



**HAL**  
open science

# Fonctions moléculaires des hélicases ARN DDX5 et DDX17 dans la biologie du muscle dans un contexte sain et pathologique

Micaela Polay Espinoza

► **To cite this version:**

Micaela Polay Espinoza. Fonctions moléculaires des hélicases ARN DDX5 et DDX17 dans la biologie du muscle dans un contexte sain et pathologique. Sciences agricoles. Université Claude Bernard - Lyon I, 2014. Français. NNT : 2014LYO10041 . tel-00988051

**HAL Id: tel-00988051**

**<https://theses.hal.science/tel-00988051>**

Submitted on 7 May 2014

**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.



N° Ordre : 41 - 2014

Année : 2014

THESE DE L'UNIVERSITE DE LYON  
Délivrée par  
L'UNIVERSITE CLAUDE BERNARD LYON 1

ECOLE DOCTORALE BMIC

DIPLOME DE DOCTORAT  
(arrêté du 7 Août 2006)

Soutenue publiquement le 21 mars 2014

Par

**Micaela POLAY ESPINOZA**

TITRE :

Fonctions moléculaires des hélicases ARN DDX5 et DDX17 dans la  
biologie du muscle dans un contexte sain et pathologique

JURY :

Joëlle MARIE	Directrice de Recherche, Inserm	Rapportrice
Sylvie TUFFERY-GIRAUD	Ingénieure de Recherche, UM1	Rapportrice
Philippe BOUVET	Professeur des Universités, ENS	Examinateur
Caroline MOYRET-LALLE	Maître de Conférences, UCBL1	Examinatrice
Cyril BOURGEOIS	Chargé de Recherche, Inserm	Examinateur - Encadrant
Didier AUBOEUF	Directeur de Recherche, Inserm	Directeur de thèse



# Remerciements

Tout d'abord, je tiens à remercier les membres de mon jury de thèse : Joëlle Marie, Sylvie Tuffery-Giraud, Philippe Bouvet et Caroline Moirey-Lalle pour avoir accepté d'évaluer mon travail de thèse. Je remercie tout particulièrement mes rapportrices Joëlle Marie et Sylvie Tuffery-Giraud qui ont donné de leur temps pour lire et porter un regard critique sur mon manuscrit.

Ma gratitude va également aux membres de mon Comité de Suivi de Thèse : Laurent Shaeffer, Vincent Lacroix et Alexandre Méjat qui, tout au long de ces années, ont suivi mon travail et m'ont conseillé dans l'évolution de ce dernier.

Je remercie mon directeur de thèse Didier Auboeuf de m'avoir accueilli au sein de son équipe pendant près de 4 ans. En particulier, je lui suis reconnaissante de m'avoir permis de découvrir autant d'aspects d'un laboratoire de biologie moléculaire. En tant que bio-informaticienne, je le remercie de m'avoir laissé l'opportunité de passer à la paillasse. J'ai beaucoup appris par cette expérience et elle m'a permis, je pense, de faire de moi une meilleure bio-informaticienne.

Bien sûr, je remercie très chaleureusement Cyril Bourgeois. J'ai bien conscience que je lui dois beaucoup aussi bien quant à la gestion du projet « différenciation », la rédaction de ce manuscrit que plus globalement dans mon expérience de thèse.

Je remercie les membres de l'équipe « Epissage alternatif et progression tumorale » : Fatima, Amandine, Marie-Pierre, Clara, Emilie, Camille, Etienne, Hussein, Léo, FO... Grâce à eux, l'ambiance du labo a toujours été bonne. Je remercie tout particulièrement Fatima, Amandine, Hussein et Leo pour nos discussions (profondes et beaucoup moins profondes!). Une pensée spéciale pour Etienne, mon « compagnon de thèse » avec qui nous avons partagé nos incertitudes et remises en question tout au long de ces années. Je n'oublie bien évidemment pas les anciens membres de l'équipe qui ont aussi beaucoup compté dans mon expérience de doctorante : Sophie, Simon et Eleonora ; j'espère que nous continuerons à garder contact pendant encore longtemps!

Sur un plan plus personnel, je remercie ma mère qui a toujours été là pour moi, à



---

la fois écoute et soutien inamovible de ma vie. Je sais que pour cette amoureuse des études, cette thèse est une réalisation importante et c'est la raison principale de la fierté que j'en tire. Merci d'avoir toujours été là et d'être un esprit libre et « fou ». Last but not least, merci à Guillaume Boule. Je sais que je suis quelqu'un qui peut être difficile, en particulier en période de stress. Merci pour ton soutien, ta présence, ta personnalité tout au long de ses années.

Enfin, je remercie bien sûr l'AFM pour m'avoir fait confiance en financant mon doctorat et en rendant possible la réalisation de ce projet.

# Résumé

Les ARN hélicases homologues à boîte DEAD DDX5 et DDX17 sont des protéines « multi-tâches », elles sont impliquées dans de nombreuses étapes de la régulation du métabolisme des ARNs dont la transcription, l'épissage et la dégradation des ARNs. Lors de processus biologiques complexes tels que la différenciation myogénique, les programmes d'expression génique sont profondément modifiés. Durant mon travail de thèse, j'ai contribué à montrer que DDX5 et DDX17 sont des protéines orchestratrices de la différenciation en coordonnant de manière directe et dynamique plusieurs niveaux de régulation génique. DDX5 et DDX17 contrôlent l'activité du facteur de transcription MyoD, régulateur majeur de la myogénèse ainsi que des microARNs spécifiques du muscle miR-1 et miR-206. Ceux-ci ciblent et régulent en retour l'expression de DDX5 et DDX17 mettant en place une boucle de rétrocontrôle négative induisant la diminution d'expression de ces deux protéines au cours de la différenciation. Enfin, cette diminution d'expression permet la mise en place d'un programme d'épissage participant à l'acquisition de phénotypes morphologiques des cellules différenciées. D'un point de vue mécanistique, il apparaît qu'un sous-groupe des événements d'épissage régulés durant la différenciation est contrôlé par la coopération de DDX5 et DDX17 avec le facteur d'épissage hnRNP H/F. Cette coopération assure la reconnaissance des sites 5' d'épissage en amont de séquence riches en GC. D'autre part, DDX5 a aussi été impliqué dans un contexte pathologique du muscle. Cette hélicase interagit avec la mutation responsable de la Dystrophie Myotonique de type 1 (DM1). Durant ma thèse, j'ai produit des résultats préliminaires suggérant un rôle de DDX5 dans la mise en place des défauts d'épissage observés dans cette pathologie.



# Table des matières

<b>Principales abbréviations utilisées</b>	<b>15</b>
<b>Table des figures</b>	<b>18</b>
<b>Liste des tableaux</b>	<b>19</b>
<b>Avant-propos</b>	<b>23</b>
<b>1 Introduction</b>	<b>25</b>
I L'épissage . . . . .	27
A Préambule . . . . .	27
B Importance de l'épissage alternatif . . . . .	28
C Mécanisme de la réaction d'épissage . . . . .	30
1 Sites de reconnaissance des exons et introns . . . . .	30
2 Réaction d'épissage . . . . .	31
3 Assemblage du spliceosome . . . . .	32
D Epissage alternatif et régulation . . . . .	36
1 L'épissage alternatif . . . . .	36
2 Régulation de l'épissage . . . . .	38
a Les éléments cis-régulateurs . . . . .	38
b Les protéines trans-régulatrices SRs et hnRNPs . . . . .	40
c Autres protéines trans-régulatrices . . . . .	42
d Couplage à la transcription . . . . .	44
e Influence de la structure chromatinienne . . . . .	47
f Influence des structures secondaires . . . . .	49
E Analyse à large échelle et émergence du « code de l'épissage »	50
F Epissage et pathologies . . . . .	51
G Conclusion . . . . .	53

II	Les hélicases ARN DDX5 et DDX17 . . . . .	54
A	Préambule . . . . .	54
B	Les hélicases à boîte DEAD . . . . .	55
1	Structure générale . . . . .	55
2	Mécanisme de remodelage de l'ARN double brin . . . . .	58
3	Diversité de fonctions . . . . .	59
C	DDX5 et DDX17 . . . . .	60
1	Structure générale . . . . .	61
2	Expression et localisation . . . . .	62
3	Fonctions moléculaires . . . . .	63
a	Rôle dans la transcription . . . . .	63
b	Rôle dans l'épissage . . . . .	66
c	Rôle dans la biogénèse des micro-ARNs . . . . .	69
d	Rôle dans l'export cytoplasmique . . . . .	70
e	Rôle dans le Non-sense Mediated Decay (NMD) . . . . .	71
f	Rôle dans la biogénèse des ribosomes . . . . .	71
g	Modifications post-traductionnelles . . . . .	71
	Phosphorylation . . . . .	72
	Sumoylation . . . . .	73
	Acétylation . . . . .	73
4	Fonctions cellulaires . . . . .	74
a	Rôle dans le contrôle de la prolifération . . . . .	74
b	Rôle dans la différenciation cellulaire . . . . .	75
5	Rôle dans la pathologie . . . . .	76
D	Conclusion . . . . .	77
III	La myogénèse . . . . .	78
A	Préambule . . . . .	78
B	Vue d'ensemble . . . . .	79
C	La lignée murine C2C12 : un modèle d'étude de la myogénèse . . . . .	81

D	Les MRFs contrôlent le programme transcriptionnel de la myogénèse . . . . .	83
1	Structure et fonctions des MRFs . . . . .	83
2	Inhibition du programme de transcription myogénique . . . . .	85
3	Programme transcriptionnel myogénique et régulation épigénétique . . . . .	86
E	Autres mécanismes de la régulation génique . . . . .	87
1	Les micro-ARNs . . . . .	87
2	L'épissage alternatif . . . . .	90
IV	DM1 . . . . .	93
A	Préambule . . . . .	93
B	Différentes formes . . . . .	94
1	Forme congénitale . . . . .	94
2	Forme infantile . . . . .	95
3	Forme adulte . . . . .	95
4	Forme tardive . . . . .	96
C	Aspects cliniques . . . . .	96
D	Aspects moléculaires . . . . .	97
1	Mutation pathologique . . . . .	97
2	Toxicité et gain de fonction de l'ARN pathologique . . . . .	99
a	Dérégulation de l'épissage . . . . .	100
	Perte de fonction de MBNL1 . . . . .	100
	Gain de fonction de CUGBP1 . . . . .	102
	Perturbation d'autres facteurs d'épissage . . . . .	102
	La DM1 : une spliceopathie . . . . .	103
b	Autres dérégulations . . . . .	107
	Altération de l'expression génique . . . . .	108
	Altération du métabolisme des micro-ARNs . . . . .	108
	Altération de la traduction . . . . .	109
	Transcription bidirectionnelle . . . . .	109
	Expression de petits ARN CUG . . . . .	109

	Traduction RAN . . . . .	109
E	Stratégies thérapeutiques . . . . .	112
<b>2</b>	<b>Objectifs</b>	<b>113</b>
<b>3</b>	<b>Résultats</b>	<b>117</b>
I	Publication en révision dans Molecular Cell . . . . .	119
II	Discussion . . . . .	173
A	Mécanisme de régulation de l'épissage . . . . .	174
B	Remodelage du cytosquelette . . . . .	175
C	Perspectives . . . . .	176
<b>4</b>	<b>Résultats complémentaires</b>	<b>179</b>
I	Préambule . . . . .	181
II	Problématiques . . . . .	182
III	DDX5/DDX17 et Mbnl1 . . . . .	183
A	Régulation de l'expression de MBNL1 par DDX5/DDX17 . . . . .	183
B	Localisation de MBNL1 . . . . .	185
C	Interaction DDX5 entre MBNL1 . . . . .	186
D	Expression de MBNL1 durant la différenciation myogénique . . . . .	187
IV	Rôle de DDX5/DDX17 dans les défauts d'épissage observés dans la DM1 ? . . . . .	188
A	Démarche . . . . .	188
B	1ère catégorie d'exons . . . . .	190
C	2ème catégorie d'exons . . . . .	193
D	Perspectives . . . . .	195
<b>5</b>	<b>Annexes</b>	<b>199</b>
I	Préambule . . . . .	201
II	Les puces ADN « tout exon » : généralités et méthodologie . . . . .	201
A	Généralités . . . . .	201
B	Annotations . . . . .	204
C	Système d'analyse mis en place . . . . .	205

---

1	Prétraitement . . . . .	206
2	Traitement . . . . .	207
	a Niveau de régulation globale du gène . . . . .	208
	b Niveau de régulation de l'exon . . . . .	209
III	Articles . . . . .	211
	A Publication dans NSMB . . . . .	211
	B Publication dans NAR . . . . .	221
	C Publication dans Genome research . . . . .	233
	<b>Bibliographie</b>	<b>245</b>





# Principales abbréviations utilisées

3'UTR	3' Untranslated Region
5'UTR	5' Untranslated Region
ADN	Acide désoxyribonucléique
ADNc	acide désoxyribonucléique complémentaire
Ago	protéines « Argonaute »
ARN	Acide Ribonucléique
ARNdb	Acide RiboNucléique double brin
ARN Pol II	ARN Polymérase II
ARNi	mécanisme d'ARN interférence
ARNm	ARN messenger
ARNt	ARN de Transfert
ATP	Adénosine Triphosphate
ATPase	Adénine Tri-Phosphatase
BBP	Branch-Point Bridging Protein
bp	base pairs
bHLH	Basic Helix-Loop-Helix
CELF	CUGBP- and ETR3-like factor
CBF	C-Promoter Binding Factor
ChIP	Chromatin Immunoprécipitation
CLIP	Cross-Linking Immunoprecipitation
CTD	C-Terminal Domain
CTCF	CCTC Binding Factor
CUGBP1	CUG Binding Protein 1
DAPI	4',6'-diamidino-2-phénylindole
DM1	Dystrophie Myotonique de type 1
DMPK	Dystrophia Myotonica Protein Kinase
ESE	Exonic Splicing Enhancer
ESS	Exonic Splicing Silencer

---

FC	Fold Change
FISH	Fluorescent In Situ Hybridization
GTPase	guanosine triphosphatase
HAT	Histone Acetyltransferase
HDAC	Histone Deacetylase
hnRNP	heterogeneous nuclear Ribonucleoprotein
ISE	Intronic Splicing Enhancer
ISS	Intronic Splicing Silencer
KO	Knock-Out
MBNL1	Muscleblind-like 1
Mef	Myocyte Enhancer Factor
MHC	Myosin Heavy Chain
miARN	micro ARN
MPS	Metaprobese
MRF	Myogenic Regulatory Factor
MyoD	Myogenic Differentiation
NMD	Non-sense mRNA Decay
nPTB	neuronal Polypyrimidine Tract Binding Protein
NTP	Nucleotide Tri Phosphate
ORF	Open Reading Frame
PLA	Proximity Ligation Assay
PKC	Protein Kinase C
pré-mARN	ARNm précurseur
pré-miARN	miARN précurseur
pri-miARN	miARN primaire
PS	Probeset
PTB	Polypyrimidine Tract Binding Protein
PTC	codon stop prématuré
RISC	RNA induced silencing complex
RIP	RNA Immunoprecipitation
RNP	RiboNucleoProtein

RRM	RNA Recognition Motif
RS	Arginine-Serine-Rich domain
RT-PCR	Retro-transcriptase Polymerase Chain Reaction
shRNA	small hairpin RNA
SI	Splicing Index
siARN	small interfering RNA (petit ARN interférant)
SRA	steroid receptor RNA activator
TBP	TATA box binding protein
U2AF	U2 Auxilliary Factor



# Table des figures

1	Des « switch » d'épissage contrôlent la différenciation cellulaire . . . . .	29
2	Sites de reconnaissance des introns et exons . . . . .	30
3	Réaction d'épissage . . . . .	31
4	Assemblage dynamique du spliceosome durant la réaction d'épissage .	33
5	Hélicases ARN durant la réaction d'épissage . . . . .	35
6	Différents type d'épissage alternatif . . . . .	37
7	Différences entre exons constitutifs et alternatifs . . . . .	38
8	Séquences cis-régulatrices de l'épissage . . . . .	39
9	Séquences et facteurs de régulation de l'épissage . . . . .	41
10	Recrutement de facteurs d'épissage par la machinerie transcriptionnelle	45
11	Vitesse d'élongation de l'ARN polymérase II et l'épissage alternatif .	46
12	Modifications d'histones et épissage alternatif . . . . .	48
13	Impact des structures secondaires sur l'épissage alternatif . . . . .	49
14	Superfamilles SF1 et SF2 des hélicases . . . . .	55
15	Motifs des hélicase à boîte DEAD . . . . .	56
16	Remodelage de l'ARN double brin par les hélicases ARN . . . . .	59
17	Domaines des hélicases ARN hautement conservées DDX5 et DDX17	62
18	DDX5 et DDX17 : protéines régulatrices de la transcription . . . . .	65
19	Régulation de l'épissage alternatif par DDX5 et/ou DDX17 . . . . .	68
20	Rôles de DDX5/DDX17 dans la biogénèse des micro-ARNs . . . . .	70
21	Modifications post-traductionnelles de DDX5 et impact sur ses interactions et fonctions . . . . .	72
22	Multiplés fonctions moléculaires des hélicases ARN DDX5 et DDX17	74
23	Rôle de DDX5 et DDX17 dans la différenciation myogénique . . . . .	76
24	Etapes de la myogénèse . . . . .	80
25	Myogénèse des C2C12 . . . . .	82
26	Structure des MRFs . . . . .	84

27	Régulation épigénétique des promoteurs myogéniques . . . . .	87
28	MyomiRs . . . . .	89
29	Motifs d'épissage enrichis durant la différenciation musculaire . . . . .	92
30	Mutation pathologique de la DM1 . . . . .	94
31	Répétitions de triplets CUG dans des myoblastes humains . . . . .	99
32	Altérations de l'épissage alternatif dans la DM1 . . . . .	107
33	Mécanismes moléculaires de la pathogénèse de la DM1 . . . . .	111
34	Modèle proposé du rôle de DDX5 dans la DM1 . . . . .	182
35	DDX5/DDX17 régulent l'expression de MBNL1 . . . . .	184
36	La déplétion de DDX5/DDX17 n'impacte pas la localisation de MBNL1. . . . .	186
37	Interaction de DDX5 et MBNL1 révélé par Proximity Ligation Assay (In situ PLA) . . . . .	187
38	Expression protéique de DDX5/DDX17 et MBNL1 durant la différenciation . . . . .	188
39	Evénements d'épissage régulés au cours de la différenciation et inversement régulés lors de la déplétion de DDX5/DDX17 et dans la DM1 . . . . .	192
40	Evénements d'épissage régulés au cours de la différenciation, de la même manière lors de la déplétion de DDX5/DDX17 et inversement dans la DM1 . . . . .	195
41	Différences de conception entre les puces classiques et les puces « tout exon » . . . . .	203
42	Workflow d'analyse au niveau du gène . . . . .	209
43	Workflow d'analyse au niveau de l'exon . . . . .	210

# Liste des tableaux

1	Liste des protéines SRs et hnRNPs régulatrices de l'épissage . . . . .	42
2	Liste des protéines auxiliaires régulatrices de l'épissage . . . . .	43
3	Evénements d'épissage impactés par les structures secondaires de l'ARN	50
4	Partenaires transcriptionnels identifiés de DDX5 et DDX17 . . . . .	64
5	Tableau récapitulatif des événements d'épissage dérégulés dans la DM1	105
6	Gènes dont l'épissage est régulé au cours de la différenciation et inversement régulés lors de la déplétion de DDX5/DDX17 et dans la DM1 . . . . .	191
7	Gènes dont l'épissage est régulé au cours de la différenciation ainsi que lors de la déplétion de DDX5/DDX17 et inversement régulés dans la DM1 . . . . .	194





# Avant-propos

L'équipe de Didier Auboeuf, dans laquelle j'ai effectué ma thèse, s'intéresse à l'épissage alternatif. Cette étape de la maturation des ARNs apparaît déterminante dans le contrôle de l'expression génique. Il est maintenant admis qu'à la manière des programmes transcriptionnels, il existe des programmes d'épissage participant activement au déroulement de processus biologiques sains mais aussi pathologiques. Les hélicases ARN sont des protéines multitâches impliquées dans le contrôle de toutes les étapes de la régulation génique de la transcription à la traduction en passant par l'épissage. L'équipe de Didier Auboeuf s'intéresse tout particulièrement aux hélicases ARN DDX5 et DDX17. Ces deux protéines homologues ont été essentiellement étudiées dans le cancer car il apparaît qu'elles participent à la progression tumorale. Mais DDX5 et DDX17 ont été aussi démontrées être des co-activatrices transcriptionnelles de MyoD, facteur de transcription majeur de la myogénèse. Ayant été impliqué dans d'autres étapes de la régulation des gènes et tout particulièrement l'épissage, un projet visant à déterminer le rôle à large échelle de DDX5 et DDX17, notamment dans l'épissage, lors de la différenciation myogénique a donc logiquement vu le jour au sein de l'équipe.

C'est dans ce contexte que j'ai débuté ma thèse en bioinformatique. Mon objectif premier a donc été d'élaborer un système d'analyse dédié aux puces ADN « tout exon » permettant de connaître l'impact à large échelle de la déplétion de DDX5 et DDX17 durant la myogénèse de cellules murines de la lignée C2C12.

Les résultats obtenus ont permis de confirmer leur activité co-transcriptionnelle mais ont aussi mis en valeur leur rôle de régulateur de l'épissage et plus globalement d'« orchestratrices » de différents niveaux de régulation des gènes durant la myogénèse. Le projet a donc rapidement pris une tournure plus « biologique » que « bioinformatique ». J'ai alors été amenée à assurer les manipulations expérimentales du projet. C'est pourquoi ce manuscrit s'intéressera plus aux mécanismes moléculaires dans lesquelles sont impliquées DDX5 et DDX17 qu'à la méthodologie bioinformatique mise en place.

D'autre part, DDX5 a été impliquée dans un contexte musculaire pathologique : la Dystrophie Myotonique de type 1 (DM1). Celle-ci est caractérisée par de multiples défauts d'épissage résultant en grande partie de la séquestration du facteur d'épissage MBNL1 au niveau de la mutation responsable de la maladie. Il a été montré que DDX5, de par son activité hélicase, aide à la séquestration pathologique de MBNL1. Il existe des multiples jeux de données de puces « tout exon » réalisées sur des échantillons de patients. J'ai donc été amenée à les analyser afin d'établir une série d'événements d'épissage candidats dérégulés dans la pathologie et potentiellement dépendants de DDX5 et DDX17. Cette partie est néanmoins beaucoup moins avancée et constitue une étude complémentaire du rôle de DDX5 qui devra être poursuivie.

Mon introduction bibliographique sera donc constituée de quatre chapitres :

- **Chapitre I : L'épissage.** J'y présenterai son importance biologique, ses aspects moléculaires, sa régulation et l'apport des technologies à haut débit dans sa compréhension globale.
- **Chapitre II : Les ARN hélicases à boîte DEAD : DDX5 et DDX17.** J'y présenterai les hélicases à boîte DEAD en général puis je me focaliserai sur les hélicases homologues DDX5 et DDX17 : leur expression, fonctions moléculaires, rôles biologiques et enfin rôle dans la pathologie.
- **Chapitre III : La myogénèse.** J'y présenterai les différentes étapes et protagonistes moléculaires permettant la mise en œuvre de ce processus en me focalisant en dernier lieu sur l'épissage.
- **Chapitre IV : La Dystrophie Myotonique de type 1.** J'y présenterai les aspects cliniques puis moléculaires de cette pathologie en insistant sur l'importance de la dérégulation de l'épissage.

Les résultats de la première partie de ma thèse portant sur le rôle « orchestrateur » de DDX5 et DDX17 au cours de la myogénèse seront présentés sous la forme d'une publication actuellement en révision dans *Molecular Cell*.

Les résultats de la deuxième partie portant sur le rôle potentiel de DDX5 et DDX17 dans la DM1 seront présentés sous la forme de résultats complémentaires non soumis pour publication.

Enfin mon travail en bioinformatique m'a permis de participer à d'autres projets de l'équipe qui ont dans certains cas donné lieu à des publications. En annexes, une courte présentation méthodologique présente le système d'analyse mis en place, suivie par les publications auxquelles j'ai contribué en y précisant à chaque fois quelle y a été ma contribution.



**1**

# **Introduction**



# I L'épissage

## A Préambule

L'expression génique est constituée d'un ensemble de processus généralement présentés comme séquentiels : transcription de l'ARN à partir de l'ADN modèle, maturation, export cytoplasmique, traduction... Ces étapes forment un ensemble continu, chacune étant liée à la suivante et pouvant l'influencer (Braunschweig et al., 2013). La maturation de l'ARN est elle-même découpée en plusieurs sous-étapes : ajout de la coiffe en 5', épissage des introns, clivage et ajout de la queue polyA en 3' pour l'export cytoplasmique.

En particulier, le rôle de l'épissage apparaît comme déterminant. En effet, les gènes eucaryotes présentent une structure morcelée. Schématiquement les séquences codantes (exons) sont entrecoupées de séquences non codantes (introns). Ces dernières peuvent constituer jusqu'à 90% des ARNs non épissés chez l'humain (Stamm et al., 2005). L'épissage consiste en l'excision des séquences introniques et la ligature des exons de l'ARN pré-messager. On distingue l'épissage dit « constitutif » (lorsqu'un intron est toujours éliminé et un exon toujours conservé dans l'ARN messenger mature), de l'épissage dit « alternatif » (lorsqu'un exon ou portion d'exon peut être inclus ou exclus de l'ARN mature). A sa découverte dans les années 1970s, ce mécanisme semblait marginal (Berget et al., 1977; Chow et al., 1977) mais on estime maintenant que près de 95% des gènes humains subissent l'épissage alternatif (Pan et al., 2008; Barash et al., 2010).

Plus de 90% des gènes humains sont annotés comme multi-exoniques (Cusack et al., 2011), il est donc possible par épissage alternatif de produire une multitude d'ARNs messagers. On estime que plus de 80% des événements d'épissage alternatif touchent la région codante des ARNs (Modrek and Lee, 2002), ce mécanisme constitue donc une source de diversité protéique énorme. Mais ce n'est pas son seul rôle ; par exemple, l'épissage alternatif régule de manière quantitative l'expression génique par la production de transcrits non fonctionnels. D'autre part, des défauts d'épissage ont été observés dans de nombreuses pathologies et des études de plus en plus nombreuses montrent l'importance fonctionnelle de ces événements dans la



pathogénèse.

## B Importance de l'épissage alternatif

En permettant de générer plusieurs transcrits potentiellement fonctionnels à partir d'un même gène, l'épissage alternatif est la source majeure à l'origine de la diversité protéomique. Les isoformes produites à partir d'un même gène peuvent varier en activité, localisation ou encore affinité (Izquierdo et al., 2005 ; Carosella et al., 2003 ; Miki et al., 1992).

Outre cette régulation qualitative, l'épissage alternatif permet aussi une régulation quantitative via la voie du NMD (Nonsense-Mediated mRNA decay). En effet, l'épissage peut générer l'apparition d'un codon de terminaison prématuré (Rebbapragada and Lykke-Andersen, 2009). Ce transcrit non fonctionnel sera alors dégradé par la machinerie NMD aboutissant en une baisse de l'expression protéique. Il est intéressant de noter que ce mécanisme permet notamment l'auto-régulation de nombreux facteurs d'épissage (Cline, 1984 ; Damianov and Black, 2010, Dembowski and Grabowski, 2009 ; Gates et al., 2011).

De plus en plus d'études montrent que des « switch » d'épissage peuvent être des événements majeurs dans le contrôle du devenir cellulaire. En effet, un nombre limité d'exons régulés peut suffire à agir sur des programmes entiers de régulation, en modifiant uniquement un domaine conférant au facteur une activité particulière (remodelage de la chromatine, activité de transcription...) (Gabut et al., 2011 ; Theodorou et al., 2009). La **Figure 1** donne deux exemples de « switch » d'épissage clés lors de processus de différenciation.

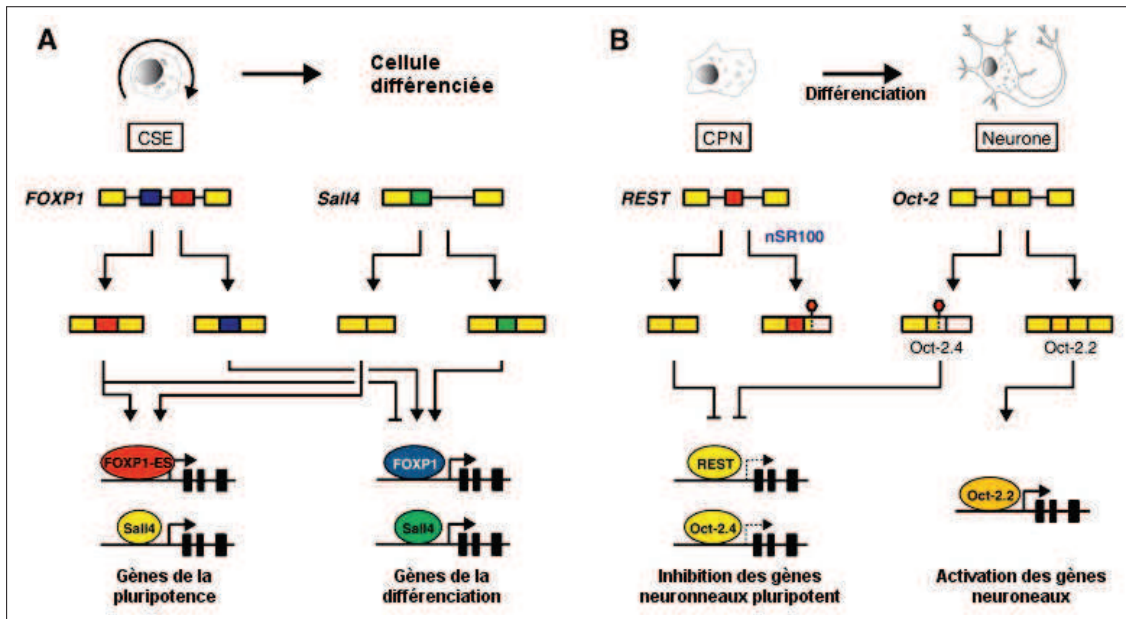


FIGURE 1: *Des « switch » d'épissage contrôlent l'activité de facteurs clés durant la différenciation cellulaire.* (A) CSE : Cellule Souche Embryonnaire. Des isoformes spécifiques des cellules souches embryonnaires des facteurs de transcription FOXP1 et Sall4 induisent l'expression de gènes de la pluripotence tout en réprimant ceux de la différenciation. A l'inverse, des isoformes de ces facteurs qui activent les gènes de la différenciation sont cette fois exprimées dans les cellules différenciées (Gabut et al., 2011). (B) CPN : Cellule Progénitrice Neuronale. Durant la différenciation neuronale, l'épissage alternatif du facteur de transcription inhibiteur REST produit une isoforme non fonctionnelle permettant aux gènes neuronaux de s'exprimer. Un autre facteur de transcription négatif Oct2 subit aussi un épissage alternatif le rendant cette fois activateur des gènes neuronaux durant la différenciation (Raj et al., 2011). (Adapté de Irimia and Blencowe, 2012).

Ainsi, durant la différenciation de cellules souches embryonnaires, un « switch » d'épissage du facteur de transcription FOXP1 modifie son motif d'interaction à l'ADN. Dans les cellules souches embryonnaires, le motif généré par épissage permet l'activation transcriptionnelle de facteurs clés de la pluripotence tels que OCT4 et NANOG, tout en réprimant également les gènes de la différenciation. A l'inverse, dans les cellules différenciées, le motif d'interaction à l'ADN alternatif

permet l'activation des gènes de la différenciation (Figure 1A) (Gabut et al., 2011). On trouve d'autres exemples d'épissage impactant l'activité transcriptionnelle de facteurs régulateurs clés dans divers processus de différenciation, notamment durant la neurogénèse (Figure 1B (Raj et al., 2011)). Les facteurs impactés par épissage alternatif sont cette fois-ci les facteurs de transcription REST et Oct-2. Les modifications d'activité qui en résultent contrôlent les programmes transcriptionnels responsables de la différenciation neuronale.

## C Mécanisme de la réaction d'épissage

### 1 Sites de reconnaissance des exons et introns

Afin d'assurer la production d'ARNs messagers matures fonctionnels, le mécanisme d'épissage se doit d'être extrêmement spécifique et fidèle. Pour se faire, la reconnaissance des exons et des introns nécessite un ensemble de séquences caractéristiques : le site 5' d'épissage correspondant à la jonction exon-intron (aussi appelé site donneur), le site de branchement situé dans l'intron, une séquence riche en pyrimidines située juste en amont du site 3' d'épissage correspondant à la jonction intron-exon (ou site accepteur) (Figure 2).

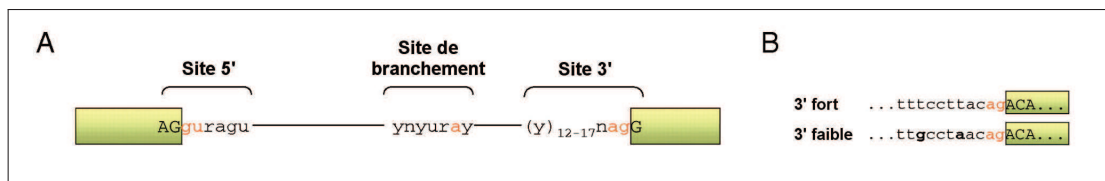


FIGURE 2: *Sites de reconnaissance des introns et exons.* (A) Séquences consensus du site 5' d'épissage, du site de branchement et du site 3' d'épissage. En orange : les nucléotides très conservés et caractéristiques des séquences. (B) Exemples de deux sites 3' d'épissage : l'un fort, l'autre faible. En gras : les nucléotides non consensus affaiblissant le site 3' (Adapté de Srebrow and Kornblihtt, 2006).

## 2 Réaction d'épissage

La réaction d'épissage aboutit à la ligature de deux exons adjacents et à la libération de l'intron intermédiaire sous la forme d'une structure en « lasso ». Elle consiste en deux étapes de trans-estérification (**Figure 3**).

Lors de la première réaction, le groupement 2'-OH de l'adénosine du site de branchement attaque la liaison phosphodiester de la jonction exon-intron en 5'. Ceci aboutit à la formation d'un exon libre en 5' et d'un intron formant une structure en lasso contenant l'exon en 3'. Lors de la deuxième réaction, c'est le groupement 3'-OH de l'exon libre en 5' qui attaque la liaison phosphodiester de la jonction intron-exon de l'exon en 3'. Les produits de ces deux réactions sont deux exons ligaturés et un intron cyclisé au niveau de l'adénosine du site de branchement (Spierling et al., 2008). Celui-ci sera, par la suite, pris en charge par des nucléases afin d'être recyclé (Brow, 2002).

