. Les structures en G-quadruplex peuvent être requises pour le recrutement de certaines protéines comme les hnRNP F (Huang et al., 2017).

### b. Les facteurs *trans*

Les expériences de iCLIP ont permis de caractériser un grand nombre de RBP appartenant à différentes classes de protéines (Fu and Ares, 2014). Les deux familles principales sont les protéines SR et les hnRNP qui jouent un rôle majeur dans la régulation de l'épissage alternatif. Elles seront détaillées ci-après. D'autres RBP sont exprimées plus spécifiquement dans certains tissus ou types cellulaires comme les protéines NOVA dans le cerveau ou MBNL1 dans le muscle. Des hélicases en dehors de celles présentes dans le spliceosome ont été décrites comme jouant un rôle dans l'épissage alternatif, notamment DDX5 et DDX17 (ou p68 et p72) (Bourgeois et al., 2016). Ces hélicases à ARN ont la capacité de pouvoir dérouler l'ARN et de rendre accessible des séquences régulatrices et peuvent agir de manière coordonnée avec

d'autres RBP, comme démontré dans le cas de l'hélicase DDX5 avec les protéines hnRNP A1 ou H/F (Dardenne et al., 2014; Lee et al., 2018). Ces protéines régulent plus d'un millier d'épissages alternatifs en commun dans les cellules humaines et partagent des sites de liaison proches de séquences montrant des structures de l'ARN (Lee et al., 2018). La régulation de l'épissage alternatif d'un exon ou d'un groupe d'exons dépend le plus souvent d'une combinaison de RBP qui agissent de manière coopérative ou compétitive. Le niveau respectif d'expression nucléaire de ces protéines doit être finement contrôlé pour permettre la mise en place des programmes d'épissage spécifiques requis au cours du développement ou de la différenciation cellulaire.

### *i) Les protéines SR*

La famille des protéines SR est caractérisée par la présence d'un ou deux motifs de liaison à l'ARN (RRM) à l'extrémité N-terminale des protéines et d'un domaine répété de dipeptides arginine et sérine (RS) à l'extrémité C-terminale, engagé dans des interactions protéine-protéine (Wagner and Frye, 2021). Il existe aussi des protéines possédant des domaines RS apparentés (*SR-like*) (Figure 11) (Chen and Manley, 2009; Manley and Krainer, 2010).

Gènes	Autres noms	Domaines	Sites de liaison
<b>Protéines SR canoniques</b>			
SRSF1	ASF, SF2, SRp30a	RRM, RRMH et RS	RGAAGAAC
SRSF2	SC35, PR264, SRp30b	RRM et RS	UGCUGUU
SRSF3	SRp20	RRM et RS	GCUCUCUUC
SRSF4	SRp75	RRM, RRMH et RS	GAAGGA
SRSF5	SRp40, HRS	RRM, RRMH et RS	AGGAGAAGGGA
SRSF6	SRp55, B52	RRM, RRMH et RS	GGCAGCACCUUG
SRSF7	9G8	RRM, doigt de zinc et RS	(GAC) <sub>n</sub>
SRSF8	SRp46	RRM et RS	Non Déterminé
SRSF9	SRp30c	RRM, RRMH et RS	CUGGAUU
SRSF10	SRp38, SRp40, FUSIP1, TASR1	RRM et RS	AAAGACAAA
SRSF11	SRp54, p54	RRM et RS	Non Déterminé
SRSF12	SRp35	RRM et RS	Non Déterminé
<b>Autres protéines SR</b>			
RNPS1		RRM et RS	Non Déterminé
SRp86	SREK1, SRp508	RRM et RS	Non Déterminé
TRA2α	-	RRM et 2 domaines riche en Arg	GAAARGARR
TRA2β	-	RRM et 2 RS	(GAA) <sub>n</sub>
RBM5	-	RRM et RS	Non Déterminé
CAPER (RBM39)	-	RRM et RS	Non Déterminé



*Figure 11. Liste des différentes protéines SR et SR-like chez l'homme. Ce tableau mentionne les domaines de liaison à l'ARN (RNA Recognition Motifs, RRM et RRMH) et d'interaction protéine-protéine (domaine RS, riche en arginine et sérine), ainsi que leurs sites de liaison.*

La liaison des protéines SR au transcrit naissant modifie l'inclusion des exons. Les protéines SR qui se fixent sur des séquences exoniques favorisent, comme nous l'avons vu, la stabilisation des interactions protéine-protéine à travers l'exon et l'assemblage du spliceosome alors qu'une liaison intronique peut inhiber ce recrutement (Chen and Manley, 2009). Des compétitions entre les protéines SR peuvent exister. Deux protéines SR qui se lient à des exons adjacents entrent en compétition pour le recrutement de la snRNP U2 dans le choix du site d'épissage 3'. La protéine SR la plus apte à recruter U2 détermine l'endroit où l'épissage de l'exon a lieu (Han et al., 2011). Ces protéines sont aussi requises pour le traitement, l'exportation, la surveillance par le NMD et la traduction de l'ARNm. De nombreuses études ont démontré le rôle des protéines SR dans différentes maladies (Alzheimer, Huntington, autisme), défauts du développement ou cancers. Des dérégulations de leur niveau d'expression induit de nombreux défauts d'épissage (Fernández-Nogales et al., 2014; Gao et al., 2007; Gonatopoulos-Pournatzis et al., 2018; Jordan et al., 2019; Quesnel-Vallières et al., 2015; Wang et al., 2005; Yu et al., 2004).

#### *ii) Les protéines hnRNP*

Les protéines hnRNP sont caractérisées par des domaines de liaison à l'ARN conservés : le motif de reconnaissance de l'ARN (RRM) (ou domaine KH chez l'homme), le domaine contenant des glycines entrecoupées de résidus arginine (RGG) et le domaine riche en glycine qui permet de renforcer la liaison hnRNP-ARN (Xie et al., 2021). Les hnRNP ont été nommées à l'origine dans l'ordre alphabétique de A1 à U (Figure 12). Elles sont fortement impliquées dans l'épissage alternatif, comme en témoigne leur capacité à réguler globalement l'épissage des exons dans différents types cellulaires ou de tissus grâce à leurs domaines d'interaction avec l'ARN ou avec d'autres facteurs d'épissage comme U2AF (Gueroussov et al., 2017; Tavanez et al., 2012). La régulation de l'épissage alternatif par les hnRNP peut aussi être médiée par des liaisons sur des régions non-codantes (introns et 3'UTR) comme c'est le cas pour TDP-43 (Tollervey et al., 2011). Ces protéines sont aussi impliquées dans les processus métaboliques de l'ARN comme la circulation nucléocytoplasmique de l'ARNm (Low et al., 2021).

Gènes	Autres noms	Domaines	Sites de liaison
hnRNP A1	-	RRM, RGG et G	UAGGGA/U
hnRNP A2	-	RRM, RGG et G	(UUAGGG)n
hnRNP B1	-	Non déterminé	Non déterminé
hnRNP C1	AUF1	RRM	U rich
hnRNP C2	-	Non déterminé	Non déterminé
hnRNP F	-	RRM, RGG et GY	GGGA et G rich
hnRNP G	-	RRM et SRGY	AAGU
hnRNP H	DSEF1	RRM, RGG, GYR et GY	GGGA et G rich
hnRNP H'	-	Non déterminé	Non déterminé
hnRNP I	PTB	RRM	UCUU and CUCUCU
hnRNP L	-	RRM	C et A rich
hnRNP LL	SRRF	RRM	C et A rich
hnRNP M	-	RRM et GY	ND
hnRNP Q	-	RRM et RGG	CC(A/C)

Figure 12. Liste des différentes hnRNP chez l'Homme.

Les hnRNP possèdent des domaines de liaison à l'ARN (RRM ; RGG, riche en arginine ; GY, riche en glycine). Leurs séquences de liaison à l'ARN ainsi que leurs gènes cibles sont aussi mentionnés dans le tableau.

### iii) Les facteurs de transcription

Une étude récente est venue étendre encore la liste des facteurs *trans* capables de réguler l'épissage alternatif en révélant le rôle inattendu des facteurs de transcription. Une technique de séquençage nouvelle génération (SPAR-Seq, *Systematic Parallel Analysis of Endogenous RNA regulation coupled to Barcode Sequencing*) a montré qu'un tiers des épissages alternatifs était contrôlé par des facteurs de transcription ou de liaison à l'ADN dans des cellules embryonnaires et neurales de souris (Han et al., 2017). Ces facteurs agissent soit en se liant directement à l'ARN à proximité de l'exon régulé, soit de manière indirecte en modulant l'expression des facteurs d'épissage régulant le même exon alternatif. Il a récemment été décrit que le facteur de transcription CTCF, un régulateur majeur de l'expression génique, contribue également par des mécanismes très variés à la régulation de l'épissage alternatif. Au niveau du transcriptome, il agit sur l'élongation de la transcription, la méthylation de l'ADN, l'architecture de la chromatine, les modifications des histones, et la régulation de l'expression et de l'assemblage des facteurs d'épissage (Alharbi et al., 2021). Notre groupe a par ailleurs contribué à démontrer le rôle dans l'épissage de la protéine FUBP1 (*Far Upstream Element binding*

*protein 1*) (Miro et al., 2015). Cette protéine est principalement connue comme facteur de transcription du gène MYC et est impliquée dans de nombreux cancers hématopoïétiques, avant que son rôle dans la régulation de l'épissage ne soit décrit (Debaize and Troadec, 2019).

## B. Autres processus de régulation de l'épissage

L'épissage des gènes humains est intimement lié à l'étape de transcription. En conséquence, tous les aspects de la régulation de la transcription peuvent influencer l'épissage tels que l'architecture des promoteurs ou encore la vitesse d'élongation de l'ARN polymérase II, elle-même contrôlée par la structure et les modifications de la chromatine et l'occupation des nucléosomes (Baralle and Giudice, 2017; Giono and Kornblihtt, 2020). Certains événements d'épissage semblent nécessiter une plus grande assistance de la structure de la chromatine que d'autres, comme le montre le positionnement accru des nucléosomes dans les exons avec des sites d'épissage faibles qui peut favoriser un ralentissement ou une pause de la Pol II (Tilgner et al., 2009). L'épissage est aussi lié aux processus post-transcriptionnels, notamment dans la surveillance des ARNm matures. Ce mécanisme physiologique qui fait intervenir le système NMD couplé à l'épissage sert à la régulation de l'expression des transcrits, notamment dans la régulation des facteurs d'épissage. Certaines RBP possèdent des codons STOP prématuré (PTC) dans des exons alternatifs dont elles contrôlent l'inclusion afin de réguler leur propre expression. Ce processus appelé « AS-NMD » influence le choix des sites d'épissage du transcriptome en fonction des besoins cellulaires (Titus et al., 2021) comme décrit plus loin.

### 1. Le couplage transcription/épissage

Classiquement, il est considéré que deux modèles non mutuellement exclusifs peuvent expliquer le mécanisme par lequel le processus de transcription influence l'épissage : le modèle de recrutement et le modèle cinétique (Kornblihtt et al., 2013).

#### a. La phosphorylation du CTD de la Pol-II : le modèle de recrutement

Dans le modèle de recrutement, la phosphorylation du CTD de la Pol II permet l'interaction de la coiffe avec les facteurs d'épissage, notamment les protéines SR (Giono and Kornblihtt, 2020). Ces derniers vont alors être présents en forte concentration à proximité des sites d'épissage et vont impacter les décisions d'inclure ou non les exons dans le transcrit mature. Le CTD possède une répétition de 52 résidus YSPTSPS phosphorylés qui permettent à la Pol II de quitter le promoteur et commencer la synthèse du pré-ARNm (Figure 13) (Schier and Taatjes, 2020). L'épissage des exons de la fibronectine *via* le recrutement de SRSF3 sur le CTD phosphorylé de la Pol II est un exemple fonctionnel de ce modèle (de la Mata and Kornblihtt,

2006). Il a été montré que le changement du niveau de phosphorylation du CTD induit par UV (hyper-phosphorylation) conduit à des défauts d'épissage alternatif de la troponine T mais aussi dans de nombreux gènes (Muñoz et al., 2017; Nieto Moreno et al., 2020).

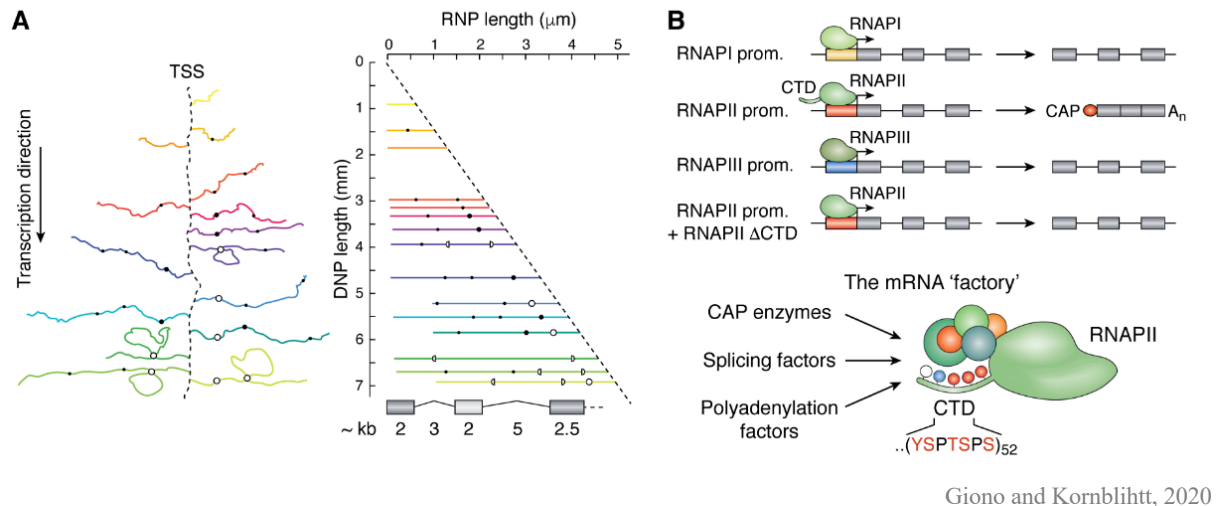


Figure 13. Le couplage des processus de transcription et d'épissage. (A) Schéma représentatif en "arbre de Noël" montrant l'élongation de la transcription à partir du site d'initiation (TSS), couplée à l'épissage avec la présence de structures en lasso d'intron (données de Beyer et Osheim sur des embryons de drosophile). (B) Schéma mettant en évidence le rôle de la Pol II et de son domaine C-terminal (CTD) dans la maturation de l'ARNm.

### b. La pause proximale du promoteur de la Pol-II

Après l'initiation de la transcription qui mène à un pré-ARNm naissant de 20 à 60 pb, la Pol II s'arrête. Cette pause proximale du promoteur va limiter l'activité transcriptionnelle afin d'assurer la mise en place de la coiffe 5' sur l'ARN naissant (Mayer et al., 2015; Tome et al., 2018). Ceci empêche aussi la ré-initiation de la transcription par une autre Pol II (Shao and Zeitlinger, 2017). Le promoteur reste dans un état libre sans nucléosome afin de ne pas inhiber totalement la transcription (Schier and Taatjes, 2020). La pause transcriptionnelle va permettre de recruter la snRNP U1 sur un site 5' naissant non épissé. Le couplage transcription/épissage semble être bidirectionnel, ce qui implique que les facteurs d'épissage stimulent aussi l'initiation de la transcription. La snRNP U1 qui se fixe sur un site 5' non épissé stimule réciproquement l'association des facteurs de transcription sur le promoteur afin de réactiver la transcription (Damgaard et al., 2008).

### c. La vitesse d'élongation de la Pol II : le modèle cinétique

L'élongation de la transcription après la pause proximale est activée par le facteur pTEFb (Muniz et al., 2021). Celui-ci interagit notamment avec le facteur d'épissage SRSF2 (Lin et al., 2008). La vitesse d'élongation de la Pol II peut influencer l'épissage en modulant la disponibilité de séquence agissant en cis et le choix des sites d'épissage ainsi que le recrutement des facteurs d'épissage par la machinerie de transcription (Dvinge, 2018). En règle générale, une élongation

de la Pol II plus rapide va favoriser le saut d'exon tandis qu'une vitesse plus lente va aider à la reconnaissance de sites d'épissage faibles et à l'inclusion d'exons alternatifs. Des expériences de ChIP-Seq montrent que la réduction de la vitesse d'élongation de la Pol II avec la Camptothécine (inhibiteur de la topoisomérase I) entraîne une accumulation des différents facteurs d'épissage (snRNP U1, U5, U2AF65, protéines SR, hnRNP et RBP) sur les sites d'épissage. La vitesse de synthèse du pré-ARNm peut également affecter la formation des structures secondaires et donc l'accessibilité aux sites d'épissage et aux autres séquences régulatrices présentes à proximité (Giono and Kornblihtt, 2020).

## 2. Le couplage épissage et marques épigénétiques

La structure de la chromatine est déterminée par les modifications d'histones et la méthylation de l'ADN. Les différentes régions de la chromatine possèdent des profils uniques de méthylation de l'ADN où il existe plus de 50 modifications d'histones qui sont spécifiques des tissus (Naftelberg et al., 2015). Ce code épigénétique est couplé à la transcription mais aussi à l'épissage alternatif (Giono and Kornblihtt, 2020).

### a. Les modifications d'histones

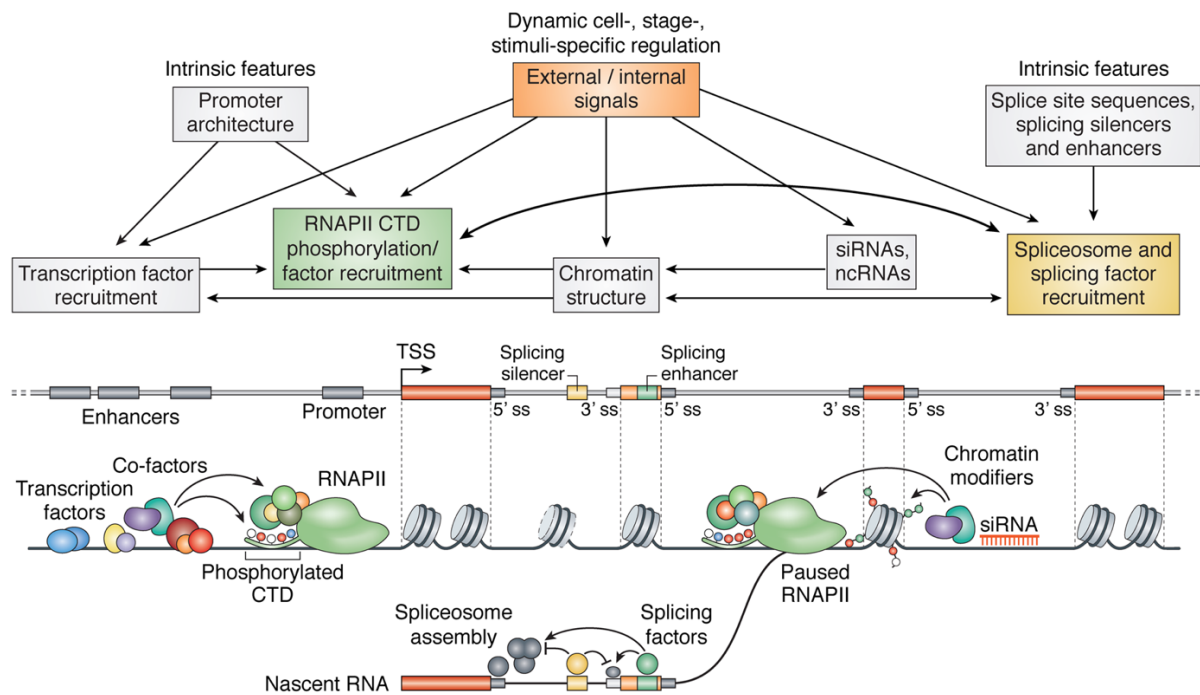
Les modifications d'histones des séquences promotrices au niveau des sites de début de transcription permettent de prédire le niveau d'expression des gènes. De même les modifications de la méthylation, de l'acétylation ou de la phosphorylation des histones qui sont enrichies au niveau des exons, influencent le niveau d'inclusion des exons dans les ARNm (Karlić et al., 2010). Les différentes modifications des histones agissent soit en modulant la vitesse d'élongation de la Pol II par relâchement de la chromatine soit en recrutant des facteurs directement ou *via* des protéines de liaison à la chromatine. Ainsi, H3K36me3 recrute MRG15 ou Psip1, qui recrutent des facteurs d'épissage tels que la snRNP U5, U2AF65, des protéines SR ou PTB. Ceux-ci favorisent la répression de l'épissage des exons dépendants de PTB. Ces trimères forment des complexes co-transcriptionnels d'épissage-chromatine induisant un épissage alternatif d'exons de différents gènes (Guo et al., 2014; Khan et al., 2014; Luco et al., 2011; Naftelberg et al., 2015; Pradeepa et al., 2012). A l'échelle du génome, l'inhibition des histones désacétylases modifie l'épissage de plus de 700 gènes en réduisant la liaison des facteurs d'épissage (Hnilicová et al., 2011). L'hyper-acétylation des histones induit une augmentation de l'élongation de la Pol II corrélant à une diminution d'inclusion des exons alors que la pause transcriptionnelle, corrélée à la méthylation d'histones, permet la ré-inclusion de ces exons (Fuchs et al., 2014). A l'inverse, l'épissage module les marques épigénétiques de la chromatine. La liaison de snRNP U1 au site 5'ss impacte l'organisation des nucléosomes par

des modifications de méthylation d'histones. La surexpression de protéines SR favorise la suppression des nucléosomes tandis que la surexpression de hnRNP augmente ces structures (de Almeida et al., 2011; Keren-Shaul et al., 2013).

### b. La méthylation de l'ADN

La méthylation se produit plus fréquemment sur les séquences exoniques que dans les introns. Les exons non reconnus par la machinerie d'épissage ont des niveaux plus faibles de méthylation d'ADN que les exons inclus de manière constitutive (Gelfman et al., 2013). Les protéines de l'hétérochromatine, comme HP1, recrutent des facteurs d'épissage tels que la snRNP U2, des protéines SR, notamment SRSF3 et des hnRNP dans les cellules de mammifères (Yearim et al., 2015).

L'illustration du lien entre la structure chromatinienne, la transcription et la machinerie du spliceosome associé aux RBP, montre la complexité de la régulation spatio-temporelle de l'épissage. L'interconnexion entre ces divers processus biologiques est déterminante pour établir les niveaux d'expression des différentes isoformes des gènes, en fonction des besoins cellulaires (Figure 14).



Giono and Kornblihtt, 2020

Figure 14. Les différents niveaux de régulation de la transcription et de l'épissage alternatif. La régulation de la transcription et l'épissage alternatif est dépendante de multiples paramètres incluant en premier lieu des caractéristiques intrinsèques de séquence permettant de recruter des facteurs de transcription au niveau du promoteur ou les composants du spliceosome ainsi que les facteurs d'épissage sur les séquences régulatrices (sites 5' et 3', les activateurs et répresseurs d'épissage). Les facteurs de transcription et leurs coactivateurs ainsi que les remodeleurs de la chromatine agissent sur le recrutement de la Pol II, son état de phosphorylation ou le recrutement de cofacteurs régulant ainsi l'initiation et l'élongation de la transcription. L'ensemble des mécanismes associés à l'ARN et la structure de la chromatine participent à la régulation de l'élongation de la transcription et de l'épissage.

### 3. L'épissage alternatif couplé au système NMD

Un des mécanismes décrit récemment est l'épissage alternatif couplé au système de dégradation des ARNm (AS-NMD) afin de réprimer l'activité d'un gène. Cette régulation physiologique repose sur l'introduction par épissage alternatif d'un codon STOP prématuré (PTC) dans l'ARNm afin de générer des transcrits « cibles du NMD » (Titus et al., 2021). Le NMD est le mécanisme de surveillance des ARNm après maturation et exportation dans le cytoplasme afin d'éliminer les transcrits portant des PTC (Lejeune, 2022). Les PTC peuvent être introduits de différentes manières : i) par l'inclusion d'un exon poison (PE) ; par l'exclusion d'un exon induisant un décalage du cadre de lecture ; par la rétention d'un intron poison (PI) ou enfin, par l'utilisation d'un site alternatif d'épissage, comme le site 5'ss (Figure 15) (Titus et al., 2021).

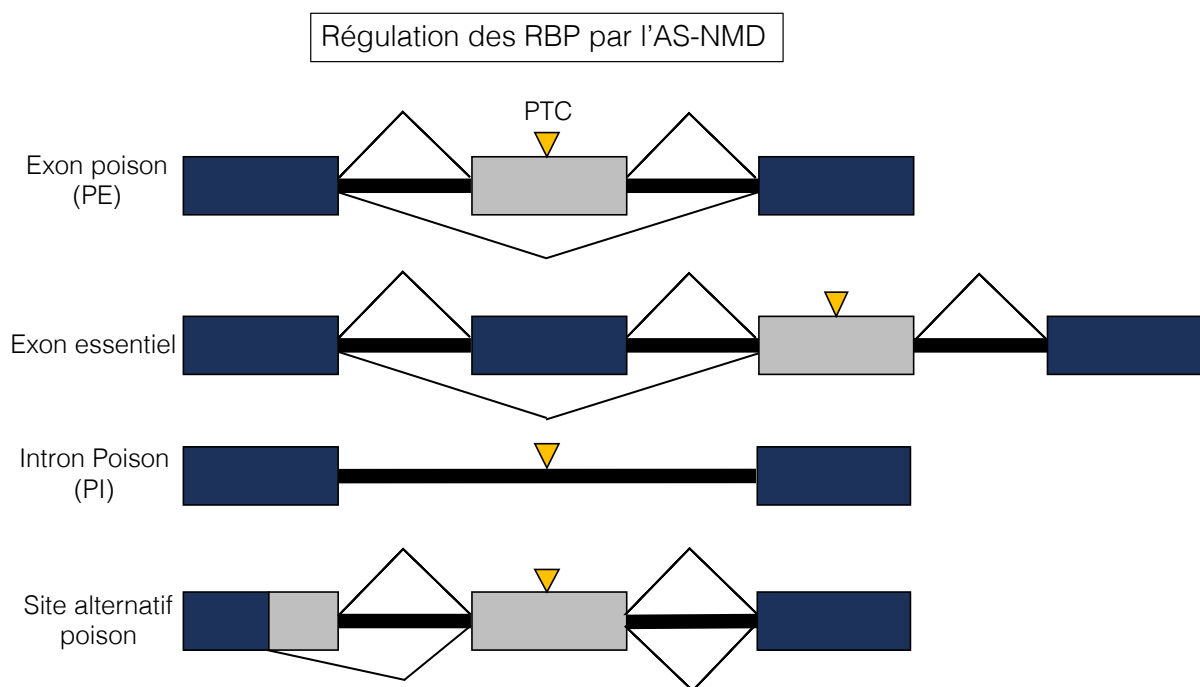


Figure 15. Exemples d'épissage alternatif conduisant à l'introduction d'un PTC dans les ARNm.

Illustration de différents mécanismes d'épissage alternatif conduisant à la dégradation de transcrits cibles par le système du NMD : inclusion d'un exon "poison" contenant un PTC ; épissage alternatif d'un exon essentiel décalant le cadre de lecture ; rétention d'un intron "poison" contenant un PTC ou utilisation d'un site d'épissage alternatif induisant un décalage du cadre de lecture.

Le mécanisme de l'AS-NMD est principalement décrit dans la régulation de l'expression des facteurs d'épissage. Le séquençage de transcriptomes, des analyses de CLIP-Seq et des cribles par ARN interférent (ARNi) montrent qu'il existe plus de 1000 PE prédits comme cibles du NMD, majoritairement dans des gènes codant pour des RBP (Yan et al., 2015). De nombreuses protéines SR et hnRNP se fixent sur leurs propres pré-ARNm durant l'épissage afin d'autoréguler négativement ou positivement leur expression en incluant ou non un PE en fonction des besoins cellulaires (Dreumont et al., 2010; Leclair et al., 2020; Rossbach et al., 2009). C'est le cas par exemple, du facteur SRSF2 qui autorégule sa propre expression par des

événements d'épissage dans son 3'UTR, les protéines hnRNP H et TDP-43 étant des antagonistes de cette régulation (Dreumont et al., 2010). Les protéines SR et hnRNP peuvent aussi réguler l'inclusion de PE dans d'autres RBP comme décrit pour SRSF3 et hnRNP M qui répriment l'inclusion de l'exon poison 6A dans le transcrit de l'hélicase DHX9, favorisant l'expression de la protéine DHX9 impliquée dans une forme de sarcome (Palombo et al., 2020).

## C. L'épissage dans le muscle

### 1. Les facteurs spécifiques du muscle strié

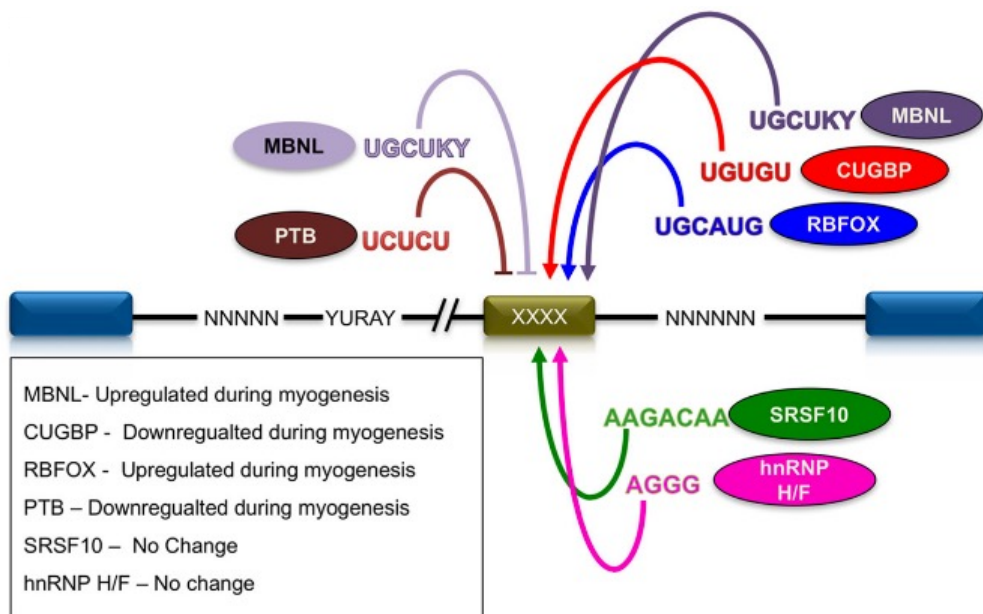
Avec le cerveau, le muscle strié (squelettique et cardiaque) est l'un des tissus à présenter le plus d'évènements d'épissage alternatif (Ellis et al., 2012). L'épissage a été particulièrement étudié durant le développement cardiaque, notamment dans la phase de transition post-natale au cours de laquelle les cellules doivent s'adapter à des conditions d'oxygène et de métabolisme très différentes afin de finaliser leur maturation en cardiomyocytes. Ces transitions reposent en partie sur changements du profil d'épissage de gènes impliqués dans le trafic vésiculaire et le remodelage de la membrane plasmique médiés par des variations du niveau d'expression de nombreuses RBP, notamment celles de la famille MBNL et CELF. Les facteurs MBNL1/2 et CELF1/2 régulent l'inclusion d'exons de gènes impliqués dans la cardiogenèse, notamment la troponine T (Giudice et al., 2014). Un exemple emblématique est celui du gène de la titine (*TTN*) dont l'ARNm comportant 363 exons code pour la plus grande protéine humaine (3.7 MDa). L'inclusion des exons 50-219, qui codent pour la région (I-band) définissant l'élasticité de la protéine, est régulée au cours du développement. La forme longue (N2BA) présente chez les nouveau-nés disparaît au profit de l'isoforme courte cardiaque (N2B) au stade adulte. Le rapport des deux isoformes détermine la longueur des sarcomères et la tension passive des cardiomyocytes, et donc la rigidité de la paroi du myocarde pendant le remplissage ventriculaire. Le facteur RBM20 régule cet épissage alternatif en réprimant l'inclusion des exons 50-219 (Li et al., 2013). L'analyse par RNA-Seq de l'épissage alternatif de l'isoforme N2A exprimée dans le muscle squelettique adulte a montré également la présence d'un nombre élevé d'épissages alternatifs (n=498) (Jordan et al., 2019; Savarese et al., 2018).

### 2. L'épissage alternatif du muscle squelettique

Le muscle squelettique subit aussi d'importants changements physiologiques au cours de son développement afin de produire des muscles adultes entièrement fonctionnels (Bland et al., 2010; Hinkle et al., 2019). Des vagues coordonnées de transitions d'épissage se produisent tout au long de la différenciation myogénique selon un processus hautement régulé et conservé



(mammifères, oiseaux) (Bland et al., 2010). Ces modifications d'épissage apparaissent dès le stade myoblaste, avant même l'apparition des marqueurs précoces de la différenciation. Cela s'accompagne de changements importants du niveau d'expression de différentes RBP qui corrélerent aux différents clusters temporels de transitions d'épissage. Selon une étude réalisée par RNA-Seq sur 18 biopsies musculaires d'individus sains, il y aurait en moyenne 2.12 isoformes par gène. Mais un groupe de 113 gènes, dont le gène *TTN* cité plus haut, expriment 10 ou plus d'isoformes dans le muscle squelettique (Lindholm et al., 2014). Plusieurs RBP comme MBNL1, CELF1, RBFOX2, PTB (ou hnRNP I) et QKI ont été décrites comme des régulateurs de l'épissage alternatif dans le muscle squelettique, notamment dans le modèle murin (Brinegar et al., 2017; Hall et al., 2013; Orengo et al., 2011; Singh et al., 2014). MBNL1 est capable d'interagir avec UA2F65 (Dvinge, 2018; Warf et al., 2009) et CELF1 est un antagoniste de MBNL1 dans la régulation de l'épissage alternatif (Wang et al., 2015). Des analyses bio-informatiques montrent un enrichissement des motifs de liaison pour ces différents régulateurs dans les régions introniques en amont ou en aval des exons alternatifs. Des motifs de fixation sont décrits comme riches en GU pour les protéines RBFOX, CELF1, MBNL1 ou riches en CU pour PTB (Figure 16) (Castle et al., 2008; Llorian, 2011; Nakka et al., 2018). La protéine SRSF10 joue aussi un rôle dans la différenciation musculaire. Elle régule l'épissage alternatif de différents gènes selon son état de phosphorylation, dont le facteur musculaire *Lrrfip1* chez la souris (Wei et al., 2015).



Nakka et al., 2018

Figure 16. Les principaux régulateurs de l'épissage alternatif dans le muscle. Selon la position de leur site de fixation par rapport à l'exon alternatif qu'ils régulent, les facteurs trans d'épissage peuvent avoir un effet inhibiteur (en amont) ou activateur (en aval) sur l'inclusion de l'exon. L'expression de certaines des protéines mentionnées est régulée au cours de la myogenèse.

## Partie 2 : Le complexe de jonction d'exons (EJC)

Le complexe de jonction exon-exon (EJC) est un constituant et un régulateur essentiel des particules ribonucléoprotéiques (mRNP) chez les métazoaires (Schlautmann and Gehring, 2020). Les EJC, recrutés dans le noyau par les spliceosomes, marquent les ARNm épissés non traduits et sont cruciaux pour leur cycle de vie (Kwon et al., 2021). Ils restent liés aux ARNm exportés dans le cytoplasme où ils seront désassemblés lors de la traduction.

La découverte du complexe EJC découle de l'observation de l'importance de la présence d'introns pour l'expression des gènes dans les cellules eucaryotes (Hir et al., 2016; Le Hir et al., 2003). Des études réalisées dans différents modèles avaient en effet montré une baisse significative d'expression (Hamer and Leder, 1979; Khoury et al., 1979) et de l'efficacité du transport des ARNm vers le cytoplasme dans les gènes sans introns, ne subissant pas d'épissage, comparé aux gènes avec introns (Luo and Reed, 1999). La démonstration du lien existant entre l'épissage nucléaire des introns et l'activation du mécanisme du NMD qui dégrade sélectivement les ARNm portant un codon stop prématuré (PTC) a conduit à l'hypothèse que l'épissage du pré-ARNm laisserait une signature moléculaire au niveau des jonctions exon-exon. Lors de la traduction, ce mécanisme permet de différencier un PTC situé en amont de la dernière jonction exon-exon du codon stop naturel (Belgrader et al., 1993; Carter et al., 1996; Cheng et al., 1994; Thermann, 1998; Zhang et al., 1998).

La composition de cette marque déposée 20-24 nt en amont des jonctions exon-exon par le spliceosome a pu être caractérisée au fil des années, révélant initialement la présence des protéines SRm160, DEK, RNPS1, Y14 et du facteur d'export des ARNm REF (Le Hir et al., 2000). Ce n'est que plus tard que la présence d'eIF4A3 a été identifiée dans les complexes spliceosomaux purifiés et que son rôle majeur dans la formation de l'EJC et le recrutement du système NMD a été établi (Chan et al., 2004; Ferraiuolo et al., 2004; Shibuya et al., 2004).

Le complexe EJC est défini par un cœur de protéines fixées de manière stable sur l'ARNm qui inclut eIF4A3, l'hétérodimère Y14/MAGOH et MLN51 (Ballut et al., 2005; Degot et al., 2004; Reichert et al., 2002; Tange et al., 2005). Un ensemble de protéines dites périphériques interagit de manière directe ou indirecte avec l'EJC mais qui ne sont que faiblement liées à l'ARNm et/ou présentes de façon transitoire (Hir et al., 2016; Tange et al., 2005) (Figure 17).

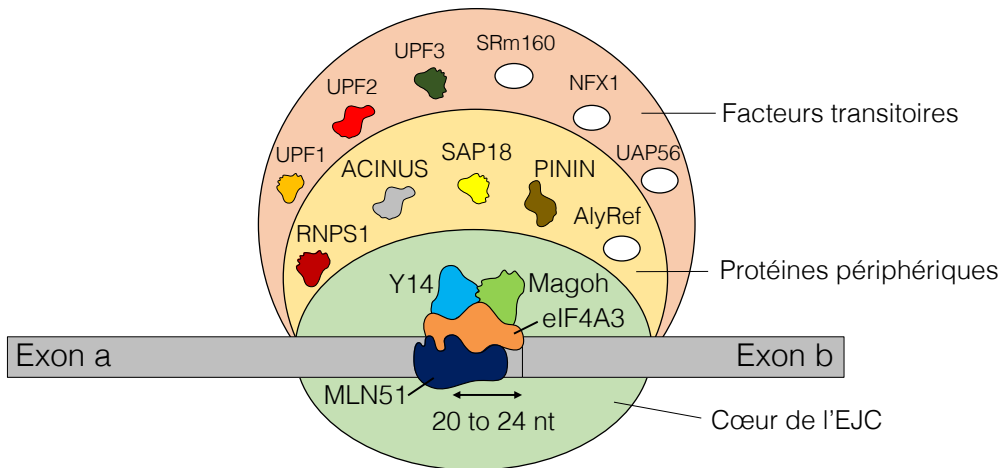


Figure 17. Les trois niveaux d'assemblage des facteurs de l'EJC et des protéines associées. Les quatre protéines cœur de l'EJC : eIF4A3, Y14, MAGOH et MLN51 (vert). EIF4A3 assure la liaison avec l'ARNm. Ces facteurs cœur sont des protéines navette qui accompagnent l'ARNm du noyau vers le cytoplasme. Les protéines périphériques (jaune) se lient à l'EJC de manière stable et forment un complexe trimérique. Les facteurs transitoires (orange) sont des protéines qui interagissent de manière dynamique avec les protéines cœur ou périphériques de l'EJC.

Ainsi, les EJC marquent une période précise du cycle de vie de l'ARNm entre l'épissage nucléaire et la traduction cytosolique. Durant cette période, les EJC contribuent à la régulation de l'épissage et au recrutement des facteurs d'export nucléaire. Dans le cytoplasme, les EJC sont impliqués dans l'efficacité de la traduction des ARNm et leur stabilité en activant le NMD en présence de transcrits porteurs d'un PTC pour éviter la production de protéines tronquées potentiellement toxiques dans la cellule (Figure 18) (Gerbracht and Gehring, 2018; Kwon et al., 2021; Schlautmann and Gehring, 2020; Woodward et al., 2017).

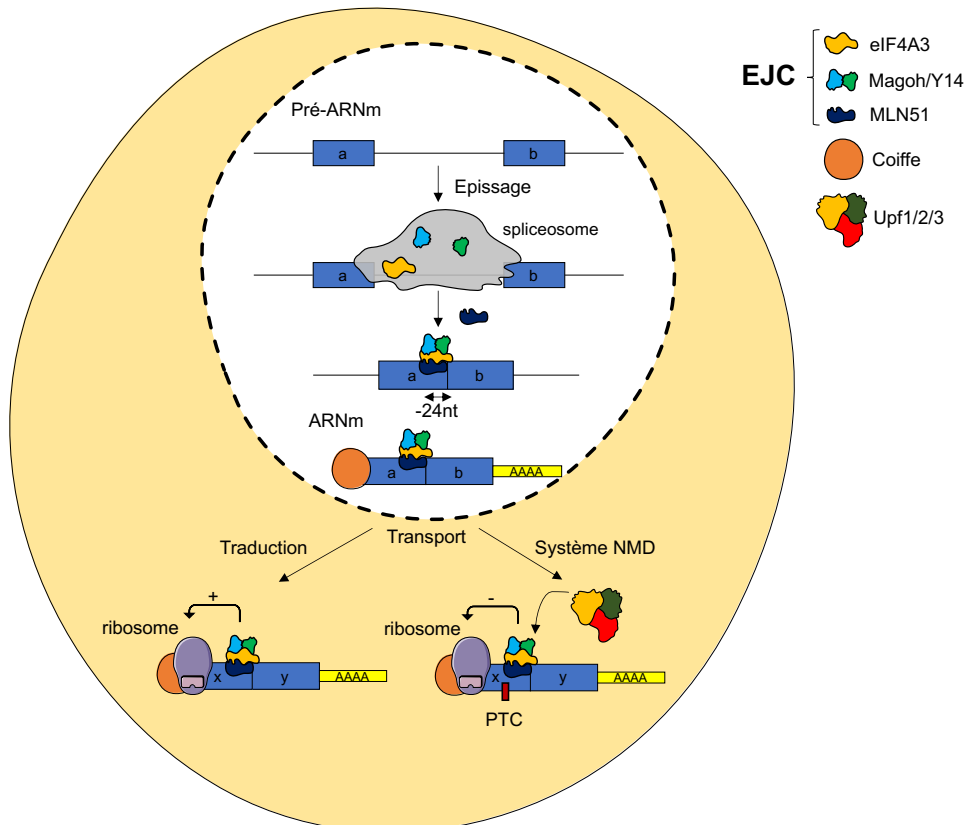


Figure 18. Assemblage de l'EJC durant l'épissage et ses rôles cytoplasmiques connus.

L'EJC s'assemble durant l'épissage en interaction avec le spliceosome dans le noyau. Les quatre protéines eIF4A3, Y14, MAGOH et MLN51 forme l'EJC déposé à 24 nucléotides de la jonction exon-exon. L'EJC joue également un rôle dans l'export de l'ARNm, l'initiation de la traduction dans le cytoplasme ainsi que le recrutement du système NMD.

## I. Les composants du cœur de l'EJC

Le cœur de l'EJC est constitué d'un complexe hétéro-tétramerique composé des protéines eIF4A3, MAGOH, Y14 qui forment un pré-complexe EJC au sein du spliceosome et de MLN51 qui rejoint le complexe plus tardivement. L'association systématique de MLN51 avec les EJC ainsi que son rôle au sein du complexe restent discutés en raison de sa faible abondance dans la cellule comparée aux autres composants du cœur de l'EJC et de sa localisation préférentiellement cytoplasmique.

### A. eIF4A3 (DDX48)

La protéine eIF4A3 est une hélicase à ARN connue aussi sous le nom de DDX48 car faisant partie des protéines à boîte DEAD (Asp-Glu-Ala-Asp,) impliquées dans de nombreux aspects du métabolisme de l'ARN. Elle appartient à la famille du facteur d'initiation eucaryote 4A (eIF4A) incluant les facteurs cytoplasmiques d'initiation de la traduction eIF4A1 et eIF4A2 avec lesquels elle partage 60% d'homologie de séquence (Lu et al., 2014). EIF4A3 est codée par le gène *eIF4A3* de 2524 pb contenant 12 exons. Elle possède 411 acides aminés (aa) pour un poids moléculaire de 47 kilodalton (kDa) (Chan et al., 2004). Elle comporte deux domaines Rec-A like en tandem (RecA1 et RecA2) caractéristiques des hélicases DEAD box, leur permettant de passer d'une conformation ouverte à une conformation fermée au prix de l'hydrolyse de l'ATP. Elle possède aussi les neuf motifs conservés au sein des protéines eIF4A (Q, I, Ia, Ib, II, III, IV, V, VI) (Ye et al., 2021). Les motifs I à III du domaine RecA1 et le motif VI du domaine RecA2 dans la partie C-terminale d'eIF4A3 permettent la liaison de l'ATP ainsi que la liaison au pré-ARNm (Figure 19) (Bourgeois et al., 2016; Linder and Jankowsky, 2011; Lu et al., 2014). EIF4A3 possède par ailleurs huit motifs spécifiques (motifs A-H) conservés entre différentes espèces et absents des protéines eIF4A1 et eIF4A2, qui lui confèrent des fonctionnalités spécifiques au sein de l'EJC dans le noyau (Shibuya et al., 2006; Woodward et al., 2017) (Figure 19). Les domaines Ia, VI et H contribuent à la liaison d'eIF4A3 sur l'ARNm mais aussi à la formation de l'EJC, au recrutement du système NMD et permettent l'interaction avec l'hétérodimère Y14-MAGOH.

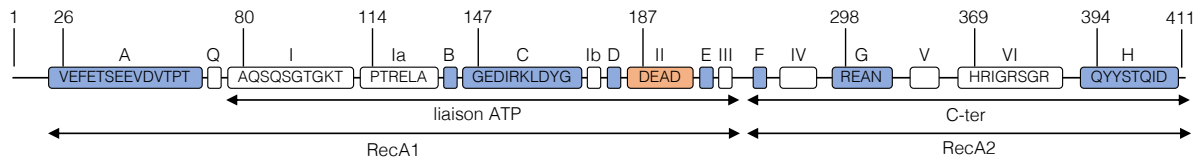


Figure 19. Représentation schématique des domaines protéiques de eIF4A3. Les motifs spécifiques d'eIF4A3 et de la famille des eIF4A sont représentés respectivement en bleu et blanc. Le motif DEAD spécifique d'eIF4A3 et des hélicases est représenté en orange.

## B. Y14/MAGOH

Y14, aussi connu sous le nom de RBM8A, est une protéine de 174aa (23kDa). Elle est codée par le gène *RBM8A* possédant 6 exons. Chez les mammifères, deux gènes paralogues situés sur des chromosomes différents, *MAGOH* et *MAGOHB*, codent pour des protéines de 146aa (17kDa) et 148aa (18kDa) (Schlautmann and Gehring, 2020; Singh et al., 2013). Les deux protéines peuvent être incorporées dans l'EJC avec la même efficacité au cours de l'épissage (Singh et al., 2013). Le domaine de liaison à l'ARN (RBD) dans la région centrale de Y14 n'est pas impliqué directement dans une liaison à l'ARN, mais dans une interaction protéine-protéine. MAGOH (*mago-nashi homolog*) se lie au niveau du RBD d'Y14 avec une forte affinité, par l'intermédiaire des motifs conservés RNP1 et RNP2, afin de former un hétérodimère (Y14/MAGOH) (Fribourg et al., 2003; Lau et al., 2003). Cette hétérodimérisation est obligatoire chez les organismes eucaryotes (Gong et al., 2014). Y14 et MAGOH ont co-évolué lentement, avec des séquences très conservées entre les différents organismes (levures, plantes, drosophiles et mammifères), afin d'assurer leur fonction de régulation du métabolisme, de localisation et de stabilité des ARNm du noyau au cytoplasme (Chuang et al., 2015).

## C. MLN51

La protéine MLN51 (*Metastatic Lymph Node 51*) aussi connu sous le nom de Barentsz (BTZ) ou CASC3 (*Cancer Susceptibility Candidate 3*) est une protéine de 703aa. Contrairement aux autres protéines du cœur de l'EJC essentiellement nucléaires, MLN51 est une protéine nucléocytoplasmique faisant la navette entre les deux compartiments cellulaires mais présente principalement dans le cytoplasme (Daguenet et al., 2012; Degot et al., 2004). Elle comporte deux domaines NLS dans la région N-terminale et un domaine NES ayant un rôle respectivement dans la localisation nucléaire et dans l'export nucléaire pour son maintien dans le cytoplasme (Daguenet et al., 2012; Macchi et al., 2003; Woodward et al., 2017). MLN51 est caractérisée par un faible taux de conservation de séquence entre l'homme et le poisson zèbre (52%) contrairement aux autres protéines de l'EJC (Woodward et al., 2017) à l'exception du

domaine SELOR (*SpEckle Localizer and RNA-binding module*) conservé à 95% avec les poissons et les amphibiens et à plus de 65% avec les invertébrés. Ce domaine, situé entre les acides aminés 148 et 246 (78aa), possède les propriétés nécessaires pour son interaction avec l'EJC (interaction avec eIF4A3, liaison à l'ARN et localisation dans les speckles nucléaires) (Degot et al., 2004).

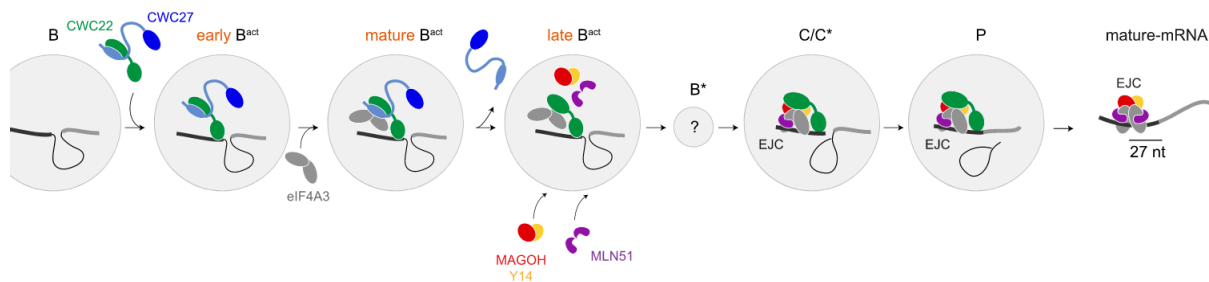
## II. Cycle de vie de l'EJC

### A. Recrutement de l'EJC dépendant de l'épissage

Le recrutement des facteurs de l'EJC, à 20-24 nt en amont des jonctions d'exons sur les pré-ARNm est dépendant de l'épissage et ne nécessite pas de séquences spécifiques (Asthana et al., 2022; Gerbracht and Gehring, 2018; Hir et al., 2016). L'assemblage se déroule par étapes successives au sein des différents complexes spliceosomaux, les facteurs étant recrutés de manière séquentielle et indépendamment du dépôt de l'EJC sur l'ARN. Le lien direct entre le dépôt de l'EJC et l'épissage des pré-ARNm a été établi par l'identification de l'interaction de eIF4A3 avec CWC22, une protéine présente dans le noyau catalytique du spliceosome où elle interagit avec la protéine PRP8 *via* son domaine MA3 (Alexandrov et al., 2012; Barbosa et al., 2012; Steckelberg et al., 2012). Cette protéine joue un rôle majeur dans l'épissage des pré-ARNm comme le montre notamment l'accumulation de transcrits non épissés en condition de déplétion de CWC22 dans les cellules HeLa (Steckelberg et al., 2015). En recrutant eIF4A3 au sein du spliceosome, la protéine CWC22 apparaît être également un acteur clé de la formation de l'EJC en initiant son assemblage lors des premières étapes de l'épissage.

### B. Assemblage de l'EJC

Les modèles de cinétique de recrutement et d'assemblage des composants du cœur de l'EJC (eIF4A3, Y14/MAGOH et MLN51) au sein du spliceosome ont évolué au cours de ces dernières années en fonction des approches expérimentales (*in vitro* ou *in vivo*) et des techniques utilisées (Gerbracht and Gehring, 2018; Schlautmann and Gehring, 2020). La Cryo-microscopie électronique (Cryo-EM) en particulier, a été d'un apport considérable en fournissant la structure du complexe C\* du spliceosome (Zhang et al., 2017). Des zones d'ombre persistent néanmoins en raison de l'absence de données précises concernant les étapes de recrutement du dimère Y14/MAGOH et de MLN51, pour lesquelles l'accès à la structure du complexe B\* en Cryo-EM serait nécessaire. Le modèle le plus récent de l'assemblage de l'EJC incluant la participation d'une nouvelle protéine, CWC27 en association avec CWC22 (Busetto et al., 2020), est présenté ci-après (Figure 20).



Busetto et al., 2020

Figure 20. Formation du complexe EJC.

L'hétérodimère CWC22/CWC27 recrute eIF4A3 lors de la formation du complexe B<sup>act</sup>. L'EJC est ensuite formé dans le complexe C avec le recrutement de Y14/MAGOH et de MLN51 et est déposé sur l'ARNm en amont de la jonction exon-exon.

## 1. Recrutement d'eIF4A3 par CWC22/CWC27

Il a classiquement été décrit que le recrutement d'eIF4A3 par CWC22 constitue l'étape initiale de la formation du complexe EJC. CWC22 est présente dans le spliceosome dès la formation du complexe B<sup>act</sup>. Elle interagit avec les deux domaines RecA1/2 d'eIF4A3 *via* son domaine MIF4G (*Middle of Initiation Factor 4G*) pour former un complexe distinct à ce stade, du complexe EJC assemblé. La structure Cryo-EM du complexe C\* humain (Zhang et al., 2017) montre que la localisation du domaine MIF4G de CWC22 au niveau de l'exon 5' permet de positionner eIF4A3 à environ -27 nt de l'extrémité de l'exon 5' et de guider la formation de l'EJC en amont de la jonction exon-exon (Schlautmann and Gehring, 2020; Wilkinson et al., 2020). Dans le dimère CWC22/eIF4A3 initial, eIF4A3 est maintenue dans une conformation ouverte inactive, qui l'empêche de se lier à l'ATP et au pré-ARNm (Alexandrov et al., 2012; Barbosa et al., 2012 ; Buchwald et al., 2013; Steckelberg et al., 2015).

Selon des données récentes, CWC22 pourrait ne pas agir seule lors du recrutement d'eIF4A3 mais en association avec un autre facteur d'épissage, CWC27, dont elle est la principale protéine partenaire (Busetto et al., 2020). La protéine CWC27 est une peptidyl-prolyl *cis-trans* isomérase (PPIase) dont le domaine catalytique PPI est inactif, mais qui conserve sa capacité de liaison aux prolines. Elle jouerait un rôle de protéine chaperonne facilitant les interactions protéine-protéine lors de l'assemblage du spliceosome. Les mutations du gène codant CWC27 sont responsables d'une maladie à transmission autosomique récessive, caractérisée par un large spectre clinique allant de formes isolées de rétinite pigmentaire à des formes syndromiques incluant des anomalies du développement crânio-facial et squelettique selon la nature et la localisation des mutations (Xue et al., 2021).

Des études par cristallographie aux rayons X de la structure 3D du complexe ternaire CWC27/CWC22/eIF4A3 ont permis d'établir que le domaine C-terminal de CWC27 se lie étroitement au domaine MIF4G de CWC22, formant un hétérodimère stable avant d'être intégré

dans le complexe B<sup>act</sup> du spliceosome. Ce dimère sert de plateforme de recrutement à eIF4A3, qui se lie au côté opposé du domaine MIF4G de CWC22 par son domaine RecA2. EIF4A3 adopte alors une conformation semi-ouverte pour permettre son intégration dans le spliceosome (Figure 21) (Busetto et al., 2020). CWC27 est ensuite libérée du complexe B<sup>act</sup> lors de la conversion en complexe B\* avant l'assemblage de l'EJC, la présence simultanée de CWC27 et Y14 dans le spliceosome étant incompatible sur le plan structural. CWC22 reste liée à eIF4A3 dans le complexe C, y compris sur l'EJC assemblé dans les structures des complexes C, C\* et P (Bertram et al., 2017; Haselbach et al., 2018; Zhang et al., 2019). En présence d'ATP, les deux domaines RecA d'eIF4A3 adopte une conformation fermée qui interagit avec le squelette phosphate-ribose de six nucléotides. En absence de contact avec les bases, cette interaction d'eIF4A3 sur l'ARNm est indépendante de la séquence (Andersen et al., 2006; Bono et al., 2006; Gerbracht and Gehring, 2018; Zhang et al., 2019).

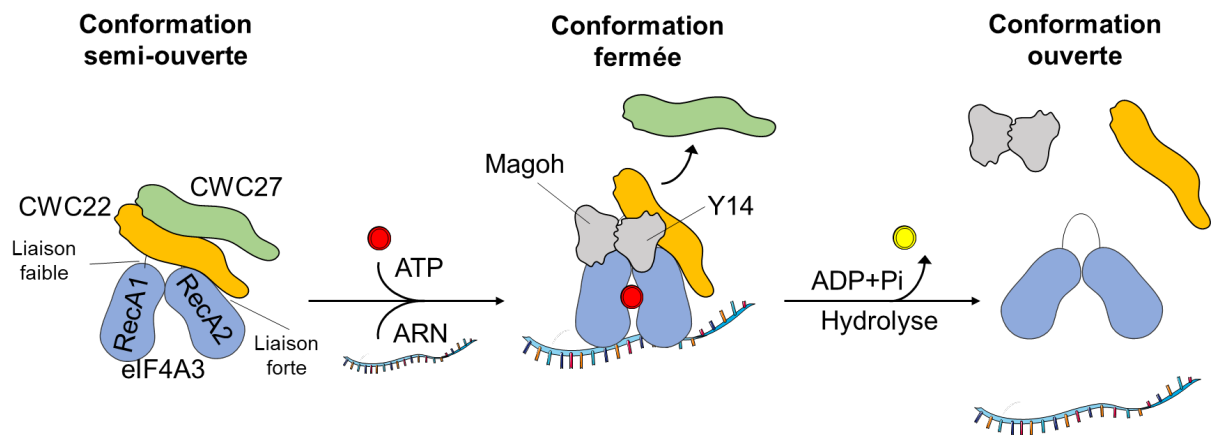


Figure 21. Mécanisme de fixation de l'hélicase à l'ARN eIF4A3.

L'hétérodimère CWC22/27 place eIF4A3 en conformation semi-ouverte grâce à une liaison forte associée à une liaison faible de CWC22 sur eIF4A3. Le recrutement de Y14/MAGOH sur eIF4A3 ainsi que la fixation d'ATP au niveau des domaines RecA1/2, engendre une conformation fermée de l'hélicase sur l'ARNm. L'hydrolyse de l'ATP qui ouvre eIF4A3 dissocie ensuite le complexe de l'ARNm.

## 2. Fixation de l'hétérodimère Y14/MAGOH sur eIF4A3

Y14 et MAGOH forment un hétérodimère dans le cytoplasme avant d'être importé dans le noyau par l'importine 13 (Chuang et al., 2015). La fixation de l'hétérodimère sur eIF4A3 maintient sa conformation fermée empêchant ainsi une libération rapide de l'ATP hydrolysé et stabilise l'association du noyau trimérique lié à l'ATP sur l'ARN (Figure 21) (Bourgeois et al., 2016; Nielsen et al., 2009; Schlautmann and Gehring, 2020). Les mécanismes et la cinétique de recrutement du dimère Y14/MAGOH par le spliceosome ne sont pas complètement élucidés et pourraient mettre en jeu une interaction avec une autre protéine, l'hélicase Aquarius (aussi appelée IBP160) avant la fixation sur eIF4A3 (Woodward et al., 2017). Aquarius se fixe de manière séquence-indépendante dans les introns lors de la formation du complexe C. Elle a été



décrite comme recrutant eIF4A3, Y14 et MAGOH au sein du spliceosome *in vitro* et chez *C.elegans* (Ideue et al., 2007; Shiimori et al., 2013). Le domaine MIF4G de CWC22 interagit également avec Y14 dans le complexe C (Gerbracht and Gehring, 2018).

### C. EJC trimérique vs tétramérique

Durant le complexe C d'épissage, après la formation du pré-EJC (eIF4A3-Y14/MAGOH), MLN51 vient se fixer sur eIF4A3 *via* son domaine SELOR afin de s'incorporer à l'EJC (Busetto et al., 2020). Cependant, l'interaction reste flexible et incomplète puisque seul le fragment correspondant aux acides aminés 170-194 de MLN51 interagit avec le domaine RecA2 d'eIF4A3 (Gerbracht and Gehring, 2018). Ce n'est qu'après le départ de CWC22 dans le complexe P après la seconde réaction d'épissage que MLN51 peut se replier totalement sur eIF4A3 et se fixer sur l'ARN (Hir et al., 2016). La partie N-terminale (aa 168-196) entre en contact avec le domaine RecA2 tandis que la région des aa 214-245 interagit avec le domaine RecA1 d'eIF4A3 (Busetto et al., 2020; Gerbracht and Gehring, 2018).

MLN51 a été décrite comme faisant partie intégrante du cœur de l'EJC sur la base d'études *in vitro* qui ont montré son interaction stable avec eIF4A3 (Ballut et al., 2005; Tange et al., 2005). Sa liaison à l'EJC trimérique est observée dans la structure cristalline du complexe C\* du spliceosome humain (Bono et al., 2006; Tange et al., 2005; Zhang et al., 2019, 2017). Son rôle nucléaire au sein de l'EJC reste cependant discuté dans la littérature en raison de sa présence en quantité sub-stoechiométrique dans la cellule comparé aux autres facteurs nucléaires du cœur de l'EJC (Singh et al., 2012). Comme mentionné précédemment, sa localisation est principalement cytoplasmique, bien qu'elle fasse la navette entre les deux compartiments cellulaires. Elle colocalise avec les autres composants de l'EJC dans le noyau si son export nucléaire est inhibé (Daguenet et al., 2012). D'autre part, l'invalidation de MLN51 dans des lignées cellulaires humaines n'a pas d'effet sur la formation de l'EJC ni sur les événements d'épissage dépendant de l'EJC mais stabilise les transcrits cibles du NMD (Gerbracht et al., 2020). L'analyse des interactomes de MLN51 ou RNPS1 obtenus par des expériences d'immunoprécipitation dans des cellules HEK293 ou HeLa indiquent que MLN51 n'est pas présente dans tous les EJC mais incluse de manière mutuellement exclusive avec RNPS1 (Mabin et al., 2018).

Tous ces éléments tendent à indiquer que MLN51 ne joue pas de rôle prépondérant lors des étapes initiales de formation de l'EJC dans le noyau. MLN51 est présentée comme une protéine périphérique de l'EJC assurant un rôle principalement cytoplasmique dans la voie du NMD

(Gerbracht et al., 2020). Selon le modèle proposé, MLN51 se fixerait principalement sur les EJC situés dans les régions 3'UTR afin de les stabiliser sur l'ARNm et renforcer leur capacité à activer la voie du NMD. MLN51 contribuerait ainsi à réduire la quantité résiduelle de transcrits porteurs de PTC ayant échappé à la dégradation lors du cycle initial de traduction grâce à la persistance des EJC qui augmenteraient la probabilité de clivage par la protéine SMG6 lors des cycles ultérieurs de traduction. Il est en effet montré que des transcrits cibles du NMD peuvent subir plusieurs cycles de traduction avant d'être dégradés (Hoek et al., 2019).

## D. Recrutement des protéines périphériques de l'EJC

Des facteurs dits périphériques se lient directement ou indirectement au cœur de l'EJC sous forme de complexes dont la composition change de façon dynamique en fonction des étapes du métabolisme des ARN. Le remodelage des protéines périphériques entre le noyau et le cytoplasme permet de réguler les processus dépendant de l'EJC dans ces compartiments cellulaires (Gerbracht and Gehring, 2018; Woodward et al., 2017).

### 1. Les complexes ASAP/PSAP

Dans le noyau, deux complexes importants s'associent à l'EJC. Ils sont composés des protéines RNPS1, SAP18 avec soit ACINUS, appelé complexe ASAP (*apoptosis and splicing-associated protein*), soit avec PININ, appelé complexe PSAP, (Figure 22) (Murachelli et al., 2012; Schlautmann and Gehring, 2020). Lors de la formation du complexe ASAP, le motif de reconnaissance de l'ARN (RRM) de RNPS1 se lie à un motif conservé d'ACINUS (RSB) avec un mode de reconnaissance similaire observé dans les protéines d'épissage U2AF ou la snRNP U1. L'association ACINUS-RNPS1 recrute ensuite le domaine UBL de type ubiquitine de SAP18. PININ possède le même motif RSB qu'ACINUS pour former le complexe PSAP avec RNPS1 et SAP18. Ces trimères possèdent des propriétés de liaison à l'ARN et aux protéines

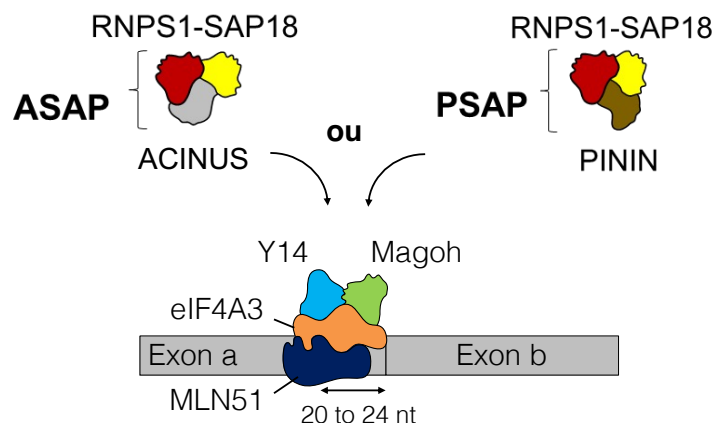


Figure 22. Composition des complexes ASAP et PSAP liés à l'EJC. Le complexe EJC formé sur l'ARNm recrute le complexe ASAP ou PSAP pour la régulation de l'inclusion des exons dans les transcrits matures.

Les deux complexes participent à la régulation d'évènements d'épissage alternatif dépendant de l'EJC comme l'utilisation de sites cryptiques d'épissage ou l'activation d'épissage d'introns avec des sites d'épissage faibles (Schlautmann and Gehring, 2020). Les mécanismes de recrutement des complexes ASAP/PSAP dans l'EJC ne sont pas encore totalement élucidés, notamment si leur interaction nécessite des protéines supplémentaires (Gerbracht and Gehring, 2018).

## 2. Les protéines SR

La plupart des protéines SR classiques et d'autres protéines de type SR, ont été récemment trouvées en interaction avec l'EJC. Des motifs riches en purines GAAGA ont été identifiés comme étant des plateformes de recrutement de, SRSF3, SRSF4, SRSF7, Tra2a/b, SRSF9 SRSF10 (Hauer et al., 2016; Mabin et al., 2018; Obrdlik et al., 2019). Des techniques d'immunoprécipitation ont montré que ces protéines s'associent physiquement avec eIF4A3 et MAGOH (Mabin et al., 2018; Saulière et al., 2012). De plus, des motifs riches en CG et AG présents autour des EJC ont une affinité particulière avec SRSF3, SRSF1 et Tra2b (Singh et al., 2012).

## 3. Les protéines d'export des ARNm

D'autres protéines périphériques de l'EJC du complexe contribuent aux fonctions de l'EJC. Les protéines UAP56 et AlyREF qui permettent la liaison de MLN51 à eIF4A3 lors de la formation de l'EJC (Gromadzka et al., 2016) sont impliquées dans l'export des ARNm vers le cytoplasme. De plus, il est intéressant de noter que le domaine SELOR de MLN51 est également présent dans d'autres protéines interagissant avec l'EJC.

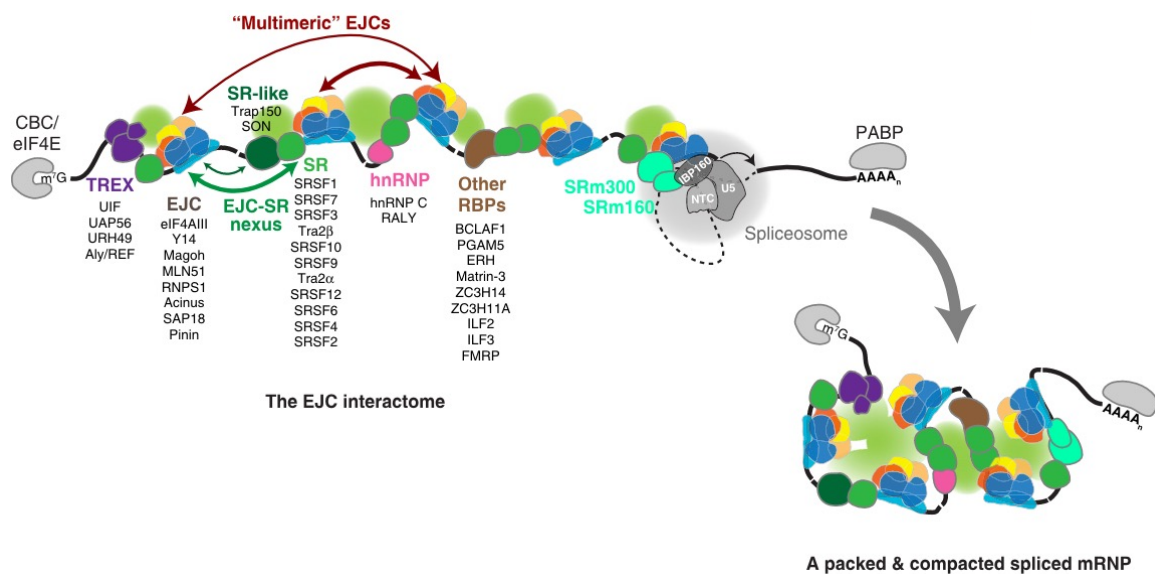
## 4. Les facteurs à motif EBM

Les interactions les mieux comprises avec le cœur de l'EJC sont celles impliquant les protéines possédant un motif EBM (*EJC Binding Motif*) en partie N ou C-terminale facilitant leur liaison. C'est le cas notamment des protéines UPF3B et UPF3A, facteurs du système NMD (Buchwald et al., 2010). D'autres protéines humaines possédant le motif EBM interagissent aussi directement avec l'EJC, notamment les protéines SMG6 (autre facteur du NMD), TDRD3 (protéine de liaison à la chromatine lors de la transcription) ou encore RRP12 (RBP des ARN ribosomiaux) (Kashima et al., 2010).

## E. Mise en place des structures mRNP autour de l'EJC

L'EJC façonne la structure, la composition et le devenir des particules ribonucléoprotéiques (mRNP) d'ARNm. Ils forment de très grandes structures compactées comportant de nombreuses

protéines (Figure 23) (Mabin et al., 2018; Metkar et al., 2018; Singh et al., 2012; Woodward et al., 2017). Différentes études ont pu établir les interactions existantes entre l'EJC, les protéines SR, ACINUS, RNPS1 et MLN51 dans la dynamique de remodelage des mRNP. Les EJC se multimérisent par des interactions faibles avec les autres composants des mRNP afin de former des complexes flexibles de très grande taille (mégadalton) qui protègent les ARNm nouvellement synthétisés de l'action des endonucléases (Metkar et al., 2018). RNPS1 et ACINUS pourraient servir de pont entre le cœur de l'EJC et les protéines SR liées à distance (Rodor et al., 2016). Cela pourrait expliquer le taux variable d'occupation des EJC aux jonctions exon-exon et le biais observé pour les ARNm alternativement épissés qui nécessitent une régulation par les protéines SR chez les mammifères.

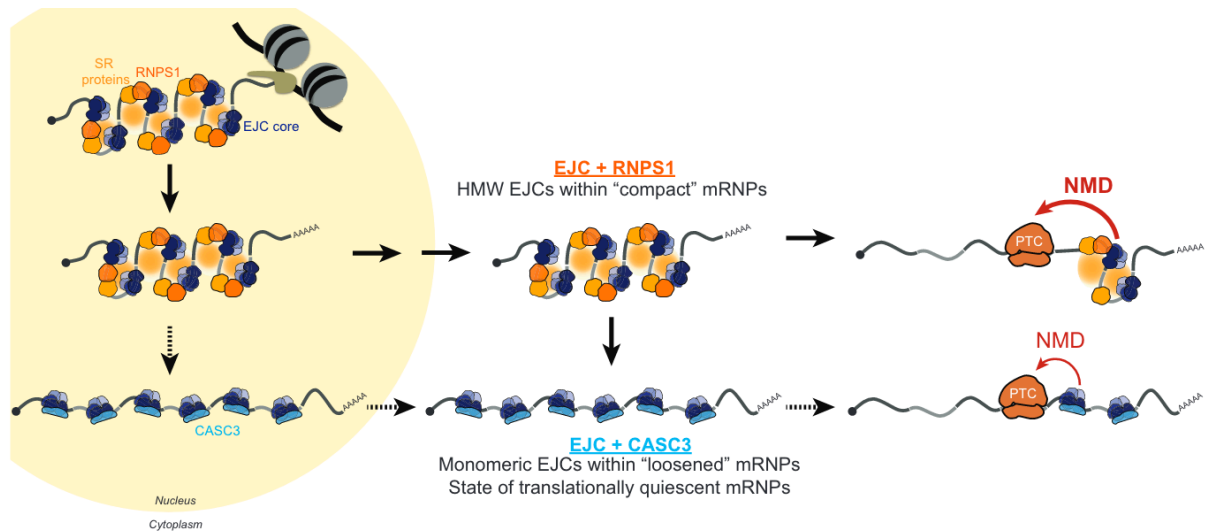


Singh et al., 2012

Figure 23. Interactome de l'EJC favorisant la formation des structures compactes de mRNP. La mise en place des structures mRNP passe par une interaction de l'EJC avec de nombreuses protéines de liaison à l'ARN (RBPs, RNA Binding Proteins) impliquées dans l'épissage (SR, hnRNP) ou l'export des ARNm (TREX).

Une étude récente a démontré que les mRNP construites autour de l'EJC subissaient un changement de conformation entre le noyau et le cytoplasme (Figure 24). Les protéines RNPS1 et SR associées à des EJC multimériques (EJC-RNPS1-SR) et l'EJC monomérique lié à MLN51 (EJC-MLN51) sont présentes de façon mutuellement exclusive entre le noyau (présence majoritaire de EJC-RNPS1-SR) et le cytoplasme (présence majoritaire de EJC-MLN51). Les EJC qui interagissent avec RNPS1 et les SR ou MLN51 se lient à des mRNP fonctionnellement différentes, notamment en ce qui concerne l'activité du NMD ou le désassemblage de l'EJC dépendant de l'initiation de la traduction (Gerbracht et al., 2020; Mabin et al., 2018). Cette observation est en accord avec la localisation principalement cytoplasmique de MLN51.

Globalement, ces résultats montrent que l'EJC est une plateforme d'interaction et de régulation des mRNP indispensable dans la régulation du métabolisme des ARNm.



Mabin et al., 2018

Figure 24. Dynamique de régulation des structures mRNP en fonction de l'interactome de l'EJC.

Le changement de composition des EJC impacte la conformation des mRNP et détermine une activité différente du système NMD : les mRNP associant les EJC multimériques à RNPS1 et aux protéines SR dans une structure compactée permettent un fort recrutement du système NMD sur tous les transcrits cibles tandis que les mRNP peu structurés comportant des EJC monomériques associés à MLN51 réduiraient l'activité du NMD et agirait sur un nombre restreint de transcrits.

### III. Localisation de l'EJC

#### A. Dépôt des EJC sur le transcriptome

##### 1. Position des EJC canoniques et non-canoniques

Depuis 2012, différentes techniques d'analyse à large échelle des sites d'interaction protéines-ARN ont été utilisées telles que le CLIP-Seq (*crosslinking and immunoprecipitation of RNA-protein complexes*) ou et le RIP (*RNA immuno-precipitation*) couplées au séquençage haut débit afin de déterminer la position des dépôts d'EJC dans le transcriptome humain. La première étude de CLIP-Seq d'eIF4A3 (Saulière et al., 2012) avait rapporté que seuls 50% des pics EJC étaient retrouvés à la position attendue en -20 à 24 nt en amont de la jonction d'épissage (EJC canoniques, cEJC). La seconde moitié des pics étaient identifiés à d'autres positions dans les exons (EJC non canoniques, ncEJC). Ces dépôts étaient retrouvés préférentiellement dans les exons bordés par de larges introns. Par ailleurs, 20% des exons semblaient dépourvus de marques, suggérant un dépôt différentiel des EJC selon le transcrit ou les jonctions.

Le développement de nouvelles approches et l'amélioration des techniques (iCLIP, eCLIP, xRIPiT) appliquées à l'étude de protéines du complexe EJC en interaction directe avec l'ARN (RBP) ou indirecte via des RAF (*RBP-associated factors*) ont permis depuis de confirmer que la quasi-totalité des EJC sont déposés *in vivo* sur les sites canoniques dans les cellules humaines

et que la majorité des exons (>85%) portent un EJC (Hauer et al., 2016; Hocq et al., 2018; Kovalak et al., 2021; Patton et al., 2020). Des différences de distribution sont néanmoins observées, avec un enrichissement au niveau des exons soumis à un épissage alternatif et au contraire, un déficit sur les ARNm codant pour des protéines ribosomiques (Hauer et al. 2016). Les protéines de l'EJC (eIF4A3, MLN51, RNPS1 et UPF3B) co-localisent aux mêmes sites de liaison sur l'ARN correspondant au site de dépôt des cEJC. Seules les trois protéines cœur eIF4A3 et Y14/MAGOH sont retrouvées dans les 10% de ncEJC (Hauer et al., 2016). MLN51 est absente confirmant que cette protéine ne fait pas partie de tous les EJC et a un rôle particulier au sein de l'EJC. Des sites de liaison indépendants du dépôt EJC ont été aussi identifiés pour RNPS1, liés potentiellement à des fonctions spécifiques de la protéine (Hauer et al., 2016). La technique de xRIPiT-seq, (RIP en tandem couplée à l'utilisation de formaldéhyde pour l'étape de cross-link) particulièrement bien adaptée pour l'étude des RAF, a mis en évidence une augmentation de sites de liaison pour RNPS1 dans les exons immédiatement en aval d'exons subissant un épissage récursif, et dont l'inclusion est dépendante de l'EJC (Figure 25) (Patton et al., 2020). Ce résultat est à mettre en relation avec la fonction connue de RNPS1 dans la répression de l'utilisation de sites 5' d'épissage aux jonctions d'exons récursifs qui sera détaillée dans le chapitre « Rôle de l'EJC dans l'Épissage ».

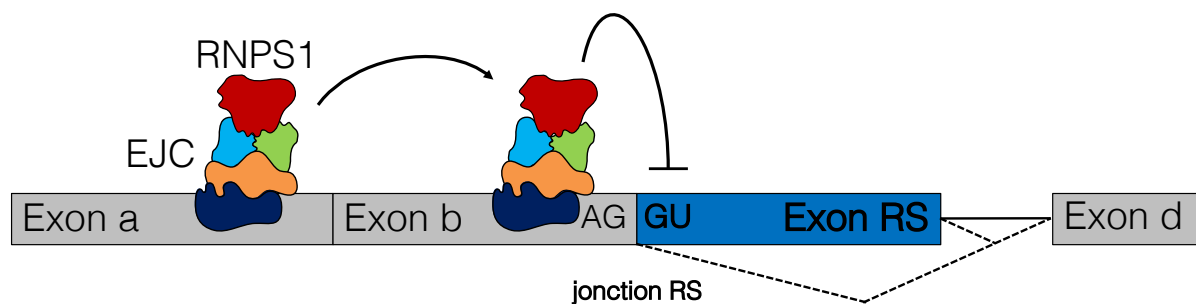


Figure 25. Représentation schématique de l'épissage d'un exon récursif (RS) régulé par l'EJC associé à RNPS1. Le complexe EJC associé à RNPS1 dans l'exon b en amont de la jonction RS inhibe l'utilisation du site donneur (GU) de l'exon RS. De plus, le complexe EJC-RNPS1 déposé en amont (exon a) stabilise le complexe EJC-RNPS1 en aval (exon b).

Le principal déterminant du dépôt de l'EJC est sa localisation en -24nt en amont de la jonction exon-exon, indépendamment de la séquence. Cette caractéristique est conservée entre espèces, notamment chez la drosophile et le poisson zèbre chez lesquels les cEJC sont très largement majoritaires (>95%) (Gangras et al., 2020; Obrdlik et al., 2019). Dans les cellules humaines, des co-facteurs tels que les protéines SR contribuent très probablement à ce mécanisme compte tenu de la fréquence élevée de motifs riches en purines reconnus par certaines protéines SR à proximité des dépôts de RNPS1, UPF3B, MLN51, ACINUS ou de l'EJC (Hauer et al., 2016; Rodor et al., 2016).

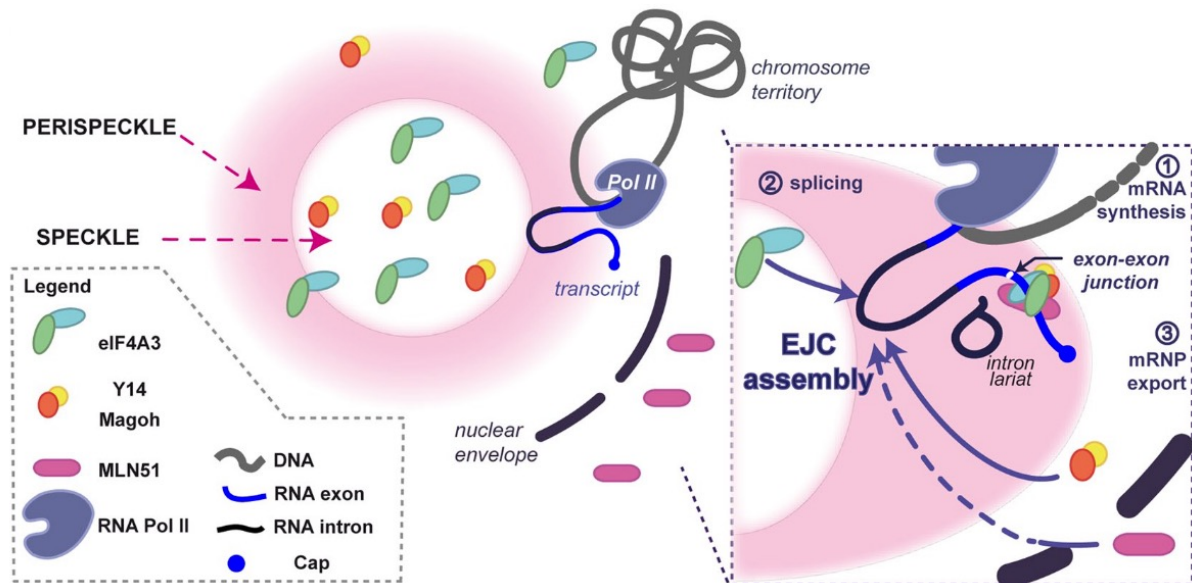
## 2. Structure secondaire de l'ARN et dépôt de l'EJC

Chez les mammifères comme chez la drosophile, les cEJC sont déposés sur des séquences ayant une faible capacité à former des structures secondaires (Obrdlik et al., 2019; Saulière et al., 2012). La présence d'une structure secondaire de l'ARN masquant la position -24 peut entraîner le déplacement du site de dépôt de 10 nt en amont ou en aval et en conséquence, une baisse du degré d'occupation par les EJC (Haberman et al., 2017). Chez la drosophile, cependant, la présence d'une structure double brin d'ARN à proximité de la position canonique favoriserait le dépôt de l'EJC, un effet non retrouvé chez les mammifères (Obrdlik et al., 2019).

### B. Localisation de l'EJC dans la cellule

La première étude de localisation de l'EJC a permis de montrer que les facteurs de l'EJC ainsi que les protéines périphériques UAP56, Aly/REF, NXF1 et RNPS1 interagissaient avec les corps nucléaires (*Nuclear speckles*) (Schmidt et al., 2009). Dans le noyau, ces structures sont des corps dynamiques ponctuels sans membrane enrichis en facteurs protéiques nécessaires aux différentes étapes de l'expression des gènes, notamment les RBP qui interviennent dans l'épissage et l'export des ARNm cellulaires. Elles peuvent jouer un rôle dans le positionnement des gènes sur des repères nucléaires spécifiques en améliorant le traitement de l'expression de l'ARN (Faber et al., 2022). En effet, ces speckles fonctionnent comme des compartiments de stockage, d'assemblage et de modification qui fournissent les facteurs d'épissage aux sites de transcription actifs qui abritent les pré-ARNm. Une étude ultérieure basée sur l'utilisation de différentes techniques de microscopie de fluorescence notamment le FRET (Transfert d'énergie par résonance de fluorescence) a permis de préciser que l'assemblage du cœur de l'EJC (eIF4A3, Y14/MAGOH) dans la cellule se déroulait dans une zone périphérique des speckles appelée paraspeckles dans lesquelles les pré-ARNm sont épissés (Daguenet et al., 2012) (Figure 26). Des protéines périphériques associées à l'EJC notamment le complexe ASAP et UPF3B sont aussi retrouvées au niveau de ces paraspeckles. Plus récemment, il a été montré que l'EJC joue un rôle crucial dans l'enrichissement spatial d'ARNm spécifiques dans un compartiment cellulaire précis des cellules humaines (Kwon et al., 2021). Il a en effet été observé que les protéines du cœur de l'EJC étaient particulièrement abondantes autour des centrosomes à la base des cils primaires des cellules monociliées quiescentes chez la souris et dans une lignée de cellules épithéliales rétinienne humaine (RPE1). Cette accumulation est ARN-dépendante et assurée par une voie dépendante des microtubules. Elle survient pendant l'état de quiescence, au cours duquel des transcrits non (ou partiellement)

traduits liés par l'EJC s'accumulent au niveau des centrosomes, et diminue lorsque ces cellules se différencient ou reprennent leur croissance.



Daguenet et al., 2012

Figure 26. Localisation subcellulaire de l'assemblage de l'EJC co-transcriptionnel. Les pré-ARNm naissants synthétisés par la Pol II (étape 1) sont transformés en mRNP (étape 2) dans les paraspeckles. Après l'assemblage du spliceosome, eIF4A3 et Y14/MAGOH sont recrutés sur l'ARN. Après la fin de l'épissage, MLN51 rejoint le complexe. Le noyau stable de l'EJC recrute les facteurs périphériques (étape 3) pour l'export de la mRNP.

## IV. Les rôles de l'EJC

Les composants principaux et périphériques de l'EJC contribuent à des étapes nucléaires ou cytoplasmiques différentes de la régulation post-transcriptionnelle de l'expression des gènes, telles que l'épissage, la localisation des ARNm, la traduction et la dégradation des ARNm par la voie du NMD. Nous développerons plus en détail le rôle de l'EJC et des protéines associées dans l'épissage, axe plus particulièrement étudié au cours de ma thèse. Les données récentes obtenues aussi bien chez la drosophile que dans les cellules humaines viennent confirmer le rôle déterminant de l'EJC dans le maintien de la fidélité de l'épissage des pré-ARNs, la régulation de l'épissage alternatif et par conséquent dans les programmes d'expression génique.

### A. Rôle nucléaire de l'EJC

#### 1. Rôle de l'EJC et des protéines associées dans l'épissage

##### a. Drosophile et autres modèles

Les premières études rapportant un rôle de l'EJC dans l'épissage ont été réalisées chez la drosophile qui ont montré que les composants de l'EJC (eIF4A3, Mago, Y14) et RNPS1 influençaient l'épissage des longs introns de gènes situés dans des régions d'hétérochromatine et interféraient avec les mécanismes de reconnaissance des exons. Cela a été observé



notamment pour le gène *mapk* de la voie de signalisation RAS/MAPK impliquée dans la prolifération, la différenciation et la survie cellulaire (Ashton-Beaucage et al., 2010; Roignant and Treisman, 2010). La déplétion des facteurs du pré-EJC trimérique (eIF4A3, Y14 et MAGOH) et des protéines ACINUS et RNPS1 du complexe ASAP, induit la rétention d'introns ayant des sites faibles d'épissage comme l'intron 4 de *piwi*. Leur épissage est dépendant du dépôt d'EJC à des jonctions exon-exon situées à proximité (environ 250 nt) (Hayashi et al., 2014). Un défaut d'épissage de l'intron 4 conduit à la dérégulation de deux éléments transposables normalement réprimés par la voie des piRNA (*Piwi-interacting RNA*) dans les ovaires chez la drosophile (Hayashi et al., 2014; Malone et al., 2014).

Le rôle de l'EJC dans l'élimination des introns a été également rapporté dans un modèle de Paramécie qui contient essentiellement de petits introns (Contreras et al., 2018) et chez le Xénope où eIF4A3 est requis pour l'épissage correcte de l'intron 101 du gène du récepteur à la ryanodine (Ryr) (Haremakei and Weinstein, 2012). Chez *C. elegans*, la formation d'un pré-EJC assure une fonction de surveillance de l'épissage des introns en interaction avec des composants du spliceosome, permettant la rétention dans le noyau des transcrits incomplètement épissés. En absence de eIF4A3, Y14, MAGOH ou UAP56, les transcrits non ou mal épissés sont exportés prématurément dans le cytoplasme (Shiimori et al., 2013).

#### b. Chez les mammifères et dans les cellules humaines

Les études menées à partir de 2012 ont confirmé le rôle de l'EJC dans l'épissage des gènes humains. Les protéines eIF4A3 et Y14, les composants du complexe ASAP (ACINUS, RNPS1 et SAP18) ainsi que PININ jouent un rôle dans l'épissage alternatif du pré-mRNA Bcl-x<sub>L</sub> (augmentation de la forme pro-apoptotique Bcl-x<sub>S</sub>) (Leu et al., 2012; Michelle et al., 2012) et plus généralement des transcrits impliqués dans le contrôle de l'apoptose et du cycle cellulaire (Michelle et al., 2012). Plusieurs études ont confirmé depuis ces données notamment dans les gènes possédant de petits introns (< 500 nt). En condition de KD de Y14 dans les cellules HeLa, on observe la rétention de ces petits introns (épissés selon le modèle de définition d'introns compte tenu de leur taille) dans de nombreux gènes notamment AURKB, MAP4K4, MDM2. Ces épissages aberrants conduisent à un arrêt du cycle au point de contrôle G2/M et à l'apoptose (Fukumura et al., 2016). Le KD de RNPS1 reproduit certains de ces épissages (Fukumura et al., 2018). Une inhibition de l'activité de eIF4A3 par une approche pharmacologique a confirmé aussi le rôle d'eIF4A3 dans la régulation de l'expression des gènes du cycle cellulaire et a révélé un nouveau rôle d'eIF4A3 dans la formation des granules de stress (Mazloomian et al., 2019).

L'analyse transcriptomique de cellules HeLa déplétées pour les protéines eIF4A3, Y14 et MLN51 a permis par ailleurs d'établir qu'un grand nombre d'exons alternatifs (cassette exons) mais aussi dans une moindre proportion d'exons constitutifs, était dépendant de la présence du complexe EJC (Wang et al., 2014). Les défauts d'épissage observés sont principalement des sauts d'exons (généralement flanqués par de grands introns) et dans une moindre proportion des rétentions d'introns caractérisés par des sites d'épissage faibles ou l'utilisation de sites alternatifs 5' et 3' dans les exons.

## 2. MLN51 controversé comme facteur d'épissage

Les protéines du cœur de l'EJC ne régulent pas l'épissage de manière individuelle mais sous la forme du complexe trimérique (eIF4A3, Y14/MAGOH) formé précocement au sein du spliceosome. Les profils d'épissage obtenus par la déplétion individuelle de ces trois protéines sont généralement très comparables. La contribution de MLN51 qui rejoint le complexe à un stade plus tardif, reste discuté en raison de données contradictoires dans la littérature obtenues par des approches expérimentales différentes.

Contrairement à l'étude de Wang et al. (2014) dans les cellules HeLa, plusieurs études réalisées chez la drosophile ou dans les cellules de mammifères n'ont pas pu mettre en évidence d'effet de la déplétion de MLN51 sur des événements d'épissage régulés par le complexe trimérique et/ou certaines protéines périphériques associées (Ashton-Beaucage et al., 2010; Hayashi et al., 2014; Malone et al., 2014; Michelle et al., 2012; Roignant and Treisman, 2010). Récemment, une étude dans les cellules HEK293 n'a pas permis de reproduire en condition de KD de MLN51 ou d'inactivation de l'expression du gène MNL51 par CRISPR/Cas9 (*knock-out*, KO) les défauts détectés auparavant dans les gènes RER1, OCIAD1, et MRPL3 dans les cellules HeLa (Gerbracht et al., 2020). Plus globalement, les données de cette étude montrent que l'absence d'expression de MLN51 n'a pas d'effet sur les événements d'épissage régulés par l'EJC dans le noyau ni sur la composition du complexe. Elle augmente en revanche de façon significative le niveau des transcrits qui incluent un exon alternatif contenant un PTC et l'expression différentielle de nombreuses cibles du NMD. Ces données confortent le rôle décrit de MLN51 dans l'efficacité du mécanisme NMD dépendant de l'EJC dans le cytoplasme.

## 3. Rôle des complexes ASAP/PSAP

Les complexes ASAP et PSAP interagissent avec l'EJC *via* leur composant individuel, ACINUS ou PININ. Leur fixation sur le l'EJC est mutuellement exclusive. Les complexes sont fonctionnellement différents et ont des rôles sur l'épissage non interchangeables (Wang et al.,

2018). Ainsi, les épissages aberrants induits par le KD de PININ et donc en absence du complexe PSAP, ne peuvent pas être corrigés par l'expression d'ACINUS et le complexe ASAP. Ainsi, les combinaisons de complexes EJC-ASAP ou EJC-PSAP peuvent réguler des événements d'épissage différents, d'autres épissages alternatifs étant régulés par l'EJC uniquement, indépendamment des complexes ASAP ou PSAP (Wang et al., 2018).

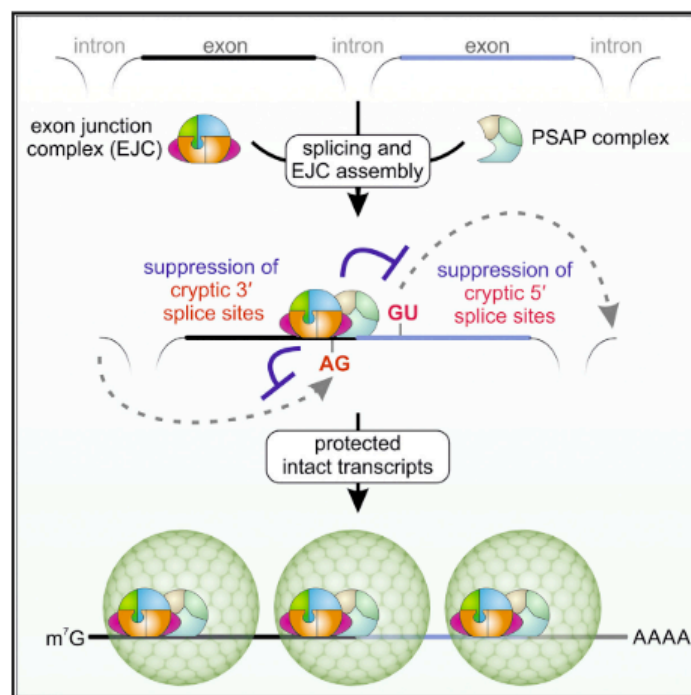
Au sein du complexe PSAP, RNPS1 participe à la formation du complexe A du spliceosome en interagissant avec PININ qui est en interaction avec la snRNP U2. Le domaine RRM de RNPS1 est nécessaire à la formation du complexe ASAP ou PSAP. Il suffit à corriger la majorité des épissages aberrants induits par la déplétion de RNPS1 dans des lignées cellulaires humaines mais l'épissage dans certains gènes nécessitent la protéine RNPS1 complète (Schlautmann et al., 2021). Le domaine RRM interagit également avec des protéines SR et la snRNP U1. La protéine RNPS1 peut aussi être impliquée dans des interactions en dehors de l'EJC, comme avec les protéines SRRM4 et SRSF11 au niveau de séquences régulatrices ISE, pour l'inclusion correcte dans le cerveau de micro-exons impliqués dans l'autisme (Gonatopoulos-Pournatzis et al., 2018).

La protéine PININ ne possède pas de domaine RRM ou RS. La déplétion de cette protéine régule négativement le niveau de protéines SR influençant en conséquence le choix des sites d'épissage (Leu, 2021). PININ peut interagir avec les protéines SR mais aussi avec un complexe ASAP déjà formé avec ACINUS afin de moduler l'épissage des exons. PININ est connu également pour participer à la régulation de l'épissage alternatif des gènes ESRP1, PAX6, FOXJ3, ARHGEF11 et SLC37A2 dans les cellules épithéliales de cornées humaines (HCET) afin de maintenir ou améliorer les caractéristiques épithéliales (Joo et al., 2013, 2014). PININ est aussi un modulateur de l'épissage alternatif de Bcl-x, induisant la forme pro-apoptotique Bcl-x<sub>s</sub> (Leu et al., 2012).

#### 4. Fonction conservée de l'EJC : Répression du mécanisme de ré-épissage

Les premières études ont décrit principalement la survenue de sauts d'exon et de rétentions d'introns en condition de KD des facteurs de l'EJC. Les données récentes basées sur de nouvelles expérimentations ou la réanalyse de données transcriptomiques montrent qu'une fonction majeure de l'EJC, conservée de la drosophile à l'homme, est d'inhiber l'utilisation de sites cryptiques dans les introns et de bloquer le **ré-épissage** dans les transcrits déjà épissés (Blazquez et al., 2018; Boehm et al., 2018; Gehring and Roignant, 2021; Joseph and Lai, 2021).

Les sites canoniques d'épissage 5'(AG|GURAGU) et 3'(YAG|GU) contiennent des nucléotides susceptibles de reconstituer de nouveaux sites d'épissage après l'élimination de l'intron. Les début et fin d'exons sont enrichis en nucléotides compatibles avec la formation d'un nouveau site 5' d'épissage. Le dinucléotide GT est le dinucléotide le plus fréquent au début des exons (Sibley et al., 2016) et il est estimé que 16,2% des transcrits humains contiennent au moins un exon qui peut reconstituer un 5'ss à la jonction exon-exon (Blazquez et al., 2018). La déplétion des protéines du cœur de l'EJC, de RNPS1 ou de PININ conduit à l'utilisation de ces sites reconstitués (Leung and Johnson, 2018) avec une implication plus fréquente de RNPS1 et du complexe PSAP au niveau des sites 5' d'épissage et de l'EJC au niveau des sites 3' d'épissage (Figure 27).



Boehm et al., 2018

Figure 27. Régulation des sites cryptiques d'épissage par l'EJC associé au complexe PSAP.

Au cours de l'épissage, l'EJC déposé en amont des jonctions exon-exon ainsi que les complexes PSAP associés empêchent l'utilisation de sites cryptiques 3' ou 5' d'épissage dans les transcrits déjà épissés, préservant leur intégrité.

Le dépôt de l'EJC bloque l'épissage des sites 3' cryptiques présents à proximité de son dépôt dans l'exon ou bien reconstitués lors de l'épissage entre une séquence polypyrimidique de l'exon en amont et un site 3' AG présent au début de l'exon en aval.

Le recrutement du complexe PSAP par l'EJC réprime quant à lui les sites cryptiques 5' situés en aval du dépôt de l'EJC dans le transcrit déjà épissé, généralement dans les 40 premiers nucléotides ou directement au début d'un exon. En absence de l'EJC, leur utilisation entraîne l'élimination de la séquence exonique en aval du site cryptique qui apparaît alors comme un saut d'exon ou à un épissage 5' alternatif conduisant à l'inclusion de micro-exons dans le

transcrit mature. Une étude par xRIPiT-seq a montré une densité plus élevée de la protéine RNPS1 liée à l'EJC sur les exons en amont d'exons soumis fortement à un épissage récursif (ou ré-épissage) (Patton et al., 2020).

Il n'est pas possible sans investigation complémentaire de distinguer si l'exclusion d'un exon dans le transcrit mature est dû à un épissage alternatif (cassette exon) ou à un ré-épissage à un site présent en début d'exon (Joseph and Lai, 2021). Les évènements de ré-épissage produisant des transcrits aberrants sont particulièrement fréquents dans les cellules cancéreuses, qui ont pu être attribués dans le cas de la lignée de cancer du sein MCF-7 à une baisse d'expression de Y14 (Otani et al., 2021). La relaxation de la structure des mRNP due à un défaut d'expression d'un ou plusieurs composants de l'EJC pourrait faciliter la survenue d'évènements de ré-épissage en permettant le réassemblage du spliceosome sur des régions rendues accessibles du transcrit déjà épissé.

Blazquez et al. (2018) a également décrit le rôle de l'EJC dans le contrôle de l'épissage des nombreux exons RS (cryptiques ou constitutifs) présents dans le transcriptome. Dans ce cas, l'EJC bloque la reconnaissance par le spliceosome des sites 5' d'épissage d'exons RS constitutifs pour favoriser au contraire l'inclusion de centaines d'exons RS authentiques qui participent au contrôle de l'expression du transcriptome chez les mammifères, notamment dans le cerveau (Figure 28).

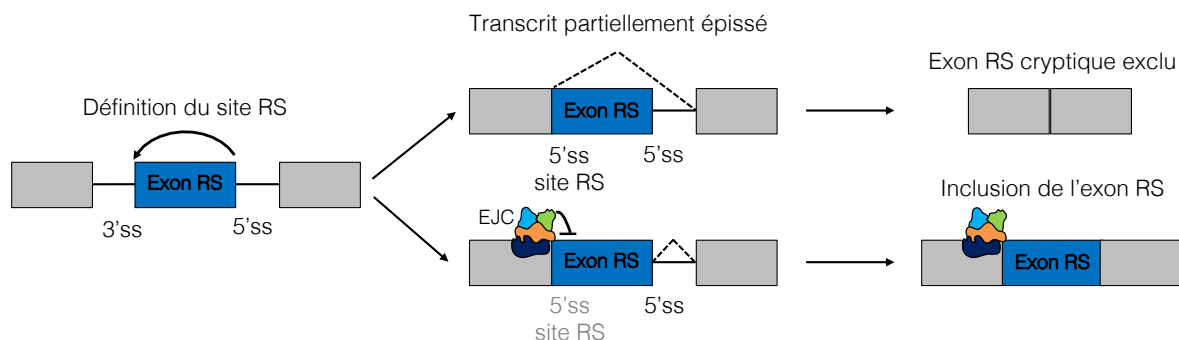


Figure 28. Régulation de l'épissage des exons RS par l'EJC.

La première étape de l'épissage liant l'exon RS (en bleu) à l'exon en amont recrée un site 5' RS au début de l'exon RS. L'EJC est déposé à la jonction exon-exon. L'exon RS est exclu si le site 5' RS est utilisé en l'absence de l'EJC ou inclus si le site 5' canonique est préférentiellement utilisé en présence de l'EJC qui bloque le site 5'ss RS lors la deuxième étape d'épissage.

Les données actuelles indiquent que l'EJC joue un rôle crucial dans l'ordre temporel des évènements d'épissage. Une fonction ancestrale de la position conservée du dépôt de l'EJC serait d'assurer un contrôle qualité de l'épissage en empêchant l'activation accidentelle de sites cryptiques ou de sites d'épissage reconstitués dans le transcrit mature (Joseph and Lai, 2021). Cette fonction doit cependant être adaptée dans la cellule pour préserver l'inclusion de micro-

exons impliqués dans le développement et les fonctions du système nerveux central (Blazquez et al., 2018; Boehm et al., 2018; Gonatopoulos-Pournatzis et al., 2018; Leung and Johnson, 2018).

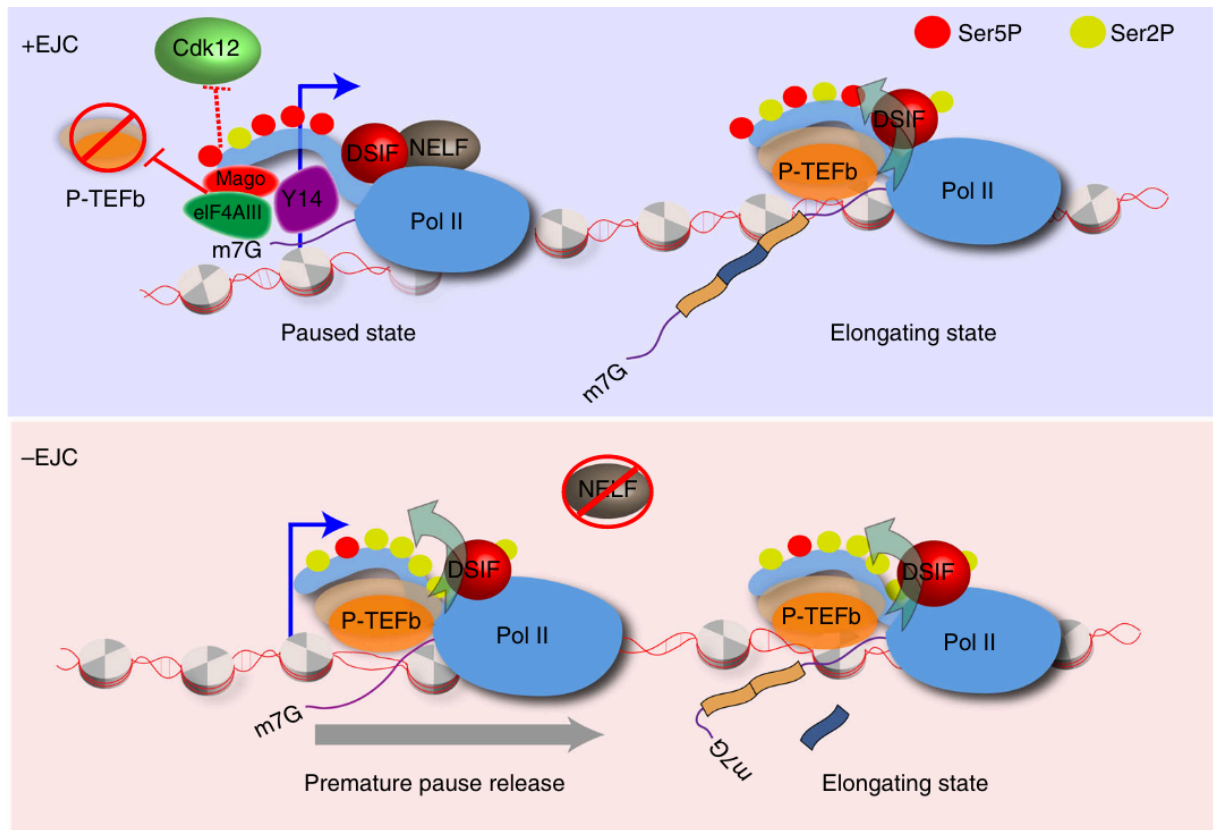
## 5. Rôle de l'EJC dans l'épissage co-transcriptionnel

Bien que le rôle de l'EJC dans le noyau soit essentiellement relié à son recrutement au sein du spliceosome et au processus d'épissage, l'hypothèse d'une possible connexion avec la machinerie de transcription et l'ARN Pol II a été confortée par des données récentes (Schlautmann and Gehring, 2020). Chez la drosophile, les facteurs de l'EJC ont été détectés sur les transcrits naissants qu'ils contiennent ou non des introns et indépendamment de CWC22 pour la formation du complexe (Choudhury et al., 2016). La protéine eIF4A3 comme le dimère Y14/MAGOH peuvent être présents isolément.

En 2019, Akhtar et al. ont montré chez la drosophile et dans les cellules HeLa que la liaison d'un pré-EJC (eIF4A3, Y14/MAGOH) non canonique sur les ARNm naissants dans les régions promotrices contribuait à stabiliser la Pol II dans son état de pause en limitant la liaison du facteur activateur d'élongation de la transcription (pTEFb). La pause proximale de la Pol II après le démarrage de la transcription est un mécanisme très conservé essentiel au contrôle et à la synchronisation de l'expression des gènes, notamment au cours du développement.

La déplétion des facteurs du pré-EJC favorise la fixation du complexe pTEFb à la Pol II et induit sa libération prématurée du promoteur en phosphorylant le facteur répresseur d'élongation (NELF) et la Ser2 de la Pol II (Figure 29). Cela s'accompagne par la survenue d'un grand nombre de défauts d'épissage en particulier dans les transcrits contenant de grands introns, démontrant l'importance de cette étape de régulation dans le contrôle de l'épissage des pré-ARNs. En accord avec ces nouvelles données, des études précédentes avaient rapporté que l'inhibition de l'élongation de la Pol II et de sa phosphorylation augmentait l'inclusion de cassette exons dont l'épissage est dépendant de l'EJC (Lu et al., 2017; Wang et al., 2014).

Ainsi, l'EJC joue un rôle direct au cours de la transcription sur la définition des exons lors de l'étape ultérieure de l'épissage en évitant une reprise prématurée de l'élongation et en stabilisant la pause de la Pol II. Cela favorise le recrutement sur le domaine CTD de la Pol II des facteurs auxiliaires nécessaires à la bonne réalisation de l'épissage, en particulier dans les grands gènes (Akhtar et al., 2019).

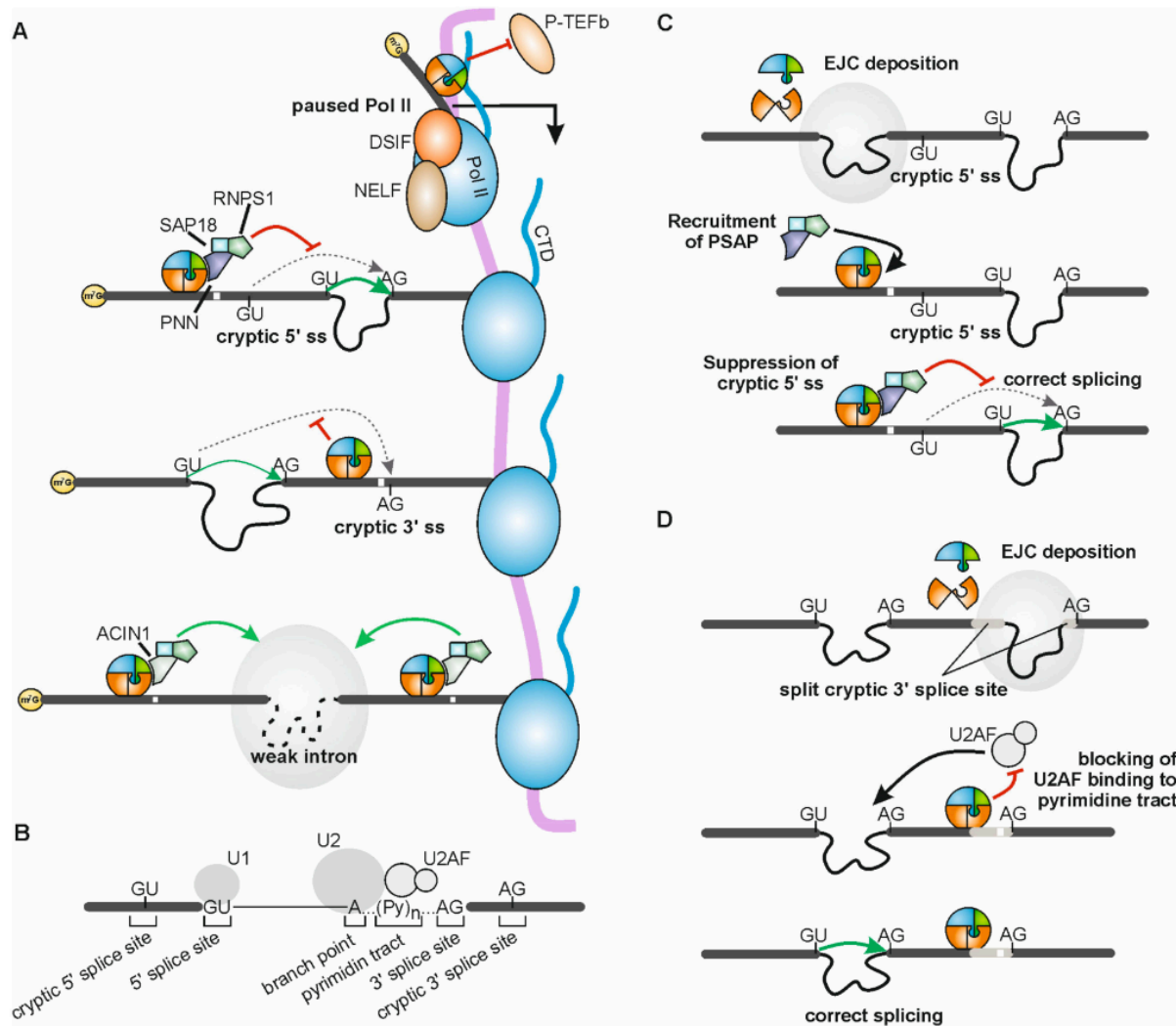


Akhtar et al., 2019

Figure 29. Rôle de l'EJC dans l'épissage co-transcriptionnel.

Le pré-EJC (eIF4A3, Y14/MAGOH) stabilise la pause proximale de la Pol II en limitant la liaison de pTEFb au promoteur et en séquestrant CDK12. Cette pause proximale est requise pour la reconnaissance correcte des exons. La déplétion des facteurs du pré-EJC favorise la fixation du complexe pTEFb à la Pol II et induit une reprise prématurée de l'élongation, induisant des défauts d'épissage (sauts d'exons) dans les transcrits.

Bien qu'il ne soit pas encore bien compris, un nouveau mécanisme se nommant « activation des débuts de transcription médiée par les exons » (EMATS) impliquant l'EJC a été décrit récemment où l'épissage des exons internes influencerait le choix des promoteurs pour un début de transcription alternatif et la production de nouvelles isoformes (Fiszbein et al., 2019).



Schlautmann and Gehring, 2020

Figure 30. Schéma global de la régulation de l'épissage par l'EJC.

(A) Les EJC non canoniques peuvent se lier à l'ARNm au niveau des promoteurs et interagir avec le domaine C-terminal (CTD) de la Pol II. Cela empêche la liaison du facteur d'élongation de la transcription pTEFb et maintient ainsi la Pol II dans un état de pause, garantissant un épissage correct. De plus chez la drosophile, les EJC déposés sur les sites canoniques améliorent l'épissage des introns faibles situés à proximité, potentiellement par l'intermédiaire du complexe ASAP. L'EJC réprime l'utilisation des sites d'épissage 3' et RNPS1 en tant que composant PSAP réprime les sites d'épissage 5' cryptiques situés en aval de l'EJC. (B) Représentation schématique des séquences introniques impliquées dans l'épissage. Les sites d'épissage cryptiques 5' et 3' dans les exons ressemblent aux sites d'épissage authentiques, mais sont situés dans les exons. (C) Modèle de la répression du site d'épissage cryptique 5' par l'EJC et le complexe PSAP. (D) Modèle de la répression du site d'épissage cryptique 3' par l'EJC. L'épissage de l'intron en aval reconstitue un site d'épissage 3' à cheval entre les deux exons. Pendant l'épissage, l'EJC est déposé à la jonction exon-exon, masquant le tract de polypyrimidines et empêchant la fixation du facteur U2AF.

## B. Rôle de l'EJC dans l'export des ARNm

L'EJC accompagne les transcrits épissés dans le cytoplasme et facilite leur export en interagissant avec le complexe de TRanscription/EXportation (TREX). Ce complexe est composé d'un sous-complexe THO, de l'ARN hélicase UAP56 (*U2AF65-associated protein 56* ou DDX39B), des facteurs d'exportation Aly/REF et d'un co-adaptateur CHTOP (Schlauthmann and Gehring, 2020). D'autres protéines sont présentes dans le complexe TREX



comme CBP80, CBP20 du *Cap Binding Complex* (CBC), ou encore NCBP3 qui interagit aussi avec le CBC et l'EJC (Heath et al., 2016; Rambout and Maquat, 2021). Aly/REF est recrutée dans un premier temps par le CBC avant d'être intégrée à l'EJC en interagissant avec eIF4A3 *via* une courte séquence peptidique (WxHD) qui est homologue à la séquence de liaison de MLN51 à eIF4A3. Ceci est confirmé par l'observation que Aly/REF se lie à l'ARNm à des positions directement en amont des sites de liaison de l'EJC (Viphakone et al., 2019). Lorsque tous les facteurs d'exportation sont déposés, REF interagit avec l'hétérodimère NXF1/NXT1, qui établit à son tour des contacts spécifiques avec les composants du complexe du pore nucléaire, activant ainsi l'exportation des ARNm polyadénylés (Le Hir et al., 2001).

## C. Rôle cytoplasmique de l'EJC

### 1. Rôle de l'EJC dans la traduction

Comme discuté auparavant, les ARNm épissés sont exprimés à des niveaux plus élevés et produisent plus de protéines que les ARNm sans intron qui ne subissent pas d'épissage. Le rôle de l'EJC dans l'augmentation de la traduction des transcrits serait de favoriser l'assemblage de la machinerie d'initiation de la traduction sur les ARNm nouvellement synthétisés et exportés dans le cytoplasme plutôt que sur les transcrits déjà traduits (Schlautmann and Gehring, 2020; Woodward et al., 2017).

Différents modèles ont été proposés mais les mécanismes mis en jeu restent encore mal connus. Un premier mécanisme implique la protéine MLN51. Elle est décrite comme un activateur de la traduction grâce à sa capacité à interagir avec les composants du complexe d'initiation de la traduction eIF3 (composé de eIF3A et eIF3D) *via* son domaine SELOR (Chazal et al., 2013). Le complexe eIF3 relie les facteurs d'initiation eIF4G et eIF4B liés à l'ARNm au complexe de pré-initiation (PIC). MLN51 liée à EJC pourrait recruter eIF3 dès l'entrée des ARNm dans le cytoplasme afin d'en augmenter la traduction.

La protéine SKAR (*S6K1 Aly/REF-like target*), un composant périphérique de l'EJC, est un composant essentiel de la régulation de l'initiation de la traduction dépendante de mTORC1 et peut également interagir avec la kinase S6K1. Lors de la synthèse protéique, le complexe mTORC1 est recruté dans le complexe eIF3 et PIC où il phosphoryle S6K1, ce qui entraîne la libération d'eIF3. Une fois activée, S6K1 est recrutée par SKAR sur les mRNP. Ceci la place à proximité des protéines ribosomales RPS6, eIF4B pour leur phosphorylation.

On ne sait pas si les activations de la traduction dépendant de MLN51 ou de SKAR sont indépendants ou si ces deux systèmes peuvent coopérer.

Le dernier mécanisme, plus hypothétique, suggère que la protéine PYM1 pourrait jouer un rôle en établissant un lien entre l'EJC et le complexe de pré-initiation 48S. Néanmoins, cette hypothèse reste encore à démontrer compte tenu que PYM1 ne se lie pas à l'EJC complet mais seulement au dimère Y14/MAGOH, qui dans cette configuration ne peut pas interagir avec eIF4A3 en raison d'un encombrement stérique (Bono et al., 2004; Gehring et al., 2009a).

## 2. Rôle de l'EJC dans la surveillance des ARNm

### a. Rôles du NMD

Un contrôle qualité post-transcriptionnel des ARNm matures qui arrivent dans le cytoplasme est obligatoire pour la fabrication de protéines fonctionnelles au sein des organismes. Le rôle le mieux décrit de l'EJC dans la littérature est sa fonction dans le recrutement du système NMD (Woodward et al., 2017). Le système NMD est un mécanisme de contrôle qualité chez les eucaryotes qui détecte et dégrade les ARNm porteurs de PTC qui résultent de mutations ou d'erreurs au cours de l'expression (transcription, épissage) des gènes. Le NMD protège ainsi la cellule de la présence de protéines tronquées non fonctionnelles ou potentiellement toxiques (Lejeune, 2022). En dehors de cette fonction, il a été identifié plus récemment que le NMD régule aussi finement l'expression normale des gènes par dégradation de transcrits physiologiques dits cibles du NMD endogène, porteurs de PTC, qui représentent 10% du transcriptome chez les mammifères.

Le NMD possède également un rôle physiologique important dans les processus de différenciation dans le SNC mais aussi dans le muscle où le NMD coopère avec une autre voie de dégradation pour contrôler la différenciation musculaire (voir partie 4).

Les trois protéines UPF (*up-frameshift*) conservées UPF1, UPF2 et UPF3 et les facteurs SMG1, 5, 6, 7, 8 et 9 sont au cœur de cette voie de régulation des substrats par le NMD (Yi et al., 2021).

### b. Substrats du NMD

Différents mécanismes sont responsables de l'introduction d'un PTC dans le transcrit mature (Yi et al., 2022) (Figure XX) : i) la survenue de mutations non-sens ou décalant le cadre de lecture (*frameshift*) dans les gènes (Hoek et al., 2019) ; ii) la présence de 3'UTR longs. De nombreux transcrits endogènes avec des codons de terminaison normaux mais des 3'UTR plus longs que la moyenne sont également reconnus par la machinerie NMD; iii) des épissages alternatifs comme des inclusions ou sauts d'exons, des rétentions d'intron internes ou l'utilisation de sites cryptiques 5' ou 3' (McGlinicy and Smith, 2008) ; iv) la présence d'introns dans le 3'UTR où le codon stop terminal naturel est considéré comme un PTC. Les jonctions

exon-exon 3'UTR, peuvent fonctionner en augmentant la probabilité qu'un événement de terminaison soit reconnu comme aberrant lors de chaque tour de traduction (Bicknell et al., 2012) et v) les uORF en amont des ORF normales, présentes dans 40% du transcriptome dans les 5'UTR des ARNm, qui permettent un début de traduction alternative où un PTC présent donne des transcrits cibles du NMD en fonction des besoins cellulaires (Lykke-Andersen and Jensen, 2015).

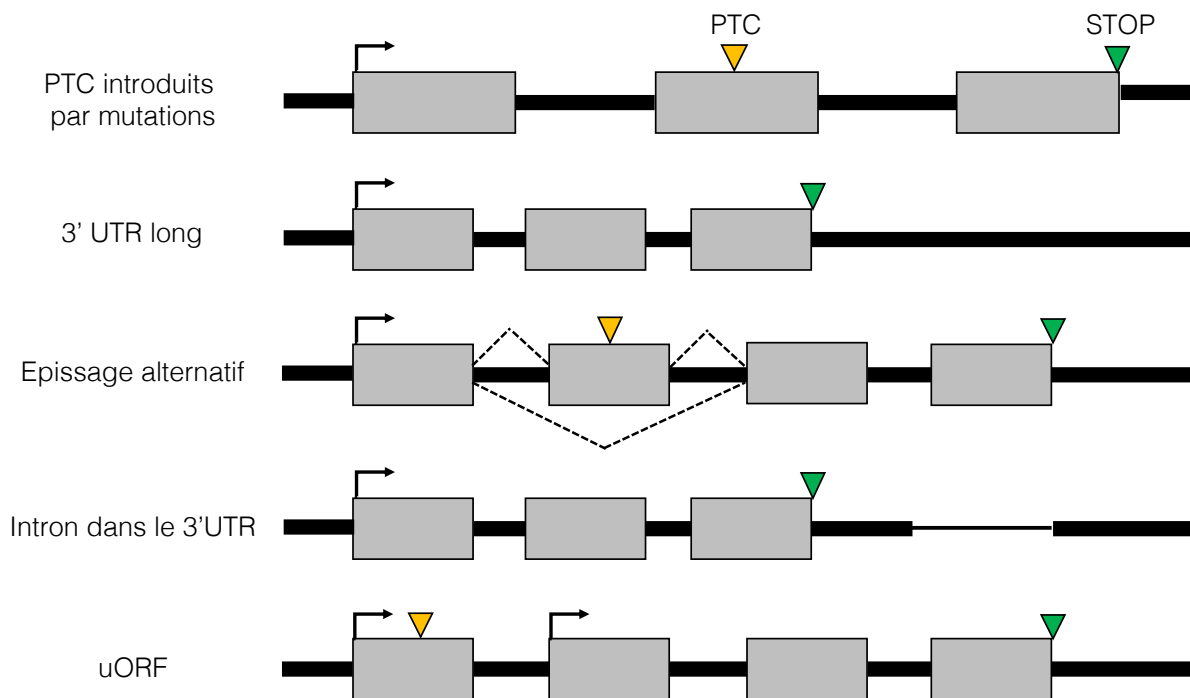


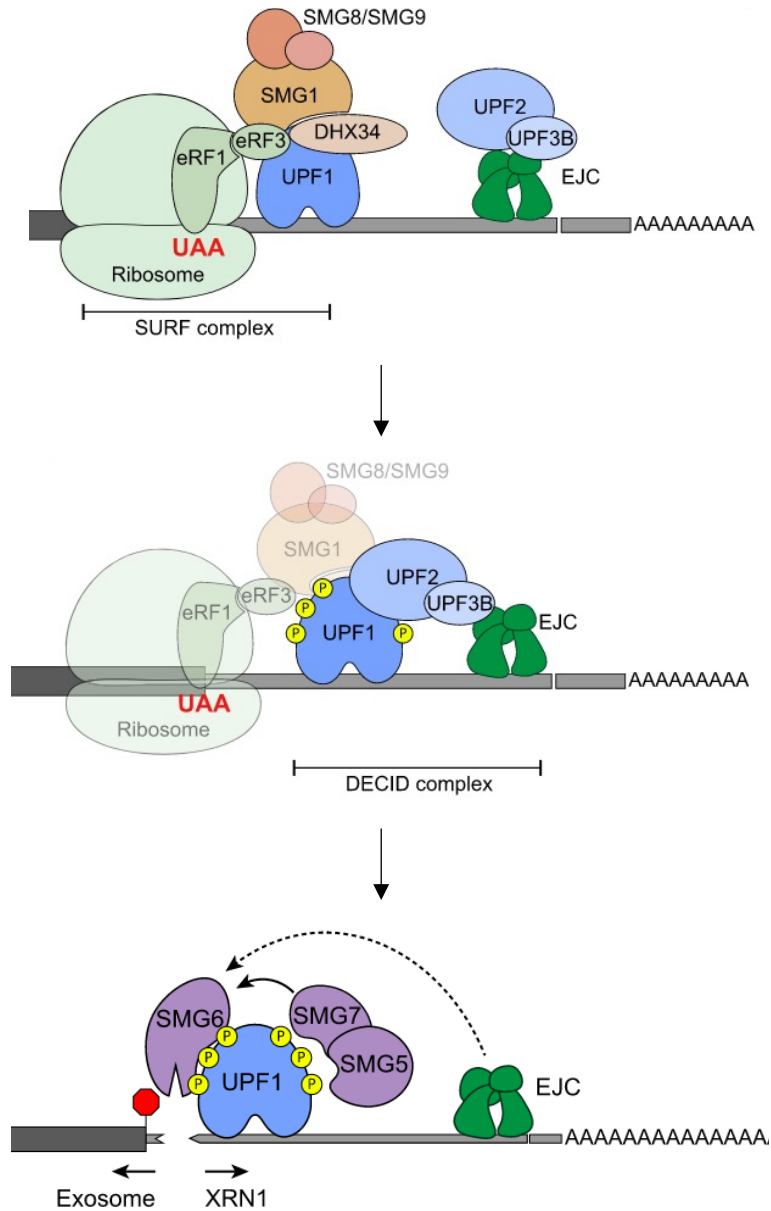
Figure 31. Substrats activant le système NMD. Différents types d'événements génèrent des transcrits qui deviennent cibles du NMD en fonction de la localisation du PTC (triangle orange) ou du codon stop naturel (triangle vert).

### c. Recrutement du système NMD

Deux mécanismes principaux du NMD co-existent dans la cellule (Lejeune, 2022). Le premier est la voie du NMD dit « classique » recruté par l'EJC. L'EJC se positionne 20-24nt en amont des jonctions exon-exon des ARNm en tant que marque du bon déroulement de l'épissage. Chez l'homme, les ribosomes effectue le contrôle qualité du transcrit durant le premier cycle de traduction par balayage à partir de la coiffe en 5' jusqu'à la queue polyA en 3' (Rufener and Mühlemann, 2013; Yi et al., 2021). En l'absence de PTC, tous les EJC sont éliminés de l'ARNm pour être recyclés. Si le ribosome rencontre un PTC à une distance d'au moins 50 nt en amont de la dernière jonction exon-exon, celui-ci recrute les facteurs UPF qui interagissent avec le ribosome afin d'activer le système NMD (Lejeune, 2022; Lindeboom et al., 2016; Mailliot et al., 2022; Yi et al., 2021). Il est possible que des PTC se situant à une distance inférieure à 50

nt en amont de la dernière jonction d'exon puissent activer le NMD par des EJC se situant dans les 3'UTR, comme cela a été montré chez le poisson zèbre (Gangras et al., 2020).

L'EJC va d'abord se lier à UPF3B qui permet de recruter UPF2 à proximité du ribosome. De son côté, lors du premier cycle de traduction, l'hélicase UPF1 se lie à l'ARN et forme le complexe SURF en interaction avec SMG1, SMG8/9, le ribosome, les facteurs de libération eRF1/3 et DHX34 (Yamashita et al., 2009). Cette interaction facilite le recrutement de UPF1 en aval avec les autres protéines UPF2 et UPF3B associées à l'EJC. SMG1 active ensuite UPF1 par phosphorylation (Kashima et al., 2006), ce qui libère les facteurs SMG1, SMG8/9 et les eRF1/3. UPF1 phosphorylé va se lier à UPF2, UPF3B et l'EJC pour former le complexe DECID (*decayinducing*) afin de procéder à la dégradation de l'ARNm. Cette fonction est assurée par les trois protéines SMG5, SMG6 et SMG7 (Figure 32). Chacune de ces protéines peut se lier à la protéine UPF1 phosphorylée (Chakrabarti et al., 2014). SMG6 clive l'ARNm en deux fragments au niveau ou immédiatement en aval du PTC par sa fonction d'endonucléase grâce à son domaine PIN (Lykke-Andersen et al., 2014). L'hétérodimère SMG5/SMG7 est recruté par UPF1 et une interaction directe entre SMG7 et CNOT8 connecte l'ARNm cible au complexe de dégradation CCR4/NOT (Loh et al., 2013). Le recrutement de CCR4/NOT vers les ARNm ciblés par le NMD effectue une dégradation rapide, suivie d'un décapage de l'ARNm et de sa dégradation par XRN1, une exonucléase 5' → 3'. Il est intéressant de noter que les ARNm cibles du NMD contenant des PTC semblent être plus sensibles aux niveaux d'expression de SMG6 (Ottens et al., 2017). L'ajout d'EJC sur un long 3'-UTR conduit à un clivage plus important des ARNm médié par SMG6 (Boehm et al., 2014). La présence d'EJC en aval confère une sensibilité accrue à la dégradation par la voie SMG6. Il est montré par analyses de molécules d'ARNm isolées que l'efficacité du NMD par les facteurs de clivage est aussi influencée par la distance entre le PTC et l'EJC. Plus la distance est longue, plus le clivage de l'ARNm est retardé. L'augmentation du nombre d'EJC présents en aval du PTC accentue le clivage des ARNm et les séquences autour de l'EJC en aval du PTC influencent la reconnaissance des facteurs de clivage (Hoek et al., 2019).



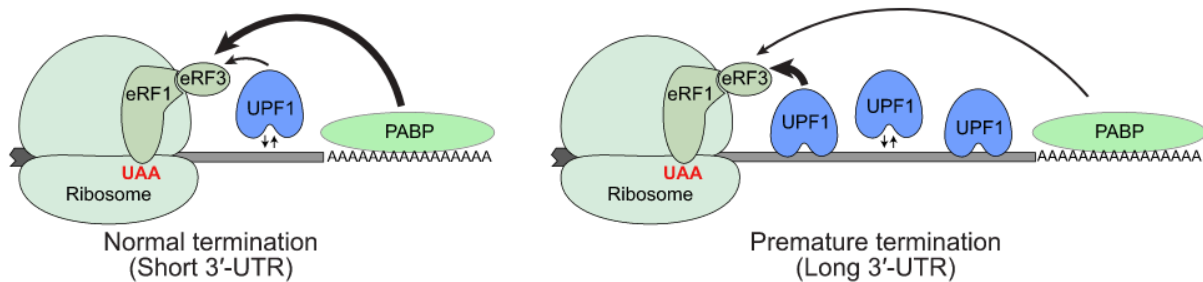
Adaptée de Yi et al., 2021

Figure 32. Recrutement du système NMD "classique" par l'EJC.

En présence d'un PTC détecté par le ribosome, UPF1, ainsi que eRF1 et eRF3, interagissent avec SMG1, DHX34 et SMG8/9 pour former le complexe SURF. L'EJC recrutant UPF2/UPF3B, facilite la formation du complexe UPF1/2/3B, et donc la formation du complexe DECID, conduisant à la phosphorylation de UPF1 et à l'activation de NMD. L'endonucléase SMG6 recrutée sur l'UPF1 phosphorylée agit de manière SMG5/SMG7-dépendante pour cliver l'ARNm à proximité du PTC. L'action de SMG6 pourrait être renforcée par la présence de l'EJC. Après le clivage des ARNm, les fragments 5' et 3' sont dégradés par les exosomes et l'exonucléase XRN1 respectivement.

Le deuxième modèle d'activation du NMD, indépendant de l'EJC, est lié à la distance entre le PTC et les protéines PABP fixées sur la queue polyA (Lejeune, 2022; Singh et al., 2008; Yepiskoposyan et al., 2011; Yi et al., 2021). Plus cette distance est grande, plus la probabilité d'activation du NMD est élevée. Dans ce modèle, les protéines UPF1 s'accumulent sur le 3'UTR et entrent en compétition avec les protéines PABP situées sur la queue polyA pour le recrutement du complexe de terminaison de la traduction. Si les PABP recrutent ce complexe,

la terminaison de la traduction n'entraîne pas l'activation du NMD et mène à l'initiation de nouveaux cycles de traduction. Au contraire, dans les longs 3'UTR, si UPF1 recrute le complexe *via* eRF3, le système NMD se met en place et dégrade l'ARNm (Ivanov et al., 2016) (Figure 33). Cependant, les transcrits possédant un long 3'UTR ne sont pas toujours la cible du NMD. Il existe des mécanismes de protection sur les séquences 3'UTR par la fixation de facteurs agissant en *trans* comme TRAM1 ou PTBP1 (Ge et al., 2016; Kishor et al., 2020; Toma et al., 2015). On peut noter aussi que des 3'UTR longs sont régulés par UPF2 (Bao et al., 2016).



Yi et al. 2021

Figure 33. Mécanisme d'activation du NMD en présence de longs 3'UTR.

Sur les 3'UTR courts (à gauche), l'interaction de PABP avec eRF3 est favorisée (flèche plus épaisse) au dépend de l'interaction UPF1-eRF3 (flèche plus fine), ce qui conduit à une terminaison de la traduction normale. De longs 3'UTR (à droite) favorisent l'accumulation d'UPF1 sur les ARN et l'interaction de UPF1 avec eRF3 au dépend de PABP située à distance, ce qui conduit à une terminaison prématurée et à l'activation du NMD.

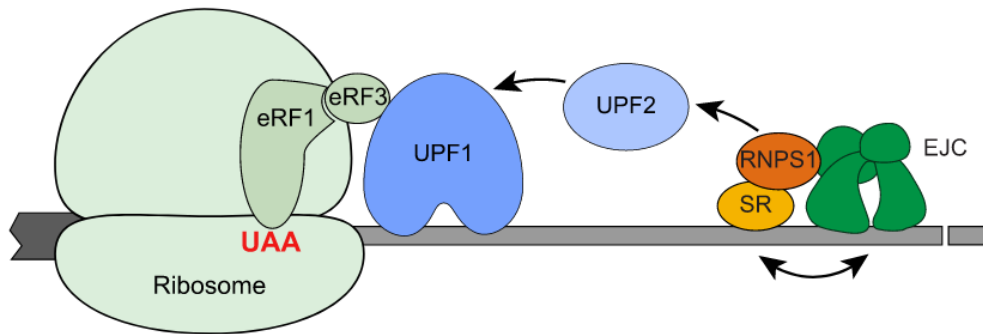
#### d. Différents types de NMD classiques recrutés par l'EJC

Le NMD dit « classique » n'est pas toujours la voie impliquée dans la dégradation des ARNm, même si celle-ci est majoritaire. Le recrutement du système NMD par l'EJC peut se faire par des mécanismes alternatifs grâce à d'autres protéines associées à EJC dans les mRNP, notamment les protéines RNPS1 et MLN51 (Mabin et al., 2018; Yi et al., 2021).

##### i. EJC-RNPS1-SR et NMD indépendant d'UPF3B

La première branche alternative du NMD, dépendant de l'EJC chez les mammifères, est le système NMD dépourvu d'UPF3B (Chan et al., 2007). Contrairement à UPF1 et UPF2, l'absence d'UPF3B n'est pas létale mais affecte la fonction et la différenciation neuronale (Alrahbeni et al., 2015; Jolly et al., 2013). Les mutations de UPF3B sont responsables d'un retard mental chez les patients qui peuvent être atteints aussi d'autisme et de déficiences intellectuelles (Laumonnier et al., 2010; Xu et al., 2013). Les ARNm cibles du NMD connues, sont régulées à la hausse chez des souris viables déplétées de UPF3B par oligonucléotides antisens (ASO) mais aussi dans des cellules de patients et des cellules HeLa dépourvus d'UPF3B (Chan et al., 2007; Huang et al., 2018). Le système NMD indépendant d'UPF3B se lie préférentiellement au ARNm liés par le complexe EJC-RNPS1-SR qui ont une interaction préférentielle pour UPF2 (Figure 34) (Mabin et al., 2018). De plus, il est montré que RNPS1

peut s'associer aux 3'UTR et engendrer une forte activité du NMD même en l'absence d'UPF3B soit indirectement en favorisant le dépôt stable de l'EJC sur l'ARNm soit plus directement par des interactions des protéines EJC et UPF (Gehring et al., 2005). Il reste à savoir si les ARNm enrichis avec des protéines SR sont des cibles privilégiées du NMD dépendant de l'EJC associé à RNPS1. Il semble donc que le NMD indépendant d'UPF3B pourrait jouer un rôle important chez les mammifères, tandis que le NMD dépendant d'UPF3B pourrait avoir une fonction spécialisée.



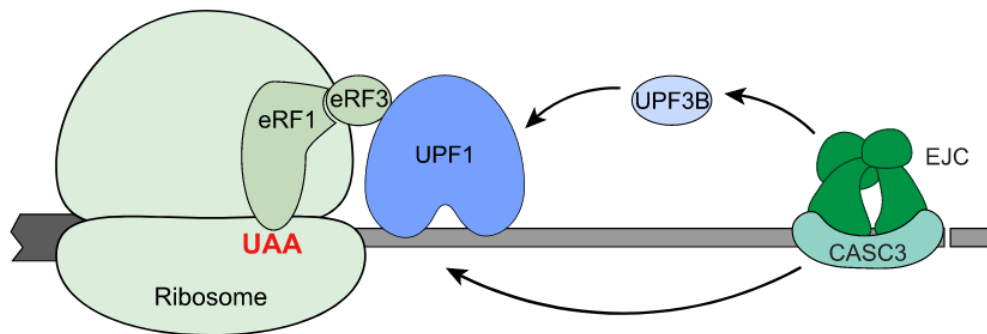
Yi et al., 2021

Figure 34. Système NMD indépendant d'UPF3B.

L'interaction entre l'EJC contenant RNPS1 et UPF2 induit une activation du NMD dépendant de l'EJC et indépendant d'UPF3B.

#### ii. EJC-MLN51 et NMD indépendant d'UPF2

Une deuxième voie de NMD alternatif, dépendant de l'EJC, se produit en l'absence d'UPF2. Il a été décrit que des ARNm cibles du NMD sont insensibles à l'élimination de UPF2 dans les cellules HeLa. Un mutant de la protéine UPF1, empêchant la liaison à UPF2 mais qui peut encore être phosphorylée afin d'activer la dégradation des ARNm, peut remplacer efficacement l'UPF1 endogène (Gehring et al., 2005; Ivanov et al., 2008). Le système NMD indépendant d'UPF2 se lie préférentiellement aux EJC lié à MLN51, lors du changement de composition des mRNP (EJC-RNPS1-SR en EJC-MLN51) (Figure 35) (Mabin et al., 2018). De plus, la liaison de UPF3B ou de MLN51 à UPF1 en aval d'un codon stop déclenche une forte augmentation du NMD dépendant d'UPF1 dans les cellules HeLa déplétées de UPF2. Les mutants UPF1 qui ne peuvent pas s'associer avec UPF2, peuvent interagir avec UPF3B et MLN51 (Gehring et al., 2009).



Yi et al., 2021

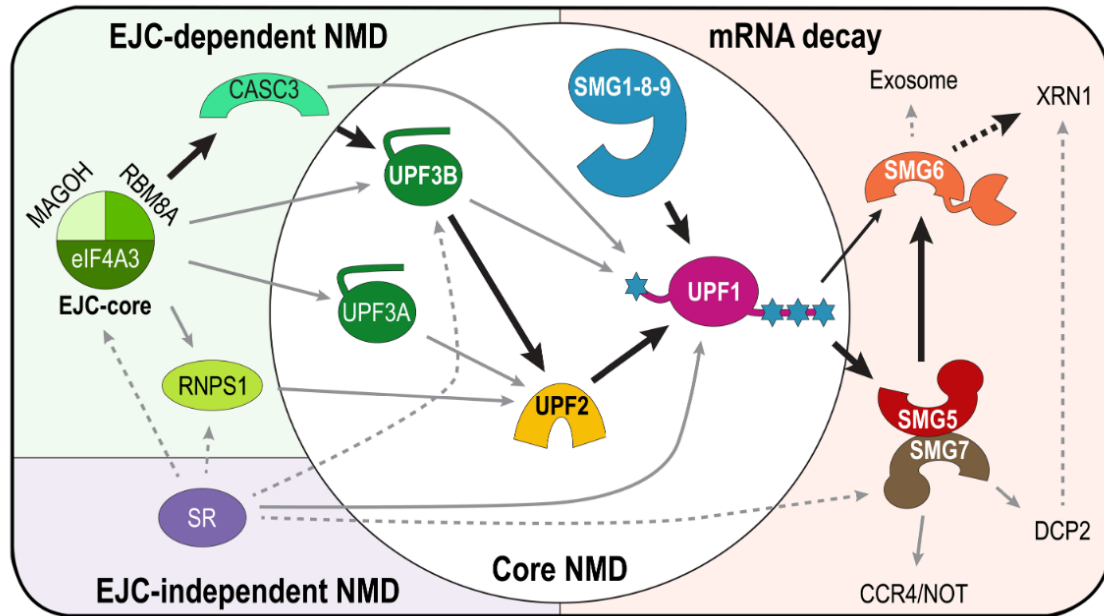
Figure 35. Système NMD indépendant d'UPF2.

UPF3B peut interagir directement avec UPF1 et déclencher le NMD. Les EJC liés à MLN51 semblent favoriser cette voie.

L'EJC-RNPS1-SR représente probablement un stade principalement nucléaire et cytoplasmique précoce de l'EJC avant ou pendant la traduction (Mabin et al., 2018) avant de passer sous la forme EJC-MLN51 plus tardive. Les cibles du NMD dépendant de l'EJC-RNPS1-SR ou de l'EJC-MLN51 sont potentiellement influencées par la vitesse du changement de composition de l'EJC dans les mRNP et par la vitesse d'entrée de l'ARNm dans la traduction. Le rapport d'expression RNPS1/MLN51 pourrait également influencer l'efficacité du NMD. Un niveau d'expression élevé de RNPS1 est lié à l'efficacité du NMD dépendant d'EJC (Viegas et al., 2007). Au contraire, les niveaux d'ARNm de MLN51 sont beaucoup plus faibles que ceux des autres facteurs de l'EJC (Zetoune et al., 2008) et la surexpression de MLN51 ne diminue que modestement l'efficacité du NMD dans les cellules HeLa (Mabin et al., 2018). Récemment, deux études sont venues renforcer l'existence des EJC et NMD alternatifs. En effet, AKT, faisant partie de la voie de signalisation des kinases PI3K/AKT/mTOR, phosphoryle directement UPF1 et permet l'activation du système NMD en remplaçant UPF2 tout en étant associé préférentiellement à des EJC liés à MLN51 et UPF3B (Cho et al., 2022; Palma et al., 2021).

Il est à noter qu'il existe un paralogue de UPF3B, UPF3A. Celui-ci présente une affinité de liaison à l'EJC plus faible. UPF3A peut activer de manière moins importante le NMD en l'absence de UPF3B mais pourrait servir de suppresseur de l'activation du NMD en séquestrant UPF2 loin de UPF3B. UPF3B assure la principale fonction d'activation du NMD, tandis que UPF3A pourrait agir comme un faible activateur de la NMD ou comme un suppresseur (Chan et al., 2009; Shum et al., 2016; Yi et al., 2021).





Mailliot et al., 2022

Figure 36. Mécanismes de régulation du système NMD.

La voie du NMD classique nécessitant UPF1, UPF2 et UPF3B est conservée de la levure à l'homme. UPF3A est un activateur du NMD lors d'un défaut en UPF3B. Dans la voie du NMD indépendante de UPF3B, RNPS1 stimule le NMD en recrutant UPF2. Le NMD indépendant de UPF2 est renforcé par MLN51 (CASC3), qui se lie à UPF3B et UPF1. UPF2 active le complexe kinase SMG1-8-9 pour phosphoryler UPF1. La dégradation de l'ARNm repose sur le recrutement de SMG6 et SMG5-7 par UPF1 phosphorylé. SMG5-7 interagit avec CCR4/NOT responsable de la dégradation de l'ARNm, DCP2 (décapage de l'ARNm), et active l'endonucléase SMG6. Les fragments d'ARNm non protégés qui en résultent sont dégradés par l'exonucléase XRN1 et le complexe exosome. Les protéines SR (en particulier SRSF1) renforcent le NMD en interagissant avec l'EJC et ses protéines associées ainsi qu'avec les facteurs du NMD. Les protéines SR peuvent également favoriser la dégradation des ARNm d'une manière indépendante de l'EJC.

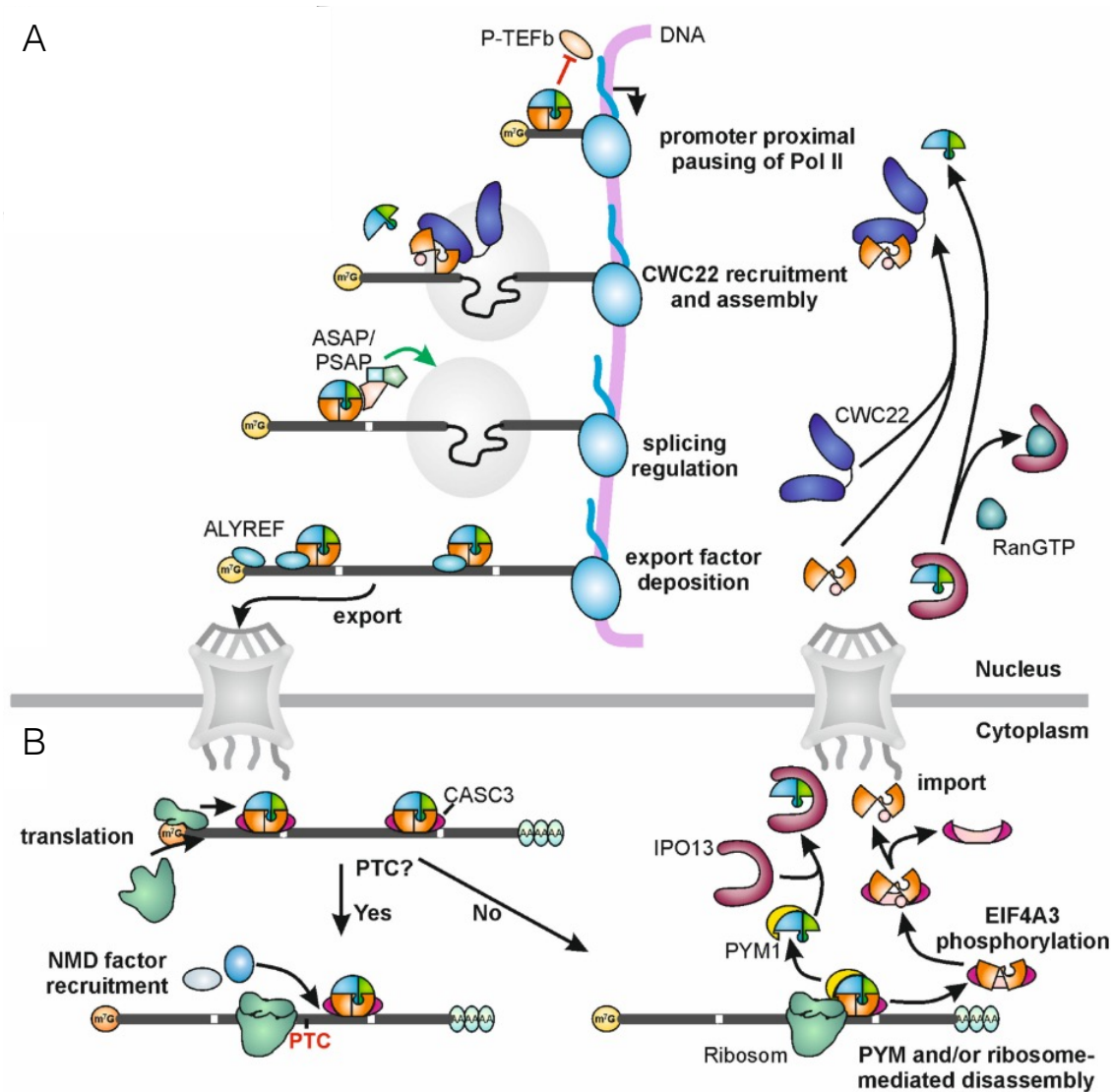
## V. Désassemblage et recyclage de l'EJC

Lorsque l'EJC a accompli ses différentes fonctions du noyau jusqu'au cytoplasme, il doit finalement être désassemblé de l'ARNm afin que chacun de ses composants puisse être recyclé et réimporté dans le noyau pour former de nouveaux complexes EJC qui seront déposés sur les ARNs naissants (Schlautmann and Gehring, 2020). Cette étape repose sur un processus de dissociation active.

Selon un premier modèle, les EJC sont détachés lors du premier cycle de traduction par la processivité des ribosomes qui possèdent une activité hélicase de l'ARNm (Dostie and Dreyfuss, 2002; Gerbracht and Gehring, 2018; Takyar et al., 2005). L'hydrolyse de l'ATP en ADP+Pi permet le changement de conformation d'eIF4A3 qui s'ouvre et conduit au désassemblage de l'EJC.

Une deuxième voie implique la protéine PYM1, qui peut agir soit en étant liée au ribosome soit indépendamment de la traduction, ce qui permet d'atteindre les EJC situés dans les régions non traduites (Gehring et al., 2009b). PYM1 se lie avec l'hétérodimère Y14/MAGOH et le dissocie de l'EJC. En absence de l'hétérodimère, eIF4A3 retrouve une conformation ouverte et est

libérée de l'ARNm. Elle est alors phosphorylée par les protéines kinases 1 et 2 dépendantes de la cycline (CDK1/2) d'une manière dépendante du cycle cellulaire (Ryu et al., 2019). Elle se détache de MLN51 et est réimportée dans le noyau où elle va être guidée vers de nouveaux spliceosomes actifs et se lier à CWC22. Le dimère Y14/MAGOH est importé dans le noyau par l'importine 13 (IPO13) afin d'être intégré également dans de nouveaux complexes EJC (Bono et al., 2010).



Adaptée de Schlautmann and Gehring 2020

Figure 37. Rôles et cycle de vie de l'EJC.

(A) Les EJC non canoniques améliorent la pause proximale du promoteur de la Pol II en empêchant l'association de la Pol II avec pTEFb. Les EJC canoniques sont recrutés par le spliceosome via l'interaction entre CWC22 et eIF4A3 et sont déposés 24nt en amont d'une jonction exon-exon. Ces EJC peuvent influencer l'épissage avec les complexes ASAP et PSAP et stimuler le dépôt de facteurs d'exportation ALYREF sur l'ARNm. (B) Avant, pendant ou après le transport des ARNm associés aux EJC vers le cytoplasme, les EJC subissent un changement de composition et sont retirés de l'ARNm pendant la traduction par les ribosomes liés à PYM1 ou par PYM1 seul. La présence d'un PTC situé suffisamment en amont d'un EJC conduit à l'activation du NMD en recrutant des facteurs UPF. Les EJC qui sont éliminés par PYM1 sont désassemblés. EIF4A3, qui est dans sa conformation ouverte, peut être phosphorylé par CDK1/2 et réimporté dans le noyau par un mécanisme inconnu. L'hétérodimère Y14-MAGOH est importé dans le noyau par IPO13. MLN51 peut faire la navette entre les deux compartiments.

## VI. Rôle physiologique de l'EJC en pathologies

Chez la souris, il a été démontré que l'haplo-insuffisance en MAGOH, Y14 ou eIF4A3 est suffisante pour provoquer une microcéphalie, ce qui indique que ces composants sont nécessaires à la neurogenèse (Mao et al., 2016).

Chez l'homme, les mutations des composants de l'EJC sont responsables de différentes pathologies, notamment des défauts du développement neuronal (Asthana et al., 2022).

Les variations du nombre de copies (duplications ou micro-délétions) au locus du gène Y14 (1q21.1) induisent le syndrome TAR (*thrombocytopenia absent radius*), des troubles du développement neurologique (troubles moteurs et cognitifs), des caractéristiques dysmorphiques comme des microcéphalies et des cardiopathies congénitales (Bernier et al., 2016). Ces patients présentent souvent un large éventail d'anomalies du squelette (Boussion et al., 2020). Il existe des patients présentant un syndrome TAR sans micro-délétion du gène Y14. De plus, Y14 est impliqué dans le développement des organes sexuels féminins. Des expériences de ARNi entraînent la létalité embryonnaire et la masculinisation de la lignée germinale hermaphrodite chez *C. elegans* (Kawano et al., 2004). Des patients ayant le syndrome TAR peuvent être aussi associés au syndrome de Mayer-Rokitansky-KüsterHauser présentant une absence congénitale de l'utérus et de la partie supérieure du vagin. Comme il a été démontré que les mutations d'autres facteurs d'épissage entraînent une masculinisation, Y14 est probablement impliqué dans la différenciation sexuelle en régulant les événements d'épissage alternatif (Tewes et al., 2015).

Le facteur eIF4A3 a également été impliqué dans des maladies génétiques, notamment le syndrome de Richieri-Costa Pereira. Une expansion non codante d'eIF4A3 identifiée chez ces patients est censée interférer avec UPF3B et provoquer un sous-développement crânio-facial et des membres (Bertola et al., 2018). Des recherches ultérieures sur les implications mécanistiques de cette expansion ont révélé que les embryons de souris haplo insuffisantes pour eIF4A3 présentent une micrognathie et une fusion mandibulaire altérées (Miller et al., 2017).

En plus des principaux facteurs EJC, la dégradation protéolytique de UPF1 par DUX4 dans la dystrophie musculaire facio-scapulo-humérale (FSHD) est associée à une forte inhibition du NMD résultant en une accumulation d'ARNs aberrants dans les cellules musculaires (Feng et al., 2015).

Récemment, MLN51 a aussi été observé comme contrôlant la croissance des synapses des motoneurone au niveau des jonctions neuro-musculaires, indépendamment de l'EJC (Ho et al., 2022).

## Partie 3 : Le gène de la Dystrophine

### I. Le gène DMD (Duchenne Muscular Dystrophy)

D'une longueur de 2.2 Mb, le gène *DMD* identifié en 1987 sur le chromosome X (Xp21.2) est le plus long gène humain (Koenig et al., 1987). Le gène contient plusieurs promoteurs régulant l'expression tissu-spécifique d'isoformes de taille variable. Le transcrit principal d'une taille de 14kb contient les 79 exons du gène (séquence codante de 11.3kb, 0,7% du gène) et code dans le muscle squelettique pour la dystrophine, une protéine filamenteuse de 427 kDa localisée sous le sarcolemme qui relie la matrice extra-cellulaire au cytosquelette de la cellule musculaire (Hoffman et al., 1987; Koenig et al., 1988).

### II. Les dystrophinopathies

#### A. Les mutations du gène DMD

Le taux de mutation est particulièrement élevé dans le gène *DMD* en raison de sa taille exceptionnelle. La proportion de néomutations dans les cas sporadiques de *DMD* (absence d'antécédents familiaux) est estimée à un tiers (Helderman-van den Enden et al., 2009). On observe une grande hétérogénéité allélique dans la DMD comme le montre le très grand nombre de mutations différentes rapportées dans les bases de données (Bladen et al., 2015; Tuffery-Giraud et al., 2009), incluant de grands réarrangements génomiques (délétions ou duplications de un ou plusieurs exons) dans environ 78% des cas et des mutations ponctuelles (mutations non-sens, petites délétion/insertion, variants d'épissage, faux sens) dans les 22% de cas restants (Fratter et al., 2020).

Les mutations du gène *DMD* sont responsables de deux formes alléliques de dystrophies musculaires de gravité différente, la dystrophie musculaire de Duchenne (DMD, OMIM #310200) et la dystrophie musculaire de Becker (BMD, OMIM #300376), transmises sur le mode récessif lié au chromosome X.

La sévérité de l'atteinte musculaire n'est pas corrélée à la taille ou au siège de la mutation mais à l'impact de la mutation sur le cadre de lecture (Monaco et al., 1988). Dans le cas de la DMD, les mutations décalent le cadre de lecture et introduisent un codon stop prématuré dans les transcrits qui seront majoritairement dégradés par le mécanisme du NMD mais également diminués par la répression de la transcription (García-Rodríguez et al., 2020). Dans la BMD en revanche, les mutations (essentiellement de grandes délétions en phase) sont compatibles avec le maintien d'un cadre ouvert de lecture permettant la production en quantité variable

d'une protéine partiellement fonctionnelle. Les exceptions (environ 10%) à cette règle dite du cadre de lecture résultent de mécanismes variés impliquant notamment une ré-initiation de la traduction ou des épissages alternatifs qui modulent la sévérité du phénotype (Fratter et al., 2020; Tuffery-Giraud et al., 2017). On peut citer les cas de sauts d'exons porteurs d'une mutation non-sens observés dans la région centrale du gène (bloc d'exons 25 à 42 en phase). Le changement nucléotidique introduit par la mutation modifie la combinaison des éléments *cis*-régulateurs dans ces exons dont la définition est sub-optimale. L'épissage alternatif (le plus souvent partiel) de l'exon muté permet l'expression de transcrits en phase qui changent le phénotype d'une forme DMD à BMD (Flanigan et al., 2011). L'étude par des approches fonctionnelles et biochimiques des exons 31 et 39 ont permis à notre équipe de démontrer que la perte de reconnaissance de l'exon était dû à la création d'un élément répresseur d'épissage (ESS) recrutant les protéines hnRNPA1, hnRNPA2 et DAZAP1 (Disset et al., 2006; Miro et al., 2015).

## B. Les formes cliniques de DMD et BMD

### 1. La DMD

La **DMD** est la forme la plus grave et la plus fréquente des maladies neuromusculaires de l'enfant, elle touche environ 1 garçon sur 5000 à la naissance dans le monde (Fortunato et al., 2021). L'absence de la dystrophine conduit à une rupture du sarcolemme, une fragilisation de la fibre musculaire et induit des cycles de dégénérescence/régénération induits par la contraction menant à une faiblesse musculaire progressive dès la petite enfance et au remplacement progressif du tissu musculaire par du tissu adipeux et fibrosé dans les stades avancés de la maladie. La perte de la marche survient avant l'âge de 12 ans (âge moyen entre 9 et 10 ans) et le décès généralement entre la deuxième et troisième décennie par insuffisance respiratoire et/ou cardiaque liée à une cardiomyopathie dilatée du ventricule gauche.

La dystrophine étant également exprimée dans les muscles lisses, la rétine, les glandes endocrines et le système nerveux central (SNC), la prise en charge de la DMD doit être multidisciplinaire (Tsuda and Fitzgerald, 2017). Des atteintes cognitives sont fréquemment rapportées. On note chez les patients DMD un décalage d'un écart-type de la moyenne de QI (moyenne = 85) par rapport à la population générale (Doorenweerd, 2020). La prévalence de la déficience intellectuelle est de 20,9% chez les patients DMD (versus 3% dans la population générale). Des problèmes d'apprentissage et de comportement peuvent survenir également chez les patients présentant ou non un décalage du QI, et sont parfois détectés très tôt dans le

développement de l'enfant constituant le premier signe clinique qui conduit au diagnostic génétique. Ainsi de nombreux rapports font état de comorbidités telles que les troubles du spectre autistique (TSA) (3-15 %), le trouble déficitaire de l'attention avec hyperactivité (TDAH) (11-32 %), les troubles obsessionnels compulsifs (TOC) (5-60 %), l'anxiété, la dépression et les troubles émotionnels (27 %) et l'épilepsie (6 %) (Doorenweerd, 2020; Naidoo and Anthony, 2020).

Les atteintes cognitives sont corrélées avec une fréquence plus élevée de mutations localisées dans la région 3' du gène qui altèrent l'expression de toutes les isoformes de dystrophine (longues et courtes) dont celles exprimées dans le cerveau (Dp140 et Dp71) (Daoud et al., 2009; Moizard et al., 2000; Ricotti et al., 2016a).

## 2. La BMD

La **BMD** est une forme clinique modérée de la maladie et de début généralement plus tardif (Fortunato et al., 2021). La distribution de la faiblesse musculaire est semblable à celle de la myopathie de Duchenne, mais l'atteinte est moins sévère et d'évolution moins rapide. La perte de la marche, si elle survient, est par définition après l'âge de 16 ans. Une atteinte intellectuelle est possible, comme dans la DMD et l'insuffisance cardiaque due à la cardiomyopathie est la principale cause de décès (Tsuda and Fitzgerald, 2017).

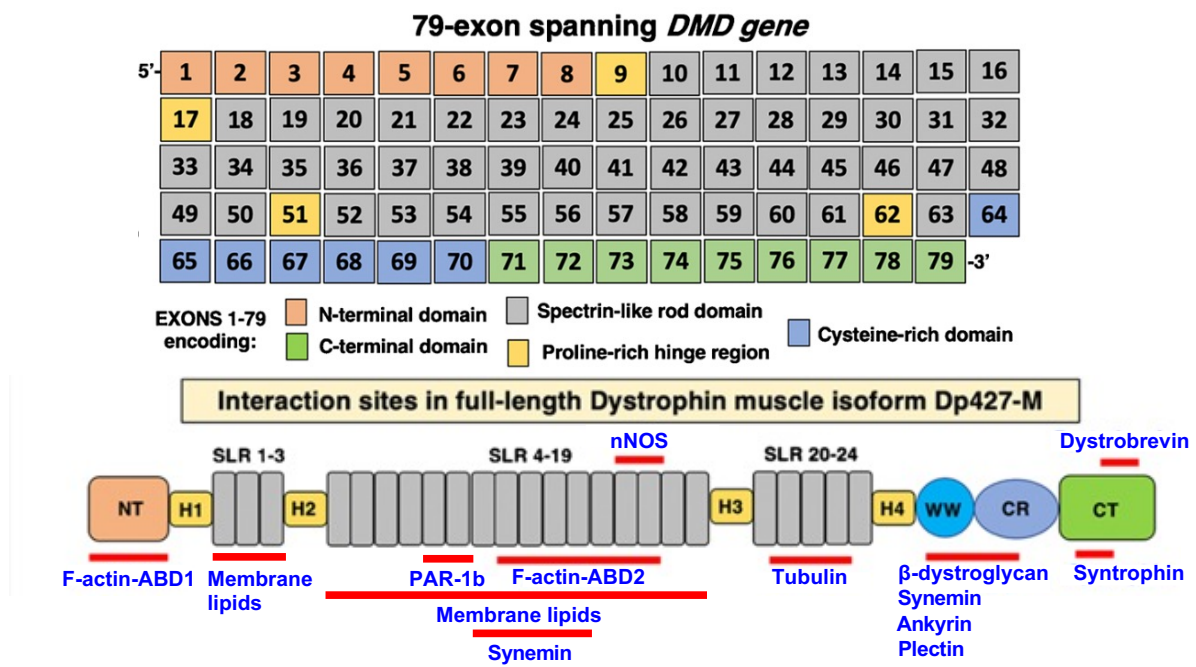
Contrairement à la DMD dont le phénotype clinique est relativement homogène, le spectre clinique de la BMD est beaucoup plus large avec un continuum de gravité déterminée par la quantité et la fonctionnalité de la protéine dystrophine mutée exprimée dans le muscle des patients (Rumeur, 2015). On trouve ainsi des formes BMD sévères jusqu'à des formes peu symptomatiques (crampes et myalgies d'effort) voire asymptomatiques (élévation isolée des enzymes musculaires sans manifestation clinique associée) (Beggs et al., 1991; Nicolas et al., 2012).

Enfin, il existe de rares cas de mutations du gène *DMD*, principalement des mutations ponctuelles en début de gène, retrouvées dans des formes sévères de cardiomyopathie dilatée isolée sans atteinte du muscle squelettique chez l'adolescent ou le jeune adulte (OMIM #302045) pouvant nécessiter une transplantation cardiaque (Muntoni et al., 2003).

L'ensemble des phénotypes associés à un défaut primaire d'expression de la dystrophine est regroupé sous le terme de dystrophinopathies, avec une répartition d'environ 1/3 de BMD et 2/3 de DMD (Bushby et al., 1991; Hughes et al., 1996).

### III. Structure de la dystrophine

La dystrophine musculaire (Dp427m) est une grande protéine de 3685 acides aminés (427 kDa) classée parmi les protéines musculaires géantes. Bien que jouant un rôle structural important dans ce tissu, elle ne représente qu'environ 0.002 % des protéines totales du muscle strié (Hoffman et al., 1987). Elle est composée de quatre domaines structurels majeurs interagissant avec différentes classes de protéines qui incluent des protéines du cytosquelette, transmembranaires, extracellulaires, cytosoliques de signalisation et d'échafaudage, ou endocytiques (Figure 38) (Duan et al., 2021; Ohlendieck and Swandulla, 2021).



Adaptée de Ohlendieck and Swandulla, 2021

Figure 38. Domaines d'interaction de l'isoforme musculaire de la dystrophine (Dp427m).

En haut, les 79 exons du gène DMD sont représentés selon un code couleur en fonction des domaines de la dystrophine pour lesquels ils codent. En bas, les protéines qui interagissent avec les différents domaines de la dystrophine sont indiquées.

**Le domaine N-terminal** (Actin Binding Domain 1, ABD1) est codé par les exons 1 à 8 (aa 1-246) formant deux sous domaines homologues de la calponine, CH1 et CH2 (Koenig et al., 1988; Norwood et al., 2000). C'est un domaine de liaison à l' $\alpha$ -actine du cytosquelette (F-actine) lorsque celle-ci est sous forme de filament (Figure 38).

**Le domaine central** (ou rod domain) est codé par les exons 9 à 61 (aa 253-3040). Il est composé de 24 répétitions homologues à la spectrine (ou spectrine like, SRL 1 à 24) entrecoupées de quatre régions charnières riches en proline (Hinge, H1 à H4) et représente 75% de la dystrophine. Cette région est à l'origine de la nature filamenteuse en forme de bâtonnet de la dystrophine (Ameziane-Le Hir et al., 2014). Ce très long domaine interagit avec les lipides

membranaires via deux domaines au niveau des répétitions R1-R3 (lipid binding domain 1, LBD1) et R4-R19 (LBD2) qui maintiennent la dystrophine en contact avec le sarcolemme (Legardinier et al., 2009), avec la F-actine (domaine ABD2, SLR 11-17) (Amann et al., 1998), la sérine/thréonine kinase Par-1b (SLR 8-9) qui sert de stabilisation avec la  $\beta$ -dystroglycane au niveau du sarcolemme (Yamashita et al., 2010), les filaments intermédiaires comme la synémine (Bhosle et al., 2006; Stone et al., 2005) ou encore les microtubules (SLR 20 à 24) (Belanto et al., 2014).

**Le domaine riche en Cystéines** (domaine CR) codé par les exons 62 à 69 (aa 3080-3360) est le domaine d'interaction avec le complexe des protéines trans-membranaires associées à la dystrophine (DAPC, *Dystrophin-associated protein complex*). Il est composé de quatre sous-domaines incluant la fin d'un domaine WW (2 tryptophanes séparées par 21 acides aminés), deux domaines en « main EF » et d'un domaine en doigt de zinc (ZZ) (Le Rumeur et al., 2010). La dystrophine se lie au  $\beta$ -dystroglycane, qui en interagissant avec l' $\alpha$ -dystroglycane qui se lie à la laminine, l'agrine et le perlecan, permet l'ancrage du sarcolemme à la Matrice Extracellulaire (MEC) (Hnia et al., 2007; Huang et al., 2000). La dystroglycane est à la fois une protéine d'adhésion cellulaire et un récepteur de signalisation (Allen et al., 2016). D'autres protéines interagissent avec le domaine CR telles que l'ankyrine, la synémine  $\alpha$  et la plectine (Le Rumeur et al., 2010).

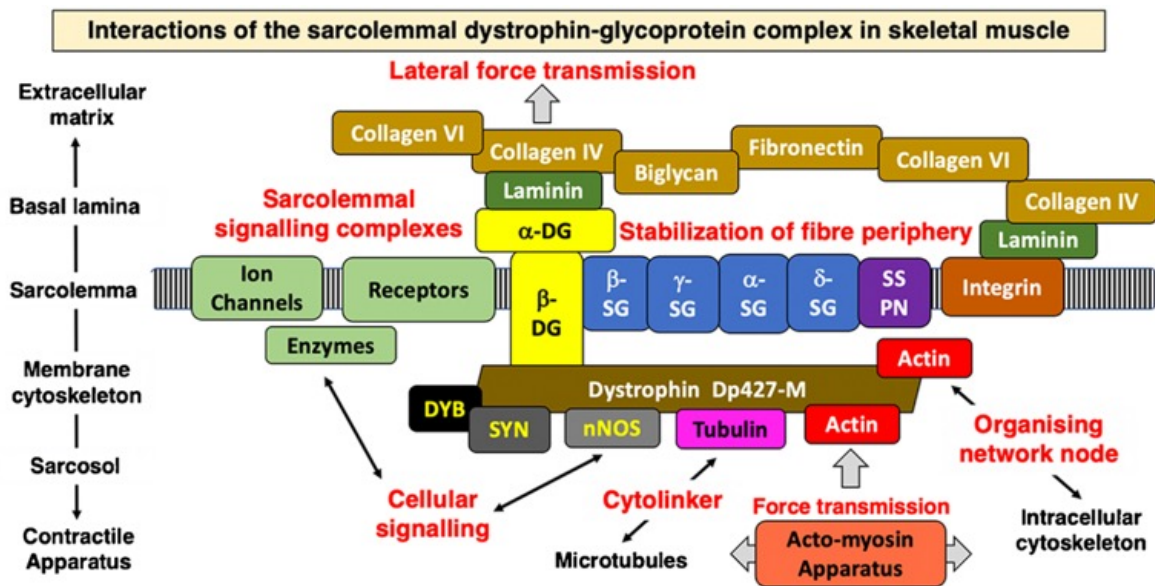
Les derniers exons (70 à 79) codent pour le **domaine C-terminal** de la dystrophine (domaine CT) qui forme une plateforme d'interactions avec des protéines cytoplasmiques, les syntrophines  $\alpha$  et  $\beta$ , des protéines adaptatrices contenant des domaines PDZ qui facilitent le recrutement de protéines de signalisation comme nNOS et l' $\alpha$ -dystrobrevine, une protéine homologue de la dystrophine exprimée dans le muscle squelettique et cardiaque (Adams et al., 2018; Nawrotzki et al., 1998; Pilgram et al., 2010).

#### IV. Le complexe dystrophine et protéines associées (DAPC)

L'isoforme musculaire Dp427m est étroitement liée à un large complexe de protéines qui peuvent être glycosylées ou non (Ervasti and Sonnemann, 2008). Ce complexe DAPC peut être subdivisé en trois sous-complexes : le complexe dystroglycanes (comprenant l' $\alpha$ -dystroglycane extracellulaire et la  $\beta$ -dystroglycane transmembraire), le complexe sarcoglycane-sarcospane (composé des protéines transmembranaires  $\alpha$ -,  $\beta$ -,  $\delta$ - et  $\gamma$ -sarcoglycane et de la sarcospane) et le complexe des protéines cytoplasmiques incluant la dystrophine,  $\alpha$ -dystrobrevine,  $\alpha$ -,  $\beta$ 1-,  $\beta$ 2-syntrophines et l'oxyde nitrique synthase neuronale (nNOS) (Figure 39). Le complexe



DAPC assure un lien mécanique entre l'appareil contractile et la matrice extracellulaire à travers la membrane cellulaire, protégeant le sarcolemme de la rupture durant le stress induit par la contraction musculaire (Ervasti, 2013).



Ohlendieck and Swandulla, 2021

Figure 39. Fonctions et interactions cellulaires de la Dp427m. L'environnement cellulaire de la dystrophine Dp427m au niveau du sarcolemme est représenté détaillant les interactions avec ses principaux partenaires protéiques et les fonctions ou rôles associés. Abréviations : DG, dystroglycane, DYB, dystrobrevine, nNOS, oxyde nitrique synthase, SG, sarcoglycane, SSPN, sarcospane, SYN, syntrophine.

Un réseau plus large de protéines s'organise autour du complexe DAPC, interagissant directement avec la dystrophine ou d'autres composants du DAPC. Ainsi, le complexe de protéines associées à la dystrophine constitue un nœud central dans l'organisation du cytosquelette et conjointement avec l'axe intégrine-laminine, il est largement impliqué dans la transmission de la force latérale de la contraction sarcomérique à l'endomysium et les couches de la MEC. La dystrophine est fortement enrichie au niveau des sites musculaires qui subissent le plus de contraintes mécaniques, tels que la jonction myotendineuse, le costamère (qui relie le sarcomère au sarcolemme) et le réseau du disque Z (Duan et al., 2021). Ce complexe fournit également la structure nécessaire à l'échange efficace d'ions, de métabolites et de molécules de signalisation au sein du système contractile permettant notamment de réguler les processus impliquant l'oxyde nitrique, la voie de signalisation de l'insuline, les voies de signalisation des kinases et le couplage excitation-contraction (Allen et al., 2016).

Compte tenu du rôle essentiel du DAPC dans le maintien de l'intégrité structurelle et de l'activité contractile des cellules musculaires, le désassemblage de ce complexe dû à l'absence de dystrophine chez les patients DMD a des conséquences très diverses sur la fonction des cellules musculaires (Dowling et al., 2021; Forcina et al., 2017).

## V. Les différentes isoformes du gène *DMD*

Il existe une grande diversité des produits d'expression du gène *DMD* due à la présence de différents promoteurs, de deux sites de polyadénylation et de nombreux transcrits alternatifs (Muntoni et al., 2003).

### A. Les promoteurs alternatifs du gène *DMD*

Sept promoteurs indépendants régulent la spécificité tissulaire de l'expression du gène *DMD*. Trois de ces promoteurs sont situés en amont du premier exon et permettent la synthèse de dystrophine pleine longueur de 427 kDa alors que les 4 autres promoteurs internes produisent des isoformes de tailles réduites (Dp260, Dp140, Dp116, Dp71) (Figure 40) (Muntoni et al., 2003; Waite et al., 2012).

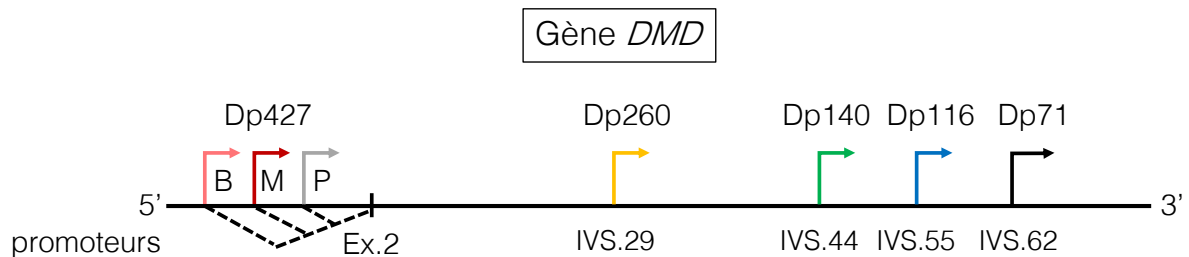


Figure 40. Les différents promoteurs du gène *DMD*.

Trois promoteurs situés à l'extrémité 5' ayant un exon 1 spécifique (indiqués par les flèches rose, rouge et grise) contrôlent l'expression d'isoformes pleine longueur (Dp427) dans le cerveau (B), le muscle (M) et les cellules de Purkinje (P). Des promoteurs internes situés dans les introns (IVS) 29, 44, 55 et 62 (flèches jaune, verte, bleue et noire) produisent des isoformes plus courtes de 260kDa, 140kDa, 116kDa et 71 kDa.

#### 1. Les isoformes longues

Trois promoteurs distincts (B, M, P) situés dans la région 5'UTR contrôlent l'expression d'isoformes longues dans le cerveau (Dp427b appelée aussi Dp427c), les muscles (Dp427m) et les cellules de Purkinje (Dp427p). Elles ne diffèrent que par les quelques acides aminés correspondant à leur exon 1 spécifique à leur extrémité N-terminale, les 78 autres exons étant identiques.

**La Dp427m** est présente dans le muscle squelettique, cardiaque et lisse ainsi que dans les cellules gliales à un faible niveau (Chelly et al., 1990; Tyers et al., 2018). Comme mentionné précédemment, la dystrophine ne représente qu'une faible partie des protéines musculaires (~0,002%) et les niveaux de transcrits dans le muscle squelettique sont par conséquent tout aussi modestes. Il avait initialement été estimé que seulement 5 à 10 molécules étaient présentes par noyau contre 25 000 à 50 000 copies pour les autres protéines du muscle (Muntoni and Strong, 1989; Tennyson et al., 1996). L'utilisation récente d'une technique d'hybridation *in situ* très sensible permettant la détection de molécules individuelles de transcrits (naissants ou

matures) à une résolution subcellulaire confirment ce niveau d'expression. Les noyaux qui expriment la dystrophine contiennent 20 à 40 transcrits naissants. La proportion de transcrits matures est bien moindre (30% des transcrits totaux) et leur demi-vie ne serait que de 2 ou 3 heures ce qui peut paraître surprenant compte tenu du temps de transcription du gène estimé à 10h (Singh and Padgett, 2009). Cela tend à indiquer que le niveau de production de transcrits dépasse les besoins de la cellule en protéine (Hildyard et al., 2020).

**La Dp427c** est une isoforme cérébrale qui s'exprime dans les neurones du cortex et l'hippocampe (Lidov et al., 1990; Nudel et al., 1989). Son promoteur se situe 130kb en amont du promoteur musculaire. La Dp427c apparaît dans le cerveau adulte, elle n'est pas exprimée au stade fœtal (Doorenweerd et al., 2017; Hildyard et al., 2020). Elle est aussi détectable dans l'œil, le cœur et très faiblement dans le muscle squelettique (Holder et al., 1996; Muntoni et al., 1995; Nishida et al., 2015).

**La Dp427p** est l'isoforme des cellules de Purkinje dans le cervelet décrite chez la souris par (Gorecki et al., 1992; Muntoni et al., 2003). Son expression a été rapportée par la suite chez l'homme dans un échantillon de cortex cérébral, et plus faiblement dans le muscle squelettique et cardiaque mais n'a pas été détectée par des techniques de PCR dans le cerveau au stade fœtal (20 semaines) (Holder et al., 1996). Les auteurs ont décrit l'existence de deux variants d'épissage (Dp427p1 et Dp427p2) qui diffèrent par l'insertion de 27 nucléotides flanquant directement l'exon 1 dans la Dp427p2 conduisant à un arrêt prématuré de la traduction seulement 8 acides aminés en aval du codon ATG utilisé dans Dp427p1 (Holder et al., 1996). Cet épissage alternatif se produit durant le développement lors du passage de la forme fœtale à la forme adulte de la dystrophine. Cependant aucune autre étude n'a validé ces résultats. Des données plus récentes viennent remettre en question le profil d'expression de la Dp427p. En effet, l'étude de l'expression des isoformes de dystrophine dans le cerveau humain au cours du développement et dans les tissus adultes (données transcriptomiques de BrainSpan atlas et Allen Human Brain Atlas) montrent que la Dp427p n'est pas détectable dans le cerveau et le muscle, au stade fœtal et adulte (Doorenweerd et al., 2017). Elle est aussi absente au stade embryonnaire chez le chien (Hildyard et al., 2020). Elle est cependant présente dans le cervelet de la souris suggérant un rôle potentiellement différent de cette isoforme dans cette région du cerveau entre l'homme et le modèle murin (Snow et al., 2014). L'expression de la Dp427p dans la rétine et les photorécepteurs a également été rapportée récemment chez la souris (García-Cruz et al., 2022).

## 2. Les isoformes courtes

Les promoteurs intra-géniques produisent des isoformes dénommées en fonction de leur poids moléculaire (Dp260, Dp140, Dp116 et Dp71/Dp40) auxquelles il manque le domaine N-terminal de liaison à l'actine et une partie ou la totalité des 24 répétitions de type spectrine des isoformes longues, mais qui conservent les sites de liaison pour la  $\beta$ -dystroglycane, les syntrophines et les dystrobrevines.

**La Dp260** dont le promoteur est situé dans l'intron 29 est exprimée à des concentrations élevées dans la rétine, où elle coexiste avec les isoformes Dp427c et Dp427m (Pillers et al., 1993). L'exon 1 spécifique code 13 nouveaux acides aminés hydrophiles à l'extrémité N-terminale suivis d'un domaine central composé des répétitions de spectrines 11 à 24 jusqu'au domaine C-terminal. La dystrophine est présente dans la couche plexiforme externe de la rétine et est nécessaire à une fonction rétinienne normale. Les mutations en aval de l'exon 30 qui affectent l'expression de Dp260, sont associées à une réduction de l'amplitude de l'onde b à l'électrorétinogramme et à une certaine déficience de la vision des couleurs rouge-vert chez les patients atteints de DMD (Ricotti et al., 2016b).

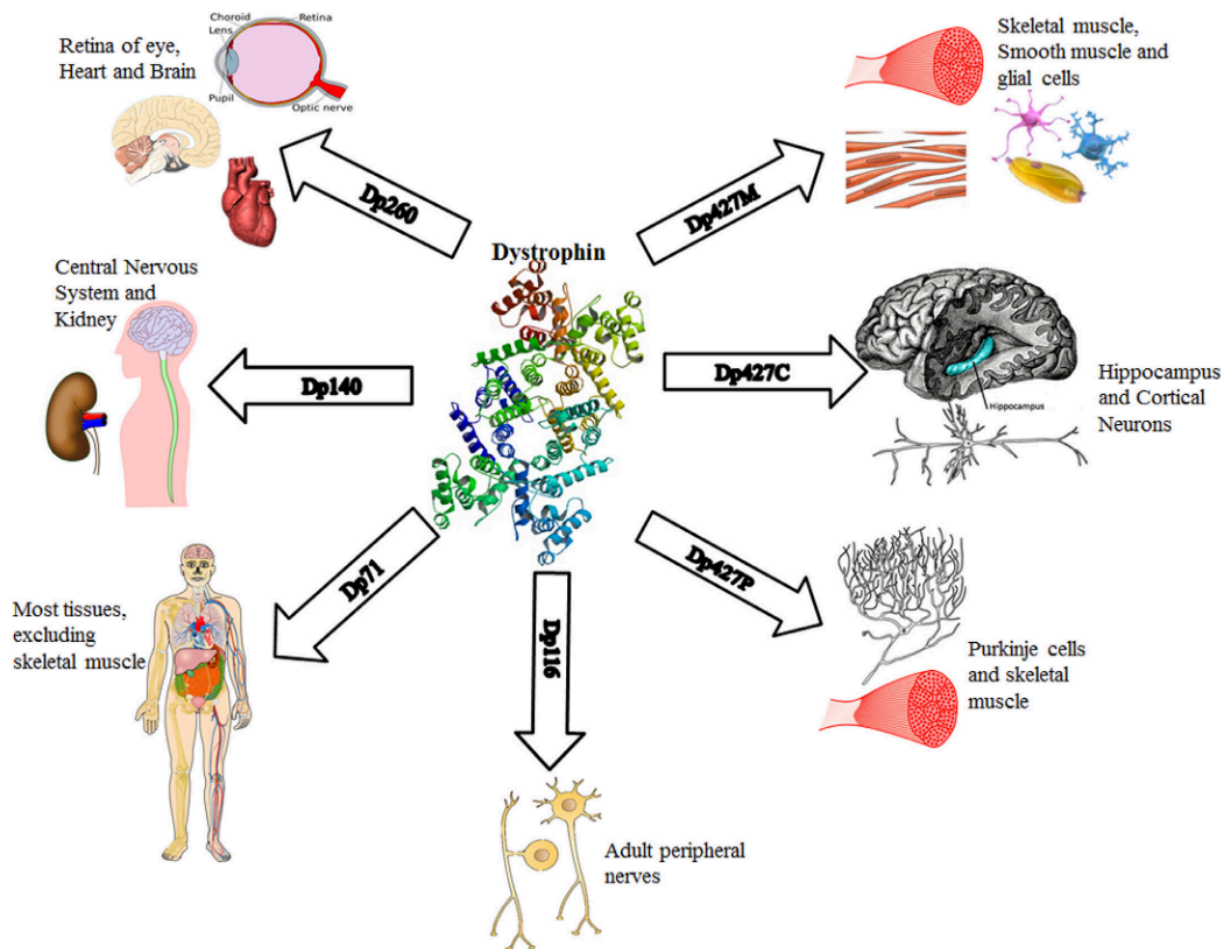
**La Dp140** est présente dans tout le SNC (cortex cérébral, le cervelet, l'hippocampe, le tronc cérébral, la moelle épinière et le bulbe olfactif). Elle a la particularité d'avoir un très long 5'UTR, son promoteur étant localisé dans l'intron 44 et le codon d'initiation de la traduction dans l'exon 51. Elle est aussi présente dans l'œil, les reins (où elle est impliquée dans la tubulogénèse rénale) (Durbeej et al., 1997; Hildyard et al., 2020; Lidov et al., 1995; Lidov and Kunkel, 1998). La Dp140 est l'isoforme la plus exprimée dans le cerveau au stade fœtal chez l'humain et le chien (Doorenweerd et al., 2017; Hildyard et al., 2020).