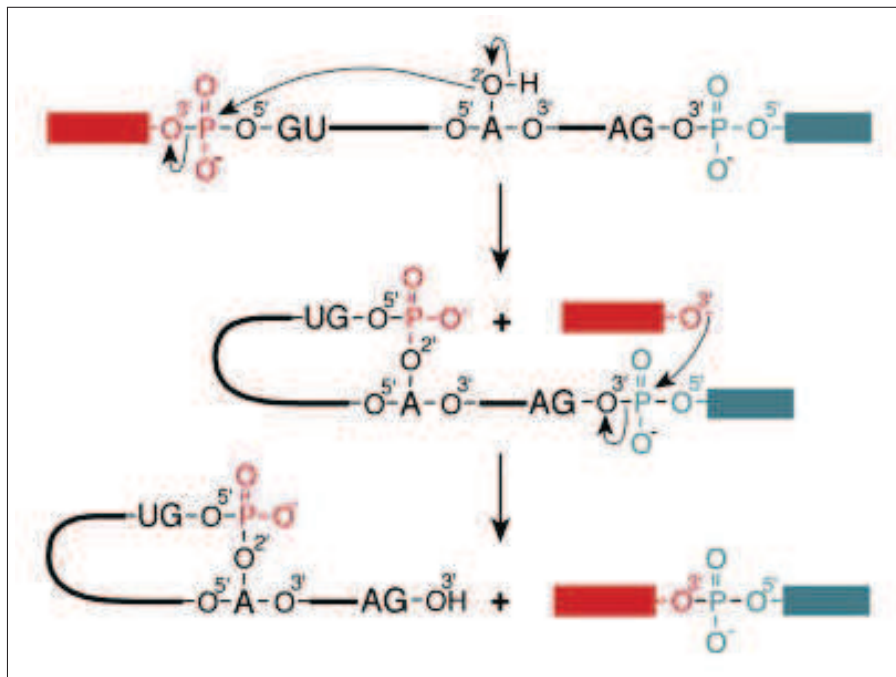


FIGURE 3: **Réaction d'épissage.** La réaction d'épissage consiste en deux réactions de trans-estérification. La première libère l'exon en 5', cyclise l'intron toujours rattaché à l'exon en 3'. La deuxième lie les deux exons et libère l'intron en lasso. (Spierling et al., 2008).

### 3 Assemblage du spliceosome

Certains introns ne nécessitent pas d'autres éléments pour catalyser leur propre épissage (ribozymes) : ce sont les introns de groupe 1 et 2 (Haugen et al., 2005 ; Pyle et al., 2007). Par exemple, les ARNs pré-ribosomiques ou de transfert appartiennent à cette catégorie. Ce n'est cependant pas le cas de la plupart des ARNs pré-messagers transcrits par l'ARN polymérase II. Ceux-ci nécessitent l'intervention d'un large complexe ribonucléoprotéique : le spliceosome.

Le spliceosome est un complexe dynamique faisant intervenir de manière séquentielle différents éléments structuraux et enzymatiques au cours de la réaction d'épissage. Ce complexe implique notamment à minima cinq snRNPs (small nuclear RiboNucleoProtein) : U1, U2, U4, U5 et U6 (U4 et U6 formant une même entité U4/U6). Une snRNP est formée d'un ARN (de 100 à 200 nucléotides de longueur) recouvert de protéines dont le nombre varie d'un type de snRNP à l'autre (de quelques protéines à plusieurs dizaines). On estime que plus de 200 protéines sont associées au spliceosome humain (Hegele et al., 2012). Parmi celles-ci, près de 140 font partie du « core spliceosome », elles sont considérées comme des éléments centraux du complexe, interagissant spécifiquement avec les snRNPs.

Les snRNPs sont recrutées au niveau des séquences précédemment décrites. En soit, elles n'ont pas de rôle catalytique mais ont un rôle structural puisqu'elles permettent le repliement de l'ARN pré-messager rendant possible les deux réactions de transestérifications du mécanisme d'épissage (Hoskins and Moore, 2012).

Brièvement, la réaction d'épissage réalisée par le spliceosome se produit en plusieurs étapes par une succession dynamique de « complexes spliceosomaux » intermédiaires. La première snRNP à se fixer est U1 qui interagit avec le site 5', puis les protéines BBP (Branchpoint Bridging Point) et Mud2 au point de branchement. Ce premier complexe constitue le spliceosome précoce (aussi appelé complexe E). U2 vient alors déplacer BBP/Mud2 au point de branchement, ce qui constitue le complexe A. Puis U4/U6 associé à U5 viennent se fixer entre U1 et U2 formant le complexe B. Des changements de conformation importants de l'appariement des snRNAs et du pré-messager provoquent la libération des U1 et U4 formant le complexe Bact puis l'association du complexe Prp19 et la dissociation

de protéines associées à U2 (SF3) mène à la formation du complexe B\*. Celui-ci permet de former le complexe actif du spliceosome, capable de catalyser la réaction d'épissage : le complexe C. Après clivage du site 5' (première réaction de transestérification), on obtient le complexe C1 qui change de nouveau de conformation pour former le complexe C2 qui effectue la deuxième réaction de transestérification. Les exons ligaturés sont alors relargués, l'intron sous forme de lasso dégradé et le spliceosome dissocié (Figure 4) (Hoskins and Moore, 2012).

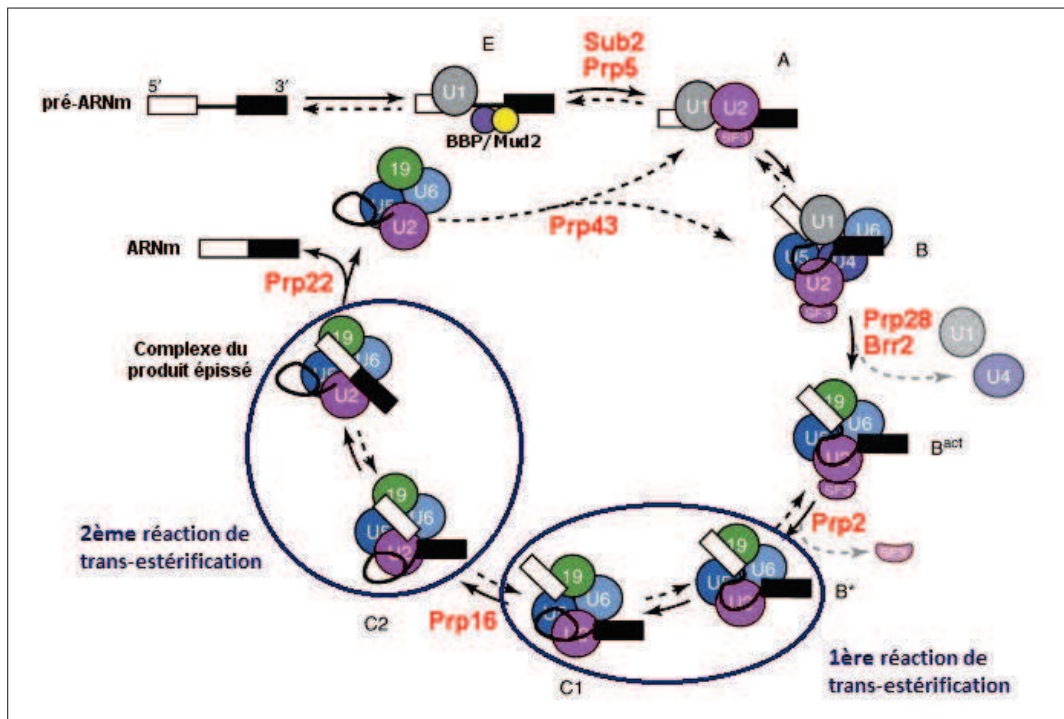


FIGURE 4: *Assemblage dynamique du spliceosome durant la réaction d'épissage.* Les différents complexes spliceosomaux sont indiqués par les lettres correspondantes (E, A, B, etc). 19 : complexe Prp19. En gris : dissociation de protéines. En rouge : hélicases ATPases DExD/H-box. Entouré en bleu : les deux étapes de transestérifications. Double flèche : étapes réversibles. Flèche simple : étapes irréversibles. (Adapté de Hoskins and Moore, 2012).

Des études à la fois in vitro et in vivo ont été réalisées afin d'étudier chacun des complexes spliceosomaux. Elles ont pu mettre en évidence la réversibilité de la plupart des transitions, ce qui laisse penser que la réaction d'épissage peut être

régulée à chaque étape de la formation du spliceosome (Hoskins et al., 2011).

Chacune des phases d'association ou dissociation d'une snRNP implique une hydrolyse d'ATP. L'énergie produite par cette réaction est nécessaire aux réarrangements structuraux qui aboutissent à la formation du site catalytique actif du spliceosome.

Parmi les nombreuses protéines enzymatiques qui gravitent autour du spliceosome en rendant possible les réarrangements ARN-ARN, ARN-protéine et protéine-protéine, on distingue des ARN hélicases (Figure 4).

Huit hélicases DExD/H participent à la formation du spliceosome, parmi celles-ci trois appartiennent à la famille à boîte DEAD, quatre à la famille à boîte DEAH et une à la famille Ski2-like (Will and Lührmann, 2011) (Figure 5). Les hélicases à boîte DEAD ne sont requises que durant l'étape d'assemblage du spliceosome alors que les autres hélicases participent à l'activation, au remodelage et au recyclage du spliceosome catalytique (Figure 5).



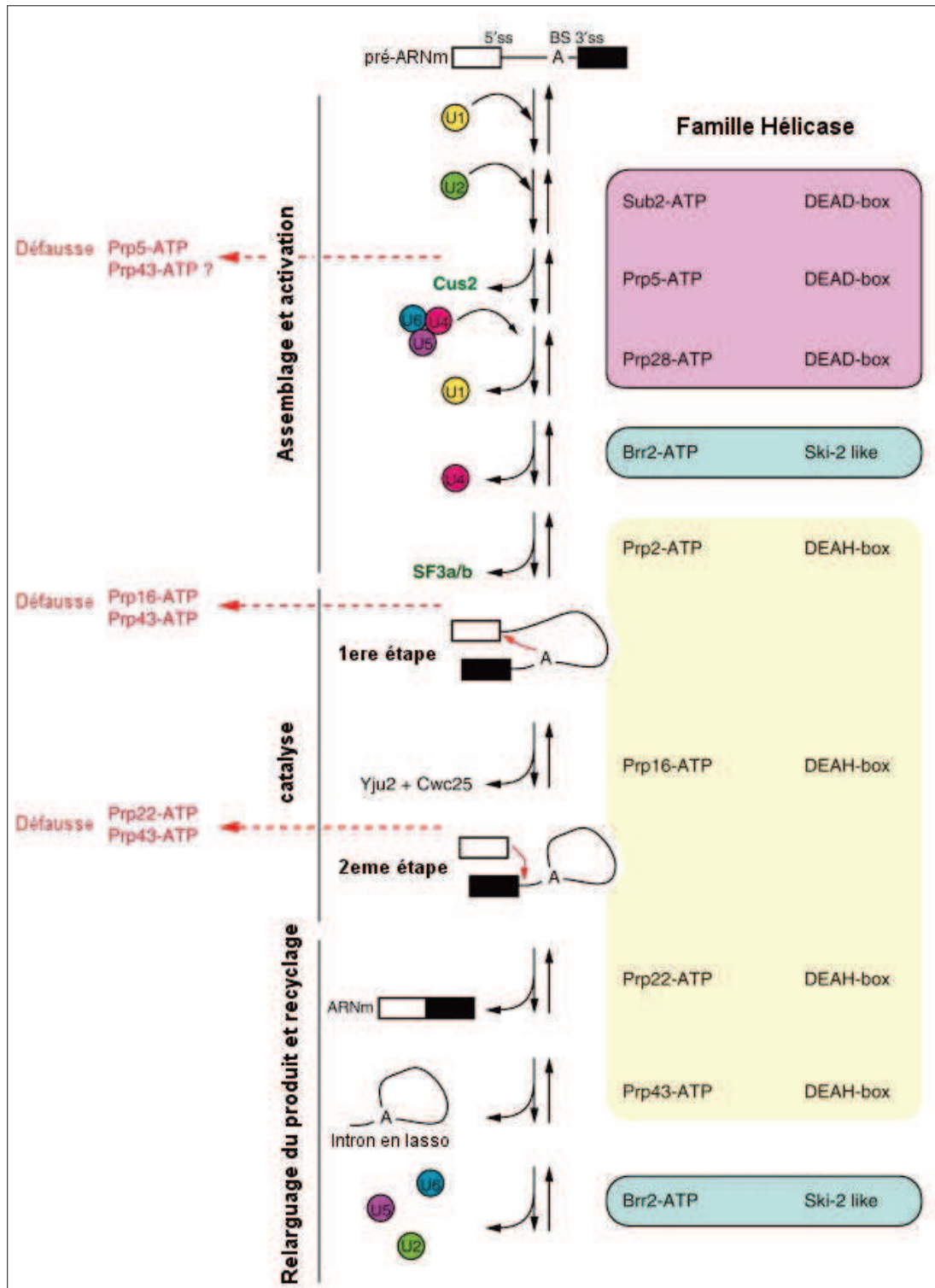




FIGURE 5: **Hélicases ARN durant la réaction d'épissage.** Les hélicases ARN facilitent les transitions conformationnelles lors de l'assemblage et l'activation du spliceosome (en haut), de la catalyse (milieu) et du relarguage des produits ainsi que du recyclage des snRNPs (en bas). Les hélicases impliquées appartiennent à trois familles : à boîte DEAD (en violet), à boîte DEAH (en jaune) et Ski2-like (en bleu) (Adapté de *Cordin et al., 2012*).

## D Epissage alternatif et régulation

### 1 L'épissage alternatif

A la différence des exons constitutifs qui sont toujours conservés dans les ARNs messagers matures, certains exons (ou portions d'exons) ne sont pas systématiquement épissés et conservés dans l'ARNm mature. L'épissage alternatif est le processus selon lequel un gène peut produire différents ARNm matures en fonction des combinaisons d'exons (ou portion d'exons) possibles.

Il existe plusieurs types d'épissages alternatifs (**Figure 6**). L'évènement « élémentaire » le plus courant est le saut d'exon : l'exon est soit inclus soit exclu de l'ARN mature. Il a été estimé que plus de la moitié des événements d'épissage alternatif correspondent à des sauts d'exons (*Sammeth et al., 2008*). Deux exons consécutifs sont dits mutuellement exclusifs si l'un est toujours exclu dans l'ARNm lorsque l'autre est inclus et réciproquement. Deux autres évènements d'épissage alternatif correspondent à la sélection différentielle de sites d'épissage respectivement accepteur (en 3') et donneur (en 5'). En fonction du site d'épissage sélectionné, une partie de l'exon pourra être incluse ou exclue de l'ARNm. La rétention d'intron correspond à l'inclusion d'un intron dans l'ARNm. La délétion est l'épissage interne à un exon. Une délétion peut être considérée comme un intron rarement épissé et un intron-rétention à une région exonique rarement incluse.

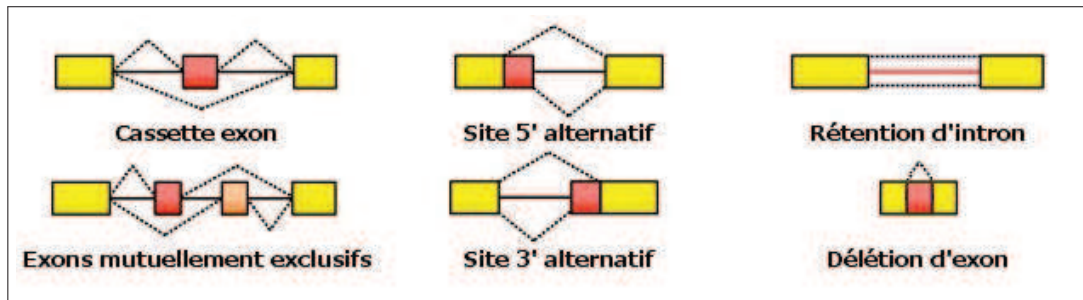


FIGURE 6: *Différents type d'épissage alternatif.* En rose : exon (ou portion d'exon) alternativement inclus(e) dans l'ARN messager mature. (Adapté de *Irimia and Blencowe, 2012*).

En plus de ces événements d'épissage, l'utilisation de promoteurs et polyA alternatifs entraîne la synthèse d'ARNm ayant des extrémités 5' et 3' différentes. Ceci peut imposer le choix de l'épissage réalisé.

Des analyses bioinformatiques ont montré que les exons constitutifs et alternatifs ne partagent pas les mêmes caractéristiques (**Figure 7**) (*Kim et al., 2008*). En effet, les exons alternatifs apparaissent être significativement plus courts que les exons constitutifs et être flanqués par des introns plus longs. Ceci est susceptible de rendre les exons alternatifs moins facilement reconnus par la machinerie d'épissage. D'autre part, la force de leurs sites d'épissage (à la fois 3' et 5') est statistiquement plus faible, ceci conforte l'idée que les exons alternatifs sont des exons « sous-optimaux » pour être épissés.

De manière plus surprenante, la séquence des exons alternatifs et de leurs introns flanquants est nettement plus conservée entre espèces que dans le cas des exons constitutifs (**Figure 7**, ici : entre homme et souris, *Kim et al., 2008*). Ceci met en évidence l'importance majeure d'autres séquences régulatrices dites séquences « cis » dans la reconnaissance des exons, en particulier alternatifs. Celles-ci facilitent ou, au contraire, empêchent l'inclusion/exclusion de l'exon par la machinerie d'épissage (mécanisme décrit en détails dans la sous-partie suivante).

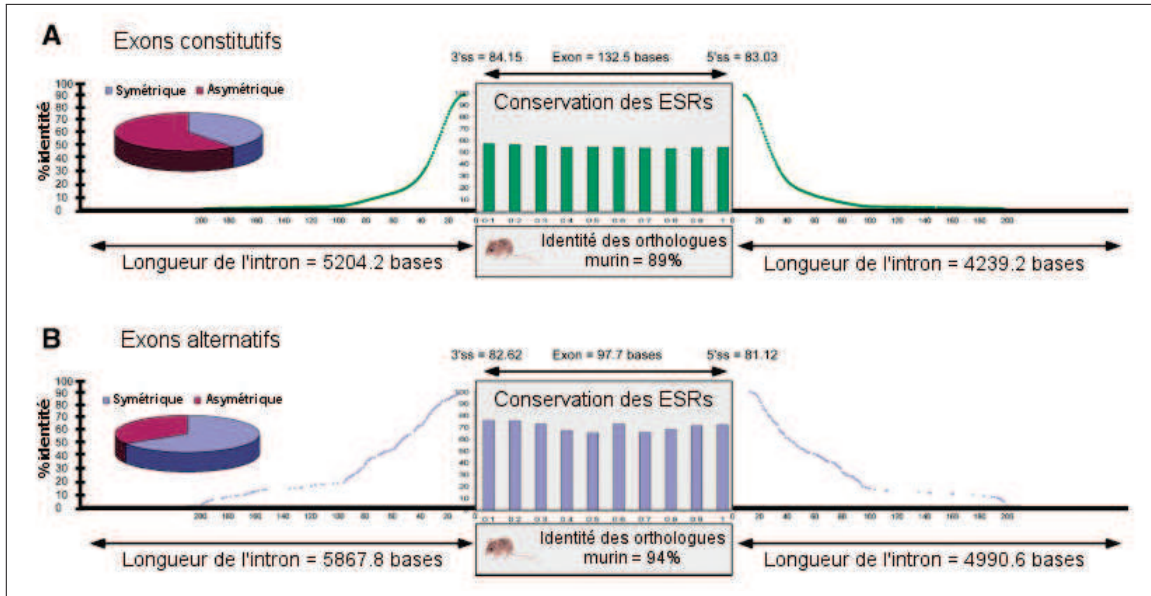


FIGURE 7: *Caractéristiques différent entre exons constitutifs et alternatifs chez l'homme. (A) Exons constitutifs. (B) Exons alternatifs. Seuls les exons conservés entre l'homme et la souris ont été analysés. Les caractéristiques suivantes ont été analysées : longueur de l'exon, force des sites d'épissage, conservation des séquences régulatrices exoniques (ESR), pourcentage d'identité entre homme et souris, longueurs des introns flanquants et leur conservation entre homme et souris, et fraction d'exons symétriques. (Adapté de Kim et al., 2008)*

D'autre part, il est intéressant de noter que la proportion d'exons symétriques (dont le nombre de nucléotides est un multiple de trois et dont l'inclusion ou l'exclusion ne change pas le cadre de lecture) est nettement supérieure dans le cas des exons alternatifs. Dans la plupart des cas, ceux-ci ne seront donc pas à l'origine d'un codon stop prématuré aboutissant à la production de transcrits non fonctionnels.

## 2 Régulation de l'épissage

### a Les éléments cis-régulateurs

Comme précédemment décrit, les exons peuvent présenter des sites d'épissages très dégénérés (en particulier dans le cas des exons alternatifs), rendant leurs sites

faiblement reconnus par la machinerie d'épissage. D'autre part, on sait qu'il existe un très grand nombre de pseudo-exons dans le génome humain : ceux-ci sont définis par des sites d'épissage potentiels plus ou moins dégénérés mais qui ne définissent pas un exon reconnu par le spliceosome (Sun and Chasin, 2000). Leur nombre dépasserait même celui des vrais exons (Zhang et al., 2005). La force des sites d'épissage ne suffit donc pas à déterminer si un exon sera reconnu ou non par le spliceosome.

Les éléments « cis » sont de courtes séquences auxiliaires (de 5 à 10 nucléotides) conservées et dégénérées présentes au sein de l'exon ou dans ses introns flanquants. Elles sont capables de favoriser ou au contraire inhiber le recrutement du spliceosome. Ces séquences régulatrices sont nommées en fonction de leur position et de leur activité : les séquences activatrices sont les ESE (Exonic Splicing Enhancer) et ISE (Intronic Splicing Enhancer), les séquences inhibitrices sont les ESS (Exonic Splicing Silencer) et ISS (Intronic Splicing Silencer) (Figure 8). Différents types de séquences peuvent être présentes dans l'environnement d'un même exon, on peut aussi observer de multiples occurrences d'une même séquence.

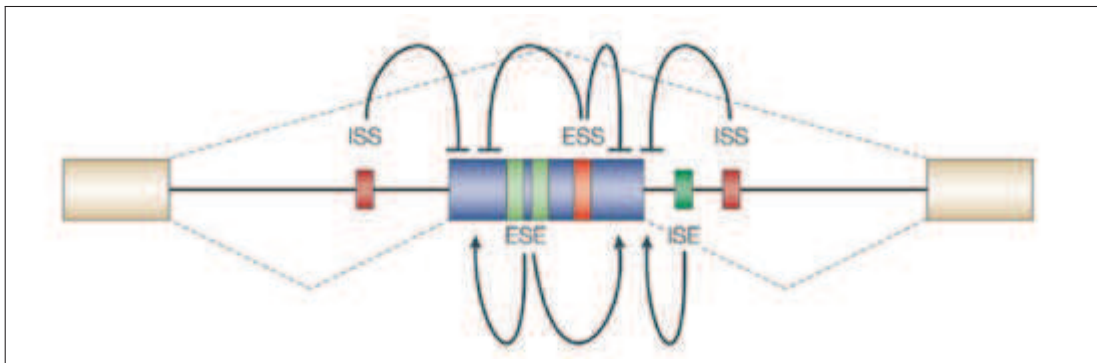


FIGURE 8: *Séquences cis-régulatrices de l'épissage.* Les séquences cis-régulatrices sont nommées en fonction de leur position et activité. (Matlin et al., 2005)

Ces éléments sont particulièrement importants pour discriminer les vrais exons des pseudos-exons, les sites d'épissage compétitifs ou encore reconnaître les exons faibles.

## b Les protéines trans-régulatrices SRs et hnRNPs

Les séquences cis-régulatrices permettent le recrutement de protéines régulatrices appelées facteurs « trans » qui vont réguler l'épissage. Deux familles majeures se dégagent : la famille des protéines SRs et celle des protéines hnRNPs.

Pendant longtemps, les protéines SRs ont été considérées comme des protéines exclusivement activatrices de l'épissage. Montrées nécessaires à la fois à l'épissage constitutif et alternatif (Schaal and Maniatis, 1999), elles possèdent un domaine RS (Serine-Arginine) qui a donné son nom à la famille. Elles fixent les séquences cis-régulatrices via leur motif RRM de fixation à l'ARN (Kohtz1994) (Figure 9). On estime que leur mécanisme d'action est essentiellement la reconnaissance des sites d'épissage : de la snRNP U1 au niveau du site 5', et du complexe U2AF et de la snRNP U2 au niveau du site 3' d'épissage (Graveley, 2000) (Figure 9). C'est pourquoi, de très nombreuses études attribuent un rôle activateur à ces protéines, bien que l'on sache maintenant qu'il ne s'agit pas d'une règle générale (Pandit2013).

A l'inverse, les protéines hnRNPs ont longtemps été considérées comme inhibitrices de l'épissage à la fois d'exons constitutifs et alternatifs (Wang2008). Cette famille de protéines est très hétérogène et a été impliquée dans de nombreuses fonctions biologiques en plus de l'épissage (transcription, stabilité, traduction...). Elles se fixent aux éléments cis-régulateurs de l'ARN via leur motif RRM (pour certaines, il s'agit d'un motif KH) et on estime que la plupart du temps, elles empêchent par encombrement stérique l'intervention de la machinerie d'épissage (Figure 9). Celui-ci est réalisé par oligomérisation le long des séquences introniques et exoniques (Zhu et al., 2001) ou encore blocage du recrutement de snRNPs (Tange et al., 2001 ; House and Lynch, 2006). Cependant de plus en plus d'études montrent que les hnRNPs peuvent aussi favoriser l'inclusion d'exons dans certains contextes (Witten and Ule, 2011 ; Huelga et al., 2012).

L'une des protéines hnRNPs la plus étudiée est PTB (ou hnRNP I) qui interagit avec la séquence riche en pyrimidines en amont du site 3' d'épissage bloquant ainsi l'accès au complexe U2AF (Saulière et al., 2006).

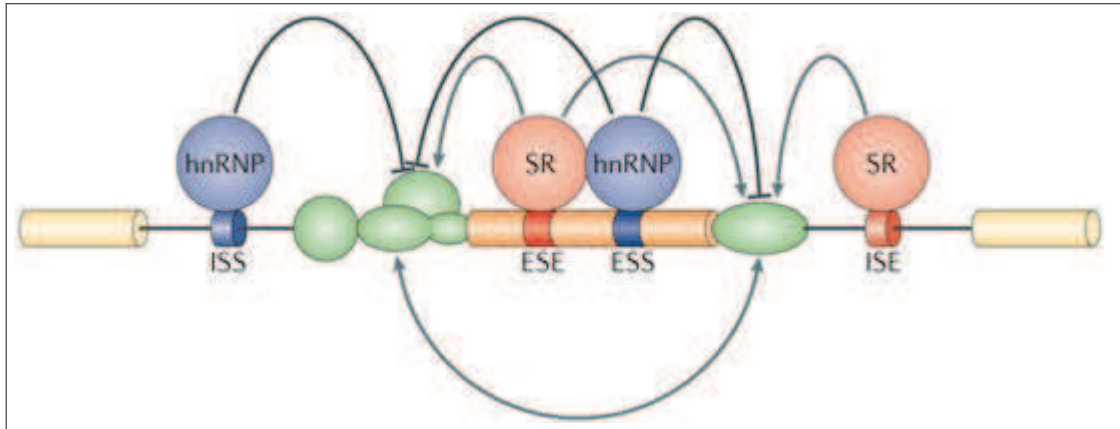


FIGURE 9: *Séquences et facteurs de régulation de l'épissage.* Les deux principales familles de régulateurs de l'épissage sont représentées : les protéines SR (en rose) et les protéines hnRNP (en bleu). Leur activité est représentée selon le modèle communément établi. En vert : les composants du spliceosome. (Kornblihtt et al., 2013)

Si pendant longtemps, les protéines SRs et hnRNPs ont été considérés comme ayant une seule activité de régulation de l'épissage (respectivement activatrice ou inhibitrice) ; de nouvelles études, notamment à large échelle, montrent qu'en réalité c'est la position à laquelle les protéines sont recrutées (par rapport à l'exon) qui détermine leur activité.

Une liste non exhaustive des protéines SRs et hnRNPs est présentée [Tableau 1](#).

	Nom	Autres Noms	Domaines Protéiques	Sites de liaison	Gènes cibles
Protéines SRs	SRSF1	Sfrs1, ASF, SF2	RRM, RS	RGAAGAAC	HipK3, CaMKII $\delta$ , HIV RNAs
	SRSF2	Sfrs2, SC35	RRM, RS	UGCUGUU	ACHe
	SRSF3	Sfrs3, X16, SRp20	RRM, RS	GCUCCUCUUC	SRp20, CT/CGRP
	SRSF4	Sfrs4, SRp75	RRM, RS	GAAGGA	FN1, E1A, CD45
	SRSF5	Sfrs5, HRS, SRp40	RRM, RS	AGGAGAAGGGA	HipK3, PKC $\beta$ -II, Fibronectin
	SRSF6	Sfrs6, SRp55	RRM, RS	GGCAGCACCUUG	cTnT, CD44
	SRSF7	Sfrs7, 9G8	RRM, RS, C2HC Znf	(GAC)n	Tau, GnRH, 9G8
	SRSF8	Sfrs8, SWAP	RS, SWAP	ND	SWAP, CD45, Tau
	SRSF9	Sfrs9, SRp30c	RRM, RS	CUGGAUU	Glucocorticoid receptor
Protéines hnRNPs	hnRNP A1	Hnrnpa1	RRM, RGG	UAGGGA/U	HipK3, SMN2, c-H-ras
	hnRNP A2/B1	Hnrnpa2b1, Hnrnpa2	RRM, RGG	(UUAGGG)n	4.1R, HIV Tat, IKBKAP
	hnRNP C	Hnrnpc, Hnrnpc1/c2	RRM	U-rich	$\beta$ -amyloid receptor
	hnRNP F	Hnrpf	RRM, RGG, GY	GGGA, G-rich	PLP, c-SRC, Bcl-x
	hnRNP G	Rbmxrt, Hnrnpg	RRM, RGG, SRGY	AAGU	SMN2, $\alpha$ -tropomyosin
	hnRNP H	Hnrph1	RRM, RGG, GYR, GY	GGGA, G-rich	PLP, HIV tat, Bcl-x
	hnRNP L	Hnrnpl	RRM	C/A-rich	eNOS, CD45
	PTB	Ptbp1, Hnrpi	RRM	UCUU, CUCUCU	nPTB, c-SRC, Fas, cTNT, CGRP, NMDA, CLBC, hnRNP
	nPTB	Ptbp2, brPTB	RRM	CUCUCU	c-SRC, GlyR $\alpha$ 2

TABLE 1: *Liste non exhaustive des protéines SRs et hnRNPs régulatrices de l'épissage chez les mammifères. Les domaines de fixation à l'ARN, les sites consensus d'interaction, ainsi que des cibles régulées par ces facteurs sont indiqués (Adapté de Gabut et al., 2008 d'après Manley and Krainer, 2010).*

### c Autres protéines trans-régulatrices

Si les protéines SRs et hnRNPs sont les plus connues, il existe une multitude d'autres facteurs régulant l'épissage. En particulier, les facteurs de régulation tissu-spécifiques tel que NOVA (Jelen et al., 2007), ESRP (Warzecha et al., 2009) ou FOX (Lee et al., 2009) assurent la mise en place d'événements d'épissage tissu-spécifiques. Une liste non exhaustive de ces facteurs est présenté [Tableau 2](#).



	Nom	Autres Noms	Domaines Protéiques	Sites de liaison	Gènes cibles
Autres facteurs	RBFOX1	A2bp1, Fox1	RRM	(U)GCAUG	NMHC-B, CGRP, F1 $\gamma$
	RBFOX2	Rbm9, Fox2	RRM	(U)GCAUG	NMHC-B, 4.1R, FGFR2
	CELF1	Cugbp1, Brunol2	RRM	U/G-rich	cTnT, Insulin Receptor
	CUGBP2	ETR-3, Brunol3	RRM	U/G-rich	cTnT, Tau, Cox-2
	CELF4	Brunol4	RRM	U/G-rich	Mtmr1, cTnT
	ELAVL4	HuD	RRM	U-rich	Ikaros, CGRP, AChE
	NOVA1	Nova-1	KH	YCAAY	GlyR $\alpha$ 2, GABA $\alpha$
	NOVA2	Nova-2	KH	YCAAY	JNK, GIRK, neogenin
	TIA1	mTia1	RRM	U-rich	MYPT1, Fas, CGRP, FGFR2, TIAR, IL-8, VEGF
	TIAL1	TIAR	RRM	U-rich	TIA1, CGRP, TIAR
	MNBL1	Mbnl	C3H1 Znf	YGCU(U/G)Y	cTnT, Insulin receptor, Clcn1, Tnnt3
	KHDRBS3	T-STAR, SIm-2	KH	UAAA	CD44, VEGF-A
	QKI	Qk, Quaking	KH	ACUAAY[...]UAAY	MAG, PLP
	SFPQ	PSF	RRM	ND	CD45, CoAA
	RBM17	SPF45	RRM, G patch	ND	Sxl, Fas
	RBM4	Rbm4a, Lark	RRM, C2HC Znf	C/U-rich	MAPT, $\alpha$ -tropomyosin, Tau
	SF3B1	SAP155, SF3b155	RRM, HEAT	ND	Bcl-x
	KHDRBS1	Sam68	KH	A/U-rich	Bcl-x

TABLE 2: *Liste non exhaustive des protéines auxiliaires régulatrices de l'épissage chez les mammifères. Les domaines de fixation à l'ARN, les sites consensus d'interaction ainsi que des cibles régulées par ces facteurs sont indiqués (Adapté de Gabut et al., 2008).*

Par exemple, le facteur PTB est exprimé dans les précurseurs de neurones mais son expression diminue au cours de la différenciation au profit de nPTB. Ce « switch » d'expression contrôle la mise en place d'un programme d'épissage spécifique au stade de différenciation dans les neurones (Coutinho-Mansfield et al., 2007).

Outre le facteur d'épissage impliqué, le positionnement de celui-ci par rapport à l'exon régulé apparaît déterminant (Wang et al., 2012b). D'autre part, on trouve très fréquemment la présence de sites de fixation pour différents facteurs d'épissage dans l'environnement d'un exon alternatif. Son épissage est donc la résultante des effets combinatoires ou compétitifs des facteurs activateurs et répresseurs (Chabot, 1996, Crawford and Patton, 2006, Sureau et al., 2011).

L'épissage alternatif apparaît être le fruit de multiples mécanismes pouvant tant être synergiques que concurrents. Il dépend à la fois de la présence de séquences



régulatrices, de leur positionnement et du taux d'expression des facteurs régulateurs correspondants.

Limiter l'épissage à la combinatoire des éléments cités ci-dessus est cependant erroné. On sait maintenant que l'épissage est couplé à d'autres niveaux de régulation totalement indépendants de la séquence en nucléotides ou de l'expression des facteurs régulateurs. Il est nécessaire de prendre en compte chacun de ces nouveaux éléments afin de comprendre dans sa globalité la complexité de régulation de l'épissage alternatif.

#### d Couplage à la transcription

Pendant longtemps la transcription et l'épissage ont été considérés comme deux événements indépendants et séquentiels. Cependant de nombreuses études montrent que les trois étapes de maturation de l'ARN (ajout de la coiffe, épissage et ajout de la queue polyA) surviennent de manière co-transcriptionnelle et que ces mécanismes sont couplés. En particulier en ce qui concerne l'épissage, des ARNs épissés matures, des composants du spliceosome et des facteurs d'épissage ont ainsi été retrouvés dans les fractions chromatiniennes avant leur relarguage dans le nucléoplasme (Listerman et al., 2006; Pandya-Jones and Black, 2009; Tilgner et al., 2012). Néanmoins si l'épissage constitutif semble être exclusivement co-transcriptionnel, l'épissage alternatif ne l'est pas toujours (Vargas et al., 2011).