**La Dp116** est spécifique des cellules de Schwann au niveau des axones neuronaux chez l'homme. Elle est très peu exprimée comparée aux autres isoformes de dystrophine. Son promoteur est dans l'intron 55 et son exon 1 est épissé avec l'exon 56 (Byers et al., 1993; Matsuo et al., 2017). La Dp116 ne conserve que trois répétitions de type spectrine (22 à 24) avant son domaine C-terminal. Son expression dans le muscle cardiaque humain a été rapportée récemment. Les patients DMD porteurs de mutations dans la séquence codante de la Dp116 auraient un meilleur pronostic concernant l'atteinte cardiaque (Yamamoto et al., 2018). Chez la souris, une expression plus large de la Dp116 dans le SNC a été rapportée, cette isoforme étant détectée dans certaines cellules (neurones, astrocytes, oligodendrocytes) et dans les photorécepteurs matures de la rétine (García-Cruz et al., 2022). Un site alternatif d'initiation de

la traduction a été identifié en amont de l'exon 1 de la Dp116 dans l'intron 55 qui coderait pour une isoforme (Dp118) contenant 14 aa supplémentaires (Matsuo et al., 2017).

**La Dp71** est exprimée dans tous les tissus, excepté le muscle squelettique (Lederfein et al., 1992; Naidoo and Anthony, 2020). Son promoteur dans l'intron 62 du gène *DMD* code pour une protéine qui ne possède que les domaines Cystéine riche et C-terminal (Muntoni et al., 2003). Bien que son expression soit ubiquitaire, elle est plus fortement exprimée dans le SNC où elle a été particulièrement étudiée (Aragón et al., 2018; Doorenweerd et al., 2017; Naidoo and Anthony, 2020; Rani et al., 2019; Tadayoni et al., 2012). Cette isoforme joue des rôles dans divers processus cellulaires, notamment l'homéostasie de l'eau, l'architecture nucléaire, l'adhésion cellulaire, ainsi que la division et la survie des cellules (Farea et al., 2020; Tadayoni et al., 2012). Allant à l'encontre des descriptions initiales, la présence de la Dp71 a pu être mise en évidence récemment dans le muscle squelettique humain grâce à l'utilisation d'une technique très sensible de western blot basée sur une détection par un système capillaire. Son expression reste cependant faible. Le transcrit Dp71 ne montre pas d'épissage alternatif dans le muscle squelettique (les 18 exons sont inclus) contrairement aux isoformes cérébrales (Kawaguchi et al., 2018).

La plus petite isoforme de dystrophine Dp40 contient uniquement le site de liaison à la  $\beta$ -dystroglycane et est dépourvue du domaine C-terminal (Tinsley et al., 1993). Son promoteur est identique à la Dp71 mais son 3'UTR se situe dans l'intron 70 qui contient un site polyA alternatif. Le codon stop de la Dp40 est directement à la jonction d'épissage exon/intron 70 (Aragón et al., 2015; Feener et al., 1989; Romo-Yáñez et al., 2020). Comme la Dp71, la Dp40 est exprimée principalement dans le SNC.

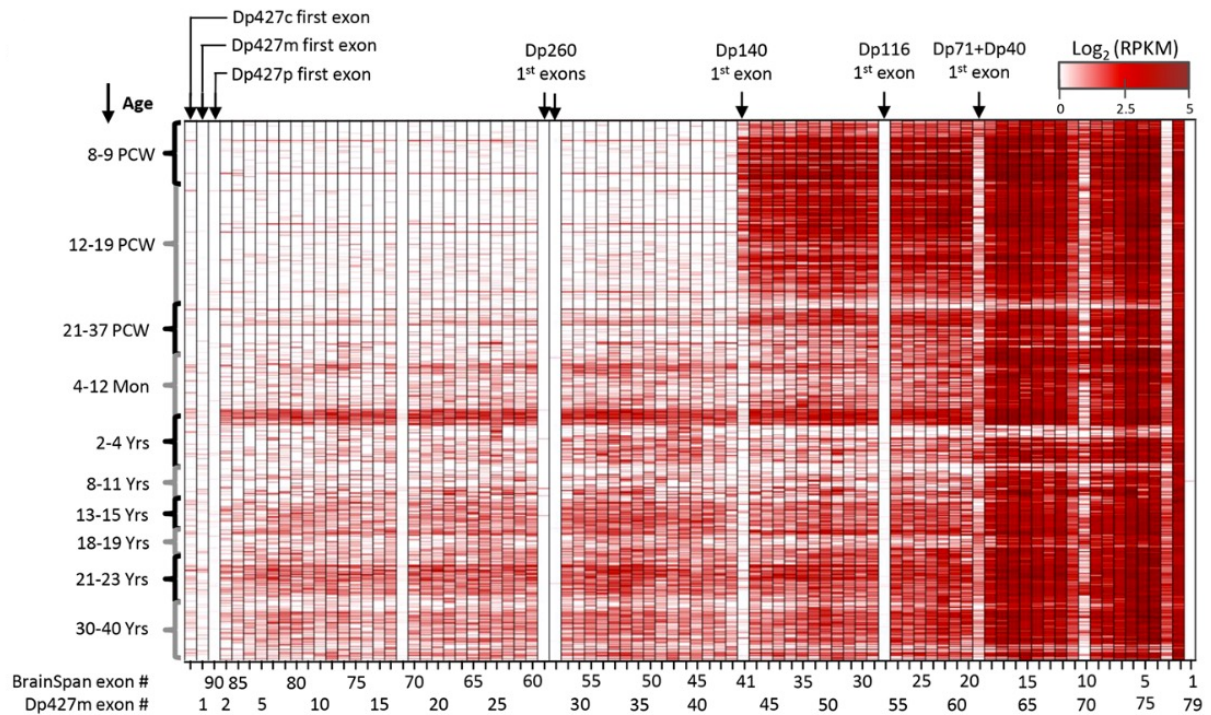


Hildyard et al, 2020 ; Tyers et al, 2018

Figure 41. Spécificité tissulaire de l'expression des isoformes du gène DMD.

Les tissus d'expression des sept isoformes de dystrophine sont représentés. La Dp427m est exprimée dans le muscle squelettique, lisse, et dans les cellules gliales. La Dp427c est exprimée dans le cerveau au niveau des neurones corticaux et de l'hippocampe. La Dp427p est exprimée dans les cellules de Purkinje et dans le muscle squelettique. La Dp260 est exprimée dans la rétine, dans le cœur et le cerveau. La Dp140 est exprimée dans le système nerveux central et les reins. La Dp116 est exprimée dans les nerfs périphériques adultes et la Dp71 est exprimée de façon ubiquitaire à l'exception du muscle squelettique.

Le système nerveux central présente la plus grande variété d'isoformes de dystrophine, qui sont impliquées dans la modulation synaptique, l'excitabilité neuronale et l'intégration des signaux (Doorenweerd et al., 2017; Naidoo and Anthony, 2020). La forme la plus abondante est la Dp71 exprimée dans les neurones et les cellules gliales du gyrus denté de l'hippocampe, sauf pendant le développement où la Dp140 prédomine. La Dp427c apparaît au stade adulte et est présente dans les neurones du cortex cérébral et dans les cellules de Purkinje du cervelet (Figure 42). L'absence d'expression de la Dp260 et de la Dp116 confirme la spécificité de l'expression de ces isoformes respectivement dans la rétine et les nerfs périphériques malgré un rapport récent indiquant que la Dp260 pourrait également être exprimée dans le cerveau (Ricotti et al., 2016a).



Doorenweerd et al., 2017

Figure 42. Expression des isoformes de dystrophine au cours du développement du cerveau humain. Données issues de BrainSpan atlas. Les stades de développement sont indiqués à gauche en semaines post-conceptuelles (PCW), et en mois (Mon) ou années (Yrs) après la naissance. La correspondance entre les numéros des exons dans l'atlas BrainSpan et la Dp427m est indiquée sous la figure.

De manière intéressante, les analyses de co-expression réalisées à partir de données disponibles dans les banques BrainSpan atlas et Allen Human Brain Atlas database (AHBA) ont montré une forte association entre l'expression des isoformes courtes ou longues de la dystrophine et certains gènes impliqués dans des troubles neurodéveloppementaux, notamment la déficience intellectuelle et les troubles du spectre autistique. Ceci pourrait expliquer la comorbidité élevée de ces troubles chez les patients atteints de DMD. L'association de la Dp140, exprimée de façon précoce au cours du développement, avec des gènes liés au neurodéveloppement précoce *via* la régulation de la différenciation des neurones et de la morphogenèse de la projection des neurones est particulièrement significative (Doorenweerd et al., 2017).

## VI. Spécificité tissulaire et cellulaire du DAPC

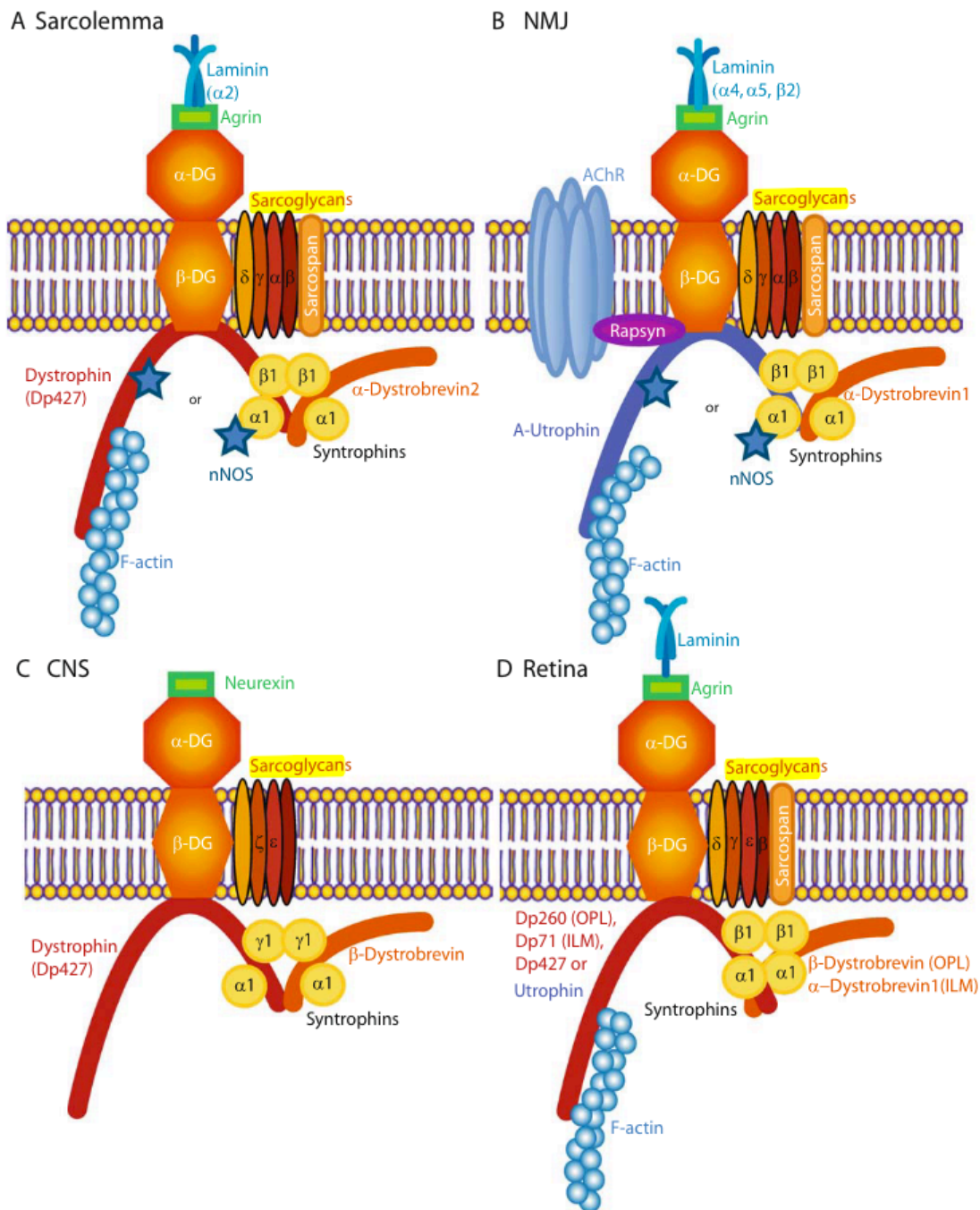
La formation du DAPC est dynamique et sa composition varie en fonction des tissus, des cellules voire même au sein d'un même myocyte selon les isoformes de dystrophine exprimées (Duan et al., 2021). Il est donc difficile d'attribuer une fonction générale à ce complexe transmembranaire. Cependant, dans tous les organismes étudiés à jour, les complexes DAPC sont localisés à la membrane plasmique et contribuent au maintien des différents types

cellulaires et à la régulation des voies de signalisation spécifiques des cellules ou tissus d'expression où ils s'expriment (Figure 43) (Pilgram et al., 2010).

Nous avons vu précédemment les rôles du complexe DAPC associé à l'isoforme Dp427m dans le muscle squelettique. Dans le **cerveau**, le DAPC est impliqué dans l'organisation des récepteurs GABAA (GABAAR) dans les neurones et des complexes protéiques contenant l'aquaporine-4 (AQP4) dans les cellules gliales. Contrairement au muscle, il existe de multiples formes de complexes DAPC en raison de la présence des nombreux isoformes et variants d'épissage de la dystrophine ainsi que d'isoformes différentes de dystrobrevine et de syntrophines que celles présentes dans le muscle (Waite et al., 2012). Dans le muscle squelettique, il existe quatre sites potentiels de liaison à la syntrophine par complexe de dystrophine : deux sur la dystrobrevine et deux sur la dystrophine ou l'utrophine, une protéine homologue de la dystrophine exprimée à la jonction neuromusculaire. Par un épissage alternatif, la dystrobrevine peut réguler la stœchiométrie de recrutement des syntrophines et en conséquence, moduler les fonctions de signalisation du DAPC (Newey et al., 2000). Dans le cerveau, l'épissage alternatif de la dystrophine contribue également à la modulation du nombre de molécules de syntrophine qui peuvent s'associer au DAPC, l'épissage alternatif des exons 71-74 supprimant deux sites de liaison à la syntrophine.

Dans la **rétilne**, le complexe DAPC est nécessaire à l'activité électrique et la formation de la rétine chez les mammifères tandis que dans les **jonctions neuromusculaires** il permet le regroupement des récepteurs des neurotransmetteurs (Pilgram et al., 2010).





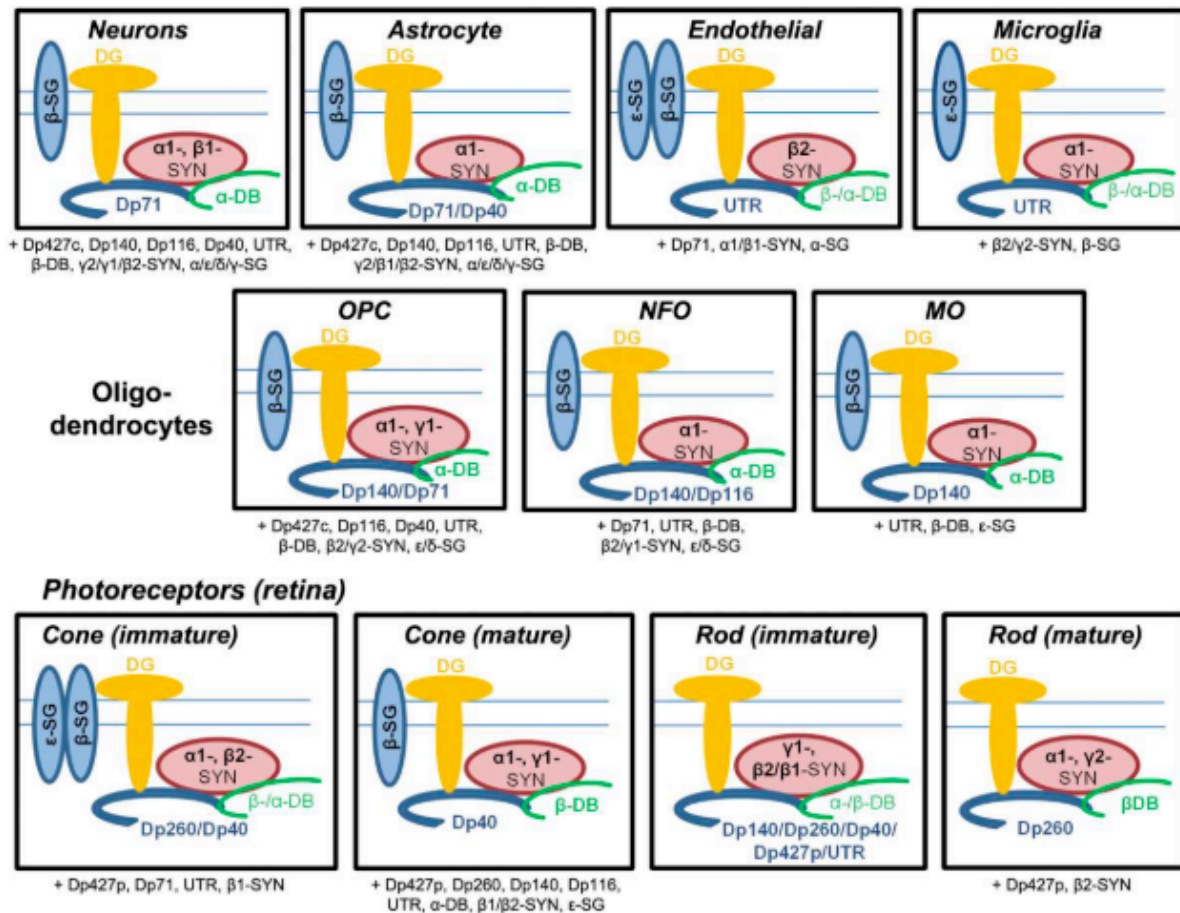
Pilgram et al., 2010

Figure 43. Les différents complexes DAPC en fonction des tissus.

Les compositions des complexes de protéines associées à la dystrophine ou à la protéine homologue, l'utrophine, sont représentées : (A) au niveau du sarcolemme des muscles des vertébrés, (B) au niveau de la jonction neuromusculaire (NMJ), (C) dans le système nerveux central (SNC) et (D) dans la rétine.

L'analyse de données RNA-Seq du niveau d'expression des différentes isoformes de dystrophine et des protéines associées dans les cellules du cerveau et les photorécepteurs rétiniens de la souris, a permis récemment de proposer différents modèles de DAPC dont la composition varie en fonction des sous-domaines ou sous-populations cellulaires (García-Cruz et al., 2022). Cette étude illustre la grande complexité de régulation et la spécificité des

fonctions des complexes DAPC au sein des cellules où s'expriment les isoformes de dystrophine (Figure 44).



Garcia-Cruz et al., 2022

Figure 44. Modèles de composition des DAPC dans différents types cellulaires. Représentation schématique des complexes élaborée à partir du niveau d'expression des gènes de la dystrophine et des protéines associées dans différents types de cellules du cerveau et de la rétine. Les noms des isoformes de dystrophine et protéines ayant un faible niveau d'expression comparé aux autres composants du DAPC sont indiqués sous les schémas.

## VII. Épissages alternatifs des isoformes de dystrophine

Les transcrits synthétisés à partir des différents promoteurs subissent divers épissages alternatifs ajoutant un niveau de régulation supplémentaire dans le contrôle de l'expression spatio-temporelle du gène *DMD*. Ce mécanisme est particulièrement présent en partie distale des isoformes exprimées dans le SNC (Hildyard et al., 2020).

L'évolution des techniques et des approches, notamment l'utilisation du RNA-Seq, permet d'explorer de manière toujours plus approfondie la diversité des variants d'épissage du gène *DMD* et d'affiner les profils d'expression des différentes isoformes exprimées ou co-exprimées dans un tissu, une sous-population de cellules voire dans une cellule unique. Cependant, le rôle biologique des variants d'épissage présents parfois en faible quantité, reste à établir dans un grand nombre de cas. Par ailleurs, les variants d'épissage très nombreux détectés dans des

isoformes de dystrophine dans divers contextes pathologiques ne permettent pas d'extrapoler leur présence à la condition physiologique.

Une description exhaustive de tous les variants d'épissage caractérisés dans les différentes isoformes de dystrophine étant difficile à réaliser, les données les plus récentes ou les plus significatives par rapport aux données des travaux de thèse sont rapportées ci-après.

## A. Épissage alternatif des isoformes non musculaires

### 1. La Dp427c

L'épissage alternatif de la Dp427c dans le cerveau a été initialement rapportée par Feener et al., (1989) qui ont décrit les sauts des exons 71, 71-72, 71-74 associés à une exclusion totale de l'exon 78. Plus récemment, l'analyse du profil d'épissage par RT-PCR de la Dp427c dans la lignée cellulaire humaine de neuroblastome SH-SY5Y (Nishida et al., 2015a) a révélé de nouveaux épissages alternatifs incluant les sauts partiels des exons 9, 37, 38, 71, l'inclusion d'un pseudo exon issu de l'intron 1 (pseudo exon 1a) et l'utilisation d'un site accepteur cryptique fort dans l'intron 40 (851 pb) menant à l'inclusion des 568 derniers nucléotides de l'intron 40 définissant un nouvel exon 41 (appelé 41e) (Figure 45). La régulation de ce petit intron est particulière. L'analyse de 20 tissus montre sa rétention à des niveaux variables dans les différents tissus en fonction des isoformes (faible <5% dans le muscle et le cerveau et jusqu'à 36 % dans le rein) (Nishida et al., 2015b). Cette rétention de l'intron 40 est aussi présente dans une lignée de rhabdomyosarcome (Niba et al., 2017).

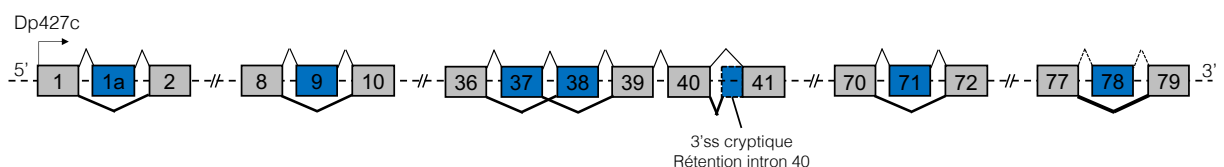


Figure 45. Épissages alternatifs de la Dp427c dans la lignée humaine neuronale SH-SY5Y.

Sept épissages alternatifs ont été détectés dans la Dp427c (rétention du pseudo-exon 1a, sauts des exons 9, 37, 38, 71 et 78, et la rétention de la région 3' de l'intron 40). Les produits d'épissage alternatifs majoritaires de cette isoforme sont décrits (boîtes bleues et traits épais). Un saut complet de l'exon 78 a été observé (traits en pointillés vs traits très épais).

### 2. Les Dp260, Dp140, Dp116

La Dp260 présente les épissages alternatifs des exons 73 ou 73-76, spécifiques de cette isoforme (Pillers et al., 1993). Récemment, l'inclusion d'un nouvel exon de 36 pb (51b) issu de l'intron 51 a été décrit dans la rétine chez la souris (principalement dans les photorécepteurs immatures) dans la Dp260 mais également dans la Dp427 et la Dp140 (García-Cruz et al., 2022). Cette même étude a rapporté l'identification de transcrits avec des insertions d'une partie de l'intron 40 ou 41 générant de nouvelles extrémités 3'UTR et des protéines tronquées dans la rétine

(Dp260) chez la souris et l'homme mais également à faible niveau dans la Dp427 cérébrale. Ces isoformes diffèrent de l'isoforme Dp234 identifiée récemment dans des cardiomyocytes immatures dérivés de cellules iPS qui utilise un site alternatif de polyadénylation dans l'intron 41 (Rani et al., 2020). Chez la souris, un variant d'épissage au niveau de l'exon 1 produit une deuxième isoforme (Dp260-2) comportant une extrémité N-terminale différente incluant 13 aa supplémentaires à l'extrémité N-terminale (D'Souza et al., 1995).

Dans la Dp140, les données sont plus anciennes et reposent sur une étude par RT-PCR montrant les sauts des exons 71, 78 et 71-74 dans le cerveau alors que dans le rein, le saut de l'exon 78 prédomine (Lidov and Kunkel, 1997).

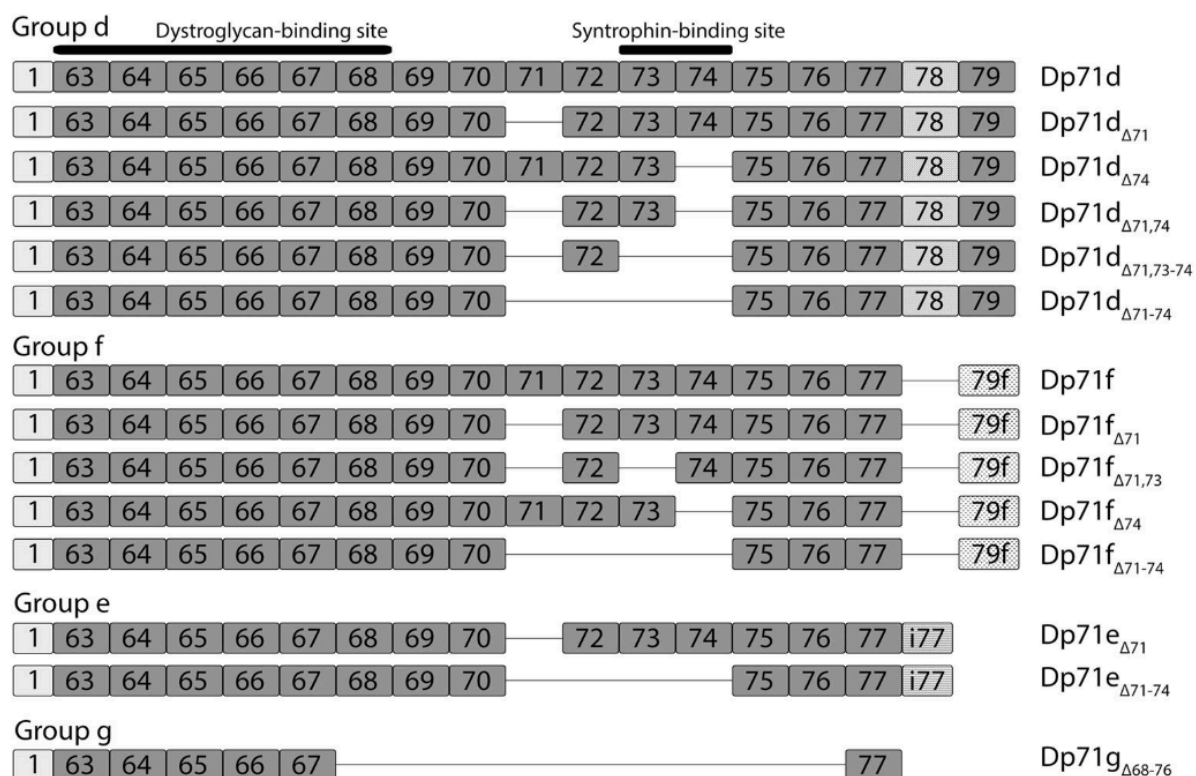
Comme nous l'avons vu précédemment, la Dp116 est exprimée de manière très spécifique dans les cellules de Schwann et quelques sous populations de cellules dans le SNC. Les épissages alternatifs physiologiques de cette isoforme sont peu connus. Certains ont été identifiés par RT-PCR dans une lignée de cellules de glioblastome (U-251) incluant les sauts des exons 68-69, 71, 78, 71-74 et l'activation d'un site cryptique 3' dans l'exon 75 (Mahyooob Rani et al., 2019).

### 3. Cas particulier de la Dp71

#### i) Épissages alternatifs de la Dp71

L'étude de la régulation de l'expression de la Dp71, isoforme majoritaire de la dystrophine dans le SNC, est d'un intérêt particulier compte tenu de son implication dans l'atteinte cognitive et les troubles neurodéveloppementaux observés chez les patients DMD porteurs d'une mutation dans les exons 63 à 79 du gène DMD (Naidoo and Anthony, 2020).

La complexité de l'épissage alternatif dans la Dp71 a conduit à l'établissement d'une nomenclature permettant de décrire les différents événements et leur association au sein d'isoformes spécifiques dans le SNC (Figure 46) (Aragón et al., 2018; Niba et al., 2021; Paúl-González et al., 2021; Rani et al., 2019). On distingue en particulier les isoformes incluant l'exon 78 (groupe d) de celles qui l'excluent (groupe f). Les événements d'épissages dans ces deux groupes sont en phase et permettent l'expression d'isoformes de Dp71 ayant des fonctions distinctes. Le saut de l'exon 78 présent dans de nombreux tissus décale la position du codon stop naturel dans l'exon 79, remplaçant les 13 derniers résidus hydrophiles de l'extrémité C-terminale de la protéine par une nouvelle séquence de 31 acides aminés ayant des propriétés physico-chimiques différentes (extrémité hydrophobe) (Howard et al., 1999).



Aragon et al., 2018; Naidoo and Anthony, 2020

Figure 46. Nomenclature des épissages alternatifs de la Dp71.

Les isoformes d'épissage de la Dp71 sont regroupées en fonction de leur extrémité C-terminale. Le groupe d inclut les isoformes dans lesquelles l'exon 78 est présent et le groupe f, celles dépourvues de l'exon 78. Les exons 78 et 79 sont absents dans les isoformes des groupe e et g qui contiennent un codon stop prématuré dans l'intron 77 (i77) ou l'exon 77. Les épissages alternatifs présents dans les différents groupes sont décrits. Les positions des sites de liaison à la dystroglycane et à la syntrophine sont indiquées.

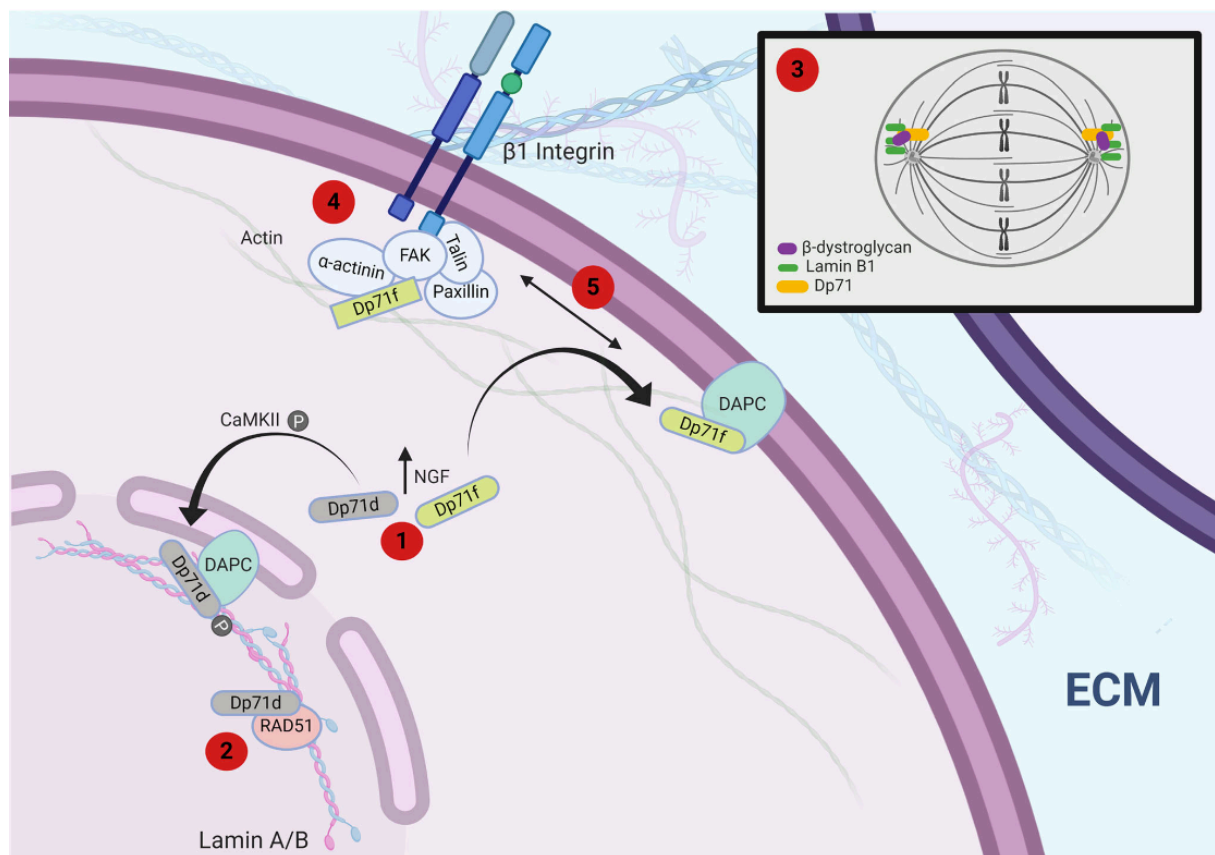
Les isoformes du "groupe e" incluent 34 nucléotides de l'intron 77, ce qui entraîne un décalage du cadre de lecture et l'apparition d'un codon stop prématuré empêchant la traduction des exons 78 et 79. L'extrémité C-terminale des Dp71e est hydrophile et composée de 10 acides aminés. La Dp71g est déletée pour les exons 68 à 76 et contient un codon stop dans l'exon 77.

Un épissage alternatif de la Dp40 a été identifié durant la différenciation neuronale et gliales du SNC chez le rat impliquant les exons 63-64 et des exons 64-67 dans les cellules non différenciées (Paúl-González et al., 2021).

## ii) Fonction des sous-isoformes de la Dp71

Les épissages alternatifs donnent lieu à des protéines ayant une localisation subcellulaire et des fonctions différentes (Naidoo and Anthony, 2020). Dans des cellules PC12, la Dp71f<sub>Δ71</sub> est associé au DAPC uniquement dans les cellules indifférenciées (Romo-Yáñez et al., 2007). L'expression des sous-groupes Dp71d, Dp71f et Dp71e<sub>Δ71</sub> est aussi associée à la différenciation neuronale (Acosta et al., 2004; García-Cruz et al., 2019; Herrera-Salazar et al., 2016). La Dp71f<sub>Δ71</sub> est exclusivement cytoplasmique alors que la Dp71d<sub>Δ71</sub> est nucléaire

dans plusieurs types cellulaires (HeLa, C2C12, N1E-115 et PC12) où elles participent aux complexes d'adhésion avec les intégrines (Figure 47) (González et al., 2000; Marquez et al., 2003). La Dp71d $\Delta$ 71 est aussi impliquée dans la division cellulaire des neurones avec la lamine B1 et le  $\beta$ -dystroglycane (Acosta et al., 2004; Higginson et al., 2008; Villarreal-Silva et al., 2011). La diminution de Dp71d $\Delta$ 71 mène à une croissance réduite des neurones, un retard de transition G0/G1, des défauts de localisation des protéines nucléaires et une augmentation de l'apoptose impactant la différenciation neuronale (Villarreal-Silva et al., 2011). Les Dp71d et f sont aussi observées dans l'appareil de Golgi et les mitochondries (Aleman et al., 2001; Chávez et al., 2000; Niba et al., 2021).



Naidoo and Anthony, 2020

Figure 47. Différentes fonctions des sous-isoformes de la Dp71 en fonction de leur localisation subcellulaire.

Rôles non-synaptiques multiples de Dp71. (1) Lors de la différenciation induite par le NGF, la Dp71d est phosphorylée par CaMKII et se déplace vers le noyau où elle intègre un DAPC nucléaire et peut servir de médiateur aux réponses nucléaires pendant la signalisation du NGF. En revanche, la Dp71f se déplace vers la surface cellulaire et se localise avec la DAPC dans des zones telles que le cône de croissance pour jouer un rôle dans la croissance des neurites. (2) La Dp71d nucléaire se lie à la protéine de réparation de l'ADN, RAD51. Elle joue également un rôle pendant la division cellulaire (3). (4) La Dp71f est un facilitateur de l'adhésion cellulaire et se lie au complexe d'adhésion  $\beta$ 1-intégrine. (5) Dp71 aide à la communication croisée entre le DAPC et le complexe d'adhésion  $\beta$ 1-intégrine.

## B. Régulation de l'épissage de la Dp427 musculaire

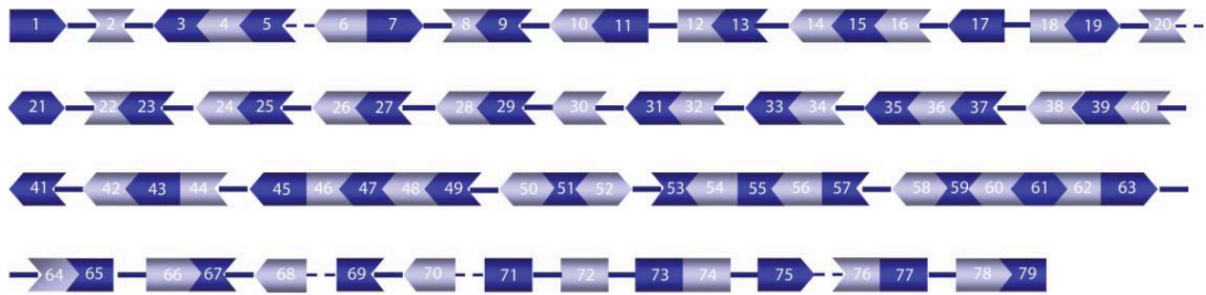
### 1. Épissage récursif de la Dp427m

Le gène *DMD* possède de très longs introns (taille moyenne de 28 kb, compris entre 107 pb - 360 kb) dont 24 ont une taille supérieure à 10kb et 5 une taille supérieure à 100kb (introns 1, 2, 7, 44, 55) suggérant fortement une élimination de ces introns par épissage récursif comme précédemment décrit chez la drosophile (Sibley et al., 2016). Néanmoins, il s'est avéré difficile de vérifier cette hypothèse. Étant donné le faible niveau d'expression du gène *DMD*, les formes intermédiaires de transcrits résultant potentiellement d'un épissage récursif sont présentes dans des proportions extrêmement faibles dans la cellule.

Une équipe a réalisé un séquençage ciblé à très grande profondeur de l'ARN pré-messager de *DMD* ("Capture-pre-mRNA-Seq") (Gazzoli et al., 2016). L'ARN nucléaire isolé de lignées de cellules musculaires humaines a été hybridé à des sondes biotinylées couvrant tous les exons, les introns, les promoteurs et les UTR. Les auteurs ont pu démontrer que seuls 31 des 78 introns étaient épissés en une seule étape. La majorité d'entre eux ont une taille inférieure à 10 kb (moyenne 6.4 kb, de 107 pb à 38,4 kb) contrairement aux 47 introns épissés en plusieurs étapes qui ont une longueur moyenne de 40 kb (650 pb - 248,4 kb). Ces données confirment les précédentes observations que l'épissage récursif a lieu principalement dans les introns les plus longs. Il existe cependant des exceptions comme on peut le voir ici dans le gène *DMD* avec l'intron 48 (38,4 kb) épissé en une seule étape alors que les deux petits introns 10 (650 pb) et 40 (851 pb) subissent un épissage récursif.

### 2. Épissage non séquentiel du transcrit Dp427m

La même étude de Gazzoli et al. (2016) a permis d'aborder la question d'un épissage séquentiel ou non séquentiel des introns de *DMD*. Il apparaît qu'environ 40 % des introns sont épissés de manière non séquentielle. Ces introns sont épissés seulement après que d'autres introns en aval aient été éliminés, potentiellement post-transcriptionnellement. L'épissage ne suit donc pas strictement un ordre 5'-3' et la vitesse d'élimination des introns n'est pas dépendante de leur taille. Il en résulte la formation de transcrits intermédiaires contenant des blocs de trois exons ou plus, en fonction de la vitesse d'épissage des introns situés entre ces exons (Figure 48).



Gazzoli et al., 2016

Figure 48. Épissage séquentiel et non séquentiel des introns de la Dp427m. Les lignes pleines et en pointillés entre 2 exons indiquent un épissage non séquentiel préférentiel (lent et intermédiaire, respectivement) tandis que les introns épissés séquentiellement ne sont pas représentés. La forme de l'exon reflète la mise en phase des exons basée sur les triplets de codons au début de l'exon.

Il n'y a pas de biais de distribution dans le gène des introns épissés en plusieurs étapes ou de manière non séquentielle. On peut noter cependant qu'un groupe d'introns en fin de gène (68 à 75) a un épissage plus lent bien qu'il s'agisse d'introns courts (< 5 kb, à l'exception de l'intron 74). Comme décrit précédemment, ces exons sont alternatifs dans de nombreux tissus. Cette observation est en accord avec la plus faible efficacité d'épissage co-transcriptionnel rapportée pour les introns entourant des exons alternatifs (de la Mata et al., 2010; Tilgner et al., 2012).

### 3. Épissage alternatif de la Dp427m

#### a. Premières descriptions

Contrairement à la Dp71, l'épissage alternatif physiologique de la Dp427m a été peu décrit. La majorité des épissages alternatifs ont été identifiés par l'analyse des transcrits musculaires ou lymphocytaires chez des patients qui ont mis en évidence des épissages aberrants dans les transcrits mutés indépendamment de la localisation de la mutation. Leur proportion est souvent surestimée en raison des techniques PCR utilisées (Bougé et al., 2017, Supplemental information S1).

Plusieurs études basées sur des techniques de RT-PCR (couplées ou non à des PCR nichées) de régions ciblées du gène ont mis en évidence des épissages alternatifs chez l'homme: (i) des exons 71 et 78 dans le muscle squelettique et cardiaque et de l'exon 68 dans le muscle lisse de l'estomac au stade fœtal (Feener et al., 1989), (ii) de l'exon 4 dans des ARN de muscle squelettiques (Torelli and Muntoni, 1996), (iii) différentes combinaisons des exons 1 à 18 (Sironi et al., 1997) ainsi que dans la région centrale du gène autour des exons 26 à 53 (Sironi et al., 2002). Ces approches ont également mis en évidence l'inclusion à des niveaux faibles de pseudo exons provenant de l'intron 1 et de l'intron 2 dans les muscles contrôles.



## b. Données récentes

En l'absence de données claires du profil d'épissage de la Dp427m, notre équipe a entrepris en 2017 de séquencer la totalité du transcrit Dp427m par une approche de RNA-Seq ciblé afin d'obtenir un fort taux de couverture (>1000X) pour ce transcrit faiblement exprimé (Bougé et al., 2017). La séquence codante (11,3 kb) a été amplifiée en un seul fragment à l'aide d'amorces spécifiques de l'isoforme Dp427m à partir de 4 ARN de muscle squelettique d'individus normaux et analysée sur un séquenceur 454 Junior (longueur de reads>300pb).

De manière surprenante, un nombre restreint d'épissages alternatifs a été identifié, tous à un faible niveau ( $\leq 3\%$ ) (Figure 49), démontrant que dans le muscle squelettique, les 79 exons sont inclus pour permettre la synthèse d'une protéine pleine longueur. Une régulation spécifique est en place pour inclure à un niveau proche de 100% les exons alternatifs des isoformes cérébrales (exons 71, 74, 78) dans la Dp427m.

Event name	Muscle mean ASE (%)	Predicted reading frame
del9	0.63	in frame
<b>del71</b>	<b>1.80</b>	<b>in frame</b>
del73	0.32	in frame
del74	0.49	in frame
<b>del78</b>	<b>2.48</b>	<b>out of frame/elongated C-term</b>
<b>PE1a (162 nt)</b>	<b>3.06</b>	<b>out of frame</b>
PE21X (66 nt)	0.40	in frame
PE51X (84 nt)	0.21	in frame
3'ss exon 20 (-2 nt)	0.18	out of frame
3'ss exon 54 (-3 nt)	1.12	in frame
3'ss exon 76 (-60 nt)	0.33	in frame
3'ss exon 78 (-4 nt)	0.43	out of frame

Bougé et al., 2017

Figure 49. Épissage alternatif physiologique dans le muscle squelettique.

Les trois principaux Alternative Splicing Events (ASEs) détectés dans le muscle squelettique sont indiqués en gras. Le saut de l'exon 78 de DMD remplace les 13 acides aminés C-terminaux de la dystrophine par 31 nouveaux acides aminés (C-terminal allongé). La taille des pseudo-exons (PE) insérés dans les transcrits matures de la dystrophine est indiquée entre parenthèses (nt, nucléotides). Le nombre de nucléotides exoniques (nt) délétés dans le transcrit mature en raison de l'utilisation de sites d'épissage 3' alternatifs (3'ss) est indiqué entre parenthèses.

## c. Les régulateurs de l'épissage alternatif de Dp427m

Les facteurs régulant l'épissage de la Dp427m restent largement inconnus à ce jour. Notre équipe a contribué à identifier certains des facteurs *trans* impliqués dans l'inclusion des exons dans l'isoforme musculaire en réalisant un crible fonctionnel. Seize facteurs d'épissage généraux ou spécifiques du muscle ont été déplétés avec des si-ARN dans une lignée cellulaire musculaire humaine (C25C148) différenciées pendant 3 jours. L'analyse du profil d'épissage

du transcrit Dp427m par RNA-Seq ciblé comme décrit précédemment a permis d'identifier des facteurs impliqués dans la régulation de l'épissage des exons 9, 37, 38, 41, 71, 74 et 78 et du pseudo exon 1A, dont le facteur MBNL1 pour l'exon 78 (Miro et al., 2020).

L'équipe a également identifié le rôle de la protéine FUBP1 (*Far Upstream element-Binding Protein 1*) plus connue comme facteur de transcription (Debaize and Troadec, 2019), dans l'épissage de l'exon 39. La protéine se fixe à une région riche en UG dans l'intron 38 et agit comme un activateur de l'inclusion de l'exon 39 dans le transcrit mature musculaire (Miro et al., 2015).

#### 4. Épissage de la Dp427m au cours de la myogenèse

##### a. Durant le développement du muscle

L'épissage de la Dp427m durant le développement musculaire est peu décrit dans la littérature à l'exception de la transition de l'inclusion de l'exon 78 entre la forme embryonnaire et la forme adulte de dystrophine. Chez l'homme et la souris, l'exon 78 est inclus à partir de 20 semaines de développement dans les transcrits du muscle squelettique et cardiaque adulte (Bies et al., 1992; Feener et al., 1989). Cet épissage qui décale la position du codon stop naturel comme indiqué précédemment modifie la structure de l'extrémité C-terminale de la dystrophine passant d'une extrémité hydrophobe dans les tissus embryonnaires à une extrémité hydrophile dans les tissus adultes, modifiant les interactions de la dystrophine avec les protéines du complexe DAPC (Bies et al., 1992; Howard et al., 1999). L'utilisation d'un oligonucléotide induisant le saut de l'exon 78 dans le poisson zèbre montre que cette transition est nécessaire au cours du développement pour une fonction musculaire normale (Rau et al., 2015). L'inclusion de l'exon 78 est dépendante du facteur MBNL1 comme observé chez les patients atteints de dystrophie myotonique de type 1 (DM1). La séquestration de la protéine MBNL1 dans les transcrits mutants du gène DMPK (porteur de répétitions pathologiques du motif CUG) diminue la quantité de protéine nucléaire et induit des anomalies d'épissage des exons dépendant de MBNL1, dont l'exon 78 de DMD dans le muscle squelettique des patients atteints de DM1 (Rau et al., 2015).

##### b. Durant la différenciation musculaire

Les études qui ont analysé par RT-PCR l'épissage alternatif du transcrit dystrophine au cours de développement ou de la différenciation musculaire ne permettent pas de s'assurer de l'identité de l'isoforme analysée (Dp71 ou Dp427m). Il est possible que le changement de l'extrémité C- hydrophobe à l'extrémité C- hydrophile qui a été décrit, reflète en fait un

changement d'isoforme de Dp71 (avec saut de l'exon 78) dans les cellules myogéniques précoces à Dp427m (inclusion de l'exon 78) dans les myofibres matures (Howard et al., 1999).

Le changement d'isoforme durant la différenciation musculaire a pu être établi à différents stades de la mise en place du muscle, à partir des données de RNA-Seq total obtenues par l'analyse de cellules humaines embryonnaires (hESC) et pluripotentes induites (hiPSC) qui ont été différenciés jusqu'à la formation de myotubes (Choi et al., 2020). La Dp71 représente l'isoforme majoritaire (90% à 83 %) dans les cellules embryonnaires non différenciées, les cellules pré somitiques, et les progéniteurs du muscles, associé à fort taux du saut de l'exon 78 dans les transcrits (98% à 87%). Au contraire, lors de l'engagement de la différenciation musculaire, la Dp427m représente 89% des transcrits *DMD* totaux dans les myoblastes et 64% des transcrits possède le saut de l'exon 78. A la fin de la différenciation, dans les myotubes, l'isoforme Dp427m représente 97% des transcrits exprimés avec un saut de l'exon 78 de seulement 14%.

Ceci montre un rôle distinct des deux isoformes lors de la myogenèse (la Dp71 est présente précocement à partir de 10 semaines de développement) et la différenciation avec un changement d'expression des isoformes de dystrophine durant ces processus. La Dp71 s'exprimerait lors de la prolifération des myoblastes et au début de la différenciation lors de la fusion des myotubes puis son expression est réprimée dans le muscle adulte (Howard et al., 1999). La Dp427m s'exprime seulement à partir du déclenchement de la différenciation en augmentant son expression pour être la seule isoforme présente dans le muscle squelettique adulte et établir sa fonction de maintien de la fibre musculaire. L'apparition de la Dp427m lors de la différenciation est validée par des publications plus récentes. A partir de cellules musculaires immortalisées, l'apparition de la Dp427m a été détectée à partir de trois jours de différenciation (Goswami et al., 2021; Miro et al., 2020).

Récemment, une nouvelle dystrophine embryonnaire Dp412e spécifique des premiers stades de la différenciation induite par BMP4 dans les cellules souches pluripotentes humaines lors de la formation du mésoderme a été mise en évidence (Massouridès et al., 2015). Elle possède un exon 1 spécifique et un début de traduction dans l'exon 6, délétant une partie du domaine N-terminal de liaison à l'actine dans la protéine. Les processus de régulation et les fonctions de cette nouvelle isoforme restent à déterminer. Son identification fournit cependant un modèle d'étude potentiellement intéressant d'une isoforme longue de la dystrophine dont la synthèse peut être obtenue rapidement à partir de lignées de cellules souches pluripotentes humaines traitées par BMP4 pour induire la différenciation mésodermique.

## Partie 4 : La différenciation du muscle squelettique

Trois types musculaires sont présents dans notre organisme : le muscle lisse, le muscle cardiaque et le muscle squelettique. Bien que le gène DMD soit exprimé dans ces différents tissus, nous présentons la différenciation dans le muscle squelettique, principal tissu touché dans les dystrophies musculaires liées à une anomalie d'expression de la dystrophine et qui correspond au modèle cellulaire utilisé au cours de mes travaux de thèse pour étudier la régulation physiologique de l'épissage de DMD.

### I. Le muscle squelettique

Chez l'homme, le muscle squelettique est le tissu le plus abondant de l'organisme, puisqu'il représente environ 40% de la masse corporelle avec 640 muscles différents qui interviennent dans la motricité ainsi que le maintien de la posture du corps. A la différence des muscles cardiaque et lisse, les muscles squelettiques sont sous le contrôle du système nerveux central. Leurs contractions sont donc volontaires. Chaque muscle est composé de faisceaux de myofibrilles (ou myocytes) striées et allongées qui constituent la cellule musculaire fonctionnelle. Ces cellules multinucléées sont issues de la fusion de dizaines de myoblastes au cours du développement embryonnaire ou de la régénération musculaire.

Les noyaux aplatis, sont situés en périphérie sous la membrane plasmique, le sarcolemme, elle-même entourée par une membrane basale. En plus d'un réticulum endoplasmique et de différents éléments nécessaires à la production d'énergie (mitochondries, granules de glycogène) ou la fourniture d'oxygène (myoglobine), le cytoplasme (sarcoplasme) des myofibrilles comporte en son centre de très nombreuses myofibrilles. Ces longs myofilaments sont formés d'une succession d'unités contractiles, les sarcomères, disposés selon une organisation géométrique qui lui donne son aspect strié. Les sarcomères sont constitués de filaments fins d'actine associés à la tropomyosine et la troponine, et de filaments épais de myosine, orientés selon l'axe de la myofibrille. (Figure 50) (Chal and Pourquié, 2017; Frontera and Ochala, 2015). La titine est associée aux filaments épais de myosine et contribue à l'élasticité du sarcomère.

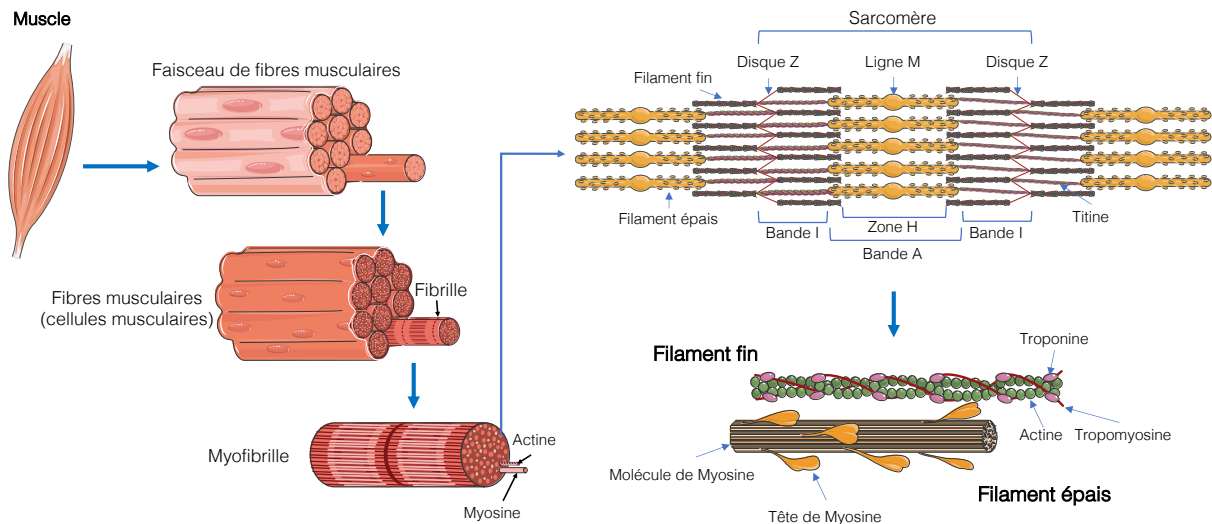


Figure 50. Structure tissulaire et cellulaire du muscle squelettique.

Le muscle squelettique est constitué de faisceaux de fibres musculaires qui représentent les cellules musculaires. Elles sont constituées de myofibrilles dans lesquels les sarcomères permettent la contraction musculaire grâce à la présence de filaments fins et épais d'actomyosine.

## A. La myogénèse

Les muscles squelettiques proviennent du mésoderme para-axial, un tissu qui se forme pendant la gastrulation à côté du tube neural. Le mésoderme para-axial naissant constitue le mésoderme pré-somitique à l'extrémité postérieure de l'embryon. Le mésoderme pré-somitique est un tissu transitoire qui possède une région antérieure engagée qui se segmente pour former les somites. Une population de cellules dérivées de somites qui expriment Pax3 et Pax7 est responsable de la croissance musculaire au cours du développement et donne naissance aux cellules satellites (CS) des muscles squelettiques postnataux. La myogénèse squelettique est initiée avec la spécification des progéniteurs pré-myogéniques et des myoblastes. Plusieurs phases de prolifération et de différenciation conduisent à la formation de myofibres multinucléés à partir de la fusion de myocytes mononucléés (Chal and Pourquié, 2017).

Les CS formées à partir des progéniteurs embryonnaires constituent une population de cellules souches non différenciées du muscle squelettique adulte. Elles restent dans un état quiescent, positionnées sous la lame basale des myofibres. Elles sont activées pour répondre à des besoins de régénération musculaire (en cas de lésion, de dégénérescence liée à des pathologies musculaires), de croissance ou de renouvellement (homéostasie).

Elles sont alors capables de se diviser de manière asymétrique permettant d'une part i) un auto-renouvellement pour reconstituer le réservoir de CS quiescentes en sortant une partie des cellules du cycle cellulaire et d'autre part ii) l'entrée dans le programme de différenciation myogénique de myoblastes qui vont après prolifération, fusionner avec des fibres musculaires

existantes ou en former de nouvelles (Dumont et al., 2015; Lepper and Fan, 2010; Nikonova et al., 2020).

## B. La différenciation musculaire

Les cellules satellites, formées à partir des progéniteurs embryonnaires, expriment le facteur de transcription Pax7, ainsi que son homologue Pax3 dans certains muscles. Les progéniteurs Pax7<sup>+</sup> assurent le développement musculaire jusqu'à la taille adulte du muscle. Pax7 est essentiel pour le maintien des cellules satellites dans un état indifférencié (Buckingham and Relaix, 2007), processus dans lequel la voie de signalisation Notch joue également un rôle déterminant en inhibant la différenciation des CS et en favorisant leur auto-renouvellement afin de maintenir l'homéostasie du tissu musculaire (Wen et al., 2012; Mayeuf and Relaix, 2011). La protéine Pax3 est essentielle pour la spécification myogénique en amont de MyoD (Chal and Pourquié, 2017).

Les cellules satellites sont en arrêt mitotique (phase G0 du cycle cellulaire). En réponse à différents signaux provenant du microenvironnement, les CS entrent dans le cycle cellulaire et prolifèrent. Pax3 est inhibé et Myf5 commence à s'exprimer (El Haddad et al., 2017). Les CS différenciées en myoblastes expriment Myf5, MyoD et Pax7, ce dernier étant inhibé avant l'entrée dans la phase de différenciation (Singh and Dilworth, 2013). MyoD active la myogénine (MyoG) entraînant la transition des myoblastes en myocytes mononucléés allongés et la sortie des cellules en phase G1 du cycle cellulaire (Chang and Rudnicki, 2014). Conjointement avec le facteur régulateur spécifique du muscle 4 (MRF4), les myocytes mononucléés vont ensuite fusionner en myotubes polynucléés afin de former les fibres musculaires. L'expression de MyoD et MyoG diminue alors afin de laisser place aux facteurs myogéniques plus tardifs MRF4 et MEF2 qui vont initier la différenciation terminale des cellules et induire l'expression des gènes spécifiques du muscle comme la myosine à chaîne longue (MyHC), la troponine T, ACTA1 et CKM (Figure 51) (Chal and Pourquié, 2017; Meyer et al., 2021; Mukund and Subramaniam, 2020; Nakka et al., 2018; Nikonova et al., 2020; Singh and Dilworth, 2013).

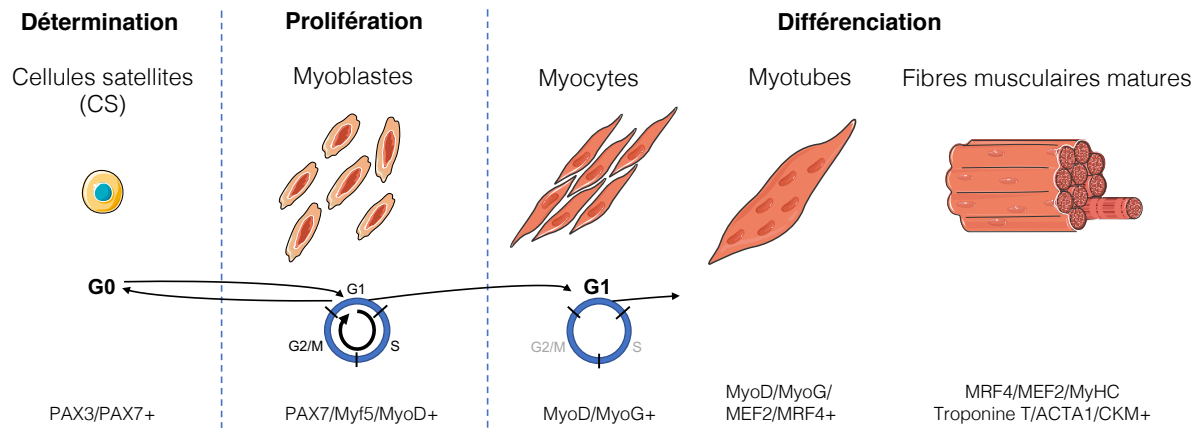


Figure 51. Myogenèse et différenciation du muscle squelettique.

Les cellules satellites (CS) quiescentes au stade G0 expriment PAX3 et PAX7. Lorsque les CS sont activées, elles entrent en état prolifératif sous forme de myoblastes qui prolifèrent et expriment Myf5 et MyoD. La sortie du cycle cellulaire en G1 ainsi que l'expression de MyoG, MEF2, MRF4, MyHC, Troponine T, ACTA1 et CKM permettent la différenciation. Les myoblastes fusionnent en myotubes multinucléés puis en fibres musculaires matures.

## II. Hélicases et différenciation musculaire

### A. UPF1

Les facteurs NMD jouent un rôle essentiel dans l'embryogenèse des mammifères. Le knock-out de presque chacun de ces gènes est létal au niveau embryonnaire. Leur expression, dont le niveau varie au cours du développement, est déterminant pour la spécification des feuillets embryonnaires (ectoderme, mésoderme et endoderme) (Nasif et al., 2018). Ainsi, des cellules souches embryonnaires de souris invalidées pour le gène *Smg6* ont pu être cultivées et ont proliféré, mais ne se différencient pas vers les trois types de feuillets (Li et al., 2015). Dans une autre étude, il a été montré que les hélicases UPF1 et UPF3B favorisent la différenciation des cellules souches embryonnaires humaines (hESCs) en mésoderme et inhibent leur différenciation en endoderme définitif ainsi que le niveau des ARNm codants pour les facteurs des voies de signalisation TGF-beta et BMP (Lou et al., 2016). Ces voies sont inversement régulées par les facteurs du NMD qui permettent la transition entre ces deux voies. Après la formation du mésendoderme (étape intermédiaire avant la formation des endoderme et mésoderme définitifs), la signalisation TGF-beta déclenche la différenciation de l'endoderme, tandis que la signalisation BMP, qui passe par l'activation de BMP4, permet la différenciation en mésoderme (Guo and Wang, 2009; C. Wang et al., 2015). L'analyse des niveaux d'expression des facteurs du NMD dans les 3 feuillets montre une forte expression de UPF1 et UPF3B dans le mésoderme comparé à l'endoderme dans lequel ils sont peu ou pas exprimés (Lou et al., 2016). Les expériences de surexpression transitoire ou de déplétion de UPF1 dans des cellules hESCs confirment le rôle de cette protéine dans l'activation de la voie de

signalisation BMP et la formation du mésoderme. On peut noter que l'expression des facteurs de l'EJC (eIF4A3, Y14 et MAGOH) est diminuée dans le mésoderme.

## 1. Balance NMD/SMD

Deux mécanismes de dégradation ciblée des ARNm peuvent être activés dans la cellule musculaire : le système du NMD classique et le SMD (*Staufen 1 (STAU1)-Mediated mRNA Decay*). Contrairement au NMD, le SMD n'est pas dépendant de la présence d'un PTC en amont d'un EJC mais cible des transcrits qui possèdent dans la région 3'UTR un site de liaison pour la protéine Staufen-1 (*SBS, STAU1-binding site*). Ce site est localisé à environ 20 à 25 nt en aval du codon STOP naturel de terminaison (Figure 52) (Kim et al., 2005).

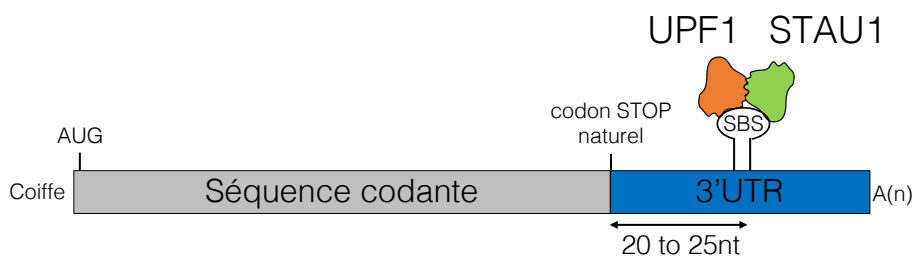


Figure 52. Interaction d'UPF1 et Staufen-1 dans le système SMD.

UPF1 est recruté par STAUI indépendamment d'un EJC et se lie au 3'UTR de l'ARNm au niveau d'un site de liaison de STAUI (SBS). Par analogie avec le NMD dépendant de l'EJC avec la présence d'un PTC, le SBS se trouve à 20-25 nt en aval du codon STOP normal afin d'induire le SMD.

La protéine UPF1 intervient dans les deux voies pour enclencher le processus de dégradation des ARNm *via* son interaction avec UPF2 (NMD) ou STAUI (SMD). Les sites de liaison de ces deux protéines sur UPF1 étant chevauchants, leur fixation est mutuellement exclusive comme démontré par des expériences de co-immunoprécipitation (Gong et al., 2009). Les voies de NMD et de SMD sont donc en compétition et l'intégration d'UPF1 dans l'un ou l'autre des systèmes permet son activation au détriment du second (Figure 53).

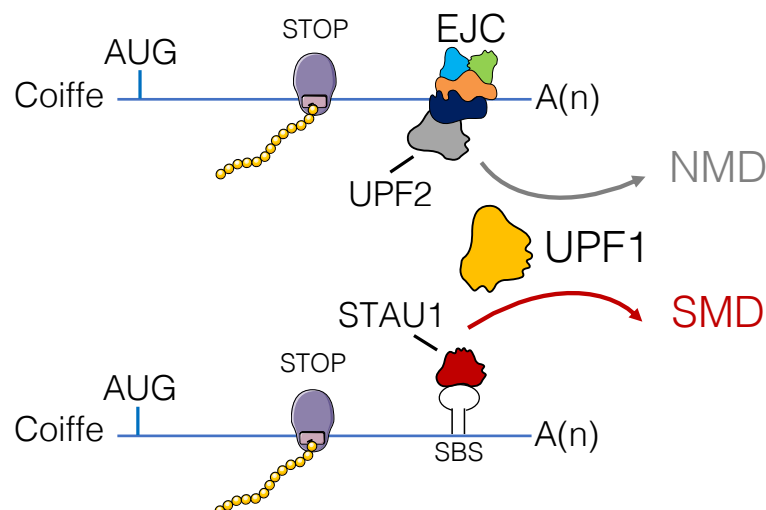




Figure 53. Balance NMD/SMD dépendant de UPF1.

UPF2, facteur de la voie classique du NMD et STAU1 qui induit la voie SMD sont en compétition pour lier UPF1 sur l'ARNm. UPF1 est présente dans les deux voies pour induire la dégradation de l'ARNm lorsque la traduction se termine suffisamment en amont d'un EJC, dans le cas du NMD, ou d'un SBS, dans le cas du SMD. Le codon STOP peut être prématuré ou normal.

Les voies du SMD et du NMD sont actives au cours de la myogenèse (Gong et al., 2009). Dans la lignée C2C12, la différenciation des myoblastes en myotubes s'accompagne d'une augmentation de l'efficacité du SMD au détriment du NMD. La baisse drastique d'expression de UPF2 dans les myotubes, contrairement au rapport STAU1/UPF1 qui reste relativement stable, peut expliquer cette baisse d'efficacité du NMD au cours de la différenciation myogénique. Cette transition entre NMD et SMD agit directement sur le répertoire de gènes exprimés dans chacune des conditions. Pour illustration, Gong et al. (2009) ont pu établir que le transcrite de PAX3, le facteur de maintenance des myoblastes comme vu précédemment, était une cible du SMD (dégradation dans les myotubes) alors que le facteur de différenciation Myogénine est une cible du NMD (dégradation dans les myoblastes).

De plus, la contribution d'un système NMD alternatif, insensible à la régulation négative de UPF2 au cours de la différenciation myogénique car dépendant de UPF3X, a été mise en évidence qui peut également entrer en compétition avec le SMD (Figure 54).

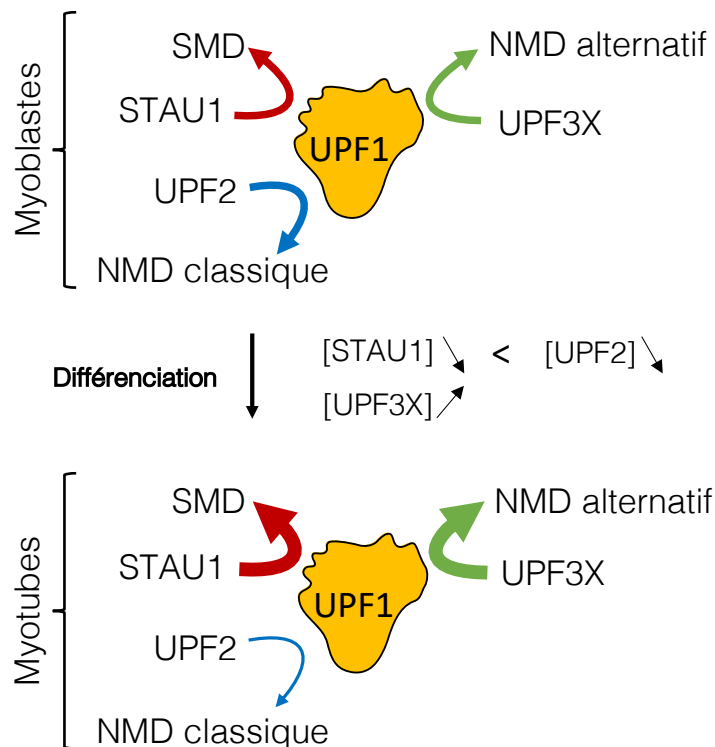


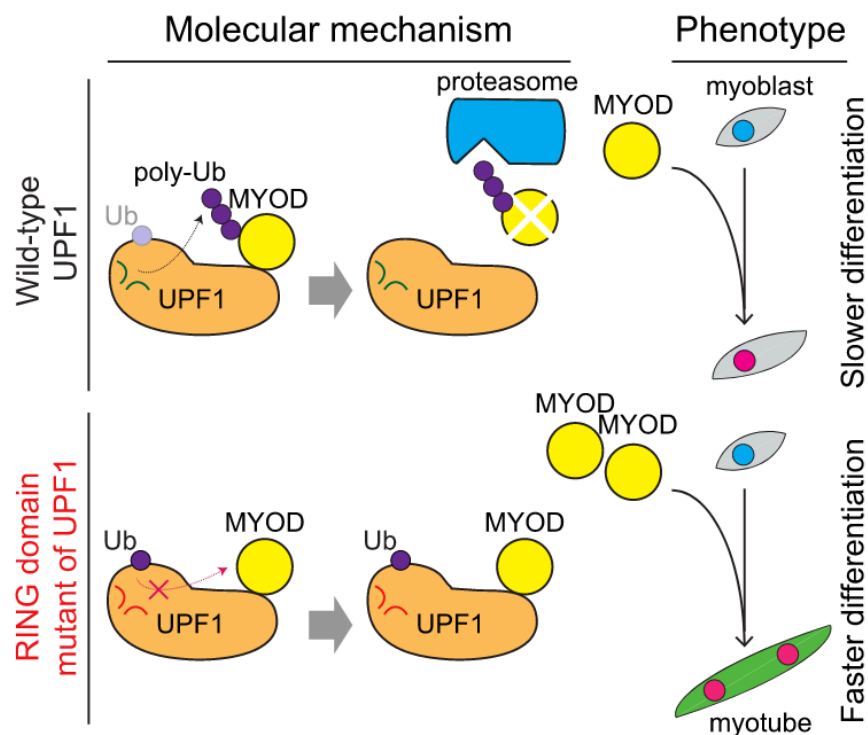
Figure 54. Balance des voies de NMD/SMD dans la différenciation musculaire.

Lors de la différenciation des myoblastes en myotubes, l'efficacité du SMD augmente, l'efficacité du NMD classique diminue, et l'efficacité de la voie alternative du NMD qui dépend de UPF3X (indépendante de UPF2) augmente. UPF2 et STAU1 sont en compétition pour interagir avec UPF1. Au cours de la myogenèse, la diminution importante du niveau de UPF2 favorise la liaison de STAU1 à UPF1 et l'augmentation du niveau de UPF3X renforce l'efficacité de la voie NMD alternative.

## 2. Interaction d'UPF1 avec MyoD

UPF1 est un répresseur de la différenciation musculaire dans les cellules humaines. Par son activité E3 ubiquitine ligase, UPF1 induit la dégradation par le protéasome de la protéine MyoD, un facteur intervenant au stade précoce de la différenciation myogénique, sans altérer le niveau de transcrit du gène (Feng et al., 2017a). Des mutations dans le domaine RING de UPF1 qui empêchent l'ubiquitination de MyoD, corrigent l'effet répressif de UPF1 sur la myogenèse. Des expériences de déplétion d'UPF1 dans des lignées ou des cellules primaires musculaires humaines montrent toutes un phénomène d'accélération de la différenciation (Figure 55). Les cellules mûrissent plus vite en myotubes (marquage positif de la chaîne lourde de Myosine, MyHC) par rapport aux cellules contrôles après le passage en milieu de différenciation. Dans ces conditions, les analyses transcriptomiques montrent que l'expression de MyoD ainsi que des gènes régulés par ce facteur de transcription sont augmentés.

Pour la première fois, un rôle direct d'UPF1 dans la myogenèse est montré grâce à la caractérisation de son action sur l'expression d'un facteur majeur de la différenciation des cellules musculaires. UPF1 peut non seulement agir sur les transcrits, substrats du NMD mais aussi catalyser l'ubiquitination et la dégradation de protéines comme illustré ici pour MYOD.



Feng et al., 2017

Figure 55. Mécanisme d'action de UPF1 sur l'expression de MyoD durant la myogenèse. L'activité ubiquitine E3 ligase (Ub) du domaine RING de UPF1 favorise la dégradation de la protéine MYOD par le protéasome dans les myoblastes (poly-Ub). La mutation du domaine RING de UPF1 stabilise la protéine MYOD, ce qui favorise la myogenèse.

Une étude récente réalisée chez le porc décrit un nouveau rôle de UPF1 dans la myogenèse en démontrant que la protéine facilite l'adhésion et la migration des cellules satellites non différenciées dans le muscle squelettique et pourrait ainsi être impliquée dans les premières étapes du processus de régénération musculaire (Tan et al., 2021).

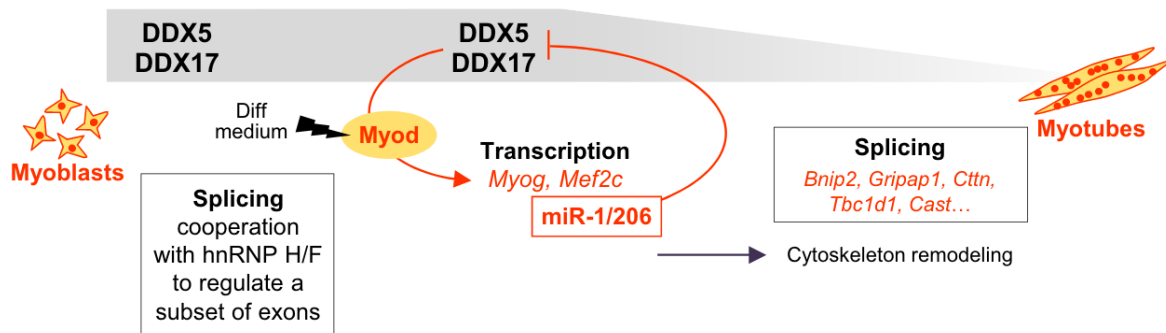
### 3. Les hélicases DDX5/17

Les hélicases DDX5 et DDX17 (ou p68 et p72) régulent l'expression des gènes en agissant à différents niveaux (régulation de facteurs de transcription, épissage alternatif ou expression de miRNA) notamment lors de la différenciation cellulaire (Bourgeois et al., 2016).

Dans les cellules musculaires, elles coopèrent avec les hnRNP H/F pour la mise en place de sous-programmes d'épissage spécifiques aux myoblastes en facilitant leur liaison aux régions hautement structurées, enrichies en quadruplex G, en aval des exons régulés (Dardenne et al., 2014). Elles agissent également de concert avec les deux facteurs d'épissage spécifiques du muscle, MBNL et RBM4, dont l'épissage est aussi modifié en l'absence des DDX5/17.

Une diminution du niveau d'expression par ARNi des DDX5/17 dans des myoblastes en culture entraîne un défaut d'expression du facteur de transcription MyoD et une différenciation incomplète (Caretta et al., 2006). En effet, DDX5 et DDX17 corégulent l'expression de ce facteur de transcription qui est directement impliqué dans l'expression de facteurs myogéniques comme la myogénine ou MEF2C et des miARN musculaires, miR-1 et miR-206 (Figure 56) (Caretta et al., 2006; Dardenne et al., 2014). Au cours de la différenciation des myoblastes en myotubes, l'activation des miARN induit une boucle de rétrocontrôle négatif de l'expression des DDX5 et DDX17.

Ainsi, les DDX5 et DDX17 contribuent au maintien d'épissages spécifiques des myoblastes. Elles sont également nécessaires à l'initiation d'un programme de transcription spécifique de la différenciation myogénique au cours de laquelle leur diminution d'expression va favoriser la transition vers un programme d'épissage spécifique des myotubes (Bland et al., 2010; Bourgeois et al., 2016).



Adaptée de Dardenne et al., 2014

Figure 56. Rôle de DDX5/17 dans la différenciation musculaire.

Dans les myoblastes, DDX5 et DDX17 coopèrent avec hnRNP H/F pour contrôler l'épissage d'un sous-ensemble spécifique d'exons. Lors de la différenciation, DDX5 et DDX17 corégulent l'activité transcriptionnelle dépendante de MyoD, contrôlant directement les effecteurs de la différenciation musculaire MyoG et Mef2c. Les miRNAs musculaires (miR1 et miR206) répriment à leur tour directement l'expression des deux hélicases et contribuent aux changements phénotypiques qui se produisent pendant la différenciation.

#### 4. L'EJC dans la différenciation musculaire

Le rôle du complexe EJC, facteur essentiel à l'activation du système NMD, dans la différenciation musculaire n'est pas décrit. Cependant, il a été décrit que MLN51 interagirait avec Staufen-1 (Macchi et al., 2003). Cette étude se concentrait principalement sur la localisation des ARNm médiée par Staufen-1 se liant à MLN51 dans les mRNP des neurones. Il pourrait être d'intérêt d'étudier l'interaction de MLN51 avec Staufen-1 au cours de la différenciation musculaire afin d'établir si l'EJC, et en particulier MLN51 connue pour son rôle dans le NMD, pourrait interagir avec le système SMD dans le muscle squelettique.

## Objectifs de la thèse

L'épissage du gène *DMD* a beaucoup été étudié par la caractérisation des différentes isoformes produites par épissage alternatif, notamment dans le SNC. Dans le muscle squelettique très peu de données existent et sont plutôt anciennes. L'étude de la régulation de l'épissage du gène présente un grand intérêt pour une meilleure compréhension des mécanismes physiopathologiques impliqués dans les dystrophinopathies, et dans le cadre des stratégies thérapeutiques actuelles développées pour la DMD basées sur la redirection de l'épissage (Aartsma-Rus et al., 2017) .

Avant mon arrivée au sein du laboratoire, l'équipe a conçu une stratégie appropriée pour l'étude du profil d'épissage du transcrit Dp427m dans le muscle qui s'appuie sur un protocole de séquençage de l'ARN (RNA-Seq) ciblant le transcrit Dp427m Cette stratégie a été choisie afin d'obtenir un taux de couverture suffisant pour l'analyse et la quantification des épissages alternatifs dans ce transcrit faiblement exprimé dans le muscle squelettique humain.

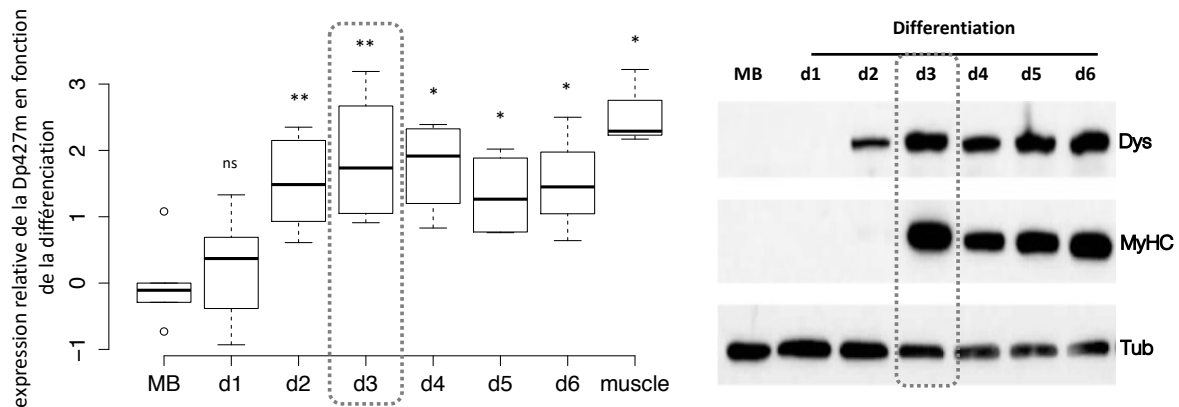
Le protocole expérimental repose sur l'amplification de la totalité de la séquence codante (11.3 kb) par PCR Long-Range (LR-PCR) à l'aide d'amorces spécifiques de l'isoforme situées dans les exons 1 et 79 à partir d'ADNc synthétisés en présence d'oligo(dT). Le séquençage des amplicons a été réalisé sur séquenceur 454-GS-Junior. Après alignement des reads sur le génome hg19 avec l'outil STAR, le logiciel IPSA (*Integrative Pipeline for Splicing Analyses*) (Pervouchine et al., 2013) a été utilisé pour analyser le nombre de reads aux jonctions et calculer le niveau d'inclusion des exons **PSI** (*Percent Splice In*).

L'analyse suivant ce protocole de quatre échantillons d'ARN provenant d'individus sains (hommes de 25 ans) a permis d'établir une référence du profil d'épissage de l'isoforme Dp427m dans le muscle squelettique adulte humain (Bougé et al., 2017). Un fort niveau d'inclusion ( $\geq 97\%$ ) des 79 exons a été observé, quelle que soit leur taille, leur position dans le transcrit, la force de leurs sites d'épissage ou la longueur des introns environnants. Un faible niveau basal d'exclusion des exons épissés alternativement dans les isoformes cérébrales de la dystrophine (Dp427c ou Dp71) a été détecté, illustrant l'existence d'une régulation spécifique de l'épissage dans le muscle pour permettre l'expression d'une isoforme complète de dystrophine.

Afin d'étudier les mécanismes moléculaires mis en jeu dans cette régulation, l'équipe a utilisé un modèle de lignée cellulaire musculaire humaine normale (C25C148), issu de l'immortalisation de cellules primaires d'individus sains de 25 ans (Mamchaoui et al., 2011). Cette lignée immortalisée conserve un programme myogénique comparable aux cellules

primaires et ne présente pas de défauts sur d'autres processus cellulaires, ce qui en fait un bon modèle cellulaire pour l'étude des gènes musculaires (Thorley et al., 2016).

L'équipe a caractérisé l'expression du gène *DMD* dans la lignée C25C148 au cours de la différenciation des myoblastes et a défini que l'expression optimale des transcrits Dp427m et de la dystrophine musculaire s'établissait à 3 jours de différenciation (Miro et al., 2020) (Figure 57).



Miro et al., 2020

Figure 57. Analyse de l'expression de la Dp427m dans les cellules C25C148 au cours de la différenciation myogénique. Expression des ARNm de la Dp427m dans les C25C148 pendant la différenciation (jour 1 à jour 6) par rapport aux cellules indifférenciées (myoblastes, MB) et au muscle squelettique (panel de gauche) et expression des niveaux protéiques de Dp427m (panel de droite).

Le profil d'épissage des cellules à 3 jours de différenciation a été comparé par RNA-Seq ciblé au muscle squelettique (Miro et al., 2020). La lignée immortalisée montre un profil d'épissage globalement comparable au muscle squelettique, sans apparition de nouveaux évènements majeurs. La seule différence se situe dans le niveau plus élevé des exons alternatifs connus (pseudo exon 1a (PE1A), exons 71 et 78) (Figure 58). Le pseudo exon 1a, issu de l'intron 1, a été initialement identifié dans les lymphocytes où il est inclus dans plus de 50% des transcrits DMD dits illégitimes (Roberts et al., 1993). L'inclusion de cet exon supplémentaire à proximité du promoteur pourrait représenter un contrôle post-transcriptionnel dans les cellules qui n'expriment normalement pas la protéine dystrophine (comme les lymphocytes), alors qu'un faible taux d'inclusion est observé dans les transcrits du muscle squelettique (3%).

Alternative Splicing Event (ASE)	C25C148 d3-diff (%)	skeletal muscle <sup>a</sup> (%)	predicted impact on the reading frame
del E71	37.32	1.80	in frame
del E78	27.30	2.48	elongated C-term
PE1a	12.69	3.06	out of frame
del E38	4.93	ND	In frame
del E37	2.97	0	In frame
del E6-41	2.18	0	in frame
3'ss exon 54 (-3 nt)	1.75	1.12	in frame
del E9	1.55	0.63	in frame
del E2	1.01	ND	out of frame
PE51X	0.56	0.21	in frame
del E71-74	0.42	ND	in frame
del E39	0.43	ND	in frame
3'ss exon 78 (-4 nt)	0.26	0.43	out of frame
del E17-44	0.33	0	in frame
del E41	0.38	ND	in frame
del E73	0.28	0.32	in frame
del E19-44	0.27	0	in frame
del E56-78	0.26	0	in frame
3'ss exon 76 (-60 nt)	0.32	0.33	in frame
del E68	0.25	ND	out of frame
PE12X	0.25	0	in frame
PE2a	0.12	ND	out of frame

Bougé et al., 2017 ; Miro et al., 2020

Figure 58. Validation du modèle cellulaire C25C148 comparé au muscle squelettique.

Liste des ASEs par ordre décroissant du pourcentage détecté par RNA-Seq dans les cellules C25C148 au jour 3 de la différenciation (d3-diff) et leurs valeurs respectives dans le muscle squelettique.

La réalisation d'un crible fonctionnel par ARNi ciblant 16 RBP dans les C25C148 a permis d'établir pour la première fois le rôle de facteurs d'épissage généraux (SRSF1, SRSF2, hTra2B, hnRNPA1, hnRNPA2/B1, DAZAP1 et TDP-43), d'hélicases spécifiques du programme myogénique (DDX5/17) et de protéines spécifiques du muscle (MBNL1, CELF1, RBFOX1, RBFOX2, RBM4, PTBP1 et QKI-5) dans l'épissage du transcrit Dp427m. La transfection des si-ARN a été effectuée dans les cellules en prolifération (stade myoblaste) et les ARN totaux ont été extraits après 3 jours de différenciation. Le RNA-Seq ciblé de la Dp427m réalisé dans chacune des conditions a révélé la contribution de plusieurs des RBP testées dans l'épissage des exons 9, 37, 38, 41, 71, 74, 78 et du pseudo exon (Miro et al., 2020).

Bien que je n'ai pas participé aux expériences et à la mise en place des analyses bio-informatiques du RNA-Seq ciblé de Dp427m, mon arrivée au laboratoire en 2018 m'a permis de m'impliquer dans l'analyse des résultats de RNA-Seq ciblé (notamment celles en relation avec mon projet de thèse) et de figurer parmi les auteurs de l'article de Miro et al., en 2020 (Annexe 1).

Lors de la réalisation du crible, l'équipe avait initialement testé les facteurs centraux de l'EJC (eIF4A3 et Y14) en raison du rôle récemment décrit du complexe dans l'épissage, notamment dans des gènes possédants de longs introns avec des exons alternatifs et/ou récursifs, dans la

rétention de petits introns, ou encore dans la suppression des sites cryptiques d'épissage 5' et 3' en association avec ses complexes associés ASAP/PSAP (Blazquez et al., 2018; Boehm et al., 2018; Fukumura et al., 2016; Hayashi et al., 2014; Schlautmann and Gehring, 2020; Wang et al., 2018, 2014).

Mon projet de thèse s'est orienté vers l'étude du rôle du complexe EJC dans la régulation de l'épissage du gène *DMD* sur la base des données préliminaires, non publiées, obtenues par l'équipe.

**Le premier objectif** de ma thèse consistait à déterminer et à valider l'impact de la diminution d'expression des facteurs de l'EJC (eIF4A3, Y14) sur l'épissage des exons du transcrit Dp427m dans les C25C148 différenciées. En raison du manque d'efficacité dans les cellules musculaires, les transfections de vecteurs d'expression en condition de KD d'eIF4A3 et Y14 ont été réalisées dans un modèle cellulaire différent (HeLa) qui exprime l'isoforme Dp71 (Figure 59). Cela nous a permis d'étendre nos observations initiales à d'autres isoformes exprimées par le gène.

**Le deuxième objectif** était d'étudier l'impact des protéines associées aux complexes EJC dans la régulation de l'épissage des transcrits Dp427m dans les C25C148. Notamment, les complexes ASAP/PSAP qui sont décrits comme régulant les sites cryptiques d'épissage de nombreux gènes (Boehm et al., 2018; Wang et al., 2018).

**Le dernier objectif** de ma thèse visait à déterminer l'impact de la déplétion des facteurs de l'EJC et du système NMD sur l'expression de la dystrophine et la différenciation myogénique.

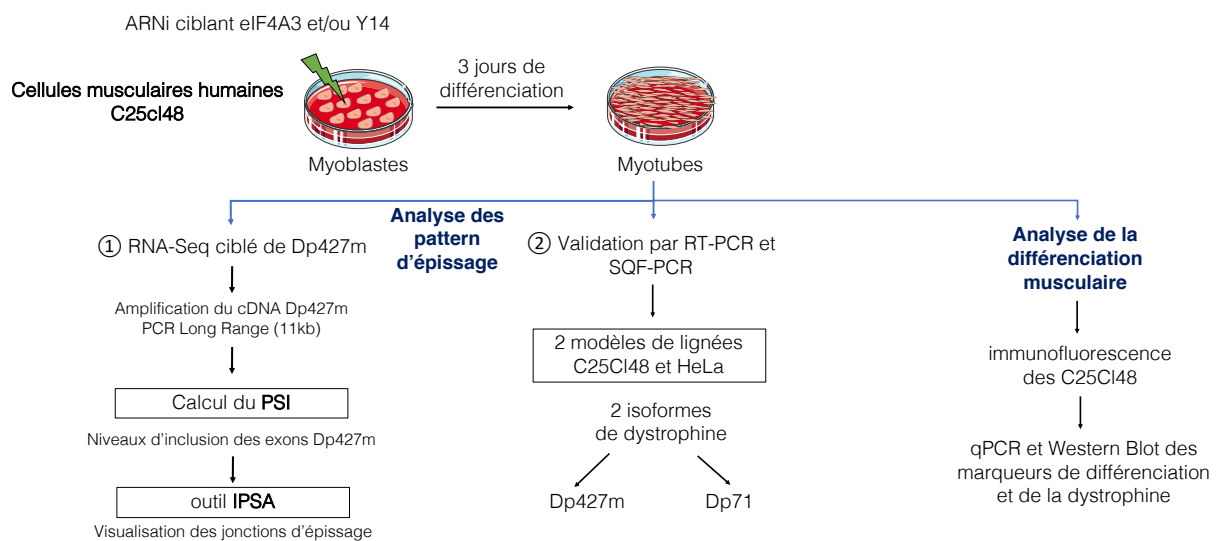


Figure 59. Méthodologie du projet de thèse.



## Principaux résultats obtenus durant la thèse

- Le complexe EJC régule l'épissage de l'exon 9 et d'exons dans la partie distale de l'isoforme Dp427m et Dp71. La déplétion d'eIF4A3 et Y14 entraîne une cascade de dérégulations incluant des sauts d'exon simples ou multiples (exons 68, 69, 69-70, 69-71, 72-74, 71-74, 73-74 et 78), la rétention de l'intron 70 et l'activation de sites donneurs d'épissage cryptiques dans l'exon 9 et 70.
- La déplétion d'eIF4A3 a un effet plus fort sur le niveau des épissages aberrants détectés dans l'isoforme Dp427m dans les C25C148 que la déplétion d'Y14 contrairement aux cellules HeLa qui montrent un effet similaire de la déplétion des deux facteurs dans la Dp71.
- La déplétion du facteur central du système NMD (UPF1) nous a permis de démontrer que les épissages observés en condition de déplétion d'eIF4A3 et/ou Y14 ne sont pas dus à la diminution de l'efficacité du système NMD et la stabilisation des transcrits possédant un PTC.
- Des expériences de sauvetage, avec un vecteur exprimant eIF4A3 ou Y14 de type sauvage et avec un vecteur eIF4A3 mutant pour la formation du complexe EJC, montrent qu'eIF4A3 et Y14 agissent indépendamment sur l'épissage du gène *DMD*. De plus, la formation du complexe EJC est nécessaire à cette régulation.
- La protéine MLN51 qui se lie à l'EJC après l'épissage et les protéines ACINUS, PININ, RNPS1, SAP18 des complexes ASAP/PSAP, régulent seulement l'épissage des deux plus petits exons 71 et 78 du gène *DMD*.
- Les déplétions des facteurs de l'EJC entraînent une diminution de l'expression de la dystrophine ainsi qu'une diminution des niveaux des facteurs de différenciation musculaires, montrant un arrêt de l'engagement dans la différenciation des myotubes. En comparaison, la diminution d'UPF1 accélère le processus de différenciation.

# Résultats

Cette partie de mon manuscrit intègre l'ensemble des résultats obtenus durant mon projet de thèse sur le rôle de l'EJC et de ses protéines dans la régulation de l'épissage du pré-ARNm du gène *DMD* ainsi que dans la différenciation myogénique. Elle comprend : (i) une première partie rédigée en anglais sous la forme d'un article scientifique correspondant au matériel et méthodes et résultats d'un article en préparation, (ii) des résultats complémentaires au projet concernant le rôle de l'EJC et de ses protéines associées dans l'épissage du transcrit *DMD* et dans la différenciation musculaire qui nous ont paru intéressants de discuter mais non aboutis pour la publication de l'article.

## Materials and methods

### Cell culture

The human immortalized myoblast cell line C25C148, established from a muscle biopsy of a 25-years-old healthy individual (Mamchaoui et al., 2011; Thorley et al., 2016), was cultured in proliferation medium as previously described (Miro et al., 2015). Myogenic differentiation was induced for 3 days by serum deprivation and elevation of insulin concentration to 10  $\mu\text{g}/\text{mL}$  in the medium. HeLa cells were grown in DMEM/F-12 medium (Gibco) containing 10% foetal bovine serum (Eurobio), 1% of L-glutamine (Gibco) and 1% of non-essential amino-acids (Gibco). All cells were cultivated at 37°C and 5% CO<sub>2</sub> in a humidified incubator.

### siRNA knockdown

For *DMD*-targeted RNA-Seq experiments, siRNA transfections were performed in C25C148 cells as previously described (Miro et al., 2020). Briefly, cells were reverse transfected with the eIF4A3 or Y14 siRNA (Dharmacon, Lafayette, CO, USA) at a final concentration of 37.5 nM in 10 cm cell culture dishes using 44.25  $\mu\text{L}$  Lipofectamine RNAiMAX (Thermo Fisher Scientific) according to the manufacturer's instructions. After 48 hours, a second round of siRNA transfection was performed and the cells were placed in differentiation medium for 3 days. The siGENOME non-targeting siRNA#5 (Dharmacon) was used as a negative control. For others siRNA experiments targeting EJC, ASAP/PSAP and NMD components (from Dharmacon or Eurofins Genomics, Ebersberg, Germany), a single round of reverse transfection was performed.  $1.5 \times 10^5$  C25C148 cells were plated in six well-plates. Cells were harvested in Trizol for RNA analysis and in Laemmli sample buffer supplemented with 5% of  $\beta$ -mercaptoethanol or RIPA buffer (10 mM Tris-HCl [pH 7.4], 150 mM NaCl, 1% NP-40, 1 mM EDTA, 0.1% SDS, 1 mM dithiothreitol) when specified, for protein analysis. The knockdown

(KD) efficiency was assessed by RT-qPCR and Western blot analysis. The siRNA target sequences are listed in the Suppl Table S3.

### **DMD-targeted RNA-Seq experiments and computational analyses**

DMD-targeted RNA-seq was performed on four biological replicates for control and duplicates for eIF4A3 and Y14 KD, as previously described (Bougé et al., 2017; Miro et al., 2020). The full-length DMD cDNA produced by priming with oligo(dT) (Superscript II, Thermofisher, scientific) was amplified as a single long fragment (11.3 kb) by long-range PCR (GoTaq® Long PCR Master Mix, Promega) using primers located in exon 1 of the muscle isoform (Dp427m) (5'-CTTTCCCCCTACAGGACTCAG-3') and in the 3'UTR (5'-CCAAATCATCTGCCATGTGG-3'). The libraries were prepared from 1.6 ug of purified LR-PCR products using the Rapid Library Preparation Kit (Roche, Basel, Switzerland) and multiplexed libraries were sequenced at a minimum read depth of 400X on the Roche GS Junior 454 sequencer. After adapter and quality-based trimming, cleaned reads of mean length of 260 bp were aligned against the GRCh37/hg19 human X chromosome reference sequence using the STAR read aligner (v. 2.3) (Dobin et al., 2013). The number of split reads supporting exon inclusion and exclusion in control and KD conditions was used by the Integrative Pipeline for Splicing Analyses (IPSA) for calculation of exon-centric and intron-centric splicing metrics. The inclusion of the 79 annotated exons of the Dp427m isoform was assessed by using the exon centric metrics (Percent-Spliced-In, or PSI ratio). Average PSI values were compared between the different experimental conditions (control, eIF4A3 and Y14 KD) in Supplementary Table S1A.

The intron centric metrics from the IPSA calculation allowed quantification of complex splicing events as it is based on the reads mapped to splice junctions (SJs) and it measures independently the splicing occurring at the 5' and 3' end of the intron (psi5 and psi3, respectively) (Pervouchine et al., 2019, 2013). The differential usage of novel, alternative splice junctions (SJs) in eIF4A3 or Y14 KD conditions compared to the control condition was calculated using  $\Delta SJ = SJ(KD) - SJ(CTRL)$  as previously described (Miro et al., 2020). Novel junctions supported by at least 10 reads in both replicates were considered. We report all alternative splicing events (ASEs) with a variation of 5% or 1% in Supplementary Table S1B. Alternative splicing sashimi plots were generated from replicate means of each condition using the ggsashimi tool (Garrido-Martín et al., 2018).

### **Reverse transcription and real-time quantitative PCR**

Total RNA was isolated from cells using Direct-zol RNA MiniPrep (Ozyme, Zymo Research) according to the manufacturer's instructions. Reverse transcription (RT) was performed on 300ng total RNA using 300 ng of random hexamer primers and SuperScript III reverse transcriptase (Life Technologies) following the manufacturer's conditions. For quantitative PCR (qPCR), 1  $\mu$ L of 1:6 diluted cDNA was amplified in triplicate technical repeats for four biological replicates using the SYBR Green I Master Mix and LightCycler 480 (Roche) in a final volume of 5  $\mu$ L containing 0.5  $\mu$ M of primers targeting the proteins of interest (Suppl Table S3). Amplified transcripts were quantified using the comparative CT method normalized to RPLP0 expression (Hs\_RPLP0\_2\_SG QuantiTect Primer Assay, Qiagen, Courtaboeuf, France). The results are presented as mean  $\pm$  SD of normalized fold expression changes ( $2^{-\Delta\Delta CT}$ ) or expression levels ( $2^{-\Delta CT}$ ) for dystrophin and differentiation factors expression.

### **End-point PCR and semi-quantitative fluorescent PCR (SQF-PCR)**

All PCR experiments involved four independent biological replicates. For conventional end-point PCR, primers in exons flanking the alternative splicing event (Suppl Table S3), were used to amplify 1  $\mu$ L of cDNA with the Taq DNA polymerase (New England Biolabs) for 35 cycles in a 25  $\mu$ L total volume. RT-PCR products were resolved by gel electrophoresis in 2% agarose gels on ChemiDoc <sup>TM</sup> XRS+ System (Bio-Rad). Bands were quantified using the Image Lab 6.0.1 software (Bio-Rad). The identity of the amplified products was verified by Sanger sequencing (Suppl Fig. S1D). For semi-quantitative analysis of fluorescently marked PCR products, the same protocol was applied except that FAM-labeled forward primers were used and the cycling conditions were as follows: 94°C for 3 min followed by 26 cycles at 94°C for 30 s, 59°C for 30 s and 72°C for 45 s and a final extension step at 72°C for 20 min. Fluorescently labeled PCR fragments were sized and quantified by capillary electrophoresis using 3500 Dx Genetic Analyzer equipment and GeneMapper v6.0 software (Applied Biosystems, Foster City, USA). Ratios of splicing isoforms were determined as the peak area of one specific isoform divided by the total peak areas for the other detected isoforms. Primer sequences for RT-PCR are provided in Supplementary Table S3.

### **siRNA rescue experiments**

HeLa cells were reverse transfected in 6-well plates ( $1.8 \times 10^5$  cells/well) with control, eIFA43 and Y14 siRNAs at a final concentration of 37.5 nM with 7.5  $\mu$ l Lipofectamine RNAiMAX (Thermo Fisher Scientific) and were transfected 24 hr later with 1  $\mu$ g of plasmids expressing

FLAG-tagged proteins or empty vector (negative control) using 6ul Lipofectamine® LTX (Life Technologies) according to the manufacturer's protocol. The siRNA-resistant plasmids for the wild-type eIF4A3 protein (FLAG-eIF4A3-WT) induced by site-directed mutagenesis for regions targeted by siRNA eIF4A3 and the mutant form (D401KE402R) that does not form EJC (FLAG-eIF4A3-Mut) are gifts from H. Le Hir (Wang et al., 2014). The FLAG-Y14-WT vector, built in our laboratory, was created by cloning into the NheI/XhoI sites of a Flag-pcDNA3 plasmid, the PCR amplified coding region of Y14 using a Taq Polymerase (New England Biolabs) (RT using 300 ng of random hexamer primers from 500 ng of HeLa RNA) and the following primers: Flag-pcDNA3-NheI-Y14-WT-F (5'-CATCGCTAGCATGGATTACAAGGACGACGACGATAAGGCGGACGTGCTAGATCTTCAC-3') and Flag-pcDNA3-XhoI-Y14-WT-R (5'-TCGACTCGAGTCAGCGACGTCTCCGGTCTGG-3'). To make the Y14 mRNA resistant to the siRNA, four silent third-codon point mutations were introduced in the target region of the siRNA using the Quick-Change site-directed mutagenesis (Agilent) kit and the following primers: 5'-GGGGTGGATTCTCTTTGTAAGTGGAGTCC-3' (forward) and 5'-TCCACCGAGCGTTGTGGTCCGGGTTTCATC-3' (reverse). All constructs were verified by Sanger sequencing. Cells were harvested 24 hr post-transfections of the plasmids. Expression of the tagged protein was checked by western blot analysis and purified RNA was used for reverse transcription.

### Western blot analysis

After ultra-sonication of cell lysates, equal amount of total proteins were separated on 10% or 4–15% precast SDS–Polyacrylamide gels (Thermo Fisher Scientific) depending on the protein probed. Proteins were transferred onto Polyvinylidene fluoride (PVDF) or nitrocellulose membranes using Trans-Blot Turbo Transfer System (Biorad) and probed with specific antibodies against eIF4A3, Y14, MLN51, RNPS1, SAP18, ACINUS, PININ, UPF1, dystrophin, myosin heavy chain (MyHC) and Troponin T. Tubulin or GAPDH was used as loading controls. The origin of the antibodies and dilutions used are described in Supplementary Table S3. After washing, membranes were probed with secondary antibodies either conjugated to horseradish peroxidase (HRP) or labeled with LI-COR IRDye fluorescent dye (ThermoFisher). Protein signals were detected on an Azure Sapphire™ imager (Azure Biosystems, Dublin, United States) using Enhanced chemiluminescence ECL reagents (Pierce, Rockford, IL, USA) or visualized by an Odyssey infrared imaging system (LI-COR Biosciences, Lincoln, NE; Cooper, Lo, & North, 2003). Western Blotting images were collected

and stacked on ImageJ software (v2.0.0) and the densitometry was performed using Image Lab 6.0.1 software (Bio-Rad).

### **Immunofluorescence microscopy and myotubes analysis**

Three-day differentiated C25Cl48 cells were fixed with 2% paraformaldehyde (PFA; Sigma-Aldrich) for 5 min at room temperature, washed and permeabilized in PBS/0.25% Triton X-100 for 5 min at room temperature (RT). After washing, cells were incubated with mouse monoclonal anti-troponin T (1/200; Sigma-Aldrich, St. Quentin Fallavier, France) antibody at room temperature for 1 hr, then washed for three times and incubated with the Alexa 555-conjugated anti-mouse secondary antibody (1/1000; Fisher Scientific, Illkirch, France) for 30 min at room temperature. A final concentration of 0.1 mg/ml DAPI (Sigma) was included to stain nuclei. Images (eight fields for each condition) were acquired with a Zeiss Axio imager M1 epifluorescence microscope (Carl Zeiss SAS-Microscopy, Marly le Roi, France), with 10X objective and 1388 x 1040 pixels. Myotube morphology analyses were performed using the ImageJ software as recently described for five random fields (Meyer et al., 2021). DAPI-positive cells were analysed to determine the number of nuclei per image, normalized to the total image area to compare all conditions. The nuclear fusion index was calculated as the number of nuclei incorporated into troponin T-positive myotubes (with at least four nuclei) expressed as a percentage of the total visible nuclei in the image frame. The average surface area of troponin T-stained myotubes was calculated as the mean of five measurements taken along the myotube length for a representative measure.

### **Statistical analyses**

All the data were analyzed with GraphPad Prism v9.4.1 (Graph-Pad Software, Inc., La Jolla, USA) software and expressed as the mean  $\pm$  SD. Sample number (n) indicates the number of independent biological samples in each experiment. Groups were compared with unpaired non-parametric multiple Mann-Whitney tests for all bar graphs shown in this study. Differences in means were considered statistically significant at  $p \leq 0.05$ . For all graphs, n.s. denotes not significant, \*  $p \leq 0.05$ , \*\*  $p \leq 0.01$ , \*\*\*  $p \leq 0.001$ . For correlations, a simple linear regression between two groups was performed to obtain the coefficients of correlation R squared (value  $>0.7$  is considered as “strong” correlation,  $0.70 >$  “right”  $>0.50 >$  “moderate”  $>0.3 >$  “poor” correlation (Hazra and Gogtay, 2016).

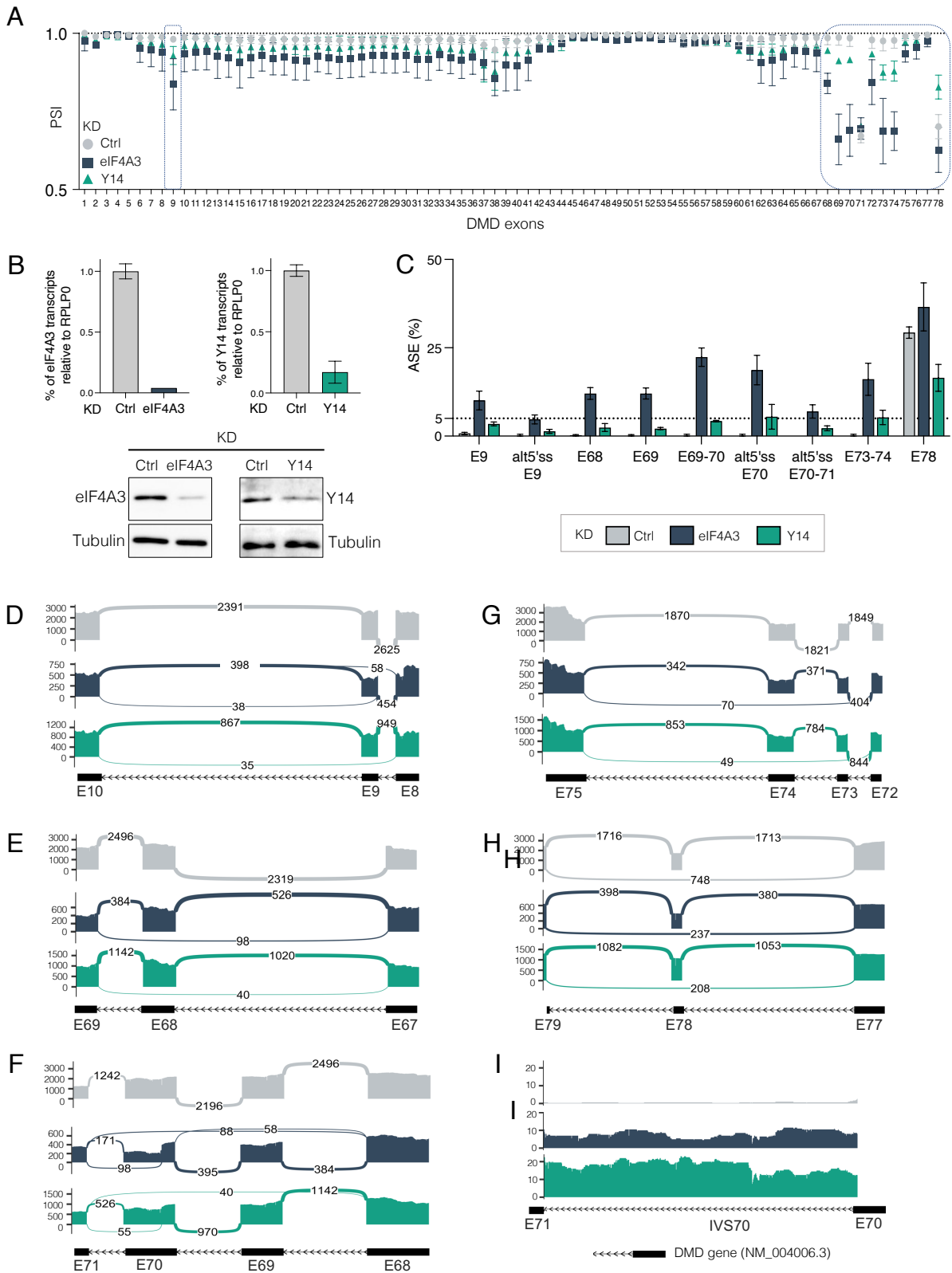
## Results

### Depletion of eIF4A3 or Y14 affects splicing of the 3' end of the Dp427m transcript

In order to determine whether the EJC plays a role in the regulation of splicing of the long DMD pre-mRNA transcript, we analyzed the splicing changes induced by depletion of the two endogenous eIF4A3 or Y14 EJC core components using a deep sequencing approach (Fig. 60A). Proliferating human C25C148 muscular cell line was transfected with one rounds of control or specific eIF4A3 and Y14 validated siRNA followed by 3 days of differentiation in order to obtain expression of Dp427m dystrophin at both the mRNA and protein level (Miro et al., 2020). The knockdown (KD) efficiency was verified by RT-qPCR and western blotting (Fig. 60B). Two independent biological replicates were analyzed in parallel to four control conditions by *DMD*-targeted RNA-seq as previously described (Bougé et al., 2017; Miro et al., 2020). This approach of 454 amplicon sequencing of the complete cDNA (11.3 kb) amplified as a single long fragment by long-range PCR allows in-depth analysis of splicing changes in the lowly expressed *DMD* gene. Exon-inclusion levels, given as percent spliced-in (PSI) values, were computed for all the 79 annotated exons of the Dp427m isoform using IPSA software (Fig.60A) (Pervouchine et al., 2019, 2013). The PSI profile showed that, with the exception of exon 9, eIF4A3 and Y14 KD selectively affects the inclusion of a cluster of exons at the 3' end of the transcript (from exon 68 to 78) compared to the control condition (si-Ctrl) (Fig.60A, Suppl Table S1A). The inclusion level of exon 71, which has a basal level of alternative splicing in the cell line, was apparently not changed in the depleted conditions. In order to identify putative novel alternative junctions and the Alternative Splicing Events (ASEs) leading to decreased exon-inclusion, we performed a differential splice junction (SJ) usage analysis between EJC KD and control condition as previously described (Miro et al., 2020). By selecting a change in junction usage above 5% ( $|\Delta SJ| \geq 0.05$ ) in the eIF4A3 KD condition that led to more significant changes, nine ASEs were identified consisting in single or double exon skipping and the use of an alternative 5' splice site (alt5'ss) in exon 9 and in exon 70 (Fig. 60C; Suppl Table S1B). A similar set of deregulated exons was observed upon depletion of Y14 but to a much lesser extent, which was unexpected due to the role of the two core EJC factors for stable EJC formation. Overall, the changes were in the same direction (*ie* more exon skipping and use of alt5'ss), except for exon 78, which showed marked re-inclusion in Y14 KD condition. The splicing patterns of ASEs and junction reads are depicted in Sashimi plots in Fig. 60D-H. Although, the amplicon-based protocol used for *DMD*-targeted RNA-seq is not suitable for detecting retention of large introns, it can reveal retention of small introns. The



analysis of read coverage of the 9 introns of less than 1 kb in the transcript revealed a low level of intron 70 retention (698bp) (Fig. 60I).



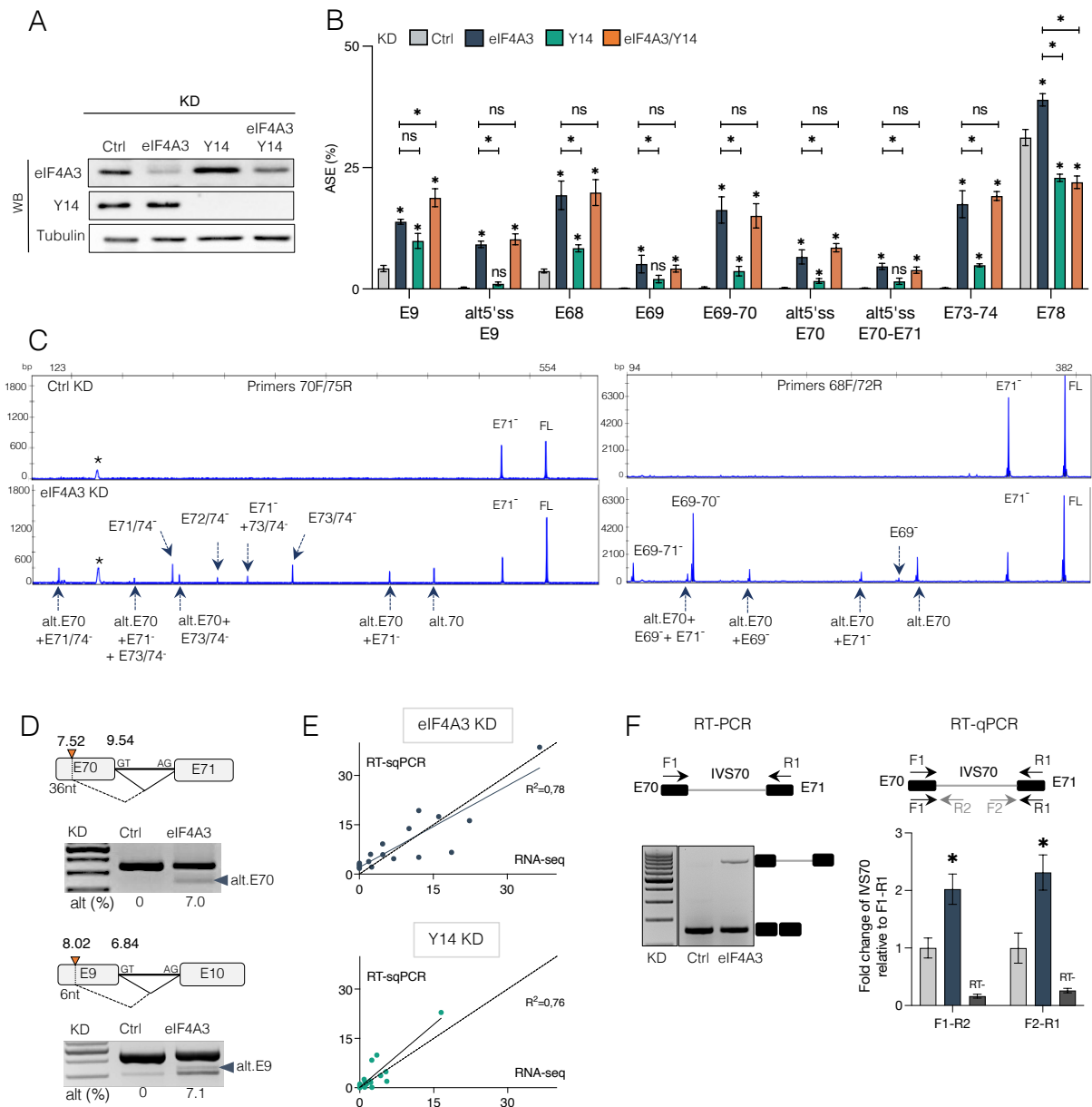
**Figure 60. Splicing profile of the Dp427m isoform in eIF4A3- or Y14-depleted human C25CL48 muscle cells.**

(A) Inclusion level of the 79 DMD exons measured as Percent Spliced In (PSI) from DMD-targeted RNA-seq data obtained after 3 days of differentiation in C25CL48 cells knocked down (KD) for eIF4A3 (dark blue) and Y14 (green) compared to a control siRNA (Ctrl, grey). (B) The knockdown efficiency of eIF4A3 and Y14 was assessed by RT-qPCR of transcripts levels using RPLPO as endogenous reference gene (upper panel). The histograms represent the means  $\pm$  standard deviations of two

biological replicates. The depletion of the targeted proteins was verified by immunoblotting, with tubulin as a loading control. (C) Alternative Splicing Events (ASEs) above 5% (dotted line) identified with the IPSA software in the eIF4A3 KD condition are depicted in the bar graph and compared to levels detected in the Y14 KD and control (Ctrl) conditions. (D-I) Sashimi plots obtained by ggsashimi and color-coded by condition (si Ctrl, si-eIF4A3, siY14) depicting (D-H) the differentially spliced exons and (I) intron 70 (IVS70) retention. The raw junction read counts (mean of two replicates) are shown on top of each junction represented as curved lines. Schematic of the primary transcript (ENST00000367455) for the DMD gene from the Ensembl database, version 75, with exons shown as boxes, introns shown as lines, and arrows indicating the direction of transcription.

To support the targeted RNA-seq data, we analyzed the effect of eIF4A3, Y14 or the double eIF4A3/Y14 KD on DMD splicing by RT-PCR. Four independent knockdown experiments were performed in the C25Cl48 muscular cell line. High knockdown efficiency was achieved at RNA and protein level in all conditions (Fig. 61A, Suppl Fig. S1A). The splicing changes were detected either by conventional RT-PCR (Suppl Fig. S1B) or semi-quantitative fluorescent PCR (RT-SQF-PCR) (Fig. 61B, Suppl Fig. S1C) using primers located in the exons flanking the splicing events (Suppl Table S3). In eIF4A3 and Y14 KD conditions, we could validate all nine ASEs that were previously detected by targeted RNA-seq, including the use of alt5'ss in exon 9 and 70 leading to inclusion of shortened exon 9 (6-nt) and exon 70 (36nt). in the mature transcripts (Fig. 61B and 61D). The identity of the splicing events was verified by Sanger sequencing (Suppl Fig. S1D). As previously observed, reduced expression of Y14 affected splicing of the DMD exons more modestly. The splicing profile in the double KD cells was superimposable to that obtained upon eIF4A3 depletion, except for exon 78 which appeared to be regulated differently and showed a Y14 KD like behavior (Fig. 61B). We took advantage of fluorescent PCR products separation on the sequencer, which allows nucleotide discrimination of larger fragments according to their size, to further characterize the pattern of ASEs in the distal end of the transcripts. The use of primer pairs located in exons 68 and 72 or in exon 70 and 75 uncovered the complexity of events occurring upon depletion of EJC components (Fig. 61C). There was an accumulation of mis-spliced transcripts as previously reported in other genes (Boehm et al., 2018). Such complex splicing patterns are less easily identified by RT-PCR and agarose gel electrophoresis because of the difficulty in discriminating fragment sizes. Many of these ASEs involved the use of the alt5'ss in exon 70 (Alt.E70), albeit weaker than the natural 5'ss splice site (MaxEnt 7.52 and 9.54, respectively) (Fig. 61D). A change in the level of inclusion of exon 71 was also noted in eIF4A3 KD condition (Fig. 61C), in contrast to the PSI analysis (Fig. 61A). This apparent decrease in single exon 71 skipping was counterbalanced by the presence of other alternative transcript isoforms. These alternative transcripts associating the E71 skipping with other more complex splicing events showed that the overall level of transcripts including exon 71 remained unchanged in the cell, consistent with the PSI analysis. We therefore extended the list to 18 EJC-dependent splicing changes (Suppl Table S2) consisting of combinations of the initially detected splicing

events by SQF-PCR. We were able to detect these additional ASEs in RNA-Seq data, but at a level below the 5% threshold initially considered (Table Suppl.S1B,  $0.05 > \Delta_{SJ} \geq 0.01$ ). Overall, a good correlation between the two targeted RNA-Seq and RT-SQF-PCR approaches was observed for both eIF4A3 KD ( $R^2 = 0.78$ ) and Y14 KD ( $R^2 = 0.76$ ) conditions (Fig. 61E, Fig. S2A, Suppl. Table S2). Finally, we performed validation experiments by RT-PCR and agarose gel electrophoresis to confirm the low-level of intron 70 retention seen in RNA-Seq data in the eIF4A3 KD condition (Fig. 61F, left panel). We accurately determined the intron 70 retention level (IVS70) by RT-qPCR using three different primer sets to specifically amplify either the correctly spliced (F1-R1) or the intron-retaining (F1-R2 and F2-R1) transcripts in eIF4A3 and Y14 KD conditions. Although they represent only a minor fraction, a significantly increased level of retained intron-containing transcripts was seen with both primer pairs compared to control for eIF4A3 KD (Fig. 61F, right panel) and Y14 KD (Fig. S2B). Taken together, these data support a role for EJC in the efficiency and fidelity of splicing in this region of the DMD pre-mRNA.



**Figure 61. Validation of Alternative Splicing Events in the DMD gene upon depletion of EJC components.**

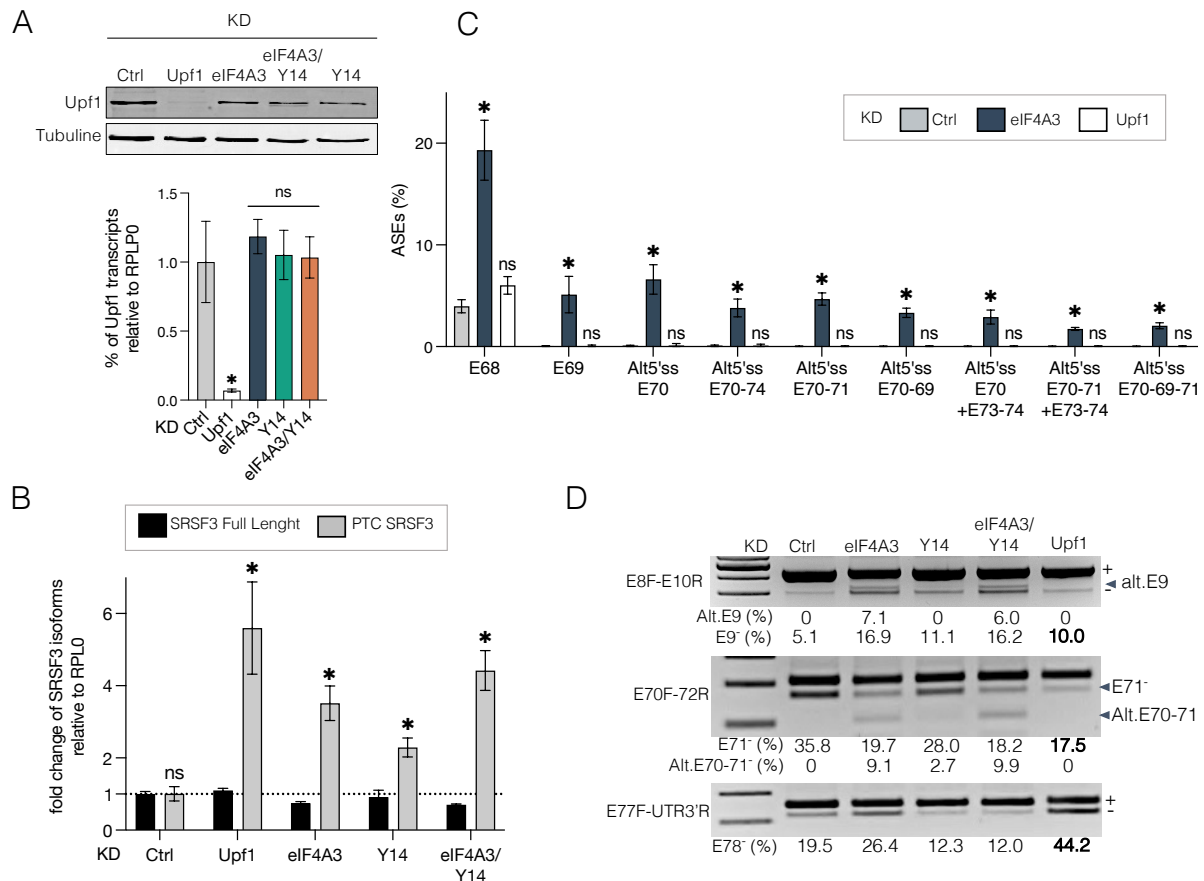
C25C148 cells were transfected in 4 replicates with a control siRNA (Ctrl) and with siRNAs targeting either eIF4A3 or Y14 or the two proteins (eIF4A3/Y14). Total RNA and protein were extracted after 3 days of differentiation. (A) Western blot analysis of eIF4A3, Y14 and Tubulin as loading control in the different KD conditions. (B, C) Alternative Splicing Events (ASEs) in the DMD gene were analyzed by fluorescent semi-quantitative RT-PCR on a capillary sequencer. (C) Representative electropherograms ( $n=4$ ) of large-sized amplified fragments with primers in exons 70 and 75 (70F/75R) or in exons 68 and 72 (68F/72R) showing the resolution of multiple fragments in the eIF4A3 KD condition compared to the control (Ctrl) condition. The traces are a visual representation of the raw data showing the size distribution in base pairs (x axis) and fluorescence signal intensity (y axis). The identity of each band determined by the size obtained with the GeneMapper software is indicated (arrows). Exon 71 skipping (E71-), full-length RT-PCR product (FL). Non-specific band (\*). (B). Quantification of ASEs ( $n=4$ ) are shown as means  $\pm$  SD ( $*p < 0.05$ ; multiple Mann-Whitney test). ns, non-significant. (D) Representative 2% agarose gels allowing visualization of the alternative (alt) donor splice site (arrow) in exon 70 and in exon 9 activated in eIF4A3 KD condition (values in % below the gel). The pattern of each splicing event is shown at the top of the gels. The MaxEnt scores for the natural and alternative donor splice sites are given. (E) Correlation of ASEs levels calculated from DMD-targeted RNA-seq and fluorescent RT-PCR data. The Pearson correlation coefficients ( $R^2$ ) and linear regression lines are indicated. (F) Validation of intron 70 retention (IVS70, 698 bp) in eIF4A3 KD condition by RT-PCR using primers in exons 70 and 71 (F1/R1) (left panel) or quantitative RT-PCR (RT-qPCR) (right panel). Fold changes were assessed ( $n=4$ ) using three different primer sets (F1/R1, F1/R2, F2/R1) as represented in the top scheme. No reverse transcriptase control (RT-). Data are mean  $\pm$  SD ( $*p < 0.05$ ; multiple Mann-Whitney test).

## Deregulation of DMD splicing is not due to repression of EJC-dependent NMD

In mammalian cells, EJC deposition is pivotal for the process of nonsense-mediated mRNA decay (NMD), that recognizes and degrades aberrant transcripts carrying premature termination codons (PTC). The EJC serves as a binding platform for the NMD factors, Up-frameshift *protein* UPF3b, UPF2 and UPF1 (Chamieh et al., 2008). Thus, we sought to verify whether the splicing changes observed on the DMD transcript upon eIF4A3 and Y14 depletions might have result from the stabilization of PTC-containing mRNAs due to less-efficient EJC-dependent NMD instead of revealing a true role in splicing by the EJC components. Therefore, we depleted the essential NMD factor UPF1 using siRNA to stabilize NMD substrates. In knockdown cells, mRNA and protein expression levels of UPF1 were drastically reduced with no significant change in eIF4A3 and Y14 KD (Fig. 62A). We assessed the NMD efficiency in the cells by quantifying the NMD-sensitive splice variant of the SRSF3 gene that includes the PTC-containing exon 4 (SRSF3-PTC) (Kano et al., 2014; Shadle et al., 2017). Consistent with previous studies, the level of the SRSF3-PTC isoform was significantly increased in the UPF1 KD condition compared to the control, whereas the level of the SRSF3 isoform encoding the functional full-length protein remained unchanged (Fig. 62B). The SRSF3-PTC isoform also escaped degradation upon eIF4A3 and/or Y14 KD to variable extent, consistent with the known NMD-activating function of the EJC. Importantly, the depletion of UPF1 failed to reproduce the splicing changes observed in eIF4A3-depleted cells for the 9 ASEs previously characterized by SQF-PCR (Fig. 62C), indicating that these ASEs detected upon EJC KD are not linked to NMD efficiency decrease.

The role of EJC components in DMD splicing is further supported by the fact that eIF4A3 KD also induces in-frame events that are not expected to be NMD substrates (Suppl Table S2). These in-frame ASEs were not detected in UPF1-depleted cells (Supp. Fig. S3A, B), except for a modest increase of exon 9 skipping relative to control (Fig. 62D, Supp. Fig. S3A). One striking feature of this analysis is the change in splicing pattern observed for exons 71 and 78. A 50% decrease in exon 71 skipping (from 35.8% to 17.5% in control and UPF1 KD condition, respectively) was shown by RT-PCR using the primer pair E70F-E72R (Fig. 62D, Supp. Fig. S3A). We were able to confirm by SQF-PCR analysis of a larger fragment using primers E70F-75R that this decrease was indeed due to the re-inclusion of the exon 71 (39 bp) and was not related to the presence of more complex splicing events including exon 71 as was observed for eIF4A3 KD (data not showed). In addition, we observed that depletion of UPF1 strongly increased the skipping of exon 78 (44.2%) compared with control (19.5%) and eIF4A3 KD

(26.4%) conditions, (Fig. 62D, Supp. Fig. S3A). The skipping of the 32-bp penultimate exon 78 changes the open reading frame of the last exon 79. A more downstream stop codon is used that modifies the C-terminal region of the dystrophin protein (Howard et al., 1999; Rau et al., 2015). Hence, the absence of UPF1 appear to have a direct or indirect impact on splicing of these two short in-frame DMD exons.



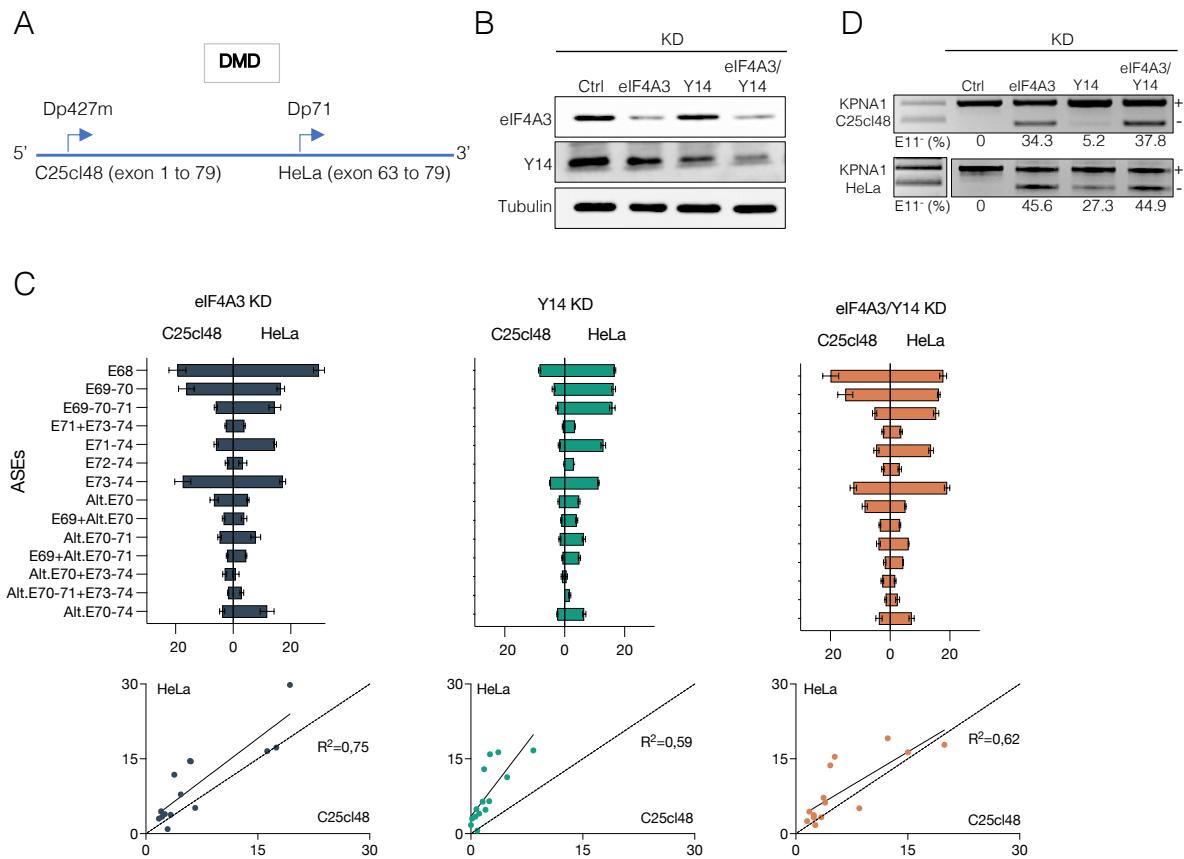
**Figure 62. Effect of UPF1 knockdown on EJC-dependent splicing events.**

Splicing of NMD-sensitive transcripts was studied in C25C148 cells transfected with specific siRNAs against UPF1 or control (Ctrl). (A) UPF1 expression level in UPF1, eIF4A3, Y14 and control KD conditions was determined at protein level by western using Tubulin as loading control (upper panel, representative image) and mRNA level by RT-qPCR using RPLP0 for normalization. Data are mean  $\pm$  SD ( $n=4$ ) ( $*p < 0.05$ ; multiple Mann-Whitney test). ns, non-significant. (B) The level of the SRSF3 NMD-sensitive transcript isoform (SRSF3-PTC containing exon 4) was assessed as a control of reduced NMD efficiency in the various KD conditions. Bar graph shows fold changes in functional full-length protein (SRSF3-FL) and PTC-containing (SRSF3-PTC) isoforms measured by isoform specific qPCR as previously described (Kano et al., 2014) using RPLP0 as endogenous reference gene. Data are mean  $\pm$  SD ( $n=4$ ) ( $*p < 0.05$ ; multiple Mann-Whitney test). (C) Bar graph showing levels of out-of-frame ASEs introducing a PTC in DMD transcripts assessed by fluorescent semi-quantitative RT-PCR in UPF1 KD compared to eIF4A3 and control KD conditions. Data are mean  $\pm$  SD of four independent replicates. ( $*p < 0.05$ ; multiple Mann-Whitney test). (D) Representative 2% agarose gels showing changes in the inclusion level of in frame DMD exon 9, exon 71 and exon 78 upon UPF1 depletion. The primers used are shown on the left side of the gel and the exon skipping levels (E9-, E71-, E78-) quantified by the Image Lab software below the gel. (+) exon inclusion, (-) exon exclusion, (Alt) alternative donor splice sites in exon 9 and 70.

### The EJC-mediated splicing regulation of DMD exons is conserved in different isoforms

We next wanted to test whether the eIF4A3 and/or Y14 siRNA-induced splicing defects in the Dp427m transcript could be rescued by re-expression of the wild-type proteins from siRNA-resistant eIF4A3 and Y14 expression constructs (Flag-eIF4A3-WT and Flag-Y14-WT). As we had previously experimented with other splicing factors, we were unable to perform effective rescue of Flag-eIF4A3-WT in the C25Cl48 muscle cell line, where plasmids are difficult to transfect (data not shown). Therefore, we changed the cell model and used HeLa cells in which eIF4A3 KD rescue experiments had previously been successfully performed (Wang et al., 2014). The short dystrophin isoform (Dp71) is expressed in HeLa cells from an alternative promoter located in intron 62 (Fig. 63A). Dp71 includes the 17 last exons of the muscular Dp427m isoform (*ie* all the exons that were deregulated in eIF4A3 KD condition except exon 9). First, we validated that depletion of eIF4A3 and/or Y14 (Fig. 63B) caused splicing changes in Dp71 like those observed in muscle cells. A good correlation in splicing pattern was observed between the two cell lines in eIF4A3 KD (Fig. 63C). The KD of Y14 had a greater impact on splicing in HeLa cells, close to that of the eIF4A3 KD, in agreement with other studies (Wang et al., 2014), despite a level of Y14 depletion comparable to the one in the muscle cells (Fig. 61A, Fig. 63B, Fig. S1A-B). This difference between the 2 cell lines was not limited to the *DMD* gene, it was also visible for the control gene KPNA1 (Wang et al., 2014), pointing to a cell-type specific behavior to variation in intracellular concentration of EJC components (Fig. 63D).

Rescue experiments conducted in HeLa cells, showed significant, but not complete, rescue of splicing defects in cells lacking eIF4A3 upon expression of Flag-eIF4A3-WT but not Flag-Y14-WT by SQF-PCR (Fig. 64A, compare dark grey bars in upper and lower panel). Conversely, rescue of Y14 siRNA-induced splicing defects by Flag-Y14-WT but not Flag-eIF4A3-WT. Only the E68 and E73-74 skipping event showed rescue under both conditions (Fig. 64A, compare green bars in upper and lower panel). In double KD cells, transfection of either protein led to an intermediate, yet mostly significant, degree of rescue. Overall, these data tend to establish the specific role for the eIF4A3 and Y14 components of the EJC in the correct splicing of DMD distal exons. Whether this regulation of splicing is dependent on the EJC assembly was further investigated by ectopically expressing a mutant Flag-eIF4A3-MUT (D401KE402R) that lacks the binding capacity to the MAGOH-Y14 heterodimer. This mutant was unable to reverse the effect of knocking the eIF4A3 protein down (Fig. 64B, upper panel).



**Figure 63. Validation of the HeLa cell line as a model to study EJC-dependent splicing events in the DMD gene.** (A) Scheme depicting the C25Cl48 and HeLa cell specific DMD isoforms (Dp427m and Dp71) and their promoter position in the gene. (B) HeLa cells were transfected with the siRNAs against eIF4A3, Y14 or a mixture of both (eIF4A3/Y14) and a control siRNA (Ctrl). At 48 h post-transfection, whole cell extracts were subjected to Western blot analysis to assess knockdown efficiency. (C) Comparison of the levels of ASEs induced by EJC components in C25Cl48 and in HeLa cells in the three KD conditions. Pearson correlation coefficients ( $R^2$ ) and linear regression lines are displayed below the bar graphs. (D) RT-PCR analysis of KPNA1 exon 11 as a control for increased exon skipping in KD EJC in HeLa cells (Wang et al., 2014) to confirm the weaker impact of Y14 KD in C25Cl48 compared to HeLa cells. Values below the gels are for exon 11 skipping (%).

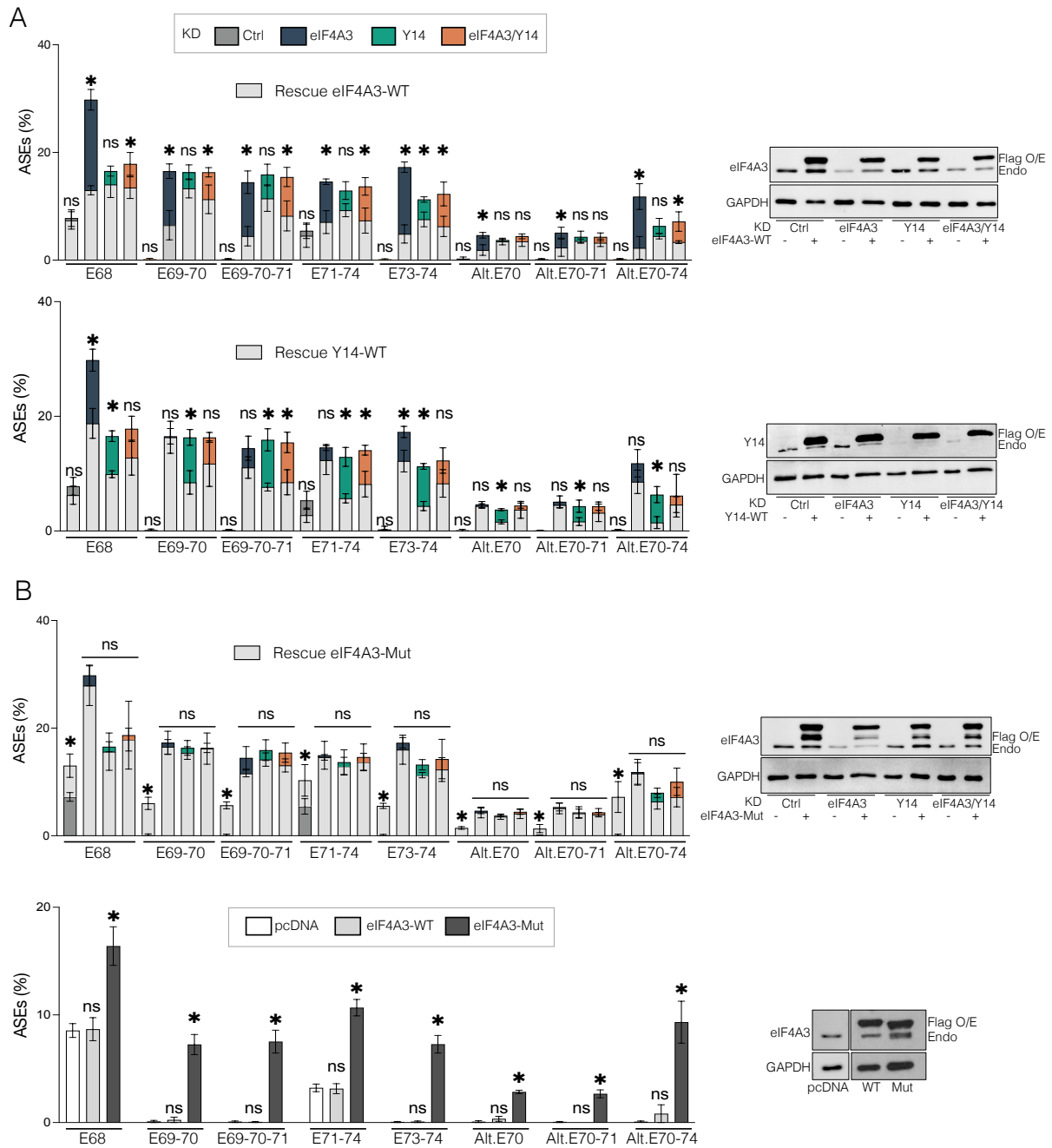
In addition, its expression in untreated cells exerted a dominant negative effect on the endogenous wild-type protein, reproducing to some extent the aberrant splicing events detected in eIF4A3-depleted cells (Fig. 64B, bottom panel). These experiments indicate that the regulation of DMD pre-mRNA splicing is dependent of the Exon Junction Complex formation. Note that the expression of exogenous Flag-tagged proteins in all rescue experiments was verified by western blotting (Fig. 64A, B, right panels).

### EJC peripheral protein complexes modulate splicing of DMD exons 71 and 78

To better understand how EJC influences DMD pre-mRNA splicing, we aimed to determine whether peripheral protein MLN51 and complexes ASAP/PSAP play a role in this regulation. In particular, we postulated that RNPS1 bound to EJC may be required to inhibit the use of the alternative 5' splice sites in DMD exon 9 and 70, in agreement with its reported role in suppressing cryptic or reconstituted 5' splice sites (Blazquez et al., 2018; Boehm et al., 2018).



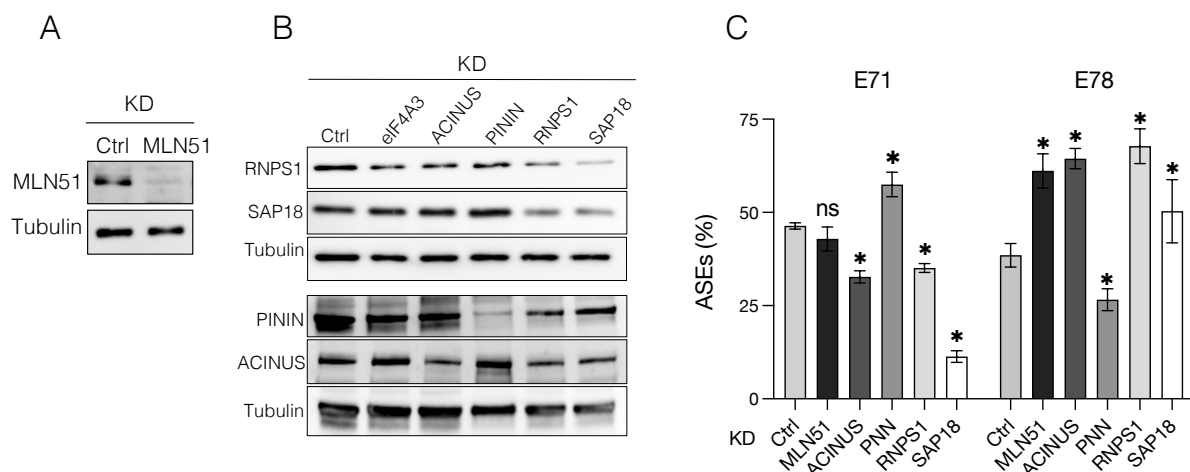
We performed siRNA-mediated depletion in C25C148 muscle cells of MLN51 and of each component of the ASAP and PSAP complexes.



**Figure 64. eIF4A3 and Y14 individually contribute to the inclusion of DMD exons but complex formation is required for correct DMD splicing.**

(A) HeLa cells depleted for eIF4A3, Y14 or the two proteins (eIF4A3/Y14) were transfected with a siRNA-resistant wild type (WT) FLAG-eIF4A3 (upper panel) or FLAG-Y14 (lower panel) expression vector and (B, upper panel) with a mutant (Mut) siRNA-resistant FLAG-eIF4A3 that do not form EJC. Histograms show the quantification of ASEs by fluorescent semi-quantitative RT-PCR in rescue conditions (light gray bars) compared to levels in eIF4A3, Y14, eIF4A3/Y14 and control siRNA (Ctrl) KD conditions. Data are mean  $\pm$  SD ( $n=4$ ) (\* $p < 0.05$ ; multiple Mann-Whitney test). ns, non-significant. (Endo) endogenous protein. (B, lower panel). The wild-type (WT) or mutant (Mut) siRNA-resistant FLAG-eIF4A3 were overexpressed in untreated HeLa cells and ASEs were quantified as described above and compared to the control condition (empty pcDNA vector-transfected cells). Representative western blots showing Flag protein overexpression (Flag O/E) using antibodies against native proteins and GAPDH as a loading control are shown on the right side of the histograms.

ASAP and PSAP are composed of RNPS1 and SAP18 and differ by one of the two specific proteins, namely ACINUS and PININ for ASAP and PSAP, respectively. The efficiency of siRNAs and the impact on the expression level of these proteins were assessed by RT-qPCR (Fig. S4A, B) and western blot analysis (Fig. 65A, B). The splicing pattern of distal DMD exons was analyzed by SQF-PCR and agarose gel electrophoresis of RT-PCR products. First, we did not detect any effect of RNPS1 depletion in the use of the exonic cryptic 5' ss in exon 9 and exon 70 (Fig. S4C). Furthermore, we observed that the EJC-dependent splicing events previously characterized under eIF4A3 and Y14 KD, were not found under silencing of MLN51 or ASAP/PSAP components (Fig. 65C, Suppl Fig. S4D) However, depletion of each protein of ASAP and PSAP complexes had a significant impact on the level of inclusion of exons 71 and 78, different from that observed in the absence of the core components of the EJC (Fig. 65C, Suppl Fig. S4E). We found that, in contrast to PININ KD, the silencing of MLN51 as well as ASAP components (ACINUS, RNPS1, SAP18) significantly increased the skipping of exon 78. The opposite situation was observed for exon 71, where the KD of ASAP components enhanced the inclusion of exon 71, specifically SAP18, while the KD of PININ increased the skipping. These results indicate specific and independent roles of the ASAP and PSAP complexes on the inclusion of these DMD exons, as previously described (Wang et al., 2018). The exon 78 being dependent on the presence of the ASAP complex, and exon 71 being more dependent on the PSAP complex.



**Figure 65. The EJC-peripheral proteins regulate splicing of DMD exons 71 and 78.**

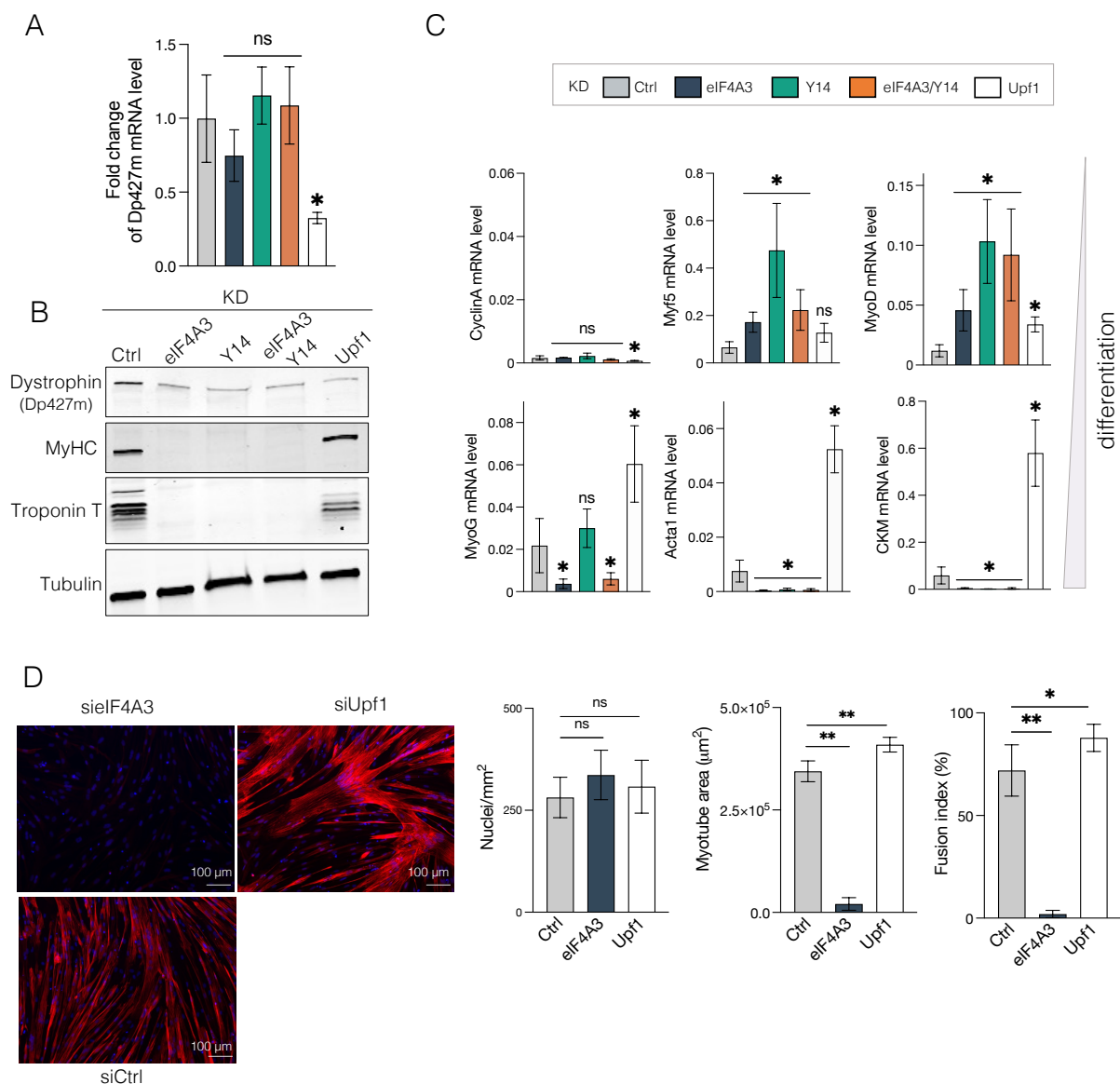
The role on DMD splicing of MLN51 and components of the ASAP/PSAP complexes (ACINUS, PININ, RNPS1 and SAP18) was assessed in C25C148 cells by transfecting specific siRNAs against the different proteins together with a siRNA control (Ctrl). After 3 days of differentiation, total RNA and protein were extracted from siRNA-treated cells. (A, B) The depletion efficiency in the control different KD conditions was assayed by western blot using Tubulin as loading control. (C) Quantification of exon 71 and exon 78 skipping level (ASE, %) by fluorescent semi-quantitative RT-PCR from 4 independent experiments. Data are mean  $\pm$  SD (\* $p$  < 0.05; multiple Mann-Whitney test). ns, non-significant.

### Reduced expression of EJC factors impairs C25C148 myogenic differentiation

We analyzed the expression of the dystrophin Dp427m isoform at the mRNA and protein levels in eIF4A3, Y14 and UPF1 KD in the differentiated C25C148 muscle cells. Dp427m protein expression is optimal at 3 days of differentiation in C25C148 model (Miro et al., 2020). While the Dp427m mRNA level was globally unchanged in the eIF4A3 and/or Y14 KD condition, the western blot showed a decrease in the expression of Dp427m (Fig. 66A-B) which nevertheless remained higher than that of undifferentiated myoblasts (Suppl Fig. S5A). Regarding UPF1 KD, the decrease of Dp427m expression correlated with a decrease in mRNA expression. We were able to verify that this decrease in Dp427 mRNA was not associated with a change in the expression of the Dp71 isoform which is expressed in undifferentiated myoblasts (de León et al., 2005) (Suppl Fig. S5B). Strikingly, western blot analysis also revealed that troponin T and myosin heavy chain (MHC), two cytoskeletal proteins normally expressed in multinucleated myotubes as observed in the control, were completely absent in eIF4A3 and Y14 KD conditions, conversely to the UPF1 KD condition (Fig. 66B).

We therefore sought to further characterize the differentiation state of the cells in the different EJC KD conditions. We profiled the expression of muscle-specific marker genes by RT-qPCR. First, we analyzed the level of cyclin A2, which is required for cell cycle progression in proliferating cells. The low level detected in all KD conditions, which was comparable or even lower (UPF1 KD) than the control, indicated a cell cycle arrest that usually precedes differentiation (Fig. 66C). The control condition provided us with the expression profile of different myogenic regulatory factors in differentiated cells at 3 days. The myogenic commitment factors Myf5 and MyoD, the early differentiation factor myogenin (MyoG) as well as the muscle-specific proteins expressed at a later stage of differentiation, actin alpha 1 (ACTA1) and creatine kinase M-type (CKM) were tested. Compared to the control, a dramatically different pattern was observed in cells silenced for eIF4A3 and Y14, resembling undifferentiated myoblasts with higher levels of Myf5 and MyoD. A very low level of MyoG for eIF4A3 KD and an absence of ACTA1 and CKM were observed. However, MyoG is still present in Y14 KD condition at similar level compared to the control condition, suggesting that Y14 depleted cells are non-fused myocytes and not myoblasts as in eIF4A3 KD. Conversely and consistent with the described role of UPF1 in repressing myogenesis by promoting MyoD protein degradation (Feng et al., 2017a), the depletion of UPF1 accelerated myogenesis and enhanced expression of myogenic differentiation genes, MyoG, ACTA1 and CKM. These results were confirmed by immunofluorescence studies that evidenced significant

morphological differences between KD conditions (Fig. 66D). We could observe that UPF1 KD cells matured faster into myotubes compared with the control, as evidenced by strong troponin T staining of the cells and an increase in cell surface area and fusion index (Fig. 66D). Conversely cells lacking eIF4A3 showed a complete absence of myotube formation, with no change in the number of nuclei which showed no impact on the proliferation. All together these results show that knocking down of EJC components impacts myogenic differentiation and maintains myoblasts in an immature stage. Furthermore, we were able to show that while the UPF1 KD induces accelerated differentiation, it does not allow normal expression of the muscle Dp427m dystrophin isoform in myotubes.



**Figure 66. Depletion of EJC components in muscular C25C148 cells decreases dystrophin expression and impairs myogenic differentiation.**

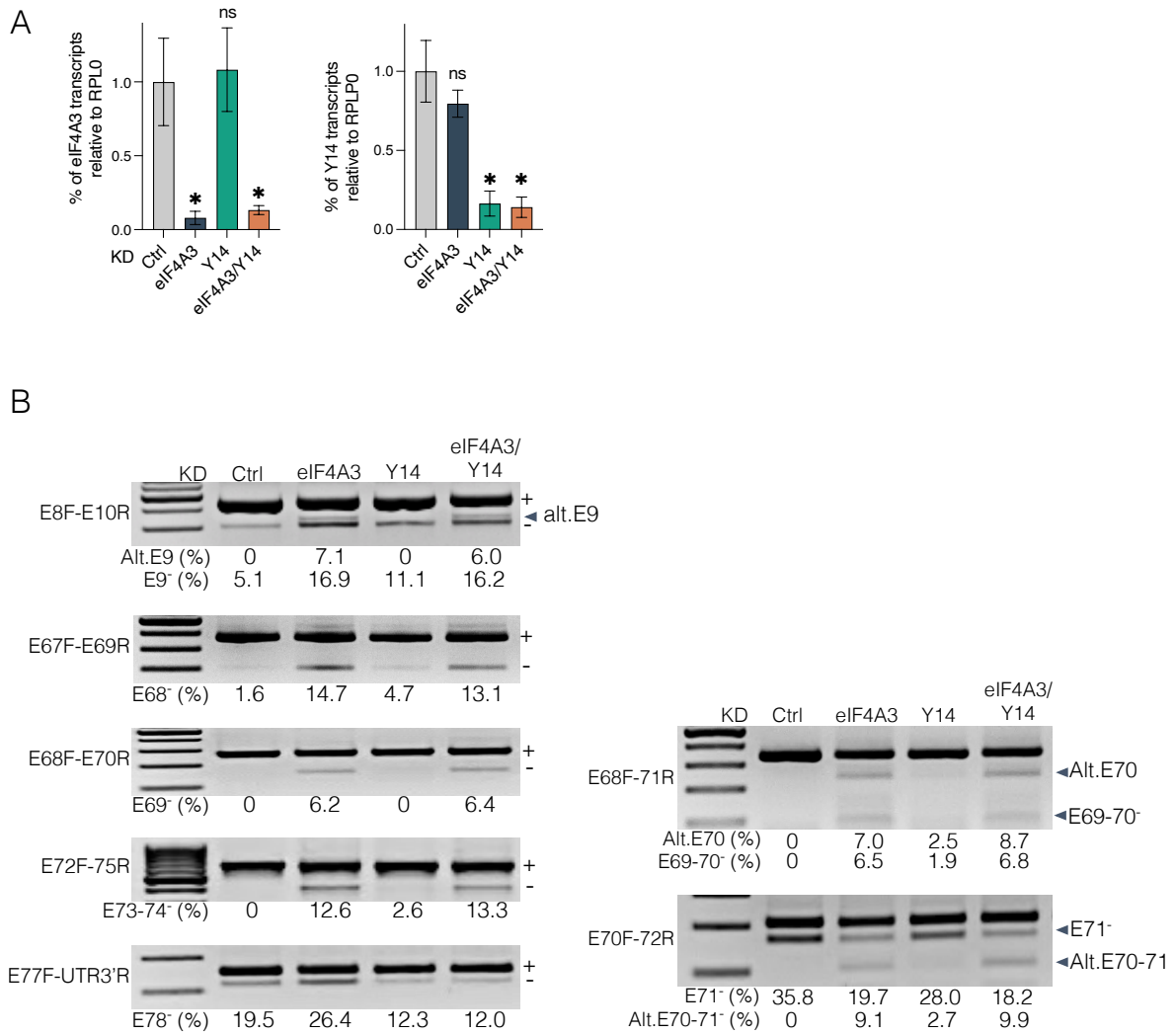
The DMD gene expression level in C25C148 cells depleted for EJC (eIF4A3, Y14, eIF4A3/Y14) and UPF1 was assessed (A) at mRNA level by measuring the level of Dp427m by RT-qPCR (n=4) and (B) at protein level using the Dys1 antibody against the rod domain of dystrophin. Antibodies against the muscle Myosin heavy chain (MyHC) and Troponin T proteins were used to

assess the differentiation state of the cells. (C) The mRNA level of factors indicative of the proliferative stage (cyclin A), myogenic commitment (Myf5, MyoD), and early (Myogenin, MyoG) or late (Acta1, CKM) differentiation was measured by RT-qPCR. Data are mean  $\pm$  SD (n=4) (\*p < 0.05; multiple Mann-Whitney test). ns, non-significant. (D) Cells were fixed and immunostained for Troponin-T and myoblast differentiation was observed by fluorescence microscopy (red, Troponin-T; blue, DAPI). Scale bars = 100  $\mu$ m. Myotube morphology was assessed by calculating mean myotube area (Troponin T expressing cells with at least 4 merged nuclei), total number of nuclei and fusion index (nuclei per myotube/total nuclei). (\*p < 0.05; \*\*p < 0.01).

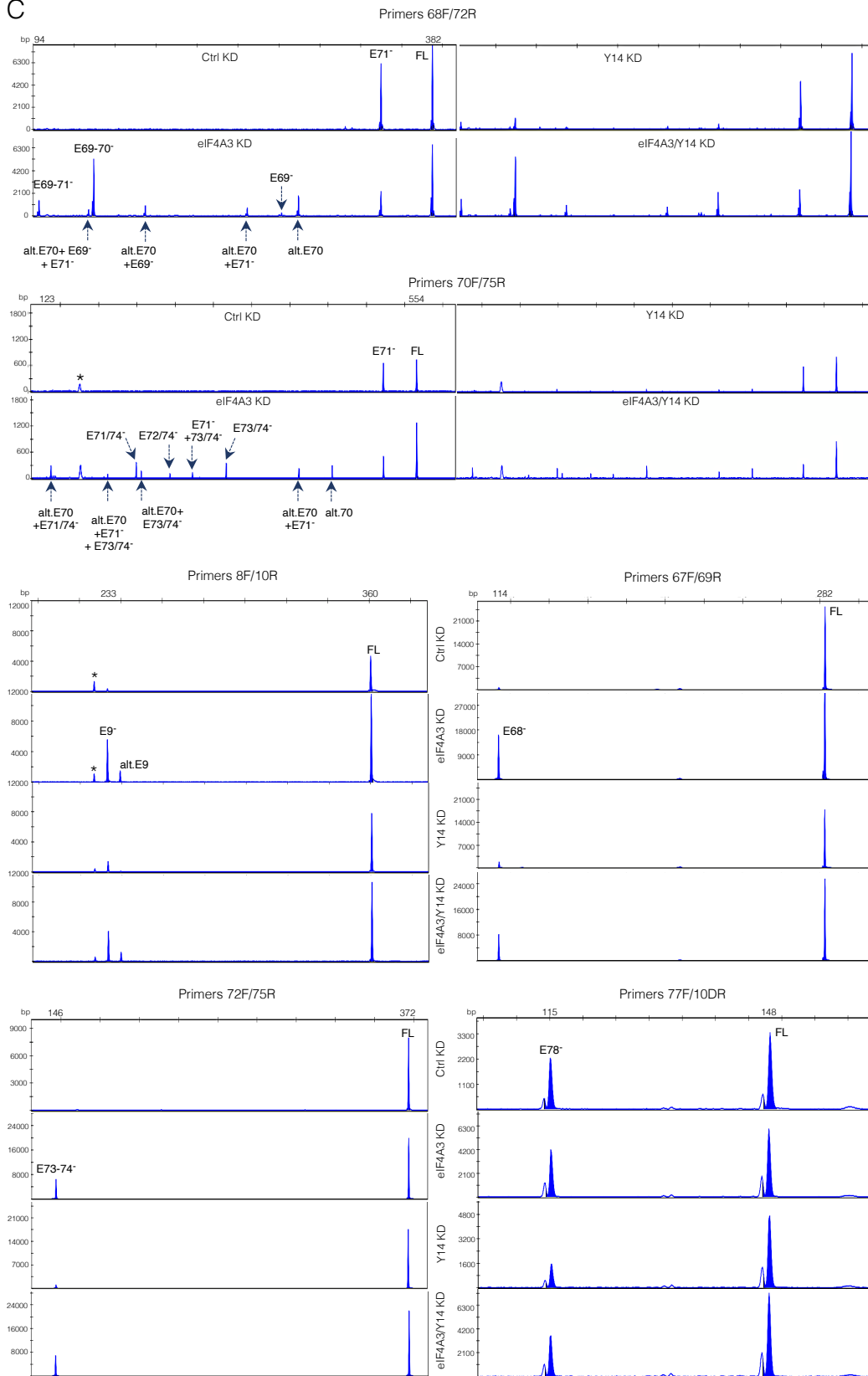
# Supplementary information

## Supplementary Figures

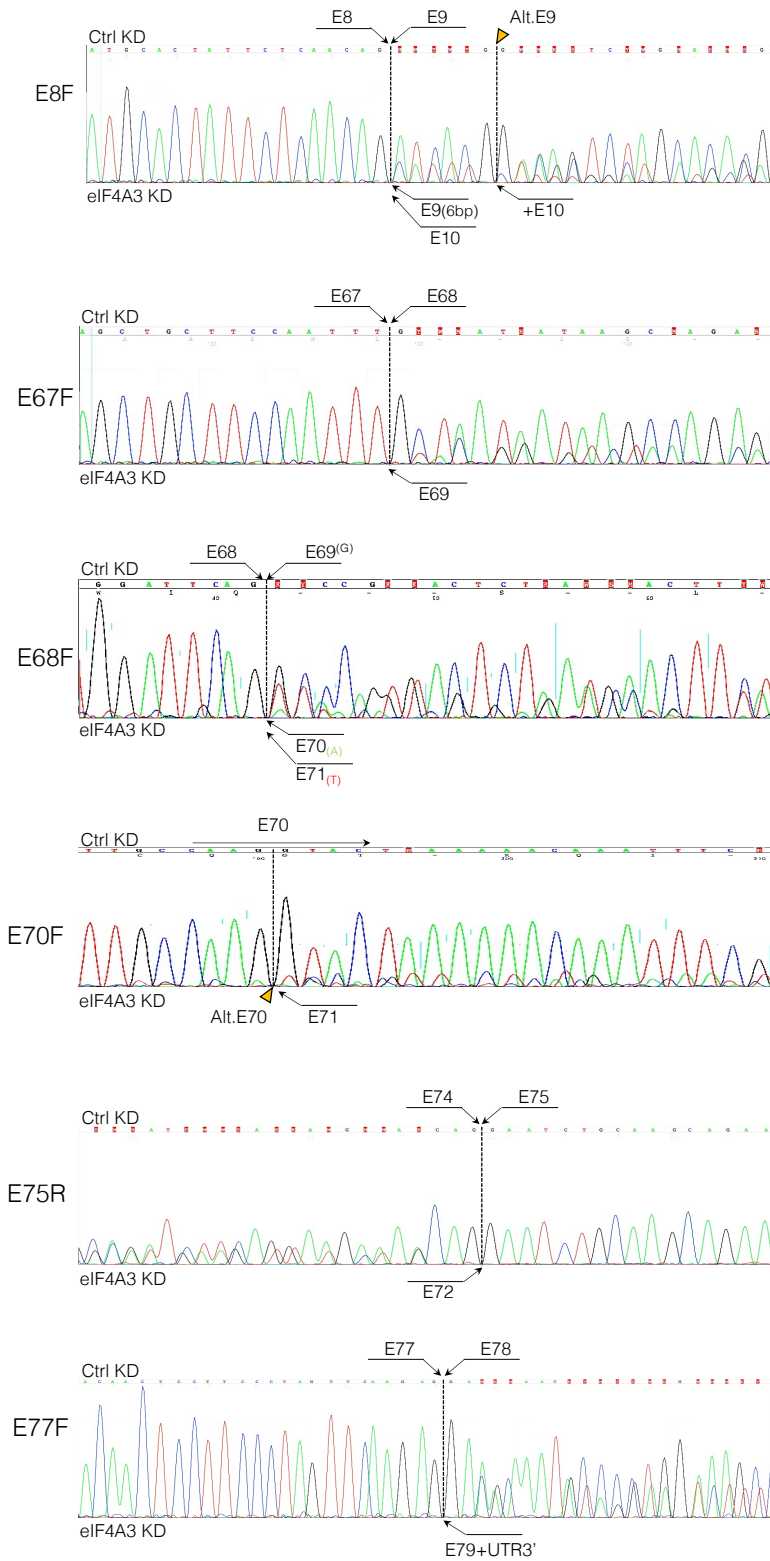
Supplemental Figure 1



C



D





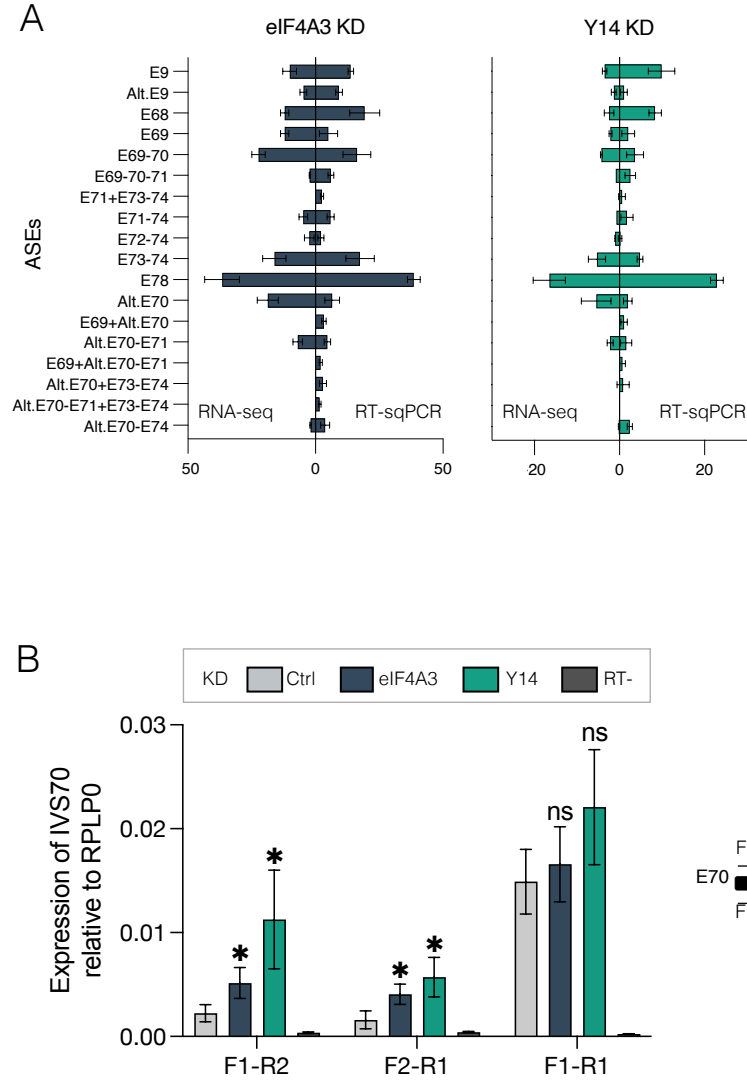
**(A)** Efficiency of eIF4A3 and Y14 KD was assessed at mRNA level by RT-qPCR using RPLP0 for normalization. Data are mean  $\pm$  SD (n=4) (\*p < 0.05; multiple Mann-Whitney test). ns, non-significant.

**(B)** Representative agarose gels of alternative splicing events (ASE) analyzed by RT-PCR upon depletion of eIF4A3, Y14, or the two proteins (eIF4A3/Y14) compared to the control siRNA (Ctrl). Primers used for amplification are indicated on the left side of the gels. The percentages of exon exclusion (E9<sup>-</sup>, E68<sup>-</sup>, E69<sup>-</sup>, E73-74<sup>-</sup>, E78<sup>-</sup>, E69-70<sup>-</sup>, E71<sup>-</sup>) or use of alternative splice sites (Alt.E9, AltE70) are indicated below the gels. The + and - signs indicate exon inclusion and exclusion, respectively.

**(C)** Electropherograms from fluorescent semi-quantitative RT-PCR experiments. Primers used for amplification are indicated above each panel, and the identity of each peak is shown (arrows). Full-length RT-PCR product (FL).

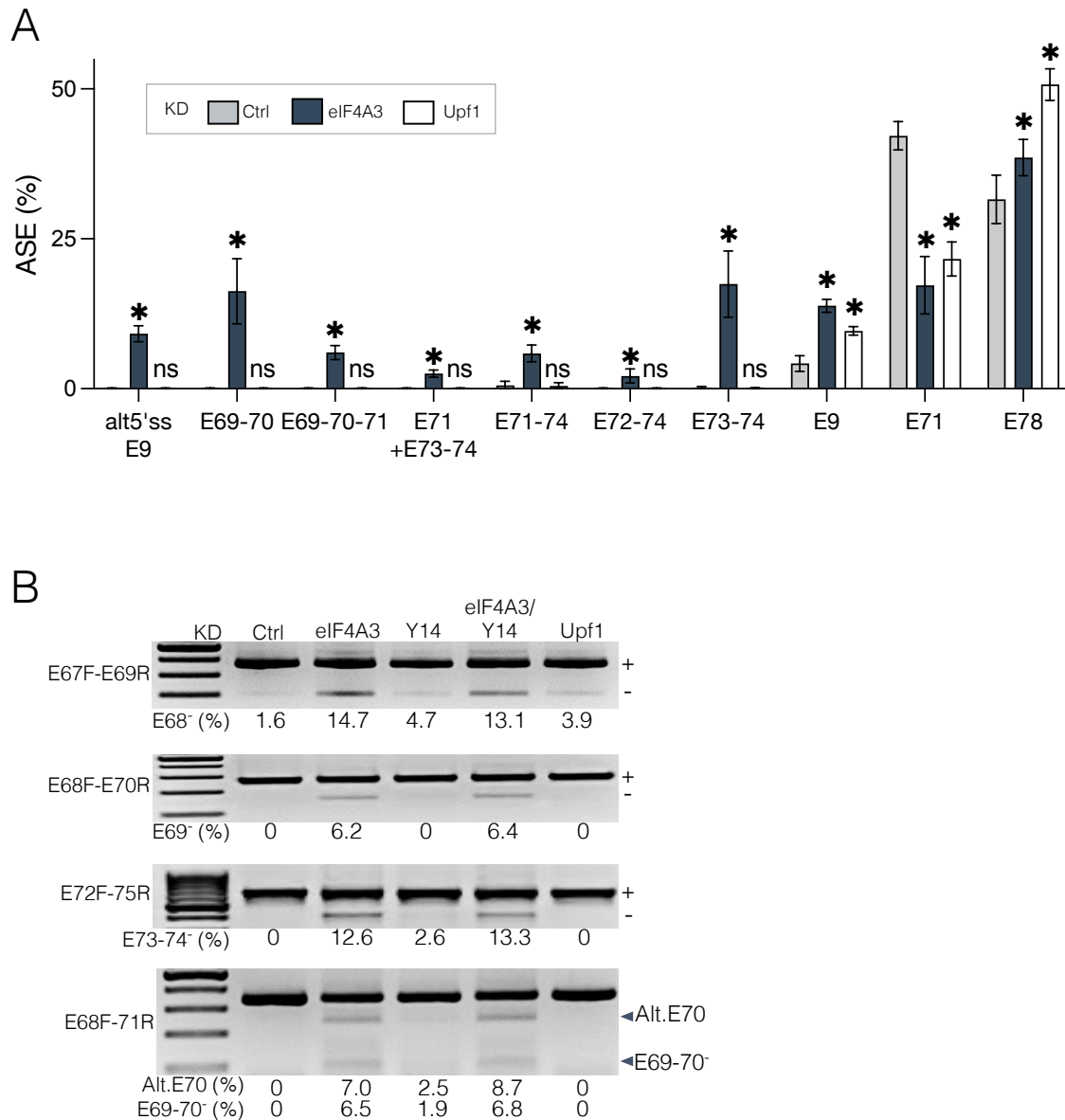
**(D)** Sanger sequencing of RT-PCR products. The position of the abnormal junction in the eIF4A3 KD condition is indicated by a dotted line. The sequence shift is visible beyond this position. The primers used for sequencing is shown on the left side of the sequences.

## Supplemental Figure 2



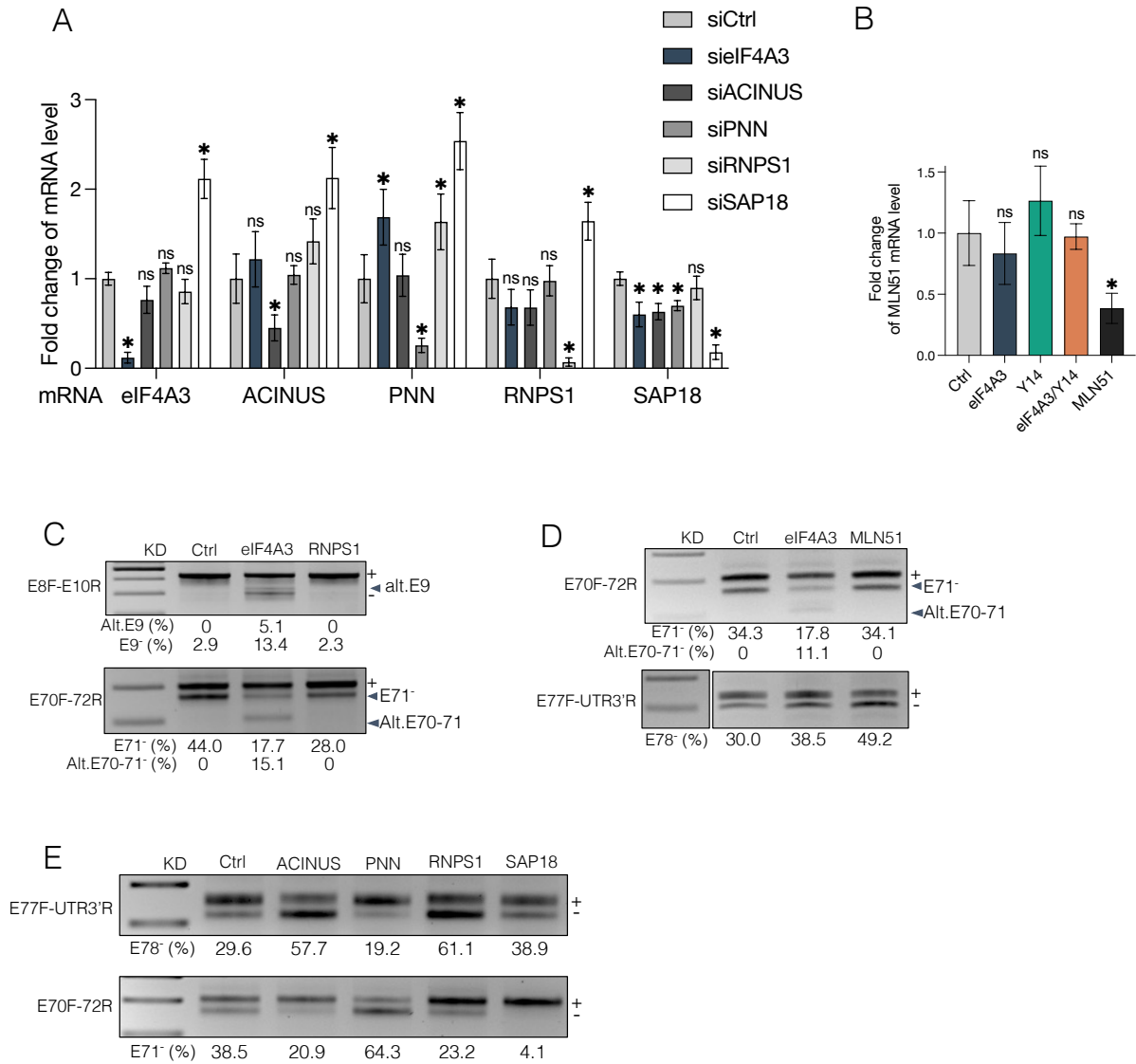
**(A)** Comparison of ASEs levels determined by DMD-targeted RNA-seq and fluorescent semi-quantitative RT-PCR under KD eIF4A3 and Y14 conditions. **(B)** Determination of intron 70 retention level in eIF4A3 KD and Y14 KD conditions by RT-qPCR. Fold changes were assessed (n=4) using three different primer sets (F1/R1, F1/R2, F2/R1) as represented in the top scheme using RPLP0 for normalization. No reverse transcriptase control (RT-). Data are mean  $\pm$  SD (\*p < 0.05; multiple Mann-Whitney test).

### Supplemental Figure 3



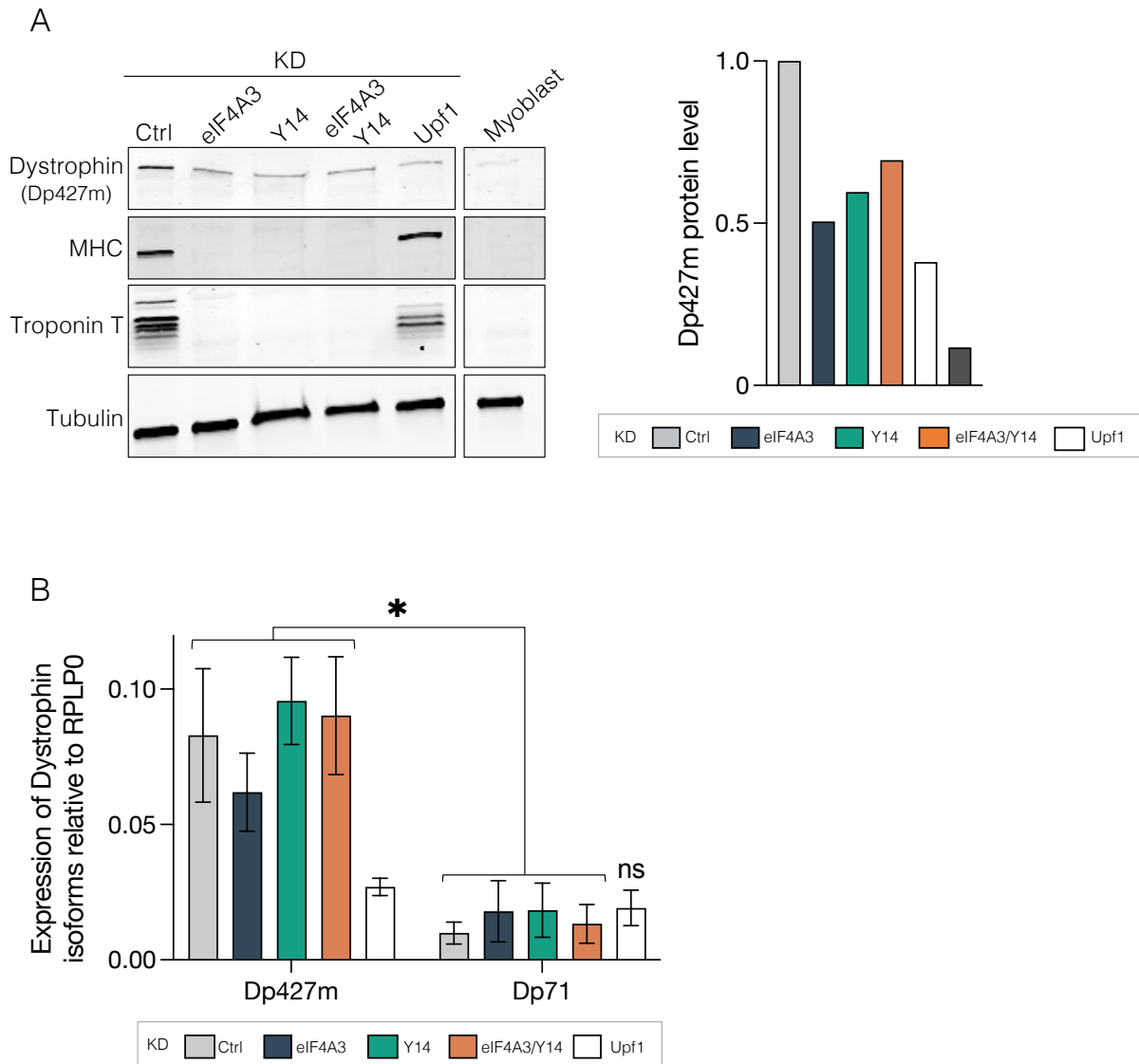
**(A)** Quantification of ASEs by fluorescent semi-quantitative PCR (n=4) in eIF4A3 and UPF1 KD compared to control condition (Ctrl) are shown as means  $\pm$  SD (\*p < 0.05; multiple Mann-Whitney test). ns, non-significant. **(B)** Representative agarose gels of out-of-frame alternative splicing events (ASE) analyzed by RT-PCR in UPF1 KD compared to core EJC components KD and control siRNA (Ctrl). The annotations are defined as in Suppl Fig.1B.