Actuellement, l'influence de la transcription sur l'épissage est expliquée par deux mécanismes non mutuellement exclusifs. Le premier est le recrutement par la machinerie transcriptionnelle de facteurs impliqués dans l'épissage. En effet, le domaine CTD de l'ARN polymérase II peut s'associer et aider au recrutement de nombreux facteurs d'épissage (de la Mata and Kornblihtt, 2006; Das et al., 2007; Muñoz et al., 2010) (Figure 10A). Cette interaction peut aussi se produire via un intermédiaire : le complexe Médiateur qui interagit avec la machinerie de transcription a été montré comme interagissant avec le facteur d'épissage hnRNP L (Huang et al., 2012) (Figure 10B).

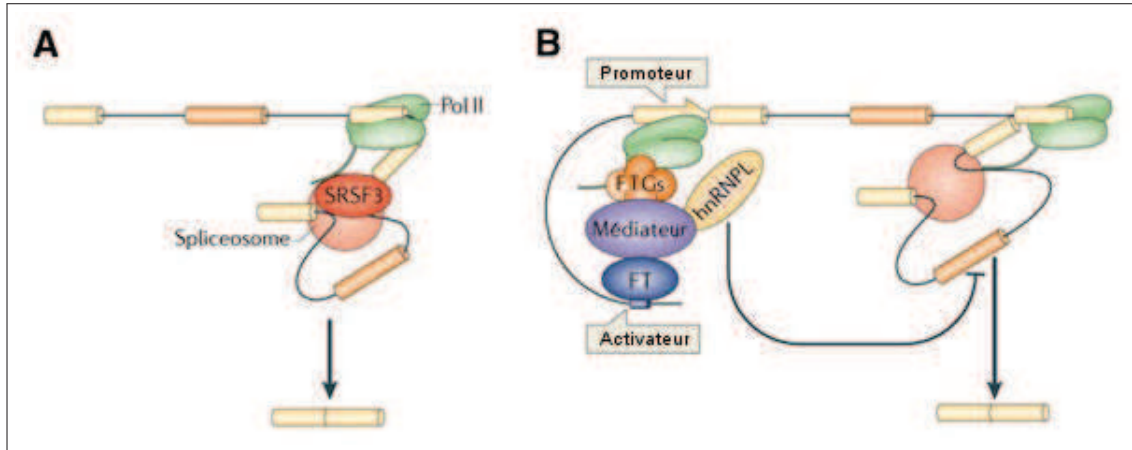


FIGURE 10: *Recrutement de facteurs d'épissage par la machinerie transcriptionnelle.* (A) Le domaine CTD de l'ARN polymérase II est nécessaire au recrutement du facteur d'épissage SRSF3 induisant une exclusion d'exon. (B) Le complexe Médiateur s'associe à des facteurs de transcription généraux (GTFs) et spécifiques (TFs) mais peut aussi s'associer au facteur d'épissage négatif hnRNPL induisant une exclusion d'exon (Adapté de Kornblihtt et al., 2013)

Le deuxième mécanisme expliquant l'influence de la transcription sur l'épissage est la cinétique d'élongation de l'ARN polymérase II. En effet, le fait de favoriser les pauses de la polymérase II ou de diminuer sa vitesse d'élongation (par exemple par traitement de drogue) est associé avec une augmentation du taux d'inclusion d'exons (Nogués et al., 2003, Ip et al., 2011). On estime que la fenêtre de temps au cours de laquelle un exon « faible » peut recruter le spliceosome pour entraîner son inclusion est ainsi augmentée. A l'inverse, une vitesse d'élongation plus importante a été associée à l'exclusion d'exons (Nogues et al., 2002) (Figure 11A).

Cependant dans certains contextes particuliers, une faible vitesse d'élongation apparaît au contraire favoriser l'exclusion d'exons (Ip et al., 2011; Dutertre et al., 2010). En effet, des facteurs d'épissage inhibiteurs ont alors le temps de se lier à l'ARN nouvellement synthétisé (Figure 11B) et entraînent l'exclusion de l'exon du transcrit mature.

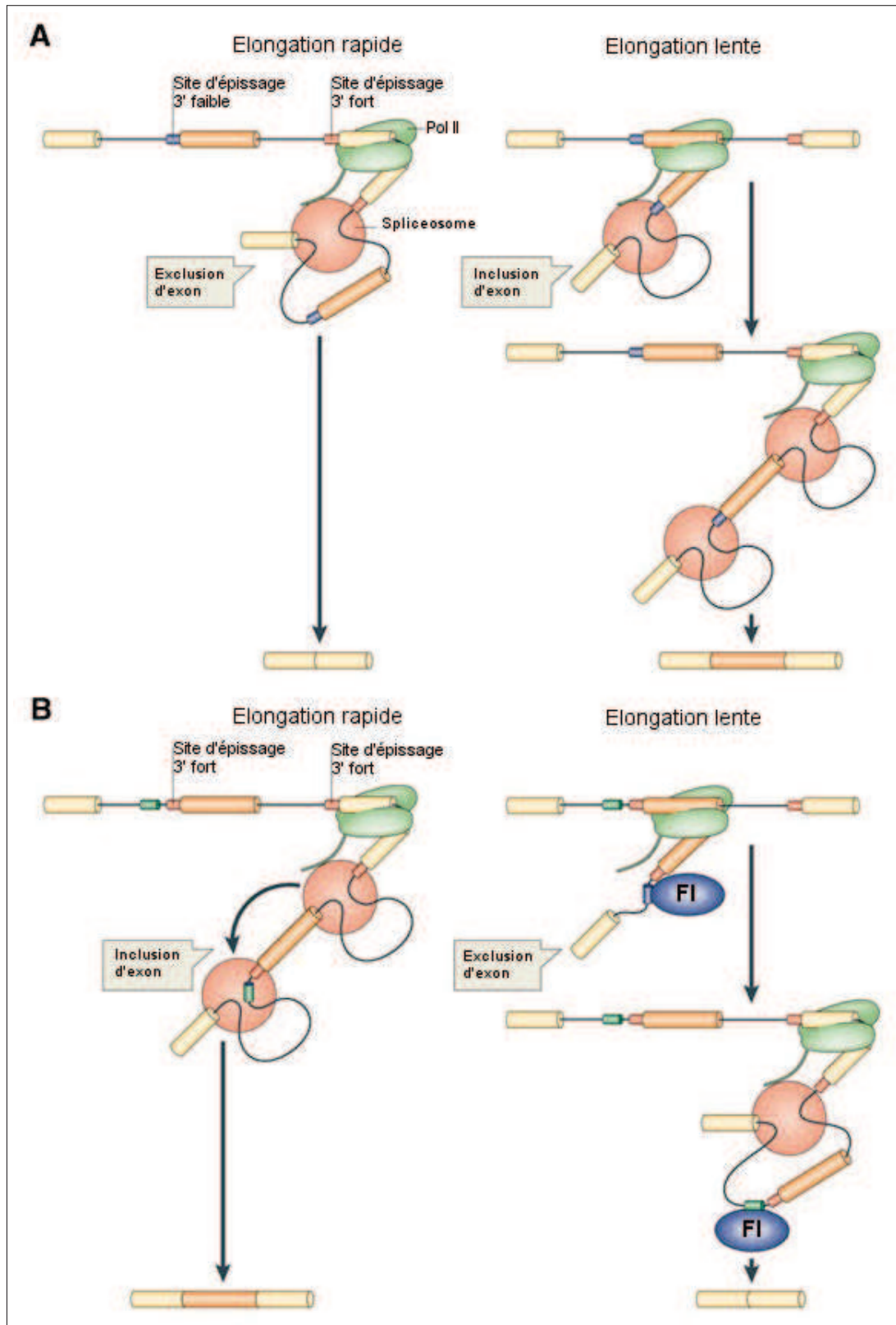


FIGURE 11: ***Influence de la vitesse d'élongation de l'ARN polymérase II sur l'épissage alternatif.*** (A) Une vitesse d'élongation rapide de l'ARN polymérase II entraîne la reconnaissance des sites d'épissage forts au détriment des faibles entraînant une exclusion de l'exon faible (gauche). Au contraire, une vitesse réduite permet au spliceosome de reconnaître les deux exons. (B) Dans le cas de deux exons consécutifs forts, une vitesse d'élongation rapide favorise la reconnaissance des deux exons (gauche) alors qu'une vitesse faible permet à des facteurs inhibiteurs (FI) de se fixer entraînant une exclusion d'exon (droite). (Adapté de [Kornblihtt et al., 2013](#)).

Si l'influence de la transcription sur l'épissage est bien établie, il est intéressant de noter que de récentes études montrent qu'elle est réciproque et que l'épissage influence aussi directement la transcription ([Lin et al., 2008](#); [Feng et al., 2012](#); [Mercer and Mattick, 2013](#)).

#### e Influence de la structure chromatinienne

La structure de la chromatine apparaît être aussi un élément clé dans la régulation de l'épissage alternatif ([Alló et al., 2010](#); [Luco et al., 2011](#)). En particulier, deux éléments ont été identifiés : les modifications d'histones et le positionnement des nucléosomes.

Premièrement, les modifications post-traductionnelles d'histones peuvent être divisées en deux catégories : celles associées à une transcription active et celles associées à une inhibition de transcription. Par exemple, l'exon 18 du gène NCAM est soumis à ce type de régulation dans les cellules neuronales. La dépolarisation membranaire entraîne l'acétylation de la Lys9 de l'histone 3 (H3K9ac), marque épigénétique qui promeut un état ouvert de la chromatine et, donc, facilite l'élongation de la polymérase II. En conséquence on observe une exclusion de l'exon. A l'inverse, lors de la différenciation neuronale, les marques H3K9m2 et H3K9m3 (respectivement double et tri-méthylations de la même lysine) induisent

une inhibition d'élongation et une inclusion d'exon (Schor et al., 2009) (Figure 12A). D'autre part, les modifications d'histones ont été impliquées dans la régulation de l'épissage via recrutement de facteurs d'épissage. Par exemple, la marque H3K36m3 a été identifiée comme interagissant avec une protéine adaptatrice qui à son tour s'associe avec le facteur d'épissage PTB (Luco et al., 2010) (Figure 12B) ou encore SRSF1 (Pradeepa et al., 2012).

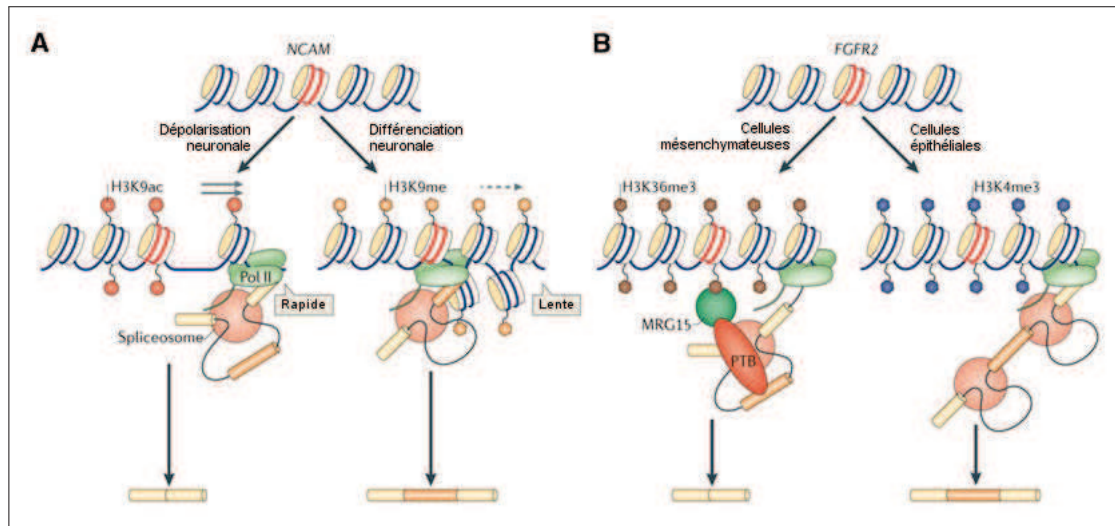


FIGURE 12: *Deux mécanismes par lesquels les modifications d'histones influencent l'épissage alternatif.* (A) Exemple de l'influence des modifications d'histones sur la vitesse d'élongation de la polymérase II dans les cellules neuronales. (B) Exemple de recrutement de facteur d'épissage par des modifications d'histones dans les cellules mésenchymateuses (Adapté de Kornblihtt et al., 2013).

Deuxièmement, de récentes études ont montré que les nucléosomes sont préférentiellement positionnés au niveau des exons (Spies et al., 2009; Tilgner et al., 2009). La longueur de la séquence d'ADN compactée au niveau d'un nucléosome est en moyenne de 150 nucléotides, ce qui correspond à la longueur moyenne d'un exon. On estime que ce fait est à l'origine de pauses par l'ARN polymérase II au début des exons laissant ainsi à la machinerie d'épissage le temps de reconnaître le site 3' d'épissage. De manière intéressante, ce positionnement particulier des nucléosomes est significativement plus observé au niveau des exons alternatifs que constitutifs. D'autre part, des études de « molécule unique » de l'ARN polymérase II ont montré

que sa vitesse d'élongation n'est pas constante et que les nucléosomes constituent des ralentisseurs pour l'enzyme (Hodges et al., 2009).

## f Influence des structures secondaires

Les molécules d'ARN peuvent adopter de nombreuses conformations grâce aux appariements intramoléculaires. On estime que la plupart des ARNs possèdent seulement des structures secondaires locales (Graveley et al., 1998). Celles-ci peuvent avoir un impact sur l'épissage. Elles peuvent, par exemple, empêcher l'assemblage du spliceosome et donc inhiber l'épissage d'un exon ou, au contraire, promouvoir l'inclusion d'un exon en masquant un site répresseur voire en rapprochant des séquences régulatrices (Jin et al., 2011) (Figure 13).

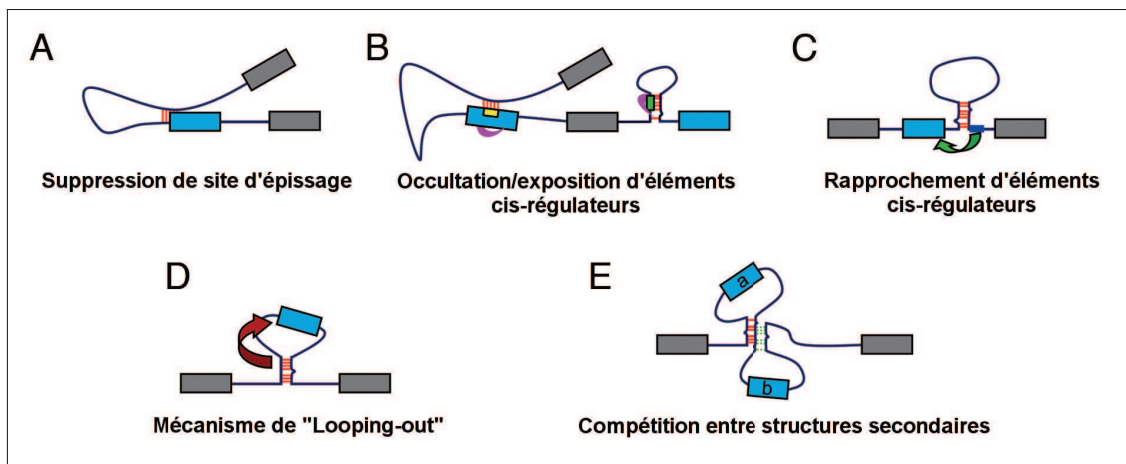


FIGURE 13: *Impact des structures secondaires sur l'épissage alternatif.* (A) *Suppression de site d'épissage.* (B) *Masquage/exposition d'éléments de régulation cis.* (C) *Rapprochement d'éléments de régulation cis.* (D) *Mécanisme de « Looping-out » : des interactions inter-introniques font sortir l'exon entraînant son exclusion.* (E) *Compétition entre structures secondaires* (Adapté de Jin et al., 2011).

Sans être un mécanisme très étudié, l'influence des structures secondaires sur l'épissage alternatif ne semble pas marginale. On trouve dans la littérature plusieurs

exemples de ce type de régulation ([Tableau 3](#)).

	Mécanisme	Exemples de gènes cibles	Références
<b>A</b>	Suppression de site d'épissage	GH, Tau, SMN1 et SMN2	Estes et al., 1992 ; Donahue et al., 2006 ; Singh et al., 2007
<b>B</b>	Masquage/exposition d'éléments de régulation cis	cTnt, SMN1 et SMN2	Warf et al., 2009 ; Miyaso et al., 2003 ; Liu et al., 2010
<b>C</b>	Rapprochement d'éléments de régulation cis	FGFR2	Baraniak et al., 2003 ; Muh et al., 2002 ; Kreaehling and Graveley, 2005
<b>D</b>	Mécanisme de « looping loop »	FGF-R1	Miriami et al., 2003
<b>E</b>	Compétition entre structures secondaires	FGF-R1	Miriami et al., 2003

TABLE 3: *Evénements d'épissage impactés par les structures secondaires de l'ARN chez l'humain.* (Adapté de [Jin et al., 2011](#)).

Dans ce contexte, les ARN hélicases apparaissent particulièrement intéressantes. Il s'agit de protéines possédant une activité enzymatique leur permettant de modifier les structures secondaires de l'ARN. Plusieurs hélicases interviennent dans la réaction d'épissage et il a été montré qu'elles peuvent jouer un rôle dans la régulation de l'épissage alternatif ([Hönig et al., 2002](#) ; [Camats et al., 2008](#) ; [Kar et al., 2011](#)).

## **E Analyse à large échelle et émergence du « code de l'épissage »**

Les technologies à haut débit développées ces dernières années ont permis l'analyse à large échelle de l'épissage alternatif. Les puces à exons ou le séquençage massif facilitent une analyse globale des événements régulés dans divers contextes biologiques (stade de différenciation, de développement, pathologiques, de traitements, de déplétions de facteurs...). Ces techniques permettent ainsi de définir de façon de plus en plus fiable le profil d'épissage alternatif dans des conditions



physiologiques comme pathologiques (Hallegger et al., 2010).

D'autre part, les séquences régulées ont ainsi pu être analysées afin d'en dégager des enrichissements de motifs (de sites d'épissage plus ou moins forts, de potentielles séquences régulatrices...); et, plus globalement, les combinaisons d'éléments cis permettant de discriminer et contrôler épissage constitutif et alternatif (Blencowe et al., 2009).

Parallèlement le développement de technologies permettant d'identifier les interactions entre protéines et ARN a considérablement complété notre connaissance de l'épissage. Ces méthodes, telles que le CLIP-seq ou sa variation le PAR-CLIP, permettent d'identifier précisément l'interaction directe des protéines susceptibles de réguler l'épissage (Darnell, 2010). C'est notamment de cette manière que l'on sait aujourd'hui que les protéines hnRNP ou SR, pendant longtemps considérées uniquement respectivement inhibitrices ou activatrices de l'épissage, peuvent en fait avoir les deux activités (Witten and Ule, 2011; Huelga et al., 2012).

La combinaison de ces différents types de données associée à des algorithmes bioinformatiques et biostatistiques a permis d'établir des « règles de l'épissage » ou « code de l'épissage » (Yeo et al., 2005; Barash et al., 2010). Ce code permet de prédire l'activité activatrice ou inhibitrice de l'épissage pour une série de facteurs dans un contexte donné. Bien que globalement bien validé (de 65 à 90% des événements tissu-spécifiques validés), ce modèle ne permet pas d'expliquer tous les événements d'épissage.

En effet, comme abordé précédemment, les éléments cis et trans ne sont pas les seuls à influencer sur l'épissage (transcription, chromatine, structures...). Il est donc nécessaire de mettre au point des techniques d'analyse à large échelle de tous ces niveaux puis de les intégrer afin de mieux comprendre comment fonctionne le mécanisme de l'épissage alternatif.

## F Epissage et pathologies

L'épissage alternatif apparaît crucial pour de nombreuses fonctions cellulaires. Des défauts de ce mécanisme peuvent donc être à l'origine directe ou être liées à de



nombreuses pathologies.

Par exemple, une mutation ponctuelle dans un gène peut modifier la structure secondaire de son ARN, modifier un site d'épissage ou un élément cis-régulateur, voire même être à l'origine d'un nouvel site d'épissage ou de cis régulation. Les conséquences d'une telle mutation peuvent alors impacter sur le contenu final du transcrit et générer la production de protéines aberrantes ou non fonctionnelles. On estime que près d'un tiers des maladies causées par des SNPs (mutation d'un seul nucléotide) a pour conséquence une modification d'épissage ([Singh and Cooper, 2012](#)). De nombreux exemples de ce type sont maintenant décrits. Parmi ceux-ci, le syndrome de progeria Hutchinson-Gilford (HGPS) dont la mutation dans l'exon 11 du gène codant pour la lamine A entraîne la création d'un nouveau site d'épissage 5' fort. L'utilisation de ce site au détriment du site naturel aboutit à la production de la progérine, une protéine ne possédant pas un domaine essentiel à sa fonction ([Sandre-Giovannoli and Lévy, 2006](#)). Un autre exemple est celui de la démence frontotemporale avec Parkinsonisme, des mutations au niveau de l'exon 10 du gène Tau affecte son inclusion normalement régulé par de multiples facteurs ([Andreadis, 2006](#)).

Les mutations impactant les éléments cis ne sont pas les seules à pouvoir altérer l'épissage, la dérégulation d'éléments trans a été démontrée dans de nombreuses pathologies. L'effet de ces altérations peuvent être pléiotropique et d'autant plus préjudiciable. Par exemple, la dystrophie myotonique de type 1 est causée par une expansion anormale de CTG dans la région 3'UTR du gène DMPK. Ces répétitions forment des agrégats nucléaires dans lesquels sont séquestrés des facteurs d'épissage reconnaissant ce type de séquence, tels que les facteurs MBNL1 et CUGBP1. Cette séquestration entraîne une perte de fonction de ces deux facteurs entraînant de multiples défauts d'épissage directement liés aux caractéristiques phénotypiques de la maladie ([Lee and Cooper, 2009](#)).

De nombreux événements d'épissage apparaissent également dérégulés dans divers cancers même si le lien de cause à effet est encore actuellement peu connu. Suite à une mutation ou à une altération d'expression de facteur d'épissage, on observe des épissages aberrants de l'oncogène CD44 ([Zerbe et al., 2004](#)) ou des gènes

suppresseurs de tumeurs comme BRCA1 ou KLF6 (Liu et al., 2001, Narla et al., 2005).

## G Conclusion

Des études récentes montrent que l'épissage influence les différents aspects de la régulation des gènes ; structure chromatinienne, transcription, stabilité ou encore transport des ARNs sont couplés à, et en partie coordonnés par, la machinerie d'épissage (Braunschweig et al., 2013). Cette complexité met en avant la nécessité d'intégration des différents niveaux de régulation des gènes afin de mieux comprendre comment l'épissage s'intègre dans ce réseau complexe. Dans cette optique, l'étude de protéines « multitâches » est d'un intérêt tout particulier. En effet, les protéines impliquées à différents niveaux de régulation sont de bons candidats pour comprendre l'intégration de tous ces aspects. Des données récentes laissent penser que ce type de protéines est beaucoup plus fréquent que ce qui est aujourd'hui admis (Baltz et al., 2012 ; Castello et al., 2012) et laisse entrevoir leur importance.

## II Les hélicases ARN DDX5 et DDX17

### A Préambule

Les hélicases ARN sont des enzymes utilisant l'ATP pour se lier à l'ARN, remodeler les structures secondaires ou les complexes ribonucleoprotéiques (RNP) formés par l'ARN. Ubiquitaires et extrêmement conservées ([Anantharaman et al., 2002](#)), elles constituent la plus grande classe des protéines du métabolisme des ARNs ([Tanner and Linder, 2001](#)). Elles ont été impliquées à toutes les étapes, de la transcription à la traduction.

Les hélicases (ADN et ARN) ont été classées en six grandes superfamilles (SFs) en fonction de leur séquence, structure et fonctions biologiques ([Singleton et al., 2007](#)). On peut dégager deux grandes catégories parmi ces six superfamilles : les hélicases formant des anneaux (SF3 à 6) et celles qui n'en forment pas (SF1 et 2). Les hélicases ARN eucaryotes appartiennent toutes à cette deuxième catégorie ([Jankowsky, 2011](#)).

En particulier, les hélicases ARN sont présentes dans six familles des superfamilles SF1 et 2 : Upf1-like de la superfamille SF1 ; Ski2-like, RIG-I-like, NS3/NPH-II, DEAH/RHA et à boîte DEAD (ou DEAD-box) qui appartiennent à la superfamille SF2 ([Figure 14](#)). La sous-famille d'hélicases ARN à boîte DEAD constitue la plus grande famille des hélicases ARN et comporte, par exemple, 37 membres chez l'humain.

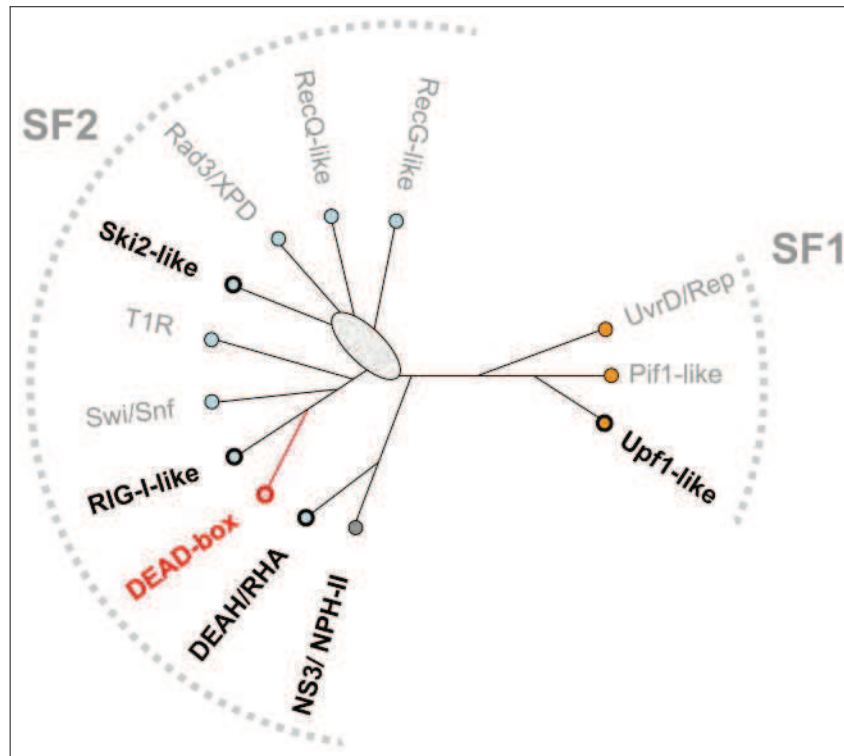


FIGURE 14: *Superfamilles SF1 et SF2 des hélicases.* En gras : les familles contenant des hélicases ARN. En rouge : les hélicases ARN à boîte DEAD. L'ovale indique une incertitude de relation dans le cladogramme (Adapté de [Jankowsky, 2011](#)).

## B Les hélicases à boîte DEAD

### 1 Structure générale

Les hélicases ARN à boîte DEAD sont constituées d'un core hélicase très conservé présentant des motifs impliqués dans la liaison et l'hydrolyse de l'ATP, la liaison à l'ARN et les interactions intramoléculaires. Ce core hélicase est constitué de deux domaines virtuellement identiques proches de la recombinaison A (RecA) chez la bactérie ([Figure 15A](#)). Ces domaines sont reliés par une séquence de liaison intermédiaire et sont relâchés en l'absence de liaison à l'ARN ([Fairman-Williams et al., 2010](#)).

Au sein des deux domaines du core hélicase, il a été identifié au moins 12 motifs caractéristiques conservés à la fois au niveau de leur séquence et de leur position

relative (Figure 15B). Néanmoins on observe des variations au sein de ces motifs entre hélicases ARN. Ils ne doivent pas être considérés individuellement, c'est leur ensemble ainsi que leur agencement qui permet de définir une hélicase ARN.

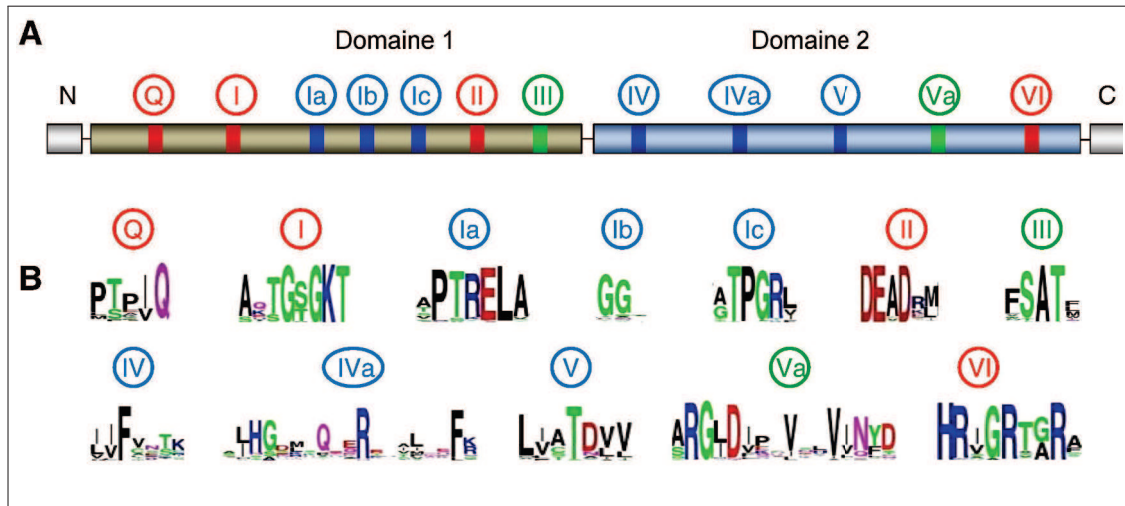


FIGURE 15: *Motifs des hélicase à boîte DEAD.* (A) Schéma de la structure primaire des hélicases à boîtes DEAD. Les différents motifs sont indiqués par des cercles de couleurs. Les couleurs utilisées indiquent leur fonction biochimique. En rouge : liaison et hydrolyse de l'ATP, en bleu : liaison à l'ARN, en vert : coordination entre la liaison à l'ATP et à l'ARN. (B) Conservation des séquences des motifs des hélicases à boîtes DEAD. La hauteur des lettres des proportionnelle à la conservation des acides-aminés. Les couleurs utilisées indiquent les propriétés chimiques des acides-aminés (vert et violet : polaire, bleu : basique, rouge : acide, noir : hydrophobe) (Adapté de Putnam and Jankowsky, 2013).

Les motifs permettant la liaison et l'hydrolyse de l'ATP d'une part et la liaison à l'ARN d'autre part sont bien identifiés (Figure 15).

Les motifs Q, I, II et VI sont impliqués dans les fonctions liées à l'ATP. Si les hélicases peuvent se lier et hydrolyser les nucléotides triphosphates (NTP) en général, elles possèdent une affinité particulière pour l'ATP. Celle-ci est essentiellement conférée par le motif Q qui se lie de manière spécifique à l'adénine. Des mutations du motif I (ou Walker A) abolissent les fonctions ATPase et hélicase de l'enzyme. En particulier, on sait que la lysine (K) forme une liaison hydrogène avec le  $\beta$ -

phosphate du complexe Mg-ATP. Le motif II (ou Walker B) contient le motif Asp-Glu-Ala-Asp (D-E-A-D) qui a donné son nom à la famille. Le premier acide aspartique (D) et le glutamate (E) interagissent avec l'ion magnésium complexé avec l'ATP, ce qui entraîne un changement conformationnel du core de l'hélicase et leur rapprochement aux motifs II et VI. Ceci permet la formation d'une poche constituant un site enzymatique actif pour l'hydrolyse de l'ATP. Les résidus impliqués dans la liaison et l'hydrolyse de l'ATP font partie des mêmes sites et sont interconnectés, l'introduction de mutations ciblant spécifiquement l'une ou l'autre fonction (liaison ou hydrolyse) est par conséquent complexe (Putnam and Jankowsky, 2013).

Les motifs d'interaction à l'ARN sont eux aussi très conservés et impliquent les motifs Ia, Ib, Ic, IV, IVa et V. Cette liaison se produit via le squelette ribose-phosphate de l'ARN, il n'y a donc pas d'interaction directe avec les acides nucléiques. Ceci explique au moins en partie pourquoi les hélicases ARN ne présentent pas de motifs d'interaction préférentiels à l'ARN (Jankowsky, 2010 ; Linder and Jankowsky, 2011).

La communication entre les sites liés à l'ATP et ceux liés à l'ARN est mal identifiée. Néanmoins il a été démontré que les motifs III, IV et Va sont essentiels à cette communication. Dans quelle mesure la liaison à l'ARN affecte l'activité ATPase (et vice-versa) semble dépendante des hélicases. Dans certains cas, elle la promeut (Lorsch and Herschlag, 1998) ; dans d'autres, c'est le contraire (Theissen et al., 2008 ; Henn et al., 2012). Les résultats contradictoires de différentes études thermodynamiques plus poussées laissent penser que les détails de ce communication sont complexes et encore non compris.

A la différence du core hélicase, les extrémités N-terminale et C-terminale ne sont pas conservées entre hélicases ARN. Elles constituent des domaines auxiliaires pouvant varier en taille (de quelques acides-aminés à plusieurs centaines) et en séquence. On considère que ces domaines sont potentiellement très importants pour la spécificité de fonction des hélicases ARN.

## 2 Mécanisme de remodelage de l'ARN double brin

Le remodelage de l'ARN par les hélicases à boîte DEAD est un processus qui nécessite l'hydrolyse d'ATP (Jarmoskaite and Russell, 2011). Ce mécanisme diffère de celui des hélicases ADN et des hélicases ARN virales (Pyle, 2008). Au lieu de se lier et parcourir l'un des deux brins que forme le duplex d'ARN, les hélicases à boîte DEAD se fixent directement au niveau de la structure double brin (Figure 16).

La liaison au duplex d'ARN peut impliquer un monomère d'hélicase, plusieurs protomères d'hélicases ou encore une hélicase associée à d'autres protéines co-facteurs (Yang et al., 2007b). La plupart des hélicases à boîte DEAD utilisent un ARN simple brin adjacent au duplex d'ARN pour leur recrutement sur la structure double brin (Tijerina et al., 2006). L'orientation particulière du simple brin, tout comme son interaction physique au duplex, n'est pas nécessaire à la réaction (Yang and Jankowsky, 2006). Le recrutement de l'hélicase à l'ARN double brin ouvre localement la structure; cette étape nécessite que l'hélicase interagisse avec une molécule d'ATP mais pas qu'elle l'hydrolyse. L'hydrolyse d'ATP est, en fait, nécessaire au relargage de l'hélicase une fois les deux brins séparés localement (Liu et al., 2008; Linder and Jankowsky, 2011) (Figure 16).