Supplemental Figure 4



Efficiency of **(A)** ASAP/PSAP components KD and **(B)** MLN51 KD was assessed at mRNA level by RT-qPCR using RPLP0 for normalization. Data are mean  $\pm$  SD (n=4) (\* $p$  < 0.05; multiple Mann-Whitney test). **(C-E)** Representative agarose gels of alternative splicing events (ASE) analyzed by RT-PCR under depletion of **(C)** RNPS1 **(D)** MLN51 and **(E)** ASAP/PSAP components compared to the control siRNA (Ctrl). The annotations are defined as in Suppl Fig. 1B.

## Supplemental Figure 5



**(A)** Western blot analysis of dystrophin, muscle Myosin Heavy Chain (MyHC) and Troponin T expression level in 3 day differentiated C25Cl48 cells depleted for EJC (eIF4A3, Y14, eIF4A3/Y14) and UPF1 compared to Myoblast (left panel) and bar graph of quantification using tubulin as a loading control (right panel). **(B)** Dp427m and Dp71 mRNA level was measured by RT-qPCR (n=4) in eIF4A3, Y14, eIF4A3/Y14, UPF1 KD conditions. Data are mean  $\pm$  SD (n=4) (\* $p$  < 0.05; multiple Mann-Whitney test).

Supplementary Tables

Table S1A

DMD_ Exons	ExonSize	PSI_siCtrl_rep1	PSI_siCtrl_rep2	PSI_siCtrl_rep3	PSI_siCtrl_rep4	PSI_MEAN_siCtrl	PSI_SD_siCtrl
1	31				1	1,00000	0
2	62	0,98487	0,98635	0,99207	0,99470	0,98950	0,00465
3	93	0,98728	0,99675	0,99595	0,99744	0,99436	0,00476
4	78	0,98454	0,99643	0,99611	0,99648	0,99339	0,00590
5	93	0,98744	0,99720	0,99658	0,99476	0,99400	0,00449
6	173	0,97840	0,98522	0,99377	0,98660	0,98600	0,00630
7	119	0,97906	0,98697	0,99433	0,98658	0,98674	0,00624
8	182	0,98705	0,98789	0,99456	0,98596	0,98887	0,00388
9	129	0,98021	0,98092	0,98253	0,98151	0,98129	0,00098
10	189	0,98551	0,98455	0,99173	0,98316	0,98624	0,00379
11	182	0,98283	0,98283	0,99207	0,98375	0,98537	0,00449
12	151	0,98075	0,97984	0,99221	0,98485	0,98441	0,00564
13	120	0,97943	0,97636	0,98836	0,98325	0,98185	0,00517
14	102	0,97562	0,97281	0,98896	0,98269	0,98002	0,00727
15	108	0,97020	0,97055	0,98854	0,97938	0,97717	0,00869
16	180	0,97443	0,97640	0,98906	0,98240	0,98057	0,00660
17	176	0,97312	0,97565	0,98920	0,98126	0,97981	0,00713
18	124	0,97238	0,97588	0,99041	0,98233	0,98025	0,00793
19	88	0,97155	0,97694	0,99108	0,98566	0,98131	0,00873
20	242	0,97079	0,97688	0,99046	0,98468	0,98070	0,00864
21	181	0,97013	0,97612	0,99018	0,98268	0,97978	0,00862
22	146	0,97159	0,97763	0,99048	0,98359	0,98082	0,00809
23	213	0,97057	0,97765	0,99012	0,98142	0,97994	0,00814
24	114	0,96737	0,97616	0,98755	0,98033	0,97785	0,00842
25	156	0,96681	0,97591	0,98696	0,98140	0,97777	0,00859
26	171	0,96823	0,97639	0,98775	0,98298	0,97884	0,00847
27	183	0,96976	0,97680	0,98803	0,98227	0,97922	0,00779
28	135	0,96859	0,97674	0,98636	0,98394	0,97891	0,00800
29	150	0,96898	0,97708	0,98524	0,98473	0,97901	0,00766
30	162	0,97428	0,97894	0,98791	0,98232	0,98086	0,00574
31	111	0,96916	0,97533	0,98627	0,97812	0,97722	0,00710
32	174	0,96942	0,97663	0,98642	0,98201	0,97862	0,00732
33	156	0,97223	0,97985	0,98796	0,98540	0,98136	0,00696
34	171	0,97055	0,97911	0,98767	0,98465	0,98050	0,00752
35	180	0,96989	0,97792	0,98582	0,98421	0,97946	0,00723
36	129	0,97146	0,97884	0,98529	0,98423	0,97996	0,00633
37	171	0,95267	0,96341	0,96352	0,97221	0,96295	0,00800
38	123	0,92953	0,94506	0,94469	0,96970	0,94725	0,01663
39	138	0,95899	0,97358	0,97882	0,97730	0,97217	0,00906
40	153	0,95996	0,97558	0,98209	0,98135	0,97475	0,01028
41	183	0,96494	0,97798	0,98252	0,98476	0,97755	0,00887
42	195	0,97123	0,98831	0,99171	0,98910	0,98509	0,00935
43	173	0,97711	0,99112	0,99411	0,99027	0,98815	0,00754
44	148	0,97491	0,99183	0,99430	0,99080	0,98796	0,00882
45	176	0,98916	0,99481	0,99467	0,99405	0,99317	0,00270
46	148	0,98990	0,99445	0,99507	0,99544	0,99372	0,00258
47	150	0,99043	0,99471	0,99516	0,99582	0,99403	0,00244

48	186	0,98780	0,99658	0,99457	0,99581	<b>0,99369</b>	<b>0,00401</b>
49	102	0,98977	0,99718	0,99518	0,99823	<b>0,99509</b>	<b>0,00377</b>
50	109	0,99150	0,99753	0,99681	0,99837	<b>0,99605</b>	<b>0,00310</b>
51	233	0,99264	0,99731	0,99554	0,99708	<b>0,99564</b>	<b>0,00215</b>
52	118	0,99053	0,99651	0,99515	0,99790	<b>0,99502</b>	<b>0,00320</b>
53	212	0,98376	0,99437	0,99504	0,99578	<b>0,99224</b>	<b>0,00568</b>
54	155	0,98021	0,99532	0,99112	0,99636	<b>0,99075</b>	<b>0,00738</b>
55	190	0,97583	0,99568	0,99060	0,99552	<b>0,98941</b>	<b>0,00935</b>
56	173	0,96537	0,99650	0,98909	0,99552	<b>0,98662</b>	<b>0,01454</b>
57	157	0,96632	0,99656	0,98922	0,99558	<b>0,98692</b>	<b>0,01411</b>
58	121	0,96748	0,99481	0,99029	0,99469	<b>0,98682</b>	<b>0,01306</b>
59	269	0,96802	0,99385	0,98988	0,99378	<b>0,98638</b>	<b>0,01238</b>
60	147	0,96696	0,99421	0,98631	0,98970	<b>0,98430</b>	<b>0,01200</b>
61	79	0,96438	0,99166	0,98783	0,98948	<b>0,98334</b>	<b>0,01274</b>
62	61	0,96242	0,99132	0,98739	0,98968	<b>0,98270</b>	<b>0,01362</b>
63	62	0,96138	0,99339	0,98759	0,99167	<b>0,98351</b>	<b>0,01495</b>
64	75	0,96523	0,99386	0,98827	0,99247	<b>0,98496</b>	<b>0,01336</b>
65	202	0,96602	0,99412	0,98955	0,99490	<b>0,98615</b>	<b>0,01362</b>
66	86	0,96410	0,99425	0,98848	0,99486	<b>0,98542</b>	<b>0,01450</b>
67	158	0,96652	0,99449	0,98855	0,99516	<b>0,98618</b>	<b>0,01344</b>
68	167	0,96920	0,98948	0,99011	0,99249	<b>0,98532</b>	<b>0,01082</b>
69	112	0,96872	0,99403	0,98969	0,99487	<b>0,98683</b>	<b>0,01228</b>
70	137	0,96493	0,99233	0,98848	0,99336	<b>0,98478</b>	<b>0,01340</b>
71	39	0,67798	0,68343	0,66266	0,66164	<b>0,67143</b>	<b>0,01095</b>
72	66	0,95998	0,98842	0,97844	0,98893	<b>0,97894</b>	<b>0,01353</b>
73	66	0,95249	0,98624	0,97770	0,98498	<b>0,97535</b>	<b>0,01570</b>
74	159	0,96015	0,98542	0,97646	0,98846	<b>0,97762</b>	<b>0,01271</b>
75	244	0,98052	0,99456	0,99298	0,99353	<b>0,99040</b>	<b>0,00662</b>
76	124	0,98431	0,99503	0,99855	0,99423	<b>0,99303</b>	<b>0,00611</b>
77	93	0,98771	0,99473	0,99914	0,99657	<b>0,99454</b>	<b>0,00490</b>
78	32	0,67480	0,69958	0,71381	0,71737	<b>0,70139</b>	<b>0,01932</b>

DMD_ Exons	ExonSize	PSI_sielF4A3 _rep1	PSI_sielF4A3 _rep2	PSI_MEAN_sielF4A3	PSI_SD_sielF4A3
1	31	0,95286	0,99569	0,97428	0,03029
2	62	0,96504	0,95926	0,96215	0,00409
3	93	0,99583	0,99596	0,99590	0,00009
4	78	0,99456	0,99648	0,99552	0,00136
5	93	0,98814	0,99584	0,99199	0,00544
6	173	0,92072	0,98607	0,95340	0,04621
7	119	0,90922	0,98579	0,94751	0,05414
8	182	0,89718	0,98417	0,94068	0,06151
9	129	0,77895	0,89504	0,83700	0,08209
10	189	0,89402	0,97663	0,93533	0,05841
11	182	0,89935	0,98081	0,94008	0,05760
12	151	0,90047	0,98438	0,94243	0,05933
13	120	0,87608	0,97673	0,92641	0,07117
14	102	0,87801	0,96812	0,92307	0,06372
15	108	0,84884	0,96304	0,90594	0,08075
16	180	0,87687	0,96824	0,92256	0,06461
17	176	0,89362	0,96516	0,92939	0,05059
18	124	0,88197	0,95482	0,91840	0,05151
19	88	0,88312	0,96350	0,92331	0,05684
20	242	0,86980	0,96068	0,91524	0,06426
21	181	0,86813	0,95699	0,91256	0,06283
22	146	0,89164	0,96203	0,92684	0,04977
23	213	0,88543	0,95905	0,92224	0,05206
24	114	0,87319	0,95799	0,91559	0,05996
25	156	0,87544	0,96163	0,91854	0,06095
26	171	0,87908	0,96825	0,92367	0,06305
27	183	0,89134	0,96943	0,93039	0,05522
28	135	0,88921	0,96883	0,92902	0,05630
29	150	0,88856	0,96733	0,92795	0,05570
30	162	0,89165	0,96727	0,92946	0,05347
31	111	0,87438	0,96005	0,91722	0,06058
32	174	0,86913	0,96205	0,91559	0,06570
33	156	0,89586	0,96670	0,93128	0,05009
34	171	0,88579	0,95988	0,92284	0,05239
35	180	0,87227	0,95431	0,91329	0,05801
36	129	0,88467	0,96186	0,92327	0,05458
37	171	0,86798	0,94656	0,90727	0,05556
38	123	0,81633	0,89330	0,85482	0,05443
39	138	0,84047	0,95812	0,89930	0,08319
40	153	0,84139	0,95342	0,89741	0,07922
41	183	0,86751	0,95923	0,91337	0,06486
42	195	0,94349	0,96178	0,95264	0,01293
43	173	0,94306	0,95831	0,95069	0,01078
44	148	0,97347	0,96356	0,96852	0,00701
45	176	0,98571	0,98857	0,98714	0,00202
46	148	0,98616	0,98347	0,98482	0,00190



47	150	0,98780	0,98761	<b>0,98771</b>	<b>0,00013</b>
48	186	0,97926	0,98046	<b>0,97986</b>	<b>0,00085</b>
49	102	0,97484	0,98094	<b>0,97789</b>	<b>0,00431</b>
50	109	0,97876	0,98355	<b>0,98116</b>	<b>0,00339</b>
51	233	0,98319	0,98548	<b>0,98434</b>	<b>0,00162</b>
52	118	0,98571	0,98544	<b>0,98558</b>	<b>0,00019</b>
53	212	0,98034	0,98162	<b>0,98098</b>	<b>0,00091</b>
54	155	0,98168	0,98554	<b>0,98361</b>	<b>0,00273</b>
55	190	0,96085	0,98186	<b>0,97136</b>	<b>0,01486</b>
56	173	0,96364	0,97549	<b>0,96957</b>	<b>0,00838</b>
57	157	0,96893	0,98117	<b>0,97505</b>	<b>0,00865</b>
58	121	0,97387	0,98192	<b>0,97790</b>	<b>0,00569</b>
59	269	0,97771	0,98686	<b>0,98229</b>	<b>0,00647</b>
60	147	0,95722	0,96512	<b>0,96117</b>	<b>0,00559</b>
61	79	0,92308	0,96718	<b>0,94513</b>	<b>0,03118</b>
62	61	0,85972	0,95357	<b>0,90665</b>	<b>0,06636</b>
63	62	0,85600	0,97107	<b>0,91354</b>	<b>0,08137</b>
64	75	0,87546	0,97347	<b>0,92447</b>	<b>0,06930</b>
65	202	0,90761	0,98234	<b>0,94498</b>	<b>0,05284</b>
66	86	0,90196	0,98066	<b>0,94131</b>	<b>0,05565</b>
67	158	0,89432	0,97729	<b>0,93581</b>	<b>0,05867</b>
68	167	0,81654	0,86284	<b>0,83969</b>	<b>0,03274</b>
69	112	0,60278	0,71754	<b>0,66016</b>	<b>0,08115</b>
70	137	0,63040	0,74773	<b>0,68907</b>	<b>0,08296</b>
71	39	0,67101	0,71984	<b>0,69543</b>	<b>0,03453</b>
72	66	0,79257	0,89459	<b>0,84358</b>	<b>0,07214</b>
73	66	0,61261	0,76096	<b>0,68679</b>	<b>0,10490</b>
74	159	0,64276	0,73074	<b>0,68675</b>	<b>0,06221</b>
75	244	0,91599	0,95444	<b>0,93522</b>	<b>0,02719</b>
76	124	0,92857	0,98210	<b>0,95534</b>	<b>0,03785</b>
77	93	0,96241	0,98954	<b>0,97598</b>	<b>0,01918</b>
78	32	0,67803	0,57572	<b>0,62688</b>	<b>0,07234</b>

<b>DMD_</b> <b>Exons</b>	<b>ExonSize</b>	<b>PSI_siY14</b> <b>_rep1</b>	<b>PSI_siY14</b> <b>_rep2</b>	<b>PSI_MEAN_</b> <b>siY14</b>	<b>PSI_SD_</b> <b>siY14</b>
1	31	0,99639	0,99890	<b>0,99765</b>	<b>0,00177</b>
2	62	0,98604	0,97327	<b>0,97966</b>	<b>0,00903</b>
3	93	0,99590	0,98792	<b>0,99191</b>	<b>0,00564</b>
4	78	0,99293	0,98506	<b>0,98900</b>	<b>0,00556</b>
5	93	0,99275	0,99039	<b>0,99157</b>	<b>0,00167</b>
6	173	0,98112	0,96183	<b>0,97148</b>	<b>0,01364</b>
7	119	0,98204	0,95866	<b>0,97035</b>	<b>0,01653</b>
8	182	0,98022	0,94971	<b>0,96497</b>	<b>0,02157</b>
9	129	0,94976	0,90852	<b>0,92914</b>	<b>0,02916</b>
10	189	0,97493	0,94608	<b>0,96051</b>	<b>0,02040</b>
11	182	0,97676	0,94819	<b>0,96248</b>	<b>0,02020</b>
12	151	0,97697	0,94632	<b>0,96165</b>	<b>0,02167</b>
13	120	0,97977	0,93482	<b>0,95730</b>	<b>0,03178</b>
14	102	0,97386	0,93568	<b>0,95477</b>	<b>0,02700</b>
15	108	0,96840	0,92649	<b>0,94745</b>	<b>0,02963</b>
16	180	0,97354	0,93974	<b>0,95664</b>	<b>0,02390</b>
17	176	0,97278	0,93976	<b>0,95627</b>	<b>0,02335</b>
18	124	0,97033	0,93310	<b>0,95172</b>	<b>0,02633</b>
19	88	0,97587	0,94044	<b>0,95816</b>	<b>0,02505</b>
20	242	0,97618	0,93972	<b>0,95795</b>	<b>0,02578</b>
21	181	0,97670	0,93631	<b>0,95651</b>	<b>0,02856</b>
22	146	0,97867	0,93949	<b>0,95908</b>	<b>0,02770</b>
23	213	0,97815	0,93099	<b>0,95457</b>	<b>0,03335</b>
24	114	0,97988	0,92361	<b>0,95175</b>	<b>0,03979</b>
25	156	0,97936	0,93571	<b>0,95754</b>	<b>0,03087</b>
26	171	0,98060	0,94167	<b>0,96114</b>	<b>0,02753</b>
27	183	0,98113	0,94343	<b>0,96228</b>	<b>0,02666</b>
28	135	0,98221	0,94490	<b>0,96356</b>	<b>0,02638</b>
29	150	0,97898	0,94294	<b>0,96096</b>	<b>0,02548</b>
30	162	0,97375	0,93961	<b>0,95668</b>	<b>0,02414</b>
31	111	0,96769	0,91771	<b>0,94270</b>	<b>0,03534</b>
32	174	0,96800	0,91951	<b>0,94376</b>	<b>0,03429</b>
33	156	0,97478	0,93254	<b>0,95366</b>	<b>0,02987</b>
34	171	0,97167	0,92995	<b>0,95081</b>	<b>0,02950</b>
35	180	0,96991	0,92189	<b>0,94590</b>	<b>0,03396</b>
36	129	0,96903	0,92714	<b>0,94809</b>	<b>0,02962</b>
37	171	0,93679	0,86000	<b>0,89840</b>	<b>0,05430</b>
38	123	0,92093	0,83558	<b>0,87826</b>	<b>0,06035</b>
39	138	0,96619	0,90999	<b>0,93809</b>	<b>0,03974</b>
40	153	0,96278	0,91212	<b>0,93745</b>	<b>0,03582</b>
41	183	0,96524	0,92258	<b>0,94391</b>	<b>0,03017</b>
42	195	0,97826	0,95430	<b>0,96628</b>	<b>0,01694</b>
43	173	0,97475	0,95128	<b>0,96302</b>	<b>0,01660</b>
44	148	0,98254	0,95089	<b>0,96672</b>	<b>0,02238</b>
45	176	0,98481	0,98608	<b>0,98545</b>	<b>0,00090</b>
46	148	0,98575	0,98984	<b>0,98780</b>	<b>0,00289</b>

47	150	0,98605	0,98844	<b>0,98725</b>	<b>0,00169</b>
48	186	0,98794	0,98712	<b>0,98753</b>	<b>0,00058</b>
49	102	0,99261	0,97525	<b>0,98393</b>	<b>0,01228</b>
50	109	0,99537	0,97623	<b>0,98580</b>	<b>0,01353</b>
51	233	0,99551	0,98277	<b>0,98914</b>	<b>0,00901</b>
52	118	0,99267	0,98948	<b>0,99108</b>	<b>0,00226</b>
53	212	0,98927	0,98149	<b>0,98538</b>	<b>0,00550</b>
54	155	0,98639	0,97927	<b>0,98283</b>	<b>0,00503</b>
55	190	0,98326	0,97534	<b>0,97930</b>	<b>0,00560</b>
56	173	0,98277	0,96595	<b>0,97436</b>	<b>0,01189</b>
57	157	0,98382	0,96589	<b>0,97486</b>	<b>0,01268</b>
58	121	0,97838	0,96374	<b>0,97106</b>	<b>0,01035</b>
59	269	0,97276	0,96435	<b>0,96856</b>	<b>0,00595</b>
60	147	0,94808	0,94206	<b>0,94507</b>	<b>0,00426</b>
61	79	0,95714	0,93639	<b>0,94677</b>	<b>0,01467</b>
62	61	0,95799	0,92136	<b>0,93968</b>	<b>0,02590</b>
63	62	0,96093	0,93564	<b>0,94829</b>	<b>0,01788</b>
64	75	0,95647	0,93981	<b>0,94814</b>	<b>0,01178</b>
65	202	0,96402	0,95343	<b>0,95873</b>	<b>0,00749</b>
66	86	0,96149	0,95730	<b>0,95940</b>	<b>0,00296</b>
67	158	0,96449	0,95810	<b>0,96130</b>	<b>0,00452</b>
68	167	0,94175	0,94874	<b>0,94525</b>	<b>0,00494</b>
69	112	0,91528	0,90966	<b>0,91247</b>	<b>0,00397</b>
70	137	0,92177	0,90886	<b>0,91532</b>	<b>0,00913</b>
71	39	0,70206	0,70536	<b>0,70371</b>	<b>0,00233</b>
72	66	0,92879	0,93900	<b>0,93390</b>	<b>0,00722</b>
73	66	0,85773	0,89186	<b>0,87480</b>	<b>0,02413</b>
74	159	0,85611	0,90198	<b>0,87905</b>	<b>0,03243</b>
75	244	0,98059	0,96439	<b>0,97249</b>	<b>0,01146</b>
76	124	0,98359	0,97189	<b>0,97774</b>	<b>0,00827</b>
77	93	0,98666	0,98832	<b>0,98749</b>	<b>0,00117</b>
78	32	0,80103	0,85512	<b>0,82808</b>	<b>0,03825</b>

Table S1B

siRNA	Junction	Event	EventSize (nucleotides)			Mean(psi5, psi3)_rep1			Mean(psi5, psi3)_rep2			Mean(psi5, psi3)_rep3			Mean(psi5, psi3)_rep4			SD(psi5, psi3)_siCtrl			Mean%(psi5, psi3)_siCtrl		
			psi5_rep1	psi3_rep1	psi3_rep1	psi5_rep2	psi3_rep2	psi3_rep2	psi5_rep3	psi3_rep3	psi3_rep3	psi5_rep4	psi3_rep4	psi3_rep4	psi5_rep4	psi3_rep4	psi3_rep4	psi5_siCtrl	psi3_siCtrl	psi3_siCtrl	psi5_siCtrl	psi3_siCtrl	psi3_siCtrl
siCtrl	junction_E8-E10	E9 skip	lost of 129 nt	0,0070	0,0077	0,0073	0,0065	0,0073	0,0069	0,0114	0,0120	0,0117	0,0041	0,0045	0,0043	<b>0,0075</b>	<b>0,0031</b>	<b>0,75%</b>					
siCtrl	junction_E9+6bp-E10	Alt5'ss-E9	lost of 123 nt	N/A	0,0000	0,0000	N/A	0,0004	0,0004	N/A	0,0000	0,0000	N/A	0,0000	0,0000	<b>0,0001</b>	<b>0,0002</b>	<b>0,01%</b>					
siCtrl	junction_E67-E69	E68 skip	lost of 167 nt	0,0014	0,0013	0,0013	0,0054	0,0048	0,0051	0,0000	0,0000	0,0000	0,0028	0,0023	0,0025	<b>0,0022</b>	<b>0,0022</b>	<b>0,22%</b>					
siCtrl	junction_E68-E70	E69 skip	lost of 112 nt	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	<b>0,0000</b>	<b>0,0000</b>	<b>0,00%</b>					
siCtrl	junction_E68-E71	E69-70 skip	lost of 249 nt	0,0005	0,0009	0,0007	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	<b>0,0002</b>	<b>0,0004</b>	<b>0,02%</b>					
siCtrl	junction_E70-E72	E71 skip	lost of 39 nt	0,2907	0,2866	0,2887	0,3017	0,3144	0,3081	0,3145	0,3292	0,3218	0,3132	0,3499	0,3316	<b>0,3125</b>	<b>0,0186</b>	<b>31,25%</b>					
siCtrl	junction_E70+36bp-E71	Alt5'ss-E70	lost of 101 nt	N/A	0,0000	0,0000	N/A	0,0000	0,0000	N/A	0,0000	0,0000	N/A	0,0000	0,0000	<b>0,0000</b>	<b>0,0000</b>	<b>0,00%</b>					
siCtrl	junction_E70+36bp-E72	Alt5'ss-E70 E71 skip	lost of 140 nt	N/A	0,0000	0,0000	N/A	0,0000	0,0000	N/A	0,0000	0,0000	N/A	0,0000	0,0000	<b>0,0000</b>	<b>0,0000</b>	<b>0,00%</b>					
siCtrl	junction_E72-E75	E73-74 skip	lost of 225 nt	0,0006	0,0006	0,0006	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	<b>0,0001</b>	<b>0,0003</b>	<b>0,01%</b>					
siCtrl	junction_E77-E79	E78 skip	lost of 32 nt	0,3164	0,3131	0,3147	0,2921	0,2970	0,2946	0,2835	0,2855	0,2845	0,2777	0,2787	0,2782	<b>0,2930</b>	<b>0,0160</b>	<b>29,30%</b>					

siRNA	Junction	Event	EventSize	psi5_rep1	psi3_rep1	Mean(psi5, psi3)_rep1	psi5_rep2	psi3_rep2	Mean(psi5, psi3)_rep2	Mean(psi5, psi3)_sielF4A3	SD(psi5, psi3)_sielF4A3	Mean(psi5, psi3)_siCtrl	SD(psi5, psi3)_siCtrl	ΔSJ	ΔSJ (%)
sielF4A3	junction_E8-E10	E9 skip	lost of 129 nt	0,1206	0,1188	0,1197	0,0781	0,0859	0,0820	<b>0,1008</b>	<b>0,0267</b>	0,0075	0,0031	0,09	9,33%
sielF4A3	junction_E9+6bp-E10	Alt5'ss-E9	lost of 123 nt	N/A	0,0383	0,0383	N/A	0,0563	0,0563	<b>0,0473</b>	<b>0,0127</b>	0,0001	0,0002	0,05	4,72%
sielF4A3	junction_E67-E69	E68 skip	lost of 167 nt	0,0899	0,1276	0,1087	0,1160	0,1485	0,1322	<b>0,1205</b>	<b>0,0166</b>	0,0022	0,0022	0,12	11,82%
sielF4A3	junction_E68-E70	E69 skip	lost of 112 nt	0,1094	0,1542	0,1318	0,1015	0,1175	0,1095	<b>0,1206</b>	<b>0,0158</b>	0,0000	0,0000	0,12	12,06%
sielF4A3	junction_E68-E71	E69-70 skip	lost of 249 nt	0,1844	0,2995	0,2419	0,1498	0,2605	0,2051	<b>0,2235</b>	<b>0,0260</b>	0,0002	0,0004	0,22	22,33%
sielF4A3	junction_E70-E72	E71 skip	lost of 39 nt	0,1739	0,0968	0,1353	0,1553	0,0829	0,1191	<b>0,1272</b>	<b>0,0115</b>	0,3125	0,0186	-0,19	-18,53%
sielF4A3	junction_E70+36bp-E71	Alt5'ss-E70	lost of 101 nt	N/A	0,1574	0,1574	N/A	0,2163	0,2163	<b>0,1869</b>	<b>0,0417</b>	0,0000	0,0000	0,19	18,69%
sielF4A3	junction_E70+36bp-E72	Alt5'ss-E70 E71 skip	lost of 140 nt	N/A	0,0565	0,0565	N/A	0,0829	0,0829	<b>0,0697</b>	<b>0,0187</b>	0,0000	0,0000	0,07	6,97%
sielF4A3	junction_E72-E75	E73-74 skip	lost of 225 nt	0,2159	0,1702	0,1930	0,1153	0,1424	0,1288	<b>0,1609</b>	<b>0,0454</b>	0,0001	0,0003	0,16	16,08%
sielF4A3	junction_E77-E79	E78 skip	lost of 32 nt	0,3254	0,3096	0,3175	0,4164	0,4115	0,4139	<b>0,3657</b>	<b>0,0682</b>	0,2930	0,0160	0,07	7,27%

siRNA	Junction	Event	EventSize	psi5_rep1	psi3_rep1	Mean(Psi5, psi3)_rep1	psi5_rep2	psi3_rep2	Mean(Psi5, psi3)_rep2	Mean(psi5, psi3)_siY14	SEM(psi5, psi3)_siY14	Mean(psi5, psi3)_siCtrl	SD(psi5, psi3)_siCtrl	ΔSJ	ΔSJ %
siY14	junction_E70-E72	E71 skip	lost of 39 nt	0,2478	0,2252	0,2365	0,2472	0,2156	0,2314	0,2340	<b>0,0036</b>	0,3125	0,0186	-0,08	-7,9%
siY14	junction_E70+36bp-E71	Alt5'ss-E70	lost of 101 nt	N/A	0,0794	0,0794	N/A	0,0300	0,0300	0,0547	<b>0,0349</b>	0,0000	0,0000	0,05	5,5%
siY14	junction_E72-E75	E73-74 skip	lost of 225 nt	0,0682	0,0661	0,0671	0,0393	0,0380	0,0387	0,0529	<b>0,0201</b>	0,0001	0,0003	0,05	5,3%
siY14	junction_E77-E79	E78 skip	lost of 32 nt	0,1925	0,1911	0,1918	0,1399	0,1361	0,1380	0,1649	<b>0,0380</b>	0,2930	0,0160	-0,13	-12,8%

siRNA	Junction	Event	EventSize	psi5_rep1	psi3_rep1	Mean(Psi5, psi3)_rep1	psi5_rep2	psi3_rep2	Mean(Psi5, psi3)_rep2	psi5_rep3	psi3_rep3	Mean(Psi5, psi3)_rep3	psi5_rep4	psi3_rep4	Mean(Psi5, psi3)_rep4	Mean(Psi5, psi3)_siCtrl	SD(Psi5, psi3)_siCtrl	Mean%(Psi5, psi3)_siCtrl
siCtrl	junction_E1-E1a	pseudo E1a inclusion from IVS1	gain of 163nt	0,1798	N/A	0,1795	0,1194	N/A	0,1301	0,1008	N/A	0,1088	0,1001	N/A	0,1046	<b>0,1120</b>	<b>0,0104</b>	<b>13,08%</b>
siCtrl	junction_E1a-E2	E2 skip	lost of 62nt	N/A	0,1792		N/A	0,1408		N/A	0,1167		N/A	0,1092				
siCtrl	junction_E1-E3	pseudo E2a inclusion from IVS2	gain of 140nt	0,0101	0,0106	0,0104	0,0010	N/A	0,0010	0,0006	N/A	0,0006	0,0013	N/A	0,0013	<b>0,0078</b>	<b>0,0034</b>	<b>0,84%</b>
siCtrl	junction_E2-E2a	pseudo E7x inclusion from IVS7	gain of 159nt	N/A	0,0010	0,0010	N/A	0,0023	0,0017	N/A	0,0032	0,0019	N/A	0,0026	0,0020	<b>0,0015</b>	<b>0,0007</b>	<b>0,16%</b>
siCtrl	junction_E7-E7x	E9 skip	lost of 129nt	0,0022	N/A	0,0015	0,0004	N/A	0,0002	0,0019	N/A	0,0017	0,0026	N/A	0,0013	<b>0,0009</b>	<b>0,0006</b>	<b>0,12%</b>
siCtrl	junction_E7x-E8	Alt5'ss-E9	lost of 123nt	N/A	0,0008		N/A	0,0000		N/A	0,0014		N/A	0,0000				
siCtrl	junction_E8-E10	Alt5'ss-E14	lost of 99nt	0,0070	0,0077	0,0073	0,0065	0,0073	0,0069	0,0114	0,0120	0,0117	0,0041	0,0045	0,0043	<b>0,0076</b>	<b>0,0040</b>	<b>0,75%</b>
siCtrl	junction_E9+6bp-E10	Alt5'ss-E34x inclusion from IVS34	gain of 89nt	N/A	0,0000	0,0000	N/A	0,0004	0,0004	N/A	0,0000	0,0000	N/A	0,0000	0,0000	<b>0,0000</b>	<b>0,0000</b>	<b>0,01%</b>
siCtrl	junction_E14+3bp-E15	pseudo E34x inclusion from IVS34	gain of 89nt	N/A	0,0064	0,0064	N/A	0,0060	0,0060	N/A	0,0025	0,0025	N/A	0,0267	0,0267	<b>0,0146</b>	<b>0,0171</b>	<b>1,04%</b>
siCtrl	junction_E34-E34x	Alt5'ss-E35	lost of 136nt	0,0038	N/A	0,0040	0,0025	N/A	0,0015	0,0005	N/A	0,0005	0,0033	N/A	0,0037	<b>0,0031</b>	<b>0,0009</b>	<b>0,24%</b>
siCtrl	junction_E34x-E35	E37 skip	lost of 171nt	N/A	0,0041		N/A	0,0005		N/A	0,0005		N/A	0,0041				
siCtrl	junction_E35+34bp-E36	E38 skip	lost of 123nt	N/A	0,0007	0,0007	N/A	0,0000	0,0000	N/A	0,0000	0,0000	N/A	0,0000	0,0000	<b>0,0000</b>	<b>0,0000</b>	<b>0,02%</b>
siCtrl	junction_E36-E38	E69-70 skip	lost of 249nt	0,0149	0,0173	0,0161	0,0158	0,0190	0,0174	0,0198	0,0237	0,0218	0,0120	0,0142	0,0131	<b>0,0175</b>	<b>0,0055</b>	<b>1,71%</b>
siCtrl	junction_E37-E39	E69-71 skip	lost of 288nt	0,0358	0,0361	0,0359	0,0326	0,0322	0,0324	0,0384	0,0377	0,0380	0,0115	0,0123	0,0119	<b>0,0274</b>	<b>0,0136</b>	<b>2,96%</b>
siCtrl	junction_E67-E69	E69-74 skip	lost of 579nt	0,0014	0,0013	0,0013	0,0054	0,0048	0,0051	0,0000	0,0000	0,0000	0,0028	0,0023	0,0025	<b>0,0026</b>	<b>0,0027</b>	<b>0,22%</b>
siCtrl	junction_E68-E70	E71 skip	lost of 140nt	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	<b>0,0000</b>	<b>0,0000</b>	<b>0,00%</b>
siCtrl	junction_E68-E71	E75 skip	lost of 431nt	0,0005	0,0009	0,0007	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	<b>0,0000</b>	<b>0,0000</b>	<b>0,02%</b>
siCtrl	junction_E68-E72	E71-74 skip	lost of 330nt	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	<b>0,0000</b>	<b>0,0000</b>	<b>0,00%</b>
siCtrl	junction_E68-E75	Alt5'ss-E70 E71 skip	lost of 140nt	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	<b>0,0000</b>	<b>0,0000</b>	<b>0,00%</b>
siCtrl	junction_E70+36bp-E72	Alt5'ss-E70 E75 skip	lost of 431nt	N/A	0,0000	0,0000	N/A	0,0000	0,0000	N/A	0,0000	0,0000	N/A	0,0000	0,0000	<b>0,0000</b>	<b>0,0000</b>	<b>0,00%</b>
siCtrl	junction_E70+36bp-E75	E71-74 skip	lost of 330nt	N/A	0,0000	0,0000	N/A	0,0000	0,0000	N/A	0,0000	0,0000	N/A	0,0000	0,0000	<b>0,0000</b>	<b>0,0000</b>	<b>0,00%</b>
siCtrl	junction_E70-E75	E71-74 skip	lost of 330nt	0,0088	0,0076	0,0082	0,0033	0,0034	0,0033	0,0080	0,0076	0,0078	0,0026	0,0030	0,0028	<b>0,0046</b>	<b>0,0026</b>	<b>0,55%</b>
siCtrl	junction_E71-E75	E72-74 skip	lost of 291nt	0,0005	0,0003	0,0004	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	<b>0,0000</b>	<b>0,0000</b>	<b>0,01%</b>
siCtrl	junction_E75-76+59bp	Alt3'ss-E76	lost of 59nt	0,0016	N/A	0,0016	0,0050	N/A	0,0050	0,0007	N/A	0,0007	0,0017	N/A	0,0017	<b>0,0034</b>	<b>0,0023</b>	<b>0,22%</b>

siRNA	Junction	Event	EventSize	psi5_rep1	psi3_rep1	Mean(Psi5, psi3)_rep1	psi5_rep2	psi3_rep2	Mean(Psi5, psi3)_rep2	Mean(Psi5, psi3)_rep_siY14	SD(Psi5, psi3)_siY14	Mean(Psi5, psi3)_Ctrl	SD(Psi5, psi3)_siCtrl	ΔSJ	ΔSJ %
siY14	junction_E1-E1a	pseudo E1a inclusion from IVS1	gain of 163nt	0,1273	N/A	0,1348	0,1653	N/A	0,1705	<b>0,1527</b>	<b>0,0252</b>	<b>0,1308</b>	<b>0,0344</b>	0,02	2,19%
siY14	junction_E1a-E2			N/A	0,1424		N/A	0,1758							
siY14	junction_E2-E2a	pseudo E2a inclusion from IVS2	gain of 140nt	0,0063	N/A	0,0056	0,0086	N/A	0,0172	<b>0,0114</b>	<b>0,0082</b>	<b>0,0016</b>	<b>0,0004</b>	0,01	0,98%
siY14	junction_E2a-E3			N/A	0,0048		N/A	0,0259							
siY14	junction_E7-E7x	pseudo E7x inclusion from IVS7	gain of 159nt	0,0279	N/A	0,0162	0,0414	N/A	0,0237	<b>0,0200</b>	<b>0,0052</b>	<b>0,0012</b>	<b>0,0007</b>	0,02	1,88%
siY14	junction_E7x-E8			N/A	0,0046		N/A	0,0059							
siY14	junction_E8-E10	E9 skip	lost of 129nt	0,0295	0,0322	0,0308	0,0374	0,0391	0,0383	<b>0,0346</b>	<b>0,0053</b>	<b>0,0075</b>	<b>0,0031</b>	0,03	2,70%
siY14	junction_E9+6bp-E10	Alt5'ss-E9	lost of 123nt	N/A	0,0175	0,0175	N/A	0,0095	0,0095	<b>0,0135</b>	<b>0,0057</b>	<b>0,0001</b>	<b>0,0002</b>	0,01	1,34%
siY14	junction_E14+3bp-E15	Alt5'ss-E14	lost of 99nt	N/A	0,0000	0,0000	N/A	0,0000	0,0000	<b>0,0000</b>	<b>0,0000</b>	<b>0,0104</b>	<b>0,0110</b>	-0,01	-1,04%
siY14	junction_E36-E38	E37 skip	lost of 171nt	0,0305	0,0350	0,0328	0,0627	0,0941	0,0784	<b>0,0556</b>	<b>0,0323</b>	<b>0,0171</b>	<b>0,0036</b>	0,04	3,85%
siY14	junction_E37-E39	E38 skip	lost of 123nt	0,0557	0,0477	0,0517	0,0948	0,0934	0,0941	<b>0,0729</b>	<b>0,0300</b>	<b>0,0296</b>	<b>0,0120</b>	0,04	4,34%
siY14	junction_E67-E69	E68 skip	lost of 167nt	0,0340	0,0308	0,0324	0,0173	0,0152	0,0163	<b>0,0243</b>	<b>0,0114</b>	<b>0,0022</b>	<b>0,0022</b>	0,02	2,21%
siY14	junction_E68-E70	E69 skip	lost of 112nt	0,0217	0,0252	0,0235	0,0165	0,0214	0,0190	<b>0,0212</b>	<b>0,0032</b>	<b>0,0000</b>	<b>0,0000</b>	0,02	2,12%
siY14	junction_E68-E71	E69-70 skip	lost of 249nt	0,0300	0,0577	0,0439	0,0257	0,0560	0,0409	<b>0,0424</b>	<b>0,0021</b>	<b>0,0002</b>	<b>0,0004</b>	0,04	4,22%
siY14	junction_E70+36bp-E72	Alt5'ss-E70+E71 skip	lost of 140nt	N/A	0,0270	0,0270	N/A	0,0170	0,0170	<b>0,0220</b>	<b>0,0071</b>	<b>0,0000</b>	<b>0,0000</b>	0,02	2,20%
siY14	junction_E75-76+59bp	Alt3'ss-E76	lost of 59nt	0,0131	N/A	0,0131	0,0133	N/A	0,0133	<b>0,0132</b>	<b>0,0001</b>	<b>0,0022</b>	<b>0,0019</b>	0,01	1,09%

siRNA	Junction	Event	EventSize	psi5_rep1	psi3_rep1	Mean(Psi5, psi3)_rep1	psi5_rep2	psi3_rep2	Mean(Psi5, psi3)_rep2	Mean(Psi5, psi3)_rep_siIF4A3	SD(Psi5, psi3)_siIF4A3	Mean(Psi5, psi3)_Ctrl	SD(Psi5, psi3)_siCtrl	ΔSJ	ΔSJ %
siIF4A3	junction_E1-E1a	pseudo E1a inclusion from IVS1	gain of 163nt	0,1319	N/A	0,1205	0,1858	N/A	0,1822	<b>0,1513</b>	<b>0,0436</b>	<b>0,1308</b>	<b>0,0344</b>	0,02	2,06%
siIF4A3	junction_E1a-E2			N/A	0,1091		N/A	0,1785							
siIF4A3	junction_E1-E3	E2 skip	lost of 62nt	0,0200	0,0224	0,0212	0,0265	0,0253	0,0259	<b>0,0236</b>	<b>0,0033</b>	<b>0,0084</b>	<b>0,0028</b>	0,02	1,52%
siIF4A3	junction_E14+3bp-E15	Alt5'ss-E14	lost of 99nt	N/A	0,0478	0,0478	N/A	0,0332	0,0332	<b>0,0405</b>	<b>0,0103</b>	<b>0,0104</b>	<b>0,0110</b>	0,03	3,01%
siIF4A3	junction_E34-E34x	pseudo E34x inclusion from IVS34	gain of 89nt	0,0172	N/A	0,0155	0,0134	N/A	0,0151	<b>0,0153</b>	<b>0,0003</b>	<b>0,0024</b>	<b>0,0017</b>	0,01	1,29%
siIF4A3	junction_E34x-E35			N/A	0,0138		N/A	0,0167							
siIF4A3	junction_E35+34bp-E36	Alt5'ss-E35	lost of 136nt	N/A	0,0217	0,0217	N/A	0,0407	0,0407	<b>0,0312</b>	<b>0,0134</b>	<b>0,0002</b>	<b>0,0003</b>	0,03	3,11%
siIF4A3	junction_E37-E39	E38 skip	lost of 123nt	0,0554	0,0655	0,0604	0,0748	0,0752	0,0750	<b>0,0677</b>	<b>0,0103</b>	<b>0,0296</b>	<b>0,0120</b>	0,04	3,82%
siIF4A3	junction_E68-E72	E69-71 skip	lost of 288nt	0,0188	0,0242	0,0215	0,0203	0,0276	0,0240	<b>0,0227</b>	<b>0,0018</b>	<b>0,0000</b>	<b>0,0000</b>	0,02	2,27%
siIF4A3	junction_E68-E75	E69-74 skip	lost of 579nt	0,0188	0,0179	0,0183	0,0165	0,0226	0,0195	<b>0,0189</b>	<b>0,0009</b>	<b>0,0000</b>	<b>0,0000</b>	0,02	1,89%
siIF4A3	junction_E70+36bp-E75	Alt5'ss-E70+E75 skip	lost of 431nt	N/A	0,0209	0,0209	N/A	0,0174	0,0174	<b>0,0191</b>	<b>0,0025</b>	<b>0,0000</b>	<b>0,0000</b>	0,02	1,91%
siIF4A3	junction_E70-E75	E71-74 skip	lost of 330nt	0,0507	0,0209	0,0358	0,0777	0,0417	0,0597	<b>0,0477</b>	<b>0,0169</b>	<b>0,0055</b>	<b>0,0029</b>	0,04	4,22%
siIF4A3	junction_E71-E75	E72-74 skip	lost of 291nt	0,0465	0,0299	0,0382	0,0127	0,0104	0,0116	<b>0,0249</b>	<b>0,0188</b>	<b>0,0001</b>	<b>0,0002</b>	0,02	2,48%
siIF4A3	junction_E75-76+59bp	Alt3'ss-E76	lost of 59nt	0,0378	N/A	0,0378	0,0148	N/A	0,0148	<b>0,0263</b>	<b>0,0163</b>	<b>0,0022</b>	<b>0,0019</b>	0,02	2,40%

Supplementary Table S1. Data from *DMD*-targeted RNA-Seq. (A) *DMD* Exon inclusion levels in Ctrl, eIF4A3 and Y14 KD based on PSI calculation. (B) Splicing Junction (SJ) usage in Ctrl compared to eIF4A3 and Y14 KD analyzed with IPSA software ( $|\Delta SJ| \geq 0.05$  and  $(0.05 \geq |\Delta SJ| \geq 0.01)$ ).

Table S2

ASE	impact on reading frame	C25cl48 cell line						$\Delta$ SJ (%)			
		si-Ctrl		si-eIF4A3		si-Y14		si-eIF4A3 vs si-Ctrl		si-Y14 vs si-Ctrl	
		RNA-seq n=4	RT-QFPCR n=4	RNA-seq n=2	RT-QFPCR n=4	RNA-seq n=2	RT-QFPCR n=4	RNA-seq	RT-QFPCR	RNA-seq	RT-QFPCR
E9	IF	0,75	4,22	10,08	13,84	3,46	9,90	9,33	9,62	2,70	5,68
alt5'ssE9	IF	0,01	0,00	4,73	9,19	1,35	1,05	4,72	9,19	1,34	1,05
E68	OOF	0,22	3,71	12,05	19,31	2,43	8,38	11,82	15,60	2,21	4,67
E69	OOF	0,00	0,00	12,06	5,11	2,12	2,07	12,06	5,11	2,12	2,07
E69-70	IF	0,02	0,00	22,35	16,26	4,24	3,67	22,33	16,26	4,22	3,67
E69-70-71	IF	0,00	0,00	2,27	6,02	0,87	2,55	2,27	6,02	0,87	2,55
E71_E73-74	IF	0,00	0,06	0,00	2,54	0,00	0,60	0,00	2,48	0,00	0,54
E71-74	IF	0,55	0,59	4,77	5,90	1,05	1,77	4,22	5,31	0,49	1,18
E72-74	IF	0,01	0,00	2,49	2,15	0,99	0,15	2,48	2,15	0,98	0,15
E73-74	IF	0,01	0,18	16,09	17,47	5,29	4,86	16,08	17,30	5,28	4,69
E78	elongated C-term	29,30	31,59	36,57	38,58	16,49	22,89	7,27	6,99	-12,81	-8,70
alt5'ssE70	OOF	0,00	0,00	18,69	6,60	5,47	1,95	18,69	6,60	5,47	1,95
E69_alt5'ssE70	OOF	0,00	0,00	0,00	3,33	0,00	1,10	0,00	3,33	0,00	1,10
alt5'ssE70-71	OOF	0,00	0,00	6,97	4,68	2,20	1,58	6,97	4,68	2,20	1,58
E69_alt5'ssE70-71	OOF	0,00	0,00	0,00	2,05	0,00	0,74	0,00	2,05	0,00	0,74
alt5'ssE70_E73-74	OOF	0,00	0,00	0,00	2,90	0,00	0,85	0,00	2,90	0,00	0,85
alt5'ssE70-71_E73-74	OOF	0,00	0,00	0,00	1,74	0,00	0,00	0,00	1,74	0,00	0,00
alt5'ssE70-74	OOF	0,00	0,08	1,91	3,80	2,46	0,13	1,91	3,72	2,46	0,05

Supplementary Table S2. Comparison of all ASEs detected by fluorescent semi-quantitative RT-PCR (RT-QFPCR) and *DMD*-targeted RNA-Seq in Ctrl, eIF4A3 and Y14 KD. The  $\Delta$ SJ shows the changes in the ASE levels between EJC KD and the Ctrl condition. (IF) In frame, (OOF) Out of frame.



Table S3

A

Target site sequences of siRNA	
Name	Sequences
#5 (Ctrl)	UGGUUUACAUGUCGACUAA
eIF4A3	AGACAUGACUAAAGUGGAAUU
Y14	CGCUCUGUUGAAGGCUGGAUU
RNPS1	GCAUCCAGCCGCUCAGGAA
UPF1	GAUGCAGUUC CGCUCCAUUUU
MLN51	CCAGCAUACAUACCUCGGAUU
ACINUS	GCUCGCUGCCCAAUCAUUUU
PININ	GGUAAGGUGGCUCAGCGAGUU
SAP18	GGAUGCAACCUUGAAAGAACUGACAUU

B

RT-PCR primers of <i>DMD</i> gene	
Name	sequences
Exon 8 F *	AGTGAGCATTGAAGCCATCC
Exon 10 R	GCAATGTGTCTCAGCAGAA
Exon 67 F *	CAGTTGGGTGAAGTTGC
Exon 69 R	GCCTTTTGCAACTCGACCAGA
Exon 68 F *	CTTCCTAGACTGGATGAGACTG
Exon 70 R	TCTCGAACATCTTCTCCTGATG
Exon 68 F	GCAGAAACTGCCAAGCATCAG
Exon 71 R	CAGAAGTTGATCAGAGTAACGG
Exon 70 F *	CATCAGGAGAAGATGTTTCGAGA
Exon 72 R	TCATCGTGTGAAAGCTGAGG
Exon 72 F *	CCTCAGCTTTCACACGATGA
Exon 75 R	AACCACTCGGAGCAGCATAG
Exon 77 F *	AGTCCTCCCAGGACACAA
Exon 10dR	CAAATCATCTGCCATGTGG

\*FAM modification in each forward primers for SQF-PCR

C

qPCR primers	
Name	Sequences
eIF4A3-F	CCAGCCGAGTGCTTATTTCTAC
eIF4A3-R	CCTGATCTCCCAATTCTGTGTA
Y14-F	GGACCACAACGCTCTGTTG
Y14-R	CAACTAGAGTATACCCCTTCAG
RNPS1-F	GGACAAAACCCGAAAGAGGC
RNPS1-R	CCTCCTGTTGTCGTGTCTGC
UPF1-F	AGATCACGGCACAGCAGAT
UPF1-R	TGGCAGAAGGGTTTTCTT
MLN51-F	AGCCTGACACCAAAGCACT
MLN51-R	TAGGGCCCTTTTTACCCACT
PININ-F	ACAAGAATCCACTGTTGCTACTG
PININ-R	CAAAAGCCGCAGTTCTGTCT
ACINUS-F	CAGCCACCACACAGAAGAAA
ACINUS-R	GCGAGAGTCATCAGCATGAA
SAP18-F	AATGTACCGTCCAGCGAGTT
SAP18-R	GCCCTTCTTTCTAGCTTCTGG
Dp427m-F	TAGAGGACTGTTATGAAAGAGAAG
Dp427m-R	CAGGGCATGAACTCTTGTG
Dp71-F	ATGAGGGAACAGCTCAAAGG
Dp71-R	CTGGTCCAAGGCATCACAT
CyclinA-F	GCACTCTACACAGTCACGG
CyclinA-R	GTCTCTGGTGGGTTGAGGA
Myf5-F	CATGCCCGAATGTAACAGTC
Myf5-R	CCCAGGTTGCTCTGAGG
MyoD-F	ACAACGGACGACTTCTATGAC
MyoD-R	TGCTCTTCGGGTTTCAGGA
MyoG-F	ACCCCGCTTCTATGATGG
MyoG-R	ACACCGACTTCCTCTTACACA
ACTA1-F	GACTTCTCAGGACGACGAATC
ACTA1-R	CATTTTCTTCCACAGGGCTT
CKM-F	CAAGCACCCCAAGTTCG
CKM-R	GTCGTCAATGGACTGGC
SRSF3-FL-F	GTGAAAAAAGAAGTAGAAATCGTGG
SRSF3-FL-R	CTCCTTCTTGGAGATCTGCCGACGAG

D

Antibodies		
Name	Dilutions	Origin
Rabbit-anti-eIF4A3	1:20000	Kindly provided by Hervé Le Hir
Rabbit-anti-Y14	1:250	
Rabbit-anti-MLN51	1 :500	
Goat-anti-UPF1	1 :500	Bethyl
Rabbit-anti-ACINUS	1:100	Invitrogen
Mouse-anti-SAP18	1 :250	Santa Cruz
Rabbit-anti-GAPDH	1:4000	
Mouse-anti-MyHC-My32	1:8000	Sigma
Mouse-anti-Troponin T	1:2500	
Mouse-anti-Dystrophin-Dy4	1:100	GTX
Rabbit-anti-PININ	1:250	Protein Tech
Rabbit-anti-RNPS1	1:1000	
Mouse-anti-Tubulin	1:20000	

Supplementary Table S3. Tables of (A) siRNA, (B) RT-PCR primers, (B) qPCR primers and (D) antibodies used in this study.

## Résultats complémentaires

Les résultats présentés dans cette partie ont été obtenus lors d'expériences complémentaires réalisées durant mon projet de thèse afin d'explorer plus en détail 1) le rôle de l'EJC dans l'épissage de *DMD* (couplage avec la transcription, présence de dépôts canoniques et non canoniques) et 2) le rôle des facteurs UPF1 et UPF2 dans l'épissage des exons 71 et 78 de la *Dp427m* et la différenciation musculaire dans notre modèle cellulaire C25C148. Il s'agit pour partie de résultats préliminaires qui nécessitent des validations.

### I. Rôle de l'EJC dans l'épissage de *DMD* et transcription

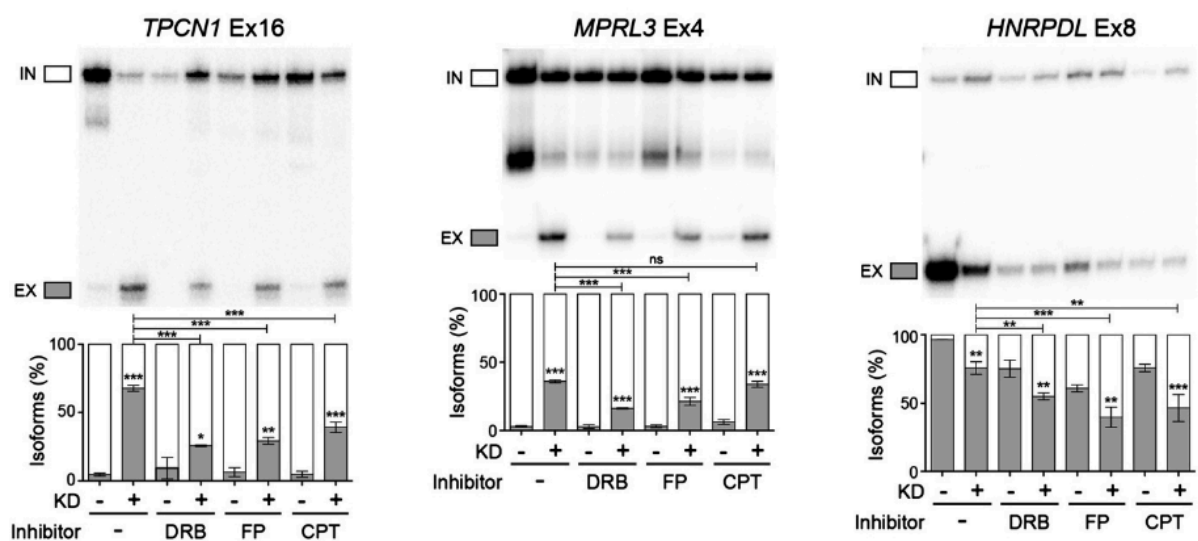
#### A. Introduction

Initialement estimé à 16 heures (Tennyson, 1996), le temps de transcription par l'ARN Pol II du gène *DMD* a été ré-évalué à environ 10 heures sur la base de l'étude de Singh & Padgett, 2009. Les auteurs ont étudié dans des lignées cellulaires la cinétique de transcription par la Pol II de différents grands gènes et montré que celle-ci, estimée à 3,8kb par minute, était constante sur de longues distances et comparable entre les gènes testés. Ils ont aussi pu déterminer que l'épissage co-transcriptionnel du pré-ARNm des introns dépendants de U2 se produisait dans les 5 à 10 minutes suivant la synthèse, quelle que soit la longueur de l'intron considéré (entre 1 kb et 240 kb).

Différents paramètres du processus de transcription tels que la pause proximale de la Pol II au promoteur, la phosphorylation du CTD de la Pol II par pTEFb ou la vitesse d'élongation de la Pol II peuvent impacter l'épissage alternatif des gènes en interférant notamment avec le recrutement de facteurs d'épissage (Braunschweig et al., 2013; Giono and Kornblihtt, 2020; Ip et al., 2011). C'est le cas notamment de l'EJC qui inhibe pTEFb et permet la pause proximale de la Pol II au promoteur comme nous l'avons vu dans l'étude bibliographique (Akhtar et al., 2019).

Différentes drogues sont utilisées pour étudier le couplage de l'épissage avec la transcription. Le DRB (5,6-dichloro-1- $\beta$ -D-ribofuranosyl-benzimidazole) et le Flavopiridol (FP) inhibent le facteur d'élongation pTEFb et la phosphorylation du domaine CTD de la Pol II (Blagosklonny, 2004; Zorio and Bentley, 2001). La Camptothécine (CPT) inhibe la topoisomérase I induisant un échappement de la Pol II du site de pause proximale et par ailleurs une hyperphosphorylation de la Pol II qui ne peut plus se fixer sur les promoteurs. Cela modifie globalement la densité de Pol II sur les promoteurs et par conséquent la transcription des gènes (Baranello et al., 2010; Darzacq et al., 2007).

Ces trois drogues ont été utilisées en condition de déplétion d'eIF4A3 dans des cellules HeLa (Wang et al., 2014). Dans ces cellules, la baisse d'expression d'eIF4A3 et de dépôt d'EJC avaient conduit à différents sauts d'exon qui ont pu être partiellement corrigés (ré-inclusion des exons) par l'inhibition de la transcription par les différentes drogues utilisées (Figure 67). L'inhibition de la transcription a permis de compenser en partie l'absence de dépôt d'EJC dans les cellules traitées par le ARNi contre eIF4A3 (KD). Comme décrit dans la littérature, une vitesse lente d'élongation de la Pol II a favorisé l'inclusion d'exons en permettant une meilleure interaction des facteurs d'épissage, notamment avec les sites faibles d'épissage des exons alternatifs, démontrant que la régulation de l'épissage par l'EJC est liée à l'élongation de la transcription et la phosphorylation de Pol II.



Wang et al., 2014

Figure 67. Rôle de l'EJC dans l'épissage couplé à la transcription. Données issues de Wang et al., 2014. RT-PCR représentant l'épissage d'exons des gènes *TPCN1*, *MRPL3* et *HNRPDL* dans les cellules traitées avec différents inhibiteurs (DRB, FP et CPT). La quantification des expériences en trois réplicats est présentée sous forme de moyenne  $\pm$  SD. \* $P < 0,05$  ; \*\* $P < 0,01$  ; \*\*\* $P < 0,001$ , one-way ANOVA.

De la même manière, nous avons cherché à établir un lien entre la régulation de l'épissage par l'EJC dans le gène *DMD* et la transcription. Nous avons analysé l'effet d'un traitement avec la CPT sur l'épissage du gène *DMD* dans des cellules musculaires C25C148 en prolifération (myoblastes) en condition de déplétion pour eIF4A3.

## B. Matériel et méthodes

Les cellules C25C148 ont été transfectées par reverse transfection avec un si-ARN ciblant eIF4A3 ou un si-ARN contrôle (si-Ctrl) à 20nM. Après 24 h, la CPT (Sigma) diluée dans le DMSO a été rajoutée dans les puits à une concentration de 6uM ou 15uM. En parallèle, trois types de contrôles non transfectés par les si-ARN ont été réalisés : cellules non traitées, cellules

traitées uniquement par le DMSO, cellules traitées uniquement par la CPT à 6 $\mu$ M. Les cellules en prolifération ont été récupérées dans du Trizol 48 h après le début des traitements pour les analyses des profils d'épissage des ARNm par RT-PCR.

## C. Résultats expérimentaux

### 1. Profil d'épissage de *DMD* dans les cellules C25Cl48 en prolifération vs différenciation

Nous avons tout d'abord observé un niveau basal d'épissage alternatif différent dans les cellules C25cl48 en prolifération comparé aux cellules différenciées. Dans nos précédentes expériences, à l'exception des exons 71 et 78, nous n'avons pas détecté d'épissages alternatifs significatifs dans les cellules après 3 jours de différenciation contrairement aux cellules en prolifération qui montrent également un niveau d'épissage alternatif relativement élevé (15 à 20%) des exons 9 et 73-74 (Figure 68). Une deuxième différence majeure avec les cellules différenciées est l'absence d'effet notable de la déplétion d'eIF4A3 sur le niveau d'inclusion de ces exons comparés au si-Ctrl dans les conditions expérimentales utilisées, à l'exception de l'activation du site alternatif dans l'exon 9.

### 2. La CPT induit des variations sur les ASE

Les cellules traitées avec le DMSO ou transfectées avec le si-Ctrl montrent un niveau d'inclusion des exons 9, 71, 73-74 et 78 comparable aux cellules non traitées (condition endogène) validant ces contrôles (Figure 68). Le traitement CPT à 6 $\mu$ M et 15 $\mu$ M augmente l'inclusion des exons 9, 71, 73-74 dans toutes les conditions testées (endogène, si-Ctrl, si-eIF4A3), à l'exception de l'exon 78 dont le niveau d'exclusion est plus élevé de façon surprenante.

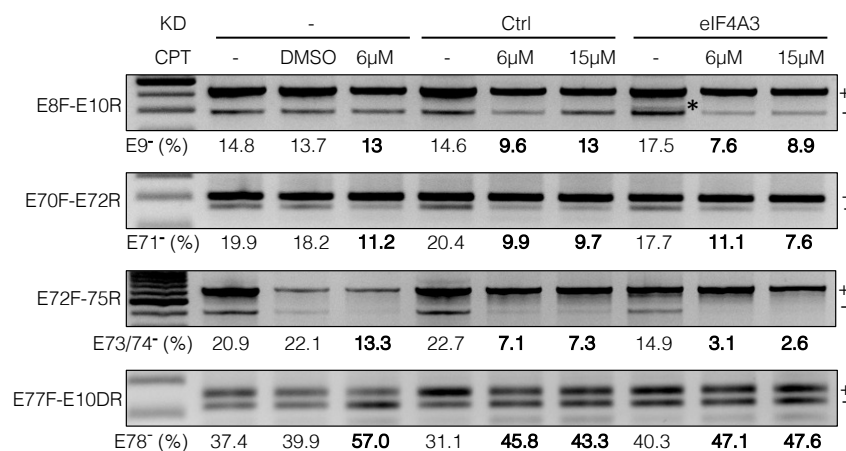


Figure 68. Gel d'agarose montrant l'effet de la CPT sur l'épissage alternatif de *DMD*. Les pourcentages de saut des exons 9, 71, 73/74 et 78 dans les différentes conditions sont indiqués sous les gels ainsi que les amorces spécifiques utilisées pour la RT-PCR (à gauche des gels). Les cellules transfectées par le siRNA Ctrl, le siRNA eIF4A3 ou non transfectées (-) ont ensuite été cultivées en présence de DMSO ou de Camptothécine (CPT) à 6 $\mu$ M ou 15 $\mu$ M. (\*) *Ati.E9*.

## D. Discussion et perspectives

On retrouve dans notre modèle de cellules musculaires un effet de la CPT sur l'inclusion des exons alternatifs de *DMD*, en accord avec le mécanisme général de régulation de l'expression génique basé sur le couplage fonctionnel de l'épissage avec la transcription (Giono and Kornblihtt, 2020).

Néanmoins, il s'agit de résultats préliminaires qui nécessitent de plus amples investigations. Différentes expériences sont à envisager :

1. Valider le profil d'épissage de l'isoforme Dp427m dans les cellules C25C148 au stade de myoblastes choisi pour nos expériences en raison d'une transcription active dans ces cellules qui prolifèrent contrairement aux myotubes qui ne se divisent plus après la fusion. Bien que la protéine dystrophine ne soit pas exprimée dans les myoblastes, les transcrits Dp427m sont déjà présents. Ces différences d'épissage peuvent s'expliquer par la cinétique d'expression des isoformes présentes dans la cellule musculaire durant la myogenèse, notamment le changement entre la Dp71 et la Dp427m (de León et al., 2005; Doorenweerd et al., 2017; Howard et al., 1999). Les analyses par RT-PCR utilisant des jeux d'amorces spécifiques des événements alternatifs à étudier ne permettent pas de déterminer dans quelle isoforme ils sont présents (la partie distale du transcrit est commune aux Dp71 et Dp427m). Néanmoins l'amplification de l'exon 9 montre que la Dp427m est présente et correctement amplifiée dans les conditions expérimentales. Les différences d'épissage peuvent également s'expliquer par une maturation différente du transcrit Dp427m dans ces précurseurs myogéniques.

2. Valider l'absence et/ou identifier les causes du manque d'efficacité de l'effet du si-eIF4A3 sur le niveau d'inclusion des exons alternatifs dans les cellules en prolifération (à l'exception du site alternatif de l'exon 9). A noter qu'un traitement de 24h seulement par ARNi a été réalisé qui peut être insuffisant comparé à 48h dans notre protocole habituel. Des validations par qPCR et Western blot de l'efficacité du traitement par le si-ARN doivent encore être réalisées. Néanmoins, l'effet de la CPT sur le niveau basal d'inclusion des exons testés a cependant pu être visualisé également dans la condition de si-eIF4A3.

3. Tester l'effet de la CPT en condition de différenciation des cellules C25C148 à 3 jours selon le protocole de culture et de traitement par les siARN bien maîtrisé dans le laboratoire.

4. Démontrer l'effet de l'inhibition de la transcription sur l'inclusion d'exons dont l'épissage est dépendant de l'EJC (dans les myoblastes et/ou les cellules différenciées à 3 jours).

5. Compte tenu de l'effet observé avec la CPT sur le niveau d'inclusion des exons alternatifs de *DMD* dans les myoblastes, valider les résultats avec des drogues dont le mécanisme d'action sur la transcription est différent (DRB, FV).

6. Nos résultats pointent à nouveau vers une régulation particulière de l'épissage de l'exon 78 (l'inhibition de la transcription augmente le saut d'exon), comme déjà observé par la déplétion des facteurs de l'EJC, des facteurs du NMD et des protéines des complexes ASAP/PSAP dans les cellules C25C148 différenciées pendant 3 jours. L'inclusion de cet exon est hautement régulée au cours du développement musculaire et dans les différents tissus d'expression des isoformes de la dystrophine. Le complexe de transcription et le dépôt de l'EJC sur les jonctions d'exons pourraient jouer un rôle décisif dans l'inclusion des exons dans la partie distale du gène, notamment de l'exon 78, le plus petit exon du gène (32 bp).

## II. Dépôt de l'EJC et régulation de l'épissage de *DMD*

### A. Introduction

L'EJC se fixe 20 à 24 nt en amont des jonctions d'exons durant l'épissage des pré-ARNm indépendamment de la séquence. Cependant, ces EJC canoniques (cEJC) ne seraient pas les seuls dépôts présents le long des ARNm. Il existerait des EJC non canoniques (ncEJC) en dehors de la position des jonctions d'exons.

Les expériences de CLIP et de RIP menées ces dernières années concernant les RBP eIF4A3 et MLN51 ont établi que les cEJC représentaient entre 80 et 95% des EJC totaux dans le transcriptome de cellules humaines (HeLa et HEK293), de la drosophile ou encore du poisson zèbre (Gangras et al., 2020; Haberman et al., 2017; Hauer et al., 2016; Mabin et al., 2018; Obrdlik et al., 2019). Ceci montre une grande conservation de cette position. Les cEJC sont composés des quatre protéines eIF4A3, Y14/MAGOH et MLN51 contrairement aux ncEJC dans lesquels MLN51 n'est pas systématiquement retrouvé, montrant le rôle particulier de MLN51 qui pourrait ne pas être présent dans tous les EJC.

La position des complexes associées à l'EJC, ASAP et PSAP, ont aussi été étudiés, notamment par des CLIP et RIP de la protéine RNPS1, introduisant la notion de RAF (RBP Associated Factors). Ces études montrent que 80% à 90% des protéines RNPS1 co-localisent avec les cEJC. Cependant, RNPS1 est aussi associée à des ncEJC sans MLN51. La protéine est détectée également en dehors des interactions avec les EJC, se fixant seule sur les ARNm parfois à des positions introniques (Hauer et al., 2016; Mabin et al., 2018; Patton et al., 2020).



La multimérisation des EJC est aussi un paramètre déterminant dans la régulation des ARNm. Un EJC associé à RNPS1 sur un exon constitutif en amont d'un exon RS liant aussi un EJC et RNPS1, permet de masquer un site cryptique d'épissage (Patton et al., 2020).

Bien que les cEJC soient déposés de façon ubiquitaire sur les jonctions d'exons dans le transcriptome, ils sont enrichis sur les gènes possédant de longs introns, ayant des exons alternatifs et RS mais aussi, dans les gènes impliqués dans le développement et la différenciation.

Nos résultats ont montré que l'EJC et RNPS1 régulent l'épissage de la partie distale du gène *DMD*, notamment l'activation du site donneur cryptique de l'exon 70, l'exon 71 et l'exon 78 dans les deux lignées testées (C25C148 et HeLa).

Nous avons cherché à établir la distribution des dépôts des EJC entre les exons 68 et 79 du gène *DMD* pour relier les résultats de nos approches fonctionnelles avec la présence de dépôts d'EJC. Nous avons travaillé à partir des données disponibles de iCLIP et de xRIPIT-seq d'eIF4A3, MAGOH, RNPS1 et MLN51. Nous avons par ailleurs réalisé des essais de RNA pull-down pour valider certaines données de CLIP.

## B. Matériel et méthodes

### 1. Analyse préliminaire des données de iCLIP et de xRIPIT-seq

Dans un premier temps, nous avons récupéré les données de iCLIP d'eIF4A3, MLN51 et RNPS1 sur la plateforme ArrayExpress : E-MTAB-4215 (Hauer et al., 2016). Ces données disponibles au format FastQ sont issues de séquençages Illumina effectués dans des cellules HeLa qui expriment l'isoforme Dp71. Nous n'avons pas trouvé de données disponibles dans des cellules musculaires qui auraient permis de faire la même analyse sur l'isoforme Dp427m. Les FastQ ont été alignés avec STAR (Dobin et al., 2013) *via* la plateforme Galaxy sur le génome de référence hg19. Les fichiers d'alignements, au format BAM, ont ensuite été visualisés sur l'outil IGV (Thorvaldsson et al., 2013). Nous avons validé notre procédure d'analyse en vérifiant en premier lieu la position d'EJC sur une série de gènes publiés par Hauer et al. (2016). Dans un second temps, nous avons récupéré les fichiers BAM (alignements effectués sur le génome de référence hg38) du xRIPIT-seq de MAGOH, MLN51 et RNPS1 provenant de cellules HEK293 qui expriment aussi la Dp71 (Mabin et al., 2018; Patton et al., 2020). Ces fichiers nous ont été fournis aimablement par G. Singh. Les fichiers d'alignements ont ensuite été visualisés sur l'outil IGV. Grâce à ces deux jeux de données, nous avons pu analyser la position des EJC sur les exons 68 à 79 du gène *DMD*.

## 2. Expériences de RNA pull-down

Les oligonucléotides sens et antisens hybridés (MWG) correspondant aux différentes sondes ARN de l'exon 70 et 78 (Figure 69) ont été clonés dans le vecteur pGEM<sup>®</sup>-3Z, digéré par EcoRI-PstI (Promega). Les matrices linéarisées par HindIII ont ensuite été transcrites *in vitro* à l'aide du kit de transcription *in vitro* RiboMAX<sup>™</sup>-T7 (Promega). Les sondes ARN (700 pmol) ont été oxydées avec du m-périodate de sodium 0.1M dans de l'acétate de sodium 1M, pH 5.0 (SigmaAldrich) sur un rotateur dans l'obscurité pendant 1 h. Les sondes d'ARN ont été purifiées avec le RNeasy Plus Mini Kit (Qiagen) en utilisant le protocole supplémentaire Qiagen 'Purification of total RNA containing miRNA' pour retenir les petits ARN. Les sondes d'ARN oxydées ont été incubées avec 1 mL de billes d'agarose de dihydrazide d'acide adipique (Sigma-Aldrich). Ensuite, 150 pmol d'ARN immobilisé a été incubé avec 62µL d'extraits nucléaires de cellules HeLa (Promega) dans 44,3µL de tampon D, complété par 3,2 mM MgCl<sub>2</sub>, 67µg d'ARNt d'*Escherichia Coli* (Sigma-Aldrich) et 0,08 µg/µL d'albumine de sérum bovin (NEB), pendant 30 min à TA. Les billes ont ensuite été culottées par centrifugation à 100g pendant 2,5 min et lavées cinq fois avec du tampon D contenant 1,5 mM de MgCl<sub>2</sub> avant l'ajout de tampon Laemmli et le chargement sur un gel de SDS-PAGE à 10%. Les protéines extraites ont été analysées par Western blot en utilisant des anticorps spécifiques d'eIF4A3 et RNPS1 (voir les références dans la partie Résultats).

SD_E70_eIF4A3_cEJC
GAATGGGCTACCTGCCAGTGCAGACTGTCTTAGAGGGGGACAACATG GAAAC
SD_E70_eIF4A3_ncEJC
ACTACATCAGGAGAAGATGTTTCGAGACTTTGCCAAGGTAATAAAAAAC AAA
SD_RNPS1_IVS77
CAGCAGAGAATAAATGGTCATCTCAACATTTGTTTGGTAAAAGAAGCA AATTGGTATGAAATTGTATTTATGCATGT

Figure 69. Séquences des sondes des exons 70 et 78 utilisées dans les expériences de RNA pull-down. SD\_E70\_eIF4A3\_cEJC et SD\_E70\_eIF4A3\_ncEJC correspondent respectivement aux sites canonique et non canonique de l'EJC prédits sur l'exon 70 de DMD. SD\_RNPS1\_IVS77 correspond à un site RNPS1 prédit dans l'intron 77 de DMD.

## C. Résultats

### 1. Visualisation des cEJC sur les exons 68 à 79 du gène *DMD*

Nous avons retrouvé un dépôt canonique de l'EJC pour les exons 68, 69, 70, 72, 73, 74, 75, 76 et 77) comme le montre la présence simultanée d'eIF4A3, MAGOH, MLN51 et RNPS1 à ces positions (Figure 70). Il convient de noter que les données disponibles ne fournissent pas de résultats fiables pour les deux exons courts 71 et 78 en raison de leur faible couverture, ces exons subissant un fort épissage alternatif dans les lignées cellulaires utilisées.

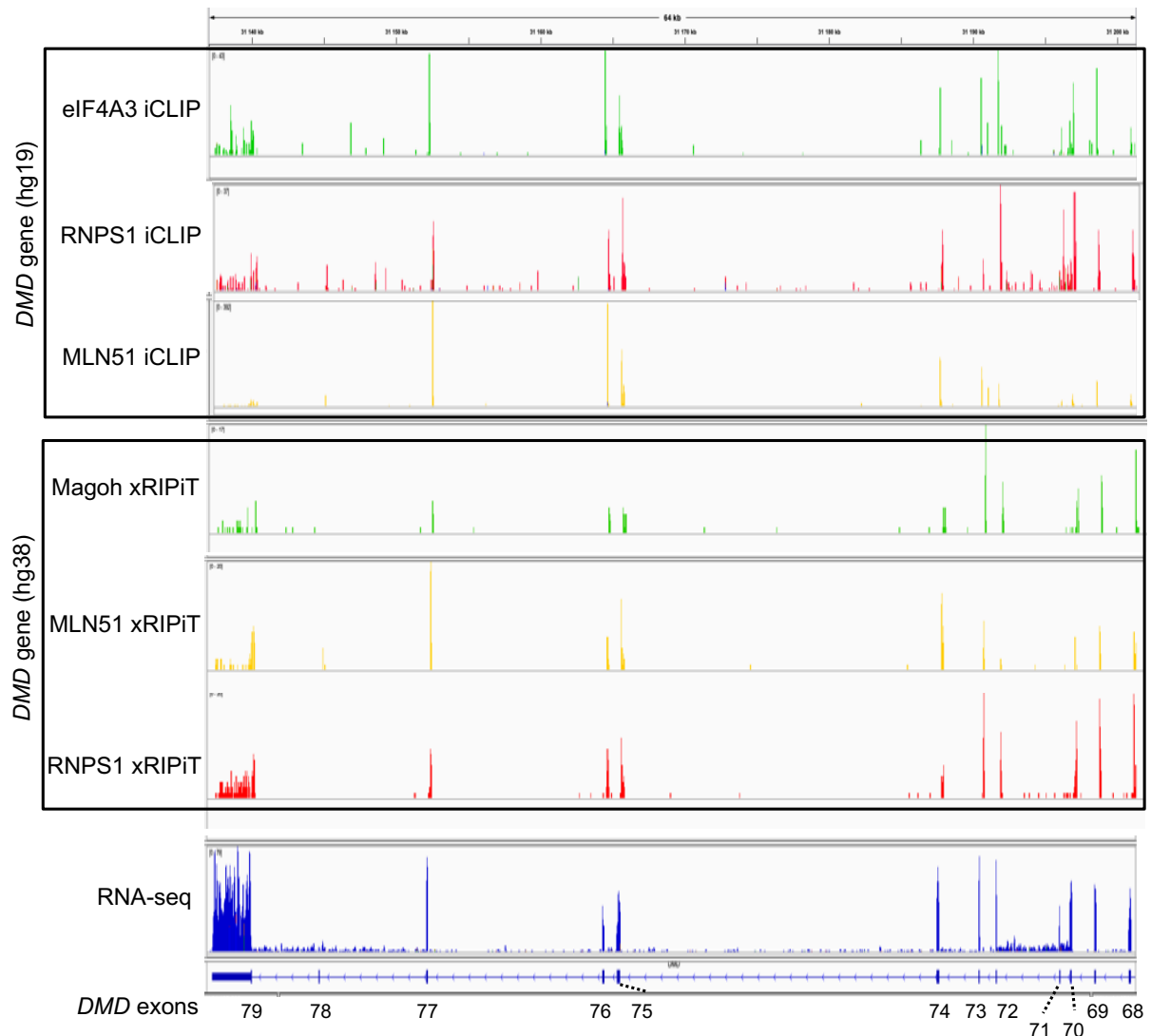


Figure 70. Distribution des dépôts EJC et de RNPS1 sur la partie distale du gène DMD (exons 68 à 79). Données issues de Hauer et al., 2016 et Mabin et al., 2018. Les fichiers BAM ont été alignés sur les génomes de référence hg19 et hg38. La visualisation des alignements des exons 68 à 79 de DMD a été réalisée à l'aide du logiciel IGV.

## 2. Hypothèse de dépôts de ncEJC dans le transcrit DMD

De façon intéressante, nous avons détecté la présence potentielle de dépôts non canoniques correspondant à une colocalisation d'eIF4A3 ou de MAGOH avec RNPS1 mais sans MLN51 :

- au début de l'exon 70 en amont du site donneur cryptique (Alt.E70) (Figure 71A)
- au début des exons 74 et 75 (données non montrées).
- dans l'intron 77 : un site de liaison (> 10 reads) pour RNPS1 (et plus faiblement pour MLN51) a été détecté en amont de l'exon 78 au niveau d'une séquence conservée (visualisée sur UCSC Genome Browser). De manière intéressante, RNPS1 régule fortement l'épissage de l'exon 78 (Figure 71B). Néanmoins, il faut noter que cette liaison de RNPS1 dans l'intron 77 n'a pas été mise en évidence à partir des données de xRIPiT.

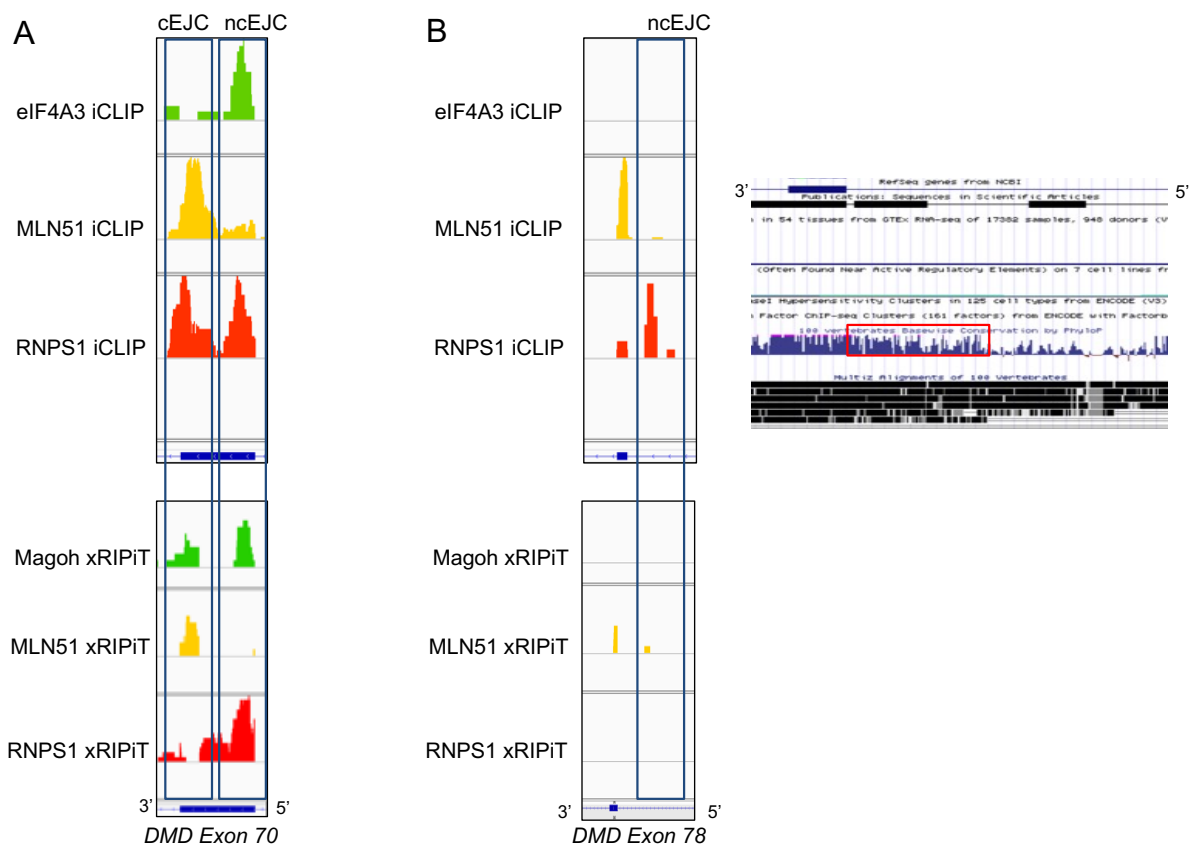


Figure 71. Analyses des données de iCLIP et xRIPiT sur l'exon 70 et l'intron 77 de DMD.

(A) Dépôts canoniques (cEJC) et non canoniques (ncEJC) des EJC et de RNPS1 sur l'exon 70 et (B) non canonique dans l'intron 77 en amont de l'exon 78 associé à la visualisation dans UCSC (à droite) de la conservation de cette séquence. Données issues de Hauer et al., 2016 et Mabin et al., 2018.

### 3. Validation du dépôt EJC en position non canonique sur l'exon 70

Nous avons effectué des expériences de RNA pull-down d'eIF4A3 afin de tenter de confirmer le dépôt du ncEJC en début de l'exon 70 du gène *DMD*. Nous avons défini des sondes dans 2 régions d'intérêt de l'exon 70 : (1) une sonde au début de l'exon 70, en amont du site donneur cryptique Alt.E70 dont l'activation est associée à de nombreux événements d'épissage alternatifs en condition de déplétion d'eIF4A3. Ce ncEJC potentiel masquerait le site Alt.E70 afin de permettre la bonne utilisation du site donneur d'épissage canonique pour l'inclusion de l'exon 70 dans le transcrit mature et/ou inhiber des événements de ré-épissage au niveau de ce site ; (2) une sonde dans la région de dépôt canonique de l'EJC couvrant la position -24nt comme contrôle positif (Figure 69). En parallèle, le dépôt de l'extrait nucléaire de cellules HeLa et de bille sans ARN dans le Western blot final sont utilisés respectivement comme contrôle positif et négatif de la détection de la protéine eIF4A3 dans les conditions expérimentales.

Les résultats n'ont pas été concluants. Si un léger signal (un peu supérieur à la condition sans ARN) correspondant à eIF4A3 a pu être visualisé avec la sonde ncEJC (SD ncEJC), l'absence

de signal pour la sonde contenant le site canonique (SD cEJC) ne permet pas d'interpréter correctement ce résultat (Figure 72).

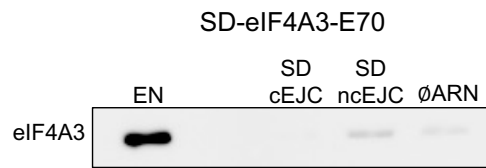


Figure 72. Western blot révélant la fixation d'eIF4A3 sur les sondes de l'exon 70 analysées en RNA pull-down. A gauche, le contrôle positif montrant le niveau d'expression d'eIF4A3 dans les extraits nucléaires (EN) de cellules HeLa utilisés pour le RNA pull-down. A droite, l'hybridation d'eIF4A3 sur les sondes de l'exon 70 de DMD. Témoin sans ARN ( $\emptyset$  ARN).

#### 4. Validations du dépôt RNPS1 dans l'intron 77

Nous avons utilisé deux approches (minigènes rapporteurs et RNA pull down) pour mettre en évidence le rôle de RNPS1 sur l'inclusion de l'exon 78 *via* une liaison dans l'intron 77.

##### a. Analyse de mutants délétionnels d'un minigène rapporteur d'épissage (pSPL3-E78).

Nous avons à notre disposition dans le laboratoire des versions du minigène pSPL3-E78 qui diffèrent par la taille des régions introniques clonées de part et d'autre de l'exon 78. La version L/L (Large/Large) inclut la plus grande région intronique, puis des combinaisons de délétions ont été réalisées aux sites M (medium) ou S (small) (Figure 73A). Nous avons transfecté les différents minigènes en condition si-Ctrl et si-RNPS1 dans les C25C148. Les profils d'épissage du minigène de l'exon 78 ont été analysés par RT-PCR en utilisant des amorces spécifiques du minigène (SD6F-5'-TCTGAGTCACCTGGACAACC-3' et SA2R-5'ATCTCAGTGGTATTTGTGAGC-3'). La migration des produits de RT-PCR dans un gel d'agarose montre une augmentation du saut de l'exon 78 en condition de déplétion de RNPS1 pour les minigènes (L/L, L/M, M/L et M/M) contenant la région de fixation potentielle de RNPS1, comme observé pour l'exon 78 endogène dans ces cellules (Figure 73A-B).

##### b. Analyse des essais de pull-down de RNPS1 dans l'intron 77

Une expérience de RNA pull-down a été réalisée avec une sonde contenant le site de fixation potentiel de RNPS1 de l'intron 77 (Figure 69). Les protéines liées par la sonde ont été analysées en western blot avec un anticorps contre RNPS1 mais aussi eIF4A3 servant de contrôle négatif, car aucun signal n'avait été observé dans les données iCLIP et xRIPiT. Le western blot a montré la présence d'une bande correspondant à RNPS1 assez significative, si on compare au niveau de la protéine dans les extraits nucléaires (EN). Un faible signal pour eIF4A3 a également été détecté (Figure 73C). Ces résultats suggèrent que la régulation de l'épissage de l'exon 78 par

RNPS1 pourrait passer par le recrutement de la protéine en amont de l'exon, et potentiellement en association avec l'EJC.

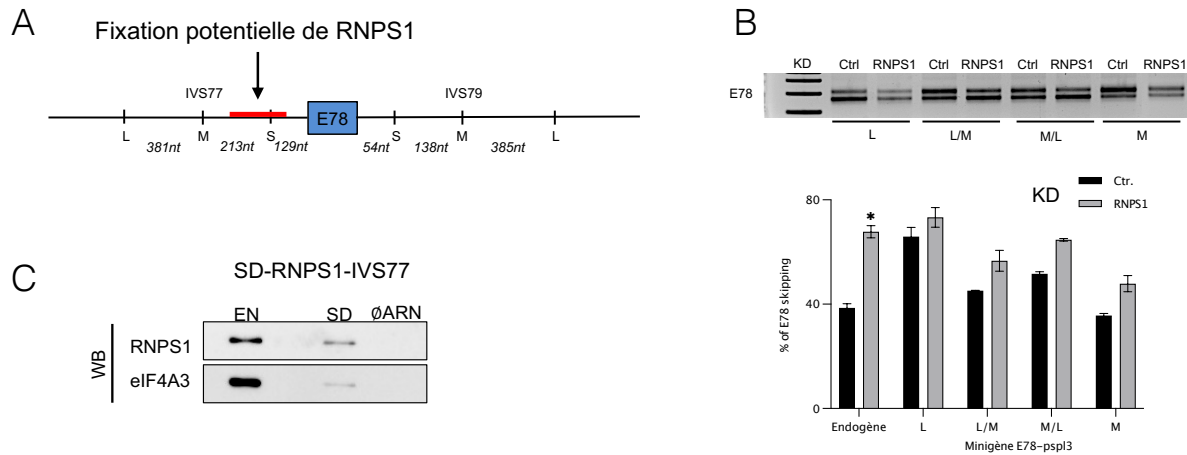


Figure 73. Étude du site de fixation de RNPS1.

(A) Représentation schématique du site de fixation potentiel de RNPS1 dans l'IVS77 (en rouge). (B) Analyse des profils d'épissage en gel d'agarose des mutants délétionnels des minigènes E78-psp13 en conditions si-Ctrl et si-RNPS1. (C) Analyse par Western blot de la fixation de RNPS1 et eIF4A3 sur la sonde IVS77 analysée en RNA pull-down. Témoin sans ARN ( $\emptyset$  ARN). EN, extraits nucléaires de cellules HeLa utilisés pour le RNA pull-down.

## D. Discussion et perspective

Nos analyses bio-informatiques montrent que le gène *DMD* possède majoritairement des EJC canoniques dans les régions analysées, ce qui est en accord avec les données de la littérature qui rapportent une fréquence de 80% à 90% de cEJC dans le transcriptome humain (Hauer et al., 2016; Mabin et al., 2018; Patton et al., 2020).

Les constructions minigènes de l'exon 78 ont pu valider le rôle de RNPS1 dans la régulation de l'épissage de l'exon 78 du gène *DMD*, comme cela avait été montré en condition endogène (voir partie Résultats). Les signaux obtenus *in vitro* dans les essais de RNA pull-down pour RNPS1 et eIF4A3 ne nous permettent pas d'affirmer la présence d'un site de liaison pour ces protéines dans l'intron 77. Il est à noter cependant que le signal pour RNPS1 était plus marqué et que les complexes ASAP/PSAP peuvent réguler des événements d'épissage indépendants de l'EJC (Boehm et al., 2018; Rodor et al., 2016; Wang et al., 2018). La démonstration de ce rôle fonctionnel de RNPS1 nécessiterait de poursuivre les investigations. La liaison de RNPS1 sur la séquence de l'exon 77 pourrait être confirmée par des expériences d'immunoprécipitation de l'ARN (RIP) dans les cellules C25Cl48 à l'aide d'anticorps spécifiques de RNPS1 mais également d'eIF4A3 ou MLN51 ainsi que l'analyse par RT-PCR de la région candidate à l'aide d'amorces spécifiques.

La caractérisation des sites de dépôts des EJC reste complexe et l'existence des dépôts non canoniques est sujet à discussion. L'estimation de la proportion des nEJC ne cesse de diminuer

avec l'évolution des techniques d'étude des interactions ARN-protéines à large échelle et l'amélioration des performances des analyses bio-informatiques.

Les approches ciblées réalisées *in vitro* comme le RNA pull down basé peuvent ne pas être les plus adaptées pour montrer la présence d'un complexe qui se forme de manière dynamique au cours de la réaction d'épissage. L'hypothèse d'un site ncEJC masquant le site alternatif de l'exon 70 est néanmoins particulièrement attirante compte tenu de la forte dérégulation de ce site dans les conditions de KD de eIF4A3 dans les cellules C25C148. La démonstration de son existence reste problématique.

### III. Résultats complémentaires d'UPF1 par l'étude d'UPF2

#### A. Introduction

Précédemment, nous avons vu que la déplétion d'UPF1 induisait des variations des niveaux d'inclusion des exons 71 et 78 ainsi qu'une accélération de la différenciation musculaire.