Selon les cas, cette hydrolyse peut être « productive » et aboutir à la séparation de l'ARN double brin ou « futile » lorsque le duplex d'ARN est reformé après dissociation de l'hélicase (Figure 16).

Les hélicases ARN peuvent séparer des duplex d'ARN jusqu'à 10 à 12 paires de bases de longueur, c'est pourquoi ce mécanisme de séparation est appelé « local strand separation unwinding » (Yang et al., 2007b). On note une diminution de l'efficacité du remodelage par l'hélicase avec la longueur du duplex.

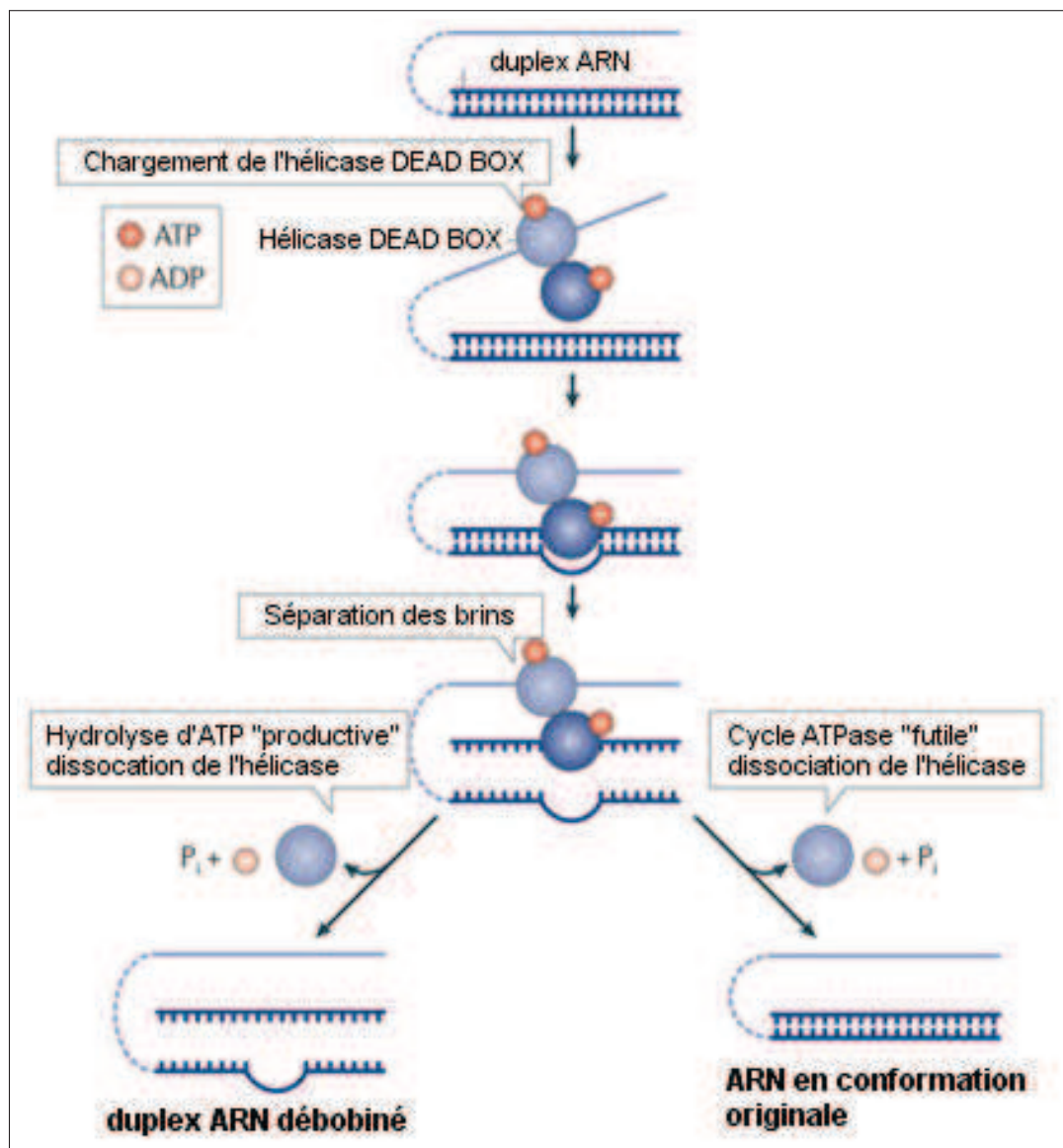


FIGURE 16: *Remodelage de l'ARN double brin par les hélicases ARN.* En foncé : région de duplex d'ARN. En clair : région simple brin. En pointillé : proximité mais pas forcément interaction physique avec le duplex d'ARN (Adapté de Linder and Jankowsky, 2011).

### 3 Diversité de fonctions

Les hélicases ARN à boîte DEAD sont des protéines multitâches, elles ont été impliquées dans quasiment toutes les étapes de régulation des gènes, et plus particulièrement du métabolisme de l'ARN. On peut dégager trois caractéristiques



biochimiques principales : leur capacité de liaison à l'ARN, de liaison aux NTPs (spécialement à l'ATP) et de liaison aux protéines.

Les fonctions moléculaires liées à l'ARN et l'ATP ont été les plus étudiées et regroupe leur capacité à :

- remodeler l'ARN (ouverture mais aussi formation de structures)
- remodeler des complexes ribonucléiques (RNP) (et notamment déplacement de protéine au sein de RNP)
- titrer certains ARNs cibles (« RNA clamping » : l'hélicase reste liée à son ARN cible et n'est plus relarguée)

Toutes ces fonctions moléculaires font intervenir leur core hélicase suggérant une thématique fonctionnelle commune à toutes les hélicases à boîte DEAD.

Néanmoins les hélicases ARN présentent également des fonctions moléculaires ARN et ATP indépendantes. En particulier, elles peuvent se lier aux protéines via leurs extrémités N-terminale et C-terminale indépendamment de leur core hélicase. Ainsi il a été démontré que certaines hélicases peuvent avoir une activité co-transcriptionnelle dans divers contextes biologiques.

Cette variété de fonctions sera illustrée de manière exhaustive dans la partie suivante traitant de deux hélicases ARN à boîte DEAD prototypiques : les protéines homologues DDX5 et DDX17.

## C DDX5 et DDX17

DDX5 et DDX17 ont été impliquées dans de multiples étapes de la régulation des gènes : transcription, épissage alternatif, export, biogénèse des ARNs ribosomiaux et biogénèse des micro-ARNs... DDX5 et DDX17 sont capables d'interagir pour former des homo ou hétérodimères, ce qui suggère qu'elles appartiennent à des voies moléculaires identiques (Ogilvie et al., 2003). Elles présentent une redondance de fonctions (Jalal et al., 2007); néanmoins, il apparaît clair qu'elles possèdent également des fonctions distinctes (Fuller-Pace, 2013).

## 1 Structure générale

DDX5 (aussi appelée p68) est l'une des premières hélicase ARN à boîte DEAD à avoir été identifiée (Lane and Hoeffler, 1980). Par la suite, DDX17 (ou p72) a été découverte car elle présente un très grand degré d'homologie avec DDX5 (Lamm et al., 1996). En effet, les deux protéines partagent 90% d'homologie au niveau du core hélicase. Les extrémités N-terminale et C-terminale sont plus divergentes et présentent respectivement 60 et 30% d'homologie (Figure 17).

D'autre part, l'ARNm de DDX17 peut être alternativement traduit à partir d'un codon d'initiation de la traduction alternatif situé en amont. Ceci entraîne la production d'une protéine plus grande de 79 acides-aminés : p82 (Uhlmann-Schiffler et al., 2002) (Figure 17). Des expériences in vitro montrent que cette protéine présente les mêmes fonctions hélicases que DDX17 (Uhlmann-Schiffler et al., 2002). Il n'existe que très peu d'études portant exclusivement sur p82, il est donc actuellement difficile de savoir si cette protéine peut, dans certains contextes, avoir des fonctions différentes de celles de DDX17 ou DDX5.

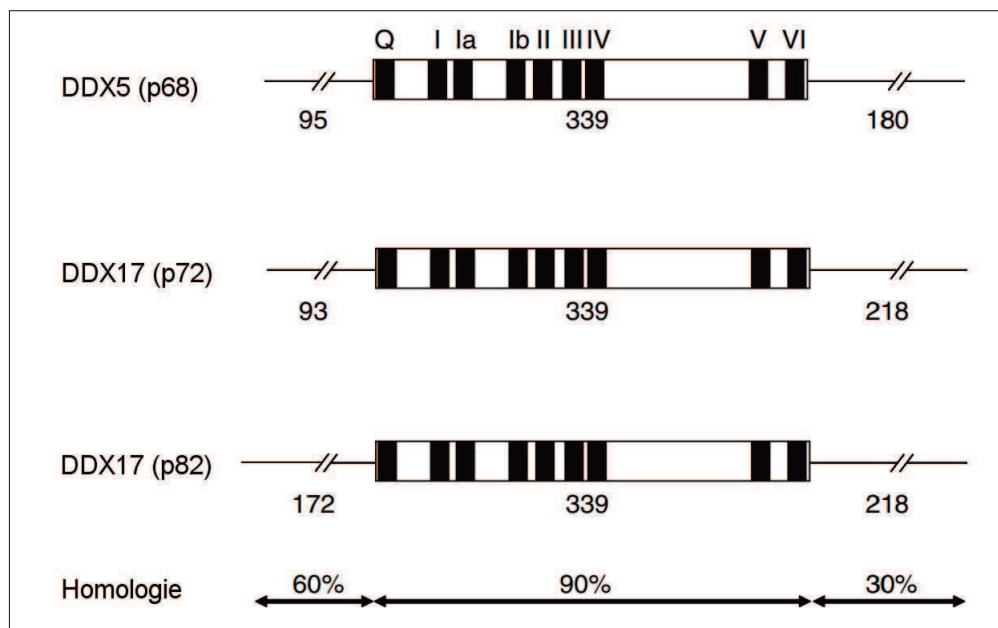


FIGURE 17: *Domaines des hélicases ARN hautement conservées DDX5 et DDX17. p82 (dérivé de l'ARNm de DDX17 mais produit à partir d'un code d'initiation en amont) est aussi représenté. En noir : les neuf motifs conservés par toutes les hélicases ARN à boîte DEAD. Le nombre d'acides-aminés des extrémités et du core hélicase est indiqué en dessous des régions correspondantes (Adapté de Fuller-Pace and Moore, 2011).*

## 2 Expression et localisation

L'expression de DDX5 et DDX17 est indispensable au développement et à la survie (Stevenson et al., 1998; Ip et al., 2000). DDX5 et DDX17 sont des protéines ubiquitaires. Néanmoins, on observe des variations importantes d'expression selon les tissus (Lamm et al., 1996, Stevenson et al., 1998). Différentes études les présentent comme majoritaires dans le cœur et le muscle squelettique (Janknecht, 2010) ou le poumon et le placenta (Rössler et al., 2000). D'autre part, leur expression apparaît régulée durant le développement (Stevenson et al., 1998; Ip et al., 2000).

Si elles sont essentiellement étudiées dans le noyau, DDX5 et DDX17 peuvent aussi être présentes dans le cytoplasme (Kahlina et al., 2004; Yang et al., 2006). Le mécanisme régulant le changement de localisation de ces deux protéines est en partie compris et implique la voie Ran GTPase (Wang et al., 2009). Une récente étude a

montré que DDX5 régule sa propre expression et celle de DDX17 par interaction avec une protéine du complexe Upf (Upf3), ce qui active de la voie NMD et entraîne la dégradation de leurs ARNs messagers (Geißler et al., 2013).

Enfin, DDX5 présente un long intron très conservé durant l'évolution (Rössler et al., 2000). Celui-ci peut être inclus par épissage alternatif; une étude a montré que les transcrits le contenant peuvent réguler négativement l'expression protéique de DDX5 (Barta and Iggo, 1995). De manière intéressante, il a plus récemment été montré que cet intron code pour un micro-ARN fonctionnel jusqu'à présent non annoté (Moore et al., 2011).

### 3 Fonctions moléculaires

#### a Rôle dans la transcription

DDX5 et DDX17 interagissent avec des facteurs de transcription et sont capables de moduler leur activité. Ce sont des co-régulateurs transcriptionnels impliqués dans l'initiation de la transcription et peuvent être co-activateurs ou co-répresseurs. Une liste des facteurs de transcription partenaires identifiés de DDX5 et DDX17 est présentée dans le [Tableau 4](#).

	Facteur	Recrutement au niveau des promoteurs ou région régulatrices cibles du facteur	Nécessaire à la transcription des gènes cibles	Autres facteurs interagissants
<b>DDX5</b>	ER $\alpha$	oui	non	SRC-1, SRA, CBP, p300, TBP, TFIIA, Brg-1, P/CAF, HDAC1/2/3
	AR	oui	oui	
	VDR			
	p53	oui	oui (pour p21WAF1)	
	Runx2	oui	oui (avec knockdown double DDX5/DDX17)	
	NF $\kappa$ b		? (mais nécessaire à la prolifération gliomale)	
	MAML1/Notch1		oui	
	E2F1	oui	oui (facteurs de réplication ADN)	
	$\beta$ -caténine	oui	oui	
nucléoplasmine	oui	oui (ADNr)		
<b>DDX17</b>	ER $\alpha$		oui	SRC-1, SRA, CBP, p300, P/CAF, HDAC1/2/3
	p53	oui sur MDM2	MDM2	
<b>DDX5 et DDX17</b>	MyoD	oui pour DDX5	oui (avec knockdown double DDX5/DDX17)	
	NFAT5		oui (avec knockdown double DDX5/DDX17)	

TABLE 4: *Fonctions de régulation de la transcription et partenaires transcriptionnels identifiés de DDX5 et DDX17. Fonctions impliquant uniquement DDX5, uniquement DDX17 et à la fois DDX5 et DDX17. (Adapté de Fuller-Pace, 2013)*

Ces deux hélicases interagissent et activent l'ARN polymérase II ainsi que des facteurs de transcription ubiquitaires tels que CBP, p300 ou PCAF, des acétyltransférases modificatrices d'histones (Rossow and Janknecht, 2003; Shin and Janknecht, 2007). D'autre part, elles interagissent et cette fois répriment la transcription via leur recrutement aux HDACs (histone déacétylases) 1, 2 et 3 (Wilson et al., 2004; Jacobs et al., 2007; Mooney et al., 2010b). Ces associations indiquent que DDX5 et DDX17 sont des éléments importants de la machinerie transcriptionnelle d'initiation notamment via leur interaction à des modificateurs

de la chromatine (Figure 18).

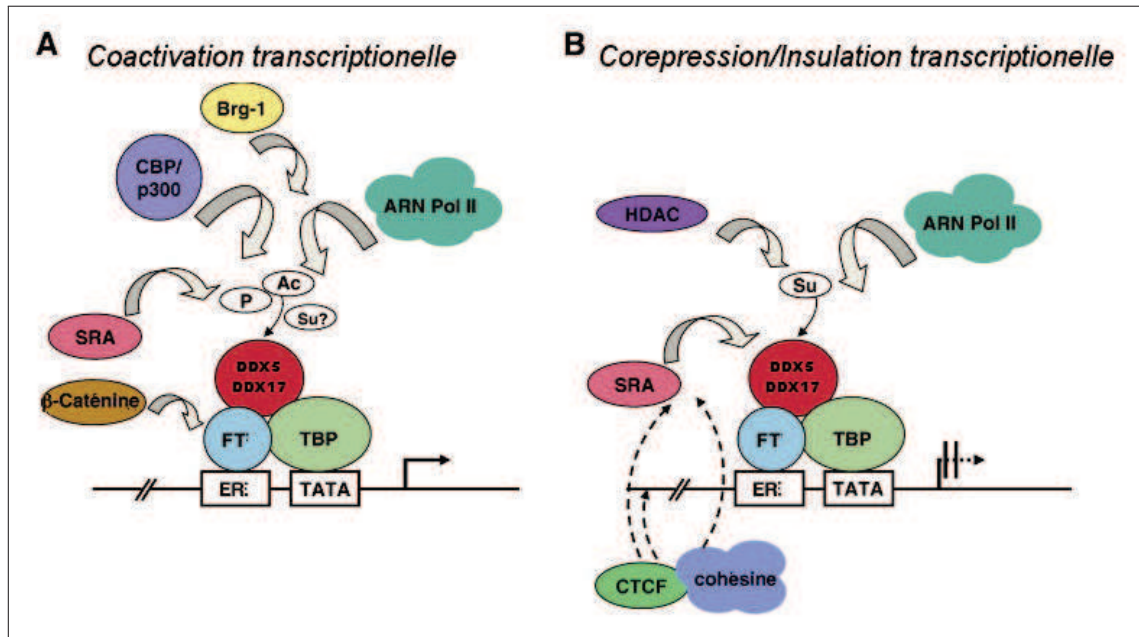


FIGURE 18: *DDX5 et DDX17 : protéines régulatrices de la transcription.* (A) *Co-activation de la transcription.* (B) *Co-répression de la transcription et insulation.* Les modifications post-traductionnelles susceptibles de modifier l'activité co-régulatrice de DDX5 et DDX17 sont aussi indiquées (Adapté de Fuller-Pace, 2013).

L'activité co-transcriptionnelle de DDX5 et DDX17 a surtout été étudiée dans les voies de signalisation de réponse aux hormones. Dans un premier temps, DDX5 et DDX17 ont été montrées être des co-activateurs de ER $\alpha$ , récepteur nucléaire de réponse aux œstrogènes (Samaan et al., 2013). Cette co-activation est décuplée par le facteur SRC1 (Steroid Receptor Co-activator 1) et est dépendante de l'interaction avec SRA. DDX5 et DDX17 apparaissent donc nécessaires à l'expression des gènes cibles de ER $\alpha$  (Wortham et al., 2009; Samaan et al., 2013). Un rôle tout à fait similaire a été décrit avec le récepteur aux androgènes (AR) (Clark et al., 2008; Samaan et al., 2013) ainsi que pour le récepteur à la vitamine D (Wagner et al., 2012).

Ddx5 est également un co-activateur du suppresseur de tumeur p53, facteur de transcription activé en réponse à divers stress cellulaires; DDX5 contrôle ainsi la

transcription de gènes impliqués dans l'arrêt du cycle cellulaire, la réparation de l'ADN et l'apoptose (Bates et al., 2005 ; Vousden and Prives, 2009).

D'autre part, DDX5 comme DDX17 en association avec SRA active l'activité transcriptionnelle de MyoD, régulateur majeur de la myogénèse (Caretto et al., 2006). DDX5 est également co-activateur de Runx2, facteur clé de la différenciation ostéoblastique (Jensen et al., 2008). Ce rôle dans la différenciation cellulaire sera détaillée dans la partie « Rôle dans la différenciation cellulaire ».

Il est intéressant de noter que la co-régulation de la transcription par DDX5 et DDX17 ne semble pas nécessiter l'activité hélicase des ces protéines (Fuller-Pace, 2013). De plus, elles ne présentent pas toujours les mêmes partenaires transcriptionnels et donc les mêmes fonctions de régulation de la transcription (Figure 18). Ces éléments mettent en évidence l'importance des extrémités N-terminale et C-terminale dans la fonction et la spécificité des hélicases ARN.

### b Rôle dans l'épissage

Des analyses par spectrométrie de masse ont montré que DDX5 est un composant du spliceosome (Neubauer et al., 1998). DDX5 et DDX17 interagissent avec la snRNP U1 qui reconnaît le site 5' d'épissage (Lee, 2002 ; Liu, 2002). Ainsi, la déplétion de DDX5 entraîne l'accumulation d'ARNs non-épissés (Lin et al., 2005). Cette fonction de régulation est hélicase dépendante dans la mesure où des mutants de DDX5 inactivant son activité hélicase ou ATPase, inhibent la dissociation de la snRNP U1 du site 5' (Lin et al., 2005).

DDX5 comme DDX17 ont été impliquées dans le contrôle de l'épissage alternatif d'une multitude de transcrits dans divers contextes cellulaires (Clark et al., 2008 ; Dardenne et al., 2012 ; Germann et al., 2012 ; Samaan et al., 2013). La régulation de l'épissage alternatif de quatre transcrits a été étudiée de manière plus poussée d'un point de vue mécanistique : CD44, H-ras, Tau et Tnnt2 (Hönig et al., 2002 ; Camats et al., 2008 ; Kar et al., 2011 ; Laurent et al., 2012). Ceux-ci ont en commun la présence d'une structure secondaire en tige-boucle au niveau de séquences cis-régulatrices qui apparaît prise en charge par DDX5 et DDX17.

Le gène CD44 présente une tige-boucle séquestrant le site 5' d'épissage en aval de son exon 4. DDX17 augmente l'inclusion de cet exon en déstabilisant cette structure (Figure 19A). DDX5 altère la structure secondaire d'une tige-boucle en aval de l'exon IDX du transcrit H-ras empêchant le facteur d'épissage hnRNP H de se fixer. DDX5 induit aussi un remodelage de la tige-boucle située en aval du site 5' de l'exon 10 du gène Tau. DDX5 permet alors à la snRNP U1 de se fixer entraînant une inclusion d'exon (Figure 19B). Enfin, DDX5 régule l'épissage de l'exon 5 de la Tnnt2 en facilitant l'interaction du facteur d'épissage MBNL1 au niveau des séquences régulatrices de l'ARNm (Figure 19C).



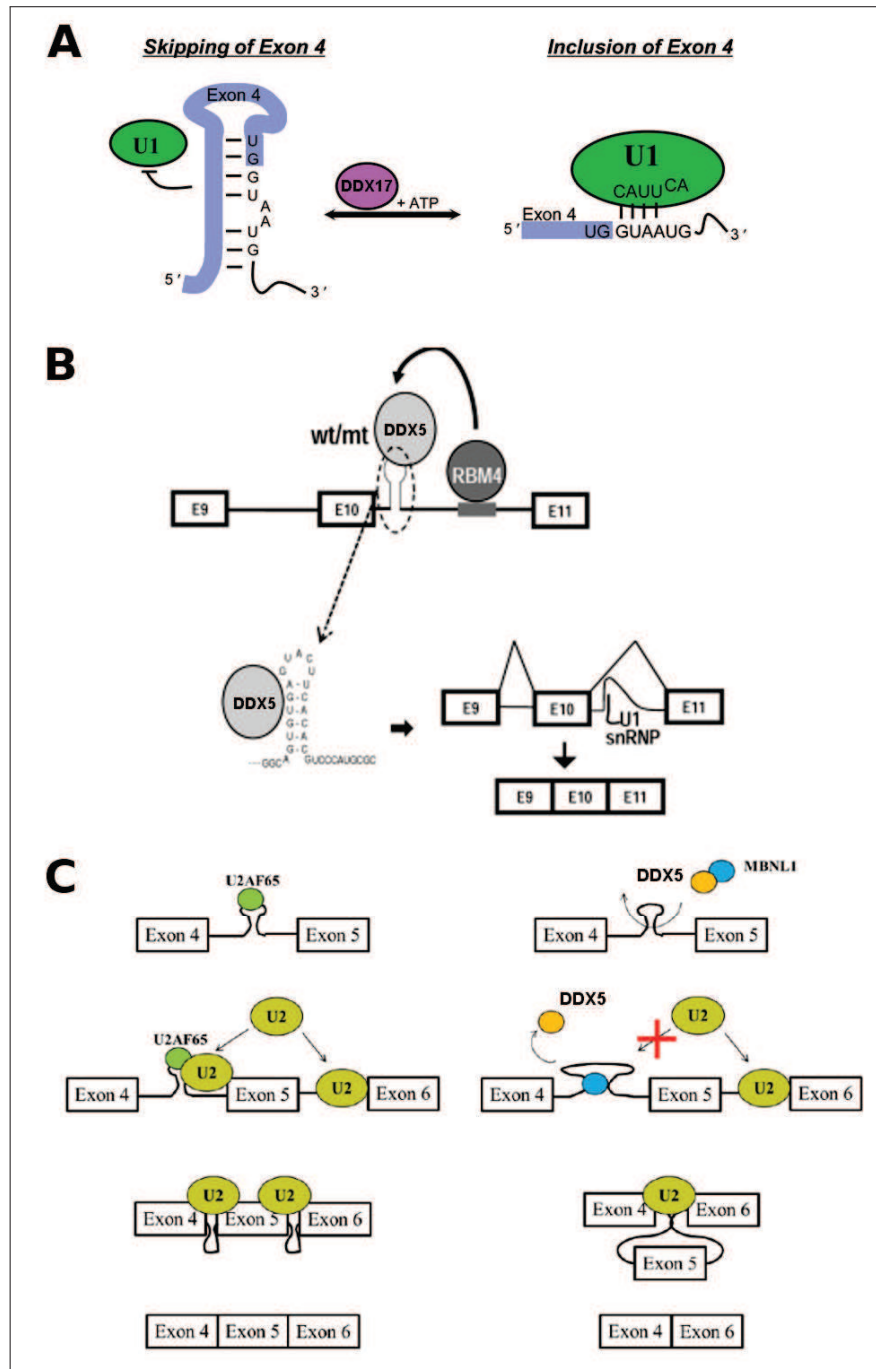


FIGURE 19: *Régulation de l'épissage alternatif par DDX5 et/ou DDX17.* (A) *Régulation de l'épissage alternatif de l'exon 4 de CD44 par DDX17.* (B) *Régulation de l'épissage alternatif de l'exon 10 de Tau par DDX5.* (C) *Régulation de l'épissage alternatif de l'exon 5 de Tnnt2 par DDX5* (Adapté de *Hönig et al., 2002, Kar et al., 2011 et Laurent, 2011*).

Ces résultats montrent que DDX5 et DDX17 sont des protéines importantes pour la régulation de l'épissage à la fois en tant que composants du spliceosome mais aussi en tant que facteurs auxiliaires pouvant faciliter ou inhiber la reconnaissance des exons.

### c Rôle dans la biogénèse des micro-ARNs

Les micro-ARNs sont de petits ARNs non codant transcrits par l'ARN polymérase II en longs transcrits primaires : les pri-micro-ARNs. Ceux-ci sont pris en charge par le complexe Drosha-DGCR8 et produisent des transcrits intermédiaires d'environ 100 nucléotides de longueur : les pre-micro-ARNs. Après export dans le cytoplasme, les pre-miR sont clivés par Dicer pour former des courts duplex d'une vingtaine de paires de bases de longueur. Enfin après séparation des deux brins, le micro-ARN mature peut interagir avec ces transcrits cibles via le complexe RISC (RNA-induced silencing complex) et entraîner leur dégradation ou inhiber leur traduction ([Winter et al., 2009](#)).

DDX5 et DDX17 participent à la biogénèse des micro-ARNs. En effet, des analyses par spectrométrie de masse ont montré que les deux protéines font parties du complexe Drosha ([Gregory et al., 2004](#) ; [Davis et al., 2010](#)). L'inactivation de DDX5 et DDX17 résulte, entre autres, en des altérations du processing de certains pri-microARNs (alors que l'expression des microARNs matures correspondants est inchangée) ([Wang et al., 2012a](#)). L'activité ATPase des hélicase est requise pour cette fonction, ce qui suggère que DDX5 et DDX17 remodèle le pri-microARN pour le rendre accessible aux coupures enzymatiques du complexe Drosha.

D'autre part, l'homologue de DDX5 chez la drosophile (Dmp68) fait partie du complexe Ago2, un composant essentiel du complexe RISC ([Ishizuka et al., 2002](#)). Ceci suggère un rôle cytoplasmique de DDX5 dans la maturation des micro-ARNs. Des données obtenues *in vitro* vont dans ce sens : DDX5 est capable de débobiner le duplex du miR let-7 ([Salzman et al., 2007](#)).

Ces résultats montrent que DDX5 et DDX17 sont des protéines importantes pour la biogénèse des micro-ARNs à la fois dans les étapes précoces se déroulant dans le

noyau et aussi dans les étapes tardives de maturation dans le cytoplasme (Figure 20).

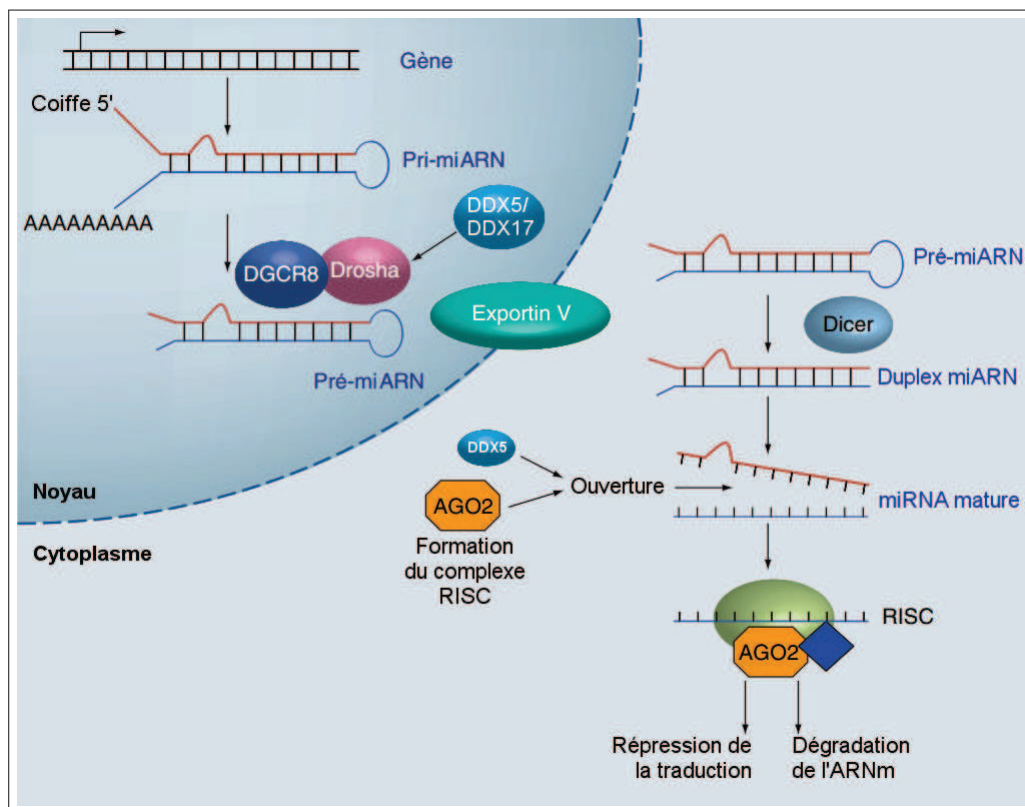


FIGURE 20: *Rôles de DDX5/DDX17 dans la biogénèse des micro-ARNs.* Les protéines DDX5 et DDX17 sont impliquées dans la biogénèse des micro-ARNs car elles font partie du complexe DROSHA. D'autre part, DDX5 interagit avec Ago2 suggérant un rôle dans le recrutement de micro-ARNs mature dans le complexe RISC (Adapté de Fuller-Pace and Moore, 2011).

#### d Rôle dans l'export cytoplasmique

Si pendant longtemps, les protéines DDX5 et DDX17 ont été considérées comme exclusivement nucléaires, des études ont montré leur présence dans le cytoplasme (Kahlina et al., 2004; Yang et al., 2006). En effet, DDX5 est capable de transiter entre le noyau et le cytoplasme grâce à des séquences de localisation nucléaire ou d'export (Wang et al., 2009). Une étude récente de notre équipe a montré l'importance de DDX5 dans l'export cytoplasmique de l'ARNm de c-fos. En assurant l'épissage co-transcriptionnel de cet ARN, DDX5 permet le recrutement et interagit avec les facteurs d'export Aly et TAP (Zonta et al., 2013).

**e Rôle dans le Non-sense Mediated Decay (NMD)**

Chez la levure, l'homologue de DDX5 (Dbp2p) a été identifiée comme interagissant avec UPF1, une hélicase ARN de la super-famille I jouant un rôle essentiel dans le mécanisme de NMD. Ainsi la déplétion de Dpd2 conduit à l'accumulation d'ARN non-sens (Bond et al., 2001). Dans des cellules humaines, une étude récente a montré que DDX5 interagit avec Upf3, une protéine du complexe intervenant dans le NMD. DDX5 est alors capable de réguler sa propre expression et celui de DDX17 (Geißler et al., 2013).

**f Rôle dans la biogénèse des ribosomes**

Une étude plus récente montre que le suppresseur de tumeur ARF inhibe l'interaction de DDX5 avec la nucléophosmine (NPM), une protéine régulant à la fois la transcription et la maturation des ARNr. DDX5 ne peut alors être recrutée au niveau des promoteurs des ADNr et limite leur expression transcriptionnelle (Saporita et al., 2011).

Ddx5 et Ddx17 sont également impliquées dans la maturation des ARNs ribosomiques chez la levure et dans des cellules humaines (Bond et al., 2001 ; Jalal et al., 2007). La déplétion de ces protéines conduit à des défauts de maturation des ARNs ribosomiques et à une perturbation des structures nucléolaires (Jalal et al., 2007).

Ces résultats montrent que DDX5 et DDX17 affecte l'expression des ARNs ribosomiques à la fois au niveau de l'initiation de leur transcription mais aussi au niveau de leur maturation.

**g Modifications post-traductionnelles**

Les modifications post-traductionnelles peuvent réguler les interactions et fonctions des protéines. DDX5 et DDX17 sont sujettes à de multiples modifications de ce type (Figure 21).

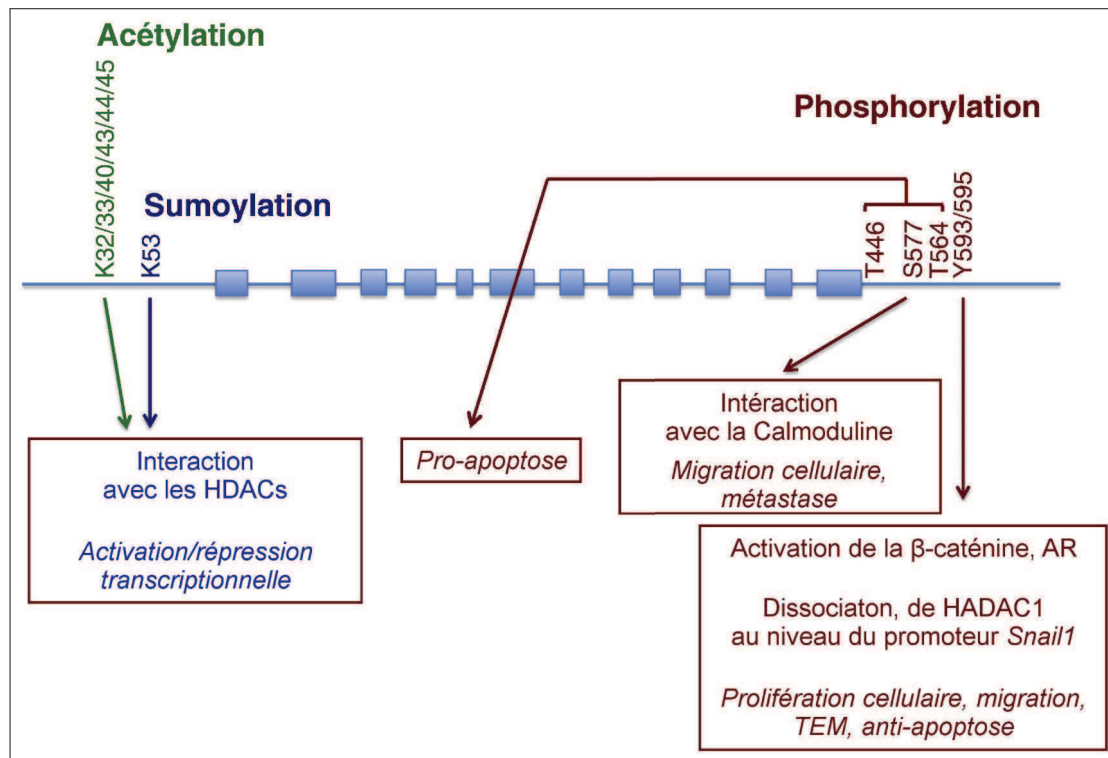


FIGURE 21: *Modifications post-traductionnelles de DDX5 et impact sur ses interactions et fonctions.* Remarque : des modifications similaires sont observées pour DDX17 (Adapté de Fuller-Pace, 2013).

### Phosphorylation

DDX5 peut être phosphorylée par PDGF (Platelet-Derived Growth Factor) en Y593, forme de la protéine qui promeut la localisation nucléaire de la  $\beta$ -caténine, protéine impliquée dans l'adhésion cellulaire et l'activation transcriptionnelle de gènes de la prolifération et de la différenciation (Yang et al., 2006). Cette phosphorylation a donc un impact sur la capacité oncogénique de DDX5.

D'autre part, DDX5 peut être phosphorylée par PKC (Protéine Kinase C) en S557 au sein d'un domaine d'interaction à la calmoduline (CaM) (Buelt et al., 1994). Bien que l'impact de cette modification n'ait pas été étudié sur l'activité de DDX5, l'interaction DDX5-CaM a été récemment décrite comme participant à la migration cellulaire dans un modèle de cancer colorectale (Wang et al., 2013). Cette modification est donc d'un intérêt tout particulier.

Globalement, la phosphorylation de DDX5 apparaît avoir un rôle pro-prolifératif,

pro-invasion et anti-apoptotique autrement dit un rôle oncogénique dans la cellule. Enfin, la phosphorylation de DDX5 en Ser(197) par MK2 a été décrite comme indispensable à sa localisation nucléaire et à son activité dans la maturation des micro-ARNs dans des cellules embryonnaires de fibroblastes (Hong et al., 2013).

### Sumoylation

DDX5 et DDX17 peuvent être sumoylées (respectivement en K53 et K50), ce qui favorise leur interaction avec HDAC1 et donc leur activité de répression de la transcription (Mooney et al., 2010b). Ainsi DDX5 sumoylée réprime la transcription de la thymidine kinase et des gènes cibles de p53 (Jacobs et al., 2007).

Cependant, dans une autre étude, la sumoylation de DDX5 a été associée à une activation de l'activité transcriptionnelle de ER- $\alpha$  sans effets sur la transcription de MDM2, bien que l'interaction avec HDAC1 est bien été vérifiée (Mooney et al., 2010b).

Il n'est donc pas encore claire si les effets de sumoylation de DDX5 et DDX17 est spécifique d'un promoteur donné ou du contexte cellulaire.

### Acétylation

Une étude a montré que DDX5 et DDX17 peuvent être acétylées au niveau de leur extrémité N-terminale (Mooney et al., 2010a). Cette modification post-traductionnelle est généralement associée à une activation de la transcription. Effectivement, l'acétylation de DDX5 et DDX17 active leur capacité à co-réguler ER- $\alpha$ , ainsi que celle de p53 (uniquement dans le cas de DDX17). De manière plus surprenante, cette étude montre que l'acétylation de DDX5 et DDX17 favorise leur interaction avec les HDACs (Mooney et al., 2010a). Comme pour la sumoylation, les mécanismes de régulation de l'acétylation de DDX5 et DDX17 ne sont pas encore bien compris.

Un récapitulatif des différentes fonctions moléculaire actuellement connues pour DDX5 et DDX17 est présenté [Figure 22](#).



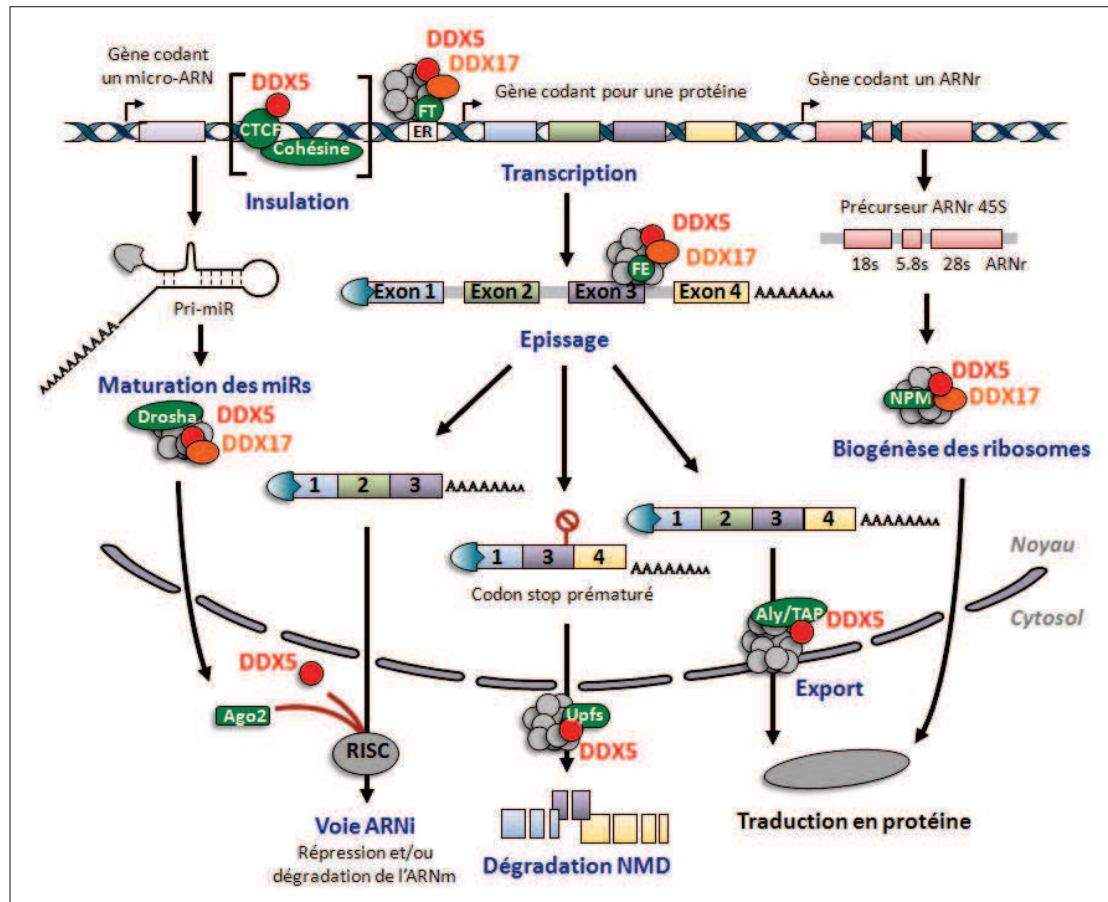


FIGURE 22: *Multiples fonctions moléculaires des hélicases ARN DDX5 et DDX17.* ER : élément de régulation de la transcription. En bleu : étapes de la régulation des gènes dans lesquelles ont été impliqués DDX5 et/ou DDX17. En rouge : DDX5. En orange : DDX17. En vert : les partenaires identifiés de DDX5 et/ou DDX17. Lorsqu'ils sont multiples des acronymes généraux ont été utilisés (FT : Facteurs de transcription, FE : Facteurs d'épissage) (Adapté de *Samaan, 2013*).

#### 4 Fonctions cellulaires

Les fonctions moléculaires multiples de DDX5 et DDX17 en font des régulateurs clés de fonctions cellulaires majeures telles que la prolifération ou la différenciation.

##### a Rôle dans le contrôle de la prolifération

Plusieurs études montrent que DDX5 coactive des facteurs de transcription

impliqués dans la régulation de la prolifération cellulaire. DDX5 interagit avec p50, sous-unité du facteur de transcription  $\text{NF}\kappa\text{B}$ , entraîne sa localisation nucléaire et active son activité transcriptionnelle dans des cellules de tumeurs gliales (Wang et al., 2012c).  $\text{NF}\kappa\text{B}$  induit l'expression de gènes de la réponse immunitaire, de l'adhésion cellulaire, de la prolifération, de la différenciation, de l'apoptose...

Bien que le mécanisme moléculaire impliqué demeure incompris, DDX5 a été montré comme interagissant avec le coactivateur transcriptionnel MAML1 dans des cellules de leucémie. MAML1 est un composant du complexe Notch1 qui active la transcription de gènes impliqués dans le développement, la prolifération et le développement (Lin et al., 2013).

D'autre part, une étude récente montre que DDX5 stimule l'expression des facteurs de réplication E2F1 en permettant le recrutement de l'ARN polymérase II au niveau de leurs promoteurs dans un modèle de cancer du sein (Mazurek et al., 2012).

Tous ces éléments montrent l'importance de DDX5 dans le contrôle de la prolifération et le développement tumorale.

### **b Rôle dans la différenciation cellulaire**

DDX5 et DDX17, associés à l'ARN non codant SRA, ont été démontrés être des activateurs transcriptionnels de MyoD, facteur de transcription majeur dans la différenciation musculaire (Carette et al., 2006). Leur déplétion induit une incapacité des cellules à entrer en différenciation montrant leur rôle essentiel dans ce processus de différenciation (Figure 23A). L'interaction de DDX5 au niveau des promoteurs des gènes musculaires a été confirmée par immunoprécipitation de chromatine en association avec le facteur de transcription TBP, l'ARN polymérase II ainsi que avec des protéines de remodelage de la chromatine (Brg-1). La déplétion de DDX5 et DDX17 entraîne une absence de recrutement de TBP, de l'ARN polymérase II et de Brg-1. Le modèle proposé est donc que DDX5 et DDX17 rendent accessible les promoteurs musculaires à MyoD et à la machinerie transcriptionnelle par recrutement de facteurs de remodelage de la chromatine (Figure 23B). Ces résultats suggèrent que DDX5 et DDX17 agissent comme des protéines adaptatrices



permettant le recrutement d'éléments de la machinerie transcriptionnelle au niveau des promoteurs à activer dans certaines conditions.

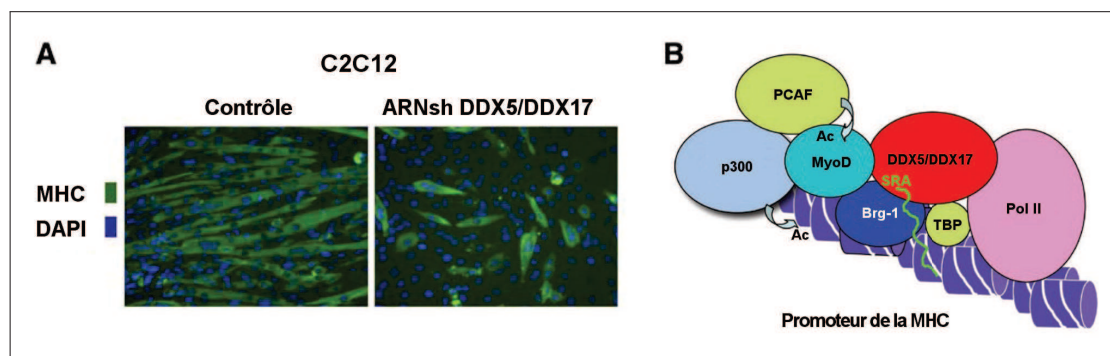


FIGURE 23: *Rôle de DDX5 et DDX17 dans la différenciation myogénique.* (A) C2C12 transfectées avec un retrovirus contrôle ou ciblant DDX5 et DDX17 puis cultivées en milieu de différenciation pendant 36h. Immunofluorescence avec un anti-corps anti-MHC. (B) Modèle d'activation transcriptionnelle de MyoD par DDX5 et DDX17 (Adapté de Caretti et al., 2006 et Caretti et al., 2007).

D'autre part, DDX5 et DDX17 sont également des co-activateurs du facteur de transcription Runx2, essentiel au développement ostéoblastique (Jensen et al., 2008). De la même manière que pour MyoD, DDX5 est recrutée au niveau des promoteurs cibles de Runx2 et la déplétion des deux protéines entraîne une inhibition de l'activité transcriptionnelle de Runx2. Néanmoins et de manière surprenante, la déplétion de DDX5 et DDX17 accélère la différenciation ostéoblastique suggérant un mécanisme de régulation plus complexe.

## 5 Rôle dans la pathologie

En accord avec le rôle de régulateur de la prolifération et de la différenciation cellulaire, il existe de nombreuses études montrant que la dérégulation de l'expression ou de la fonction de DDX5 et DDX17 joue un rôle actif dans le développement tumoral. Ces deux protéines sont surexprimées dans le cancer du sein, colorectal, de la prostate ou encore de la peau (Causevic et al., 2001; Wang et al., 2012d; Janknecht, 2010). 80% des cancers du sein expriment des marqueurs hormonaux tel

que le récepteur aux œstrogènes ER $\alpha$  dont DDX5 et DDX17 sont des co-activatrices transcriptionnelles (Wortham et al., 2009). De manière similaire, le récepteur aux androgènes AR joue un rôle important dans l'étiologie du cancer de la prostate. Tout comme dans le cancer du sein, DDX5 et DDX17 sont surexprimés et co-activatrices de AR suggérant un rôle protumoral pour ces deux hélicases (Clark et al., 2008). Dans le cancer du colon, DDX5 et DDX17 sont surexprimées de façon croissante avec la gravité des stades de cancers colo-rectaux. D'autre part, en tant que co-activatrices de la  $\beta$ -caténine, elles contrôlent l'expression des proto-oncogènes c-Myc, la cycline D1, c-jun et fra-1 (Shin et al., 2007). Enfin, DDX5 est phosphorylée dans les lignées de cancers colo-rectaux, cette phosphorylation est associée à la prolifération cellulaire (Yang et al., 2007a).

DDX5 joue aussi potentiellement un rôle dans la Dystrophie Myotonique de type 1 (DM1) (Laurent et al., 2012). Une étude récente a montré que DDX5 est capable d'aider à la séquestration pathologique du facteur d'épissage MBNL1. Cette fonction sera détaillée plus en détails dans la partie « Résultats complémentaires ».

## D Conclusion

DDX5 et DDX17 apparaissent être des protéines « multi-tâches » aux multiples fonctions moléculaires et par conséquent étant capable de contrôler des fonctions cellulaires essentielles. Ce sont par conséquent de potentielles coordinatrices de la régulation des gènes. Ce rôle est particulièrement intéressant dans des processus biologiques complexes tels que la différenciation ou le développement pathologique.

### III La myogénèse

#### A Préambule

Le tissu musculaire est un constituant fondamental de l'organisme. C'est l'un des tissus les plus abondants chez les vertébrés ; par exemple chez l'homme, il représente 30 à 40% de la masse corporelle totale. Les muscles participent à diverses fonctions essentielles tel que la respiration, la digestion, ou encore la locomotion.

Le tissu musculaire est caractérisé par sa capacité de contraction. On distingue trois principaux types musculaires ; bien qu'ayant tous des fonctions relatives à la contraction, ils se distinguent par leurs structures et fonctions. Les muscles lisses sont contrôlés par le système nerveux autonome et sont associés aux fonctions végétatives. Ainsi ils assurent le transit intestinal, la contraction des vaisseaux sanguins et globalement les fonctions de contraction involontaire (à l'exception de la respiration et des battements cardiaques). Le myocarde est constitué des cellules cardiaques assurant une contraction rythmique et continue. Il dépend aussi du système nerveux autonome. Enfin, les muscles squelettiques interviennent dans les fonctions de locomotion et de posture. Leur contrôle s'effectue via le système nerveux central et leurs contractions sont donc volontaires. Le reste de mon introduction sera focalisée en particulier sur ce dernier type de tissu musculaire.

Le muscle squelettique présente une organisation structurale caractéristique. Il est constitué d'un assemblage de milliers de fibres musculaires. Celles-ci correspondent à de longues cellules cylindriques (jusqu'à 5 cm de longueur pour 10 à 100 um de diamètre) qui résultent de la fusion de plusieurs cellules au cours du développement musculaire. Ces cellules sont donc polynucléées et leurs noyaux sont observables en périphérie.

Afin de comprendre la structure du muscle squelettique, et les pathologies qui lui sont associées (en particulier les myopathies qui sont liées à des défauts de ce type de muscle), de nombreuses études se sont portées sur la compréhension des mécanismes conduisant à sa formation : la myogénèse. Plusieurs étapes sont nécessaires au déroulement de celle-ci, ce qui en fait un processus complexe avec des acteurs divers impliquant de nombreux remaniements aussi bien au niveau de la morphologie de

la cellule, de l'architecture intracellulaire que des profils d'expression géniques.

## B Vue d'ensemble

La myogénèse est un processus ordonné constitué de plusieurs étapes successives. Chez les mammifères, lors du développement embryonnaire, le mésoderme (l'un des trois feuillets embryonnaires issus de la gastrulation) donne naissance aux somites. Ces structures contiennent des cellules précurseurs multipotentes qui sont à l'origine de plusieurs types cellulaires. Leur devenir est orienté en réponse à des signaux émis par les cellules des structures environnantes. Parmi ces morphogènes, les protéines Wnts, IGFs et Shh provoquent leur transformation en précurseurs myogéniques (Parker et al., 2003).

Ces cellules myogéniques sont aptes à quitter les somites et à coloniser le corps pour former les futurs muscles squelettiques. C'est l'étape de migration. Une fois arrivés à destination, les précurseurs myogéniques sont déterminés en myoblastes. Ces cellules spécialisées prolifèrent activement puis sortent du cycle cellulaire pour se différencier en myocytes. Ceux-ci s'alignent puis lors de l'étape de fusion, fusionnent en myotubes multinucléés. Enfin, c'est l'étape de maturation en fibres musculaires fonctionnelles (Bentzinger et al., 2012) (Figure 24).

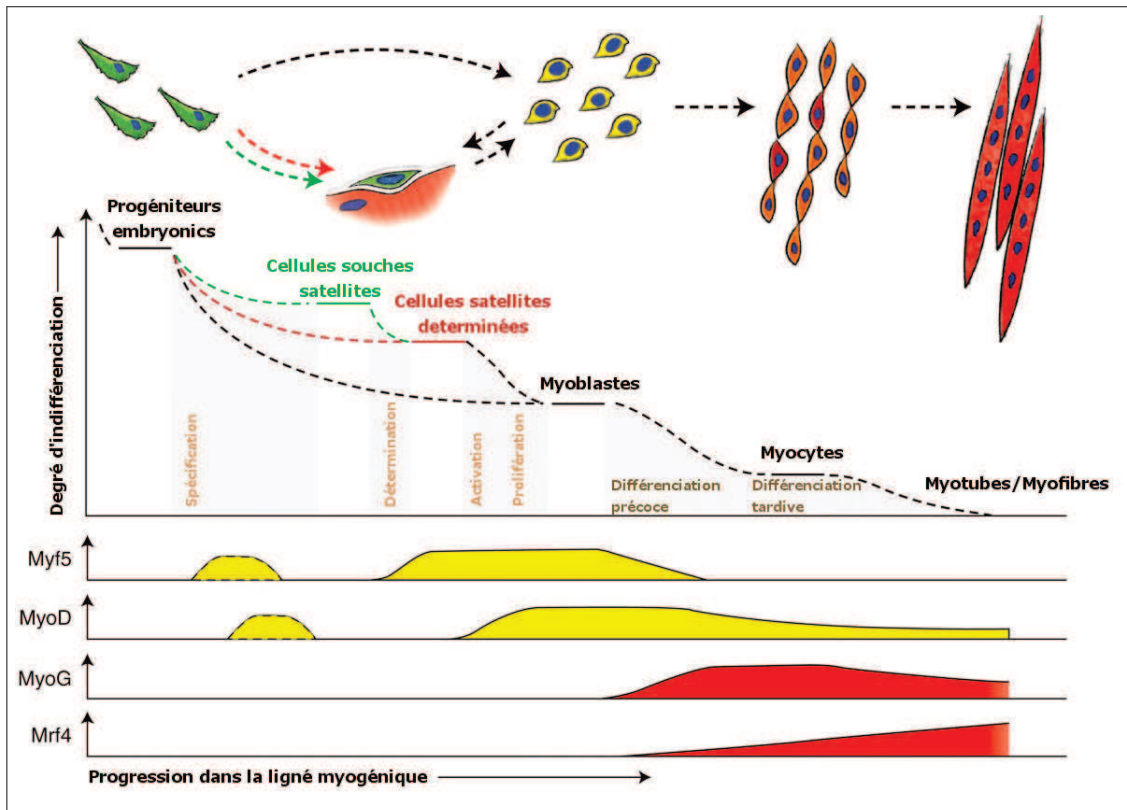


FIGURE 24: *Étapes de la myogenèse.* Grandes étapes de la myogénèse associées à l'expression séquentielle des quatre MRFs (Adapté de [Bentzinger et al., 2012](#)).

Après l'étape de migration, tous les progéniteurs ne sont pas déterminés en myoblastes. En effet, certaines cellules restent quiescentes et en périphérie des fibres musculaires. Cette localisation a donné leur nom à ces cellules dites « satellites ». On estime que ces cellules représentent entre 2 et 5% des noyaux d'une fibre musculaire mature ([Figeac et al., 2007](#)). Elles sortent de leurs quiescence et sont activées par différents signaux (stress lié à l'étirement ou à la blessure d'un muscle) pour générer des myoblastes différenciés qui vont alors poursuivre la myogénèse ([Le Grand and Rudnicki, 2007](#)) ([Figure 24](#)). Cependant une partie des cellules générées quittera de nouveau le cycle cellulaire pour assurer le stock de cellules satellites. Cette double fonction (renouvellement et capacité de différenciation) en fait des cellules souches adultes. Les cellules satellites constituent donc une source de cellules capables de réparer ou augmenter la masse musculaire dans le muscle adulte.

Les mécanismes de la myogénèse embryonnaire et de la régénération sont en

grande partie communs : en particulier, facteurs de transcription et molécules de signalisation qui sont impliqués sont les mêmes.

La myogénèse apparaît donc être un processus très ordonné et régulé, qui implique une grande variété d'acteurs moléculaires. C'est un processus complexe qui repose sur des interactions cellulaires et l'activation de voies de signalisation intracellulaires. Ces mécanismes induisent des programmes d'expression géniques spécifiques qui vont mettre en place l'identité myogénique.

## **C La lignée murine C2C12 : un modèle d'étude de la myogénèse**

L'utilisation de modèles cellulaires permet une étude simplifiée de processus de différenciation en se focalisant sur des étapes bien particulières. Si les cultures primaires ont l'avantage d'être proches de ce qui se produit in vivo dans l'organisme, elles présentent l'inconvénient de ne pouvoir être maintenues en culture que pendant un nombre de passages limité. D'autre part, elles peuvent présenter des contaminations par d'autres populations cellulaires.

C'est pourquoi l'utilisation de cellules immortalisées reste la plus fréquente dans ce genre d'étude. La lignée la plus employée dans l'étude de la myogénèse est la lignée C2C12. Il s'agit d'une lignée murine immortalisée de myoblastes. Elle a été produite à partir de cellules satellites du muscle de la cuisse de souris C3H 70h après blessure par écrasement (Yaffe and Saxel, 1977; Blau et al., 1985).

Elles ont la capacité d'entrer en différenciation en cas de privation de sérum dans le milieu de culture ou de forte densité cellulaire. Elles permettent l'étude des étapes de la myogénèse correspondant à la transformation des myoblastes en myotubes ; autrement dit : la sortie du cycle cellulaire, l'alignement, la fusion et la maturation en myotubes (Figure 25).

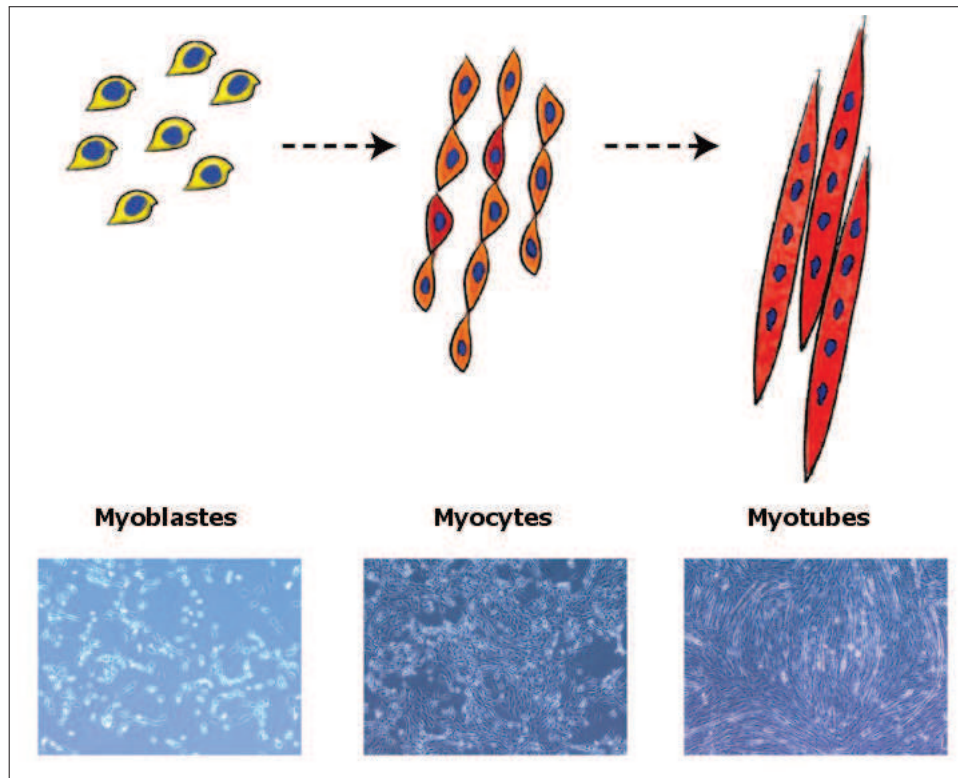


FIGURE 25: *Myogénèse des C2C12*. Etapes de la myogénèse (en haut), cellules C2C12 en culture aux étapes de différenciation correspondantes (en bas) (Adapté de Bentzinger et al., 2012).

Le comportement des cellules C2C12 en termes de réponse moléculaire mise en place lors de l'induction de la myogénèse est globalement assez proche de ceux des myoblastes primaires (Cornelison, 2008). Leur facilité d'utilisation en ont fait un modèle d'étude de choix de la myogénèse et a apporté de très nombreuses connaissances sur le déroulement de ce processus. En particulier, de nombreuses informations ont été acquises, grâce à ce modèle, sur l'expression et le rôle des facteurs moléculaires majeurs dans la différenciation myogénique : les MRFs.

## D Les MRFs contrôlent le programme transcriptionnel de la myogénèse

### 1 Structure et fonctions des MRFs

Les facteurs de transcription bHLH (basique-Hélice-Boucle-Hélice) sont des régulateurs des programmes de développement dans près de la quasi-totalité des tissus des métazoaires (Jones, 2004). Chez l'homme, on dénombre plus de 100 protéines de cette super-famille (Skinner et al., 2010). La grande majorité de ces protéines fonctionnent en dimères et reconnaissent la séquence consensus d'ADN CANNTG, appelée boîte E.

En particulier, les Muscle Regulatory Factors (MRFs) sont des protéines à domaine bHLH indispensables au développement du muscle squelettique. Chacune possède la capacité de convertir des cellules non musculaires en cellules musculaires, avec des efficacités variables, à la fois *in vitro* (Weintraub et al., 1991) et *in vivo* (Santerre et al., 1993). On dénombre quatre MRFs (MyoD, Myf5, MyoG et Mrf4); elles jouent des rôles distincts mais néanmoins chevauchants durant la myogénèse. De manière général, on estime que Myf5 et MyoD permettent la détermination des cellules progénitrices en myoblastes alors que la fusion et la maturation sont, elles, contrôlées par MyoG puis Mrf4 (Figure 24). Néanmoins, on sait maintenant que Mrf4 est exprimée au même moment que Myf5 lors de la spécification (Summerbell et al., 2002). D'autre part, en l'absence de Myf5 et MyoD, Mrf4 est aussi capable d'initier la myogénèse indiquant que Mrf4 pourrait avoir un rôle à la fois lors de la détermination et de la différenciation à proprement parler (Kassar-Duchossoy et al., 2004).

Les MRFs présentent toutes, deux motifs très conservés : un motif basique et un motif HLH (Hélice-Boucle-Hélice) (Figure 26). Le motif basique permet à ces facteurs de se fixer à l'ADN au niveau de la séquence consensus CANNTG, appelée boîte E, présente dans les promoteurs des gènes spécifiques du muscle squelettique (Massari and Murre, 2000). Le domaine bHLH des MRFs permet leur dimérisation avec d'autres facteurs à domaine HLH. Ce sont ces dimères qui se fixent au niveau des boîtes E, chaque protéine du dimère reconnaissant une moitié de motif. Les



dimères formés peuvent être des homodimères (deux protéines du même type) ou être constitués d'une MRF associée à une autre protéine bHLH appelée protéine E. Cette interaction favorise ou inhibe leur liaison à l'ADN (Olson and Klein, 1994). Les MRFs sont donc capables, avec des efficacités différentes, d'activer la transcription des gènes spécifiques du muscle.

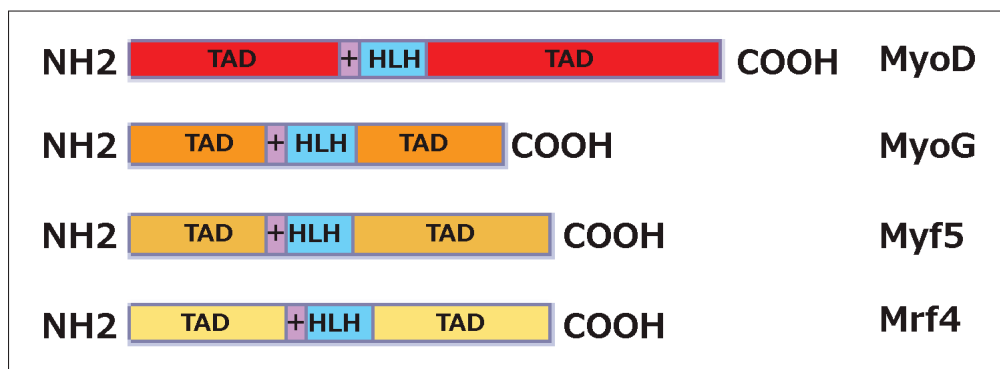


FIGURE 26: **Structure des MRFs.** Domaines structuraux des quatre MRFs *MyoD*, *MyoG*, *Myf5* et *Mrf4*. Les quatre protéines sont divergentes au niveau de leurs extrémités N-terminale et C-terminale qui contiennent les domaines d'activation de la transcription (TAD) mais présentent un degré élevé de similarité au niveau de leurs motifs basique et hélice-boucle-hélice (HLH) (Adapté de Gonçalves et al., 2011).

A la différence du motif bHLH, les extrémités N-terminal et C-terminal des quatre MRFs sont peu conservées entre elles. Elles ont été impliquées dans l'activation transcriptionnelle, le remodelage de la chromatine et les interactions protéine-protéine (Puri and Sartorelli, 2000).

Des recherches bioinformatiques montrent la possibilité de trouver des millions de boîtes E dans le génome. Pourtant, il a été montré que les MRFs ne sont capables de réguler qu'un nombre restreint de gènes. Cela montre l'importance des co-facteurs dans la régulation transcriptionnelle contrôlée par les MRFs.

L'initiation de la différenciation est également marquée par l'expression de facteurs de la famille MEF2. Chez les vertébrés, quatre gènes appartiennent à cette famille : *Mef2a*, *Mef2b*, *Mef2c* et *Mef2d*. Comme *MyoG*, ces facteurs sont absents avant l'induction de la différenciation et leur expression augmente ensuite. Ils peuvent

s'associer avec les facteurs myogéniques pour induire l'expression de gènes spécifiques du muscle et ainsi initier la différenciation terminale des myoblastes (Naya et al., 1999). Bien que n'ayant pas par elles-mêmes d'activité myogénique, les protéines de la famille MEF2 augmentent l'efficacité de conversion myogénique des MRFs (Black and Olson, 1998).

C'est donc l'action des MRFs associée à celle de leurs co-activateurs, et des protéines MEF2 qui régule l'expression de protéines spécifiques de la différenciation myogénique parmi lesquelles on trouve les marqueurs musculaires suivants : la créatine kinase musculaire (Ckm), les sous-unités du récepteur à acétylcholine (AchR), ou encore les différentes myosines (Yun and Wold, 1996).

## 2 Inhibition du programme de transcription myogénique

MyoD est exprimé dans les myoblastes avant l'entrée en différenciation et l'activation de ses gènes cibles. En effet, un certain nombre de protéines interagissent avec des protéines E ou MyoD, empêchant ainsi l'interaction de MyoD avec les protéines E et donc sa capacité d'activation de la transcription. La plupart de ces protéines appartiennent elles aussi à la famille bHLH. C'est le cas de Id, Twist, MyoR ou encore Mist-1. La protéine Id est surexprimée en conditions de culture riche en sérum (autrement dit en conditions de prolifération) et est capable de s'hétérodimériser avec MyoD. Elle séquestre ainsi MyoD l'empêchant d'activer les gènes de la différenciation (Benezra et al., 1990). Twist inhibe la myogénèse en interagissant avec les protéines E mais aussi avec les protéines Mef2 (Spicer et al., 1996). MyoR et Mist1 sont capables de former des hétérodimères avec MyoD, ces dimères sont capables de lier les boîtes E mais pas d'activer la transcription (Lemerrier et al., 1998 ; Lu et al., 1999).

D'autre part, des protéines modificatrices de la chromatine sont exprimées dans les myoblastes. Par des modifications d'histones (de type déacétylation et méthylation), elles répriment l'activité transcriptionnelle de MyoD au niveau des promoteurs des gènes de la différenciation musculaires (Puri and Sartorelli, 2000 ; McKinsey et al., 2001 ; Perdiguero et al., 2009 ; Bharathy and Taneja, 2012).

### 3 Programme transcriptionnel myogénique et régulation épigénétique

Suite aux signaux extracellulaires appropriés (morphogènes durant l'embryogénèse ou stress lié à une blessure dans le muscle adulte), la myogénèse est activée. Elle se traduit par la mise en place d'un programme transcriptionnel contrôlé par les MRFs débutant par l'activation de l'expression de p21 et une sortie irréversible du cycle cellulaire (Penn et al., 2004).

La fixation des MRFs sur leurs gènes cibles a été étudiée dans divers systèmes cellulaires à l'aide des nouvelles technologies à haut débit telles que le ChIP-chip ou le Chip-seq (Cao et al., 2010; Soleimani et al., 2012; Cao et al., 2006; Blais et al., 2005). Ces analyses à large échelle se sont particulièrement focalisées sur MyoD, facteur de transcription jouant un rôle majeur dans la myogénèse. Comme attendu, elles ont montré que MyoD se fixe sur les séquences régulatrices de gènes connus pour être surexprimés ou sous-exprimés durant la différenciation. De manière plus surprenante, elles ont aussi révélé que MyoD se fixe sur un très grand nombre de gènes qui n'apparaissent pas régulés durant la myogénèse. D'autre part, ces fixations sont observés à la fois dans les myoblastes et les myotubes différenciés.