UPF1 n'étant pas connu pour agir directement dans l'épissage des gènes, nous supposons que les changements des niveaux d'inclusion des exons 71 et 78 pourraient être dû : i) au mécanisme d'AS-NMD où dans ce cas, il pourrait exister des isoformes de *DMD* incluant des PTC dans d'autres régions du transcrite associées à ces sauts d'exons qui eux maintiennent un cadre de lecture ouvert ; ii) à un changement de rapport de présence des isoformes Dp427m et Dp71 dans la cellule, qui impacterait les niveaux des exons 71 et 78 à cause de l'épissage alternatif différentiel entre ces deux isoformes. En effet, le niveau de l'expression de la Dp427m diminue fortement en condition de siUPF1 pour atteindre un niveau comparable à la Dp71 (voir partie Résultats, Figure Supplémentaire 5A-B). Ceci pourrait impacter les niveaux d'inclusion relatifs des exons 71 et 78.

Il n'est pas clair à ce stade si UPF1 est directement responsable du changement du niveau d'inclusion des exons 71 et 78 ou si le système NMD ou encore le processus de différenciation pourraient être impliqués.

Pour cela, nous avons étudié le rôle de UPF2, le deuxième facteur de la voie classique du NMD, qui assure le lien entre l'EJC et UPF1 lors de l'activation du NMD. Des transcrits aberrants et des défauts développementaux sont observés en l'absence d'UPF2 dans de nombreux tissus (Bao et al., 2016; Weischenfeldt et al., 2012).

Cependant, certains ARNm cibles du NMD sont insensibles à la déplétion d'UPF2 dans le modèle HeLa (Gehring et al., 2005). Ceci confirme que différentes voies du NMD existent.

Notamment, le facteur de l'EJC, MLN51 qui recrute UPF3B de manière indépendante d'UPF2 (Gehring et al., 2009a).

Dans le tissu musculaire, UPF2 formant le système NMD avec UPF1 est antagoniste du système SMD, formé de Stauf1 et UPF1. Le NMD classique est décrit comme inhibant les gènes de différenciation afin de maintenir l'état de prolifération des myoblastes (Gong et al., 2009). Au contraire la voie du SMD, ainsi que le NMD alternatif passant par UPF3B permettent la mise en place des myotubes.

Dans ce contexte, il était intéressant d'étudier plus spécifiquement la contribution d'UPF2 dans la régulation de l'épissage de Dp427m ainsi que dans la différenciation musculaire et de comparer les résultats avec ceux obtenus pour UPF1 et l'EJC. Pour cela, nous avons déplété la protéine UPF2 à l'aide d'un si-ARN dans les cellules musculaires humaines C25C148 et étudié le profil d'épissage des exons 71 et 78 ainsi que la différenciation musculaire à 3 jours. Les résultats ont été comparés aux conditions de KD pour eIF4A3 et UPF1.

## B. Matériel et méthodes

Le protocole de transfection du si-ARN UPF2 (MWG) (5'-GGCUUUUGUCCCAGCCAUC-3') à 20nM dans les C25C148, les extractions d'ARN sur colonnes à partir du Trizol, les analyses des profils d'épissage des exons 71 et 78 par SQF-PCR fluorescente, les PCR quantitatives de l'expression d'UPF2, de la dystrophine, des facteurs de la myogenèse ainsi que les Western Blot ont été réalisées dans les mêmes conditions que celles décrites dans la section Matériel et méthodes de la partie Résultats.

## C. Résultats

### 1. UPF2 possède un rôle différent d'UPF1

A partir de quatre transfections indépendantes, les résultats montrent que comme UPF1, la déplétion d'UPF2 modifie l'inclusion des exons 71 et 78 du gène *DMD* avec cependant des effets différents (Figure 74A). Alors que l'absence d'UPF1 réinclut l'exon 71 d'environ 20%, le si-UPF2 augmente le saut de cet exon d'environ 10% comparé au si-Ctrl. Concernant le saut de l'exon 78, la déplétion d'UPF2 accentue l'effet observé en si-UPF1, menant à une forte exclusion de cet exon (~ 60%) comparé au si-Ctrl (~ 30%).

Nous nous sommes ensuite intéressés aux conséquences de la diminution d'UPF2 sur l'expression relative des isoformes Dp427m et Dp71 dans les C25C148. Comme pour la déplétion d'UPF1, le si-UPF2 mène à une diminution drastique des transcrits Dp427m comparé



au contrôle (Figure 74B). En revanche, nous avons observé une augmentation significative des transcrits Dp71 en condition de déplétion d'UPF2 comparé à la condition contrôle. De plus, une proportion plus élevée de transcrits Dp71 comparé aux transcrits Dp427m est observée avec le si-UPF2 (Figure 74B). Ces résultats suggèrent qu'UPF2 pourrait jouer un rôle différent d'UPF1 dans la différenciation musculaire. L'absence d'UPF2 est associée à un profil non différencié (augmentation de Dp71).

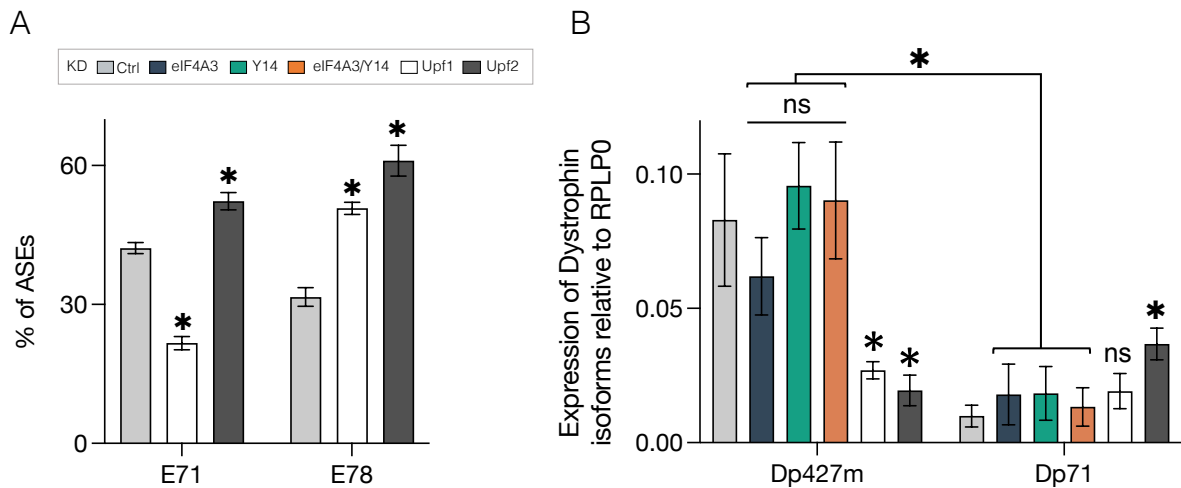


Figure 74. Effet du si-UPF2 sur l'épissage et l'expression de DMD.

(A) Quantification des sauts des exons 71 et 78 de DMD par SQF-PCR (B) Quantification de l'expression de la Dp427m et la Dp71 par qPCR.

## 2. UPF2 bloque la différenciation des C25cl48

A l'inverse du si-UPF1 qui accélère le processus de différenciation musculaire (voir partie Résultats), la déplétion d'UPF2 (n=4) a montré un défaut de la mise en place de la myogenèse. En effet, nous avons observé par RT-qPCR (n=2) une augmentation de l'expression de la cyclineA comme dans les myoblastes ainsi qu'une augmentation de Myf5 comparée au contrôle (figure 75A). Les facteurs de différenciation (MyHC, Troponine T, MyoG, ACTA1 et CKM) étaient absents en condition de déplétion d'UPF2 en qPCR ainsi qu'en western blot (Figure 75A-B). De plus, nous avons observé une forte diminution du niveau de la Dp427m par western blot, qui confirme la diminution du niveau de transcrits Dp427m observée précédemment. Le niveau d'expression semble très faible, atteignant un niveau retrouvé dans les myoblastes indifférenciés (Figure 75B).

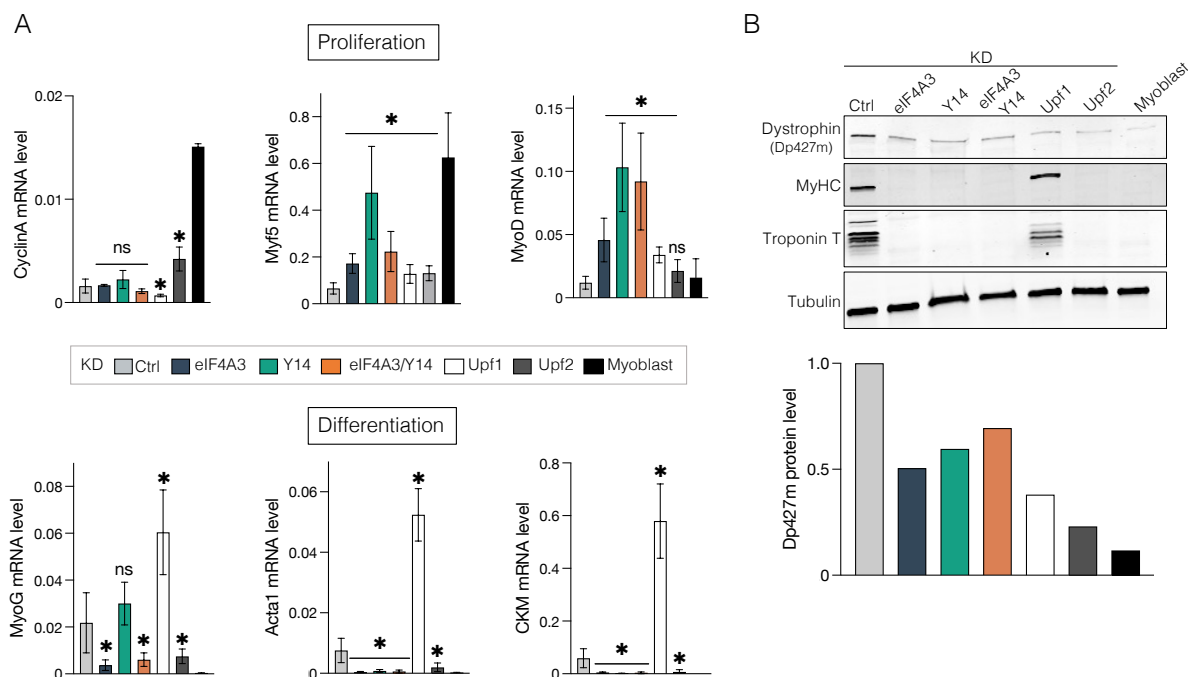


Figure 75. Quantification de l'expression des facteurs de la myogenèse et de la Dp427m.

(A) Comparaison des facteurs de différenciation par (A) qPCR, et (B) Western Blot, entre les différentes conditions de siRNA présentées dans la partie Résultats et le si-UPF2.

## D. Discussion et perspectives

Tout comme la diminution d'UPF1, la déplétion d'UPF2 engendre des variations des niveaux d'inclusion des exons 71 et 78 dans l'ARNm de la Dp427m. Cependant la régulation semble différente, avec même un effet inverse concernant l'exon 71. Les mécanismes impliqués dans cette régulation par les facteurs du NMD restent inconnus puisque le saut de ces exons n'entraîne pas la formation de codon STOP prématuré dans les ARNm.

Contrairement à l'absence d'UPF1 qui accélère la myogenèse, la déplétion d'UPF2 empêche le processus de différenciation musculaire. Cependant, ces résultats semblent être en contradiction avec la littérature. En effet, la différenciation musculaire est gérée par une balance entre les systèmes NMD et SMD (Gong et al., 2009). Dans cette étude, il est montré que la déplétion d'UPF2 inactive la voie du NMD classique par la perte d'interaction avec UPF1, favorisant la liaison d'UPF1 avec Staufen-1 dans le système SMD afin d'inhiber les gènes de prolifération et ainsi engager les myoblastes des cellules C2C12 de souris dans la différenciation. Au contraire, le système NMD inhibe les gènes de différenciation, notamment la myogénine.

Nous avons vu que les déplétions des facteurs de l'EJC (eIF4A3 et Y14) entraînent un blocage de la myogenèse (voir partie Résultats, Figure 7). En effet, le niveau élevé de Myf5 et de MyoD en condition de si-eIF4A3 suggère que l'absence de la protéine bloque les myoblastes en G1 du cycle cellulaire lors de l'engagement dans la différenciation musculaire. Le si-Y14

empêcherait la fusion des myotubes, les cellules restant au stade de myocytes mononucléés (niveau de MyoG similaire au si-Ctrl). Les marqueurs de différenciation tardifs (ACTA1, CKM et MyHC) sont absents lors des déplétions de eIF4A3 et Y14. Il est intéressant de noter que la déplétion d'UPF2 semble bloquer le processus de myogenèse plus en amont. En effet, la surexpression de la cyclinA, la surexpression de Myf5 et un niveau moins élevé de MyoD montrent que les cellules sont encore en stade prolifératif et ne sont pas engagées dans la différenciation musculaire (Figure 75A).

Nos données soulèvent de nombreuses questions, et les mécanismes par lesquels passe la régulation de la différenciation musculaire médiés par les facteurs du NMD ou de l'EJC ne sont pas élucidés. Il serait intéressant d'approfondir l'étude des balances entre le NMD classique et alternatif (UPF1, UPF2 et UPF3B) et le SMD (UPF1, Staufin-1) afin de mieux comprendre ces mécanismes de régulation de la myogenèse dans le muscle humain.

# Discussion générale

Ce projet de thèse décrit pour la première fois le rôle de l'EJC et de ses protéines associées dans la régulation de l'épissage du gène *DMD* mais aussi dans la différenciation musculaire.

### **Importance de l'étude du gène *DMD***

Depuis sa découverte en 1987, le gène *DMD* est l'objet d'une recherche active en raison de son importance en pathologie humaine. Ses mutations sont responsables de la plus fréquente des dystrophies musculaires chez l'enfant, la dystrophie musculaire de Duchenne (DMD). La taille exceptionnelle de ce gène, la complexité de sa structure et de la régulation de son expression en font un modèle d'étude particulièrement difficile. Si la pathologie moléculaire à l'origine des dystrophinopathies est relativement bien connue aujourd'hui (Tuffery-Giraud et al., 2017), les mécanismes physiopathologiques, les conséquences de l'absence de dystrophine dans la cellule musculaire et les organes dans lesquels s'exprime ses différentes isoformes protéiques, comme le cerveau ou la rétine, restent incomplètement compris. De nombreuses stratégies thérapeutiques sont développées mais ont donné pour le moment des résultats limités. Une de ces approches est basée sur la redirection de l'épissage chez les patients DMD porteurs de grandes délétions génomiques. L'objectif est de restaurer une phase ouverte du cadre de lecture dans le transcrit en induisant le saut d'un exon en amont ou en aval de la délétion à l'aide d'oligonucléotides antisens (stratégie du saut d'exon). L'épissage alternatif d'exon dans la région de la mutation survient parfois de façon spontanée chez les patients et contribue à la modulation de la sévérité du phénotype et peut expliquer également la présence des fibres révertantes observées au niveau du muscle.

L'équipe dans laquelle j'ai réalisé mes travaux de thèse s'est particulièrement intéressée à l'étude des mécanismes moléculaires liés aux défauts d'épissage dans le gène *DMD* comme cause de la pathologie ou en tant que modificateur du phénotype dans le cas des sauts d'exon associés aux mutations non-sens situées dans la région centrale du gène (Miro et al., 2015).

Si l'épissage du gène *DMD* est très étudié dans le contexte pathologique ou à des fins thérapeutiques, les connaissances concernant la régulation de l'épissage physiologique des transcrits *DMD* dans les tissus ou au cours du développement restent très limitées.

L'épissage alternatif est très présent dans le cerveau et le muscle où s'exprime le gène *DMD* (Baralle and Giudice, 2017). L'étude bibliographique dans ce manuscrit montre la très grande diversité des épissages alternatifs identifiés dans les différentes isoformes de dystrophine. Ils prédominent néanmoins dans les isoformes cérébrales. L'analyse des transcrits musculaires issus de patients atteints de DMD avaient laissé supposer une fréquence élevée de ces

événements dans le muscle squelettique. Il s'agissait cependant souvent de transcrits minoritaires pouvant correspondre à des formes intermédiaires d'épissage détectées grâce à la grande sensibilité des techniques de PCR utilisées. Lorsqu'ils sont en phase, ces transcrits peuvent être amplifiés préférentiellement par rapport au transcrit muté dégradé par le mécanisme du NMD ou diminués par une baisse de la transcription (García-Rodríguez et al., 2020).

Notre équipe a depuis démontré que tous les exons de l'isoforme musculaire du gène *DMD* (Dp427m) étaient épissés de manière constitutifs et inclus dans le transcrit mature du muscle squelettique adulte normal (Bougé et al., 2017). Ces résultats confirment ce qui avait été observé dans une étude précédente, montrant des épissages alternatifs dans l'isoforme cérébrale Dp427c mais qui n'étaient pas présents dans le muscle (Nishida et al., 2015a).

Malgré la description exhaustive de ces profils d'épissage différents dans les isoformes de la dystrophine, aucune étude n'a été réalisée afin d'identifier les mécanismes de régulation mis en jeu, à l'exception de la description du rôle de MBNL1 dans l'inclusion de l'exon 78 (Rau et al., 2015).

Cette constatation a amené mon équipe à entreprendre l'identification de facteurs *trans* impliqués dans la reconnaissance et l'inclusion des 79 exons dans le transcrit mature Dp427m exprimé dans le muscle squelettique. Un premier crible fonctionnel avait été réalisé dans la lignée de cellules musculaires humaines C25C148 qui a permis de démontrer le rôle de plusieurs RBP dans l'épissage des exons 9, 37, 38, 41, 71, 74, 78 et du pseudo exon PE1a (Miro et al., 2020).

### **Rôle de l'EJC dans l'épissage du transcrit Dp427m**

Mes travaux de thèse contribuent à la compréhension des mécanismes de régulation de l'épissage du gène *DMD* par l'étude du complexe EJC et des protéines associées. L'EJC déposé aux jonctions exon-exon lors de la maturation des pré-ARN sert de marque reliant l'épissage à la régulation post-transcriptionnelle de l'ARNm (Gerbracht and Gehring, 2018). Les mécanismes de régulation de l'épissage ont dû s'adapter chez les eucaryotes supérieurs pour prendre en charge des pré-ARNm plus complexes en termes de nombre d'exons et de longueurs d'introns. L'EJC a évolué afin de fonctionner comme un "verrou" au niveau des jonctions exon-exon, empêchant le réassemblage du spliceosome aux jonctions déjà épissées et permettre une plus grande précision de l'épissage (Blazquez et al., 2018; Boehm et al., 2018). La fonction la plus conservée de l'EJC semble d'assurer l'irréversibilité de l'épissage (Obrdlik et al., 2019).

Cette hypothèse est cohérente avec le fait que la liaison d'eIF4A3 n'est pas requise pour la réaction d'épissage elle-même. Par conséquent, on peut s'attendre à ce que des niveaux faibles d'eIF4A3 (comme lors de KD de eIF4A3) ne vont pas interférer avec l'épissage mais entraîner une diminution de l'assemblage des EJC sur les ARN épissés.

Cette notion de contrôle de la fidélité peut s'avérer importante pour l'épissage du gène *DMD*, le plus grand gène décrit chez l'homme dont la transcription excède 10 heures. L'EJC est déposé de manière co-transcriptionnelle et immédiatement après l'épissage, ce qui lui permet de bloquer rapidement les événements d'épissage aberrants. Ce mécanisme ne se limite pas aux transcrits épissés de manière co-transcriptionnelle dans un ordre séquentiel 5'→3. L'EJC protège également les transcrits dans lesquels les introns sont retirés de manière non-séquentielle, en masquant les signaux d'épissage et empêcher l'épissage d'un site 3' reconstitué (Leung et al., 2019).

Nous avons évalué le rôle des protéines centrales de l'EJC dans l'épissage de *DMD* en déplétant l'expression d'eIF4A3 qui est une RBP hélicase et de sa protéine associée Y14 définie comme une RAF, dans le modèle cellulaire C25C148. Nous avons ensuite analysé le profil d'épissage du transcrit Dp427m par la technique de séquençage ciblé mise au point dans l'équipe (Bougé et al., 2017; Miro et al., 2020). Au moment du choix de l'approche expérimentale, l'équipe a considéré qu'il existait un biais important d'analyse en faveur des gènes fortement exprimés par les techniques de RNA-Seq total disponibles. L'étude approfondie de l'expression des gènes faiblement exprimés comme le gène *DMD*, et en particulier de l'épissage qui demande une profondeur de lecture supérieure pour détecter des événements présents en quantité variable, demandait de mettre en place une analyse ciblée. Celle-ci s'est révélée performante pour détecter les ASE induits par les déplétions de eIF4A3 et Y14 de type saut d'exon, activation de sites cryptiques alternatifs ou rétention d'introns de petite taille. Il faut cependant noter que certains événements peuvent ne pas avoir été identifiés en raison des limitations de la technique telles que (i) les biais possibles d'amplification lors de l'étape de PCR, (ii) l'impossibilité de détecter des transcrits avec des rétentions d'introns dont la taille dépasse les limites techniques de la Long Range polymérase utilisée, (iii) l'impossibilité de détecter des transcrits utilisant des sites polyA alternatifs présents dans le gène (Nishida et al., 2015b; Tinsley et al., 1993) en raison du choix d'amorces spécifiques situées dans les exons 1 et 79 de l'isoforme musculaire.

En dehors de l'exon 9, la déplétion des facteurs du cœur de l'EJC régule principalement le niveau d'inclusion des exons de la partie distale du gène (exons 68 à 78). Différents types

d'évènements ont été caractérisés incluant des exons cassettes (exons 9, 68, 69, 78) ou multiples (exons 69-70, 69-71, 71-74, 72-74, 73-74, 71/73-74), la rétention de l'intron 70 (698 pb) et l'utilisation de sites d'épissage alternatifs donneurs dans l'exon 9 et 70 qui pourraient correspondre aux micro exons produits par épissage récursif décrits par Blazquez et al. (2018). L'analyse préliminaire des séquences aux jonctions d'exons ne nous a pas permis d'identifier de motifs qui permettraient de démontrer que certains des sauts d'exon sont dus à des événements de ré-épissage plutôt qu'à un épissage alternatif de cassette exons. Néanmoins, il a été montré que le ré-épissage pouvait avoir lieu au niveau de sites faibles d'épissage, peu conformes à la séquence consensus.

Le profil d'épissage associant ces différents évènements est particulièrement complexe à analyser. L'analyse des produits de RT-PCR par la technique de PCR semi-quantitative fluorescente sur séquenceur capillaire permettant de séparer des fragments jusqu'à 600 pb avec une résolution d'1 pb, a permis d'établir plus précisément la diversité des ASE présents. Cependant, nous n'avons pas utilisé d'approche de reconstruction de transcrits lors de notre analyse bio-informatique permettant de réellement définir la combinaison des évènements présents au sein de la même molécule de transcrits. L'évolution vers des techniques de séquençage long-read de molécule unique serait d'un apport considérable pour la caractérisation de l'association des différents variants d'épissage dans ces transcrits longs.

### **L'EJC, régulateur de l'épissage de plusieurs isoformes du gène *DMD***

Plusieurs de ces épissages aberrants qui maintiennent une phase ouverte de lecture, correspondent à des épissages alternatifs présents dans les autres isoformes de *DMD*, notamment dans le SNC et la rétine (Aragón et al., 2018; Mahyoob Rani et al., 2019; Nishida et al., 2015a; Paúl-González et al., 2021). Ceci suggère que l'EJC, seul ou en coopération avec d'autres RBP comme les SR protéines au sein des mRNP (Mabin et al., 2018) pourrait réguler différemment ces exons dans les isoformes exprimées dans les différents tissus (Naidoo and Anthony, 2020). Il est intéressant de noter que la détection récente du transcrit de la Dp71 dans le muscle squelettique (Kawaguchi et al., 2018) a montré que tous les exons étaient inclus comme pour la Dp427m alors que cette isoforme Dp71 (exons 63 à 79) est connue pour subir un fort épissage alternatif dans le SNC. Il serait intéressant d'étudier la régulation par l'EJC de l'épissage des différentes isoformes du gène en utilisant d'autres modèles cellulaires spécifiques (cellules neuronales, rétinienne, ...).

Nous avons pu vérifier que la détection de ces différents ASE, principalement ceux qui introduisent un PTC, n'était pas corrélée à la baisse d'efficacité du NMD induite par la



déplétion des facteurs de l'EJC. En effet, la déplétion des facteurs majeurs du NMD, UPF1 et UPF2, ne reproduit pas ces événements. Il est important de noter cependant que la diminution d'expression d'UPF1 et UPF2 modifie le niveau d'inclusion des deux exons les plus courts de DMD, l'exon 71 (39 pb) et l'exon 78 (32 pb), épissés alternativement dans les isoformes cérébrales (Aragón et al., 2018). Cette observation était plutôt inattendue, ces transcrits alternatifs n'étant pas des cibles du NMD. En effet, le saut de l'exon 71 est en phase tandis que le saut de l'exon 78 décale le codon stop terminal dans le 3'UTR du gène, mais n'introduit pas de PTC. Plusieurs hypothèses peuvent être avancées pour expliquer nos résultats : (i) Bien que UPF1 et UPF2 soient principalement cytoplasmiques, il n'est pas à exclure un rôle potentiel direct ou indirect dans l'épissage des exons 71 et 78 de Dp427m. Une autre étude a déjà rapporté dans d'autres gènes des changements du niveau d'inclusion d'exons en condition de déplétion d'UPF1 (Wang et al., 2014) ; (ii) un changement du rapport d'expression entre les isoformes Dp427m et Dp71 dans les cellules musculaires C25C148 déplétées pour UPF1 et UPF2. Nous avons en effet observé une forte diminution de la quantité de transcrits Dp427m en condition de KD de UPF1 et UPF2, ramenant son expression à un niveau comparable à celui de la Dp71 caractérisée par un niveau plus élevé d'épissage alternatif de ces exons (Figure Supplémentaire 5A-B, 74B). Les mécanismes conduisant à un changement drastique d'expression de la Dp427m et/ou de la Dp71 lorsque l'expression des facteurs du NMD est inhibée ainsi que leur rôle dans l'épissage de *DMD* restent néanmoins à déterminer.

Au cours de nos expériences, nous avons observé une différence de réponse au KD de eIF4A3 et de Y14 dans la lignée C25C148 malgré une très bonne efficacité du si-ARN dans les deux cas (Figure 60A). Cette observation ne concerne pas l'exon 78 qui semble avoir une régulation spécifique par Y14. Ceci a été également constaté pour un exon contrôle d'un autre gène (exon 11 de KPNA1, Figure 63D). Ce résultat est surprenant compte tenu que les deux protéines contribuent à part égale à la formation du complexe EJC. Ces différences n'avaient pas été vues dans de précédentes études réalisées dans les cellules HeLa (Wang et al., 2014), et effectivement nous avons pu observer une contribution égale des deux protéines sur l'épissage des exons DMD dans les expériences réalisées dans les cellules HeLa. Ces cellules ont été utilisées pour les expériences de sauvetage de l'expression de eIF4A3 et Y14 avec des vecteurs d'expression codant la protéine sauvage ou mutante pour la formation du complexe. Nous avons ainsi mis en évidence le rôle spécifique de chacune des protéines dans la régulation de l'épissage des exons DMD, et montré que la formation du complexe serait aussi nécessaire à la régulation de l'épissage des exons DMD par l'EJC. Les données montrent donc une différence

de régulation par eIF4A3 et Y14 entre les deux modèles cellulaires ou révèlent des limitations techniques particulières, non identifiées, dans la lignée cellulaire musculaire.

L'intérêt d'avoir dû choisir le modèle HeLa pour les expériences de sauvetage de l'expression de eIF4A3 et de Y14 en raison de l'impossibilité de mettre au point les expériences de sauvetage dans les C25C148, est d'avoir pu démontrer que la régulation des exons DMD par l'EJC dans la partie distale du gène n'était pas spécifique de l'isoforme Dp427m mais était aussi conservée dans l'isoforme Dp71. Les paramètres d'assemblage de l'EJC auraient pu être modifiés dans un contexte cellulaire différent et surtout au sein d'un transcrit beaucoup plus court dans lequel les durées de transcription et d'épissage sont réduites.

### **MNL51 controversé**

La déplétion du composant de l'EJC, MLN51, n'a pas montré les mêmes dérégulations du transcrit Dp427m. En effet, seule l'augmentation du saut de l'exon 78 a été observée dans cette condition. Ce résultat est en accord avec les données récentes de la littérature qui rapportent un rôle différent de MLN51 par rapport aux protéines du sous complexe trimérique eIF4A3/Y14-MAGOH. En effet, MLN51 est principalement cytoplasmique. Elle fait la navette entre ce compartiment et le noyau et ne semble pas faire partie de tous les EJC (Hauer et al., 2016; Singh et al., 2012). De plus, son inhibition n'empêche pas la formation du pré-EJC (eIF4A3, Y14/MAGOH) (Gerbracht et al., 2020). Les EJC multimériques assemblés avec des protéines SR et RNPS1, subiraient ensuite un changement de conformation pour exister principalement en tant qu' EJC monomérique contenant seulement MLN51 et diminuant l'activité du NMD (Mabin et al., 2018). Son rôle dans l'épissage reste débattu. La déplétion de MLN51 a été décrite comme régulant globalement l'épissage de gène (Wang et al., 2014) ou comme stabilisant les transcrits cibles du NMD (Gerbracht et al., 2020). Le fait que nous ayons observé que MLN51 régulait seulement l'exon 78, suggère fortement qu'il ne possède pas le même rôle que les autres protéines de l'EJC et qu'il pourrait ne pas faire probablement partie intégrante de tous les EJC de la partie distale de Dp427m. L'effet du KD de MLN51 est comparable au KD d'UPF1, indiquant un rôle potentiel en lien avec le système NMD. Des expériences d'immunoprécipitation pourraient être réalisées afin de déterminer si MLN51 se trouve sur toutes les jonctions d'exons de Dp427m dans les C25C148.

### **Analyse des complexes ASAP/PSAP**

L'étude des protéines des complexes ASAP/PSAP a été initialement guidée par notre intérêt d'étudier la protéine RNPS1. Cette protéine apparaît comme un régulateur majeur de l'épissage

(Schlautmann et al., 2021) et est décrite pour contribuer avec le complexe PSAP à la répression de l'utilisation des sites cryptiques 5' d'épissage (Boehm et al., 2018). Cependant, nous n'avons pas observé de modification de l'activation des deux sites alternatifs identifiés dans l'exon 9 et 70 en condition de KD RNPS1. L'hypothèse d'une protection par l'EJC contre un ré-épissage dans le transcrit déjà épissé reste plausible pour ces deux sites situés en début d'exon (+6 dans l'exon 9 et +36 dans l'exon 70). De manière intéressante, l'analyse des données de iCLIP ont montré la présence potentielle d'un dépôt non canonique de l'EJC (ncEJC) en amont du site alternatif, dans l'exon 70. Les expériences de RNA pull-down n'ont pas permis à ce jour de valider cette observation, mais la démonstration de l'existence des dépôts ncEJC reste une question intéressante à explorer. Comme les autres protéines des complexes ASAP/PSAP étudiées (ACINUS, PININ, SAP18), RNPS1 module uniquement l'épissage les deux plus petits exons 71 et 78 de Dp427m. De façon intéressante, ASAP permet l'exclusion de l'exon 71 et l'inclusion de l'exon 78. A l'inverse, PSAP permet d'inclure l'exon 71 et d'exclure l'exon 78. Ceci est en accord avec les données de la littérature qui montrent que ces deux complexes ne régulent pas les mêmes événements d'épissage (Wang et al., 2018). Ils semblent jouer un rôle distinct et complémentaire en lien avec l'EJC dans la régulation des transcrits en fonction des besoins cellulaires.

### **Rôle de l'EJC dans la différenciation musculaire**

Parallèlement à l'impact sur l'épissage, nous avons montré que la déplétion des facteurs de l'EJC (eIF4A3 et Y14) influençait la myogenèse et l'engagement des cellules musculaires C25C148 dans la différenciation. En effet, les niveaux protéiques de la Dp427m et les marqueurs de différenciation (MyoG, MyHC, Troponine T, ACTA1 et CKM) étaient très diminués contrairement aux marqueurs exprimés dans les myoblastes en prolifération (Myf5 et MyoD). Dans nos expériences, la prolifération ne semble pas altérée (nombre de noyaux comparable par rapport au contrôle). Par conséquent, le processus de différenciation pourrait être bloqué en G1 du cycle cellulaire lors de l'engagement des myoblastes en différenciation, ce qui expliquerait la forte expression des transcrits MyoD. Cela suggère aussi qu'il n'y a pas de processus de dédifférenciation vers des cellules non différenciées. Les cellules en condition de KD de Y14 semblent progresser plus loin dans le processus de différenciation qu'en condition de KD de eIF4A3 car le niveau de myogénine est normal. On peut supposer que les cellules atteignent le stade de myocytes sans pour autant permettre la fusion afin de former des myotubes dans la différenciation tardive. Cela montre des effets potentiellement différents entre

eIF4A3 et Y14 qui devront être validés par des expériences complémentaires d'immunomarquage dans les différentes conditions de déplétion.

Nous n'avons pas pu étudier plus en détail les mécanismes de blocage de la différenciation induit par la diminution de l'expression des facteurs de l'EJC. Celui-ci pourrait être consécutif à la dérégulation de l'expression de gènes impliqués dans la différenciation comme les hélicases DDX5/17 (Dardenne et al., 2014) ou à un impact sur le cycle cellulaire comme décrit pour eIF4A3 qui inhibe la transition M/G1 (Kanellis et al., 2021; Ryu et al., 2019) et pourrait empêcher l'engagement du processus de différenciation des myotubes en G1. Pour cela, il serait nécessaire d'étudier (i) l'effet de la déplétion de chacun des composants de l'EJC sur la différenciation (ii) la cinétique d'expression des facteurs de l'EJC du stade myoblaste jusqu'à la différenciation en myotubes dans des cellules primaires musculaires, modèle plus adapté à l'étude des processus de la myogenèse.

### **Rôle du NMD dans la différenciation musculaire**

La déplétion UPF1 a montré une augmentation globale de la différenciation comparée au contrôle, avec une surexpression très forte des marqueurs de différenciation tardifs mais aussi du facteur d'engagement des myoblastes, MyoD, dans ce processus. Ceci est accord avec la littérature qui montre que la déplétion d'UPF1 accélère le processus de la myogenèse (Feng et al., 2017). En l'absence d'UPF1, qui normalement séquestre la protéine MyoD, celle-ci est libérée afin d'activer la différenciation. Nos données suggèrent que la mise en place rapide des myotubes en l'absence d'UPF1 passe par cette voie. Néanmoins, il peut y voir une régulation complexe entre les voies NMD et SMD impliquant UPF1, qui répriment et activent respectivement la différenciation musculaire (Gong et al., 2009). La balance de régulation qui engage fortement le processus de différenciation pourrait être étudiée par l'inhibition de Staufen-1 (SMD), du facteur UPF3B de la voie alternative du NMD ainsi que des analyses complémentaires du facteur UPF2 déjà réalisées (partie III des Résultats complémentaires). Par ailleurs nous avons observé la diminution de l'expression de la protéine Dp427m en condition de déplétion d'UPF1, ce qui était inattendu compte tenu qu'elle est normalement exprimée dans les myotubes. Cette baisse du niveau de protéine est corrélée à une baisse du niveau des transcrits Dp427m contrairement aux conditions de KD des composants de l'EJC pour lesquelles une baisse similaire de niveau de dystrophine est observée sans modification du niveau de transcrit. Il s'agit de résultats nouveaux concernant le rôle de l'EJC dans la différenciation musculaire chez l'homme, qui ouvrent de nouvelles perspectives d'étude des mécanismes mis en jeu.

## **L'EJC régule l'épissage de deux domaines importants de la dystrophine.**

Contrairement à notre hypothèse de départ basée sur les données obtenues chez la drosophile (Hayashi et al., 2014), la déplétion des composants de l'EJC n'a pas eu d'effet sur l'inclusion d'exons bordés par de longs introns (soumis pour un nombre d'entre eux à un épissage récursif). Nos résultats montrent clairement un biais vers les exons de la partie distale du transcrit. Cette région des exons 68 à 78 possède des caractéristiques particulières de structure génomique et fonctionnelles car elle contient (i) essentiellement des petits introns (6/11 ont une taille inférieure à 3000 pb, dont deux < 1000 pb) (ii) un grand nombre d'exons alternativement épissés dans les autres isoformes de dystrophine (iii) le site alternatif de polyadénylation utilisé pour produire la Dp40 dans le petit intron 70 (698 pb) (iv) l'exon 78 dont l'expression est essentielle à la formation du muscle au cours de la myogenèse. De manière intéressante, nous avons montré le même rôle de l'EJC sur l'épissage des exons distaux dans l'isoforme Dp71 exprimée dans les cellules HeLa. De même, la rétention de l'intron 70 semble spécifique, car l'inspection des données de RNA-Seq n'a pas montré de rétention pour les dix autres introns dont la taille est inférieure à 1000 pb dans le gène DMD.

Les exons 68 à 79 représentent la partie codante des domaines riche en Cystéine et C-terminal de la dystrophine. Ces deux domaines sont impliqués dans des interactions protéine-protéine essentielles avec la  $\beta$ -dystroglycane au niveau de la membrane plasmique, des syntrophines et de la dystrobrevine cytoplasmiques pour les fonctions de maintien de la fibre musculaire et de signalisation cellulaire. L'épissage de cette région est régulé de manière différente dans le muscle squelettique (où tous les exons sont inclus) et dans les isoformes cérébrales où l'épissage de ces exons peut varier en fonctions des types cellulaires. L'épissage alternatif agit sur la localisation sub-cellulaire des isoformes ainsi que sur le recrutement d'effecteurs des voies de signalisation comme les des syntrophines.

Ces résultats contribuent à identifier des régulateurs de l'expression de ces différentes isoformes, mettant en évidence le rôle du complexe EJC dans le contrôle de la fidélité de l'épissage dans cette région du gène. Ce rôle est à intégrer avec les connaissances actuelles qui montrent : (i) un rôle des composants de l'EJC dans les pathologies neurodéveloppementales, (ii) une implication des défauts d'expression des isoformes cérébrales de la dystrophine dans le déficit cognitif observé chez les patients et la prévalence accrue des troubles comportementaux associés. Cela ouvre de nouvelles perspectives de recherche visant à étudier le rôle du complexe EJC dans la régulation de l'expression de la dystrophine dans le système nerveux central à l'aide de modèles cellulaires pertinents.

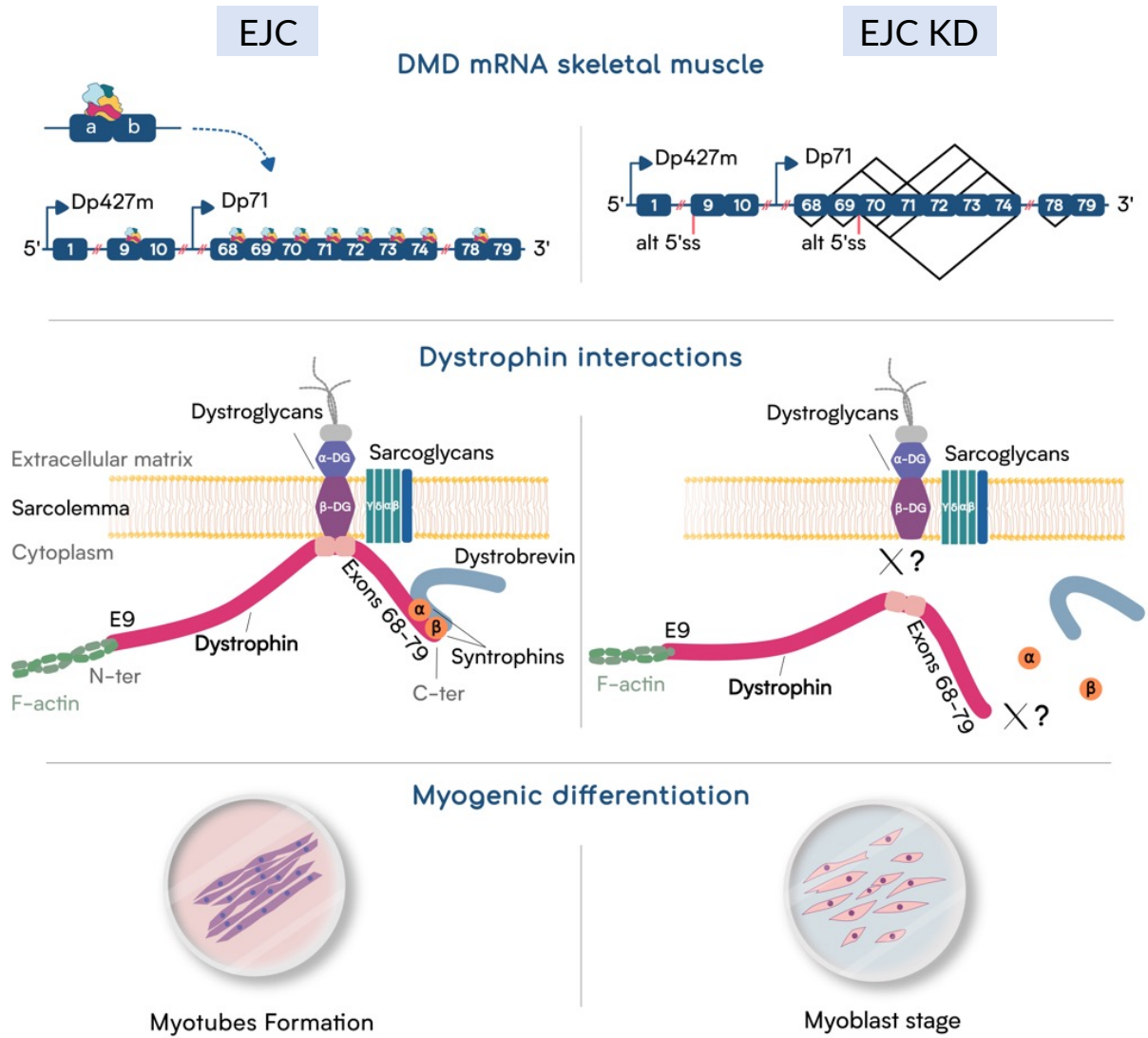


Figure 76. Résultats principaux du projet de thèse.

## Bibliographie

- Aartsma-Rus, A., Straub, V., Hemmings, R., Haas, M., Schlosser-Weber, G., Stoyanova-Beninska, V., Mercuri, E., Muntoni, F., Sepodes, B., Vroom, E., Balabanov, P., 2017. Development of Exon Skipping Therapies for Duchenne Muscular Dystrophy: A Critical Review and a Perspective on the Outstanding Issues. *Nucleic Acid Ther.* 27, 251–259. <https://doi.org/10.1089/nat.2017.0682>
- Acosta, R., Montañez, C., Fuentes-Mera, L., Gonzalez, E., Gómez, P., Quintero-Mora, L., Mornet, D., Álvarez-Salas, L.M., Cisneros, B., 2004. Dystrophin Dp71 is required for neurite outgrowth in PC12 cells. *Exp. Cell Res.* 296, 265–275. <https://doi.org/10.1016/j.yexcr.2004.01.015>
- Adams, M.E., Odom, G.L., Kim, M.J., Chamberlain, J.S., Froehner, S.C., 2018. Syntrophin binds directly to multiple spectrin-like repeats in dystrophin and mediates binding of nNOS to repeats 16–17. *Hum. Mol. Genet.* 27, 2978–2985. <https://doi.org/10.1093/hmg/ddy197>
- Akhtar, J., Kreim, N., Marini, F., Mohana, G., Brüne, D., Binder, H., Roignant, J.-Y., 2019. Promoter-proximal pausing mediated by the exon junction complex regulates splicing. *Nat. Commun.* 10, 521. <https://doi.org/10.1038/s41467-019-08381-0>
- Aleman, V., Osorio, B., Chavez, O., Rendon, A., Mornet, D., Martinez, D., 2001. Subcellular localization of Dp71 dystrophin isoforms in cultured hippocampal neurons and forebrain astrocytes. *Histochem. Cell Biol.* 115, 243–254. <https://doi.org/10.1007/s004180000221>
- Alexandrov, A., Colognori, D., Shu, M.-D., Steitz, J.A., 2012. Human spliceosomal protein CWC22 plays a role in coupling splicing to exon junction complex deposition and nonsense-mediated decay. *Proc. Natl. Acad. Sci. U. S. A.* 109, 21313–21318. <https://doi.org/10.1073/pnas.1219725110>
- Alharbi, A.B., Schmitz, U., Bailey, C.G., Rasko, J.E.J., 2021. CTCF as a regulator of alternative splicing: new tricks for an old player. *Nucleic Acids Res.* 49, 7825–7838. <https://doi.org/10.1093/nar/gkab520>
- Allen, D.G., Whitehead, N.P., Froehner, S.C., 2016. Absence of Dystrophin Disrupts Skeletal Muscle Signaling: Roles of Ca<sup>2+</sup>, Reactive Oxygen Species, and Nitric Oxide in the Development of Muscular Dystrophy. *Physiol. Rev.* 96, 253–305. <https://doi.org/10.1152/physrev.00007.2015>
- Alrahbeni, T., Sartor, F., Anderson, J., Miedzybrodzka, Z., McCaig, C., Müller, B., 2015. Full UPF3B function is critical for neuronal differentiation of neural stem cells. *Mol. Brain* 8, 33. <https://doi.org/10.1186/s13041-015-0122-1>
- Amann, K.J., Renley, B.A., Ervasti, J.M., 1998. A Cluster of Basic Repeats in the Dystrophin Rod Domain Binds F-actin through an Electrostatic Interaction\*. *J. Biol. Chem.* 273, 28419–28423. <https://doi.org/10.1074/jbc.273.43.28419>
- Ameziane-Le Hir, S., Raguénès-Nicol, C., Paboef, G., Nicolas, A., Le Rumeur, E., Vié, V., 2014. Cholesterol favors the anchorage of human dystrophin repeats 16 to 21 in membrane at physiological surface pressure. *Biochim. Biophys. Acta BBA - Biomembr.* 1838, 1266–1273. <https://doi.org/10.1016/j.bbamem.2014.01.010>
- Andersen, C.B.F., Ballut, L., Johansen, J.S., Chamieh, H., Nielsen, K.H., Oliveira, C.L.P., Pedersen, J.S., Séraphin, B., Hir, H.L., Andersen, G.R., 2006. Structure of the Exon Junction Core Complex with a Trapped DEAD-Box ATPase Bound to RNA. *Science* 313, 1968–1972. <https://doi.org/10.1126/science.1131981>
- Aragón, J., González-Reyes, M., Romo-Yáñez, J., Vacca, O., Aguilar-González, G., Rendón, A., Vaillend, C., Montañez, C., 2018. Dystrophin Dp71 Isoforms Are Differentially Expressed in the Mouse Brain and Retina: Report of New Alternative Splicing and a Novel Nomenclature for Dp71 Isoforms. *Mol. Neurobiol.* 55, 1376–1386. <https://doi.org/10.1007/s12035-017-0405-x>
- Aragón, J., Martínez-Herrera, A., Bermúdez-Cruz, R.M., Bazán, M.L., Soid-Raggi, G., Ceja, V., Coy-Arechavaleta, A.S., Alemán, V., Depardón, F., Montañez, C., 2015. EF-hand domains are involved in the differential cellular distribution of dystrophin Dp40. *Neurosci. Lett.* 600, 115–120. <https://doi.org/10.1016/j.neulet.2015.05.038>
- Ashton-Beaucage, D., Udell, C.M., Lavoie, H., Baril, C., Lefrançois, M., Chagnon, P., Gendron, P., Caron-Lizotte, O., Bonnell, É., Thibault, P., Therrien, M., 2010. The Exon Junction Complex Controls the Splicing of mapk and Other Long Intron-Containing Transcripts in *Drosophila*. *Cell* 143, 251–262. <https://doi.org/10.1016/j.cell.2010.09.014>
- Asthana, S., Martin, H., Rupkey, J., Patel, S., Yoon, J., Keegan, A., Mao, Y., 2022. The Physiological Roles of the Exon Junction Complex in Development and Diseases. *Cells* 11, 1192. <https://doi.org/10.3390/cells11071192>
- Ballut, L., Marchadier, B., Baguet, A., Tomasetto, C., Séraphin, B., Le Hir, H., 2005. The exon junction core complex is locked onto RNA by inhibition of eIF4AIII ATPase activity. *Nat. Struct. Mol. Biol.* 12, 861–869. <https://doi.org/10.1038/nsmb990>

- Bao, J., Vitting-Seerup, K., Waage, J., Tang, C., Ge, Y., Porse, B.T., Yan, W., 2016. UPF2-Dependent Nonsense-Mediated mRNA Decay Pathway Is Essential for Spermatogenesis by Selectively Eliminating Longer 3'UTR Transcripts. *PLoS Genet.* 12, e1005863. <https://doi.org/10.1371/journal.pgen.1005863>
- Bao, S., Moakley, D.F., Zhang, C., 2019. The Splicing Code Goes Deep. *Cell* 176, 414–416. <https://doi.org/10.1016/j.cell.2019.01.013>
- Baralle, F.E., Giudice, J., 2017. Alternative splicing as a regulator of development and tissue identity. *Nat. Rev. Mol. Cell Biol.* 18, 437–451. <https://doi.org/10.1038/nrm.2017.27>
- Baranello, L., Bertozzi, D., Fogli, M.V., Pommier, Y., Capranico, G., 2010. DNA topoisomerase I inhibition by camptothecin induces escape of RNA polymerase II from promoter-proximal pause site, antisense transcription and histone acetylation at the human HIF-1 $\alpha$  gene locus. *Nucleic Acids Res.* 38, 159–171. <https://doi.org/10.1093/nar/gkp817>
- Barbosa, I., Haque, N., Fiorini, F., Barrandon, C., Tomasetto, C., Blanchette, M., n.d. Human CWC22 escorts the helicase eIF4AIII to spliceosomes and promotes exon junction complex assembly. *Mol. Biol.* 9.
- Bartys, N., Kierzek, R., Lisowiec-Wachnicka, J., 2019. The regulation properties of RNA secondary structure in alternative splicing. *Biochim. Biophys. Acta BBA - Gene Regul. Mech.* 1862, 194401. <https://doi.org/10.1016/j.bbagr.2019.07.002>
- Beggs, A.H., Hoffman, E.P., Snyder, J.R., Arahata, K., Specht, L., Shapiro, F., Angelini, C., Sugita, H., Kunkel, L.M., 1991. Exploring the molecular basis for variability among patients with Becker muscular dystrophy: dystrophin gene and protein studies. *Am. J. Hum. Genet.* 49, 54–67.
- Belanto, J.J., Mader, T.L., Eckhoff, M.D., Strandjord, D.M., Banks, G.B., Gardner, M.K., Lowe, D.A., Ervasti, J.M., 2014. Microtubule binding distinguishes dystrophin from utrophin. *Proc. Natl. Acad. Sci. U. S. A.* 111, 5723–5728. <https://doi.org/10.1073/pnas.1323842111>
- Belgrader, P., Cheng, J., Maquat, L.E., 1993. Evidence to implicate translation by ribosomes in the mechanism by which nonsense codons reduce the nuclear level of human triosephosphate isomerase mRNA. *Proc. Natl. Acad. Sci.* 90, 482–486. <https://doi.org/10.1073/pnas.90.2.482>
- Bernier, R., Steinman, K.J., Reilly, B., Wallace, A.S., Sherr, E.H., Pojman, N., Mefford, H.C., Gerds, J., Earl, R., Hanson, E., Goin-Kochel, R.P., Berry, L., Kanne, S., Snyder, L.G., Spence, S., Ramocki, M.B., Evans, D.W., Spiro, J.E., Martin, C.L., Ledbetter, D.H., Chung, W.K., Simons VIP consortium, 2016. Clinical phenotype of the recurrent 1q21.1 copy-number variant. *Genet. Med. Off. J. Am. Coll. Med. Genet.* 18, 341–349. <https://doi.org/10.1038/gim.2015.78>
- Bertola, D.R., Hsia, G., Alvizi, L., Gardham, A., Wakeling, E.L., Yamamoto, G.L., Honjo, R.S., Oliveira, L. a. N., Di Francesco, R.C., Perez, B.A., Kim, C.A., Passos-Bueno, M.R., 2018. Richieri-Costa-Pereira syndrome: Expanding its phenotypic and genotypic spectrum. *Clin. Genet.* 93, 800–811. <https://doi.org/10.1111/cge.13169>
- Bertram, K., Agafonov, D.E., Liu, W.-T., Dybkov, O., Will, C.L., Hartmuth, K., Urlaub, H., Kastner, B., Stark, H., Lührmann, R., 2017. Cryo-EM structure of a human spliceosome activated for step 2 of splicing. *Nature* 542, 318–323. <https://doi.org/10.1038/nature21079>
- Bhosle, R.C., Michele, D.E., Campbell, K.P., Li, Z., Robson, R.M., 2006. Interactions of intermediate filament protein synemin with dystrophin and utrophin. *Biochem. Biophys. Res. Commun.* 346, 768–777. <https://doi.org/10.1016/j.bbrc.2006.05.192>
- Bicknell, A.A., Cenik, C., Chua, H.N., Roth, F.P., Moore, M.J., 2012. Introns in UTRs: Why we should stop ignoring them. *BioEssays* 34, 1025–1034. <https://doi.org/10.1002/bies.201200073>
- Bies, R.D., Phelps, S.F., Cortez, M.D., Roberts, R., Caskey, C.T., Chamberlain, J.S., 1992. Human and murine dystrophin mRNA transcripts are differentially expressed during skeletal muscle, heart, and brain development. *Nucleic Acids Res.* 20, 1725–1731.
- Black, D.L., n.d. Does steric interference between splice sites block the splicing of a short c-src neuron-specific exon in non-neuronal cells? 15.
- Bladen, C.L., Salgado, D., Monges, S., Foncuberta, M.E., Kekou, K., Kosma, K., Dawkins, H., Lamont, L., Roy, A.J., Chamova, T., Guerguelcheva, V., Chan, S., Korngut, L., Campbell, C., Dai, Y., Wang, J., Barišić, N., Brabec, P., Lahdetie, J., Walter, M.C., Schreiber-Katz, O., Karcagi, V., Garami, M., Viswanathan, V., Bayat, F., Buccella, F., Kimura, E., Koeks, Z., van den Bergen, J.C., Rodrigues, M., Roxburgh, R., Lusakowska, A., Kostera-Pruszczyk, A., Zimowski, J., Santos, R., Neagu, E., Artemieva, S., Rasic, V.M., Vojinovic, D., Posada, M., Bloetzer, C., Jeannet, P.-Y., Joncourt, F., Díaz-Manera, J., Gallardo, E., Karaduman, A.A., Topaloglu, H., El Sherif, R., Stringer, A., Shatillo, A.V., Martin, A.S., Peay, H.L., Bellgard, M.I., Kirschner, J., Flanigan, K.M., Straub, V., Bushby, K., Verschuuren, J., Aartsma-Rus, A., Bérout, C., Lochmüller, H., 2015. The TREAT-NMD DMD Global Database: analysis of more than 7,000 Duchenne muscular dystrophy mutations.



Hum. Mutat. 36, 395–402. <https://doi.org/10.1002/humu.22758>

Blagosklonny, M.V., 2004. Flavopiridol, An Inhibitor of Transcription: Implications, Problems and Solutions. *Cell Cycle* 3, 1537–1542. <https://doi.org/10.4161/cc.3.12.1278>

Bland, C.S., Wang, E.T., Vu, A., David, M.P., Castle, J.C., Johnson, J.M., Burge, C.B., Cooper, T.A., 2010. Global regulation of alternative splicing during myogenic differentiation. *Nucleic Acids Res.* 38, 7651–7664. <https://doi.org/10.1093/nar/gkq614>

Blazquez, L., Emmett, W., Faraway, R., Pineda, J.M.B., Bajew, S., Gohr, A., Haberman, N., Sibley, C.R., Bradley, R.K., Irimia, M., Ule, J., 2018. Exon Junction Complex Shapes the Transcriptome by Repressing Recursive Splicing. *Mol. Cell* 72, 496–509.e9. <https://doi.org/10.1016/j.molcel.2018.09.033>

Boehm, V., Britto-Borges, T., Steckelberg, A.-L., Singh, K.K., Gerbracht, J.V., Gueney, E., Blazquez, L., Altmüller, J., Dieterich, C., Gehring, N.H., 2018. Exon Junction Complexes Suppress Spurious Splice Sites to Safeguard Transcriptome Integrity. *Mol. Cell* 72, 482–495.e7. <https://doi.org/10.1016/j.molcel.2018.08.030>

Boehm, V., Haberman, N., Ottens, F., Ule, J., Gehring, N.H., 2014. 3' UTR Length and Messenger Ribonucleoprotein Composition Determine Endocleavage Efficiencies at Termination Codons. *Cell Rep.* 9, 555–568. <https://doi.org/10.1016/j.celrep.2014.09.012>

Bono, F., Cook, A.G., Grünwald, M., Ebert, J., Conti, E., 2010. Nuclear Import Mechanism of the EJC Component Mago-Y14 Revealed by Structural Studies of Importin 13. *Mol. Cell* 37, 211–222. <https://doi.org/10.1016/j.molcel.2010.01.007>

Bono, F., Ebert, J., Lorentzen, E., Conti, E., 2006. The Crystal Structure of the Exon Junction Complex Reveals How It Maintains a Stable Grip on mRNA. *Cell* 126, 713–725. <https://doi.org/10.1016/j.cell.2006.08.006>

Borao, S., Ayté, J., Hümmer, S., 2021. Evolution of the Early Spliceosomal Complex—From Constitutive to Regulated Splicing. *Int. J. Mol. Sci.* 22, 12444. <https://doi.org/10.3390/ijms222212444>

Bougé, A.-L., Murauer, E., Beyne, E., Miro, J., Varilh, J., Taulan, M., Koenig, M., Claustres, M., Tuffery-Giraud, S., 2017. Targeted RNA-Seq profiling of splicing pattern in the *DMD* gene: exons are mostly constitutively spliced in human skeletal muscle. *Sci. Rep.* 7, 39094. <https://doi.org/10.1038/srep39094>

Bourgeois, C.F., Mortreux, F., Auboeuf, D., 2016. The multiple functions of RNA helicases as drivers and regulators of gene expression. *Nat. Rev. Mol. Cell Biol.* 17, 426–438. <https://doi.org/10.1038/nrm.2016.50>

Boussion, S., Escande, F., Jourdain, A.-S., Smol, T., Brunelle, P., Duhamel, C., Alembik, Y., Attié-Bitach, T., Baujat, G., Bazin, A., Bonnière, M., Carassou, P., Carles, D., Devisme, L., Goizet, C., Goldenberg, A., Grotto, S., Guichet, A., Jouk, P.-S., Loeuillet, L., Mechler, C., Michot, C., Pelluard, F., Putoux, A., Whalen, S., Ghomid, J., Manouvrier-Hanu, S., Petit, F., 2020. TAR syndrome: Clinical and molecular characterization of a cohort of 26 patients and description of novel noncoding variants of RBM8A. *Hum. Mutat.* <https://doi.org/10.1002/humu.24021>

Braunschweig, U., Gueroussov, S., Plocik, A.M., Graveley, B.R., Blencowe, B.J., 2013. Dynamic Integration of Splicing within Gene Regulatory Pathways. *Cell* 152, 1252–1269. <https://doi.org/10.1016/j.cell.2013.02.034>

Brinegar, A.E., Xia, Z., Loehr, J.A., Li, W., Rodney, G.G., Cooper, T.A., n.d. Extensive alternative splicing transitions during postnatal skeletal muscle development are required for calcium handling functions 21.

Buchwald, G., Ebert, J., Basquin, C., Sauliere, J., Jayachandran, U., Bono, F., Le Hir, H., Conti, E., 2010. Insights into the recruitment of the NMD machinery from the crystal structure of a core EJC-UPF3b complex. *Proc. Natl. Acad. Sci.* 107, 10050–10055. <https://doi.org/10.1073/pnas.1000993107>

Buchwald, G., Schüssler, S., Basquin, C., Le Hir, H., Conti, E., 2013. Crystal structure of the human eIF4AIII–CWC22 complex shows how a DEAD-box protein is inhibited by a MIF4G domain. *Proc. Natl. Acad. Sci. U. S. A.* 110, E4611–E4618. <https://doi.org/10.1073/pnas.1314684110>

Buckingham, M., Relaix, F., 2007. The Role of Pax Genes in the Development of Tissues and Organs: Pax3 and Pax7 Regulate Muscle Progenitor Cell Functions. *Annu. Rev. Cell Dev. Biol.* 23, 645–73. <https://doi.org/10.1146/annurev.cellbio.23.090506.123438>

Busetto, V., Barbosa, I., Basquin, J., Marquet, É., Hocq, R., Hennion, M., Paternina, J.A., Namane, A., Conti, E., Bensaude, O., Le Hir, H., 2020. Structural and functional insights into CWC27/CWC22 heterodimer linking the exon junction complex to spliceosomes. *Nucleic Acids Res.* 48, 5670–5683. <https://doi.org/10.1093/nar/gkaa267>

Bushby, K.M., Thambyayah, M., Gardner-Medwin, D., 1991. Prevalence and incidence of Becker muscular dystrophy. *Lancet Lond. Engl.* 337, 1022–1024. [https://doi.org/10.1016/0140-6736\(91\)92671-n](https://doi.org/10.1016/0140-6736(91)92671-n)

- Byers, T.J., Lidov, H.G.W., Kunkel, L.M., 1993. An alternative dystrophin transcript specific to peripheral nerve. *Nat. Genet.* 4, 77–81. <https://doi.org/10.1038/ng0593-77>
- Caretti, G., Schiltz, R.L., Dilworth, F.J., Di Padova, M., Zhao, P., Ogryzko, V., Fuller-Pace, F.V., Hoffman, E.P., Tapscott, S.J., Sartorelli, V., 2006. The RNA Helicases p68/p72 and the Noncoding RNA SRA Are Coregulators of MyoD and Skeletal Muscle Differentiation. *Dev. Cell* 11, 547–560. <https://doi.org/10.1016/j.devcel.2006.08.003>
- Carter, M.S., Li, S., Wilkinson, M.F., 1996. A splicing-dependent regulatory mechanism that detects translation signals. *EMBO J.* 15, 5965–5975.
- Castle, J.C., Zhang, C., Shah, J.K., Kulkarni, A.V., Cooper, T.A., Johnson, J.M., 2008. Differential expression of 24,426 human alternative splicing events and predicted cis-regulation in 48 tissues and cell lines. *Nat. Genet.* 40, 1416–1425. <https://doi.org/10.1038/ng.264>
- Chakrabarti, S., Bonneau, F., Schüssler, S., Eppinger, E., Conti, E., 2014. Phospho-dependent and phospho-independent interactions of the helicase UPF1 with the NMD factors SMG5–SMG7 and SMG6. *Nucleic Acids Res.* 42, 9447–9460. <https://doi.org/10.1093/nar/gku578>
- Chal, J., Pourquié, O., 2017. Making muscle: skeletal myogenesis in vivo and in vitro. *Development* 144, 2104–2122. <https://doi.org/10.1242/dev.151035>
- Chamieh, H., Ballut, L., Bonneau, F., Le Hir, H., 2008. NMD factors UPF2 and UPF3 bridge UPF1 to the exon junction complex and stimulate its RNA helicase activity. *Nat. Struct. Mol. Biol.* 15, 85–93. <https://doi.org/10.1038/nsmb1330>
- Chan, C.C., Dostie, J., Diem, M.D., Feng, W., Mann, M., Rappsilber, J., Dreyfuss, G., 2004. eIF4A3 is a novel component of the exon junction complex. *RNA N. Y. N* 10, 200–209. <https://doi.org/10.1261/rna.5230104>
- Chan, W.-K., Bhalla, A.D., Le Hir, H., Nguyen, L.S., Huang, L., Gécz, J., Wilkinson, M.F., 2009. A UPF3-mediated regulatory switch that maintains RNA surveillance. *Nat. Struct. Mol. Biol.* 16, 747–753. <https://doi.org/10.1038/nsmb.1612>
- Chan, W.-K., Huang, L., Gudikote, J.P., Chang, Y.-F., Imam, J.S., MacLean, J.A., Wilkinson, M.F., 2007. An alternative branch of the nonsense-mediated decay pathway. *EMBO J.* 26, 1820–1830. <https://doi.org/10.1038/sj.emboj.7601628>
- Chang, N.C., Rudnicki, M.A., 2014. Satellite Cells, in: *Current Topics in Developmental Biology*. Elsevier, pp. 161–181. <https://doi.org/10.1016/B978-0-12-416022-4.00006-8>
- Chávez, O., Harricane, M.C., Alemán, V., Dorbani, L., Larroque, C., Mornet, D., Rendon, A., Martínez-Rojas, D., 2000. Mitochondrial expression of a short dystrophin-like product with molecular weight of 71 kDa. *Biochem. Biophys. Res. Commun.* 274, 275–280. <https://doi.org/10.1006/bbrc.2000.3118>
- Chazal, P.-E., Daguene, E., Wendling, C., Ulryck, N., Tomasetto, C., Sargueil, B., Le Hir, H., 2013. EJC core component MLN51 interacts with eIF3 and activates translation. *Proc. Natl. Acad. Sci.* 110, 5903–5908. <https://doi.org/10.1073/pnas.1218732110>
- Chelly, J., Hamard, G., Koulakoff, A., Kaplan, J.-C., Kahn, A., Berwald-Netter, Y., 1990. Dystrophin gene transcribed from different promoters in neuronal and glial cells. *Nature* 344, 64–65. <https://doi.org/10.1038/344064a0>
- Chen, M., Manley, J.L., 2009. Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nat. Rev. Mol. Cell Biol.* 10, 741–754. <https://doi.org/10.1038/nrm2777>
- Cheng, J., Belgrader, P., Zhou, X., Maquat, L.E., 1994. Introns are cis effectors of the nonsense-codon-mediated reduction in nuclear mRNA abundance. *Mol. Cell. Biol.* 14, 6317–6325. <https://doi.org/10.1128/MCB.14.9.6317>
- Cho, H., Abshire, E.T., Popp, M.W., Pröschel, C., Schwartz, J.L., Yeo, G.W., Maquat, L.E., 2022. AKT constitutes a signal-promoted alternative exon-junction complex that regulates nonsense-mediated mRNA decay. *Mol. Cell* 82, 2779–2796.e10. <https://doi.org/10.1016/j.molcel.2022.05.013>
- Choi, I.Y., Lim, H., Cho, H.J., Oh, Y., Chou, B.-K., Bai, H., Cheng, L., Kim, Y.J., Hyun, S., Kim, H., Shin, J.H., Lee, G., 2020. Transcriptional landscape of myogenesis from human pluripotent stem cells reveals a key role of TWIST1 in maintenance of skeletal muscle progenitors. *eLife* 9, e46981. <https://doi.org/10.7554/eLife.46981>
- Choudhury, S.R., Singh, A.K., McLeod, T., Blanchette, M., Jang, B., Badenhorst, P., Kanhere, A., Brogna, S., n.d. Exon junction complex proteins bind nascent transcripts independently of pre-mRNA splicing in *Drosophila melanogaster*. *eLife* 5, e19881. <https://doi.org/10.7554/eLife.19881>
- Chuang, T.-W., Lee, K.-M., Tarn, W.-Y., 2015. Function and Pathological Implications of Exon Junction Complex Factor Y14 13.

- Contreras, J., Begley, V., Marsella, L., Villalobo, E., 2018. The splicing of tiny introns of *Paramecium* is controlled by MAGO. *Gene* 663, 101–109. <https://doi.org/10.1016/j.gene.2018.04.007>
- Daguenet, E., Baguet, A., Degot, S., Schmidt, U., Alpy, F., Wendling, C., Spiegelhalter, C., Kessler, P., Rio, M.-C., Le Hir, H., Bertrand, E., Tomasetto, C., 2012a. Perispeckles are major assembly sites for the exon junction core complex. *Mol. Biol. Cell* 23, 1765–1782. <https://doi.org/10.1091/mbc.e12-01-0040>
- Daguenet, E., Baguet, A., Degot, S., Schmidt, U., Alpy, F., Wendling, C., Spiegelhalter, C., Kessler, P., Rio, M.-C., Le Hir, H., Bertrand, E., Tomasetto, C., 2012b. Perispeckles are major assembly sites for the exon junction core complex. *Mol. Biol. Cell* 23, 1765–1782. <https://doi.org/10.1091/mbc.e12-01-0040>
- Damgaard, C.K., Kahns, S., Lykke-Andersen, S., Nielsen, A.L., Jensen, T.H., Kjems, J., 2008. A 5' Splice Site Enhances the Recruitment of Basal Transcription Initiation Factors In Vivo. *Mol. Cell* 29, 271–278. <https://doi.org/10.1016/j.molcel.2007.11.035>
- Daoud, F., Angeard, N., Demerre, B., Martie, I., Benyaou, R., Leturcq, F., Cossée, M., Deburgrave, N., Saillour, Y., Tuffery, S., Urtizbera, A., Toutain, A., Echenne, B., Frischman, M., Mayer, M., Desguerre, I., Estournet, B., Réveillère, C., Penisson-Besnier, null, Cuisset, J.M., Kaplan, J.C., Héron, D., Rivier, F., Chelly, J., 2009. Analysis of Dp71 contribution in the severity of mental retardation through comparison of Duchenne and Becker patients differing by mutation consequences on Dp71 expression. *Hum. Mol. Genet.* 18, 3779–3794. <https://doi.org/10.1093/hmg/ddp320>
- Dardenne, E., Polay Espinoza, M., Fattet, L., Germann, S., Lambert, M.-P., Neil, H., Zonta, E., Mortada, H., Gratadou, L., Deygas, M., Chakrama, F.Z., Samaan, S., Desmet, F.-O., Tranchevent, L.-C., Dutertre, M., Rimokh, R., Bourgeois, C.F., Auboeuf, D., 2014. RNA helicases DDX5 and DDX17 dynamically orchestrate transcription, miRNA, and splicing programs in cell differentiation. *Cell Rep.* 7, 1900–1913. <https://doi.org/10.1016/j.celrep.2014.05.010>
- Darzacq, X., Shav-Tal, Y., de Turrís, V., Brody, Y., Shenoy, S.M., Phair, R.D., Singer, R.H., 2007. In vivo dynamics of RNA polymerase II transcription. *Nat. Struct. Mol. Biol.* 14, 796–806. <https://doi.org/10.1038/nsmb1280>
- de Almeida, S.F., Grosso, A.R., Koch, F., Fenouil, R., Carvalho, S., Andrade, J., Levezinho, H., Gut, M., Eick, D., Gut, I., Andrau, J.-C., Ferrier, P., Carmo-Fonseca, M., 2011. Splicing enhances recruitment of methyltransferase HYPB/Setd2 and methylation of histone H3 Lys36. *Nat. Struct. Mol. Biol.* 18, 977–983. <https://doi.org/10.1038/nsmb.2123>
- De Conti, L., Baralle, M., Buratti, E., 2013. Exon and intron definition in pre-mRNA splicing. *WIREs RNA* 4, 49–60. <https://doi.org/10.1002/wrna.1140>
- De, I., Bessonov, S., Hofele, R., dos Santos, K., Will, C.L., Urlaub, H., Lührmann, R., Pena, V., 2015. The RNA helicase Aquarius exhibits structural adaptations mediating its recruitment to spliceosomes. *Nat. Struct. Mol. Biol.* 22, 138–144. <https://doi.org/10.1038/nsmb.2951>
- de la Mata, M., Kornblihtt, A.R., 2006. RNA polymerase II C-terminal domain mediates regulation of alternative splicing by SRp20. *Nat. Struct. Mol. Biol.* 13, 973–980. <https://doi.org/10.1038/nsmb1155>
- de León, M.B., Montañez, C., Gómez, P., Morales-Lázaro, S.L., Tapia-Ramírez, V., Valadez-Graham, V., Recillas-Targa, F., Yaffe, D., Nudel, U., Cisneros, B., 2005. Dystrophin Dp71 Expression Is Down-regulated during Myogenesis. *J. Biol. Chem.* 280, 5290–5299. <https://doi.org/10.1074/jbc.M411571200>
- Debaize, L., Troadec, M.-B., 2019. The master regulator FUBP1: its emerging role in normal cell function and malignant development. *Cell. Mol. Life Sci. CMLS* 76, 259–281. <https://doi.org/10.1007/s00018-018-2933-6>
- Degot, S., Le Hir, H., Alpy, F., Kedinger, V., Stoll, I., Wendling, C., Seraphin, B., Rio, M.-C., Tomasetto, C., 2004. Association of the Breast Cancer Protein MLN51 with the Exon Junction Complex via Its Speckle Localizer and RNA Binding Module. *J. Biol. Chem.* 279, 33702–33715. <https://doi.org/10.1074/jbc.M402754200>
- Disset, A., Bourgeois, C.F., Benmalek, N., Claustres, M., Stevenin, J., Tuffery-Giraud, S., 2006. An exon skipping-associated nonsense mutation in the dystrophin gene uncovers a complex interplay between multiple antagonistic splicing elements. *Hum. Mol. Genet.* 15, 999–1013. <https://doi.org/10.1093/hmg/ddl015>
- Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., Gingeras, T.R., 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21. <https://doi.org/10.1093/bioinformatics/bts635>
- Dominski, Z., Kole, R., 1991. Selection of splice sites in pre-mRNAs with short internal exons. *Mol. Cell. Biol.* 11, 6075–6083.
- Doorenweerd, N., 2020. Combining genetics, neuropsychology and neuroimaging to improve understanding of brain involvement in Duchenne muscular dystrophy - a narrative review. *Neuromuscul. Disord. NMD* 30, 437–442. <https://doi.org/10.1016/j.nmd.2020.05.001>

- Doorenweerd, N., Mahfouz, A., Putten, M. van, Kaliyaperumal, R., Hoen, P.A.C. t', Hendriksen, J.G.M., Aartsma-Rus, A.M., Verschuuren, J.J.G.M., Niks, E.H., Reinders, M.J.T., Kan, H.E., Lelieveldt, B.P.F., 2017. Timing and localization of human dystrophin isoform expression provide insights into the cognitive phenotype of Duchenne muscular dystrophy. *Sci. Rep.* 7, 1–12. <https://doi.org/10.1038/s41598-017-12981-5>
- Dostie, J., Dreyfuss, G., 2002. Translation Is Required to Remove Y14 from mRNAs in the Cytoplasm. *Curr. Biol.* 12, 1060–1067. [https://doi.org/10.1016/S0960-9822\(02\)00902-8](https://doi.org/10.1016/S0960-9822(02)00902-8)
- Dowling, P., Gargan, S., Murphy, S., Zweyer, M., Sabir, H., Swandulla, D., Ohlendieck, K., 2021. The Dystrophin Node as Integrator of Cytoskeletal Organization, Lateral Force Transmission, Fiber Stability and Cellular Signaling in Skeletal Muscle. *Proteomes* 9, 9. <https://doi.org/10.3390/proteomes9010009>
- Dreumont, N., Hardy, S., Behm-Ansmant, I., Kister, L., Branlant, C., Stévenin, J., Bourgeois, C.F., 2010. Antagonistic factors control the unproductive splicing of SC35 terminal intron. *Nucleic Acids Res.* 38, 1353–1366. <https://doi.org/10.1093/nar/gkp1086>
- D'Souza, V.N., Man, N. t., Morris, G.E., Karges, W., Pillers, D.-A.M., Ray, P.N., 1995. A novel dystrophin isoform is required for normal retinal electrophysiology. *Hum. Mol. Genet.* 4, 837–842. <https://doi.org/10.1093/hmg/4.5.837>
- Duan, D., Goemans, N., Takeda, S., Mercuri, E., Aartsma-Rus, A., 2021. Duchenne muscular dystrophy. *Nat. Rev. Dis. Primer* 7, 13. <https://doi.org/10.1038/s41572-021-00248-3>
- Duff, M.O., Olson, S., Wei, X., Garrett, S.C., Osman, A., Bolisetty, M., Plocik, A., Celniker, S.E., Graveley, B.R., 2015. Genome-wide identification of zero nucleotide recursive splicing in *Drosophila*. *Nature* 521, 376–379. <https://doi.org/10.1038/nature14475>
- Dumont, N.A., Wang, Y.X., Rudnicki, M.A., 2015. Intrinsic and extrinsic mechanisms regulating satellite cell function. *Development* 142, 1572–1581. <https://doi.org/10.1242/dev.114223>
- Durbeej, M., Jung, D., Hjalt, T., Campbell, K.P., Ekblom, P., 1997. Transient Expression of Dp140, a Product of the Duchenne Muscular Dystrophy Locus, during Kidney Tubulogenesis. *Dev. Biol.* 181, 156–167. <https://doi.org/10.1006/dbio.1996.8430>
- Dvinge, H., 2018. Regulation of alternative mRNA splicing: old players and new perspectives. *FEBS Lett.* 592, 2987–3006. <https://doi.org/10.1002/1873-3468.13119>
- El Haddad, M., Notarnicola, C., Evano, B., El Khatib, N., Blaquièrre, M., Bonnieu, A., Tajbakhsh, S., Hugon, G., Vernus, B., Mercier, J., Carnac, G., 2017. Retinoic acid maintains human skeletal muscle progenitor cells in an immature state. *Cell. Mol. Life Sci.* 74, 1923–1936. <https://doi.org/10.1007/s00018-016-2445-1>
- Ellis, J.D., Barrios-Rodiles, M., Çolak, R., Irimia, M., Kim, T., Calarco, J.A., Wang, X., Pan, Q., O'Hanlon, D., Kim, P.M., Wrana, J.L., Blencowe, B.J., 2012. Tissue-Specific Alternative Splicing Remodels Protein-Protein Interaction Networks. *Mol. Cell* 46, 884–892. <https://doi.org/10.1016/j.molcel.2012.05.037>
- Ervasti, J.M., 2013. Structure and Function of the Dystrophin-Glycoprotein Complex, Madame Curie Bioscience Database [Internet]. Landes Bioscience.
- Ervasti, J.M., Sonnemann, K.J., 2008. Biology of the striated muscle dystrophin-glycoprotein complex. *Int. Rev. Cytol.* 265, 191–225. [https://doi.org/10.1016/S0074-7696\(07\)65005-0](https://doi.org/10.1016/S0074-7696(07)65005-0)
- Faber, G.P., Nadav-Eliyahu, S., Shav-Tal, Y., 2022. Nuclear speckles – a driving force in gene expression. *J. Cell Sci.* 135, jcs259594. <https://doi.org/10.1242/jcs.259594>
- Farea, M., Rani, A.Q.M., Maeta, K., Nishio, H., Matsuo, M., 2020. Dystrophin Dp71ab is monoclonally expressed in human satellite cells and enhances proliferation of myoblast cells. *Sci. Rep.* 10, 17123. <https://doi.org/10.1038/s41598-020-74157-y>
- Feener, C.A., Koenig, M., Kunkel, L.M., 1989. Alternative splicing of human dystrophin mRNA generates isoforms at the carboxy terminus. *Nature* 338, 509–511. <https://doi.org/10.1038/338509a0>
- Feng, Q., Jagannathan, S., Bradley, R.K., 2017a. The RNA Surveillance Factor UPF1 Represses Myogenesis via Its E3 Ubiquitin Ligase Activity. *Mol. Cell* 67, 239-251.e6. <https://doi.org/10.1016/j.molcel.2017.05.034>
- Feng, Q., Jagannathan, S., Bradley, R.K., 2017b. The RNA Surveillance Factor UPF1 Represses Myogenesis via Its E3 Ubiquitin Ligase Activity. *Mol. Cell* 67, 239-251.e6. <https://doi.org/10.1016/j.molcel.2017.05.034>
- Feng, Q., Snider, L., Jagannathan, S., Tawil, R., van der Maarel, S.M., Tapscott, S.J., Bradley, R.K., 2015. A feedback

loop between nonsense-mediated decay and the retrogene DUX4 in facioscapulohumeral muscular dystrophy. *eLife* 4, e04996. <https://doi.org/10.7554/eLife.04996>

Fernández-Nogales, M., Cabrera, J.R., Santos-Galindo, M., Hoozemans, J.J.M., Ferrer, I., Rozemuller, A.J.M., Hernández, F., Avila, J., Lucas, J.J., 2014. Huntington's disease is a four-repeat tauopathy with tau nuclear rods. *Nat. Med.* 20, 881–885. <https://doi.org/10.1038/nm.3617>

Ferraiuolo, M.A., Lee, C.-S., Ler, L.W., Hsu, J.L., Costa-Mattioli, M., Luo, M.-J., Reed, R., Sonenberg, N., 2004. A nuclear translation-like factor eIF4AIII is recruited to the mRNA during splicing and functions in nonsense-mediated decay. *Proc. Natl. Acad. Sci. U. S. A.* 101, 4118–4123. <https://doi.org/10.1073/pnas.0400933101>

Fica, S.M., 2020. Cryo-EM snapshots of the human spliceosome reveal structural adaptations for splicing regulation. *Curr. Opin. Struct. Biol.* 65, 139–148. <https://doi.org/10.1016/j.sbi.2020.06.018>

Fica, S.M., Oubridge, C., Wilkinson, M.E., Newman, A.J., Nagai, K., 2019. A human post-catalytic spliceosome structure reveals essential roles of metazoan factors for exon ligation. *Science* 363, 710–714. <https://doi.org/10.1126/science.aaw5569>

Fiszbein, A., Krick, K.S., Begg, B.E., Burge, C.B., 2019. Exon-mediated activation of transcription starts. *Cell* 179, 1551–1565.e17. <https://doi.org/10.1016/j.cell.2019.11.002>

Flanigan, K.M., Dunn, D.M., Niederhausern, A. von, Soltanzadeh, P., Howard, M.T., Sampson, J.B., Swoboda, K.J., Bromberg, M.B., Mendell, J.R., Taylor, L.E., Anderson, C.B., Pestronk, A., Florence, J.M., Connolly, A.M., Mathews, K.D., Wong, B., Finkel, R.S., Bonnemann, C.G., Day, J.W., McDonald, C., Weiss, R.B., 2011. Nonsense mutation-associated Becker muscular dystrophy: interplay between exon definition and splicing regulatory elements within the DMD gene. *Hum. Mutat.* 32, 299–308. <https://doi.org/10.1002/humu.21426>

Forcina, L., Pelosi, L., Miano, C., Musarò, A., 2017. Insights into the Pathogenic Secondary Symptoms Caused by the Primary Loss of Dystrophin. *J. Funct. Morphol. Kinesiol.* 2, 44. <https://doi.org/10.3390/jfmk2040044>

Fortunato, F., Farnè, M., Ferlini, A., 2021. The DMD gene and therapeutic approaches to restore dystrophin. *Neuromuscul. Disord.* 31, 1013–1020. <https://doi.org/10.1016/j.nmd.2021.08.004>

Fratte, C., Dalgleish, R., Allen, S.K., Santos, R., Abbs, S., Tuffery-Giraud, S., Ferlini, A., 2020. EMQN best practice guidelines for genetic testing in dystrophinopathies. *Eur. J. Hum. Genet.* 28, 1141–1159. <https://doi.org/10.1038/s41431-020-0643-7>

Fribourg, S., Gatfield, D., Izaurralde, E., Conti, E., 2003. A novel mode of RBD-protein recognition in the Y14–Mago complex. *Nat. Struct. Mol. Biol.* 10, 433–439. <https://doi.org/10.1038/nsb926>

Frontera, W.R., Ochala, J., 2015. Skeletal muscle: a brief review of structure and function. *Calcif. Tissue Int.* 96, 183–195. <https://doi.org/10.1007/s00223-014-9915-y>

Fu, X.-D., Ares, M., 2014. Context-dependent control of alternative splicing by RNA-binding proteins. *Nat. Rev. Genet.* 15, 689–701. <https://doi.org/10.1038/nrg3778>

Fuchs, G., Hollander, D., Voichek, Y., Ast, G., Oren, M., 2014. Cotranscriptional histone H2B monoubiquitylation is tightly coupled with RNA polymerase II elongation rate. *Genome Res.* 24, 1572–1583. <https://doi.org/10.1101/gr.176487.114>

Fukumura, K., Inoue, K., Mayeda, A., 2018. Splicing activator RNPS1 suppresses errors in pre-mRNA splicing: A key factor for mRNA quality control. *Biochem. Biophys. Res. Commun.* 496, 921–926. <https://doi.org/10.1016/j.bbrc.2018.01.120>

Fukumura, K., Wakabayashi, S., Kataoka, N., Sakamoto, H., Suzuki, Y., Nakai, K., Mayeda, A., Inoue, K., 2016. The Exon Junction Complex Controls the Efficient and Faithful Splicing of a Subset of Transcripts Involved in Mitotic Cell-Cycle Progression. *Int. J. Mol. Sci.* 17, 1153. <https://doi.org/10.3390/ijms17081153>

Gangras, P., Gallagher, T.L., Parthun, M.A., Yi, Z., Patton, R.D., Tietz, K.T., Deans, N.C., Bundschuh, R., Amacher, S.L., Singh, G., 2020. Zebrafish *rbm8a* and *magoh* mutants reveal EJC developmental functions and new 3'UTR intron-containing NMD targets. *PLoS Genet.* 16, e1008830. <https://doi.org/10.1371/journal.pgen.1008830>

Gao, L., Wang, J., Wang, Y., Andreadis, A., 2007. SR protein 9G8 modulates splicing of tau exon 10 via its proximal downstream intron, a clustering region for frontotemporal dementia mutations. *Mol. Cell. Neurosci.* 34, 48–58. <https://doi.org/10.1016/j.mcn.2006.10.004>

García-Cruz, C., Aragón, J., Lourdel, S., Annan, A., Roger, J.E., Montanez, C., Vaillend, C., 2022. Tissue and cell-specific whole-transcriptome meta-analysis from brain and retina reveals differential expression of dystrophin complexes and new

dystrophin spliced isoforms. *Hum. Mol. Genet.* ddac236. <https://doi.org/10.1093/hmg/ddac236>

García-Cruz, C., Merino-Jiménez, C., Ceja, V., Aragón, J., Siqueiros-Márquez, L., Reyes-Grajeda, J.P., Montañez, C., 2019. The dystrophin isoform Dp71e $\Delta$ 71 is involved in neurite outgrowth and neuronal differentiation of PC12 cells. *J. Proteomics* 191, 80–87. <https://doi.org/10.1016/j.jprot.2018.03.027>

García-Rodríguez, R., Hiller, M., Jiménez-Gracia, L., van der Pal, Z., Balog, J., Adamzek, K., Aartsma-Rus, A., Spitali, P., 2020. Premature termination codons in the DMD gene cause reduced local mRNA synthesis. *Proc. Natl. Acad. Sci. U. S. A.* 117, 16456–16464. <https://doi.org/10.1073/pnas.1910456117>

Garrido-Martín, D., Palumbo, E., Guigó, R., Breschi, A., 2018. ggsashimi: Sashimi plot revised for browser- and annotation-independent splicing visualization. *PLOS Comput. Biol.* 14, e1006360. <https://doi.org/10.1371/journal.pcbi.1006360>

Gazzoli, I., Pulyakhina, I., Verwey, N.E., Ariyurek, Y., Laros, J.F.J., 't Hoen, P.A.C., Aartsma-Rus, A., 2016. Non-sequential and multi-step splicing of the dystrophin transcript. *RNA Biol.* 13, 290–305. <https://doi.org/10.1080/15476286.2015.1125074>

Ge, Z., Quek, B.L., Beemon, K.L., Hogg, J.R., 2016. Polypyrimidine tract binding protein 1 protects mRNAs from recognition by the nonsense-mediated mRNA decay pathway. *eLife* 5, e11155. <https://doi.org/10.7554/eLife.11155>

Gehring, N.H., Kunz, J.B., Neu-Yilik, G., Breit, S., Viegas, M.H., Hentze, M.W., 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. <https://doi.org/10.1016/j.molcel.2005.08.012>

Gehring, N.H., Lamprinaki, S., Hentze, M.W., Kulozik, A.E., 2009a. The Hierarchy of Exon-Junction Complex Assembly by the Spliceosome Explains Key Features of Mammalian Nonsense-Mediated mRNA Decay. *PLOS Biol.* 7, e1000120. <https://doi.org/10.1371/journal.pbio.1000120>

Gehring, N.H., Lamprinaki, S., Kulozik, A.E., Hentze, M.W., 2009b. Disassembly of Exon Junction Complexes by PYM. *Cell* 137, 536–548. <https://doi.org/10.1016/j.cell.2009.02.042>

Gehring, N.H., Roignant, J.-Y., 2021. Anything but Ordinary – Emerging Splicing Mechanisms in Eukaryotic Gene Regulation. *Trends Genet.* 37, 355–372. <https://doi.org/10.1016/j.tig.2020.10.008>

Gelfman, S., Cohen, N., Yearim, A., Ast, G., 2013. DNA-methylation effect on cotranscriptional splicing is dependent on GC architecture of the exon-intron structure. *Genome Res.* 23, 789–799. <https://doi.org/10.1101/gr.143503.112>

Gerbracht, J.V., Boehm, V., Britto-Borges, T., Kallabis, S., Wiederstein, J.L., Ciriello, S., Aschemeier, D.U., Krüger, M., Frese, C.K., Altmüller, J., Dieterich, C., Gehring, N.H., 2020. CASC3 promotes transcriptome-wide activation of nonsense-mediated decay by the exon junction complex. *Nucleic Acids Res.* 48, 8626–8644. <https://doi.org/10.1093/nar/gkaa564>

Gerbracht, J.V., Gehring, N.H., 2018. The exon junction complex: structural insights into a faithful companion of mammalian mRNPs. *Biochem. Soc. Trans.* 46, 153–161. <https://doi.org/10.1042/BST20170059>

Giono, L.E., Kornblihtt, A.R., 2020. Linking transcription, RNA polymerase II elongation and alternative splicing. *Biochem. J.* 477, 3091–3104. <https://doi.org/10.1042/BCJ20200475>

Giudice, J., Xia, Z., Wang, E.T., Scavuzzo, M.A., Ward, A.J., Kalsotra, A., Wang, W., Wehrens, X.H.T., Burge, C.B., Li, W., Cooper, T.A., 2014. Alternative splicing regulates vesicular trafficking genes in cardiomyocytes during postnatal heart development. *Nat. Commun.* 5, 3603. <https://doi.org/10.1038/ncomms4603>

Gonatopoulos-Pournatzis, T., Wu, M., Braunschweig, U., Roth, J., Han, H., Best, A.J., Raj, B., Aregger, M., O'Hanlon, D., Ellis, J.D., Calarco, J.A., Moffat, J., Gingras, A.-C., Blencowe, B.J., 2018. Genome-wide CRISPR-Cas9 Interrogation of Splicing Networks Reveals a Mechanism for Recognition of Autism-Misregulated Neuronal Microexons. *Mol. Cell* 72, 510–524.e12. <https://doi.org/10.1016/j.molcel.2018.10.008>

Gong, C., Kim, Y.K., Woeller, C.F., Tang, Y., Maquat, L.E., 2009. SMD and NMD are competitive pathways that contribute to myogenesis: effects on PAX3 and myogenin mRNAs. *Genes Dev.* 23, 54–66. <https://doi.org/10.1101/gad.1717309>

Gong, P., Zhao, M., He, C., 2014. Slow Co-Evolution of the MAGO and Y14 Protein Families Is Required for the Maintenance of Their Obligate Heterodimerization Mode. *PLoS ONE* 9. <https://doi.org/10.1371/journal.pone.0084842>

González, E., Montañez, C., Ray, P.N., Howard, P.L., García-Sierra, F., Mornet, D., Cisneros, B., 2000. Alternative splicing regulates the nuclear or cytoplasmic localization of dystrophin Dp71. *FEBS Lett.* 482, 209–214. [https://doi.org/10.1016/S0014-5793\(00\)02044-5](https://doi.org/10.1016/S0014-5793(00)02044-5)

Gorecki, D., Geng, Y., Thomas, K., Hunt, S.P., Barnard, E.A., Barnard, P.J., 1991. Expression of the dystrophin gene in

mouse and rat brain. *Neuroreport* 2, 773–776. <https://doi.org/10.1097/00001756-199112000-00011>

Goswami, M.V., Tawalbeh, S.M., Canessa, E.H., Hathout, Y., n.d. Temporal Proteomic Profiling During Differentiation of Normal and Dystrophin-Deficient Human Muscle Cells 18.

Gromadzka, A.M., Steckelberg, A.-L., Singh, K.K., Hofmann, K., Gehring, N.H., 2016. A short conserved motif in ALYREF directs cap- and EJC-dependent assembly of export complexes on spliced mRNAs. *Nucleic Acids Res.* 44, 2348–2361. <https://doi.org/10.1093/nar/gkw009>

Gruss, O.J., Meduri, R., Schilling, M., Fischer, U., 2017. UsnRNP biogenesis: mechanisms and regulation. *Chromosoma* 126, 577–593. <https://doi.org/10.1007/s00412-017-0637-6>

Gueroussov, S., Weatheritt, R.J., O'Hanlon, D., Lin, Z.-Y., Narula, A., Gingras, A.-C., Blencowe, B.J., 2017. Regulatory Expansion in Mammals of Multivalent hnRNP Assemblies that Globally Control Alternative Splicing. *Cell* 170, 324-339.e23. <https://doi.org/10.1016/j.cell.2017.06.037>

Guo, R., Zheng, L., Park, J.W., Lv, R., Chen, H., Jiao, F., Xu, W., Mu, S., Wen, H., Qiu, J., Wang, Z., Yang, P., Wu, F., Hui, J., Fu, X., Shi, X., Shi, Y.G., Xing, Y., Lan, F., Shi, Y., 2014. BS69/ZMYND11 reads and connects histone H3.3 lysine 36 trimethylation-decorated chromatin to regulated pre-mRNA processing. *Mol. Cell* 56, 298–310. <https://doi.org/10.1016/j.molcel.2014.08.022>

Guo, X., Wang, X.-F., 2009. Signaling cross-talk between TGF- $\beta$ /BMP and other pathways. *Cell Res.* 19, 71–88. <https://doi.org/10.1038/cr.2008.302>

Haberman, N., Huppertz, I., Attig, J., König, J., Wang, Z., Hauer, C., Hentze, M.W., Kulozik, A.E., Le Hir, H., Curk, T., Sibley, C.R., Zarnack, K., Ule, J., 2017. Insights into the design and interpretation of iCLIP experiments. *Genome Biol.* 18, 7. <https://doi.org/10.1186/s13059-016-1130-x>

Hall, M.P., Nagel, R.J., Fagg, W.S., Shiue, L., Cline, M.S., Perriman, R.J., Donohue, J.P., Ares, M., 2013. Quaking and PTB control overlapping splicing regulatory networks during muscle cell differentiation. *RNA* 19, 627–638. <https://doi.org/10.1261/rna.038422.113>

Hamer, D.H., Leder, P., 1979. Splicing and the formation of stable RNA. *Cell* 18, 1299–1302. [https://doi.org/10.1016/0092-8674\(79\)90240-X](https://doi.org/10.1016/0092-8674(79)90240-X)

Han, H., Braunschweig, U., Gonatopoulos-Pournatzis, T., Weatheritt, R.J., Hirsch, C.L., Ha, K.C.H., Radovani, E., Nabeel-Shah, S., Sterne-Weiler, T., Wang, J., O'Hanlon, D., Pan, Q., Ray, D., Zheng, H., Vizeacoumar, F., Datti, A., Magomedova, L., Cummins, C.L., Hughes, T.R., Greenblatt, J.F., Wrana, J.L., Moffat, J., Blencowe, B.J., 2017. Multilayered Control of Alternative Splicing Regulatory Networks by Transcription Factors. *Mol. Cell* 65, 539-553.e7. <https://doi.org/10.1016/j.molcel.2017.01.011>

Han, J., Ding, J.-H., Byeon, C.W., Kim, J.H., Hertel, K.J., Jeong, S., Fu, X.-D., 2011. SR Proteins Induce Alternative Exon Skipping through Their Activities on the Flanking Constitutive Exons. *Mol. Cell. Biol.* 31, 793–802. <https://doi.org/10.1128/MCB.01117-10>

Haremaki, T., Weinstein, D.C., 2012. Eif4a3 is required for accurate splicing of the *Xenopus laevis* ryanodine receptor pre-mRNA. *Dev. Biol.* 372, 103–110. <https://doi.org/10.1016/j.ydbio.2012.08.013>

Haselbach, D., Komarov, I., Agafonov, D.E., Hartmuth, K., Graf, B., Dybkov, O., Urlaub, H., Kastner, B., Lührmann, R., Stark, H., 2018. Structure and Conformational Dynamics of the Human Spliceosomal Bact Complex. *Cell* 172, 454-464.e11. <https://doi.org/10.1016/j.cell.2018.01.010>

Hauer, C., Sieber, J., Schwarzl, T., Hollerer, I., Curk, T., Alleaume, A.-M., Hentze, M.W., Kulozik, A.E., 2016. Exon Junction Complexes Show a Distributional Bias toward Alternatively Spliced mRNAs and against mRNAs Coding for Ribosomal Proteins. *Cell Rep.* 16, 1588–1603. <https://doi.org/10.1016/j.celrep.2016.06.096>

Hayashi, R., Handler, D., Ish-Horowicz, D., Brennecke, J., 2014. The exon junction complex is required for definition and excision of neighboring introns in *Drosophila*. *Genes Dev.* 28, 1772–1785. <https://doi.org/10.1101/gad.245738.114>

Hazra, A., Gogtay, N., 2016. Biostatistics Series Module 6: Correlation and Linear Regression. *Indian J. Dermatol.* 61, 593–601. <https://doi.org/10.4103/0019-5154.193662>

Heath, C.G., Viphakone, N., Wilson, S.A., 2016. The role of TREX in gene expression and disease. *Biochem. J.* 473, 2911–2935. <https://doi.org/10.1042/BCJ20160010>

Helderman-van den Enden, A., De Jong, R., Den Dunnen, J., Houwing-Duistermaat, J., Kneppers, A., Ginjaar, H., Breuning, M., Bakker, E., 2009. Recurrence risk due to germ line mosaicism: Duchenne and Becker muscular dystrophy.

- Clin. Genet. 75, 465–472. <https://doi.org/10.1111/j.1399-0004.2009.01173.x>
- Herrera-Salazar, A., García-Villegas, R., Aragón, J., Sánchez-Trujillo, A., Ceja, V., Martínez-Herrera, A., Merino-Jiménez, C., Montañez, C., 2016. Overexpression of mutant dystrophin Dp71[INCREMENT]78–79 stimulates cell proliferation. *NeuroReport* 27, 6–11. <https://doi.org/10.1097/WNR.0000000000000475>
- Higginson, J.R., Thompson, O., Winder, S.J., 2008. Targeting of dystroglycan to the cleavage furrow and midbody in cytokinesis. *Int. J. Biochem. Cell Biol.* 40, 892–900. <https://doi.org/10.1016/j.biocel.2007.10.019>
- Hildyard, J.C.W., Crawford, A.H., Rawson, F., Riddell, D.O., Harron, R.C.M., Piercy, R.J., 2020. Single-transcript multiplex in situ hybridisation reveals unique patterns of dystrophin isoform expression in the developing mammalian embryo. *Wellcome Open Res.* 5, 76. <https://doi.org/10.12688/wellcomeopenres.15762.2>
- Hinkle, E.R., Wiedner, H.J., Black, A.J., Giudice, J., 2019. RNA processing in skeletal muscle biology and disease. *Transcription* 10, 1–20. <https://doi.org/10.1080/21541264.2018.1558677>
- Hir, H.L., Saulière, J., Wang, Z., 2016. The exon junction complex as a node of post-transcriptional networks. *Nat. Rev. Mol. Cell Biol.* 17, 41–54. <https://doi.org/10.1038/nrm.2015.7>
- Hnia, K., Zouiten, D., Cantel, S., Chazalotte, D., Hugon, G., Fehrentz, J.-A., Masmoudi, A., Diment, A., Bramham, J., Mornet, D., Winder, S.J., 2007. ZZ domain of dystrophin and utrophin: topology and mapping of a  $\beta$ -dystroglycan interaction site. *Biochem. J.* 401, 667–677. <https://doi.org/10.1042/BJ20061051>
- Hnilicová, J., Hozefi, S., Dušková, E., Icha, J., Tománková, T., Staněk, D., 2011. Histone Deacetylase Activity Modulates Alternative Splicing. *PLOS ONE* 6, e16727. <https://doi.org/10.1371/journal.pone.0016727>
- Ho, C.H., Paolantoni, C., Bawankar, P., Tang, Z., Brown, S., Roignant, J.-Y., Treisman, J.E., 2022. An exon junction complex-independent function of Barentsz in neuromuscular synapse growth. *EMBO Rep.* 23, e53231. <https://doi.org/10.15252/embr.202153231>
- Hocq, R., Paternina, J., Alasseur, Q., Genovesio, A., Le Hir, H., 2018. Monitored eCLIP: high accuracy mapping of RNA-protein interactions. *Nucleic Acids Res.* 46, 11553–11565. <https://doi.org/10.1093/nar/gky858>
- Hoek, T.A., Khuperkar, D., Lindeboom, R.G.H., Sonneveld, S., Verhagen, B.M.P., Boersma, S., Vermeulen, M., Tanenbaum, M.E., 2019. Single-Molecule Imaging Uncovers Rules Governing Nonsense-Mediated mRNA Decay. *Mol. Cell* 75, 324–339.e11. <https://doi.org/10.1016/j.molcel.2019.05.008>
- Hoffman, E.P., Brown, R.H., Kunkel, L.M., 1987. Dystrophin: the protein product of the Duchenne muscular dystrophy locus. *Cell* 51, 919–928. [https://doi.org/10.1016/0092-8674\(87\)90579-4](https://doi.org/10.1016/0092-8674(87)90579-4)
- Hogg, R., McGrail, J.C., O’Keefe, R.T., 2010. The function of the NineTeen Complex (NTC) in regulating spliceosome conformations and fidelity during pre-mRNA splicing. *Biochem. Soc. Trans.* 38, 1110–1115. <https://doi.org/10.1042/BST0381110>
- Holder, E., Maeda, M., Bies, R.D., 1996. Expression and regulation of the dystrophin Purkinje promoter in human skeletal muscle, heart, and brain. *Hum. Genet.* 97, 232–239. <https://doi.org/10.1007/BF02265272>
- Howard, P.L., Dally, G.Y., Ditta, S.D., Austin, R.C., Worton, R.G., Klamut, H.J., Ray, P.N., 1999. Dystrophin isoforms Dp71 and Dp427 have distinct roles in myogenic cells. *Muscle Nerve* 22, 16–27. [https://doi.org/10.1002/\(SICI\)1097-4598\(199901\)22:1<16::AID-MUS5>3.0.CO;2-R](https://doi.org/10.1002/(SICI)1097-4598(199901)22:1<16::AID-MUS5>3.0.CO;2-R)
- Huang, H., Zhang, J., Harvey, S.E., Hu, X., Cheng, C., 2017. RNA G-quadruplex secondary structure promotes alternative splicing via the RNA-binding protein hnRNPF. *Genes Dev.* 31, 2296–2309. <https://doi.org/10.1101/gad.305862.117>
- Huang, L., Low, A., Damle, S.S., Keenan, M.M., Kuntz, S., Murray, S.F., Monia, B.P., Guo, S., 2018. Antisense suppression of the nonsense mediated decay factor Upf3b as a potential treatment for diseases caused by nonsense mutations. *Genome Biol.* 19, 4. <https://doi.org/10.1186/s13059-017-1386-9>
- Huang, X., Poy, F., Zhang, R., Joachimiak, A., Sudol, M., Eck, M.J., 2000. Structure of a WW domain containing fragment of dystrophin in complex with  $\beta$ -dystroglycan. *Nat. Struct. Biol.* 7, 634–638. <https://doi.org/10.1038/77923>
- Hughes, M.I., Hicks, E.M., Nevin, N.C., Patterson, V.H., 1996. The prevalence of inherited neuromuscular disease in Northern Ireland. *Neuromuscul. Disord.* NMD 6, 69–73. [https://doi.org/10.1016/0960-8966\(94\)00017-4](https://doi.org/10.1016/0960-8966(94)00017-4)
- Ideue, T., Sasaki, Y.T.F., Hagiwara, M., Hirose, T., 2007. Introns play an essential role in splicing-dependent formation of the exon junction complex. *Genes Dev.* 21, 1993–1998. <https://doi.org/10.1101/gad.1557907>



- Ip, J.Y., Schmidt, D., Pan, Q., Ramani, A.K., Fraser, A.G., Odom, D.T., Blencowe, B.J., 2011. Global impact of RNA polymerase II elongation inhibition on alternative splicing regulation. *Genome Res.* 21, 390–401. <https://doi.org/10.1101/gr.111070.110>
- Ivanov, A., Mikhailova, T., Eliseev, B., Yeramala, L., Sokolova, E., Susorov, D., Shuvalov, A., Schaffitzel, C., Alkalaeva, E., 2016. PABP enhances release factor recruitment and stop codon recognition during translation termination. *Nucleic Acids Res.* 44, 7766–7776. <https://doi.org/10.1093/nar/gkw635>
- Ivanov, P.V., Gehring, N.H., Kunz, J.B., Hentze, M.W., Kulozik, A.E., 2008. Interactions between UPF1, eRFs, PABP and the exon junction complex suggest an integrated model for mammalian NMD pathways. *EMBO J.* 27, 736–747. <https://doi.org/10.1038/emboj.2008.17>
- Jolly, L.A., Homan, C.C., Jacob, R., Barry, S., Gecz, J., 2013. The UPF3B gene, implicated in intellectual disability, autism, ADHD and childhood onset schizophrenia regulates neural progenitor cell behaviour and neuronal outgrowth. *Hum. Mol. Genet.* 22, 4673–4687. <https://doi.org/10.1093/hmg/ddt315>
- Joo, J.-H., Correia, G.P., Li, J.-L., Lopez, M.-C., Baker, H.V., Sugrue, S.P., 2013. Transcriptomic Analysis of PNN- and ESRP1-Regulated Alternative Pre-mRNA Splicing in Human Corneal Epithelial Cells. *Invest. Ophthalmol. Vis. Sci.* 54, 697–707. <https://doi.org/10.1167/iovs.12-10695>
- Jordan, P., Gonçalves, V., Fernandes, S., Marques, T., Pereira, M., Gama-Carvalho, M., 2019. Networks of mRNA Processing and Alternative Splicing Regulation in Health and Disease, in: Romão, L. (Ed.), *The MRNA Metabolism in Human Disease, Advances in Experimental Medicine and Biology*. Springer International Publishing, Cham, pp. 1–27. [https://doi.org/10.1007/978-3-030-19966-1\\_1](https://doi.org/10.1007/978-3-030-19966-1_1)
- Joseph, B., Kondo, S., Lai, E.C., 2018. Short cryptic exons mediate recursive splicing in *Drosophila*. *Nat. Struct. Mol. Biol.* 25, 365–371. <https://doi.org/10.1038/s41594-018-0052-6>
- Joseph, B., Lai, E.C., 2021. The Exon Junction Complex and intron removal prevent re-splicing of mRNA. *PLOS Genet.* 17, e1009563. <https://doi.org/10.1371/journal.pgen.1009563>
- Jurica, M.S., Licklider, L.J., Gygi, S.R., Grigorieff, N., Moore, M.J., 2002. Purification and characterization of native spliceosomes suitable for three-dimensional structural analysis. *RNA* 8, 426–439.
- Kanellis, D.C., Espinoza, J.A., Zisi, A., Sakkas, E., Bartkova, J., Katsori, A.-M., Boström, J., Dyrskjöt, L., Broholm, H., Altun, M., Elsässer, S.J., Lindström, M.S., Bartek, J., 2021. The exon-junction complex helicase eIF4A3 controls cell fate via coordinated regulation of ribosome biogenesis and translational output. *Sci. Adv.* 7, eabf7561. <https://doi.org/10.1126/sciadv.abf7561>
- Kano, S., Nishida, K., Kurebe, H., Nishiyama, C., Kita, K., Akaike, Y., Kajita, K., Kurokawa, K., Masuda, K., Kuwano, Y., Tanahashi, T., Rokutan, K., 2014. Oxidative stress-inducible truncated serine/arginine-rich splicing factor 3 regulates interleukin-8 production in human colon cancer cells. *Am. J. Physiol.-Cell Physiol.* 306, C250–C262. <https://doi.org/10.1152/ajpcell.00091.2013>
- Karlič, R., Chung, H.-R., Lasserre, J., Vlahoviček, K., Vingron, M., 2010. Histone modification levels are predictive for gene expression. *Proc. Natl. Acad. Sci.* 107, 2926–2931. <https://doi.org/10.1073/pnas.0909344107>
- Kashima, I., Jonas, S., Jayachandran, U., Buchwald, G., Conti, E., Lupas, A.N., Izaurralde, E., 2010. SMG6 interacts with the exon junction complex via two conserved EJC-binding motifs (EBMs) required for nonsense-mediated mRNA decay. *Genes Dev.* 24, 2440–2450. <https://doi.org/10.1101/gad.604610>
- Kashima, I., Yamashita, A., Izumi, N., Kataoka, N., Morishita, R., Hoshino, S., Ohno, M., Dreyfuss, G., Ohno, S., 2006. Binding of a novel SMG-1–Upf1–eRF1–eRF3 complex (SURF) to the exon junction complex triggers Upf1 phosphorylation and nonsense-mediated mRNA decay. *Genes Dev.* 20, 355–367. <https://doi.org/10.1101/gad.1389006>
- Kastner, B., Will, C.L., Stark, H., Lührmann, R., 2019. Structural Insights into Nuclear pre-mRNA Splicing in Higher Eukaryotes. *Cold Spring Harb. Perspect. Biol.* 11, a032417. <https://doi.org/10.1101/cshperspect.a032417>
- Kawaguchi, T., Niba, E.T.E., Rani, A.Q.M., Onishi, Y., Koizumi, M., Awano, H., Matsumoto, M., Nagai, M., Yoshida, S., Sakakibara, S., Maeda, N., Sato, O., Nishio, H., Matsuo, M., 2018. Detection of Dystrophin Dp71 in Human Skeletal Muscle Using an Automated Capillary Western Assay System. *Int. J. Mol. Sci.* 19. <https://doi.org/10.3390/ijms19061546>
- Kawano, T., Kataoka, N., Dreyfuss, G., Sakamoto, H., 2004. Ce-Y14 and MAG-1, components of the exon–exon junction complex, are required for embryogenesis and germline sexual switching in *Caenorhabditis elegans*. *Mech. Dev.* 121, 27–35. <https://doi.org/10.1016/j.mod.2003.11.003>
- Ke, S., Chasin, L.A., 2011. Context-dependent splicing regulation. *RNA Biol.* 8, 384–388.

<https://doi.org/10.4161/rna.8.3.14458>

Keren-Shaul, H., Lev-Maor, G., Ast, G., 2013. Pre-mRNA splicing is a determinant of nucleosome organization. *PLoS One* 8, e53506. <https://doi.org/10.1371/journal.pone.0053506>

Khan, D.H., Gonzalez, C., Cooper, C., Sun, J.-M., Chen, H.Y., Healy, S., Xu, W., Smith, K.T., Workman, J.L., Leygue, E., Davie, J.R., 2014. RNA-dependent dynamic histone acetylation regulates MCL1 alternative splicing. *Nucleic Acids Res.* 42, 1656–1670. <https://doi.org/10.1093/nar/gkt1134>

Khoury, G., Gruss, P., Dhar, R., Lai, C.-J., 1979. Processing and expression of early SV40 mRNA: a role for RNA conformation in splicing. *Cell* 18, 85–92. [https://doi.org/10.1016/0092-8674\(79\)90356-8](https://doi.org/10.1016/0092-8674(79)90356-8)

Kim, Y.K., Furic, L., DesGroseillers, L., Maquat, L.E., 2005. Mammalian Staufen1 Recruits Upf1 to Specific mRNA 3'UTRs so as to Elicit mRNA Decay. *Cell* 120, 195–208. <https://doi.org/10.1016/j.cell.2004.11.050>

Kishor, A., Fritz, S.E., Haque, N., Ge, Z., Tunc, I., Yang, W., Zhu, J., Hogg, J.R., 2020. Activation and inhibition of nonsense-mediated mRNA decay control the abundance of alternative polyadenylation products. *Nucleic Acids Res.* gkaa491. <https://doi.org/10.1093/nar/gkaa491>

Koenig, M., Hoffman, E.P., Bertelson, C.J., Monaco, A.P., Feener, C., Kunkel, L.M., 1987. Complete cloning of the duchenne muscular dystrophy (DMD) cDNA and preliminary genomic organization of the DMD gene in normal and affected individuals. *Cell* 50, 509–517. [https://doi.org/10.1016/0092-8674\(87\)90504-6](https://doi.org/10.1016/0092-8674(87)90504-6)

Koenig, M., Monaco, A.P., Kunkel, L.M., 1988. The complete sequence of dystrophin predicts a rod-shaped cytoskeletal protein. *Cell* 53, 219–228. [https://doi.org/10.1016/0092-8674\(88\)90383-2](https://doi.org/10.1016/0092-8674(88)90383-2)

Kornblihtt, A.R., Schor, I.E., Alló, M., Dujardin, G., Petrillo, E., Muñoz, M.J., 2013. Alternative splicing: a pivotal step between eukaryotic transcription and translation. *Nat. Rev. Mol. Cell Biol.* 14, 153–165. <https://doi.org/10.1038/nrm3525>

Kovalak, C., Donovan, S., Bicknell, A.A., Metkar, M., Moore, M.J., 2021. Deep sequencing of pre-translational mRNPs reveals hidden flux through evolutionarily conserved alternative splicing nonsense-mediated decay pathways. *Genome Biol.* 22, 132. <https://doi.org/10.1186/s13059-021-02309-y>

Kumari, A., Sedehizadeh, S., Brook, J.D., Kozłowski, P., Wojciechowska, M., 2021. Differential fates of introns in gene expression due to global alternative splicing. *Hum. Genet.* <https://doi.org/10.1007/s00439-021-02409-6>

Kwon, O.S., Mishra, R., Safieddine, A., Coleno, E., Alasseur, Q., Faucourt, M., Barbosa, I., Bertrand, E., Spassky, N., Le Hir, H., 2021. Exon junction complex dependent mRNA localization is linked to centrosome organization during ciliogenesis. *Nat. Commun.* 12, 1351. <https://doi.org/10.1038/s41467-021-21590-w>

Lau, C.-K., Diem, M.D., Dreyfuss, G., Van Duyne, G.D., 2003. Structure of the Y14-Magoh Core of the Exon Junction Complex. *Curr. Biol.* 13, 933–941. [https://doi.org/10.1016/S0960-9822\(03\)00328-2](https://doi.org/10.1016/S0960-9822(03)00328-2)

Laumonnier, F., Shoubridge, C., Antar, C., Nguyen, L.S., Van Esch, H., Kleefstra, T., Briault, S., Fryns, J.P., Hamel, B., Chelly, J., Ropers, H.H., Ronce, N., Blesson, S., Moraine, C., Géczy, J., Raynaud, M., 2010. Mutations of the UPF3B gene, which encodes a protein widely expressed in neurons, are associated with nonspecific mental retardation with or without autism. *Mol. Psychiatry* 15, 767–776. <https://doi.org/10.1038/mp.2009.14>

Le Hir, H., Gatfield, D., Izaurralde, E., 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. <https://doi.org/10.1093/emboj/20.17.4987>

Le Hir, H., Izaurralde, E., Maquat, L.E., Moore, M.J., 2000. The spliceosome deposits multiple proteins 20–24 nucleotides upstream of mRNA exon–exon junctions. *EMBO J.* 19, 6860–6869. <https://doi.org/10.1093/emboj/19.24.6860>

Le Hir, H., Nott, A., Moore, M.J., 2003. How introns influence and enhance eukaryotic gene expression. *Trends Biochem. Sci.* 28, 215–220. [https://doi.org/10.1016/S0968-0004\(03\)00052-5](https://doi.org/10.1016/S0968-0004(03)00052-5)

Le Rumeur, E., Winder, S.J., Hubert, J.-F., 2010. Dystrophin: more than just the sum of its parts. *Biochim. Biophys. Acta* 1804, 1713–1722. <https://doi.org/10.1016/j.bbapap.2010.05.001>

Leclair, N.K., Brugiolo, M., Urbanski, L., Lawson, S.C., Thakar, K., Yurieva, M., George, J., Hinson, J.T., Cheng, A., Graveley, B.R., Anczuków, O., 2020. POISON EXON SPLICING REGULATES A COORDINATED NETWORK OF SR PROTEIN EXPRESSION DURING DIFFERENTIATION AND TUMORIGENESIS. *Mol. Cell* 80, 648–665.e9. <https://doi.org/10.1016/j.molcel.2020.10.019>

Lederfein, D., Levy, Z., Augier, N., Mornet, D., Morris, G., Fuchs, O., Yaffe, D., Nudel, U., 1992. A 71-kilodalton protein is

- a major product of the Duchenne muscular dystrophy gene in brain and other nonmuscle tissues. *Proc. Natl. Acad. Sci.* 89, 5346–5350. <https://doi.org/10.1073/pnas.89.12.5346>
- Lee, Y., Rio, D.C., 2015. Mechanisms and Regulation of Alternative Pre-mRNA Splicing. *Annu. Rev. Biochem.* 84, 291–323. <https://doi.org/10.1146/annurev-biochem-060614-034316>
- Lee, Y.J., Wang, Q., Rio, D.C., 2018. Coordinate regulation of alternative pre-mRNA splicing events by the human RNA chaperone proteins hnRNPA1 and DDX5. *Genes Dev.* 32, 1060–1074. <https://doi.org/10.1101/gad.316034.118>
- Legardinier, S., Raguénès-Nicol, C., Tascon, C., Rocher, C., Hardy, S., Hubert, J.-F., Le Rumeur, E., 2009. Mapping of the Lipid-Binding and Stability Properties of the Central Rod Domain of Human Dystrophin. *J. Mol. Biol.* 389, 546–558. <https://doi.org/10.1016/j.jmb.2009.04.025>
- Lejeune, F., 2022. Nonsense-Mediated mRNA Decay, a Finely Regulated Mechanism. *Biomedicines* 10, 141. <https://doi.org/10.3390/biomedicines10010141>
- Lepper, C., Fan, C.-M., 2010. Inducible lineage tracing of Pax7-descendant cells reveals embryonic origin of adult satellite cells. *genesis* 48, 424–436. <https://doi.org/10.1002/dvg.20630>
- Leu, S., 2021. The role and regulation of Pnn in proliferative and non-dividing cells: From embryogenesis to pathogenesis. *Biochem. Pharmacol.* 192, 114672. <https://doi.org/10.1016/j.bcp.2021.114672>
- Leu, S., Lin, Y.-M., Wu, C.-H., Ouyang, P., 2012. Loss of Pnn expression results in mouse early embryonic lethality and cellular apoptosis through SRSF1-mediated alternative expression of Bcl-xS and ICAD. *J. Cell Sci.* 125, 3164–3172. <https://doi.org/10.1242/jcs.100859>
- Leung, A.K.W., Nagai, K., Li, J., 2011. Structure of the spliceosomal U4 snRNP core domain and its implication for snRNP biogenesis. *Nature* 473, 536–539. <https://doi.org/10.1038/nature09956>
- Leung, C.S., Douglass, S.M., Morselli, M., Obusan, M.B., Pavlyukov, M.S., Pellegrini, M., Johnson, T.L., 2019. H3K36 Methylation and the Chromodomain Protein Eaf3 Are Required for Proper Cotranscriptional Spliceosome Assembly. *Cell Rep.* 27, 3760–3769.e4. <https://doi.org/10.1016/j.celrep.2019.05.100>
- Leung, C.S., Johnson, T.L., 2018. The Exon Junction Complex: A Multitasking Guardian of the Transcriptome. *Mol. Cell* 72, 799–801. <https://doi.org/10.1016/j.molcel.2018.11.030>
- Li, S., Guo, W., Dewey, C.N., Greaser, M.L., 2013. Rbm20 regulates titin alternative splicing as a splicing repressor. *Nucleic Acids Res.* 41, 2659–2672. <https://doi.org/10.1093/nar/gks1362>
- Li, T., Shi, Y., Wang, P., Guachalla, L.M., Sun, B., Joerss, T., Chen, Y., Groth, M., Krueger, A., Platzer, M., Yang, Y., Rudolph, K.L., Wang, Z., 2015. Smg6/Est1 licenses embryonic stem cell differentiation via nonsense-mediated mRNA decay. *EMBO J.* 34, 1630–1647. <https://doi.org/10.15252/embj.201489947>
- Lidov, H.G., Byers, T.J., Watkins, S.C., Kunkel, L.M., 1990. Localization of dystrophin to postsynaptic regions of central nervous system cortical neurons. *Nature* 348, 725–728. <https://doi.org/10.1038/348725a0>
- Lidov, H.G., Kunkel, L.M., 1998. Dystrophin and Dp140 in the adult rodent kidney. *Lab. Invest. J. Tech. Methods Pathol.* 78, 1543–1551.
- Lidov, H.G., Selig, S., Kunkel, L.M., 1995. Dp140: a novel 140 kDa CNS transcript from the dystrophin locus. *Hum. Mol. Genet.* 4, 329–335. <https://doi.org/10.1093/hmg/4.3.329>
- Lidov, H.G.W., Kunkel, L.M., 1997. Dp140: Alternatively Spliced Isoforms in Brain and Kidney. *Genomics* 45, 132–139. <https://doi.org/10.1006/geno.1997.4905>
- Lin, S., Coutinho-Mansfield, G., Wang, D., Pandit, S., Fu, X.-D., 2008. The splicing factor SC35 has an active role in transcriptional elongation. *Nat. Struct. Mol. Biol.* 15, 819–826. <https://doi.org/10.1038/nsmb.1461>
- Lindeboom, R.G.H., Supek, F., Lehner, B., 2016. The rules and impact of nonsense-mediated mRNA decay in human cancers. *Nat. Genet.* 48, 1112–1118. <https://doi.org/10.1038/ng.3664>
- Linder, P., Jankowsky, E., 2011. From unwinding to clamping — the DEAD box RNA helicase family. *Nat. Rev. Mol. Cell Biol.* 12, 505–516. <https://doi.org/10.1038/nrm3154>
- Lindholm, M.E., Huss, M., Solnestam, B.W., Kjellqvist, S., Lundeberg, J., Sundberg, C.J., 2014. The human skeletal muscle transcriptome: sex differences, alternative splicing, and tissue homogeneity assessed with RNA sequencing. *FASEB J.* 28, 4571–4581. <https://doi.org/10.1096/fj.14-255000>

- Liu, Q., Fang, L., Wu, C., 2022. Alternative Splicing and Isoforms: From Mechanisms to Diseases. *Genes* 13, 401. <https://doi.org/10.3390/genes13030401>
- Llorian, M., 2011. Decoding muscle alternative splicing 8.
- Loh, B., Jonas, S., Izaurralde, E., 2013. The SMG5–SMG7 heterodimer directly recruits the CCR4–NOT deadenylase complex to mRNAs containing nonsense codons via interaction with POP2. *Genes Dev.* 27, 2125–2138. <https://doi.org/10.1101/gad.226951.113>
- Lou, C.-H., Dumdie, J., Goetz, A., Shum, E.Y., Brafman, D., Liao, X., Mora-Castilla, S., Ramaiah, M., Cook-Andersen, H., Laurent, L., Wilkinson, M.F., 2016. Nonsense-Mediated RNA Decay Influences Human Embryonic Stem Cell Fate. *Stem Cell Rep.* 6, 844–857. <https://doi.org/10.1016/j.stemcr.2016.05.008>
- Low, Y.-H., Asi, Y., Foti, S.C., Lashley, T., 2021. Heterogeneous Nuclear Ribonucleoproteins: Implications in Neurological Diseases. *Mol. Neurobiol.* 58, 631–646. <https://doi.org/10.1007/s12035-020-02137-4>
- Lu, C.-C., Lee, C.-C., Tseng, C.-T., Tarn, W.-Y., 2017. Y14 governs p53 expression and modulates DNA damage sensitivity. *Sci. Rep.* 7. <https://doi.org/10.1038/srep45558>
- Lu, W.-T., Wilczynska, A., Smith, E., Bushell, M., 2014. The diverse roles of the eIF4A family: you are the company you keep. *Biochem. Soc. Trans.* 42, 166–172. <https://doi.org/10.1042/BST20130161>
- Luco, R.F., Allo, M., Schor, I.E., Kornblihtt, A.R., Misteli, T., 2011. Epigenetics in Alternative Pre-mRNA Splicing. *Cell* 144, 16–26. <https://doi.org/10.1016/j.cell.2010.11.056>
- Luo, M.J., Reed, R., 1999. Splicing is required for rapid and efficient mRNA export in metazoans. *Proc. Natl. Acad. Sci. U. S. A.* 96, 14937–14942. <https://doi.org/10.1073/pnas.96.26.14937>
- Lykke-Andersen, S., Chen, Y., Ardal, B.R., Lilje, B., Waage, J., Sandelin, A., Jensen, T.H., 2014. Human nonsense-mediated RNA decay initiates widely by endonucleolysis and targets snoRNA host genes. *Genes Dev.* 28, 2498–2517. <https://doi.org/10.1101/gad.246538.114>
- Lykke-Andersen, S., Jensen, T.H., 2015. Nonsense-mediated mRNA decay: an intricate machinery that shapes transcriptomes. *Nat. Rev. Mol. Cell Biol.* 16, 665–677. <https://doi.org/10.1038/nrm4063>
- Mabin, J.W., Woodward, L.A., Patton, R.D., Yi, Z., Jia, M., Wysocki, V.H., Bundschuh, R., Singh, G., 2018. The Exon Junction Complex Undergoes a Compositional Switch that Alters mRNP Structure and Nonsense-Mediated mRNA Decay Activity. *Cell Rep.* 25, 2431-2446.e7. <https://doi.org/10.1016/j.celrep.2018.11.046>
- Macchi, P., Kroening, S., Palacios, I.M., Baldassa, S., Grunewald, B., Ambrosino, C., Goetze, B., Lupas, A., St. Johnston, D., Kiebler, M., 2003. Barentsz, a New Component of the Staufien-Containing Ribonucleoprotein Particles in Mammalian Cells, Interacts with Staufien in an RNA-Dependent Manner. *J. Neurosci.* 23, 5778–5788. <https://doi.org/10.1523/JNEUROSCI.23-13-05778.2003>
- Mahyob Rani, A.Q., Maeta, K., Kawaguchi, T., Awano, H., Nagai, M., Nishio, H., Matsuo, M., 2019. Schwann cell-specific Dp116 is expressed in glioblastoma cells, revealing two novel DMD gene splicing patterns. *Biochem. Biophys. Rep.* 20, 100703. <https://doi.org/10.1016/j.bbrep.2019.100703>
- Mailliot, J., Vivoli-Vega, M., Schaffitzel, C., 2022. No-nonsense: insights into the functional interplay of nonsense-mediated mRNA decay factors. *Biochem. J.* 479, 973–993. <https://doi.org/10.1042/BCJ20210556>
- Malone, C.D., Mestdagh, C., Akhtar, J., Kreim, N., Deinhard, P., Sachidanandam, R., Treisman, J., Roignant, J.-Y., 2014. The exon junction complex controls transposable element activity by ensuring faithful splicing of the piwi transcript. *Genes Dev.* 28, 1786–1799. <https://doi.org/10.1101/gad.245829.114>
- Mamchaoui, K., Trollet, C., Bigot, A., Negroni, E., Chaouch, S., Wolff, A., Kandalla, P.K., Marie, S., Di Santo, J., St Guily, J.L., Muntoni, F., Kim, J., Philippi, S., Spuler, S., Levy, N., Blumen, S.C., Voit, T., Wright, W.E., Aamiri, A., Butler-Browne, G., Mouly, V., 2011. Immortalized pathological human myoblasts: towards a universal tool for the study of neuromuscular disorders. *Skelet. Muscle* 1, 34. <https://doi.org/10.1186/2044-5040-1-34>
- Manley, J.L., Krainer, A.R., 2010. A rational nomenclature for serine/arginine-rich protein splicing factors (SR proteins). *Genes Dev.* 24, 1073–1074. <https://doi.org/10.1101/gad.1934910>
- Mao, H., McMahon, J.J., Tsai, Y.-H., Wang, Z., Silver, D.L., 2016. Haploinsufficiency for Core Exon Junction Complex Components Disrupts Embryonic Neurogenesis and Causes p53-Mediated Microcephaly. *PLOS Genet.* 12, e1006282. <https://doi.org/10.1371/journal.pgen.1006282>

- Marquez, F.G., Cisneros, B., Garcia, F., Ceja, V., Velázquez, F., Depardón, F., Cervantes, L., Rendón, A., Mornet, D., Rosas-vargas, H., Mustre, M., Montañez, C., 2003. Differential expression and subcellular distribution of dystrophin Dp71 isoforms during differentiation process. *Neuroscience* 118, 957–966. [https://doi.org/10.1016/S0306-4522\(03\)00063-0](https://doi.org/10.1016/S0306-4522(03)00063-0)
- Massouridès, E., Polentes, J., Mangeot, P.-E., Mournetas, V., Nectoux, J., Deburgrave, N., Nusbaum, P., Leturcq, F., Popplewell, L., Dickson, G., Wein, N., Flanigan, K.M., Peschanski, M., Chelly, J., Pinset, C., 2015. Dp412e: a novel human embryonic dystrophin isoform induced by BMP4 in early differentiated cells. *Skelet. Muscle* 5, 40. <https://doi.org/10.1186/s13395-015-0062-6>
- Matsuo, M., Awano, H., Matsumoto, M., Nagai, M., Kawaguchi, T., Zhang, Z., Nishio, H., 2017. Dystrophin Dp116: A Yet to Be Investigated Product of the Duchenne Muscular Dystrophy Gene. *Genes* 8, 251. <https://doi.org/10.3390/genes8100251>
- Mayer, A., di Iulio, J., Maleri, S., Eser, U., Vierstra, J., Reynolds, A., Sandstrom, R., Stamatoyannopoulos, J.A., Churchman, L.S., 2015. Native Elongating Transcript Sequencing Reveals Human Transcriptional Activity at Nucleotide Resolution. *Cell* 161, 541–554. <https://doi.org/10.1016/j.cell.2015.03.010>
- Mazloomian, A., Araki, S., Otori, M., El-Naggar, A.M., Yap, D., Bashashati, A., Nakao, S., Sorensen, P.H., Nakanishi, A., Shah, S., Aparicio, S., 2019. Pharmacological systems analysis defines EIF4A3 functions in cell-cycle and RNA stress granule formation. *Commun. Biol.* 2, 165. <https://doi.org/10.1038/s42003-019-0391-9>
- McGlincy, N.J., Smith, C.W.J., 2008. Alternative splicing resulting in nonsense-mediated mRNA decay: what is the meaning of nonsense? *Trends Biochem. Sci.* 33, 385–393. <https://doi.org/10.1016/j.tibs.2008.06.001>
- Merkhofer, E.C., Hu, P., Johnson, T.L., 2014. Introduction to Cotranscriptional RNA Splicing. *Methods Mol. Biol. Clifton NJ* 1126, 83–96. [https://doi.org/10.1007/978-1-62703-980-2\\_6](https://doi.org/10.1007/978-1-62703-980-2_6)
- Metkar, M., Ozadam, H., Lajoie, B.R., Imakaev, M., Mirny, L.A., Dekker, J., Moore, M.J., 2018. Higher-Order Organization Principles of Pre-translational mRNPs. *Mol. Cell* 72, 715–726.e3. <https://doi.org/10.1016/j.molcel.2018.09.012>
- Meyer, P., Notarnicola, C., Meli, A.C., Matecki, S., Hugon, G., Salvador, J., Khalil, M., Féasson, L., Cances, C., Cottalorda, J., Desguerre, I., Cuisset, J.-M., Sabouraud, P., Lacampagne, A., Chevassus, H., Rivier, F., Carnac, G., 2021. Skeletal Ryanodine Receptors Are Involved in Impaired Myogenic Differentiation in Duchenne Muscular Dystrophy Patients. *Int. J. Mol. Sci.* 22, 12985. <https://doi.org/10.3390/ijms222312985>
- Michelle, L., Cloutier, A., Toutant, J., Shkreta, L., Thibault, P., Durand, M., Garneau, D., Gendron, D., Lapointe, E., Couture, S., Le Hir, H., Klinck, R., Elela, S.A., Prinos, P., Chabot, B., 2012. Proteins Associated with the Exon Junction Complex Also Control the Alternative Splicing of Apoptotic Regulators. *Mol. Cell. Biol.* 32, 954–967. <https://doi.org/10.1128/MCB.06130-11>
- Miller, E.E., Kobayashi, G.S., Musso, C.M., Allen, M., Ishiy, F.A.A., de Caires, L.C., Jr, Goulart, E., Griesi-Oliveira, K., Zechi-Ceide, R.M., Richieri-Costa, A., Bertola, D.R., Passos-Bueno, M.R., Silver, D.L., 2017. EIF4A3 deficient human iPSCs and mouse models demonstrate neural crest defects that underlie Richieri-Costa-Pereira syndrome. *Hum. Mol. Genet.* 26, 2177–2191. <https://doi.org/10.1093/hmg/ddx078>
- Miro, J., Bougé, A.-L., Murauer, E., Beyne, E., Da Cunha, D., Claustres, M., Koenig, M., Tuffery-Giraud, S., 2020. First Identification of RNA-Binding Proteins That Regulate Alternative Exons in the Dystrophin Gene. *Int. J. Mol. Sci.* 21, 7803. <https://doi.org/10.3390/ijms21207803>
- Miro, J., Laaref, A.M., Rofidal, V., Lagrafeuille, R., Hem, S., Thorel, D., Méchin, D., Mamchaoui, K., Mouly, V., Claustres, M., Tuffery-Giraud, S., 2015. FUBP1: a new protagonist in splicing regulation of the DMD gene. *Nucleic Acids Res.* 43, 2378–2389. <https://doi.org/10.1093/nar/gkv086>
- Moizard, M.P., Toutain, A., Fournier, D., Berret, F., Raynaud, M., Billard, C., Andres, C., Moraine, C., 2000. Severe cognitive impairment in DMD: obvious clinical indication for Dp71 isoform point mutation screening. *Eur. J. Hum. Genet. EJHG* 8, 552–556. <https://doi.org/10.1038/sj.ejhg.5200488>
- Monaco, A.P., Bertelson, C.J., Liechti-Gallati, S., Moser, H., Kunkel, L.M., 1988. An explanation for the phenotypic differences between patients bearing partial deletions of the DMD locus. *Genomics* 2, 90–95. [https://doi.org/10.1016/0888-7543\(88\)90113-9](https://doi.org/10.1016/0888-7543(88)90113-9)
- Mukund, K., Subramaniam, S., 2020. Skeletal muscle: A review of molecular structure and function, in health and disease. *Wiley Interdiscip. Rev. Syst. Biol. Med.* 12, e1462. <https://doi.org/10.1002/wsbm.1462>
- Muniz, L., Nicolas, E., Trouche, D., 2021. RNA polymerase II speed: a key player in controlling and adapting transcriptome composition. *EMBO J.* 40, e105740. <https://doi.org/10.15252/embj.2020105740>

- Muñoz, M.J., Nieto Moreno, N., Giono, L.E., Cambindo Botto, A.E., Dujardin, G., Bastianello, G., Lavore, S., Torres-Méndez, A., Menck, C.F.M., Blencowe, B.J., Irimia, M., Foiani, M., Kornblihtt, A.R., 2017. Major Roles for Pyrimidine Dimers, Nucleotide Excision Repair, and ATR in the Alternative Splicing Response to UV Irradiation. *Cell Rep.* 18, 2868–2879. <https://doi.org/10.1016/j.celrep.2017.02.066>
- Muntoni, F., Melis, M.A., Ganau, A., Dubowitz, V., n.d. Transcription of the Dystrophin Gene in Normal Tissues and in Skeletal Muscle of a Family with X-Linked Dilated Cardiomyopathy 7.
- Muntoni, F., Strong, P.N., 1989. Transcription of the dystrophin gene in Duchenne muscular dystrophy muscle. *FEBS Lett.* 252, 95–98. [https://doi.org/10.1016/0014-5793\(89\)80896-8](https://doi.org/10.1016/0014-5793(89)80896-8)
- Muntoni, F., Torelli, S., Ferlini, A., 2003. Dystrophin and mutations: one gene, several proteins, multiple phenotypes. *Lancet Neurol.* 2, 731–740. [https://doi.org/10.1016/S1474-4422\(03\)00585-4](https://doi.org/10.1016/S1474-4422(03)00585-4)
- Murachelli, A.G., Ebert, J., Basquin, C., Le Hir, H., Conti, E., 2012. The structure of the ASAP core complex reveals the existence of a Pinin-containing PSAP complex. *Nat. Struct. Mol. Biol.* 19, 378–386. <https://doi.org/10.1038/nsmb.2242>
- Naftelberg, S., Schor, I.E., Ast, G., Kornblihtt, A.R., 2015. Regulation of Alternative Splicing Through Coupling with Transcription and Chromatin Structure. *Annu. Rev. Biochem.* 84, 165–198. <https://doi.org/10.1146/annurev-biochem-060614-034242>
- Naidoo, M., Anthony, K., 2020. Dystrophin Dp71 and the Neuropathophysiology of Duchenne Muscular Dystrophy. *Mol. Neurobiol.* 57, 1748–1767. <https://doi.org/10.1007/s12035-019-01845-w>
- Nakka, K., Ghigna, C., Gabellini, D., Dilworth, F.J., 2018. Diversification of the muscle proteome through alternative splicing. *Skelet. Muscle* 8, 8. <https://doi.org/10.1186/s13395-018-0152-3>
- Nasif, S., Contu, L., Mühlemann, O., 2018. Beyond quality control: The role of nonsense-mediated mRNA decay (NMD) in regulating gene expression. *Semin. Cell Dev. Biol., Diversity of transcripts emanating from protein-coding genes* 75, 78–87. <https://doi.org/10.1016/j.semcd.2017.08.053>
- Nawrotzki, R., Loh, N.Y., Ruegg, M.A., Davies, K.E., Blake, D.J., 1998. Characterisation of alpha-dystrobrevin in muscle. *J. Cell Sci.* 111 ( Pt 17), 2595–2605. <https://doi.org/10.1242/jcs.111.17.2595>
- Newey, Sarah E., Benson, M.A., Ponting, C.P., Davies, K.E., Blake, D.J., 2000. Alternative splicing of dystrobrevin regulates the stoichiometry of syntrophin binding to the dystrophin protein complex. *Curr. Biol.* 10, 1295–1298. [https://doi.org/10.1016/S0960-9822\(00\)00760-0](https://doi.org/10.1016/S0960-9822(00)00760-0)
- Nguyen, T.H.D., Galej, W.P., Fica, S.M., Lin, P.-C., Newman, A.J., Nagai, K., 2016. CryoEM structures of two spliceosomal complexes: starter and dessert at the spliceosome feast. *Curr. Opin. Struct. Biol., Folding and binding • Nucleic acids-protein complexes* 36, 48–57. <https://doi.org/10.1016/j.sbi.2015.12.005>
- Niba, E.T.E., Awano, H., Lee, T., Takeshima, Y., Shinohara, M., Nishio, H., Matsuo, M., 2021. Dystrophin Dp71 Subisoforms Localize to the Mitochondria of Human Cells. *Life* 11, 978. <https://doi.org/10.3390/life11090978>
- Niba, E.T.E., Yamanaka, R., Rani, A.Q.M., Awano, H., Matsumoto, M., Nishio, H., Matsuo, M., 2017. DMD transcripts in CRL-2061 rhabdomyosarcoma cells show high levels of intron retention by intron-specific PCR amplification. *Cancer Cell Int.* 17, 58. <https://doi.org/10.1186/s12935-017-0428-4>
- Nicolas, A., Lucchetti-Miganeh, C., Yaou, R.B., Kaplan, J.-C., Chelly, J., Leturcq, F., Barloy-Hubler, F., Le Rumeur, E., 2012. Assessment of the structural and functional impact of in-frame mutations of the DMD gene, using the tools included in the eDystrophin online database. *Orphanet J. Rare Dis.* 7, 45. <https://doi.org/10.1186/1750-1172-7-45>
- Nielsen, K.H., Chamieh, H., Andersen, C.B.F., Fredslund, F., Hamborg, K., Le Hir, H., Andersen, G.R., 2009. Mechanism of ATP turnover inhibition in the EJC. *RNA* 15, 67–75. <https://doi.org/10.1261/rna.1283109>
- Nieto Moreno, N., Villafañez, F., Giono, L.E., Cuenca, C., Soria, G., Muñoz, M.J., Kornblihtt, A.R., 2020. GSK-3 is an RNA polymerase II phospho-CTD kinase. *Nucleic Acids Res.* 48, 6068–6080. <https://doi.org/10.1093/nar/gkaa322>
- Nikonova, E., Kao, S.-Y., Spletter, M.L., 2020. Contributions of alternative splicing to muscle type development and function. *Semin. Cell Dev. Biol.* 104, 65–80. <https://doi.org/10.1016/j.semcd.2020.02.003>
- Nishida, A., Minegishi, M., Takeuchi, A., Awano, H., Niba, E.T.E., Matsuo, M., 2015a. Neuronal SH-SY5Y cells use the C-dystrophin promoter coupled with exon 78 skipping and display multiple patterns of alternative splicing including two intronic insertion events. *Hum. Genet.* 134, 993–1001. <https://doi.org/10.1007/s00439-015-1581-2>
- Nishida, A., Minegishi, M., Takeuchi, A., Niba, E.T.E., Awano, H., Lee, T., Iijima, K., Takeshima, Y., Matsuo, M., 2015b.

- Tissue- and case-specific retention of intron 40 in mature dystrophin mRNA. *J. Hum. Genet.* 60, 327–333. <https://doi.org/10.1038/jhg.2015.24>
- Norwood, F.L., Sutherland-Smith, A.J., Keep, N.H., Kendrick-Jones, J., 2000. The structure of the N-terminal actin-binding domain of human dystrophin and how mutations in this domain may cause Duchenne or Becker muscular dystrophy. *Structure* 8, 481–491. [https://doi.org/10.1016/S0969-2126\(00\)00132-5](https://doi.org/10.1016/S0969-2126(00)00132-5)
- Nudel, U., Zuk, D., Einat, P., Zeelon, E., Levy, Z., Neuman, S., Yaffe, D., 1989. Duchenne muscular dystrophy gene product is not identical in muscle and brain. *Nature* 337, 76–78. <https://doi.org/10.1038/337076a0>
- Obrdlík, A., Lin, G., Haberman, N., Ule, J., Ephrussi, A., 2019. The Transcriptome-wide Landscape and Modalities of EJC Binding in Adult *Drosophila*. *Cell Rep.* 28, 1219–1236.e11. <https://doi.org/10.1016/j.celrep.2019.06.088>
- Ohlendieck, K., Swandulla, D., 2021. Complexity of skeletal muscle degeneration: multi-systems pathophysiology and organ crosstalk in dystrophinopathy. *Pflugers Arch.* 473, 1813–1839. <https://doi.org/10.1007/s00424-021-02623-1>
- Orengo, J.P., Ward, A.J., Cooper, T.A., 2011. Alternative splicing dysregulation secondary to skeletal muscle regeneration. *Ann. Neurol.* 69, 681–690. <https://doi.org/10.1002/ana.22278>
- Otani, Y., Fujita, K., Kameyama, T., Mayeda, A., 2021. The Exon Junction Complex Core Represses Cancer-Specific Mature mRNA Re-splicing: A Potential Key Role in Terminating Splicing. *Int. J. Mol. Sci.* 22, 6519. <https://doi.org/10.3390/ijms22126519>
- Ottens, F., Boehm, V., Sibley, C.R., Ule, J., Gehring, N.H., 2017. Transcript-specific characteristics determine the contribution of endo- and exonucleolytic decay pathways during the degradation of nonsense-mediated decay substrates. *RNA* 23, 1224–1236. <https://doi.org/10.1261/rna.059659.116>
- Pai, A.A., Paggi, J.M., Yan, P., Adelman, K., Burge, C.B., 2018. Numerous recursive sites contribute to accuracy of splicing in long introns in flies. *PLoS Genet.* 14, e1007588. <https://doi.org/10.1371/journal.pgen.1007588>
- Palma, M., Leroy, C., Salomé-Desnoullez, S., Werkmeister, E., Kong, R., Mongy, M., Le Hir, H., Lejeune, F., 2021. A role for AKT1 in nonsense-mediated mRNA decay. *Nucleic Acids Res.* 49, 11022–11037. <https://doi.org/10.1093/nar/gkab882>
- Palombo, R., Verdile, V., Paronetto, M.P., 2020. Poison-Exon Inclusion in DHX9 Reduces Its Expression and Sensitizes Ewing Sarcoma Cells to Chemotherapeutic Treatment. *Cells* 9, 328. <https://doi.org/10.3390/cells9020328>
- Pan, Q., Shai, O., Lee, L.J., Frey, B.J., Blencowe, B.J., 2008. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.* 40, 1413–1415. <https://doi.org/10.1038/ng.259>
- Patton, R.D., Sanjeev, M., Woodward, L.A., Mabin, J.W., Bundschuh, R., Singh, G., 2020. Chemical crosslinking enhances RNA immunoprecipitation for efficient identification of binding sites of proteins that photo-crosslink poorly with RNA. *RNA* 26, 1216–1233. <https://doi.org/10.1261/rna.074856.120>
- Paúl-González, S., Aragón, J., Rodríguez-Martínez, G., Romo-Yáñez, J., Montanez, C., 2021. Differential expression of Dp71 and Dp40 isoforms in proliferating and differentiated neural stem cells: Identification of Dp40 splicing variants. *Biochem. Biophys. Res. Commun.* 560, 152–158. <https://doi.org/10.1016/j.bbrc.2021.03.142>
- Pertea, M., Shumate, A., Pertea, G., Varabyou, A., Breitwieser, F.P., Chang, Y.-C., Madugundu, A.K., Pandey, A., Salzberg, S.L., 2018. CHES: a new human gene catalog curated from thousands of large-scale RNA sequencing experiments reveals extensive transcriptional noise. *Genome Biol.* 19, 1–14. <https://doi.org/10.1186/s13059-018-1590-2>
- Pervouchine, D., Popov, Y., Berry, A., Borsari, B., Frankish, A., Guigó, R., 2019. Integrative transcriptomic analysis suggests new autoregulatory splicing events coupled with nonsense-mediated mRNA decay. *Nucleic Acids Res.* 47, 5293–5306. <https://doi.org/10.1093/nar/gkz193>
- Pervouchine, D.D., Knowles, D.G., Guigo, R., 2013. Intron-centric estimation of alternative splicing from RNA-seq data. *Bioinformatics* 29, 273–274. <https://doi.org/10.1093/bioinformatics/bts678>
- Pilgram, G.S.K., Potikanond, S., Baines, R.A., Fradkin, L.G., Noordermeer, J.N., 2010. The Roles of the Dystrophin-Associated Glycoprotein Complex at the Synapse. *Mol. Neurobiol.* 41, 1–21. <https://doi.org/10.1007/s12035-009-8089-5>
- Pillers, D.-A.M., Bulman, D.E., Weleber, R.G., Sigesmund, D.A., Musarella, M.A., Murphey, W.H., Westall, C., Panton, C., Becker, L.E., Worton, R.G., Ray, P.N., n.d. Dystrophin expression in the human retina is required for nonnal function as defined by electroretinography 5.
- Pradeepa, M.M., Sutherland, H.G., Ule, J., Grimes, G.R., 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, e1002717.

<https://doi.org/10.1371/journal.pgen.1002717>

Quesnel-Vallières, M., Irimia, M., Cordes, S.P., Blencowe, B.J., 2015. Essential roles for the splicing regulator nSR100/SRRM4 during nervous system development. *Genes Dev.* 29, 746–759. <https://doi.org/10.1101/gad.256115.114>

Rambout, X., Maquat, L.E., 2021. NCBP3: A Multifaceted Adaptive Regulator of Gene Expression. *Trends Biochem. Sci.* 46, 87–96. <https://doi.org/10.1016/j.tibs.2020.09.003>

Rambout, X., Maquat, L.E., 2020. The nuclear cap-binding complex as choreographer of gene transcription and pre-mRNA processing. *Genes Dev.* 34, 1113–1127. <https://doi.org/10.1101/gad.339986.120>

Rani, A.Q.M., Farea, M., Maeta, K., Kawaguchi, T., Awano, H., Nagai, M., Nishio, H., Matsuo, M., 2019. Identification of the shortest splice variant of Dp71, together with five known variants, in glioblastoma cells. *Biochem. Biophys. Res. Commun.* 508, 640–645. <https://doi.org/10.1016/j.bbrc.2018.11.168>

Rani, A.Q.M., Yamamoto, T., Kawaguchi, T., Maeta, K., Awano, H., Nishio, H., Matsuo, M., 2020. Intronic Alternative Polyadenylation in the Middle of the DMD Gene Produces Half-Size N-Terminal Dystrophin with a Potential Implication of ECG Abnormalities of DMD Patients. *Int. J. Mol. Sci.* 21, 3555. <https://doi.org/10.3390/ijms21103555>

Rau, F., Lainé, J., Ramanoudjame, L., Ferry, A., Arandel, L., Delalande, O., Jollet, A., Dingli, F., Lee, K.-Y., Peccate, C., Lorain, S., Kabashi, E., Athanasopoulos, T., Koo, T., Loew, D., Swanson, M.S., Le Rumeur, E., Dickson, G., Allamand, V., Marie, J., Furling, D., 2015. Abnormal splicing switch of DMD's penultimate exon compromises muscle fibre maintenance in myotonic dystrophy. *Nat. Commun.* 6, 7205. <https://doi.org/10.1038/ncomms8205>

Reichert, V.L., Le Hir, H., Jurica, M.S., 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. <https://doi.org/10.1101/gad.1030602>

Reixachs-Solé, M., Eyra, E., n.d. Uncovering the impacts of alternative splicing on the proteome with current omics techniques. *WIREs RNA n/a*, e1707. <https://doi.org/10.1002/wrna.1707>

Ricotti, V., Jäggle, H., Theodorou, M., Moore, A.T., Muntoni, F., Thompson, D.A., 2016a. Ocular and neurodevelopmental features of Duchenne muscular dystrophy: a signature of dystrophin function in the central nervous system. *Eur. J. Hum. Genet.* 24, 562–568. <https://doi.org/10.1038/ejhg.2015.135>

Ricotti, V., Mandy, W.P.L., Scoto, M., Pane, M., Deconinck, N., Messina, S., Mercuri, E., Skuse, D.H., Muntoni, F., 2016b. Neurodevelopmental, emotional, and behavioural problems in Duchenne muscular dystrophy in relation to underlying dystrophin gene mutations. *Dev. Med. Child Neurol.* 58, 77–84. <https://doi.org/10.1111/dmcn.12922>

Roberts, R.G., Bentley, D.R., Bobrow, M., 1993. Infidelity in the structure of ectopic transcripts: a novel exon in lymphocyte dystrophin transcripts. *Hum. Mutat.* 2, 293–299. <https://doi.org/10.1002/humu.1380020409>

Rodor, J., Pan, Q., Blencowe, B.J., Eyra, E., Cáceres, J.F., 2016. The RNA-binding profile of Acinus, a peripheral component of the exon junction complex, reveals its role in splicing regulation. *RNA* 22, 1411–1426. <https://doi.org/10.1261/rna.057158.116>

Roignant, J.-Y., Treisman, J.E., 2010. Exon Junction Complex Subunits Are Required to Splice *Drosophila* MAP Kinase, a Large Heterochromatic Gene. *Cell* 143, 238–250. <https://doi.org/10.1016/j.cell.2010.09.036>

Romo-Yáñez, J., Ceja, V., Ilaraza-Lomelí, R., Coral-Vázquez, R., Velázquez, F., Mornet, D., Rendón, A., Montañez, C., 2007. Dp71ab/DAPs complex composition changes during the differentiation process in PC12 cells. *J. Cell. Biochem.* 102, 82–97. <https://doi.org/10.1002/jcb.21281>

Romo-Yáñez, J., Rodríguez-Martínez, G., Aragón, J., Siqueiros-Márquez, L., Herrera-Salazar, A., Velasco, I., Montañez, C., 2020. Characterization of the expression of dystrophins and dystrophin-associated proteins during embryonic neural stem/progenitor cell differentiation. *Neurosci. Lett.* 736, 135247. <https://doi.org/10.1016/j.neulet.2020.135247>

Roszbach, O., Hung, L.-H., Schreiner, S., Grishina, I., Heiner, M., Hui, J., Bindereif, A., 2009. Auto- and cross-regulation of the hnRNP L proteins by alternative splicing. *Mol. Cell. Biol.* 29, 1442–1451. <https://doi.org/10.1128/MCB.01689-08>

Rufener, S.C., Mühlemann, O., 2013. eIF4E-bound mRNPs are substrates for nonsense-mediated mRNA decay in mammalian cells. *Nat. Struct. Mol. Biol.* 20, 710–717. <https://doi.org/10.1038/nsmb.2576>

Rumeur, E.L., 2015. Dystrophin and the two related genetic diseases, Duchenne and Becker muscular dystrophies. *Bosn. J. Basic Med. Sci.* 15, 14–20. <https://doi.org/10.17305/bjbm.2015.636>

Ryu, I., Won, Y.-S., Ha, H., Kim, E., Park, Y., Kim, M.K., Kwon, D.H., Choe, J., Song, H.K., Jung, H., Kim, Y.K., 2019.



- eIF4A3 Phosphorylation by CDKs Affects NMD during the Cell Cycle. *Cell Rep.* 26, 2126–2139.e9. <https://doi.org/10.1016/j.celrep.2019.01.101>
- Sakharkar, M.K., Perumal, B.S., Sakharkar, K.R., Kanguane, P., 2005. An analysis on gene architecture in human and mouse genomes. *In Silico Biol.* 5, 347–365.
- Saulière, J., Murigneux, V., Wang, Z., Marquet, E., Barbosa, I., Le Tonquèze, O., Audic, Y., Paillard, L., Crollius, H.R., Le Hir, H., 2012. CLIP-seq of eIF4AIII reveals transcriptome-wide mapping of the human exon junction complex. *Nat. Struct. Mol. Biol.* 19, 1124–1131. <https://doi.org/10.1038/nsmb.2420>
- Savarese, M., Jonson, P.H., Huovinen, S., Paulin, L., Auvinen, P., Udd, B., Hackman, P., 2018. The complexity of titin splicing pattern in human adult skeletal muscles. *Skelet. Muscle* 8, 11. <https://doi.org/10.1186/s13395-018-0156-z>
- Schier, A.C., Taatjes, D.J., 2020. Structure and mechanism of the RNA polymerase II transcription machinery. *Genes Dev.* 34, 465–488. <https://doi.org/10.1101/gad.335679.119>
- Schlautmann, L.P., Boehm, V., Lackmann, J.-W., Altmüller, J., Dieterich, C., Gehring, N.H., 2021. Exon junction complex-associated multi-adaptor RNPS1 nucleates splicing regulatory complexes to maintain transcriptome surveillance (preprint). *Molecular Biology*. <https://doi.org/10.1101/2021.08.20.457088>
- Schlautmann, L.P., Gehring, N.H., 2020. A Day in the Life of the Exon Junction Complex. *Biomolecules* 10, 866. <https://doi.org/10.3390/biom10060866>
- Schmidt, U., Im, K.-B., Benzing, C., Janjetovic, S., Rippe, K., Lichter, P., Wachsmuth, M., 2009. Assembly and mobility of exon–exon junction complexes in living cells. *RNA* 15, 862–876. <https://doi.org/10.1261/rna.1387009>
- Schneider, M., Will, C.L., Anokhina, M., Tazi, J., Urlaub, H., Lührmann, R., 2010. Exon Definition Complexes Contain the Tri-snRNP and Can Be Directly Converted into B-like Precatalytic Splicing Complexes. *Mol. Cell* 38, 223–235. <https://doi.org/10.1016/j.molcel.2010.02.027>
- Senapathy, P., Shapiro, M.B., Harris, N.L., 1990. [16] Splice junctions, branch point sites, and exons: Sequence statistics, identification, and applications to genome project, in: *Methods in Enzymology*. Academic Press, pp. 252–278. [https://doi.org/10.1016/0076-6879\(90\)83018-5](https://doi.org/10.1016/0076-6879(90)83018-5)
- Shadle, S.C., Zhong, J.W., Campbell, A.E., Conerly, M.L., Jagannathan, S., Wong, C.-J., Morello, T.D., van der Maarel, S.M., Tapscott, S.J., 2017. DUX4-induced dsRNA and MYC mRNA stabilization activate apoptotic pathways in human cell models of facioscapulohumeral dystrophy. *PLoS Genet.* 13. <https://doi.org/10.1371/journal.pgen.1006658>
- Shao, W., Zeitlinger, J., 2017. Paused RNA polymerase II inhibits new transcriptional initiation. *Nat. Genet.* 49, 1045–1051. <https://doi.org/10.1038/ng.3867>
- Shapiro, M.B., Senapathy, P., 1987. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res.* 15, 7155–7174.
- Shi, Y., 2017. Mechanistic insights into precursor messenger RNA splicing by the spliceosome. *Nat. Rev. Mol. Cell Biol.* 18, 655–670. <https://doi.org/10.1038/nrm.2017.86>
- Shibuya, T., Tange, T.Ø., Sonenberg, N., Moore, M.J., 2004. eIF4AIII binds spliced mRNA in the exon junction complex and is essential for nonsense-mediated decay. *Nat. Struct. Mol. Biol.* 11, 346–351. <https://doi.org/10.1038/nsmb750>
- SHIBUYA, T., TANGE, T.Ø., STROUPE, M.E., MOORE, M.J., 2006. Mutational analysis of human eIF4AIII identifies regions necessary for exon junction complex formation and nonsense-mediated mRNA decay. *RNA* 12, 360–374. <https://doi.org/10.1261/rna.2190706>
- Shimori, M., Inoue, K., Sakamoto, H., 2013. A Specific Set of Exon Junction Complex Subunits Is Required for the Nuclear Retention of Unspliced RNAs in *Caenorhabditis elegans*. *Mol. Cell Biol.* 33, 444–456. <https://doi.org/10.1128/MCB.01298-12>
- Shum, E.Y., Jones, S.H., Shao, A., Dumdie, J., Krause, M.D., Chan, W.-K., Lou, C.-H., Espinoza, J.L., Song, H.-W., Phan, M.H., Ramaiah, M., Huang, L., McCarrey, J.R., Peterson, K.J., De Rooij, D.G., Cook-Andersen, H., Wilkinson, M.F., 2016. The Antagonistic Gene Paralogues Upf3a and Upf3b Govern Nonsense-Mediated RNA Decay. *Cell* 165, 382–395. <https://doi.org/10.1016/j.cell.2016.02.046>
- Sibley, C.R., Blazquez, L., Ule, J., 2016. Lessons from non-canonical splicing. *Nat. Rev. Genet.* 17, 407–421. <https://doi.org/10.1038/nrg.2016.46>
- Sibley, C.R., Emmett, W., Blazquez, L., Faro, A., Haberman, N., Briese, M., Trabzuni, D., Ryten, M., Weale, M.E., Hardy,

- J., Modic, M., Curk, T., Wilson, S.W., Plagnol, V., Ule, J., 2015. Recursive splicing in long vertebrate genes. *Nature* 521, 371–375. <https://doi.org/10.1038/nature14466>
- Singh, G., Kucukural, A., Cenik, C., Leszyk, J.D., Shaffer, S.A., Weng, Z., Moore, M.J., 2012. The Cellular EJC Interactome Reveals Higher-Order mRNP Structure and an EJC-SR Protein Nexus. *Cell* 151, 750–764. <https://doi.org/10.1016/j.cell.2012.10.007>
- Singh, G., Rebbapragada, I., Lykke-Andersen, J., 2008. A Competition between Stimulators and Antagonists of Upf Complex Recruitment Governs Human Nonsense-Mediated mRNA Decay. *PLoS Biol.* 6, e111. <https://doi.org/10.1371/journal.pbio.0060111>
- Singh, J., Padgett, R.A., 2009. Rates of in situ transcription and splicing in large human genes. *Nat. Struct. Mol. Biol.* 16, 1128–1133. <https://doi.org/10.1038/nsmb.1666>
- Singh, K., Dilworth, F.J., 2013. Differential modulation of cell cycle progression distinguishes members of the myogenic regulatory factor family of transcription factors. *FEBS J.* 280, 3991–4003. <https://doi.org/10.1111/febs.12188>
- Singh, K.K., Wachsmuth, L., Kulozik, A.E., Gehring, N.H., 2013. Two mammalian MAGOH genes contribute to exon junction complex composition and nonsense-mediated decay. *RNA Biol.* 10, 1291–1298. <https://doi.org/10.4161/rna.25827>
- Singh, R.K., Xia, Z., Bland, C.S., Kalsotra, A., Scavuzzo, M.A., Curk, T., Ule, J., Li, W., Cooper, T.A., 2014. Rbfox2-coordinated alternative splicing of Mef2d and Rock2 controls myoblast fusion during myogenesis. *Mol. Cell* 55, 592–603. <https://doi.org/10.1016/j.molcel.2014.06.035>
- Sironi, M., Cagliani, R., Pozzoli, U., Bardoni, A., Comi, G.P., Giorda, R., Bresolin, N., 2002. The dystrophin gene is alternatively spliced throughout its coding sequence. *FEBS Lett.* 517, 163–166. [https://doi.org/10.1016/S0014-5793\(02\)02613-3](https://doi.org/10.1016/S0014-5793(02)02613-3)
- Snow, W.M., Anderson, J.E., Fry, M., 2014. Regional and genotypic differences in intrinsic electrophysiological properties of cerebellar Purkinje neurons from wild-type and dystrophin-deficient mdx mice. *Neurobiol. Learn. Mem.* 107, 19–31. <https://doi.org/10.1016/j.nlm.2013.10.017>
- Srebrow, A., Kornblihtt, A.R., 2006. The connection between splicing and cancer. *J. Cell Sci.* 119, 2635–2641. <https://doi.org/10.1242/jcs.03053>
- Steckelberg, A.-L., Altmueller, J., Dieterich, C., Gehring, N.H., 2015. CWC22-dependent pre-mRNA splicing and eIF4A3 binding enables global deposition of exon junction complexes. *Nucleic Acids Res.* 43, 4687–4700. <https://doi.org/10.1093/nar/gkv320>
- Steckelberg, A.-L., Boehm, V., Gromadzka, A.M., Gehring, N.H., 2012. CWC22 Connects Pre-mRNA Splicing and Exon Junction Complex Assembly. *Cell Rep.* 2, 454–461. <https://doi.org/10.1016/j.celrep.2012.08.017>
- Stone, M.R., O'Neill, A., Catino, D., Bloch, R.J., 2005. Specific Interaction of the Actin-binding Domain of Dystrophin with Intermediate Filaments Containing Keratin 19. *Mol. Biol. Cell* 16, 4280–4293. <https://doi.org/10.1091/mbc.E05-02-0112>
- Surono, A., Takeshima, Y., Wibawa, T., Pramono, Z.A.D., Matsuo, M., 1997. Six Novel Transcripts That Remove a Huge Intron Ranging from 250 to 800 kb Are Produced by Alternative Splicing of the 5' Region of the Dystrophin Gene in Human Skeletal Muscle. *Biochem. Biophys. Res. Commun.* 239, 895–899. <https://doi.org/10.1006/bbrc.1997.7579>
- Tadayoni, R., Rendon, A., Soria-Jasso, L.E., Cisneros, B., 2012. Dystrophin Dp71: The Smallest but Multifunctional Product of the Duchenne Muscular Dystrophy Gene. *Mol. Neurobiol.* 45, 43–60. <https://doi.org/10.1007/s12035-011-8218-9>
- Takyar, S., Hickerson, R.P., Noller, H.F., 2005. mRNA Helicase Activity of the Ribosome. *Cell* 120, 49–58. <https://doi.org/10.1016/j.cell.2004.11.042>
- Tan, Y., Jin, Y., Wang, S., Cao, J., Ren, Z., 2021. The RNA surveillance factor UPF1 regulates the migration and adhesion of porcine skeletal muscle satellite cells. *J. Muscle Res. Cell Motil.* 42, 203–217. <https://doi.org/10.1007/s10974-020-09585-4>
- TANGE, T.Ø., SHIBUYA, T., JURICA, M.S., MOORE, M.J., 2005. Biochemical analysis of the EJC reveals two new factors and a stable tetrameric protein core. *RNA* 11, 1869–1883. <https://doi.org/10.1261/rna.2155905>
- Tavanez, J.P., Madl, T., Kooshapur, H., Sattler, M., Valcárcel, J., 2012. hnRNP A1 Proofreads 3' Splice Site Recognition by U2AF. *Mol. Cell* 45, 314–329. <https://doi.org/10.1016/j.molcel.2011.11.033>

- Tennyson, C., 1996. Expression of the dystrophin isoform Dp71 in differentiating human fetal myogenic cultures. *Hum. Mol. Genet.* 5, 1559–1566. <https://doi.org/10.1093/hmg/5.10.1559>
- Tennyson, C.N., Dally, G.Y., Ray, P.N., Worton, R.G., 1996. Expression of the dystrophin isoform Dp71 in differentiating human fetal myogenic cultures. *Hum. Mol. Genet.* 5, 1559–1566. <https://doi.org/10.1093/hmg/5.10.1559>
- Tewes, A.-C., Rall, K.K., Römer, T., Hucke, J., Kapczuk, K., Brucker, S., Wieacker, P., Ledig, S., 2015. Variations in RBM8A and TBX6 are associated with disorders of the müllerian ducts. *Fertil. Steril.* 103, 1313–1318. <https://doi.org/10.1016/j.fertnstert.2015.02.014>
- Thermann, R., 1998. Binary specification of nonsense codons by splicing and cytoplasmic translation. *EMBO J.* 17, 3484–3494. <https://doi.org/10.1093/emboj/17.12.3484>
- Thorley, M., Duguez, S., Mazza, E.M.C., Valsoni, S., Bigot, A., Mamchaoui, K., Harmon, B., Voit, T., Mouly, V., Duddy, W., 2016. Skeletal muscle characteristics are preserved in hTERT/cdk4 human myogenic cell lines. *Skelet. Muscle* 6, 43. <https://doi.org/10.1186/s13395-016-0115-5>
- Thorvaldsdóttir, H., Robinson, J.T., Mesirov, J.P., 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* 14, 178–192. <https://doi.org/10.1093/bib/bbs017>
- Tilgner, H., Knowles, D.G., Johnson, R., Davis, C.A., Chakraborty, S., Djebali, S., Curado, J., Snyder, M., Gingeras, T.R., Guigó, R., 2012. Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs. *Genome Res.* 22, 1616–1625. <https://doi.org/10.1101/gr.134445.111>
- Tilgner, H., Nikolaou, C., Althammer, S., Sammeth, M., Beato, M., Valcárcel, J., Guigó, R., 2009. Nucleosome positioning as a determinant of exon recognition. *Nat. Struct. Mol. Biol.* 16, 996–1001. <https://doi.org/10.1038/nsmb.1658>
- Tinsley, J.M., Blake, D.J., Davies, K.E., 1993. Apo-dystrophin-3: a 2.2kb transcript from the DMD locus encoding the dystrophin glycoprotein binding site. *Hum. Mol. Genet.* 2, 521–524. <https://doi.org/10.1093/hmg/2.5.521>
- Titus, M.B., Chang, A.W., Olesnick, E.C., 2021. Exploring the Diverse Functional and Regulatory Consequences of Alternative Splicing in Development and Disease. *Front. Genet.* 12, 775395. <https://doi.org/10.3389/fgene.2021.775395>
- Tollervy, J.R., Curk, T., Rogelj, B., Briese, M., Cereda, M., Kayikci, M., König, J., Hortobágyi, T., Nishimura, A.L., Župunski, V., Patani, R., Chandran, S., Rot, G., Zupan, B., Shaw, C.E., Ule, J., 2011. Characterizing the RNA targets and position-dependent splicing regulation by TDP-43. *Nat. Neurosci.* 14, 452–458. <https://doi.org/10.1038/nn.2778>
- Toma, K.G., Rebbapragada, I., Durand, S., Lykke-Andersen, J., 2015. Identification of elements in human long 3' UTRs that inhibit nonsense-mediated decay. *RNA* 21, 887–897. <https://doi.org/10.1261/rna.048637.114>
- Tome, J.M., Tippens, N.D., Lis, J.T., 2018. Single-molecule nascent RNA sequencing reveals regulatory domain architecture at promoters and enhancers. *Nat. Genet.* 50, 1533–1541. <https://doi.org/10.1038/s41588-018-0234-5>
- Torelli, S., Muntoni, F., 1996. Alternative splicing of dystrophin exon 4 in normal human muscle. *Hum. Genet.* 97, 521–523. <https://doi.org/10.1007/BF02267079>
- Tsuda, T., Fitzgerald, K.K., 2017. Dystrophic Cardiomyopathy: Complex Pathobiological Processes to Generate Clinical Phenotype. *J. Cardiovasc. Dev. Dis.* 4, 14. <https://doi.org/10.3390/jcdd4030014>
- Tuffery-Giraud, S., Bérout, C., Leturcq, F., Yaou, R.B., Hamroun, D., Michel-Calemard, L., Moizard, M.-P., Bernard, R., Cossée, M., Boisseau, P., Blayau, M., Creveaux, I., Guiochon-Mantel, A., de Martinville, B., Philippe, C., Monnier, N., Bieth, E., Van Kien, P.K., Desmet, F.-O., Humbertclaude, V., Kaplan, J.-C., Chelly, J., Claustres, M., 2009. Genotype–phenotype analysis in 2,405 patients with a dystrophinopathy using the UMD–DMD database: a model of nationwide knowledgebase. *Hum. Mutat.* 30, 934–945. <https://doi.org/10.1002/humu.20976>
- Tuffery-Giraud, S., Miro, J., Koenig, M., Claustres, M., 2017. Normal and altered pre-mRNA processing in the DMD gene. *Hum. Genet.* 136, 1155–1172. <https://doi.org/10.1007/s00439-017-1820-9>
- Tyers, L., Davids, L.M., Wilmshurst, J.M., Esterhuizen, A.I., 2018. Skin cells for use in an alternate diagnostic method for Duchenne muscular dystrophy. *Neuromuscul. Disord.* 28, 553–563. <https://doi.org/10.1016/j.nmd.2018.05.001>
- Viegas, M.H., Gehring, N.H., Breit, S., Hentze, M.W., Kulozik, A.E., 2007. The abundance of RNPS1, a protein component of the exon junction complex, can determine the variability in efficiency of the Nonsense Mediated Decay pathway. *Nucleic Acids Res.* 35, 4542–4551. <https://doi.org/10.1093/nar/gkm461>
- Villarreal-Silva, M., Centeno-Cruz, F., Suárez-Sánchez, R., Garrido, E., Cisneros, B., 2011. Knockdown of dystrophin Dp71 impairs PC12 cells cycle: localization in the spindle and cytokinesis structures implies a role for Dp71 in cell division.

PLoS One 6, e23504. <https://doi.org/10.1371/journal.pone.0023504>

Viphakone, N., Sudbery, I., Griffith, L., Heath, C.G., Sims, D., Wilson, S.A., 2019. Co-transcriptional Loading of RNA Export Factors Shapes the Human Transcriptome. *Mol. Cell* 75, 310-323.e8. <https://doi.org/10.1016/j.molcel.2019.04.034>

Wagner, R.E., Frye, M., 2021. Noncanonical functions of the serine-arginine-rich splicing factor (SR) family of proteins in development and disease. *BioEssays* 43, 2000242. <https://doi.org/10.1002/bies.202000242>

Waite, A., Brown, S.C., Blake, D.J., 2012. The dystrophin–glycoprotein complex in brain development and disease. *Trends Neurosci.* 35, 487–496. <https://doi.org/10.1016/j.tins.2012.04.004>

Wan, R., Bai, R., Zhan, X., Shi, Y., 2020. How Is Precursor Messenger RNA Spliced by the Spliceosome? 29.

Wan, Y., Anastasakis, D.G., Rodriguez, J., Palangat, M., Gudla, P., Zaki, G., Tandon, M., Pegoraro, G., Chow, C.C., Hafner, M., Larson, D.R., 2021. Dynamic imaging of nascent RNA reveals general principles of transcription dynamics and stochastic splice site selection. *Cell* 184, 2878-2895.e20. <https://doi.org/10.1016/j.cell.2021.04.012>

Wang, C., Li, L., Cheng, W., 2015. Anorectal malformation: the etiological factors. *Pediatr. Surg. Int.* 31, 795–804. <https://doi.org/10.1007/s00383-015-3685-0>

Wang, E.T., Ward, A.J., Cherone, J.M., Giudice, J., Wang, T.T., Treacy, D.J., Lambert, N.J., Freese, P., Saxena, T., Cooper, T.A., Burge, C.B., 2015. Antagonistic regulation of mRNA expression and splicing by CELF and MBNL proteins. *Genome Res.* 25, 858–871. <https://doi.org/10.1101/gr.184390.114>

Wang, Y., Wang, J., Gao, L., Lafyatis, R., Stamm, S., Andreadis, A., 2005. Tau Exons 2 and 10, Which Are Misregulated in Neurodegenerative Diseases, Are Partly Regulated by Silencers Which Bind a SRp30c-SRp55 Complex That Either Recruits or Antagonizes htra2 $\beta$ 1. *J. Biol. Chem.* 280, 14230–14239. <https://doi.org/10.1074/jbc.M413846200>

Wang, Z., Ballut, L., Barbosa, I., Le Hir, H., 2018. Exon Junction Complexes can have distinct functional flavours to regulate specific splicing events. *Sci. Rep.* 8. <https://doi.org/10.1038/s41598-018-27826-y>

Wang, Z., Murigneux, V., Le Hir, H., 2014. Transcriptome-wide modulation of splicing by the exon junction complex. *Genome Biol.* 15, 551. <https://doi.org/10.1186/s13059-014-0551-7>

Warf, M.B., Diegel, J.V., von Hippel, P.H., Berglund, J.A., 2009. The protein factors MBNL1 and U2AF65 bind alternative RNA structures to regulate splicing. *Proc. Natl. Acad. Sci. U. S. A.* 106, 9203–9208. <https://doi.org/10.1073/pnas.0900342106>

Wei, N., Cheng, Y., Wang, Z., Liu, Y., Luo, C., Liu, L., Chen, L., Xie, Z., Lu, Y., Feng, Y., 2015. SRSF10 Plays a Role in Myoblast Differentiation and Glucose Production via Regulation of Alternative Splicing. *Cell Rep.* 13, 1647–1657. <https://doi.org/10.1016/j.celrep.2015.10.038>

Weischenfeldt, J., Waage, J., Tian, G., Zhao, J., Damgaard, I., Jakobsen, J.S., Kristiansen, K., Krogh, A., Wang, J., Porse, B.T., 2012. Mammalian tissues defective in nonsense-mediated mRNA decay display highly aberrant splicing patterns. *Genome Biol.* 13, R35. <https://doi.org/10.1186/gb-2012-13-5-r35>

Wen, Y., Bi, P., Liu, W., Asakura, A., Keller, C., Kuang, S., 2012. Constitutive Notch Activation Upregulates Pax7 and Promotes the Self-Renewal of Skeletal Muscle Satellite Cells. *Mol. Cell. Biol.* 32, 2300–2311. <https://doi.org/10.1128/MCB.06753-11>

Wilkinson, M.E., Charenton, C., Nagai, K., 2020. RNA Splicing by the Spliceosome. *Annu. Rev. Biochem.* 89, 359–388. <https://doi.org/10.1146/annurev-biochem-091719-064225>

Will, C.L., Lührmann, R., 2011. Spliceosome Structure and Function. *Cold Spring Harb. Perspect. Biol.* 3, a003707. <https://doi.org/10.1101/cshperspect.a003707>

Woodward, L.A., Mabin, J.W., Gangras, P., Singh, G., 2017. The exon junction complex: a lifelong guardian of mRNA fate: EJC: assembly, structure, and function. *Wiley Interdiscip. Rev. RNA* 8, e1411. <https://doi.org/10.1002/wrna.1411>

Xie, W., Zhu, H., Zhao, M., Wang, L., Li, S., Zhao, C., Zhou, Y., Zhu, B., Jiang, X., Liu, W., Ren, C., 2021. Crucial roles of different RNA-binding hnRNP proteins in Stem Cells. *Int. J. Biol. Sci.* 17, 807–817. <https://doi.org/10.7150/ijbs.55120>

Xu, S.-M., Curry-Hyde, A., Sytnyk, V., Janitz, M., 2022. RNA polyadenylation patterns in the human transcriptome. *Gene* 816, 146133. <https://doi.org/10.1016/j.gene.2021.146133>

Xu, X., Zhang, L., Tong, P., Xun, G., Su, W., Xiong, Z., Zhu, T., Zheng, Y., Luo, S., Pan, Y., Xia, K., Hu, Z., 2013. Exome sequencing identifies UPF3B as the causative gene for a Chinese non-syndrome mental retardation pedigree. *Clin. Genet.*

83, 560–564. <https://doi.org/10.1111/cge.12014>

Xue, C., Gu, X., Li, G., Bao, Z., Li, L., 2021. Expression and Functional Roles of Eukaryotic Initiation Factor 4A Family Proteins in Human Cancers. *Front. Cell Dev. Biol.* 9, 711965. <https://doi.org/10.3389/fcell.2021.711965>

Yamamoto, T., Awano, H., Zhang, Z., Sakuma, M., Kitaaki, S., Matsumoto, M., Nagai, M., Sato, I., Imanishi, T., Hayashi, N., Matsuo, M., Iijima, K., Saegusa, J., 2018. Cardiac Dysfunction in Duchenne Muscular Dystrophy Is Less Frequent in Patients With Mutations in the Dystrophin Dp116 Coding Region Than in Other Regions. *Circ. Genomic Precis. Med.* 11, e001782. <https://doi.org/10.1161/CIRCGEN.117.001782>

Yamashita, A., Izumi, N., Kashima, I., Ohnishi, T., Saari, B., Katsuhata, Y., Muramatsu, R., Morita, T., Iwamatsu, A., Hachiya, T., Kurata, R., Hirano, H., Anderson, P., Ohno, S., 2009. SMG-8 and SMG-9, two novel subunits of the SMG-1 complex, regulate remodeling of the mRNA surveillance complex during nonsense-mediated mRNA decay. *Genes Dev.* 23, 1091–1105. <https://doi.org/10.1101/gad.1767209>

Yamashita, K., Suzuki, A., Satoh, Y., Ide, M., Amano, Y., Masuda-Hirata, M., Hayashi, Y.K., Hamada, K., Ogata, K., Ohno, S., 2010. The 8th and 9th tandem spectrin-like repeats of utrophin cooperatively form a functional unit to interact with polarity-regulating kinase PAR-1b. *Biochem. Biophys. Res. Commun.* 391, 812–817. <https://doi.org/10.1016/j.bbrc.2009.11.144>

Yan, Q., Weyn-Vanhenhenryck, S.M., Wu, J., Sloan, S.A., Zhang, Y., Chen, K., Wu, J.Q., Barres, B.A., Zhang, C., 2015. Systematic discovery of regulated and conserved alternative exons in the mammalian brain reveals NMD modulating chromatin regulators. *Proc. Natl. Acad. Sci.* 112, 3445–3450. <https://doi.org/10.1073/pnas.1502849112>

Yang, H., Beutler, B., Zhang, D., 2021. Emerging roles of spliceosome in cancer and immunity. *Protein Cell.* <https://doi.org/10.1007/s13238-021-00856-5>

Ye, J., She, X., Liu, Z., He, Z., Gao, X., Lu, L., Liang, R., Lin, Y., 2021. Eukaryotic Initiation Factor 4A-3: A Review of Its Physiological Role and Involvement in Oncogenesis. *Front. Oncol.* 11, 712045. <https://doi.org/10.3389/fonc.2021.712045>

Yearim, A., Gelfman, S., Shayevitch, R., Melcer, S., Glaich, O., Mallm, J.-P., Nissim-Rafinia, M., Cohen, A.-H.S., Rippe, K., Meshorer, E., Ast, G., 2015. HP1 is involved in regulating the global impact of DNA methylation on alternative splicing. *Cell Rep.* 10, 1122–1134. <https://doi.org/10.1016/j.celrep.2015.01.038>

Yeo, G., Burge, C.B., 2004. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J. Comput. Biol. J. Comput. Mol. Cell Biol.* 11, 377–394. <https://doi.org/10.1089/1066527041410418>

Yepiskoposyan, H., Aeschmann, F., Nilsson, D., Okoniewski, M., Mühlemann, O., 2011. Autoregulation of the nonsense-mediated mRNA decay pathway in human cells. *RNA* 17, 2108–2118. <https://doi.org/10.1261/rna.030247.111>

Yi, Z., Arvola, R.M., Myers, S., Dilsavor, C.N., Abu Alhasan, R., Carter, B.N., Patton, R.D., Bundschuh, R., Singh, G., 2022. Mammalian UPF3A and UPF3B can activate nonsense-mediated mRNA decay independently of their exon junction complex binding. *EMBO J.* <https://doi.org/10.15252/embj.2021109202>

Yi, Z., Sanjeev, M., Singh, G., 2021a. The Branched Nature of the Nonsense-Mediated mRNA Decay Pathway. *Trends Genet.* 37, 143–159. <https://doi.org/10.1016/j.tig.2020.08.010>

Yi, Z., Sanjeev, M., Singh, G., 2021b. The Branched Nature of the Nonsense-Mediated mRNA Decay Pathway. *Trends Genet.* 37, 143–159. <https://doi.org/10.1016/j.tig.2020.08.010>

Yu, Q., Guo, J., Zhou, J., 2004. A minimal length between tau exon 10 and 11 is required for correct splicing of exon 10. *J. Neurochem.* 90, 164–172. <https://doi.org/10.1111/j.1471-4159.2004.02477.x>

Zetoune, A.B., Fontanière, S., Magnin, D., Anczuków, O., Buisson, M., Zhang, C.X., Mazoyer, S., 2008. Comparison of nonsense-mediated mRNA decay efficiency in various murine tissues. *BMC Genet.* 9, 83. <https://doi.org/10.1186/1471-2156-9-83>

Zhang, J., Sun, X., Qian, Y., LaDuca, J.P., Maquat, L.E., 1998. At Least One Intron Is Required for the Nonsense-Mediated Decay of Triosephosphate Isomerase mRNA: a Possible Link between Nuclear Splicing and Cytoplasmic Translation. *Mol. Cell. Biol.* 18, 5272–5283.

Zhang, X., Yan, C., Hang, J., Finci, L.I., Lei, J., Shi, Y., 2017. An Atomic Structure of the Human Spliceosome. *Cell* 169, 918–929.e14. <https://doi.org/10.1016/j.cell.2017.04.033>

Zhang, X., Zhan, X., Yan, C., Zhang, W., Liu, D., Lei, J., Shi, Y., 2019. Structures of the human spliceosomes before and after release of the ligated exon. *Cell Res.* 29, 274–285. <https://doi.org/10.1038/s41422-019-0143-x>

Zhang, X.-O., Fu, Y., Mou, H., Xue, W., Weng, Z., 2018. The temporal landscape of recursive splicing during Pol II transcription elongation in human cells. *PLOS Genet.* 14, e1007579. <https://doi.org/10.1371/journal.pgen.1007579>

Zhou, L., Hang, J., Zhou, Y., Wan, R., Lu, G., Yin, P., Yan, C., Shi, Y., 2014. Crystal structures of the Lsm complex bound to the 3' end sequence of U6 small nuclear RNA. *Nature* 506, 116–120. <https://doi.org/10.1038/nature12803>

Zlotorynski, E., 2019. Intron definition, exon definition and back-splicing revisited. *Nat. Rev. Mol. Cell Biol.* 20, 661–661. <https://doi.org/10.1038/s41580-019-0178-3>

Zorio, D.A.R., Bentley, D.L., 2001. Transcription elongation: The 'Foggy' is lifting.... *Curr. Biol.* 11, R144–R146. [https://doi.org/10.1016/S0960-9822\(01\)00063-X](https://doi.org/10.1016/S0960-9822(01)00063-X)

# ANNEXES

# Annexe 1





Article

# First Identification of RNA-Binding Proteins That Regulate Alternative Exons in the Dystrophin Gene

Julie Miro <sup>1</sup> , Anne-Laure Bougé <sup>1,†</sup>, Eva Murauer <sup>1</sup> , Emmanuelle Beyne <sup>1</sup>, Dylan Da Cunha <sup>1</sup>, Mireille Claustres <sup>1</sup>, Michel Koenig <sup>1,2</sup> and Sylvie Tuffery-Giraud <sup>1,\*</sup>

<sup>1</sup> LGMR, Univ Montpellier, 34093 Montpellier, France; julie.miro@inserm.fr (J.M.); anne-laure.bouge@seqone.fr (A.-L.B.); e-murauer@chu-montpellier.fr (E.M.); emmanuelle.beyne@ird.fr (E.B.); dylan.da-cunha@ext.inserm.fr (D.D.C.); Mireille.Claustres@inserm.fr (M.C.); michel.koenig@inserm.fr (M.K.)

<sup>2</sup> CHU of Montpellier, 34090 Montpellier, France

\* Correspondence: sylvie.tuffery@inserm.fr

† Current address: SeqOne, 34000 Montpellier, France.

Received: 22 September 2020; Accepted: 19 October 2020; Published: 21 October 2020



**Abstract:** The Duchenne muscular dystrophy (DMD) gene has a complex expression pattern regulated by multiple tissue-specific promoters and by alternative splicing (AS) of the resulting transcripts. Here, we used an RNAi-based approach coupled with DMD-targeted RNA-seq to identify RNA-binding proteins (RBPs) that regulate splicing of its skeletal muscle isoform (Dp427m) in a human muscular cell line. A total of 16 RBPs comprising the major regulators of muscle-specific splicing events were tested. We show that distinct combinations of RBPs maintain the correct inclusion in the Dp427m of exons that undergo spatio-temporal AS in other dystrophin isoforms. In particular, our findings revealed the complex networks of RBPs contributing to the splicing of the two short DMD exons 71 and 78, the inclusion of exon 78 in the adult Dp427m isoform being crucial for muscle function. Among the RBPs tested, QKI and DDX5/DDX17 proteins are important determinants of DMD exon inclusion. This is the first large-scale study to determine which RBP proteins act on the physiological splicing of the DMD gene. Our data shed light on molecular mechanisms contributing to the expression of the different dystrophin isoforms, which could be influenced by a change in the function or expression level of the identified RBPs.