D'un point de vue mécanistique, la régulation de l'activité transcriptionnelle des MRFs et des MEF2 est expliquée en grande partie par des modifications épigénétiques. En effet, celles-ci permettent de moduler la structure de la chromatine et donc l'efficacité de transcription (Yahi et al., 2006; Albin and Puri, 2010; Aziz et al., 2010). Les deux modifications épigénétiques les plus couramment étudiées sont les acétylations/déacétylations d'histones et les méthylation. L'acétylation est régulé par les facteurs antagonistes HDACs et HATs, et est corrélée avec une activation de la transcription. A l'inverse, la déacétylation entraîne une compaction de la chromatine et une répression de la transcription. Le rôle des méthylation apparaît plus complexe et est régulé par les HMTs et les déméthylases.

Dans les cellules non différenciées, la répression des gènes musculaires est assurée via le recrutement de HDACs et de HMTs qui interagissent avec MyoD et MEF2. Les HMTs produisent des modifications d'histones inhibitrices (H3K9me2/me3 et

H3K27me3) qui empêchent l'activité transcriptionnelle de MyoD et des MEF2. Dans les cellules différenciées, le recrutement de HATs, de HMTs, de déméthylases ainsi que des complexes de remodelage de la chromatine SWI/SNF et Brg1/Brm entraînent des modifications d'histones activatrices (H3K9ac, H3K14ac, H3K4, H3R8 et H3R17me). Celles-ci permettent une configuration ouverte de la chromatine permettant l'activation de transcription des gènes cibles de MyoD et des MEF2 (Figure 27) (Bharathy and Taneja, 2012).

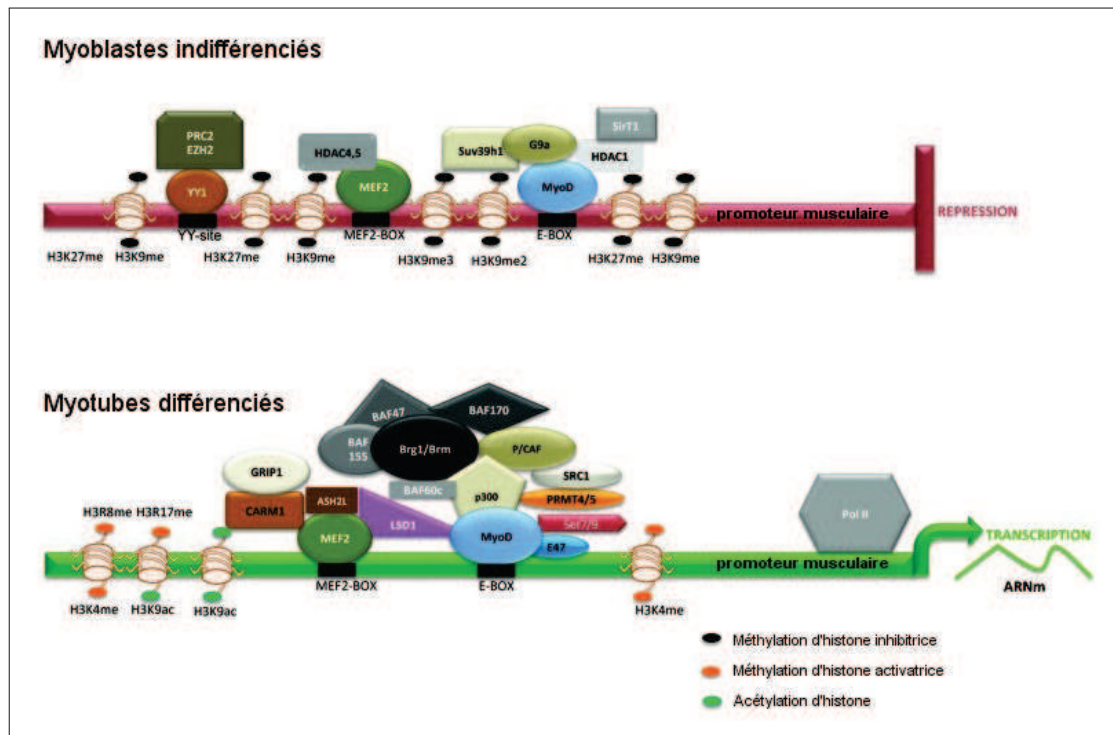


FIGURE 27: *Régulation épigénétique des promoteurs dans les myoblastes non différenciés et dans les myotubes différenciés.* (Adapté de Bharathy and Taneja, 2012).

## E Autres mécanismes de la régulation génique

### 1 Les micro-ARNs

Les micro-ARNs sont de petits ARN non codants d'une vingtaine de nucléotides très conservés dans l'évolution. Ils sont capables de réguler l'expression de gènes de manière post-transcriptionnelle en se liant à des séquences complémentaires situées

dans leurs ARN messagers cibles. Une complémentarité parfaite n'est pas nécessaire pour assurer la dégradation du transcrite ou l'inhibition de sa traduction. En effet, on sait qu'il suffit d'avoir une complémentarité au niveau d'une séquence située à l'extrémité 5' du micro-ARN : la séquence « seed ». Celle-ci peut être de 6 à 8 nucléotides de longueur (Lewis et al., 2003). Il apparaît de plus en plus évident que les micro-ARNs jouent un rôle déterminant dans de nombreux processus biologiques comprenant le développement, la différenciation cellulaire, la prolifération ou encore l'apoptose (Cremisi, 2013 ; Williams et al., 2009 ; Liu et al., 2013 ; Wang and Qian, 2014).

L'importance des micro-ARNs dans le développement musculaire est révélée, notamment, par la production de souris KO pour Dicer, enzyme essentielle à la maturation des micro-ARNs, spécifiquement dans les cellules squelettiques. Ces souris présentent un défaut de développement de la fibre musculaire et une masse musculaire réduite (O'Rourke et al., 2007).

Des analyses transcriptomiques à large échelle ont permis de révéler la variation d'expression de nombreux micro-ARNs durant la différenciation myogénique dans des lignées humaines. On dénombre 60 micro-ARNs différenciellement exprimés durant la myogénèse : 43 sont activés alors que 17 sont inhibés (Dmitriev et al., 2013). Les gènes cibles de ces miR sont impliqués dans un large éventail de fonctions cellulaires : régulation transcriptionnelle, régulation du cycle cellulaire, dégradation protéique, ubiquitination, apoptose, motilité cellulaire, transport intracellulaire... Ceci révèle le degré de complexité de la régulation par micro-ARNs durant le processus de myogénèse.

Parmi les micro-ARNs activés durant la différenciation, l'expression de miR-1, miR-133 et miR-206 est particulièrement induite et de nombreuses études ont montré leur rôle essentiel dans la différenciation musculaire squelettique (Chen et al., 2006 ; Kim et al., 2006 ; Huang et al., 2011). Ces micro-ARNs considérés comme spécifiques du muscle sont appelés MyomiRs ; ils présentent des similarités d'architecture et de fonction. Leur expression est régulée à la fois par les MRFs et les Mef2 ainsi que par d'autres facteurs de différenciation myogéniques (Rao et al., 2006 ; Sweetman et al., 2008) (Figure 28A).

Les couples miR-1-1/133a-2, miR-1-2/133a-1, and miR-206/133b sont produits en transcrits bi-cistroniques à partir de chromosomes différents (Figure 28B ). Les formes matures de miR-1 et miR-206 possèdent la même séquence seed et ; la séquence de miR-133a et miR-133b ne diffère que de deux nucléotides.

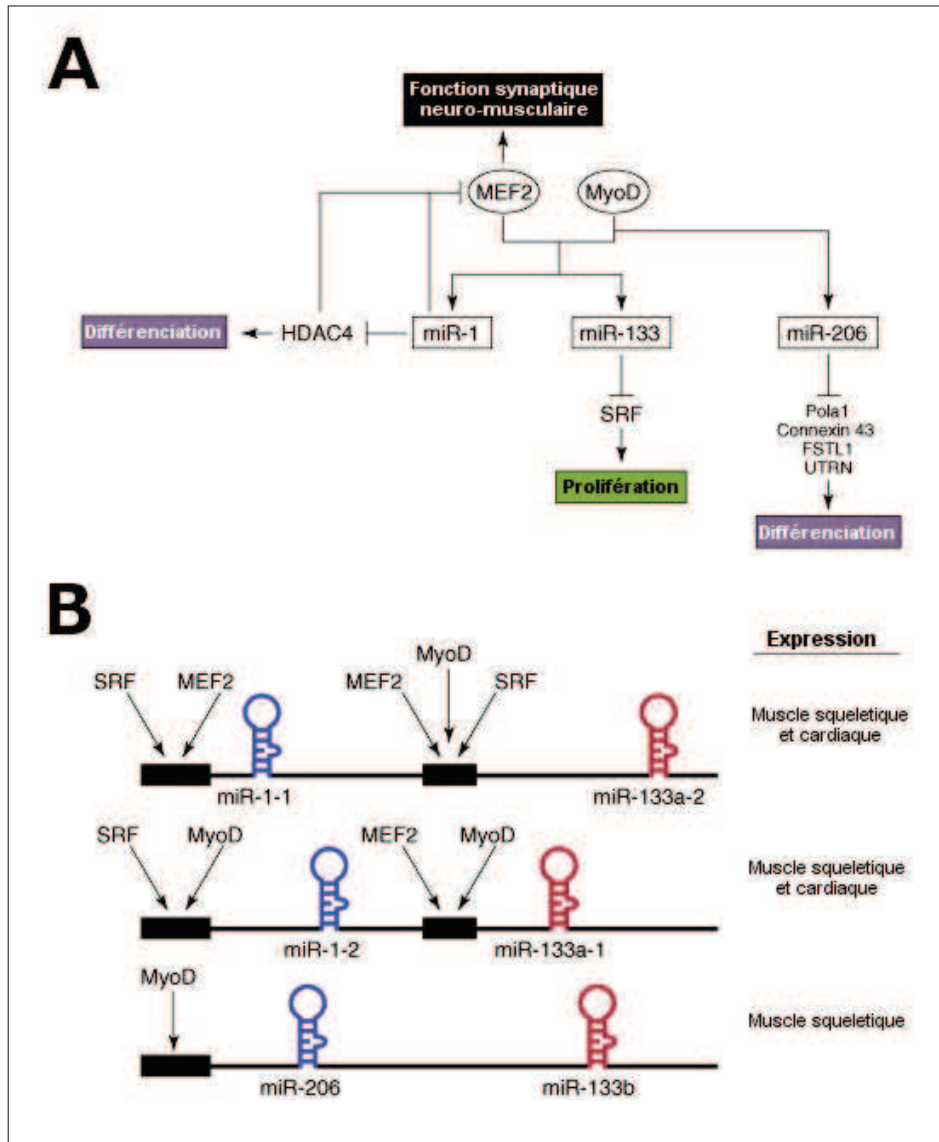


FIGURE 28: *MyomiRs*. (A) *Circuits transcriptionnels dans lesquels sont impliqués les myomiRs*. (B) *Clusters bicistroniques des myomiRs*. Les éléments régulateurs cis sont indiqués sous la forme de boîtes noires (Adapté de Williams et al., 2009).

miR-206 participe à la myogénèse en inhibant la protéine Pola1, plus grande sous-unité de l'ADN polymérase  $\alpha$  (Kim et al., 2006). L'arrêt du cycle cellulaire et

l'inhibition de synthèse d'ADN qui en découlent, sont deux événements clés à l'entrée en différenciation. D'autre part, miR-206 inhibe aussi l'expression des facteurs de transcription inhibiteurs de la différenciation Pax3 et Pax7 (Chen et al., 2010; Hirai et al., 2010).

Bien que présentant la même séquence seed et des profils d'expression similaires, miR-1 et miR-206 présentent des caractéristiques et rôles propres. En particulier, si miR-206 est exclusivement exprimé dans les muscles squelettiques, miR-1 est aussi exprimé dans le cœur. miR-1 participe à la myogénèse en inhibant HDAC4, répresseur transcriptionnel de nombreux gènes musculaires et notamment de Mef2. miR-133 participe à la différenciation myogénique en régulant l'expression de MAL1, IGF-1R et en particulier celle de nPTB. Ce facteur d'épissage contrôle les profils d'épissage lors de la différenciation neuronale et musculaire (Boutz et al., 2007).

En plus de ces miR spécifiques du muscle, d'autres micro-ARNs ont été démontrés essentiels à la différenciation myogénique tel que miR-24, miR-214 ou encore miR-26a (Sun et al., 2008, Wong and Tellam, 2008; Feng et al., 2011).

## 2 L'épissage alternatif

Les gènes codant pour les protéines de la contraction musculaire ont été parmi les premiers pour lesquels on a observé de l'épissage alternatif (Andreadis et al., 1987). Pour cette raison, de nombreux gènes du muscle ont servi de modèle pour étudier les mécanismes de l'épissage tels que les gènes de la troponine T ou de la tropomyosine (Breitbart et al., 1985; Gooding and Smith, 2008; Goo and Cooper, 2009). Ainsi de nombreuses protéines de liaison à l'ARN ont été impliquées dans l'épissage alternatif d'exons spécifiques du muscle : les membres de la famille CELF (Ladd et al., 2001), de la famille MBNL (Pascual et al., 2006), de la famille FOX (Kuroyanagi, 2009), PTB (Spellman and Smith, 2006), hnRNP H (Chen et al., 1999)...

Plus récemment, les nouvelles technologies à haut débit ont permis une meilleure compréhension des programmes d'épissage mis en place lors du développement musculaire, de la différenciation ou dans le muscle adulte. En particulier, l'analyse globale des événements d'épissage régulés dans une grande variété de tissus a permis



de confirmer que le muscle est l'un des tissus présentant le plus de variations d'épissage (Castle et al., 2008). Si la plupart de ces études transcriptomiques permettent de dégager des caractéristiques d'épissage communes aux muscles squelettiques, celles-ci n'apparaissent pas muscle-spécifiques. En effet, une grande partie d'entre elles a été montrée comme partagée par les cellules neuronales (Grabowski, 2011).

La différenciation est un processus dynamique composé d'une succession temporelle d'étapes. Les analyses à large échelle réalisées ont permis de dégager des groupes d'exons co-régulés à ces différents points de différenciation myogénique et ce, dans diverses espèces (Kalsotra et al., 2010 ; Kalsotra et al., 2008 ; Bland et al., 2010). Les événements d'épissage alternatif conservés entre espèces impliquent des gènes codant pour des protéines du cytosquelette mais aussi, et de manière plus intéressante, pour des facteurs de régulation d'épissage ou plus globalement pour des protéines de liaison à l'ARN. Ceci met en avant l'importance de la régulation de l'ARN et de l'épissage dans la mise en place du phénotype musculaire.

Les analyses de motifs réalisées ont permis de déterminer une série de motifs spécifiques de la myogénèse associés à l'inclusion ou l'exclusion d'exon (Figure 29). On observe un enrichissement net en motifs pour les protéines FOX, PTB et STAR. De manière plus surprenante, des motifs pour le facteur d'épissage hnRNP L sont aussi enrichis bien qu'il n'ait, jusqu'à maintenant, pas été associé de manière spécifique à l'épissage du muscle.



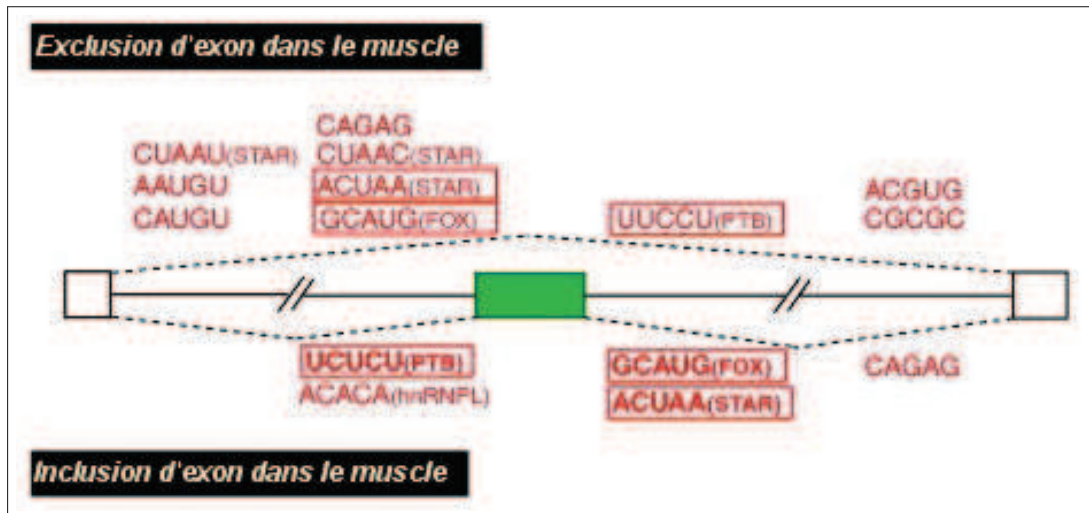


FIGURE 29: *Motifs d'épissage enrichis durant la différenciation musculaire.* Enrichissement de motifs associés à l'exclusion d'exon (en haut) et l'inclusion d'exon (en bas). En gras : les motifs conservés entre la souris et l'oiseau. Données provenant de (Barash et al., 2010 ; Kalsotra et al., 2010 ; Bland et al., 2010). Les motifs pour lesquelles les protéines d'interaction sont connues présentent le nom du facteur correspond entre parenthèses (Adapté de Llorian and Smith, 2011).

## IV DM1

### A Préambule

La dystrophie myotonique de type 1 (ou myopathie de Steinert, couramment abrégée DM1), est la forme adulte de dystrophie musculaire la plus commune. Au niveau mondial, on estime qu'en moyenne environ une personne sur 20.000 est atteinte de cette maladie (Bird, 2013). C'est une pathologie rare dont la prévalence peut varier de un cas sur 100 000 au Japon jusqu'à un cas sur 500 dans des régions spécifiques telles que le nord-est du Québec (Turner and Hilton-Jones, 2010; Yotova et al., 2005).

Elle est caractérisée principalement par un affaiblissement et une atrophie musculaire progressive, et par une myotonie (difficulté à relaxer un muscle ou un groupe de muscles une fois qu'ils ont été contractés). C'est une maladie multisystémique et dégénérative, touchant un vaste éventail de tissus en plus des muscles.

Maladie héréditaire autosomique dominante, la DM1 peut affecter autant les individus de sexe masculin que ceux de sexe féminin. L'expression des symptômes est très variable d'une personne à l'autre, même au sein d'une même famille. D'autre part, on observe un phénomène d'anticipation très marquée de la maladie c'est-à-dire une aggravation de génération en génération au sein d'une même famille et, également, une apparition de plus en plus précoce (Martorell et al., 1998).

La mutation à l'origine de la DM1 est bien caractérisée : il s'agit d'une expansion anormale de répétitions CTG située dans la région 3'UTR non codante du gène DMPK. Le nombre de ces répétitions est directement corrélé à la sévérité de la maladie. Un nombre de répétitions supérieur à 37 est associé à la pathologie avec différents degrés de sévérité (Figure 30) (Bird, 2013).

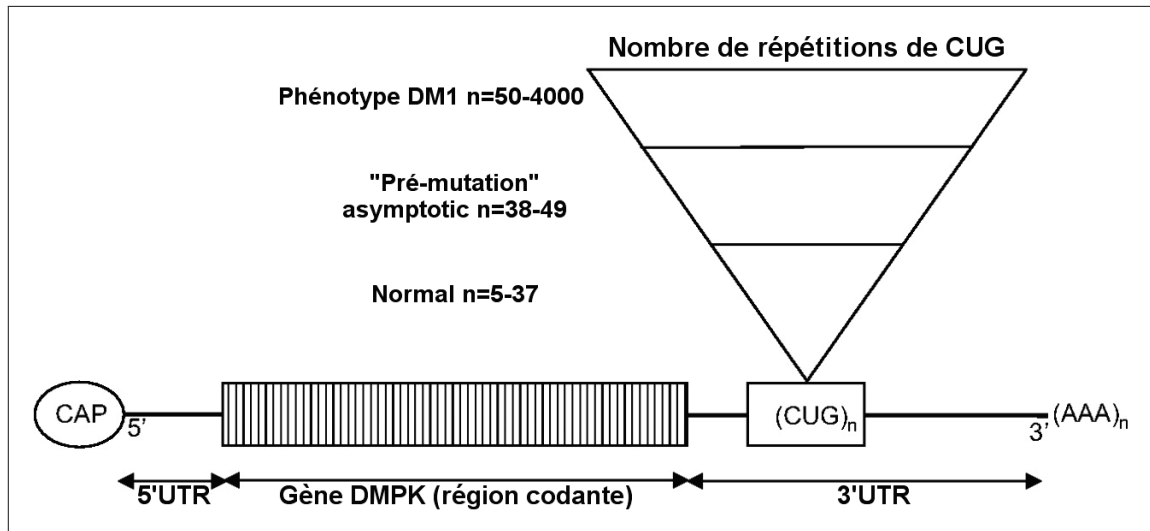


FIGURE 30: *Mutation pathologique de la DM1*. Représentation de l'ARN pré-messager de la DMPK. Les répétitions CUG sont présentes dans la région 3'UTR et une relation entre le nombre de triplets et la sévérité de la pathologie est indiquée (Turner and Hilton-Jones, 2010).

## B Différentes formes

On distingue 4 formes de DM1 selon la sévérité et l'âge des premiers symptômes. Le nombre de CTG est donné à titre indicatif. En effet, la très grande variabilité dans la manifestation des symptômes d'un sujet à l'autre empêche l'établissement d'une corrélation précise entre le nombre de CTG et la sévérité de la maladie (Harper and Harper, 2001).

### 1 Forme congénitale

C'est la forme la plus sévère des DM1. Elle est associée à une transmission majoritairement maternelle (Harper and Harper, 2001), et se caractérise par de larges expansions de répétitions d'un millier à plusieurs milliers de CTGs.

Dès la grossesse, on observe une diminution de la mobilité foetale et l'enfant présentera à la naissance une hypotonie généralisée, une détresse respiratoire, une faiblesse musculaire généralisée (notamment des muscles respiratoires et faciaux) et

des difficultés d'alimentation (liées à des troubles de la succion et de la déglutition). Si l'enfant survit à la période néonatale, on constatera souvent un retard du développement psychomoteur (Spranger et al., 1997 ; Harper and Harper, 2001). Entre 10 et 20 ans, les autres symptômes d'une myopathie de forme adulte apparaissent et les patients développent souvent de graves problèmes cardiorespiratoires autour de 30 ou 40 ans (Turner and Hilton-Jones, 2010).

## 2 Forme infantile

Les symptômes de cette forme de DM1 peuvent être visibles à partir de 10 ans. L'atteinte musculaire et neurologique peut être discrète (notamment l'absence de myotonie) et rend le diagnostic plus difficile (Turner and Hilton-Jones, 2010). En revanche, les troubles du comportement sont souvent visibles et se traduisent par des difficultés d'apprentissage, un retard scolaire et des problèmes d'insertion sociale. Certaines études retrouvent aussi un comportement autistique chez certains enfants atteints par la DM1 (Ekström et al., 2008 ; Angeard et al., 2011).

## 3 Forme adulte

Il s'agit de la forme « classique » de la DM1, elle est caractérisée par une centaine à un millier de triplets CTG. Les symptômes débutent entre 20 et 40 ans et sont très variables : faiblesse musculaire, myotonie, cataracte, résistance à l'insuline, troubles du sommeil et du comportement... Il faut noter que l'âge d'apparition, la sévérité des symptômes ainsi que le développement de la maladie varie énormément d'un individu à un autre.

Les symptômes et aspects cliniques de la forme adulte de la maladie sont décrits plus en détails dans la partie « Aspects cliniques ».

#### 4 Forme tardive

Les individus présentant cette forme de DM1 possèdent en général moins d'une centaine de répétitions CTGs. Cette forme se caractérise par des symptômes plus discrets essentiellement représentés par une cataracte, une myotonie légère parfois asymptomatique ou encore une calvitie frontale précoce. L'espérance de vie est normale (Machuca-Tzili et al., 2005; Arsenault et al., 2006). Si les symptômes de cette forme de DM1 sont beaucoup plus discrets, elle représente une source non négligeable de transmission à la descendance des formes plus graves.

D'autre part, des individus peuvent être porteurs d'un nombre conséquent de triplets CTG (une cinquantaine de répétitions) sans présenter aucun des symptômes de la maladie. Sans être cliniquement atteints, ils sont capables de transmettre la mutation pathologique, cette fois amplifiée, à leurs enfants. Ce phénomène est appelé pré-mutation (Martorell et al., 2001).

### C Aspects cliniques

L'atteinte musculaire présentée par les patients DM1 constitue un schéma bien précis et caractéristique de la maladie. Le symptôme prédominant est la faiblesse des muscles distaux, des muscles superficiels du visage et du cou. Le diaphragme, les muscles intercostaux et oculaires externes sont aussi souvent atteints. D'autre part, une myotonie musculaire est fréquemment observée et se traduit par des spasmes toniques des muscles dus à l'incapacité de relaxation après une contraction volontaire (Bird, 2013).

Les muscles lisses (ou involontaires) sont aussi très souvent atteints. Ceci se traduit par un dysfonctionnement de l'œsophage, du pharynx mais aussi du système gastro-intestinal (Bellini et al., 2006).

La DM1 est une pathologie multisystémique touchant d'autres organes que les muscles squelettiques et lisses. L'atteinte cardiaque est quasi-systématiquement présentée par les patients. Des défauts de la conduction cardiaque sont observés

à des degrés de sévérité variable, et les cardiomyopathies sont une cause fréquente de mortalité chez les patients (Petri et al., 2012).

Des troubles du système nerveux central et de la cognition sont également présents. En particulier, les apnées du sommeil, l'hypersomnie et la somnolence sont caractéristiques et bien documentés (Laberge et al., 2009; Dauvilliers and Laberge, 2012). Les troubles cognitifs présentés par les malades sont très variables et regroupent un retard intellectuel plus ou moins sévère, des difficultés d'apprentissage et d'interaction sociale. Du point de vue comportemental, on observe des patients anhédoniques, apathiques ou encore obsessionnels (Winblad et al., 2005; Antonini et al., 2006; Gaul et al., 2006).

Une dérégulation endocrinienne s'observe et peut être à l'origine d'endocrinopathies incluant l'hyperinsulinémie (pouvant entraîner un diabète de type II), le dysfonctionnement de la thyroïde, la dérégulation calcique ou encore l'atrophie testiculaire pour les hommes (Orngreen et al., 2012).

Une cataracte précoce de type polychromatique, avant l'âge de 40 ans, est une autre caractéristique de la DM1.

Enfin, les patients DM1 présentent d'autres symptômes récurrents mais difficile à lier à un organe en particulier tels que la fatigue (Kalkman et al., 2005), une prédisposition à certains cancers (Win et al., 2012), une calvitie précoce pour les hommes ou encore une réaction anormale aux médicaments (Veyckemans and Scholtes, 2013)...

## D Aspects moléculaires

### 1 Mutation pathologique

La mutation à l'origine de la DM1 est une expansion anormale de triplets CTG. Une expansion de CTG est aussi présente chez les individus sains mais le nombre de répétitions CTG varie entre 5 et 37. D'autre part, celui-ci reste stable au cours des générations. Au-delà de ce seuil, chez les individus DM1, les expansions deviennent instables et ont une forte tendance à l'amplification. Les séquences nucléotidiques

répétées augmentent en taille d'une génération à la suivante, qui présentera alors une forme plus grave de la maladie à un âge plus précoce que la génération précédente : c'est le phénomène d'anticipation ([Harper and Harper, 2001](#)).

D'autre part, le nombre de répétitions des triplets est très instable au sein des tissus d'un même individu : c'est la mosaïque intertissulaire de la mutation ([López Castel et al., 2010](#)). Celle-ci est généralement associée à des phénotypes tissu-spécifiques. Autrement dit, les tissus les plus atteints dans la maladie sont souvent ceux qui présentent un nombre de répétitions de triplets plus important. D'autre part, cette variation peut aussi être observée au sein d'un même tissu.

Plusieurs mécanismes semblent à l'origine de cette instabilité intergénérationnelle et tissulaire. Le premier est le dérapage de la fourche de réplication lors de la réplication de l'ADN. En effet, il a été montré *in vitro* que l'ADN polymérase présentait des défauts d'activité au niveau des triplets répétés à fort contenu en GC ([Pelletier et al., 2003](#)). Cependant, un autre mécanisme est nécessaire pour expliquer l'instabilité dans les cellules post-mitotiques dont font partie les cellules musculaires et neuronales. Différentes études ont montré le rôle des protéines de réparation de l'ADN dans l'instabilité des triplets. En effet, il est possible que celles-ci ne soient pas capables de réparer les structures secondaires formées par les séquences répétées, entraînant une réparation erronée de l'ADN qui résulte en une augmentation de la taille des répétitions ([Gomes-Pereira and Monckton, 2006](#)).

Le gène DMPK, sur lequel on trouve les répétitions pathologiques, code pour une protéine kinase de type sérine/thréonine ; il a été impliqué dans la dynamique du cytosquelette, l'homéostasie calcique, et la physiologie musculaire ([Kaliman and Llagostera, 2008](#)). Des études ont aussi montré que ce gène est impliqué dans l'organisation du muscle squelettique ([Reddy et al., 1996](#)) et la différenciation myoblastique ([Furling et al., 2001](#)). Chez les patients DM1, on observe une baisse de l'expression du gène DMPK ainsi que de ses gènes voisins (SIX5 et DMWD). En effet, d'une part, les triplets CTG altèrent localement la structure de la chromatine conduisant à une inhibition de transcription. D'autre part, on observe une rétention des ARNs mutants dans le noyau empêchant leur traduction. Il a alors d'abord été pensé que la maladie s'expliquait par l'inhibition d'expression des gènes DMPK,

SIX5 et DMWD. Cependant, les souris knock-out (KO) développées pour ces gènes ne présentent que quelques symptômes légers de la maladie. Il apparaît donc que ce mécanisme ne joue pas un rôle majeur dans la pathologie.

Les conséquences de l'expansion pathologique de triplets CTG sont complexes et font intervenir différents mécanismes moléculaires. Il est néanmoins maintenant établi que la majorité des symptômes est due à un gain de fonction des ARNs porteurs de l'amplification. En effet, ceux-ci sont séquestrés dans le noyau sous forme de foyers d'accumulation (ou foci) et entraînent une perturbation du métabolisme d'un large groupe d'ARNs et en particulier de leur épissage.

## 2 Toxicité et gain de fonction de l'ARN pathologique

Les répétitions pathologiques CUG de l'ARN forment des structures en tige-boucles d'ARN double-brin (Michalowski et al., 1999). Les transcrits mutants sont correctement coiffés, épissés et polyadénylés mais cette conformation particulière empêche leur export vers le cytoplasme. Ils s'accumulent donc dans le noyau des cellules sous la forme d'agrégats d'ARNs intranucléaires appelés foci (Davis et al., 1997) (Figure 31). Ces inclusions nucléaires sont observables dans les cellules musculaires et neuronales des patients DM1.

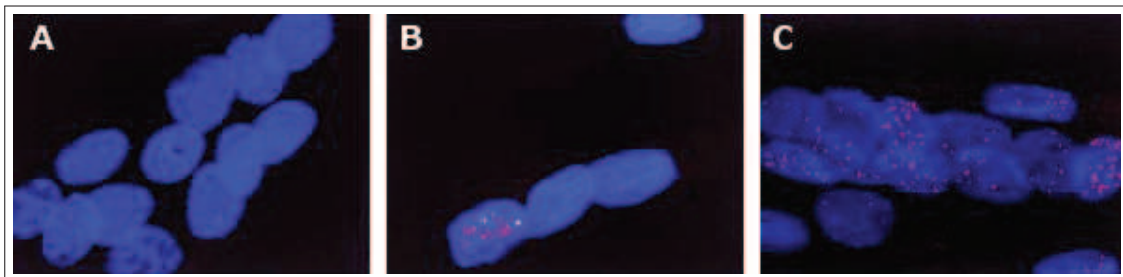


FIGURE 31: *Répétitions de triplets CUG dans des myoblastes humains.* Visualisation des répétitions CUG par RNA-FISH à l'aide d'une sonde (CAG)<sub>30</sub>-cys3 rouge. Les noyaux sont colorés au DAPI. (A) Myoblastes normaux. (B) Myoblastes d'un patient modérément atteint (150 répétitions). (C) Myoblastes d'un patient DM1 fortement atteint (Adapté de Davis et al., 1997).



Ce sont les ARNs porteurs de ces répétitions qui se révèlent toxiques et à l'origine de la DM1. En effet, l'expression de répétitions CTG non codants récapitule les symptômes majeurs de la maladie dans divers modèles animaux (Mahadevan et al., 2006; de Haro et al., 2006; Wang et al., 2008; Machuca-Tzili et al., 2011). Ces modèles n'utilisent pas forcément le gène DMPK comme vecteur des répétitions mettant ainsi en évidence l'indépendance du contexte génomique environnant. D'autre part, l'impact multisystémique des expansions CTG a aussi été validé dans les modèles animaux exprimant les répétitions dans plusieurs tissus (Seznec et al., 2001).

L'accumulation nucléaire des ARNm contenant les répétitions CUG affecte non seulement leur transport mais altère aussi le métabolisme d'autres ARNm. En effet, les répétitions CUG de l'ARN ainsi que les structures secondaires formées constituent des motifs de liaison de diverses protéines. Celles-ci se trouvent alors séquestrées au sein des foci les empêchant de réaliser leur fonction. Parmi celles-ci, les facteurs d'épissage MBNL1 et CUGBP1 voient leur expression altérée par les expansions entraînant de multiples défauts d'épissage.

La DM1 est donc une pathologie principalement causée par les inclusions ribonucléaires formées par les répétitions conférant à l'ARN muté un « gain de fonction » par lequel il peut séquestrer des protéines.

#### a Dérégulation de l'épissage

##### **Perte de fonction de MBNL1**

MBNL1 est une protéine de liaison à l'ARN très conservée entre espèces. Son expression est ubiquitaire mais préférentielle dans le muscle, le cœur et le cerveau (Fardaei et al., 2002). C'est un facteur d'épissage notamment impliqué au cours du développement et lors de la différenciation (Kalsotra et al., 2008; Venables et al., 2013; Han et al., 2013).

Le gène MBNL1 est soumis à l'épissage alternatif et peut générer plusieurs isoformes protéiques. En particulier, l'exon 3 favorise l'affinité de liaison de MBNL1 aux ARN

et l'exon 7 a été impliqué dans l'import nucléaire de la protéine (Tran et al., 2011). De nombreuses études ont été réalisées afin d'établir la « carte de régulation » décrivant l'activité activatrice ou inhibitrice de ce facteur. La capacité de MBNL1 à inclure ou exclure un exon alternatif dépend de la localisation de son site d'interaction par rapport à cet exon. En effet, la fixation de MBNL1 en amont de l'exon alternatif empêche le recrutement du spliceosome au niveau du site 3' d'épissage aboutissant à l'exclusion de l'exon alternatif. Au contraire, la présence d'un motif de fixation en aval de l'exon alternatif favorise son inclusion (Goers et al., 2010; Wang et al., 2012b).