**Keywords:** DMD gene; skeletal muscle; RNA-binding proteins (RBPs); alternative splicing; targeted RNA-seq; TDP-43

## 1. Introduction

The Duchenne muscular dystrophy (DMD) gene, the largest human gene spanning 2.2 Mb on the X chromosome, is composed of 79 exons separated by introns whose size varies widely (107 to 248,401 nucleotides). In striated skeletal and cardiac muscle, a full-length transcript of 14 kb encodes dystrophin, a subsarcolemmal 427 kDa rod-shaped protein incorporated into a multimolecular membrane-associated protein complex that connects intracellular cytoskeleton to extracellular matrix. This complex ensures membrane stability during muscle contraction and mediates cellular signaling [1,2]. The dystrophin protein is defective in Duchenne muscular dystrophy (DMD), the most severe and frequent muscular dystrophy in children [3].

The DMD gene exhibits tissue-specific isoform expression regulated by at least seven independent promoters. Three encode the full-length isoforms (Dp427p,m,c), including the Dp427m isoform which drives gene expression in skeletal and cardiac muscles. The products of four alternative promoters in downstream introns encode N-terminally truncated proteins, named according to their molecular weight as Dp260, Dp140, Dp116 and Dp71, that are specifically expressed in various organs [4,5].

The shortest Dp71 isoform is the major DMD gene product in the brain [6,7]. Alternative splicing (AS) further increases the diversity of transcript isoforms expressed from the DMD gene, and some notable differences are seen depending on the tissues. In line with the reported high prevalence of AS in nervous system tissues of vertebrates [8–10], splice variants are preponderant in cerebral dystrophin isoforms (Dp427c, Dp140, Dp71), in which they frequently involve the skipping of exons 71, 71–74, and 78 [11–14]. In contrast, the extent of AS is limited in the muscular Dp427m isoform.

We have recently designed a DMD-targeted RNA-seq procedure to establish more accurately the splicing pattern of the Dp427m isoform in adult skeletal muscle [15]. Our data showed the virtual absence of alternative splicing events (ASEs) in Dp427m, the few events detected (among which were exon 71 and exon 78 skipping) not exceeding 3%. These results suggest that a tight regulation is in place in adult skeletal muscle to maintain the inclusion of all 79 exons in the mature transcript and to produce the full-length dystrophin protein. It is worth mentioning that the penultimate exon 78 undergoes a splicing transition during skeletal muscle development, being spliced out in the embryonic isoform while its inclusion is required for muscle structure maintenance in adult tissue [16,17].

To date, the molecular mechanisms that control the spatiotemporal AS of the DMD gene, as well as the functional role of the generated isoforms, remain largely unexplored. Alternative splicing is an essential process that generates networks of functionally coordinated and biologically important transcript isoforms with tissue specific or developmental stage-specific functions [18,19]. Splicing programs are largely controlled by recruiting RNA-binding proteins (RBPs) that recognize specific regulatory sequences embedded in the pre-mRNAs, thereby promoting or inhibiting spliceosome assembly at adjacent splice sites [20].

This study reports for the first time the identification of RBPs that regulate exon inclusion in the Dp427m isoform. For this purpose, we performed a functional RNAi screen targeting 16 muscle-specific or general RBPs in a normal human muscular cell line. The analysis of splicing changes by the DMD-targeted RNA-seq procedure revealed combinatorial regulation by distinct sets of RBPs of a series of DMD exons. Interestingly, the RBP-responsive exons (PE1a, exons 9, 37, 38, 41, 71, 74, 78) have the particularity of being alternatively spliced in other tissues and/or DMD isoforms. Among them, the two short DMD exons 71 and 78 were the most sensitive to variations in intracellular amounts of RBPs.

## 2. Results

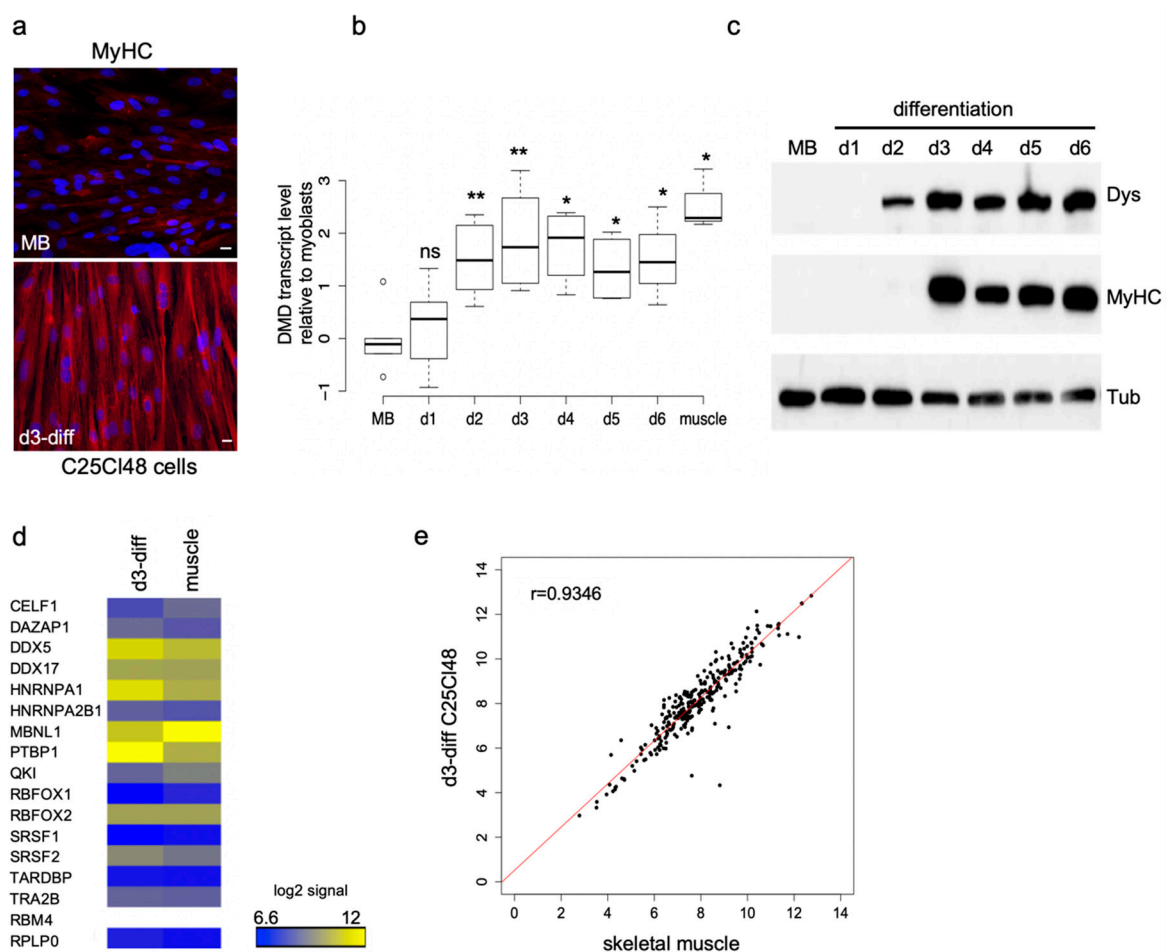
To identify trans-acting factors that play a role in DMD pre-mRNA splicing, we analyzed the splicing changes observed in the Dp427m isoform upon silencing of 16 RBPs by RNA interference (RNAi) in the normal human muscular cell line C25Cl48.

### 2.1. Selection of the Candidate RBPs

The RBPs were mainly selected based on their reported role in the regulation of muscle cell-specific ASEs during myogenic differentiation or developmentally regulated AS transitions [21]. These included the major muscle-specific regulatory factors MBNL1, CELF1, RBFOX1, RBFOX2 proteins [22], RBM4 and PTBP1 known to oppositely modulate muscle-specific exon selection during muscle cell differentiation [23], the STAR Quaking protein (QKI-5) identified as a global splicing factor in muscle cells [24] and the two RNA helicases, DDX5 and DDX17, which contribute to maintain myoblast-specific splicing subprograms [25]. In addition, we selected members of the two main RBP families involved in the regulation of alternative splicing, the serine/arginine-rich proteins (SR proteins) and the heterogeneous nuclear ribonucleoproteins (hnRNPs) [26]. Thus, SRSF1, SRSF2, hTra2 $\beta$ , hnRNPA1, hnRNPA2/B1, DAZAP1 and the TAR DNA-binding protein 43 (TDP-43), a protein involved in different aspects of RNA processing that was recently shown to play a role in the expression of muscle-specific genes [27–29], were included in the functional screen.

## 2.2. Characteristics of the Human Muscular C25Cl48 Cell Line

Since the DMD gene is transcribed only at low levels in proliferating myoblasts cultured *in vitro* [30], we first evaluated the appropriate time point during differentiation to analyze the splicing pattern of Dp427m. After three days of differentiation (d3-diff), the C25Cl48 cells were able to fuse together to form multinucleated myotubes and the expression of the late myogenic marker, myosin heavy chain (MyHC) was detected by immunostaining (Figure 1a). At this time point, Dp427m was efficiently expressed at both the RNA and the protein level in cells (Figure 1b,c). DMD-targeted RNA-seq analysis in d3-diff cells disclosed a basal alternative splicing pattern close to that of skeletal muscle tissue in terms of identity and range of detected ASEs (<5%) [15] except for three events (skipping of exon 71, exon 78 and inclusion of the pseudoexon 1a) that were detected at a higher level in the cells, as previously observed in normal cultured myotubes [31] (Supplementary Table S1). Finally, microarray experiments were performed and showed that the 16 selected RBPs were expressed at a comparable level in skeletal muscle and d3-diff cells except PTBP1 (fold change 3.36) (Figure 1d; Supplementary Table S2). Further analysis of the expression profiles of RNA splicing genes (GO:0008380) revealed a high correlation between d3-diff cells and skeletal muscle tissue (Pearson's correlation coefficient  $r = 0.93$ ) (Figure 1e, Supplementary Table S2). Considering these preliminary experiments, we chose day three of differentiation as an adequate time point to perform the subsequent splicing analysis of Dp427m in the C25Cl48 cell line.

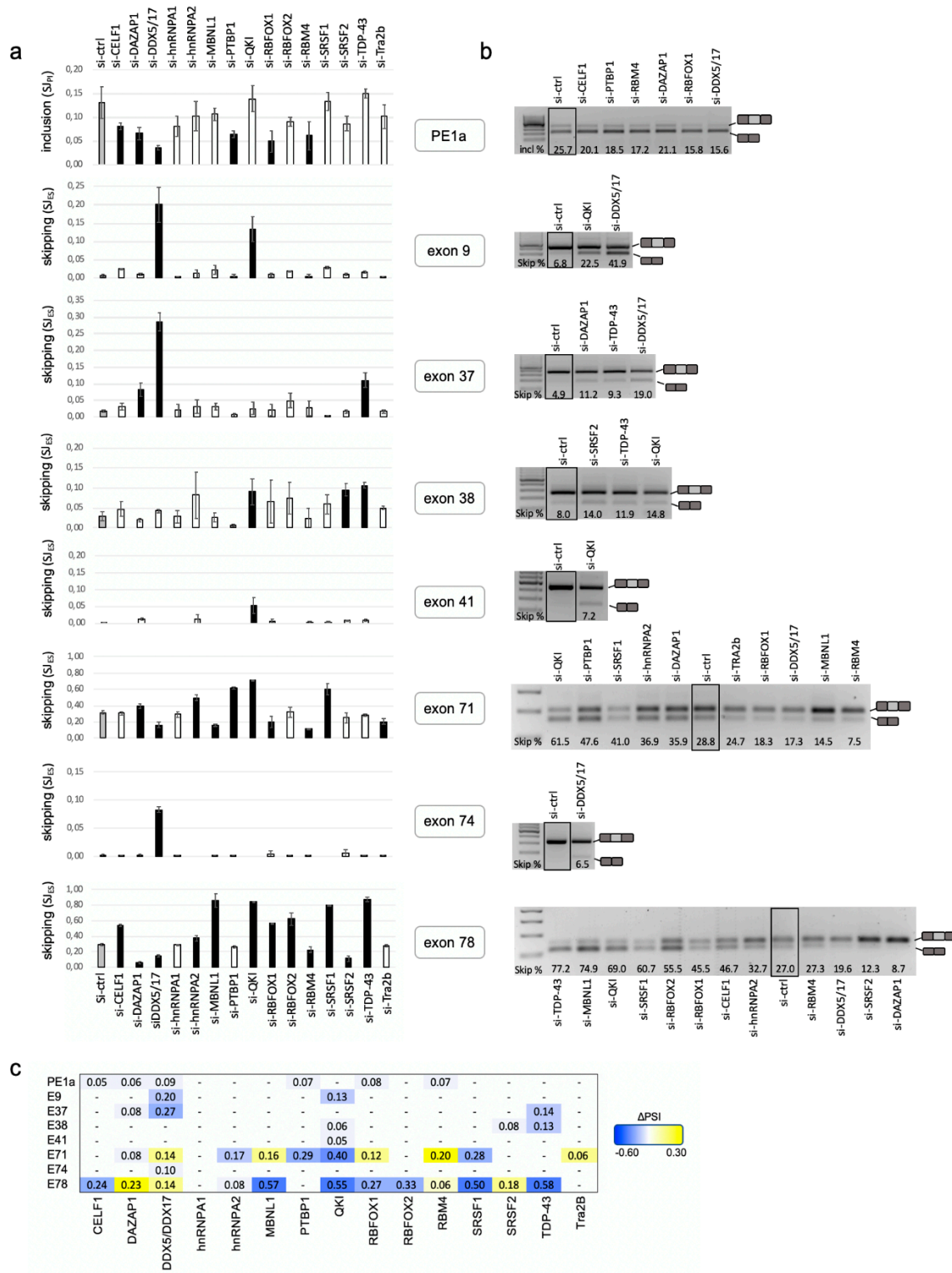


**Figure 1.** Time-course analysis of gene expression in C25Cl48 cells during myogenic differentiation. (a) Visualization of multinucleated myotubes in differentiated C25Cl48 cells. Proliferating myoblasts (MB) and three-days differentiated (d3-diff) C25Cl48 cells were fixed and stained with DAPI (cell nuclei in blue) and with the MY-32 antibody that recognizes all skeletal-muscle myosin (MyHC) isoforms (red).

Specific antibody labeling was visualized using a Cy3-conjugated secondary antibody (laser excitation wavelength of 532–557 nm). Image acquisition was performed using a Leica confocal microscope (20× objective); bar = 15  $\mu\text{m}$ . **(b,c)** Duchenne muscular dystrophy (DMD) gene expression in C25Cl48 cells during differentiation (day 1 to day 6) compared with undifferentiated cells (myoblasts, MB) and skeletal muscle tissue. **(b)** Quantitative real-time polymerase chain reaction (RT-PCR) analysis of DMD transcripts levels are normalized on RPLP0 expression and displayed as relative transcript expression to myoblasts. The results are represented as box-and-whisker plots (average of three to six biological replicates for each time point; Wilcoxon rank-sum test \*\*  $p < 0.01$ ; \*  $p < 0.05$ ). **(c)** Western blot analysis of equal amounts of total cellular protein probed with antibodies against dystrophin (Dys) and myosin heavy chain (MyHC) as a differentiation marker. Tubulin (Tub) was used as a loading control. **(d)** Microarray expression profiling of the 16 targeted RNA-binding protein (RBP) genes in C25Cl48 differentiated cells (d3-diff,  $n = 3$ ) and in skeletal muscle samples (muscle,  $n = 3$ ). The heatmap of the replicate-averaged log<sub>2</sub> signal intensity of the 16 RBP genes is shown. White boxes are for missing values (RBM4). The expression values for the standard acidic ribosomal protein P0 (RPLP0) gene are displayed. **(e)** Microarray expression profiling of RNA splicing genes (Gene Ontology Term GO:0008380). Scatterplots of replicate-averaged log<sub>2</sub> signals of C25Cl48 differentiated cells (d3-diff C25Cl48,  $n = 3$ ) versus skeletal muscle samples ( $n = 3$ ). Red line, best-fit linear regression ( $r$ , Pearson's correlation coefficient).

### 2.3. A Set of DMD Exons Is Sensitive to Changes in Cellular Abundance of RBPs

Proliferating C25Cl48 cells were treated with control siRNA or RBP siRNA and then allowed to differentiate for three days. A high level of silencing was generally achieved at both mRNA and protein levels as revealed by the RT-qPCR and Western blot experiments (Supplementary Figure S1). The splicing changes were assessed by DMD-targeted RNA-seq from two (si-RBP) and four (si-ctrl) biological replicates as previously described [15]. Briefly, this approach consisted in amplifying the whole Dp427m 11.3 kb coding sequence by long-range PCR and sequencing on a 454 GS Junior platform. Quality-based trimmed reads (mean length of 374 bp) were mapped to the human X chromosome using the Spliced Transcripts Alignment to a Reference (STAR) aligner. An average read depth of 1638X was obtained, allowing reliable detection of low-level ASEs. The summary of mapping statistics is provided in the Supplementary Table S3. Annotated, as well as novel splice junctions resulting from various splicing events (single or multiple exon skipping, use of alternative 5' ss and 3' ss within exons or introns, inclusion of pseudoexons) were detected. The splice junction (SJ) usage was determined using the intron-centric method [32], and differential analysis was performed to identify splicing alterations between si-RBP- and si-control-treated cells ( $\Delta\text{SJ} = \text{SJ}_{(\text{si-RBP})} - \text{SJ}_{(\text{si-ctrl})}$ ). Using a minimum absolute abundance change greater than or equal to 5 percentage points ( $|\Delta\text{SJ}| \geq 0.05$ ), differentially spliced junctions, corresponding to the skipping ( $\text{SJ}_{\text{ES}}$ ) of 7 DMD exons (9, 37, 38, 41, 71, 74, 78) and the insertion of a pseudoexon ( $\text{SJ}_{\text{PI}}$ ) originating from intron 1 (PE1a), were identified upon depletion of one or several RBPs. The  $\text{SJ}_{\text{ES}}$  and  $\text{SJ}_{\text{PI}}$  usage reflecting the exon skipping and pseudoexon inclusion levels, respectively, are reported in Figure 2a (see also Supplementary Table S4). All identified ASEs were validated by RT-PCR experiments and the induced exon-inclusion or skipping events detected by the DMD-targeted RNA-seq approach were well correlated with RT-PCR results (Figure 2b). The exon percent spliced in (PSI) index reflecting the global exon inclusion into a population of transcripts was also considered for the 8 RBP-responsive exons. As expected, the changes in inclusion level ( $\Delta\text{PSI} = \text{PSI}_{(\text{si-RBP})} - \text{PSI}_{(\text{si-ctrl})}$ ) were very similar to the  $\Delta\text{SJ}$  values consistent with these exons not being spliced out in other major aberrant splicing events (i.e., multiple exons skipping) (Figure 2c, Supplementary Table S4).



–  $SJ_{(si-ctrl)}$ ) above the threshold  $|\Delta SJ| \geq 0.05$ . Error bars represent the standard deviation from four (si-ctrl) or two (si-RBP) biological replicates. (b) RT-PCR validation assay of significant splicing events ( $|\Delta SJ| \geq 0.05$ ). Agarose gels are shown and the percentages of inclusion (PE1a) or skipping (exons 9, 37, 38, 41, 71, 74, 78) as determined by the Image Lab software are indicated. The control condition (si-ctrl) is in a frame. (c) Variations in the inclusion level of the eight deregulated exons calculated from the exon percent-spliced-in (PSI) values provided by the SJPIPE pipeline as follows ( $|\Delta PSI| = |PSI_{(si-RBP)} - PSI_{(si-ctrl)}|$ ). Blue boxes (negative  $\Delta PSI$  values) and yellow boxes (positive  $\Delta PSI$  values) are indicative of decreased or increased exon inclusion levels upon depletion of RBPs, respectively.

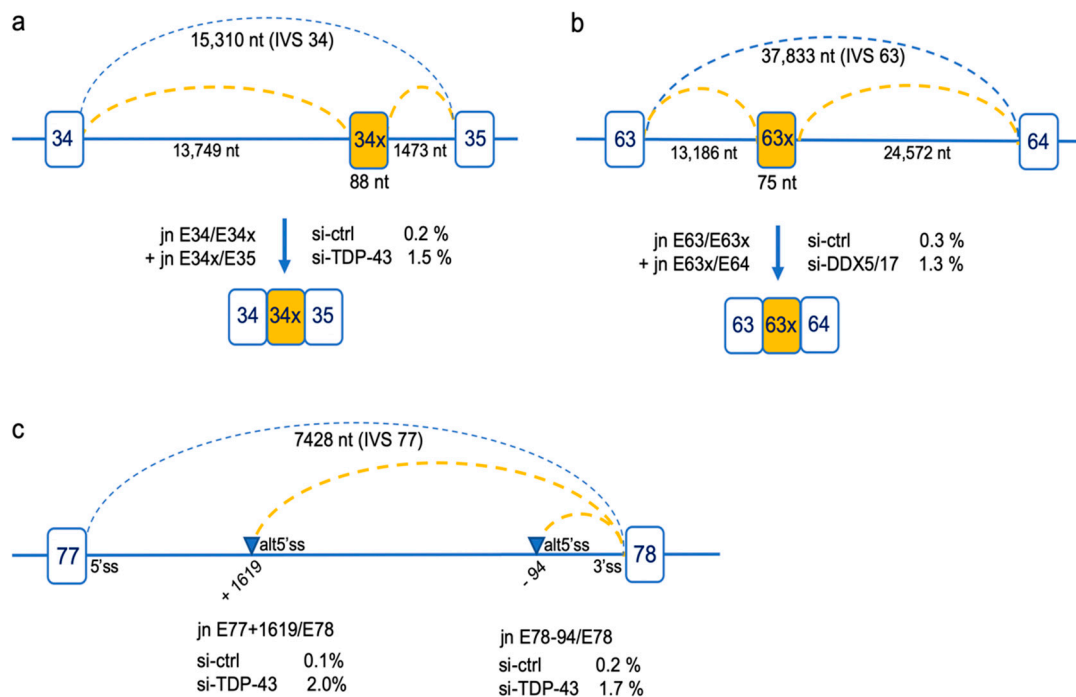
Strikingly, the deregulated exons were not random exons but those known to be alternatively spliced in other tissues or dystrophin isoforms [13,33,34] while they are almost fully included (or excluded with respect to PE1a) in the Dp427m isoform of normal skeletal muscle (Supplementary Table S1). Some notable differences were seen among these exons. A subgroup of five exons (9, 37, 38, 41, 74) were sensitive to the depletion of a limited number of RBPs ( $n = 1$  to 3) that all induced a decrease in their inclusion levels ranging from 5% to 27% ( $-0.27 \leq \Delta PSI \leq -0.05$ ), indicating that these RBPs are required for their inclusion in a physiological context. Regarding PE1a, the depletion of six different RBPs reduced its inclusion level but the changes observed were modest ( $-0.08 \leq \Delta PSI \leq -0.05$ ). Exons 71 and 78 differed from the others in that they appeared to be reactive to the depletion of most of the targeted RBPs. The knockdown (KD) of 10 (exon 71) and 12 (exon 78) out of the 16 RBPs influenced positively or negatively the splicing of these two short DMD exons (39 bp and 32 bp, respectively). Except for MBNL1, RBFOX1 and DAZAP1 acting on exons 71 and 78 splicing in opposite directions, the KD of the other RBPs generally induced splicing changes in the same direction on the two exons. We noticed that the depletion of four RBPs induced changes of high magnitude in the inclusion level of exon 78 ( $|\Delta PSI| \geq 0.5$ ) (MBNL1-, QKI-, SRSF1- and TDP-43-KD) and exon 71 (QKI-KD,  $\Delta PSI = -0.4$ ), thus identifying these RBPs as strong activators of exon 71 or 78 splicing.

The identity and number of regulatory RBPs, as well as their positive or negative activity, varied depending on the deregulated exon. Not considering the peculiar case of PE1a, this study identified a total of seven activators (CELF1, hnRNPA2, PTBP1, QKI, RBFOX2, SRSF1 and TDP-43), two repressors (RBM4 and Tra2 $\beta$ ) of DMD exon inclusion and six proteins regulating splicing in both directions, namely DAZAP1, DDX5/DDX17, MBNL1, RBFOX1 and SRSF2. This functional screen highlighted the more prominent role of QKI and DDX5/DDX17 in DMD pre-mRNA splicing among the tested RBPs as their depletion changed the inclusion level of more than half (four to six) of the eight deregulated exons, acting either as a splicing activator (QKI) or exhibiting opposite effects depending on the deregulated exon (DDX5/DDX17) (Figure 2).

#### 2.4. Rare Splicing Events Revealing Differentially Used Deep Intronic Splice Sites

In addition to the differential analyses of SJs based on a cutoff of  $|\Delta SJ| \geq 0.05$ , we had a look to less frequent splicing events considering a lower change in junction usage ( $0.05 \geq |\Delta SJ| \geq 0.01$ ) (Supplementary Table S5). We observed that other RBPs acted on the inclusion level of exons 9, 37, 38 and 41, albeit to a lesser extent. The skipping of new individual exons (exon 39 and exon 68) and multiple exons (exons 72–74 and exons 71–74) was also detected, the most represented event being the skipping of exon 39 (4.4%) upon DDX5/DDX17 KD. Finally, the analysis disclosed the occurrence of rare splicing events resulting from the use of intronic alternative splice sites. Two were frame-disrupting inclusion of pseudoexons originating from intron 34 (34 $\times$ ) and intron 63 (63 $\times$ ) into the mature transcripts upon depletion of TDP-43 and DDX5/DDX17, respectively (Figure 3a,b). The depletion of TDP-43 was also found to increase the level of splice junctions joining the authentic 3' ss of exon 78 with alternative 5' ss in intron 77 located at positions E77+1619 (2% vs 0.1% in si-ctrl) and E78-94 (1.7% vs 0.2% in si-ctrl) (Figure 3c). These aberrant junctions, present at about 2%, were not modified upon depletion of other RBPs. It is noteworthy that the positions of the 5' ss and 3' ss of pseudoexons 34 $\times$  and 63 $\times$ , as well as that of the alternative 5' ss at position E78-1619 in intron 77, are part of the intronic splice sites detected

by ultra-deep pre-mRNA-seq of nuclear dystrophin transcripts in a previous study and described as potential recursive splice sites of DMD introns [35].



**Figure 3.** Rare splicing events resulting from the usage of alternative intronic splice sites. When changes in splice junction (SJ) usage in a lower range were considered ( $0.05 \geq |\Delta SJ| \geq 0.01$ ), aberrant junctions resulting from the use of deep intronic splice sites were detected in DDX5/17 and TDP-43 KD conditions resulting from the insertion of pseudoexons originating from intron 34 or intron 63 (a,b) or resulting from the use of alternative 5' splice sites (alt5'ss) in intron 77 (blue triangles) (c). The SJ values in RBP (si-TDP-43 and si-DDX5/17) and control (si-ctrl) KD conditions are indicated.

### 3. Discussion

We report the first study aimed at determining which RBPs regulate splicing of the DMD pre-mRNA in skeletal muscle under physiological conditions. We have previously shown that all 79 exons of the Dp427m isoform are constitutively spliced in normal adult skeletal muscle, which is worth mentioning given that this tissue, together with the brain, is known to exhibit extensive AS [18,36,37]. Numerous descriptive studies have depicted tissue-specific ASEs of DMD isoforms, in particular those expressed in the central nervous system [13,14], but the molecular basis of this splicing regulation had remained unexplored to date.

Here, we assessed the role of 16 different trans-acting factors in the splicing of the DMD pre-mRNA, using RNAi in the human C25Cl48 muscular cell line coupled with DMD-targeted RNA-seq of the mature transcript. Targeted RNA-Seq is an accurate method for selecting and sequencing specific transcripts of interest. It is particularly well suited to the analysis of small changes in expression, or in splicing in low-abundance genes such as the DMD gene, owing to the high depth ( $>1000\times$ ) of sequence coverage achieved. The amplicon-based method using specific Dp427m primers efficiently detected splicing variants within the 11.3kb-coding sequence. This method also has inherent drawbacks. It can suffer from PCR amplification biases. Moreover, transcripts with retained intronic sequences whose length exceeds the maximum size limit of amplicons cannot be seen, neither those resulting from the use of alternative poly-A sites [16,38]. Nevertheless, this experimental procedure proved successful in identifying ASEs induced by the KD of the 16 RBPs and all DMD-targeted RNA-seq data were thoroughly validated by RT-PCR and Sanger sequencing.

Interestingly, it was noted that the depletion of the targeted RBPs did not change the inclusion level of constitutive exons but selectively affected exons that undergo spatio-temporal alternative splicing in the various DMD isoforms. These are all in-frame skipping events except the inclusion of PE1a harboring an in-frame stop codon, which may target the resulting transcripts for nonsense-mediated decay (NMD) when included. The inclusion of this cryptic NMD-exon, originating from intron 1 close to the transcription start site, may function as a negative feedback control of DMD gene expression in these tissues or cells as recently described for noncanonical splicing events [39]. The inclusion of PE1a is a very constant phenomenon in all nonmuscle tissues, notably in Dp427m transcripts ectopically expressed in peripheral blood lymphocytes or in the brain Dp427c isoform in neuronal cells [13,33]. Like PE1a, the DMD exon 9 is skipped in many nonmuscle tissues [34]. In-frame splicing changes are more likely to produce functionally relevant isoforms in tissue-specific contexts. Exon 9 skipping removes amino acid residues 278–320 in the first of the four flexible hinge regions that are interspersed in the midrod domain of dystrophin [40]. Alternative splicing of central DMD exons has been described in neuronal SH-SY5Y cells, in which the mutually exclusive skipping of exons 37 and 38, and the activation of an upstream cryptic 3' splice site for exon 41 creating a novel enlarged exon 41, have been identified [13]. We did not detect the usage of this cryptic 3' splice site in intron 40 upon RBPs KD in the muscular cell line, but a low-level of exon 41 skipping was observed upon depletion of QKI. Splicing variants at the 3' end of the gene characterize DMD transcripts in the central nervous system. They consist in combinations of single or multiple exon skipping involving exons 71 to 74 and exon 78, with the isoforms lacking exon 71 and/or exon 78 being by far the most represented [12,13,16,41–43]. In Dp71, the splice isoforms determine the differential subcellular localization and various cellular functions of the short dystrophin isoform in the brain [44,45]. Transcripts spliced out for exon 78 produce a protein with an elongated C-term region by replacing the last 13 amino acids in the protein with an evolutionary conserved sequence of 31 new residues encoding the ancestral alternative amphipathic C-terminal  $\alpha$ -helix.

The inclusion of exon 78 is also regulated during development in the brain and in skeletal muscle [41,46]. The proportion of transcripts lacking exon 78 falls dramatically as development progresses, and MBNL1 has been shown to regulate this splicing transition in skeletal muscle [17]. Experiments in zebrafish and mdx mouse models have confirmed that the reinclusion of exon 78 is essential for skeletal muscle development and muscle fiber organization in adult tissues. The reversion to neonatal splicing pattern of exon 78 (transcripts E78-), due to the sequestration of nuclear MBNL1 and loss of function of the protein, contributes to the progressive dystrophic process in patients with myotonic dystrophy type 1 [17]. In the DMD gene, only one case of single genomic deletion of exon 78 has been reported in a mildly affected 13-year-old BMD patient in whom only minimal impact on morphology and architecture of muscle fibers was observed [47]. Our data add significant new knowledge about splicing regulation of exon 78 by identifying new major regulators of its inclusion in the Dp427m isoform in addition to the previously known enhancer role of MBNL1 that was confirmed here. These are mainly QKI, SRSF1 and TDP-43 and to a lesser extent CELF1, RBFOX1 and RBFOX2. We observed some low-level cryptic splicing events associated with TDP-43 depletion (inclusion of cryptic exon or switching from a canonical splice site to an alternative one), particularly in intron 77. This observation is worth mentioning in light of the recently described role of TDP-43 in repressing cryptic exons and maintaining intron integrity [48,49] although the rarity of the events observed does not allow us to firmly establish this role for the DMD gene.

A remarkable finding is the large number of RBPs found to be involved in splicing regulation of exon 71 and exon 78, the two shortest exons of the gene. To some extent, these two exons share some characteristics with microexons [50]. They meet the least restrictive definition of size (6–51 nt) [51], their splicing is developmentally and/or tissue-specifically regulated, they are frame-preserving or compatible with the synthesis of a stable protein isoform and assumed to be functional as part of protein complexes. A high density of intronic enhancing signals would compensate for the difficulty of the splicing machinery to be efficiently recruited on such short exons [52]. The high number of RBPs



identified as splicing modulators of the DMD exons 71 and 78 may reflect this complex interaction between numerous regulatory sequences.

In conclusion, this study uncovered RBPs that are important for the inclusion in the muscular dystrophin isoform of a subset of DMD exons, highlighting more specifically the role of QKI and the RNA helicases DDX5/DDX17, known global regulators of alternative splicing during muscle development and cell differentiation [21,22,24,25]. The next challenge will be to dissect the molecular mechanisms underlying this regulation. Our findings form a valuable basis for further investigations to evaluate whether changes in the amount or activity of the identified RBPs contribute to splicing of the DMD gene in different cells, tissues or developmental stages. This is of particular interest regarding the cerebral dystrophin isoforms, which play a crucial role in the development and function of the human brain and are coexpressed with genes implicated in neurodevelopmental disorders [53]. DMD affects the central nervous system in about one-third of cases, mostly due to distal mutations affecting the expression of the shorter cerebral isoforms (Dp140, Dp71) [54,55]. However, the pathological mechanisms of CNS involvement have not been fully elucidated, notably in patients in whom the expression of cerebral dystrophin isoforms is preserved. This study paves the way for future work aimed at better understanding the regulation of dystrophin isoforms expression in relation to the coordinated splicing networks in tissues in physiological or pathological conditions.

## 4. Materials and Methods

### 4.1. Cell Culture

The human immortalized myoblast cell line C25Cl48 established from a muscle biopsy of a 25-years-old healthy individual [56] was cultivated as previously described [57]. The differentiation was induced by serum deprivation and elevation of insulin concentration to 10 µg/mL.

### 4.2. siRNA-Mediated Depletion of RBPs

Two consecutive siRNA transfections at a 48-h interval were performed. The first day C25Cl48 cells were seeded at  $1.4 \cdot 10^5$  cells per well in six well-plates and reverse transfected with 37.5 nM of individual siRNA or 20 nM siRNA for CELF1, MBNL1 and QKI, using 7.5 µL lipofectamine RNAiMax (Thermo Fischer Scientific, Illkirch, France) following the manufacturer's instructions. Si-RBP transfections were performed in duplicate. The siGENOME nontargeting siRNA#5 (Dharmacon, Lafayette, CO, USA) was used as a negative control ( $n = 4$ ). The differentiation was induced for three days from the day of the second siRNA treatment, then cells were harvested for protein and RNA isolation. The target sequences and origin of the siRNAs used are described in Supplementary Table S6. Knockdown was validated by RT-qPCR and Western blot analysis.

### 4.3. DMD-Targeted RNA-Seq

#### 4.3.1. Sample Preparation

Total RNA was extracted using the RNeasy plus mini kit (Qiagen, Courtaboeuf, France). cDNA was prepared from 300 ng total RNA using 2.5 µM oligo dT and Superscript II (Thermo Fisher Scientific, Illkirch, France) as described by the manufacturer.

#### 4.3.2. Long Range-Polymerase Chain Reaction (LR-PCR) and 454-Roche Library Construction and Sequencing

DMD-targeted RNA-seq was performed as previously described [15]. Briefly, the full-length DMD cDNA was amplified as a single long fragment (11.3 kb) by using the GoTaq<sup>®</sup> Long PCR Master Mix (Promega, Charbonnières-les-Bains, France) and primers located in exon 1 and in the 3' UTR of the muscle isoform (Dp427m). The library was prepared from 700 ng of purified LR-PCR products using the Rapid Library Preparation Kit (Roche, Basel, Switzerland). Three Multiplex Identifiers

(MID)-tagged samples were pooled for simultaneous amplification and sequencing on the Roche GS Junior 454 sequencer (Roche, Basel, Switzerland).

#### 4.3.3. Bioinformatics Analysis

Cleaned reads were aligned against the GRCh37/hg19 human X chromosome reference sequence using the STAR aligner (v. 2.3) [58]. Quantification of ASEs was performed based on intron-centric metrics using the integrative pipeline for splicing analyses (IPSA) package [32]. Only the reads that aligned to splice junctions, and not to exon bodies, were considered. From the output file with junction counts (SJcount) noncanonical, novel splice junctions (SJs) covered by a minimum of five reads in at least one out of the two biological replicates of siRBP-treated cells were selected. The identified SJ resulted from the occurrence of exon skipping (ES), pseudoexons insertion (PI) or use of alternative splice site (Altss) events in RBP knockdown conditions. Two separate indices, psi5 and psi3, provided by the SJPIPE pipeline for each SJ were used to calculate the SJ usage resulting from exon skipping (SJ<sub>ES</sub>), pseudoexon insertion (SJ<sub>PI</sub>) or alternative splice site (SJ<sub>Altss</sub>) (details in the calculation method are given in the Supplementary Figure S2). The change in junction usage ( $\Delta$ SJ) was determined as follows  $\Delta$ SJ = SJ<sub>(si-RBP)</sub> - SJ<sub>(si-ctrl)</sub> using the mean SJ values of the replicates (si-RBPs,  $n = 2$  and si-ctrl,  $n = 4$ ). Values greater or equal to five percentage points ( $|\Delta$ SJ| = |SJ<sub>(si-RBP)</sub> - SJ<sub>(si-ctrl)</sub>|  $\geq 0.05$ ), and not overlapping standard deviation with the control condition, were considered differentially spliced. The exon inclusion metrics or percent-spliced-in (PSI) ratio were also recovered from the SJPIPE pipeline for each individual deregulated exon and the mean values of biological replicates were then used to calculate the change in exon inclusion as follows:  $\Delta$ PSI = PSI<sub>(si-RBP)</sub> - PSI<sub>(si-ctrl)</sub>.

#### 4.3.4. Availability of Data

DMD-targeted RNA-seq data for the RBPs silencing experiments were deposited in the gene expression omnibus repository (accession number GSE153803).

#### 4.4. RT-PCR Validation and RT-qPCR

To validate RNA-seq data, independent RT reactions were performed from RBP-depleted and si-ctrl RNA samples using random primers and 300 ng of total RNA. For RT-PCR experiments, 1.5  $\mu$ L of cDNA was used as a template for PCR amplification in a 25  $\mu$ L total volume with the Taq DNA polymerase (New England Biolabs, Evry, France) and primers hybridizing to upstream and downstream exons (sequences available in Supplementary Table S6). The 30 cycles-amplified products were separated on 1.8% agarose gels and spliced products were quantified with ImageLab software (Bio-Rad, Marnes-la-coquette, France). The identity of each amplified PCR product was verified by Sanger sequencing. For quantitative PCR analysis, 1  $\mu$ L of 1/3-diluted cDNA was amplified using SYBR Green I Master kit in a final volume of 10  $\mu$ L containing 0.5  $\mu$ M of gene-specific primers (Supplementary Table S6) on a Light Cycler 480 II (Roche, Basel, Switzerland). For each biological replicate of siRNA-treated cells, two independent reverse transcription reactions were carried out and qPCR assays were run in triplicate and included no-template controls for each siRNAs. Expression levels were calculated according to the  $2^{-\Delta\Delta$ Ct method normalized to RPLP0 expression (Hs\_RPLP0\_2\_SG QuantiTect Primer Assay, Qiagen, Courtaboeuf, France). For assessment of the Dp427m transcript level, a forward primer spanning the exon1-exon2 junction (5'-TAGAGGACTGTTATGAAAGAGAAG-3') and a reverse primer in exon 4 (5'-CAGGGCATGAACTCTTG-3') were used.

#### 4.5. Western Blotting

Cells were washed twice with PBS and lysed in Laemmli sample buffer supplemented with 5%  $\beta$ -mercaptoethanol except for SRSF2 analysis. SRSF2 solubilization required incubation in RIPA buffer supplemented with protease inhibitor cocktail (Sigma-Aldrich, St Quentin, France) during 30 min on ice and under agitation, then samples were centrifuged (14,000 rpm) at 4 °C for 15 min and the supernatant

was diluted in Laemmli sample buffer supplemented with 5%  $\beta$ -mercaptoethanol. The samples were subjected to sonication prior to loading into 10% sodium dodecyl sulphate (SDS)-polyacrylamide gels. The proteins were transferred to polyvinylidene fluoride (PVDF) membranes and probed using specific antibodies against CUG-BP1 (3B1, sc-20003, Santa-Cruz, Heidelberg, Germany), DAZAP1 (gift from F. Pagani, Trieste, Italy), DDX5 (ab10261, Abcam, Paris, France), DDX17 (ab24601, Abcam), hnRNPA1 (4B10, sc-32301, Santa-Cruz, Heidelberg, Germany), hnRNPA2/B1 (DP3B3, sc-32316, Santa-Cruz), MBNL1 (MB1a, gift from G. Morris, CIND, Oswestry, UK) [59], PTBP1 (PTB-NT 4856) and RBFOX1 (1D10) (gifts from Dr D. Black, UCLA, LA, USA), QKI (A300-183A, Bethyl, Hamburg, Germany), RBFOX2 (A300-864A, Bethyl, Hamburg, Germany), RBM4 (E-3, sc-373852, Santa-Cruz, Heidelberg, Germany), SRSF1 (32-4500, Thermo Fischer Scientific, Illkirch, France), SRSF2 (PA5-12402, Thermo Fischer Scientific, Illkirch, France), TDP-43 (G400, Ozyme, Saint-Cyr-l'École, France). Anti- $\beta$ -Tubulin (AA2, Millipore, Molsheim, France) was used as loading control. The blots were developed using an enhanced chemiluminescence (ECL) kit (Pierce, Rockford, IL, USA). For gene expression analysis during myogenic differentiation, antibodies against dystrophin (MAB1694, Millipore, Molsheim, France) and fast skeletal myosin (My-32, Sigma-Aldrich, St Quentin, France) were used.

#### 4.6. Microarray Analyses

For microarray analysis, total RNA samples from three human skeletal muscles (Myobank 20316, Clontech ref#636534, lots #1404229A and #1406360A) and three biological replicates of C25C148 cells differentiated at day three were processed, hybridized on a GeneChip<sup>®</sup> Human Gene 2.0 ST Array, scanned, and quantified at the Affymetrix Service Provider and Core Facility, KFB—Center of Excellence for Fluorescent Bioanalytics (University of Regensburg, Regensburg, Germany; [www.kfb-regensburg.de](http://www.kfb-regensburg.de)).

**Supplementary Materials:** Supplementary materials can be found at <http://www.mdpi.com/1422-0067/21/20/7803/s1>. Table S1: The alternative Splicing Events (ASEs) detected by DMD-targeted RNA-Seq in C25C148 cells differentiated for 3 days; Table S2: The gene expression microarray data in d3-diff C25C148 cells and in skeletal muscle; Table S3: The Quality assessment of targeted RNA-seq data; Table S4: The DMD-targeted RNA-seq data: differential usage of splice junctions  $|\Delta SJ| \geq 0.05$ ; Table S5: The DMD-targeted RNA-seq data: differential usage of splice junctions  $0.05 \geq |\Delta SJ| \geq 0.01$ ; Table S6: The lists and sequences of siRNAs and primers used in this study; Figure S1: The assessment of siRNA-induced silencing of gene expression of the selected RNA-Binding Proteins (RBPs); Figure S2: The calculation of novel Splice Junction (SJ) usage according to the intron-centric metrics method.

**Author Contributions:** Conceptualization, S.T.-G. and J.M.; software, E.B.; formal analysis, E.B., J.M., A.-L.B. and E.M.; investigation, J.M., A.-L.B., E.M., D.D.C., S.T.-G.; data curation, E.B. and A.-L.B.; writing—original draft preparation, J.M. and S.T.-G.; writing—review and editing, M.C. and M.K.; supervision, S.T.-G.; funding acquisition, S.T.-G. and M.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Association Française contre les Myopathies (AFM), number 18413 to S.T.-G. and number 16073 to E.M.; the University of Montpellier (to A.-L.B.).

**Acknowledgments:** We thank Emanuele Buratti (ICGEB, Trieste Italy) for the pFlag-CMV-4-TDP-43 plasmid, Benoit Chabot (University of Sherbrooke, Quebec, Canada) for the pCMV-myc-SRSF2 plasmid and Glenn E. Morris (Wolfson CIND, Oswestry, UK) for the MBNL1 (MB1a) antibody. We would like to acknowledge C.F. Bourgeois and D. Auboeuf (LBMC, ENS, Lyon, France) for helpful discussions and advice in the choice of RBPs to target, and for providing some siRNAs.

**Conflicts of Interest:** The authors declare no conflict of interest.

#### References

1. Gao, Q.Q.; McNally, E.M. The Dystrophin Complex: Structure, Function, and Implications for Therapy. *Comp. Physiol.* **2015**, *5*, 1223–1239.
2. Adams, M.E.; Odom, G.L.; Kim, M.J.; Chamberlain, J.S.; Froehner, S.C. Syntrophin binds directly to multiple spectrin-like repeats in dystrophin and mediates binding of nNOS to repeats 16–17. *Hum. Mol. Genet.* **2018**, *27*, 2978–2985. [[CrossRef](#)] [[PubMed](#)]

3. Norwood, F.L.; Harling, C.; Chinnery, P.F.; Eagle, M.; Bushby, K.; Straub, V. Prevalence of genetic muscle disease in Northern England: In-depth analysis of a muscle clinic population. *Brain* **2009**, *132*, 3175–3186. [[CrossRef](#)] [[PubMed](#)]
4. Muntoni, F.; Torelli, S.; Ferlini, A. Dystrophin and mutations: One gene, several proteins, multiple phenotypes. *Lancet Neurol.* **2003**, *2*, 731–740. [[CrossRef](#)]
5. Tyers, L.; Davids, L.M.; Wilmschurst, J.M.; Esterhuizen, A.I. Skin cells for use in an alternate diagnostic method for Duchenne muscular dystrophy. *Neuromuscul. Disord.* **2018**, *28*, 553–563. [[CrossRef](#)]
6. Lederfein, D.; Levy, Z.; Augier, N.; Mornet, D.; Morris, G.; Fuchs, O.; Yaffe, D.; Nudel, U. A 71-kilodalton protein is a major product of the Duchenne muscular dystrophy gene in brain and other nonmuscle tissues. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 5346–5350. [[CrossRef](#)]
7. Bar, S.; Barnea, E.; Levy, Z.; Neuman, S.; Yaffe, D.; Nudel, U. A novel product of the Duchenne muscular dystrophy gene which greatly differs from the known isoforms in its structure and tissue distribution. *Biochem. J.* **1990**, *272*, 557–560. [[CrossRef](#)]
8. Barbosa-Morais, N.L.; Irimia, M.; Pan, Q.; Xiong, H.Y.; Gueroussov, S.; Lee, L.J.; Slobodeniuc, V.; Kutter, C.; Watt, S.; Colak, R.; et al. The evolutionary landscape of alternative splicing in vertebrate species. *Science* **2012**, *338*, 1587–1593. [[CrossRef](#)]
9. Merkin, J.; Russell, C.; Chen, P.; Burge, C.B. Evolutionary dynamics of gene and isoform regulation in Mammalian tissues. *Science* **2012**, *338*, 1593–1599. [[CrossRef](#)]
10. Raj, B.; Blencowe, B.J. Alternative Splicing in the Mammalian Nervous System: Recent Insights into Mechanisms and Functional Roles. *Neuron* **2015**, *87*, 14–27. [[CrossRef](#)]
11. Lidov, H.G.; Kunkel, L.M. Dp140: Alternatively spliced isoforms in brain and kidney. *Genomics* **1997**, *45*, 132–139. [[CrossRef](#)] [[PubMed](#)]
12. Tadayoni, R.; Rendon, A.; Soria-Jasso, L.E.; Cisneros, B. Dystrophin Dp71: The smallest but multifunctional product of the Duchenne muscular dystrophy gene. *Mol. Neurobiol.* **2012**, *45*, 43–60. [[CrossRef](#)]
13. Nishida, A.; Minegishi, M.; Takeuchi, A.; Awano, H.; Niba, E.T.; Matsuo, M. Neuronal SH-SY5Y cells use the C-dystrophin promoter coupled with exon 78 skipping and display multiple patterns of alternative splicing including two intronic insertion events. *Hum. Gene.* **2015**, *134*, 993–1001. [[CrossRef](#)] [[PubMed](#)]
14. Aragón, J.; González-Reyes, M.; Romo-Yáñez, J.; Vacca, O.; Aguilar-González, G.; Rendón, A.; Vaillend, C.; Montañez, C. Dystrophin Dp71 Isoforms Are Differentially Expressed in the Mouse Brain and Retina: Report of New Alternative Splicing and a Novel Nomenclature for Dp71 Isoforms. *Mol. Neurobiol.* **2018**, *55*, 1376–1386. [[CrossRef](#)] [[PubMed](#)]
15. Bougé, A.L.; Murauer, E.; Beyne, E.; Miro, J.; Varilh, J.; Taulan, M.; Koenig, M.; Claustres, M.; Tuffery-Giraud, S. Targeted RNA-Seq profiling of splicing pattern in the DMD gene: Exons are mostly constitutively spliced in human skeletal muscle. *Sci. Rep.* **2017**, *7*, 39094. [[CrossRef](#)] [[PubMed](#)]
16. Feener, C.A.; Koenig, M.; Kunkel, L.M. Alternative splicing of human dystrophin mRNA generates isoforms at the carboxy terminus. *Nature* **1989**, *338*, 509–511. [[CrossRef](#)]
17. Rau, F.; Lainé, J.; Ramanoudjame, L.; Ferry, A.; Arandel, L.; Delalande, O.; Jollet, A.; Dingli, F.; Lee, K.Y.; Peccate, C.; et al. Abnormal splicing switch of DMD's penultimate exon compromises muscle fibre maintenance in myotonic dystrophy. *Nat. Comm.* **2015**, *6*, 7205. [[CrossRef](#)]
18. Baralle, F.E.; Giudice, J. Alternative splicing as a regulator of development and tissue identity. *Nat. Rev. Mol. Cell. Biol.* **2017**, *18*, 437–451. [[CrossRef](#)]
19. Weyn-Vanhentenryck, S.M.; Feng, H.; Ustianenko, D.; Duffié, R.; Yan, Q.; Jacko, M.; Martinez, J.C.; Goodwin, M.; Zhang, X.; Hengst, U.; et al. Precise temporal regulation of alternative splicing during neural development. *Nat. Commun.* **2018**, *9*, 2189. [[CrossRef](#)]
20. Witten, J.T.; Ule, J. Understanding splicing regulation through RNA splicing maps. *Trends Genet.* **2011**, *27*, 89–97. [[CrossRef](#)]
21. Hinkle, E.R.; Wiedner, H.J.; Black, A.J.; Giudice, J. RNA processing in skeletal muscle biology and disease. *Transcription* **2019**, *10*, 1–20. [[CrossRef](#)]
22. Bland, C.S.; Wang, E.T.; Vu, A.; David, M.P.; Castle, J.C.; Johnson, J.M.; Burge, C.B.; Cooper, T.A. Global regulation of alternative splicing during myogenic differentiation. *Nucleic Acids Res.* **2010**, *38*, 7651–7664. [[CrossRef](#)] [[PubMed](#)]
23. Lin, J.C.; Tarn, W.Y. RBM4 down-regulates PTB and antagonizes its activity in muscle cell-specific alternative splicing. *J. Cell. Biol.* **2011**, *193*, 509–520. [[CrossRef](#)] [[PubMed](#)]

24. Hall, M.P.; Nagel, R.J.; Fagg, W.S.; Shiue, L.; Cline, M.S.; Perriman, R.J.; Donohue, J.P.; Ares, M., Jr. Quaking and PTB control overlapping splicing regulatory networks during muscle cell differentiation. *RNA* **2013**, *19*, 627–638. [[CrossRef](#)] [[PubMed](#)]
25. Bourgeois, C.F.; Mortreux, F.; Auboeuf, D. The multiple functions of RNA helicases as drivers and regulators of gene expression. *Nat. Rev. Mol. Cell Biol.* **2016**, *17*, 426–438. [[CrossRef](#)]
26. Dvinge, H. Regulation of alternative mRNA splicing: Old players and new perspectives. *FEBS Lett.* **2018**, *592*, 2987–3006. [[CrossRef](#)] [[PubMed](#)]
27. Geuens, T.; Bouhy, D.; Timmerman, V. The hnRNP family: Insights into their role in health and disease. *Hum. Genet.* **2016**, *135*, 851–867. [[CrossRef](#)]
28. Appocher, C.; Mohagheghi, F.; Cappelli, S.; Stuardi, C.; Romano, M.; Feiguin, F.; Buratti, E. Major hnRNP proteins act as general TDP-43 functional modifiers both in *Drosophila* and human neuronal cells. *Nucleic Acids Res.* **2017**, *45*, 8026–8045. [[CrossRef](#)]
29. Militello, G.; Hosen, M.R.; Ponomareva, Y.; Gellert, P.; Weirick, T.; John, D.; Hindi, S.M.; Mamchaoui, K.; Mouly, V.; Döring, C.; et al. A novel long non-coding RNA Myolinc regulates myogenesis through TDP-43 and Filip1. *J. Mol. Cell. Biol.* **2018**, *10*, 102–117. [[CrossRef](#)]
30. Dumont, N.A.; Wang, Y.X.; von Maltzahn, J.; Pasut, A.; Bentzinger, C.F.; Brun, C.E.; Rudnicki, M.A. Dystrophin expression in muscle stem cells regulates their polarity and asymmetric division. *Nat. Med.* **2015**, *21*, 1455–1463. [[CrossRef](#)]
31. Nakamori, M.; Kimura, T.; Fujimura, H.; Takahashi, M.P.; Sakoda, S. Altered mRNA splicing of dystrophin in type 1 myotonic dystrophy. *Muscle Nerve* **2007**, *36*, 251–257. [[CrossRef](#)] [[PubMed](#)]
32. Pervouchine, D.D.; Knowles, D.G.; Guigó, R. Intron-centric estimation of alternative splicing from RNA-seq data. *Bioinformatics* **2013**, *29*, 273–274. [[CrossRef](#)] [[PubMed](#)]
33. Roberts, R.G.; Bentley, D.R.; Bobrow, M. Infidelity in the structure of ectopic transcripts: A novel exon in lymphocyte dystrophin transcripts. *Hum. Mutat.* **1993**, *2*, 293–299. [[CrossRef](#)] [[PubMed](#)]
34. Reiss, J.; Rininsland, F. An explanation for the constitutive exon 9 cassette splicing of the DMD Gene. *Hum. Mol. Genet.* **1994**, *3*, 295–298. [[CrossRef](#)] [[PubMed](#)]
35. Gazzoli, I.; Pulyakhina, I.; Verwey, N.E.; Ariyurek, Y.; Laros, J.F.; 't Hoen, P.A.; Aartsma-Rus, A. Non-sequential and multi-step splicing of the dystrophin transcript. *RNA Biol.* **2016**, *13*, 290–305. [[CrossRef](#)]
36. Castle, J.C.; Zhang, C.; Shah, J.K.; Kulkarni, A.V.; Kalsotra, A.; Cooper, T.A.; Johnson, J.M. Expression of 24,426 human alternative splicing events and predicted cis regulation in 48 tissues and cell lines. *Nat. Genet.* **2008**, *40*, 1416–1425. [[CrossRef](#)]
37. Llorian, M.; Smith, C.W. Decoding muscle alternative splicing. *Curr. Opin. Genet. Dev.* **2011**, *21*, 380–387.
38. Tinsley, J.M.; Blake, D.J.; Davies, K.E. Apo-dystrophin-3: A 2.2 kb transcript from the DMD locus encoding the dystrophin glycoprotein binding site. *Hum. Mol. Genet.* **1993**, *2*, 521–524. [[CrossRef](#)]
39. Sibley, C.; Blazquez, L.; Ule, J. Lessons from non-canonical splicing. *Nat. Rev. Genet.* **2016**, *17*, 407–421. [[CrossRef](#)]
40. Le Rumeur, E.; Winder, S.J.; Hubert, J.F. Dystrophin: More than just the sum of its parts. *Biochim. Biophys. Acta* **2010**, *1804*, 1713–1722. [[CrossRef](#)]
41. Bies, R.D.; Phelps, S.F.; Cortez, M.D.; Roberts, R.; Caskey, C.T.; Chamberlain, J.S. Human and murine dystrophin mRNA transcripts are differentially expressed during skeletal muscle, heart, and brain development. *Nucleic Acids Res.* **1992**, *20*, 1725–1731. [[CrossRef](#)] [[PubMed](#)]
42. Górecki, D.C.; Monaco, A.P.; Derry, J.M.; Walker, A.P.; Barnard, E.A.; Barnard, P.J. Expression of four alternative dystrophin transcripts in brain regions regulated by different promoters. *Hum. Mol. Genet.* **1992**, *1*, 505–510. [[CrossRef](#)] [[PubMed](#)]
43. Austin, R.C.; Morris, G.E.; Howard, P.L.; Klamut, H.J.; Ray, P.N. Expression and synthesis of alternatively spliced variants of Dp71 in adult human brain. *Neuromuscul. Disord.* **2000**, *10*, 187–193. [[CrossRef](#)]
44. Aragón, J.; Martínez-Herrera, A.; Romo-Yáñez, J.; Ceja, V.; Azotla-Vilchis, C.; Siqueiros-Márquez, L.; Soid-Raggi, G.; Herrera-Salazar, A.; Montañez, C. Identification of Dp71 Isoforms Expressed in PC12 Cells: Subcellular Localization and Colocalization with  $\beta$ -Dystroglycan and  $\alpha$ 1-Syntrophin. *J. Mol. Sci.* **2016**, *58*, 201–209. [[CrossRef](#)]
45. Naidoo, M.; Anthony, K. Dystrophin Dp71 and the Neuropathophysiology of Duchenne Muscular Dystrophy. *Mol. Neurobiol.* **2020**, *57*, 1748–1767. [[CrossRef](#)] [[PubMed](#)]

46. Greener, M.J.; Sewry, C.A.; Muntoni, F.; Roberts, R.G. The 3'-untranslated region of the dystrophin gene-conservation and consequences of loss. *Eur. J. Hum. Genet.* **2020**, *10*, 413–420. [[CrossRef](#)] [[PubMed](#)]
47. Traverso, M.; Assereto, S.; Baratto, S.; Iacomino, M.; Pedemonte, M.; Diana, M.C.; Ferretti, M.; Broda, P.; Minetti, C.; Gazzero, E.; et al. Clinical and molecular consequences of exon 78 deletion in DMD gene. *J. Hum. Genet.* **2018**, *63*, 761–764. [[CrossRef](#)] [[PubMed](#)]
48. Ling, J.P.; Pletnikova, O.; Troncoso, J.C.; Wong, P.C. TDP-43 repression of nonconserved cryptic exons is compromised in ALS-FTD. *Science* **2015**, *349*, 650–655. [[CrossRef](#)]
49. Humphrey, J.; Emmett, W.; Fratta, P.; Isaacs, A.M.; Plagnol, V. Quantitative analysis of cryptic splicing associated with TDP-43 depletion. *Bmc Med. Genom.* **2017**, *10*, 38. [[CrossRef](#)]
50. Quesnel-Vallières, M.; Weatheritt, R.J.; Cordes, S.P.; Blencowe, B.J. Autism spectrum disorder: Insights into convergent mechanisms from transcriptomics. *Nat. Rev. Genet.* **2019**, *20*, 51–63. [[CrossRef](#)]
51. Li, Y.I.; Sanchez-Pulido, L.; Haerty, W.; Ponting, C.P. RBFOX and PTBP1 proteins regulate the alternative splicing of micro-exons in human brain transcripts. *Genome Res.* **2015**, *25*, 1–13. [[CrossRef](#)]
52. Ustianenko, D.; Weyn-Vanhenryck, S.M.; Zhang, C. Microexons: Discovery, regulation, and function. *Wires Rna* **2017**, e1418. [[CrossRef](#)] [[PubMed](#)]
53. Doorenweerd, N.; Mahfouz, A.; van Putten, M.; Kaliyaperumal, R.; T' Hoen, P.A.C.; Hendriksen, J.G.M.; Aartsma-Rus, A.M.; Verschuuren, J.J.G.M.; Niks, E.H.; Reinders, M.J.T.; et al. Timing and localization of human dystrophin isoform expression provide insights into the cognitive phenotype of Duchenne muscular dystrophy. *Sci. Rep.* **2017**, *7*, 12575. [[CrossRef](#)] [[PubMed](#)]
54. Daoud, F.; Angeard, N.; Demerre, B.; Martie, I.; Benyaou, R.; Leturcq, F.; Cossée, M.; Deburgrave, N.; Saillour, Y.; Tuffery, S.; et al. Analysis of Dp71 contribution in the severity of mental retardation through comparison of Duchenne and Becker patients differing by mutation consequences on Dp71 expression. *Hum. Mol. Genet.* **2009**, *18*, 3779–3794. [[CrossRef](#)]
55. Taylor, P.J.; Betts, G.A.; Maroulis, S.; Gilissen, C.; Pedersen, R.L.; Mowat, D.R.; Johnston, H.M.; Buckley, M.F. Dystrophin gene mutation location and the risk of cognitive impairment in Duchenne muscular dystrophy. *PLoS ONE* **2010**, *5*, e8803. [[CrossRef](#)]
56. Mamchaoui, K.; Trollet, C.; Bigot, A.; Negroni, E.; Chaouch, S.; Wolff, A.; Kandalla, P.K.; Marie, S.; Di Santo, J.; St Guily, J.L.; et al. Immortalized pathological human myoblasts: Towards a universal tool for the study of neuromuscular disorders. *Skelet. Muscle* **2011**, *1*, 34. [[CrossRef](#)]
57. Miro, J.; Laaref, A.M.; Rofidal, V.; Lagrèfeuille, R.; Hem, S.; Thorel, D.; Méchin, D.; Mamchaoui, K.; Mouly, V.; Claustres, M.; et al. FUBP1: A new protagonist in splicing regulation of the DMD gene. *Nucleic Acids Res.* **2015**, *43*, 2378–2389. [[CrossRef](#)]
58. Dobin, A.; Davis, C.A.; Schlesinger, F.; Drenkow, J.; Zaleski, C.; Jha, S.; Batut, P.; Chaisson, M.; Gingeras, T.R. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **2013**, *29*, 15–21. [[CrossRef](#)]
59. Holt, I.; Mittal, S.; Furling, D.; Butler-Browne, G.S.; Brook, J.D.; Morris, G.E. Defective mRNA in myotonic dystrophy accumulates at the periphery of nuclear splicing speckles. *Genes Cells* **2007**, *12*, 1035–1048. [[CrossRef](#)]

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.






© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

# Annexe 2



# RFC1 nonsense and frameshift variants cause CANVAS: clues for an unsolved pathophysiology

 Mehdi Benkirane,<sup>1,2,†</sup> Dylan Da Cunha,<sup>2,†</sup> Cecilia Marelli,<sup>3,4</sup> Lise Larrieu,<sup>1</sup> Mathilde Renaud,<sup>5</sup> Jessica Varilh,<sup>2</sup> Morgane Pointaux,<sup>1</sup>  David Baux,<sup>1</sup> Olivier Ardouin,<sup>1</sup> Charles Vangoethem,<sup>1</sup> Magali Taulan,<sup>2</sup> Benjamin Daumas Duport,<sup>6</sup> Anne Bergougnoux,<sup>1,2</sup> Anne-Gaelle Corbillé,<sup>7</sup> Mireille Cossée,<sup>1,2</sup> Raul Juntas Morales,<sup>3</sup>  Sylvie Tuffery-Giraud,<sup>2</sup> Michel Koenig,<sup>1,2</sup> Bertrand Isidor<sup>8</sup> and Marie-Claire Vincent<sup>1,2</sup>

†These authors contributed equally to this work.

Cerebellar ataxia, neuropathy and vestibular areflexia syndrome (CANVAS) is an inherited late-onset neurological disease caused by bi-allelic AAGGG pentanucleotide expansions within intron 2 of RFC1. Despite extensive studies, the pathophysiological mechanism of these intronic expansions remains elusive.

We screened by clinical exome sequencing two unrelated patients presenting with late-onset ataxia. A repeat-primer polymerase chain reaction was used for RFC1 AAGGG intronic expansion identification. RFC1 mRNA expression was assessed by quantitative reverse transcription–polymerase chain reaction.

We identified the first two CANVAS affected patients who are compound heterozygous for RFC1 truncating variants (p.Arg388\* and c.575delA, respectively) and a pathological AAGGG expansion. RFC1 expression studies in whole blood showed a significant reduction of RFC1 mRNA for both patients compared to three patients with bi-allelic RFC1 expansions.

In conclusion, this observation provides clues that suggest bi-allelic RFC1 conditional loss-of-function as the cause of the disease.

- 1 Department of Molecular Genetics, Institut Universitaire de Recherche Clinique (IURC), Montpellier Hospital, Montpellier, France
- 2 Genetics and Pathophysiology of NeuroMuscular Disorders, PhyMedExp Research Unit, CNRS, INSERM, University of Montpellier, Montpellier, France
- 3 Department of Neurology, Montpellier Hospital, Montpellier, France
- 4 Molecular Mechanisms of Neurodegenerative Dementia (MMDN), EPHE University of Montpellier, INSERM, Montpellier, France
- 5 Department of Medical Genetics, Nancy Hospital, Nancy, France
- 6 Department of Radiology, Nantes Hospital, Nantes, France
- 7 Department of Neurology, Nantes Hospital, Nantes, France
- 8 Department of Medical Genetics, Nantes Hospital, Nantes, France

Correspondence to: Professor Michel Koenig, MD, PhD Institut Universitaire de Recherche Clinique 641 Avenue du Doyen Gaston Giraud 34093 Montpellier Cedex 5, France E-mail: michel.koenig@inserm.fr

**Keywords:** CANVAS; RFC1; cerebellar ataxia; truncating variant; repeat expansion

**Abbreviations:** CANVAS = cerebellar ataxia, neuropathy and vestibular areflexia syndrome; E/E patients = patients with bi-allelic RFC1 AAGGG expansion

Received March 17, 2022. Revised June 13, 2022. Accepted July 03, 2022. Advance access publication July 27, 2022

© The Author(s) 2022. Published by Oxford University Press on behalf of the Guarantors of Brain. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com



## Introduction

Cerebellar ataxia, neuropathy and vestibular areflexia syndrome (CANVAS) is a recessive neurodegenerative disease, estimated to be one of the most common recessive ataxia.<sup>1</sup> Clinical features associated with CANVAS include late-onset cerebellar ataxia (80%), length-independent sensory neuronopathy (100%) and bilateral vestibular impairment (53%).<sup>2</sup> Less frequently, autonomic dysfunction, parkinsonism and pyramidal irritation are noticed, indicating overlap with multiple system atrophy.<sup>3</sup> Recently, bi-allelic expansions within an AluSx3 element in intron 2 of RFC1 have been identified as the cause of CANVAS.<sup>1,4</sup> To date, only bi-allelic expansions of the AAGGG pentanucleotide motif or an East Asian ACAGG motif have been reported to cause the disease, although other pathogenic motifs arrangement exist.<sup>5,6</sup> CANVAS pathophysiology has remained elusive so far, since previous studies failed to demonstrate direct loss-of-function of RFC1.<sup>1</sup> Herein we report two patients from unrelated families harbouring compound heterozygous pathogenic AAGGG expansion and RFC1 truncating mutation (p.Arg388\* and c.575delA, respectively). Studies of RFC1 messenger RNA expression in whole blood of these two patients compared to bi-allelic expansion patients provide pathophysiological clues for this enigmatic disease.

## Materials and methods

### Patient and clinical information collection

The proband of Family 1 (Patient F1-II.4) was part of a large ataxia cohort of 448 families, including 366 families previously reported.<sup>7</sup> The proband of Family 2 (Patient F2-II.1) was referred for RFC1 pathogenic expansion screening due to suspected CANVAS diagnosis. Both patients were analysed by clinical exome sequencing after exclusion of expansions causing spinocerebellar ataxias 1, –2, –3, –6, –7, –17, dentato-rubro-pallido-lusian atrophy and Friedreich's ataxia. Clinical data were collected by B.I. and A.-G.C. for Patient F1-II.4, and C.M. and R.J.M. for Patient F2-II.1, respectively. All individuals signed a written consent prior to genetic analysis. All procedures were conducted according to the guidelines of the Declaration of Helsinki.

### Clinical exome sequencing and data analysis

Clinical exome sequencing was performed using the TruSight One expanded kit ([www.illumina.com/trusightone](http://www.illumina.com/trusightone)) that targets ~6700 genes, most of them involved in human inherited diseases. Sequencing was performed using Illumina NextSeq technology with 150-bp paired-end reads (Montpellier Hospital next-generation sequencing platform). Sequence alignment, variant calling was performed following the Broad Institute's Genome Analysis Toolkit (GATK4) best practices, as described in a previous study.<sup>7</sup> Guidelines from the American College of Medical Genetics were used for clinical sequence interpretation.<sup>8</sup> BAM files were viewed with Integrative Genomics Viewer software (v.2.7.0 Broad Institute). Proband and available relatives from each family were tested for mutation segregation by Sanger sequencing.

### Repeat-primer PCR and long-range PCR analysis

Repeat-primer PCRs (RP-PCRs) were designed for each of three pentanucleotide repeat units (AAAAG/AAAGG/AAGGG). Fragment length analyses were performed on an ABI3500xl/Dx DNA Analyzer (Applied Biosystems®) and visualized through GeneMapper® v.4.0 (Applied Biosystems®). Presence of normal-sized

or expanded repeat alleles was confirmed by standard and long-range PCR (LR-PCR), using Phusion® High-Fidelity PCR Master Mix (New England Biolabs). Sanger sequencing was performed on expanded alleles to determine the repeat motif sequence. Primers, reagents and methods were designed as described in Cortese et al.<sup>1</sup>

### Total RNA Isolation and RFC1 transcripts analysis

Total RNA was purified after collection of peripheral whole blood using 2.5 ml of the PAXgene Blood RNA kit (Qiagen, Courtabouef, France) allowing intracellular RNA stabilization for quantification. First-strand cDNAs were synthesized from 300 ng of total RNA using random primers and the SuperScript III First Strand cDNA Synthesis Kit (Invitrogen). For quantitative reverse transcription PCR (RT-PCR), 1 µl of 1:6 diluted cDNA was amplified using the LightCycler 480 SYBR Green I Master Kit in a final volume of 5 µl containing 0.5 µM of primers. Three different primer pairs spanning RFC1 exons 2/3 (F1-R1) and exons 7/8 (F2-R2), previously described,<sup>1</sup> and exons 20/21 (F3-R3) were used for RFC1 amplification. Each sample was assayed in triplicate. Amplified transcripts were quantified using the comparative C<sub>T</sub> method normalized to GAPDH (glyceraldehyde-3-phosphate dehydrogenase) and RPLP0 (ribosomal protein, large, P0) (QuantiTect Primer Assay, Qiagen) as reference housekeeping transcripts, and presented as normalized fold expression change (2<sup>ΔΔCT</sup>). The differences in the expression of RFC1 transcripts between biological conditions were tested using the non-parametric Kruskal-Wallis statistical test with the GraphPad Prism 9 software (version 4.0). Technical replicate data for the three RFC1 exons were pooled for each condition (controls *n* = 36; E/E *n* = 27; Patient F2-II.1 *n* = 9; Patient F1-II.4 *n* = 9). A *P*-value ≤ 0.05 was considered statistically significant. Inclusion level of mutated RFC1 exons 6 and 10 was assessed by PCR amplification (35 cycles) using primers spanning exons 5/7 (F4-R4) and exons 9/11 (F5-R5), respectively. PCR products were analysed by agarose gel electrophoresis and Sanger sequencing (a list of primers in provided in [Supplementary Table 1](#)). Semi-quantitative allelic contribution to RFC1 transcript expression was estimated based on calibrated nucleotide peak height from Sanger sequence electropherograms. Electropherogram peak heights were adjusted with nucleotides flanking the mutations. Allele peak heights were estimated by comparison with the corresponding wild-type nucleotide of genomic DNA and cDNA from a control individual (set at the 100% level) and with wild-type and mutant nucleotides of genomic DNA from the patients (both set at the 50% level).

### Ethics declaration

The study was approved by the local ethics committee. All individuals signed a written consent before genetic analysis.

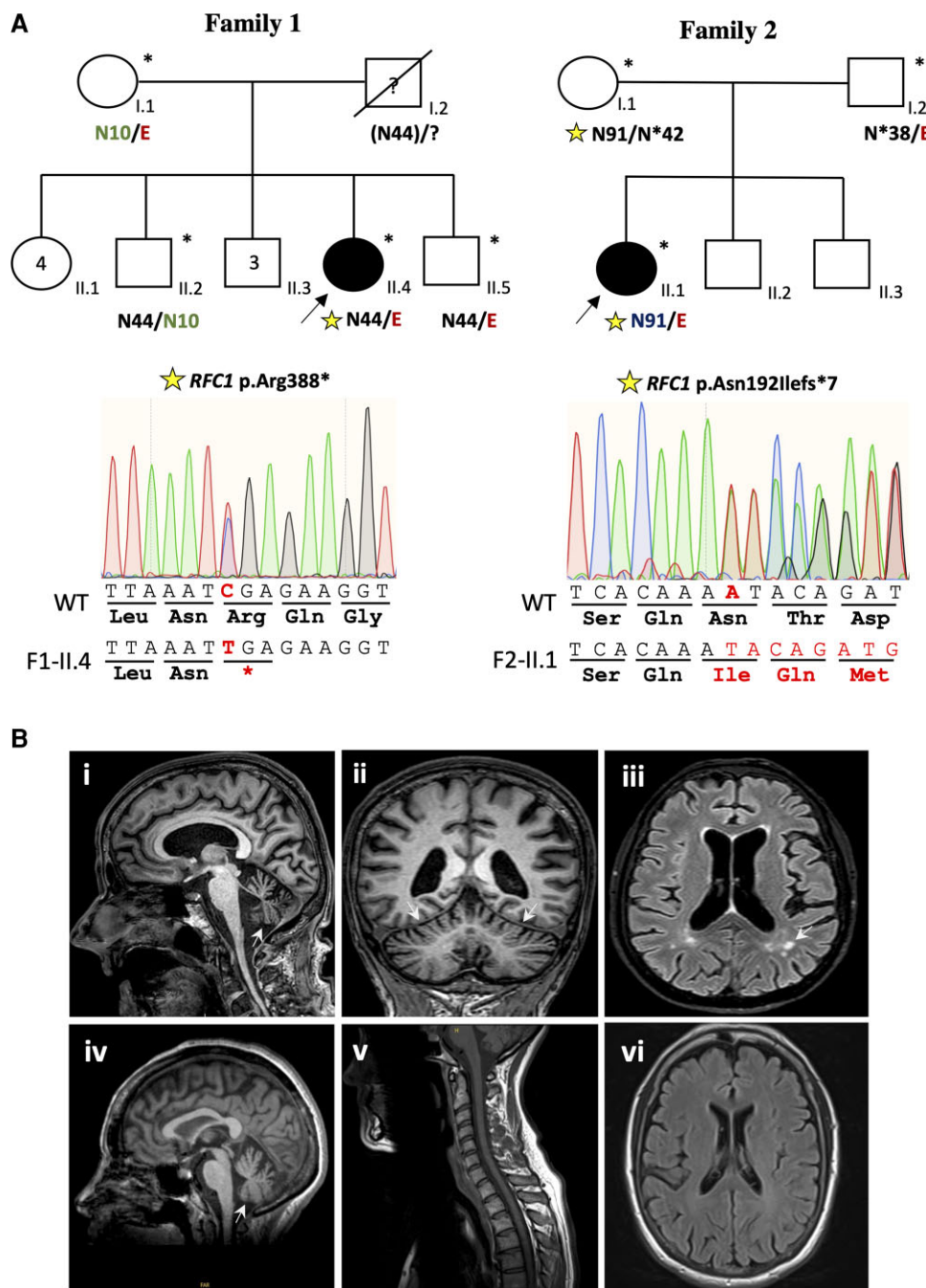
### Data availability

Readers can freely access all software used for this study by following the most recent version links mentioned in the 'Materials and methods' section. Patients' data can be made available upon reasonable request.

## Results

### Clinical findings

The proband of Family 1 (Patient F1-II.4) is a 68-year-old female born from unrelated parents who presented with progressive

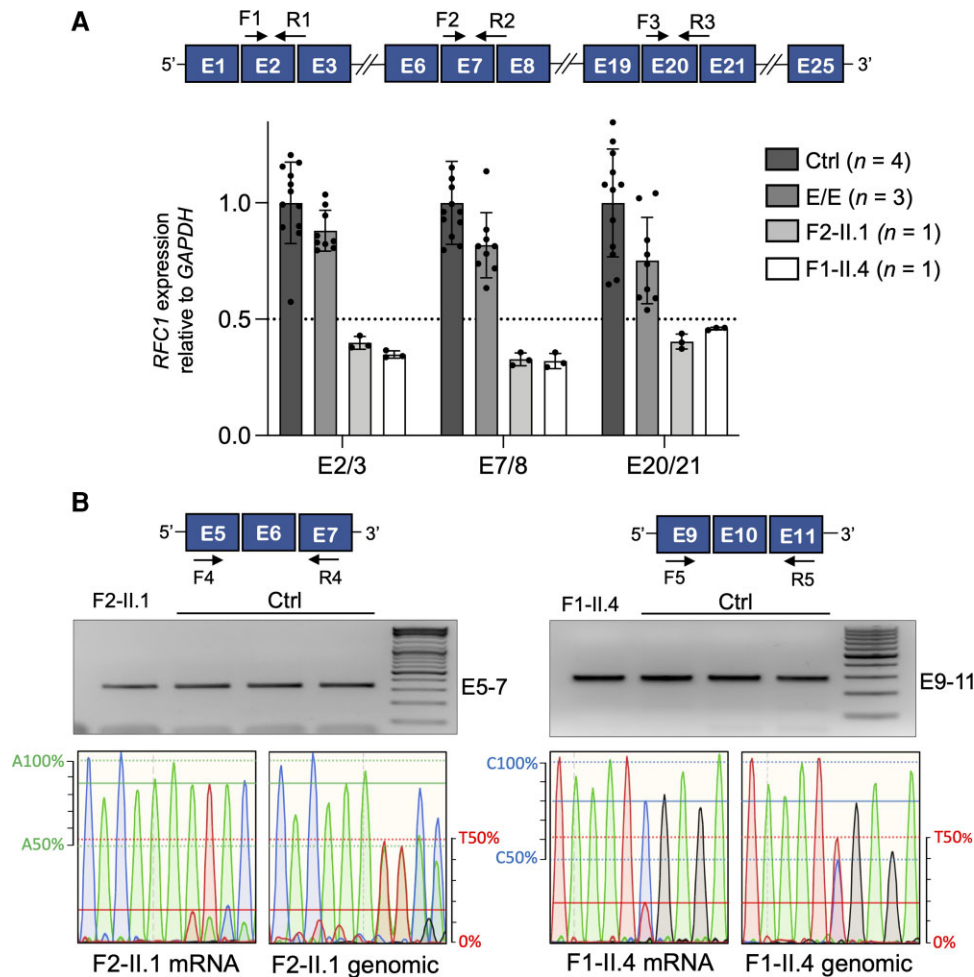


**Figure 1** Pedigrees and cerebral MRI of compound heterozygous patients. (A) Pedigree of Families 1 and 2 showing segregation of truncating RFC1 variants (p.Arg388\* and c.575delA, yellow star) and intronic AAGGG RFC1 expansion (E). N indicates normal AAAAG allele followed by the number of repeats. N\* indicates non-pathogenic expanded alleles associating different motifs followed by the number of repeats. (B) Cerebral and spinal cord MRI of both probands. Patient F1-II.4 (i–iii): sagittal (i) and coronal (ii) T<sub>1</sub>-weighted images of cerebral sections showing marked vermian and hemispheric cerebellar atrophy (arrows), respectively. Axial FLAIR image (iii) showing white matter hyperintensities (arrow). Patient F2-II.1 (iv–vi): sagittal T<sub>1</sub>-weighted images of cerebral sections (iv) and spinal cord section (v) showing moderate global cerebellar atrophy (arrow) and normal image, respectively. Normal axial FLAIR image (vi) of cerebral sections.

cerebellar and sensory ataxia at 45 years. Cerebellar signs included dysmetria, understandable dysarthria and gait imbalance leading to walking disability without support at 59 years. Neurological examination confirmed lower limbs hypoesthesia, length-dependent paraesthesia and loss of vibration sense. Deep tendon reflexes were brisk on the upper limbs and absent on the lower limbs. She later exhibited autonomic dysfunction with bladder sphincter complains, requiring permanent protection wearing.

Vestibulo–ocular reflexes were clinically preserved on last examination (at 66 years).

Similarly, a 61-year-old female (Patient F2-II:1) presented with length-dependent sensory impairment, including painful hyperesthesia and hypopallesthesia of the lower limbs since the age of 46 years, associated with chronic cough. Mild gait ataxia, brisk lower limbs reflexes and bilateral loss of vestibulo–ocular reflexes were noticed on first examination. Disease progression led to upper limb



**Figure 2 RFC1 mRNA transcript analysis.** (A) Schematic representation of RFC1 transcript (NM\_002913.5) and mapping of the primers (black arrows) used for RFC1 mRNA assessment by quantitative RT-PCR (F1-R1, F2-R2 and F3-R3) in controls ( $n = 4$ ), E/E affected individuals ( $n = 3$ ) and both reported patients (F2-II.1 and F1-II.4). GAPDH was used as reference housekeeping gene for normalization. Data of technical replicates for each tested individual are reported (black dots). The bar graphs show the mean  $\pm$  standard deviation. (B) Agarose gel electrophoresis of RT-PCR products of exons 5/7 (left) and of exons 9/11 (right) in controls (Ctrl  $n = 3$ ) and Patients F2-II.1 or F1-II.4 with the corresponding sequences of RFC1 exons 6 and 10 amplified from mRNA and genomic DNA (primers F4-R4, F5-R5). Electropherograms were normalized according to peak heights of nucleotides preceding the point mutations in exons 6 and 10. Contribution of both expanded and truncated alleles in RT-PCR product was calibrated against homozygotes (shown in Supplementary Fig. 2) and against the corresponding genomic DNA (50%). Scales used for level estimation of the mutant alleles are indicated.

sensory symptoms (49 years), dysphagia, dysarthria (57 years) and walking disability without support at 60 years. In both cases, electroneuromyography studies revealed normal motor nerve conduction and reduced nerves sensory amplitudes from the upper and lower limbs, consistent with sensory neuronopathy and dorsal root ganglia degeneration. Cerebral MRI showed global cerebellar atrophy for both patients, and non-specific diffuse white matter hyperintensities only for Patient F1-II.4 (Fig. 1B).

### Genetic investigations

A heterozygous RFC1 nonsense mutation c.1162C>T (p.Arg388\*) was identified by clinical exome sequencing in the proband of Family 1 (Patient F1-II.4) and was absent in her mother and two unaffected brothers (the father being deceased). An AAGGG expansion was subsequently found by RP-PCR in the proband, her mother and one unaffected brother, while a large normal (AAAAG)<sub>44</sub> repeat allele was identified in all three siblings but not in the mother (Fig. 1A and Supplementary Fig. 2). Flanking single-nucleotide polymorphism analysis confirmed the *de novo* occurrence of the

c.1162C>T variant, presumably on the father's allele, in cis with the large normal (AAAAG)<sub>44</sub> allele (Supplementary Table 2 and Supplementary Fig. 3). Similarly, RFC1 screening of the second patient (Patient F2-II.1) revealed a heterozygous AAGGG expansion inherited from her father, and a maternally inherited c.575delA (p.Asn192Ilefs\*7) frameshift mutation found by clinical exome sequencing (Fig. 1A).

### RFC1 transcript analyses

RFC1 mRNA quantification in whole blood of both probands, three patients with bi-allelic RFC1 AAGGG expansion (E/E) and four control individuals was assessed by quantitative RT-PCR. Using GAPDH as reference housekeeping gene, a marked reduction (>50%) of RFC1 transcripts level was observed for both compound heterozygous probands F2-II.1 and F1-II.4, while E/E patients did not show a significant change in RFC1 transcript expression compared to controls (Fig. 2A). These results were comparable for the three primer sets spanning exons 2–3, exons 7–8 and exons 20–21, and were confirmed with a second reference gene (RPLP0)

(Supplementary Fig. 1). Statistical analyses of all data-points (shown separate for each individual in Supplementary Fig. 1) indicate that RFC1 mRNA reduction in proband F1-II.4 was highly significant compared to controls ( $P < 0.01$ ) using both reference housekeeping genes for normalization. For proband F2-II.1, this reduction was also highly significant when normalized with GAPDH ( $P = 0.003$ ) but only suggestive ( $P = 0.142$ , ns) when RPLP0 was used as reference gene (Supplementary Fig. 1). RFC1 transcripts levels were also significantly reduced ( $P < 0.05$ ) in probands F1-II.4 and F2-II.1 compared to E/E patients. We next sought to estimate the contribution of each allele to this reduction by visualization of the truncating allele on sequences of the corresponding complementary DNA fragments. Semi-quantitative estimation of the respective transcripts by comparison with genomic DNA sequencing peak heights (Supplementary Fig. 2) indicates that the c.1162C>T and c.575delA alleles account for  $\approx 20$  and 15% of total RT-PCR products, respectively, reflecting a significant expression reduction of both truncating alleles down to 10% (Fig. 2B). Transcript level of the AAGGG expanded allele was deduced to be on the normal range for both patients (Patients F1-II.4 and F2-II.1) compared to controls and E/E patients. RT-PCR analysis of fragments, spanning exons 5 to 7 and exons 9 to 11, did not show abnormal splicing (Fig. 2B and Supplementary Fig. 4).

## Discussion

We expand the mutational spectrum of CANVAS causing variants by reporting the first two patients who are compound heterozygous for the pathogenic AAGGG expansion and an RFC1 mutation (p.Arg388\* and c.575delA, respectively) and who present with a late-onset form of cerebellar ataxia consistent with CANVAS diagnosis. Age of onset (45 and 46 years) and clinical presentation appear unremarkable from that of E/E affected individuals (average  $54 \pm 9$  years, range 35–73 years).<sup>1,2,9</sup> Quantification of RFC1 transcripts by quantitative RT-PCR in whole blood showed a marked reduction of expression of the p.Arg388\* and c.575delA alleles, reflecting efficient nonsense-mediated mRNA decay of the truncating alleles. E/E affected individuals did not show significant reduction mRNA expression, in accordance with previous results obtained by Cortese and colleagues<sup>1</sup> in several other tissues.

### RFC1 loss-of-function evidence?

RFC1 encodes the large subunit of Replication Factor C1, a heteropentameric AAA ATPase that acts as PCNA clamp-loader, particularly involved in mismatch repair, nucleotide excision repair and translesion synthesis.<sup>10,11</sup> Interestingly, numbers of neurodegenerative diseases leading to ataxia and neuropathy are caused by genes involved in DNA repair pathways<sup>12,13</sup> (e.g. ATM, MRE11, APTX, PNKP, ERCC6 and ERCC8), reflecting cerebellar and sensory neurons intolerance to DNA damage. Pathophysiology of CANVAS remains elusive since no detectable impact was observed on RFC1 gene expression or splicing in peripheral tissues or post-mortem brain samples (including cortex and cerebellum).<sup>1</sup> In addition, no RNA foci of either the sense (AAGGG) or antisense (CCCUU) repeated unit could be detected in any of the examined tissues.<sup>12,14</sup> Identification of truncating RFC1 variants leading to CANVAS is suggestive of a bi-allelic loss-of-function of RFC1 as the cause of the disease. The normal RFC1 mRNA transcript level found in leucocytes and other tissues of E/E affected individuals suggests an incomplete impact of the intronic RFC1 AAGGG expansion on RFC1 transcriptional regulation, by contrast with similar bi-allelic

non-coding expansion diseases such as Friedreich's ataxia.<sup>15</sup> Rather, the AAGGG intronic expansion may only conditionally interfere with RFC1 expression in stress conditions, not seen in native leucocytes or in post-mortem tissues (where fragile cells have already disappeared). Stress conditions may relate to DNA repair at the RFC1 locus and may account for the specific pathogenicity of AAGGG expansions, as opposed to AAAGG and AAAAG expansions, possibly in relation with the GC content of the expanded motifs.<sup>16</sup> Alternatively, an RNA toxic pathogenic mechanism for the AAGGG expansion cannot be formally excluded, as was suggested by foci detection in SHY-H5 cells transfected with (AAGGG)<sub>54</sub> oligonucleotides, and by the identification of intron 2 retention in RFC1 transcripts across different tissues of patients with bi-allelic expansion.<sup>1</sup> In this case, one normal allele, absent from our compound heterozygous probands, would be sufficient to balance the recessive toxic gain of function of the expanded allele. Identification of additional CANVAS patients with non-expanded mutations is needed to unravel the puzzling pathophysiology of this intriguing disease.

## Acknowledgements

We sincerely thank the patients and families for participation to this study. We thank Pr. Vincent Procaccio for contribution to the etiological investigations.

## Funding

This work was supported in part by the French patients' association 'Connaître les Syndromes Cérébelleux' (CSC). D.D.C. is supported by a fellowship from the French Association against Myopathies (AFM-Telethon).

## Competing interests

The authors report no competing interests

## Supplementary material

Supplementary material is available at Brain online.

## References

1. Cortese A, Simone R, Sullivan R, et al. Biallelic expansion of an intronic repeat in RFC1 is a common cause of late-onset ataxia. *Nat Genet.* 2019;51:649–658.
2. Cortese A, Tozza S, Yau WY, et al. Cerebellar ataxia, neuropathy, vestibular areflexia syndrome due to RFC1 repeat expansion. *Brain.* 2020;143:489–490.
3. Träschütz A, Cortese A, Reich S, et al. Natural history, phenotypic spectrum, and discriminative features of multisystemic RFC1 disease. *Neurology.* 2021; 96:e1369–e1382.
4. Rafehi H, Szmulewicz DJ, Bennett MF, et al. Bioinformatics-based identification of expanded repeats: A non-reference intronic pentamer expansion in RFC1 causes CANVAS. *Am J Hum Genet.* 2019;105:151–165.
5. Scriba CK, Beecroft SJ, Clayton JS, et al. A novel RFC1 repeat motif (ACAGG) in two Asia-pacific CANVAS families. *Brain.* 2020;143:2904–2910.
6. Beecroft SJ, Cortese A, Sullivan R, et al. A Maori-specific RFC1 pathogenic repeat configuration in CANVAS, likely due to a founder allele. *Brain.* 2020;143:2673–2680.

7. Benkirane M, Marelli C, Guissart C, et al. High rate of hypomorphic variants as the cause of inherited ataxia and related diseases: Study of a cohort of 366 families. *Genet Med*. 2021;23:2160–2170.
8. Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*. 2015;17:405–424.
9. Huin V, Coarelli G, Guemy C, et al. Motor neuron pathology in CANVAS due to RFC1 expansions. *Brain*. 2021;145:2121–2132.
10. Arbel M, Choudhary K, Tfilin O, Kupiec M. PCNA loaders and unloaders—One ring that rules them all. *Genes (Basel)*. 2021;12:1812.
11. Umar A, Buermeier AB, Simon JA, et al. Requirement for PCNA in DNA mismatch repair at a step preceding DNA resynthesis. *Cell*. 1996;87:65–73.
12. Depienne C, Mandel JL. 30 Years of repeat expansion disorders: What have we learned and what are the remaining challenges? *Am J Hum Genet*. 2021;108:764–785.
13. Renaud M, Tranchant C, Koenig M, Anheim M. Autosomal recessive cerebellar ataxias with elevated alpha-fetoprotein: Uncommon diseases, common biomarker. *Mov Disord*. 2020;35:2139–2149.
14. Malik I, Kelley CP, Wang ET, Todd PK. Molecular mechanisms underlying nucleotide repeat expansion disorders. *Nat Rev Mol Cell Biol*. 2021;22:589–607.
15. Li Y, Lu Y, Polak U, et al. Expanded GAA repeats impede transcription elongation through the FXN gene and induce transcriptional silencing that is restricted to the FXN locus. *Hum Mol Genet*. 2015;24:6932–6943.
16. Sznajder ŁJ, Thomas JD, Carrell EM, et al. Intron retention induced by microsatellite expansions as a disease biomarker. *Proc Natl Acad Sci U S A*. 2018;115:4234–4239.