MBNL1 est capable de se lier aux motifs YGCY (Y étant une pyrimidine). Ainsi il a été montré que cette protéine se lie aux expansions (CUG)<sub>n</sub> à la fois in vitro (Miller et al., 2000) et in vivo dans les cellules musculaires et neuronales de patients DM1 (Fardaei et al., 2002). MBNL1 est donc séquestrée au niveau des foci pathologiques entraînant sa déplétion du nucléoplasme et donc son impossibilité à réguler l'épissage de ses ARNs cibles. Cette perte de fonction de MBNL1 contribue au phénotype de la maladie. En effet de nombreux transcrits dérégulés chez les patients DM1 sont régulés par MBNL1 (Yuan et al., 2007). D'autre part, des souris KO pour MBNL1 portant l'exon 3 (qui code son site d'interaction à l'ARN) reproduisent la plupart des symptômes de la pathologie (Matynia et al., 2010). On estime que près de 80% des défauts d'épissage observés dans un modèle murin de DM1 sont retrouvés dans les souris KO de ce gène (Du et al., 2010).

MBNL1 appartient à la famille MBNL comportant trois homologues : MBNL1, MBNL2 et MBNL3. Si MBNL1 et MBNL2 ont une expression ubiquitaire et principalement dans le muscle, le cœur et le cerveau; MBNL3 est principalement exprimée dans le placenta. Au cours du développement, son expression diminue alors que celle de MBNL1 et MBNL2 augmente (Fardaei et al., 2002; Kanadia et al., 2003). Néanmoins, il est intéressant de noter que la colocalisation de MBNL2 comme de MBNL3 dans les foci de la DM1 a été démontrée; suggérant que tous les membres de la famille MBNL participent à la maladie (Fardaei et al., 2002).

### **Gain de fonction de CUGBP1**

La protéine CUGBP1 appartient à une famille de six paralogues : la famille CELF. Les fonctions de cette protéine sont multiples et dépendent de sa localisation subcellulaire : dans le noyau, elle agit comme un facteur d'épissage (Philips et al., 1998); dans le cytoplasme, elle est un facteur d'initiation de la traduction (Timchenko et al., 2004) ou de stabilisation de l'ARNm (Timchenko et al., 1999). Cette protéine a été beaucoup étudiée dans le cadre de la DM1 car il a été montré qu'elle lie les répétitions CUG (Timchenko et al., 2001a). Cependant à la différence de MBNL1, cette protéine n'est pas séquestrée au niveau des foci et ne subit donc pas une perte de fonction. Au contraire, la présence des ARNs contenant les répétitions CUG entraînent la stabilisation par phosphorylation de la protéine CUGBP1 dans les modèles cellulaires et animaux de la DM1 mais aussi dans les cellules et tissus de patients (Timchenko et al., 2001a; Wang et al., 2007). Cette hyperphosphorylation est médiée par la voie PKC (Protéine Kinase C) selon un mécanisme qui est encore inconnu. L'augmentation d'expression qui en résulte a des conséquences sur le profil d'épissage des transcrits cibles de CUGBP1. La surexpression de cette protéine dans différents modèles animaux induit des problèmes musculaires et cardiaques caractéristiques de la DM1 (Koshelev et al., 2010; Ward et al., 2010) suggérant un gain de fonction de CUGBP1 dans cette pathologie.

Des études suggèrent que CUGBP2 subirait de la même manière une surexpression dans la DM1 (Du et al., 2010; Hernández-Hernández et al., 2013).

### **Perturbation d'autres facteurs d'épissage**

Bien que MBNL1 et dans une moindre mesure CUGBP1 semblent être les facteurs d'épissage majeurs impliqués dans la DM1, il est possible que d'autres protéines de liaison à l'ARN soient engagées. En particulier, d'autres membres des familles MBNL et CELF présentent une expression altérée dans la pathologie et peuvent donc être également importants pour la pathogénèse.

D'autre part, le facteur d'épissage hnRNP H a aussi été associé aux répétitions des transcrits DMPK (Kim et al., 2005). Son expression semble augmentée chez les

patients DM1 et l'interaction entre MBNL1 et hnRNP H a été montrée in vivo (Paul et al., 2006). Par ailleurs, son rôle dans l'épissage alternatif de l'exon 11 du récepteur à l'insuline a été étudié suggérant une coordination physique et fonctionnelle de MBNL1 et hnRNP H dans l'épissage de cet exon (Paul et al., 2006).

### **La DM1 : une spliceopathie**

La dérégulation de l'expression des facteurs d'épissage dans la DM1 conduit à la production de profils d'épissage anormaux. Les deux facteurs majeurs altérés MBNL1 et CUGBP1 ont un rôle dans le développement normal, notamment musculaire. Durant celui-ci, l'expression de CUGBP1 diminue alors que celle de MBNL1 augmente (Kalsotra et al., 2008). C'est cette balance d'expression qui semble réguler la transition des formes d'épissage fœtales aux formes adultes pour un certain nombre de transcrits. De manière intéressante, le profil d'expression de ces deux protéines est inversé dans la DM1 et on observe un retour à un épissage de type fœtal. Une liste des événements d'épissage dérégulés dans la DM1 est présentée dans le **Tableau ??**.

Organe	Pré-ARNm	Dérégulation exon/intron	Inclusion Exclusion	Références
<b>Coeur</b>	Troponine T cardiaque (TNNT2/cTNT)	Exon 5	inclusion	Philips et al., 1998
<b>Muscle squelettique</b>	Récepteur à l'insuline (INSR)	Exon 11	exclusion	Savkur et al., 2001
	Canal chlore (CLCN1)	Intron 2	inclusion	Charlet-B et al., 2002 ; Mankodi et al., 2002
		Exon 7A	inclusion	
	Troponine T squelettique (TNNT3)	Exon foetal	inclusion	Kanadia et al., 2003
	Amphyphisine 2 (BIN1)	Exon 11	exclusion	Fugier et al., 2011
	Récepteur à la ryanodine (RyR)	Exon 70	exclusion	Kimura et al., 2005
	Sarcoplasmic/endoplasmic réticulum Ca2+ ATPase 1 (SERCA1)	Exon 22	exclusion	
	Sarcoplasmic/endoplasmic réticulum Ca2+ ATPase 2 (SERCA2)	Intron 19	inclusion	
	LIM domain inding 3 (LB3, ZASP)	Exon 11	inclusion	Lin et al., 2006
	Titin (TTN)	Zr4	inclusion	
		Zr5	inclusion	
	Calpain 3 (CAPN3)	Exon 16	exclusion	
	Nebulin-related anchoring protein (NRAP)	Exon 12	inclusion	
	Attractin-like (ATRNL1,ALP)	Exon 5a	inclusion	
		Exon 5b	inclusion	
	Forming homology 2 domain containing1 (FHOD1)	Exon 11a	exclusion	
	Glutamine-fructose-6-phosphate transaminase 1 (GFPT1)	Exon 10	exclusion	
	MBNL1	Exon 7	inclusion	
	MBNL2	Exon 7	inclusion	
	Facteur nucléaire 1	Exon 123	inclusion	Du et al., 2010
	SET and MYND domain containing 1 (SMYD1)	Exon 39	inclusion	
	Sperm associated antigen 9	Exon 39	inclusion	
	Myotubularin-related protein 1 (MTMR1)	Exon 2.1	exclusion	Buj-Bello et al., 2002 ; Ho et al., 2005
Exon 2.3		exclusion		
Alpha-dystrobrevin (DTNA)	Exon 11a	inclusion	Nakamori et al., 2008	
	Exon 12	inclusion		
<b>Cerveau</b>	Tau (MAPT)	Exon 2	exclusion	Sergeant et al., 2001
		Exon 3	exclusion	
		Exon 6c	exclusion	Leroy et al., 2006
		Exon 6d	inclusion	
		Exon 10	exclusion	Sergeant et al., 2001 ; Jiang et al., 2004
		Récepteur-N-methyl-D-aspartate (NMDAR1)	Exon 5	inclusion
	Amyloid precursor protein (APP)	Exon 7	exclusion	
	MBNL1	Exon 5	inclusion	Dhaenens et al., 2008
Exon 7		inclusion		

TABLE 5: *Tableau récapitulatif des événements d'épissage dérégulés dans la DM1.* (Adapté de *Rau, 2011*).

L'implication des défauts d'épissage dans le phénotype et en particulier les symptômes de la DM1 a été démontrée grâce à divers modèles murins.

La myotonie est causée par plusieurs défauts d'épissages dans l'ARN pré-messager du canal chlore (CLCN1), en particulier une inclusion anormale de l'intron 2 et de l'exon 7a ([Charlet-B et al., 2002](#)). Ces insertions conduisent à l'apparition de codons stop prématurés entraînant la dégradation des ARNs par la machinerie NMD (Nonsense Mediated Decay). Une absence de canal chlore CLCN1 conduit à une myotonie par hypersensibilisation de la membrane musculaire aux potentiels d'action ([Iaizzo et al., 1991](#)). L'épissage de l'intron 2 de CLCN1 est régulé par CUGBP1 ([Charlet-B et al., 2002](#)), alors que l'inclusion de l'exon 7a est régulée par MBNL1 ([Kino et al., 2009](#)). Enfin des travaux montrent que la correction, par des oligonucléotides antisens (OAS), des défauts d'épissage du canal chlore CLCN1 dans le muscle d'un modèle murin de DM1 permet de corriger la myotonie ([Mulders et al., 2009](#)). Cette étude montre donc que les défauts d'épissage du canal chlore sont la cause de la myotonie chez les patients DM1.

La faiblesse musculaire est associée à une exclusion anormale de l'exon 11 du transcrit du gène BIN1 (Bridging Integrator 1) codant pour une protéine impliquée dans la biogénèse des tubules T musculaires et essentiels au phénomène d'excitation/contraction. L'exclusion de l'exon 11 résulte en une forme inactive de la protéine et est à l'origine d'altérations des tubules T dans les muscles squelettiques des patients DM1 ([Fugier et al., 2011](#)). Cet épissage est régulé par MBNL1. Enfin, il est à noter que les épissages de RYR1, SERCA 1 et SERCA2 sont altérés chez les patients DM1 ([Kimura et al., 2005](#)) et ont été impliqués dans l'homéostasie calcique et le couplage excitation-contraction, suggérant qu'une altération de leur épissage participe à la faiblesse musculaire observée chez les patients DM1 ([Osborne et al., 2009](#)).

Les troubles de la conduction cardiaque peuvent être associés à l'inclusion anormale

de l'exon 5 du gène TNNT2 codant pour la troponine T cardiaque (Philips et al., 1998). Cet exon régule la sensibilité de la troponine T au calcium (Gomes et al., 2004); il est régulé à la fois par CUGBP1 et MBNL1 (Warf et al., 2009). Cependant, aucune démonstration directe n'existe à ce jour pour le rôle pathologique de l'épissage anormal de TNNT2.

Les patients DM1 développent entre autres un diabète de type 2 caractérisé par une résistance à l'insuline. Celle-ci est associée à une diminution de l'inclusion de l'exon 11 du gène INSR codant le récepteur à l'insuline, ce qui aboutit à la production d'un récepteur à l'insuline moins fonctionnel (Savkur et al., 2001). L'épissage de cet exon est régulé par MBNL1 et CUGBP1 (Sen et al., 2009; Sen et al., 2010).

Outre le muscle squelettique et le cœur, l'épissage alternatif de plusieurs transcrits au niveau du cerveau est également dérégulé chez les patients DM1 ce qui pourrait expliquer en partie les troubles du système nerveux central. Parmi ces gènes, on trouve Tau (MAPT), NMDAR1 ou encore APP (Sergeant et al., 2001; Jiang et al., 2004). Toutefois, on ne connaît pas exactement la fonction des exons régulés.

Un schéma récapitulant les principales altérations d'expression de facteurs d'épissage dans la DM1 et les défauts d'épissage qui en résultent associés aux phénotypes de la pathologie est présenté [Figure 32](#).

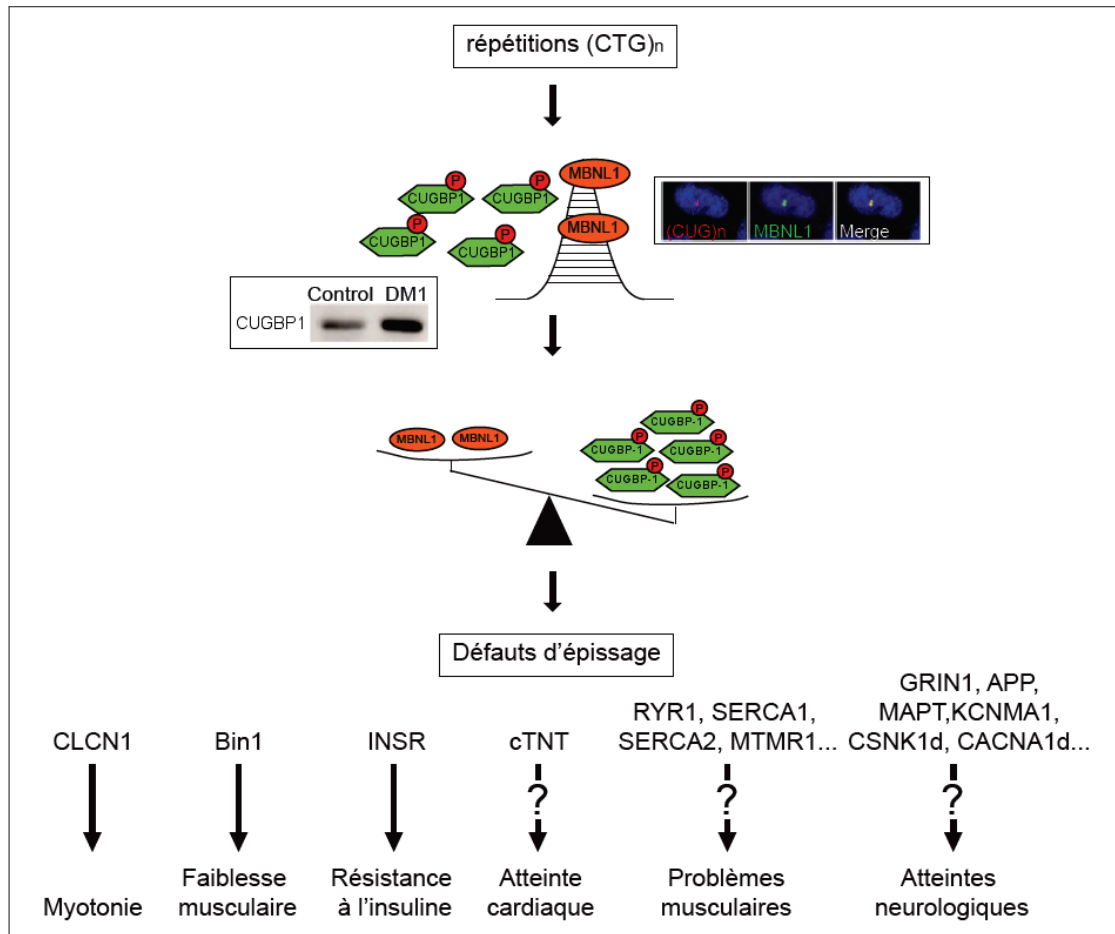


FIGURE 32: *Altérations de l'épissage alternatif dans la DM1 dû à un gain de fonction des ARNs porteurs de répétitions.* En situation normale, il existe un équilibre entre deux facteurs antagonistes de l'épissage : MBNL1 et CUGBP1. Les expansions (CUG)<sub>n</sub> observées dans la DM1 forme une structure double brin ayant au moins deux conséquences pathologiques sur ces deux familles : la séquestration de MBNL1 au sein des foci nucléaires, et une surexpression de CUGBP1. Cela résulte en une dérégulation de l'équilibre MBNL1/ CUGBP1 modifiant ainsi le profil d'épissage d'une série de gènes. Ces profils d'épissage anormaux ont été impliqués dans le développement de symptômes caractéristiques de la maladie (Adapté de Sicot, 2013).

## b Autres dérégulations

Au delà des altérations d'épissage alternatif responsables de certains symptômes importants dans la DM1, d'autres mécanismes sont aussi impliqués dans la



pathogénèse moléculaire.

### **Altération de l'expression génique**

Des analyses à large échelle ont permis de révéler l'impact des expansions CUG sur l'expression globale des gènes chez les patients DM1 (Botta et al., 2007). Une première explication à ce phénomène est la régulation de la stabilité des ARNs par MBNL1 et CUGBP1. En effet, ces deux protéines sont capables de se lier à la partie 3'UTR de transcrits et de réguler leur stabilité et leur dégradation (Masuda et al., 2012).

D'autre part, un certain nombre de facteurs de transcription ont été décrits comme séquestrés au niveau des foci ribonucléoprotéiques (Ebraldidze et al., 2004). Ce processus est appelé le « RNA leaching ». Parmi les facteurs de transcription identifiés, on peut citer Sp1, STAT1 et RAR.

Parmi les nombreux gènes dérégulés au niveau transcriptionnel, les gènes de l'homéostasie de la signalisation calcique, des fonctions mitochondriales ou encore de la matrice extracellulaire sont particulièrement représentés (Botta et al., 2007).

### **Altération du métabolisme des micro-ARNs**

La dérégulation d'un certain nombre de miARNs a été décrite dans le cœur et les muscles squelettiques (Perbellini et al., 2011 ; Rau et al., 2011). En particulier, l'expression du micro-ARN muscle spécifique miR-1 est diminuée dans le cœur de patients DM1 et est liée à la perte de fonction de MBNL1. En effet, MBNL1 est un régulateur de la biogénèse du pre-miR-1 via sa fixation sur un motif UGC localisé dans la boucle (Rau et al., 2011). La diminution de l'expression de miR-1 entraîne l'augmentation de gènes impliqués dans l'homéostasie cellulaire et pouvant donc contribuer aux altérations cardiaques observées dans la pathologie.

La dérégulation des miARNs dans la DM1 altère l'expression de leurs transcrits cibles, contribuant très certainement à la pathogénèse de la maladie.

### **Altération de la traduction**

Quelques études proposent que les répétitions CUG inhibent la traduction dans des cellules musculaires de patients DM1 (Huichalaf et al., 2010). La protéine CUGBP1 est multifonctionnelle et peut notamment activer la traduction de p21 et Mef2a (Timchenko et al., 2004). L'augmentation de sa stabilité chez les patients DM1 conduit donc à des anomalies de la traduction (Timchenko et al., 2005).

### **Transcription bidirectionnelle**

La transcription bidirectionnelle consiste en la synthèse d'un ARN à partir du brin d'ADN sens, ainsi que du brin complémentaire dit aussi antisens. Il en résulte deux types de transcrits : un transcrit sens et un transcrit antisens. Dans la DM1, l'existence d'un ARN antisens aux répétitions CUG aggraverait l'instabilité des répétitions (Nakamori et al., 2011). D'autre part, l'hybridation complémentaire des transcrits sens et antisens peut activer la voie des ARNi (Yu et al., 2011).

### **Expression de petits ARN CUG**

In vitro, les transcrits contenant de longues tige-boucles composées de répétitions CNG sont des cibles de Dicer (Krol et al., 2007). Le découpage de cet ARN double-brin par l'endonucléase induit alors la production de petits fragments CUG qui pourraient agir comme siRNA pour réguler négativement l'expression de transcrits contenant des séquences complémentaires. Ce mécanisme, s'il était validé in vivo, pourrait contribuer aux altérations d'expression des gènes observés dans la DM1 (Krol et al., 2007).

### **Traduction RAN**

Enfin, il apparaît que le transcrit DMPK pathologique peut être traduit en l'absence d'un site initiateur ATG à partir de plusieurs cadres de lecture. C'est la traduction RAN (Zu et al., 2011). Le transcrit antisens semble aussi soumis à ce mécanisme. La conséquence est l'accumulation de polypeptides potentiellement toxiques pour la

cellule. Si l'accumulation de polyglutamines a été confirmée dans un modèle murin de DM1, la toxicité de ce phénomène dans la pathologie reste à démontrer (Zu et al., 2011).

Un récapitulatif des différents mécanismes physiopathologiques proposés actuellement pour expliquer la pathogenèse de la DM1 est présenté **Figure 33**.

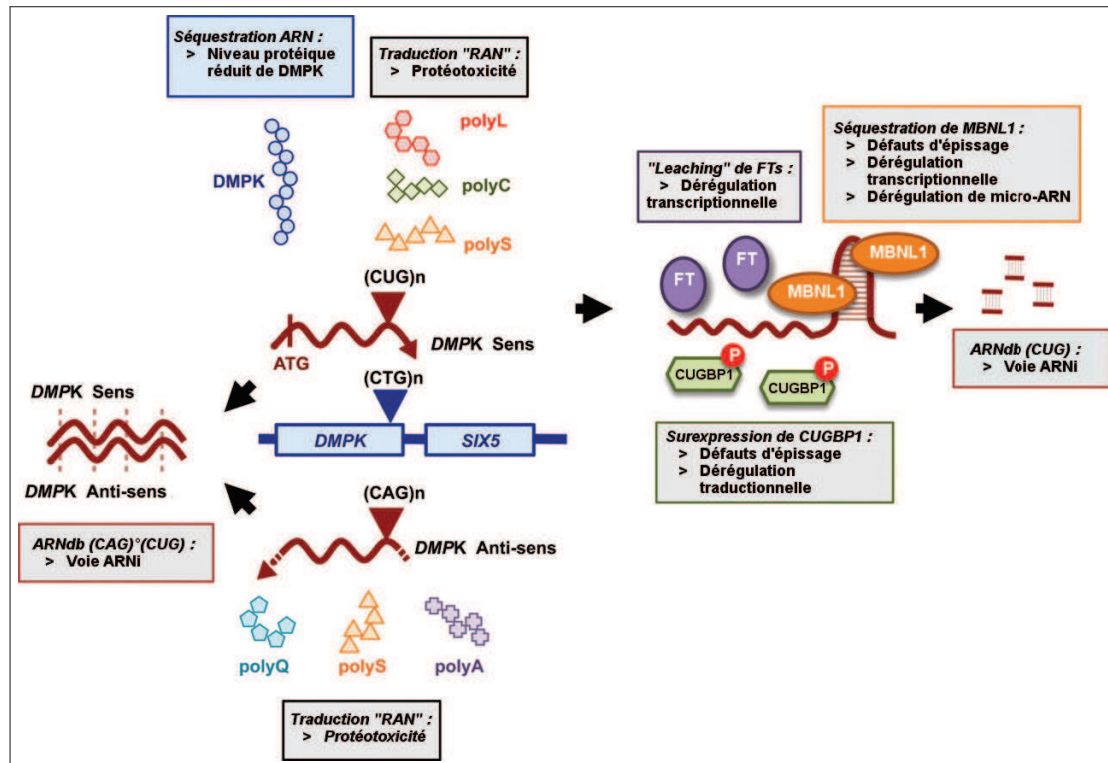


FIGURE 33: *Mécanismes moléculaires de la pathogénèse de la DM1.* Le gène *DMPK* contenant les expansions  $(CTG)_n$  est transcrit en ARN sens et antisens. Le transcrit *DMPK* muté forme une structure alternative et s'accumule dans le noyau de cellules DM1 formant des foci, provoquant la diminution de l'expression de la protéine *DMPK*. La perte de fonction de *MBNL1*, à cause de sa séquestration au sein des foci d'ARN, dérégule non seulement l'épissage alternatif mais aussi le métabolisme de la transcription et des microARNs. L'hyperphosphorylation et l'augmentation de l'expression de *CELF1* affectent l'épissage alternatif et la transcription. La retenue des facteurs de transcription par le transcrit *DMPK* muté peut aussi changer l'expression de gènes cibles. La voie des ARNs interférants peut aussi être activée par les structures ARN double brin, qui proviennent soit de la structure en épingle à cheveux du transcrit avec les expansions  $(CUG)_n$ , soit d'une hybridation complémentaire des transcrits *DMPK* sens et antisens. Les deux derniers peuvent aussi être traduits en l'absence de codon d'initiation *ATG* et générer ainsi des polypeptides, potentiellement toxiques pour la cellule (Adapté de *Sicot et al., 2011*).

## E Stratégies thérapeutiques

L'utilisation de modèles animaux a permis de mettre en évidence le caractère réversible du phénotype DM1. En effet, en éteignant l'expression des répétitions CUG au sein du muscle, le phénotype pathologique observé est supprimé (Mahadevan et al., 2006). Le tissu musculaire n'est donc pas irrémédiablement altéré et les symptômes associés sont réversibles. Ceci ouvre la porte au développement de stratégies thérapeutiques.

Une première approche est l'inhibition de l'altération des facteurs d'épissage MBNL1 ou CUGBP1. Par exemple, la surexpression par injection virale de MBNL1 dans les muscles squelettiques de modèles murins permet la restauration des défauts d'épissage et la correction de symptômes comme la myotonie (Kanadia et al., 2006). Une autre stratégie est de cibler directement l'ARN porteur des répétitions pathologiques. L'utilisation d'oligonucléotides antisens (OAS) modifiés permet d'hybrider de manière spécifique l'ARN cible et d'activer sa dégradation. L'injection subcutanée d'OAS dans un modèle murin de DM1 a permis une baisse importante du niveau du transcrit toxique et une correction des symptômes musculaires de la maladie (Wheeler et al., 2012). Ces effets bénéfiques sont apparus après 2 semaines et ont persisté jusqu'à un an après l'injection.

La limitation principale de ces stratégies thérapeutiques est de ne concerner pour l'instant que le muscle. La DM1 étant une maladie multisystémique, il est nécessaire d'améliorer les molécules existantes pour leur permettre d'accéder à tous les tissus affectés (en particulier le cœur et le cerveau). D'autre part, la durée d'efficacité doit aussi être augmentée afin d'éviter les inoculations répétées (Gao and Cooper, 2013a).

**2**

## **Objectifs**



Les hélicases ARN DDX5 et DDX17 ont été impliqués dans de multiples étapes de la régulation des gènes notamment dans le contrôle de la transcription, de l'épissage alternatif et la maturation des microARNs. Néanmoins ces différentes fonctions moléculaires ont jusqu'à maintenant été décrites dans des contextes cellulaires différents.

DDX5 et DDX17 sont des co-activateurs transcriptionnels connus de MyoD, facteur de transcription majeur de la myogénèse. Ces protéines ont été impliqué dans d'autres étapes de la régulation des gènes et tout particulièrement l'épissage. C'est pourquoi je me suis intéressée aux autres fonctions moléculaires potentielles de DDX5 et DDX17 dans ce même processus biologique : la différenciation myogénique. A l'aide de la lignée cellulaire murine myoblastique C2C12, je me suis tout particulièrement intéressée à la coordination des différentes étapes de régulation génique contrôlée par DDX5 et DDX17.

D'autre part, DDX5 a été impliquée dans un contexte musculaire pathologique : la Dystrophie Myotonique de type 1. Celle-ci est caractérisée par de multiples défauts d'épissage résultant de la séquestration du facteur d'épissage MBNL1 au niveau de la mutation responsable de la maladie. Il a été montré que DDX5, de par son activité hélicase, aide au recrutement pathologique de MBNL1. A l'aide de données à haut débit, j'ai donc cherché à établir une liste d'événements d'épissage dérégulés dans la pathologie et dépendants de DDX5 et DDX17.





**3**

## **Résultats**



## **I Publication en révision dans Molecular Cell**

Dans cet article, nous montrons que les hélicases ARN DDX5 et DDX17 coordonnent de manière directe et dynamique plusieurs niveaux de régulation des gènes lors de deux processus de différenciation différents : la Transition Epithélio-Mésenchymateuse (TEM) et la myogénèse. DDX5 et DDX17 sont des co-activateurs de l'activité transcriptionnelle de facteurs majeurs de ces deux modèles (respectivement des SMADs et de MyoD) et permettent l'expression de micro-ARNs spécifiques de la différenciation (respectivement miR-181b et miR-1/206). Ceux-ci en retour régulent l'expression de DDX5 et DDX17 révélant l'existence d'une boucle de rétro-contrôle négative. Enfin, cette diminution d'expression permet la mise en place d'un programme d'épissage participant à l'acquisition des phénotypes morphologiques des cellules différenciées. D'un point de vue mécanistique, il apparaît qu'un sous-groupe des événements d'épissage régulés durant la différenciation est contrôlé par la coopération de DDX5 et DDX17 avec le facteur d'épissage hnRNP H/F. Celle-ci pourrait alors assurer la reconnaissance des sites 5' d'épissage en amont de séquences riches en GC.

Dans ce projet, je me suis occupé d'une partie des analyses bioinformatiques ainsi que d'une très grande partie des manipulations expérimentales dans le modèle C2C12. Toute la partie MCF10A a été réalisée par un autre étudiant en thèse de l'équipe. Les deux projets étaient à l'origine indépendants mais les résultats obtenus étant très similaires, il a été décidé de combiner les deux projets afin de mettre en lumière le rôle orchestrateur de ces deux hélicases de manière indépendante du système de différenciation.

Les références citées dans cet article sont répertoriées à la fin de l'article tout comme les figures principales et supplémentaires.

# **RNA helicases DDX5 and DDX17 dynamically orchestrate transcription, microRNA and splicing programs in cell differentiation**

Etienne Dardenne<sup>1-4</sup>†, Micaela Polay Espinoza<sup>1-4</sup>†, Laurent Fattet<sup>1-4</sup>, Sophie Germann<sup>1-4</sup>, Marie-Pierre Lambert<sup>1-4</sup>, Helen Neil<sup>1-4</sup>, Eleonora Zonta<sup>1-4</sup>, Hussein Mortada<sup>1-4</sup>, Lise Gratadou<sup>1-4</sup>, Mathieu Deygas<sup>1-4</sup>, Fatima Zahra Chakrama<sup>1-4</sup>, François-Olivier Desmet<sup>1-4</sup>, Léon-Charles Tranchevent<sup>1-4</sup>, Martin Dutertre<sup>1-4</sup>, Ruth Rimokh<sup>1-4</sup>, Cyril F. Bourgeois<sup>1-4</sup>\* and Didier Auboeuf<sup>1-4</sup>\*

## **Affiliations:**

<sup>1</sup>INSERM U1052, Centre de Recherche en Cancérologie de Lyon, F-69008 Lyon, France ;

<sup>2</sup>CNRS UMR5286, Centre de Recherche en Cancérologie de Lyon, F-69008 Lyon, France ;

<sup>3</sup>Université Lyon 1, F-69000 Lyon, France ;

<sup>4</sup>Centre Léon Bérard, F-69008 Lyon, France

\*Correspondence to: [didier.auboeuf@inserm.fr](mailto:didier.auboeuf@inserm.fr) and [cyril.bourgeois@inserm.fr](mailto:cyril.bourgeois@inserm.fr)

Centre de Recherche en Cancérologie de Lyon, Centre Léon Bérard, 28 Rue Laënnec, 69008 Lyon, France. Phone: +33.4.26.55.67.46

† These authors contributed equally to this work

**RUNNING TITLE:** Orchestration of gene expression by RNA helicases

## SUMMARY

RNA helicases, including DDX5 and its paralog DDX17, are members of a large family of highly conserved proteins involved in gene expression regulation, although their *in vivo* targets and activities in biological processes like cell differentiation, that requires reprogramming of gene expression programs at multiple levels, are not well characterized. In this report, we first uncovered a new mechanism by which DDX5 and DDX17 cooperate with hnRNP H/F splicing factors to define an epithelial- and myoblast-specific splicing subprogram. We next observed that downregulation of DDX5 and DDX17 protein expression during epithelial to mesenchymal trans-differentiation and during myoblast to myotube differentiation contributes to switching splicing programs during these processes. Remarkably, this downregulation is mediated by the production of miRNAs induced upon cell differentiation in a DDX5/DDX17-dependent manner. Since DDX5 and DDX17 are required to initiate cell differentiation by acting as transcriptional coregulators of specific differentiation transcription factors, we propose to name these proteins “master orchestrators” of cell differentiation, that dynamically orchestrate several layers of gene expression.

## HIGHLIGHTS

- Regulation of alternative splicing by the cooperation between DDX5/DDX17 and hnRNP H/F at GC-rich sequences.
- DDX5/DDX17 and hnRNP H/F cooperate to express an epithelial- and myoblast-specific splicing program.
- DDX5/DDX17 are required for biogenesis of differentiation-specific microRNAs that control DDX5/DDX17 protein level through a negative feedback loop.
- DDX5 and DDX17 control the transcriptional activation of secondary master transcription factors of cell differentiation.

## Introduction

Transcriptome reprogramming during cell differentiation involves multiple layers of regulation of the gene expression process including transcription, alternative splicing that expands the gene message, and microRNAs(miRNAs) which control mRNA stability and translation (Ebert and Sharp, 2012; Kalsotra and Cooper, 2011). These different gene expression layers are most of the time analyzed independently, even though many factors participating to several of these layers have been identified. Yet, evidence that a single factor orchestrates several gene expression layers during cell differentiation is still missing.

The DEAD box RNA helicase DDX5 and its paralog DDX17 are highly conserved proteins having functions in most steps of the gene expression process, although their *in vivo* targets and activities are not well characterized (Fuller-Pace and Moore, 2011; Janknecht, 2010; Linder and Jankowsky, 2011). They are good candidates to coordinate the multiple layers of gene expression. Indeed, they are coregulators of several transcription factors, including MYOD, a master regulator of muscle differentiation (Caretta et al., 2006), and SMAD proteins (Warner et al., 2004), which mediate TGF $\beta$ -induced epithelial to mesenchymal transition (EMT), *i.e.* the transdifferentiation of epithelial cells to fibroblasts (Thiery et al., 2009). In addition, DDX5 and DDX17 control the biogenesis of miRNAs, in part via their interaction with the Drosha/DGCR8 complex (Davis et al., 2008; Fukuda et al., 2007). Finally, DDX5 and DDX17 are also components of the spliceosome and regulate alternative splicing (Dardenne et al., 2012; Germann et al., 2012; Lin et al., 2005; Samaan et al., 2013).

Two mutually non-exclusive models have been proposed to explain the role of DDX5 and DDX17 in alternative splicing. First, DDX5 and DDX17 may modulate splicing decisions owing to their ability to actively unwind the duplex between the U1 snRNA and the 5' splice site (5'ss), favoring the transition from the pre- to active spliceosome (Lin et al., 2005).

Second, DDX5 and DDX17 have been shown in two gene models (*H-ras* and *Tau*) to impact on RNA secondary structures that modulate the access either of the U1 snRNA to the 5' splice site or of the RBM4 splicing regulator to its binding site downstream of a weak 5' splice site (Camats et al., 2008; Kar et al., 2011). Whether DDX5 and DDX17 influence a specific class of secondary structure is not known. However, in both models the RNA helicase activity of DDX5 and DDX17 is required and the 5' splice site seems to be the preferential target of these factors. In addition, these models point to a role of DDX5/DDX17 in controlling the kinetics of splicing, by favoring either the transition between different states of the spliceosome or the recruitment of splicing factors downstream of the 5' splice site.

In this report, we demonstrate that DDX5 and DDX17 cooperate with hnRNP H/F splicing regulators at the level of 5' splice sites located in regions prone to form G-quadruplex structures. We show that this cooperation participates to establish epithelial- and myoblast-specific splicing programs. We next show that the downregulation of DDX5/DDX17 protein expression observed during epithelial-to-mesenchymal transition (EMT) and myogenesis contributes to the reprogramming of splicing during these processes. Importantly, DDX5 and DDX17 downregulation is mediated by specific miRNAs whose production requires both factors. Finally we show that DDX5 and DDX17 are required to initiate EMT and myogenesis as transcriptional coregulators of SMADs and MyoD, respectively, driving the production of key secondary differentiation regulators. We propose a model where DDX5 and DDX17 directly and dynamically orchestrate transcription, miRNA and splicing programs in two different models of cell differentiation.

## **Results**

### **DDX5 and DDX17 regulate alternative splicing of exons exhibiting specific features.**



In order to assess the role of DDX5 and DDX17 in splicing, a genome-wide study was performed using Affymetrix Human Exon Array on epithelial MCF-7 breast cancer cells transfected with siRNAs targeting both *DDX5* and *DDX17* (si*DDX5/17*) or a control siRNA (Figure 1A). Among the regulated single cassette exons, 63% (233) were more skipped and 37% (139) were more included following DDX5/DDX17 depletion (Figure 1B, Table S1). Hereafter, “class S” and “class I” exons will respectively refer to skipped and included exons upon DDX5/DDX17 depletion. As shown on Figures 1C and S1A, there was a good correlation between Exon Array prediction and extensive RT-PCR validation (n=121).

While analyzing DDX5/DDX17-regulated exons by RT-PCR, we observed that “class I” exons have a low basal inclusion rate in control cells (Figures 1C and 1D). This observation was confirmed by examining the computed inclusion rate of DDX5/DDX17-regulated exons using gene-normalized exon intensities from our microarray datasets (Figure S1B) or by computing their inclusion rate from the corresponding mRNAs present in public libraries (Figure S1C). Explaining the poor inclusion of “class I” exons, they have weak 5’ss and 3’ss when compared to both constitutive (CONS) and alternative (ALT) exons (Figures 1E and S1D). These exons are also flanked by large introns (Figure 1F).

“Class S” exons have very different features (Figure 1H). First, they are highly included in control cells (Figures 1C and 1D, Figures S1B and S1C). They are flanked by short introns and have strong 3’ss (Figures 1F and S1D). In addition, the 5’ss of “class S” exons is embedded in GC-rich sequences, in contrast to the 5’ss of “class I” exons (Figure 1G). However, similarly to “class I” exons, “class S” exons harbor weak 5’ss (Figure 1E).

### **DDX5/DDX17 and hnRNP H/F regulate a same subset of exons.**

Because “class S” exons are highly included in control cells despite their weak 5’ss, we hypothesized that 5’ss activators may enhance their inclusion in a DDX5/DDX17-

dependent manner. To test this hypothesis, we computed the density of previously reported intronic splicing enhancers (ISE) downstream of the 5' splice sites regulated by DDX5 and DDX17 (Lim and Burge, 2001; Zhang et al., 2005). As shown in Figure S1E, there is an enrichment of ISE downstream of "class S" exons compared to control sets of alternative and constitutive exons. We then used the SFmap software (Paz et al., 2010) to search for specific ISE motifs. As "class S" 5' splice sites were embedded in GC-rich sequences (Figure 1G), we focused our attention on splicing factors with a known affinity for GC-rich binding motifs, such as hnRNP H/F (Huelga et al., 2012; Katz et al., 2010; Xiao et al., 2009). Clearly, there is an enrichment in hnRNP H/F binding motifs downstream of "class S" exons (Figure S1E).

Supporting an important functional connection between DDX5/DDX17 and hnRNP H/F, an analysis of a previously published RNA-seq dataset (Xiao et al., 2009) showed that at least 60 of the 233 "class S" exons, *i.e.* skipped upon DDX5/DDX17 depletion, were also skipped following hnRNP H1 depletion (Table S2). In addition, analysis of existing hnRNP H/F CLIP-seq datasets (Huelga et al., 2012; Katz et al., 2010) revealed that those 60 exons, as well as the complete set of 233 "class S" exons, are flanked by a larger number of hnRNP H/F CLIP-seq reads than a control set of randomly selected alternative exons (Figure 2A and Table S2). Remarkably, joint hnRNP H/F and DDX5/DDX17 depletion (Figure S1F) cooperatively enhanced the skipping of exons containing hnRNP H/F binding sites, but not of control exons without hnRNP H/F binding sites (Figures 2B and 2C).

Intriguingly, while analyzing the environment of DDX5/DDX17-regulated exons, we observed that the 5' splice sites of exons (n-1), *i.e.* located just upstream of the included "class I" exons (n), are embedded in GC-rich sequences just as the 5' splice sites of "class S" exons (n) (Figure 2D). Furthermore, there is an enrichment in intronic hnRNP H/F binding sites downstream of the exons (n-1) of "class I" exons (Figure 2E). Based on the similar 5' splice site configuration of "class I" exons (n-1) and "class S" exons (n), and on recent kinetic models of splicing regulation, we

hypothesized that in the case of “class I” exons, the depletion of DDX5/DDX17 and/or hnRNP H/F could slow down splicing of exon (n-1) to exon (n+1), and consequently increase the time window for the recognition and inclusion of the weak exon (n). To test this model, we selected a set of “class I” exons containing or not hnRNP H/F binding sites downstream of exon (n-1). Remarkably, hnRNP H/F depletion, by itself or together with DDX5/DDX17 depletion, enhanced exon (n) inclusion only when hnRNP H/F binding sites were present downstream of exon (n-1) (Figure 2E).

We next reasoned that if DDX5/DDX17 depletion would slow down exon (n-1) to exon (n+1) splicing, this would give enough time for splicing activators to enhance weak exon (n) inclusion. As the 5' splice sites of “class I” exons are followed by GC-poor sequences, we looked for TIA1 and TIAL1 binding sites downstream of “class I” exons because these factors bind to AT-rich motifs (Del Gatto-Konczak et al., 2000; Forch et al., 2000; Wang et al., 2010b) (Figure 1F). Analysis of available CLIP-seq datasets revealed an enrichment of TIA1 binding sites downstream of “class I” exons (Figure 2F and Table S2). Furthermore, TIA1/TIAL1 depletion (Figure S1G) reduced DDX5/DDX17-mediated exon inclusion only when TIA1/TIAL1 binding sites were found downstream of the regulated exons, but not in a control exon (Figure 2G). It must be underlined that DDX5/DDX17 depletion slightly decreased TIA1 and TIAL1 protein expression level (Figure S1G). However, the effect of DDX5/DDX17 depletion on splicing could not be a consequence of TIA1/TIAL1 expression decrease because in that case, the co-depletion of TIA1/TIAL1 and DDX5/DDX17 would have had similar effects and not antagonistic effects as observed in Figure 2G.

### **Cooperation between DDX5/DDX17 and hnRNP H/F on G quadruplex structures.**

It has been previously shown that hnRNP H/F binds to G-tracts that can form G quadruplexes (Decorsiere et al., 2011). Interestingly, the high kinetic stability of a G-

quadruplex limits hnRNP H/F binding to G-rich sequences (Samatanga et al., 2013). Computational analysis revealed an enrichment in predicted G-quadruplexes downstream of “class S” exons(n) when compared to “class I” exons(n) or to a set of control exons (Figure 3A). Remarkably, an enrichment in G-quadruplexes was also predicted downstream of exons(n-1) of “class I” exons, when compared to “class S” exons(n-1) or to a set of control exons (Figure 3A). Therefore, we hypothesized that the RNA helicase activity of DDX5 and DDX17 may favor the binding of hnRNP H/F to G-tracts embedded in G-quadruplex RNA structures.

Several lines of evidence support this hypothesis. First, *in cellulo* stabilization of G-quadruplex structures using the TMPyP4 reagent had the same effect on splicing than the silencing of *DDX5/DDX17* or *HNRNPH/F* genes (Figure 3B). Second, hnRNP H1 co-immunoprecipitated with both DDX5 and DDX17 (Figure 3C), in agreement with previous reports suggestive of an interaction between these factors (Camats et al., 2008; Laurent et al., 2012). A direct interaction was further supported by an *in situ* proximity-ligation assay, a sensitive approach that reveals complexes between single molecules that are close enough (<40 nm) in unmodified cells to generate a unique spot for each complex, after PCR amplification of the complementary fluorescent antibody-linked oligonucleotide probes. PLA using DDX5 and hnRNP H1 antibodies detected multiple spots in the nuclei of MCF7 cells (Figure 3D). Third, an antibody against hnRNP H1 (but not a control antibody) immunoprecipitated their target RNAs and *DDX5/DDX17* depletion reduced this interaction (Figure 3E).

Finally, the helicase activity of DDX5 and DDX17 was required for the regulation of both exon inclusion and skipping. Stable inducible MCF-7 cells were transfected with siRNAs targeting the UTR (5' or 3') of endogenous *DDX5* and *DDX17* mRNAs (*siDDX5/17-UTR*), which had the same effect than *siDDX5/17* (compare for example Figures 1C and 3F). Re-

expression of wild-type DDX5(Samaan et al., 2013)rescued the splicing pattern of tested exons to a level near or identical to the control level (Figure3F, lanes 1-3). In contrast, re-expression of a DDX5 mutant (DDX5-KA) lacking the helicase activity could not restore, or only partially, the splicing of all tested exons (Figure 3F, lanes 4-6).Similar results were obtained with DDX17 (Figure S1H), underlining the functional redundancy between the 2 proteins in splicing regulation.

### **Cooperation between DDX5/DDX17 and hnRNP H/F contributes to establish epithelial-myoblast-specific splicing programs.**

Aiming to confirm the cooperation between DDX5/DDX17 and hnRNP H/F (Figures 2 and 3) in a different context, we used siRNAs targeting both *DDX5* and *DDX17* to silence their expression in MCF10A cells, another epithelial breast cancer cell line (Figure S2A). First, the *in situ* proximity ligation assay confirmed the interaction between DDX5 and hnRNP H1 in these cells (Figure S2B). Second, a genome-wide analysis using Affymetrix Exon Arrays, which identified hundreds of DDX5/DDX17-dependent exons in MCF10A cells (Figure 4A, Table S3), showed that these exons presented similar characteristics compared to DDX5/DDX17-regulated exons in MCF7 cells. Indeed, DDX5/DDX17-dependent exons in MCF10A have weak 5' splice sites, and the 5' splice sites of “class S” exons (n) and “class I” exons (n-1) tend to be embedded in a GC-rich environment containing more hnRNP H/F binding sites (Figures S2C, S2D and S2E). Most importantly, RT-PCR analyses revealed that a subset of DDX5/DDX17-controlled exons was regulated by a cooperation between these RNA helicases and hnRNP H/F in MCF10A cells, as in MCF-7 cells (Figure 4B).

MCF10A epithelial cells are particularly interesting because they can undergo epithelial to mesenchymal transition (EMT) upon treatment with transforming growth factor  $\beta$  (TGF $\beta$ ),

a process that could play a major role during tumor progression and metastasis formation. DDX5 and DDX17 have been described to be involved in the regulation of EMT, and DDX5 interacts with SMAD proteins (Figure S3A) (Warner et al., 2004), which mediate the cell response to TGF $\beta$ . Interestingly, we noticed that TGF $\beta$  treatment of MCF10A cells induced a decrease of DDX5 and DDX17 protein expression level (Figure 4C). This prompted us to investigate the function of DDX5 and DDX17 during this process, especially at the level of splicing, since it has been recently shown to be highly regulated during EMT.

The decreased expression of DDX5/DDX17 led us to hypothesize that some alternative exons might be similarly affected by TGF $\beta$  treatment and siRNA-mediated *DDX5/DDX17* silencing. To test this hypothesis, we analyzed the effect of si*DDX5/DDX17*, in the absence of TGF $\beta$  treatment, on the inclusion of a set of exons previously reported to be regulated during EMT (Shapiro et al., 2011; Warzecha et al., 2010). Remarkably, a number of TGF $\beta$ -regulated exons were regulated similarly by DDX5/DDX17 depletion in untreated MCF10A cells (Figure 4D). Because some of these exons were also regulated in the same manner by hnRNP H/F depletion in the absence of TGF $\beta$  (Figure 4E), we concluded that the DDX5/DDX17 and hnRNP H/F cooperation may contribute to maintain an epithelial-specific splicing pattern.

Based on previous reports (Caretta et al., 2006), we also investigated the role of DDX5/DDX17 in myogenesis using the well-studied mouse C2C12 myoblast cells. Strikingly, just as during TGF $\beta$ -mediated EMT, we observed a decrease in the expression of both Ddx5 and Ddx17 proteins upon induction of myoblast differentiation (Figure 4F). In a remarkable parallel to what we observed in MCF10A cells, a subset of exons regulated during differentiation of C2C12 cells (Bland et al., 2010) was regulated in the same manner upon *Ddx5/Ddx17* silencing in undifferentiated myoblasts (Figures 4G and S3B). Interestingly, more than half of these exons were also co-regulated by hnRNP H/F in C2C12 cells (Figure

4G, right panel). Importantly and like in MCF7 and MCF10A cells, Ddx5 and hnRNP H1 interacted directly in C2C12 cells (Figures S3C and S3D).

Altogether, these results support a model where the cooperation between DDX5/DDX17 and hnRNP H/F contributes to define an epithelial- and myoblast-specific splicing subprogram. DDX5 and DDX17 downregulation may therefore participate to the splicing changes that occur during EMT and myogenesis.

### **The downregulation of DDX5/DDX17 is part of a feedback regulatory loop involving microRNAs**

Because the DDX5/DDX17 downregulation during EMT and myogenesis may appear contradictory with previous reports indicating their implication in these processes (Caretti et al., 2006; Yang et al., 2006), we decided to investigate the molecular mechanism of this downregulation to confirm its biological relevance. Interestingly, phylogenetically conserved binding sites for two miRNAs known to be upregulated during myogenesis, miR-1 and miR-206 (Williams et al., 2009), were predicted within the 3' UTR of both *DDX5* and *DDX17* mRNAs (Figures S4A and S4B). Strikingly, overexpression of pre-miR-1 or pre-miR-206 in C2C12 cells reduced Ddx5 and Ddx17 protein levels (Figure 5A). Luciferase assays confirmed that this effect was mediated by a direct binding of miR-1/206 onto the predicted sites in mouse *Ddx5* and *Ddx17* 3' UTRs (Figure 5B).

Induction of miR-1 and miR-206 during myogenesis is mediated at least in part by MyoD, a master transcriptional regulator of myogenesis which is co-regulated by Ddx5 and Ddx17 (Caretti et al., 2006), with which it interacts in C2C12 cells (Figure S5A). Remarkably, the induction of mature miR-1 and miR-206 during differentiation of C2C12

cells was prevented by Ddx5/Ddx17 depletion (Figure 5C in yellow). A similar inhibition was observed when looking at the miRNA precursors pri-miR-1a1 and pri-miR-206 (Figure 5C in orange), indicating that the primary effect was a transcriptional inhibition of these miRNA genes. However, the partial repression of pri-miR-206 induction upon Ddx5/Ddx17 silencing could also reflect an effect on the processing of miR-206 precursors, in agreement with previous reports (Davis et al., 2008; Fukuda et al., 2007). Yet a direct transcriptional effect was strongly supported by the recruitment of Ddx5 to the promoters of both miR-1 and miR-206 coding genes (Figure 5D).

Based on these observations, we postulated that a similar feedback-loop may exist in the MCF10A cellular model. This hypothesis was supported by several lines of evidence. First, the *DDX5* 3' UTR contains two predicted binding sites for miR-181b (Figure S4A), which has been reported to be induced by TGF $\beta$  expression in MCF10A cells (Wang et al., 2010a). In agreement with this prediction, overexpression of pre-miR-181b in MCF10A cells reduced *DDX5* protein level (Figure 5E), and luciferase assays confirmed that this effect was mediated by a direct binding of miR-181b to the *DDX5* 3' UTR, at least on the most conserved of the predicted sites (Figures 5F and S4A). Intriguingly, miR-181b also repressed endogenous *DDX17* expression (Figure 5E), even though no miR-181b binding site was predicted within *DDX17* 3' UTR (Figure S4B). This suggested either the existence of a non-canonical and/or non-predicted binding site, or an indirect regulation of *DDX17* expression by miR-181b. Second, the TGF $\beta$ -induced expression of miR-181b in MCF10A cells was dependent on SMAD transcription factors (Figures S5B and S5C), which interact with their coregulator *DDX5* as described above (Figure S3A). As expected, TGF $\beta$ -induced miR-181b and pri-miR-181b expression was prevented by *DDX5/DDX17* depletion (Figure 5G) and *DDX5* was recruited



on the promoter of the miR-181b-1 gene promoter (Figure 5H), demonstrating a direct role of DDX5 in miR-181b induction.

Our results revealed that DDX5/DDX17 expression is regulated by a negative feedback loop during both myoblast differentiation and TGF $\beta$ -induced EMT: DDX5/DDX17 directly promote the expression of differentiation-specific miRNAs that in turn directly downregulate their expression. Supporting a model in which miRNA-mediated DDX5/DDX17 downregulation drives a specific splicing program, ectopic overexpression of miR-1 or miR-206 in C2C12 cells (Figure 5I), or miR-181b in MCF10A cells (Figure 5J), induced a similar splicing pattern than cell differentiation and DDX5/DDX17 depletion.

### **DDX5 and DDX17 directly control the transcriptional activation of secondary master transcription factors of cell differentiation**

Our data suggest that DDX5 and DDX17 downregulation by specific miRNAs is an important step during EMT and skeletal myogenesis (Figures 4 and 5). However the presence of DDX5 and DDX17 is required for early steps of cell differentiation since they regulate the production of miRNAs, like miR-1/206 in myoblasts, that are determinant for the fate of myogenic precursors (Figure 5). In line with previous observations (Caretta et al., 2006), *Ddx5/Ddx17* knockdown in C2C12 cells inhibited the MyoD-dependent expression of *Myog* and *Mef2c*, two master myogenic regulators, during differentiation (Figure 6A). This effect was most likely direct since Ddx5 was recruited on the promoters of *Myog* and *Mef2c* genes (Figure 6A, right). As expected, *Ddx5/Ddx17* depletion compromised the formation of myotubes (Figure 6B), confirming their key function in myogenesis.

Likewise, *DDX5/DDX17* knockdown prior to TGF $\beta$  treatment inhibited, at both RNA and protein levels, the TGF $\beta$ -induced and SMAD2/3-dependent expression of *SNAI1* and *SNAI2*, two master regulators of EMT (Thiery et al., 2009)(Figures 6C and S6A). This effect was direct as DDX5 was recruited on *SNAI1* and *SNAI2* promoters (Figure 6C, right), along with SMAD4 (Figure S6B). As expected, SNAIL1/2-target genes were not properly regulated by TGF $\beta$  in DDX5/DDX17-depleted cells, this included genes coding for proteins involved in epithelial cell junctions like Occludin and E-Cadherin (Figure S6C and S6D). As a consequence, the SNAIL-dependent loss of epithelial cell junctions, which is normally induced by TGF $\beta$ (Cano et al., 2000), and the expression of mesenchymal markers such as *FAP*, were compromised in DDX5/DDX17-depleted cells (Figures 6D, S6C and S6E). Thus, DDX5 and DDX17 are key transcriptional coregulators of SMADs in mediating the induction of EMT by TGF $\beta$ .

Collectively, these results support a model where DDX5 and DDX17 cooperate with hnRNP H/F to express a specific splicing subprogram in epithelial cells and in myoblasts, and are required to initiate differentiation as transcriptional coregulators of MyoD and SMAD transcription factors (Figure 6E). As a consequence, they elicit the biogenesis of differentiation-specific miRNAs that trigger a feedback control loop which downregulates their expression.

## Discussion

Understanding alternative splicing regulation, in particular during differentiation, is a major challenge as it is now well established that different cell types express different splicing isoforms, contributing to cell-specific functions (Merkin et al., 2012; Wang et al., 2008). While the primary sequence elements of pre-mRNAs contributing to alternative splicing are more and more characterized, the contribution of secondary RNA structures is far less understood (Jin et al., 2011). In this report we demonstrate that the DDX5 and DDX17 RNA helicases favor the recruitment of hnRNP H/F on target RNAs (Figure 3). HnRNP H/F proteins are known to recognize G-tracts (Decorsiere et al., 2011) but this binding can be limited if specially organized G-tracts assemble into G-quadruplexes, that are highly stable secondary structures (Samatanga et al., 2013). Therefore, our data support a model where the DDX5 and DDX17 RNA helicase activity may favor the binding of hnRNP H/F and assist them in their splicing enhancer function. In agreement with our results, this enhancer activity of hnRNP H/F proteins has been linked to G-runs positioned downstream of 5' splice sites (Wang et al., 2012b; Wang and Cambi, 2009; Xiao et al., 2009).

Remarkably, if DDX5/DDX17 help the recruitment of hnRNP H/F downstream of an alternative exon defined by a weak 5' splice site (class S), they enhance exon inclusion (Figure 1H). In this case, depletion of DDX5/DDX17 and/or hnRNP H/F induces exon skipping (Figures 2A-2C). However, if DDX5/DDX17 and hnRNP H/F cooperate at the level of an upstream exon (n-1), this results in the skipping of the downstream weak exon (class I) flanked by GC-poor sequences and large introns (Figure 1H). Therefore, in this case, depletion of DDX5/DDX17 and/or hnRNP H/F induces inclusion of the weak downstream exon (Figures 2D and 2E). This model is in agreement with recent kinetic models of regulation of alternative splicing, relying on the competition between splicing sites (Schor et al., 2013). Indeed, if splicing between two

exons is efficient, weak internal exons are skipped, and slowing down the splicing process increases the time window for the recognition of such exons. Supporting this model, depletion of TIA1/TIAL1 splicing activators, that act downstream of weak exons, inhibited the exon inclusion pattern induced by DDX5/DDX17 depletion (Figures 2F and 2G). Such kinetic models allow to better understand the well-known position-dependent effect of splicing factors.

Our data indicate that the cooperation between DDX5/DDX17 and hnRNP H/F participates to maintain an epithelial- and myoblast-specific splicing subprogram as some of the exons regulated during epithelial transdifferentiation and during myoblast differentiation are under the control of these factors (Figure 4). In this context, it must be underlined that exons regulated during EMT are flanked by hnRNP H/F binding motifs (Shapiro et al., 2011) and that a function of hnRNP H/F in muscle cells has previously been reported (Chen et al., 1999; Paul et al., 2011). As DDX5/DDX17 contribute to maintain an epithelial- and myoblast-specific splicing subprogram, their downregulation during EMT and myogenesis may favor the switch toward the previously reported fibroblast- and myotube-specific splicing programs (Bland et al., 2010; Shapiro et al., 2011). It is interesting to note that hnRNP H/F proteins are structurally and functionally related to ESRP proteins that recognize similar G-rich sequences and play an important role in establishing an epithelial-specific splicing program (Shapiro et al., 2011; Warzecha et al., 2010). It must also be underlined that DDX5 and DDX17 have been shown to work with RBM4 and MBNL, two splicing factors involved in the control of myogenesis (Kar et al., 2011; Laurent et al., 2012; Lin and Tarn, 2011). Therefore, DDX5/DDX17 could have a wide impact on alternative splicing during EMT and myogenesis by modulating the activity of several splicing factors involved in these processes.

In this context, one of the main phenotypes observed in both epithelial and myoblast cells under conditions of DDX5/DDX17 depletion was the formation of lamellipodia (Figure S7A), that characterize mesenchymal cells and that also participate to the alignment and fusion of myoblasts (Rochlin et al., 2010; Yilmaz and Christofori, 2009). Interestingly, we noticed that several genes that were spliced similarly upon DDX5/DDX17 downregulation and during EMT or myoblast differentiation are involved in actin cytoskeleton dynamics (Figure S7B), a process which is intimately linked to lamellipodia formation. This suggests that splicing changes elicited by DDX5/DDX17 downregulation may participate to this phenotypic change in differentiating cells, although it is likely that the decreased expression of these multi-task factors has a wider array of consequences on the various molecular processes they control. For example, it has been shown that DDX5 controls the expression of miR-182 in basal breast cancer cells, which directly affects the reorganization of the actin cytoskeleton (Wang et al., 2012a).

Finally our data demonstrate that DDX5 and DDX17 are not only required for maintaining epithelial- and myoblast-specific splicing program, but they are also required for initiating EMT and myogenesis owing to their role in transcription. Indeed, in agreement with previous reports, we demonstrate that DDX5/DDX17 are transcriptional coregulators of the MyoD and SMAD transcription factors that drive transcription programs during myogenesis and TGF $\beta$ -induced EMT. One consequence of this cooperation is the biogenesis of differentiation-specific miRNAs that in turn decrease the expression of DDX5/DDX17 (Figure 5), contributing to switch the splicing of a network of genes involved in cytoskeleton remodeling (Figure 6E).

Altogether, our results unraveled a direct and dynamic orchestration of different gene expression layers by the RNA helicases DDX5 and DDX17 during cell differentiation. The

DDX5/DDX17-dependent genetic circuitries that we characterized in two different biological processes and in two different species are remarkably similar (Figure 6E) and we propose to define such factors as master orchestrators of differentiation. Other RNA helicases such as DDX3X, DDX4/VASA and DHX9/RHA, that are also involved at multiple levels of gene expression and contribute to a variety of biological processes (Abdelhaleem, 2005; Lasko, 2013; Linder and Jankowsky, 2011), may also have a broad function in coordinating the genetic programs that unfold during biological transitions or cell differentiation processes.

## **Experimental procedures**

**Cell culture, treatment and transfection:** Cell culture of standard and stable MCF-7 cells, as well as transient transfection assays in all cell lines were performed essentially as described previously (Dardenne et al., 2012; Germann et al., 2012; Samaan et al., 2013; Zonta et al., 2013). siRNAs were transfected at a final concentration of 26.6 nM, while pre-miRNAs (Pre-miR miRNA Precursors, reference #AM17100 or negative control #1, Reference #AM17110, Life Technologies) were used at a final concentration of 10 nM. Sequences of siRNA sequences are provided in Table S4. For the analysis of the impact of potential G-quadruplexes on splicing, MCF7 cells were treated for 48 h with 150  $\mu$ M TMPyP4, a molecule that stabilizes G-quadruplexes.

C2C12 cells were cultured either in growth medium DMEM supplemented with 10% fetal bovine serum, 1% antibiotics) or in differentiation medium (DMEM supplemented with 2% horse serum, 1% antibiotics). MCF10A cells were cultured in 1:1 (DMEM)/HAM-F12 medium (Life Technologies) supplemented with 5% horse serum (Cambrex), 10 ng/ml human epidermal growth factor (PromoCell), 0.5  $\mu$ g/ml hydrocortisone (Sigma), 10  $\mu$ g/ml insulin (Sigma) and 100 ng/ml cholera toxin (Sigma). Recombinant TGF $\beta$ -1 (Peprotech) was used at

2,5 ng/ml. To block any signaling arising from autocrine production of TGF $\beta$ , experiments without TGF $\beta$  were performed in the presence of 10  $\mu$ M T $\beta$ RI kinase inhibitor SB-431542 (Sigma) for the indicated times.

Luciferase assays, using DDX5 or DDX17 reporters cloned in the psiCHEK-2 plasmid (Promega) were carried out as detailed in Supplementary Materials, and analyzed using the Dual-Luciferase<sup>®</sup> Reporter Assay System (Promega), as recommended by the manufacturer.

**RNA analysis:** RNA extraction, RT-PCR and RT-qPCR were described elsewhere (Dardenne et al., 2012; Germann et al., 2012; Samaan et al., 2013; Zonta et al., 2013). Primer sequences for PCR and qPCR are provided in Table S4. Quantification of mature miRNAs was carried out using the miRCURY LNA<sup>™</sup> Universal cDNA synthesis and SYBR<sup>®</sup> Green PCR kit (Exiqon), using specific LNA<sup>™</sup> primers sets, following manufacturer's instructions. RT-qPCR values were normalized relative to U6 snRNA (also measured using LNA<sup>™</sup> primers set, Exiqon).

### **Chromatin and RNA Immunoprecipitation**

ChIPs were performed essentially as previously described (Fattet et al., 2013; Zonta et al., 2013) using antibodies against DDX5 (4  $\mu$ g, A300-523A, Bethyl), SMAD4 (sc-7154, Santa Cruz) or control IgG (Santa Cruz). RNA-IP was performed as previously described (Bittencourt et al., 2008) using hnRNP H1 antibodies (A300-511A, Bethyl) and or rabbit IgG (sc-2027, Santa Cruz).

### **Microscopy**

Detailed protocols for immunofluorescence, *in situ* proximity ligation assays (PLA) and time-lapse videomicroscopy are given in Supplementary methods.

## Bioinformatics

Microarray analyses were performed as described previously (Mallinjou et al., 2013; Samaan et al., 2013). Exon numbering of mouse and human genes was based on the FasterDB server (<http://fasterdb.lyon.unicancer.fr/index.php>) (Mallinjou et al., 2013).

MaxEntScan was used to measure the scores of the 5' and 3' splice sites. SFmap web server (<http://sfmap.technion.ac.il/>) (Paz et al., 2010) was used to predict and map known splicing factor binding sites. UCSC LiftOver tool (<http://genome.ucsc.edu/cgi-bin/hgLiftOver>) was used to look at conservation between human and mouse exons.

Publicly available CLIP-seq datasets were retrieved from the NCBI for hnRNPH1 and hnRNPF (GSE23694 and GSE34993 respectively) and from ArrayExpress for TIA and TIAL1 (E-MTAB-432). The number of reads, overlapping the regions of interest, was computed using coverageBed (Bedtools).

For the prediction of G-quadruplex structures, the first 50 or 60 intronic nucleotides downstream of the exons of interest were searched for the previously described alternative G-quadruplex motif  $G_{n>2}-(N_{1-7}-G_{n>2})_3$  (Xiao et al., 2013). Too short introns (<100 nucleotides) were kept as is as to avoid considering exonic sequences. These sequences were then extended in each direction to avoid the artificial bell curve effect. Each sequence position was associated to the number of G-quadruplexes that overlap this nucleotide. The data points were then binned to smoothen the curve (bin size set to 5).

## ACKNOWLEDGMENTS

We are grateful to D. Furling and J. Marie for their comments on the manuscript, and to all our colleagues for inspiring discussions. We thank C. Vanbelle and C. Bouchardon (CeCIL -



SFR Santé Lyon-Est) and Dr Joël Lachuer (ProfileXpert-LCMT facility, Lyon, France) for their assistance with imaging and microarray analyses, respectively. This work was supported by grants from Inserm “Plan Cancer 2009-2013”, INCa, Fondation pour la Recherche Médicale and LNCC, from doctoral fellowships from LNCC (E.D. and S.S.), AFM (M.P.E.), the french Ministère de l'Enseignement Supérieur et de la Recherche (L.F.) and FRM (E.Z.), and from post-doctoral fellowships from the Centre Léon Bérard (H.N.) and Fondation ARC (S.G., M.P.L., F.Z.C. and L.C.T.).

## Figure legends

### Figure 1: Genome wide impact of DDX5/DDX17 on splicing defines two classes of regulated exons

(A) Western blot of DDX5, DDX17 and actin (loading control) 48 hours after MCF-7 cell transfection with a control siRNA (-) or with an siRNA targeting both *DDX5* and *DDX17* (si*DDX5/17*). (B) Pie chart showing the number of skipped and included exons following DDX5/DDX17 depletion. (C) RT-PCR validation of Exon Array prediction. Class S (skipped in the absence of DDX5/DDX17) and class I (included in the absence of DDX5/DDX17) exons are in red and in blue, respectively. The exon number is indicated. (D) Boxplot representing the percentage of inclusion obtained by RT-PCR in control siRNA (siCTRL) and si*DDX5/17* conditions. (E) Boxplot representing the 5' splice site (5'ss) strength of class S (S) and class I (I) exons or the total of both classes (S+I), as well as control sets of constitutive (CONS) or alternative (ALT) exons. (F) Boxplot representing the average length of introns upstream and downstream of the indicated exons (defined as in E). (G) Boxplot representing the percentage of GC within the 30 nt-long region downstream of the 5'ss of the indicated exons (as in E). (H) Schematic representation of DDX5/DDX17-regulated exon features. The skipped exons (n in red) have a weak 5' splice site (5'w), a strong 3' splice site (3's) and a downstream GC-rich region. The included exons (n in blue) have a weak 5' splice site (5'w), a weak 3' splice site (3'w) and a downstream AT-rich region. The introns surrounding skipped exons are smaller than introns surrounding included exons. *P*-values: \*\**P*<0.01, \*\*\**P*<0.001.

### Figure 2: The localization of GC-rich hnRNP H/F binding sites defines the splicing outcome of the two classes of DDX5/DDX17-regulated exons.

(A) 100% stacked bar chart representing the percentage of exons with hnRNP H/F CLIP-seq reads. The reads obtained from public hnRNP H/F CLIP-seq datasets were computed into regions of 500 nt upstream and downstream of the exons for the following 3 classes: a subset of 60 class S exons that have been experimentally validated as hnRNP H1-dependent in RNA-seq experiments (“skipped H1”) ; the total pool of class S exons (class S) and a panel of 985 alternative exons (Control). Distribution of the number of reads was shared into 3 groups: no read (0), 1 to 10 reads, and more than 10 reads. (\*\*\*)  $P < 0.001$ ,  $\chi^2$  test) (B) RT-PCR analysis of the effect of DDX5/DDX17 and/or hnRNP H/F depletion on class S exons containing hnRNP H1/F CLIP-seq reads in their upstream or downstream intronic regions. PCR gels are representative of three independent experiments. Quantification is shown as the percentage of alternative sequence inclusion (psi). (C) Effect of hnRNP H/F on class S exons without any CLIP-seq read enrichment. (D) Schematic representation (upper panel) of class I exons (n) with their upstream (n-1) and downstream (n+1) exons. The numbers correspond to the median %GC 30nt upstream or downstream of splice sites. The boxplot (lower panel) is as in Figure 1G except that it also includes class I exons (n-1). The median values for class S exon (n) and class I exons (n-1) are circled in red and in blue, respectively. (E) Top left panel: 100% stacked bar chart representing the percentage of hnRNP H1/F CLIP-seq reads downstream of class I exons (n-1). The CLIP-seq reads were computed into a region of 500 nt downstream of (n-1) exons for class I exons and for 985 alternative control exons. Details are as in A. Right panel: effect of hnRNP H/F depletion on class I exons harboring hnRNP H1/F CLIP-seq reads in their upstream intronic region. Bottom left panel (framed): negative control (no CLIP-seq read). (F) 100% stacked bar chart representing the percentage of TIA1/TIAL1 CLIP-seq reads downstream of class I exons (n). The reads from public TIA1/TIAL1 CLIP-seq datasets were computed into a region of 700nt downstream of exons (n) for class I exons and for a control set of 985 alternative exons. Details are as in A. (G) RT-PCR analysis of class I exons (n)

containing TIA/TIAL binding sites following control, siDDX5/17 and/or siTIA1/TIAL1 treatment. Right panel (framed): negative controls (no CLIP-seq read).

### Figure 3: DDX5, DDX17 and hnRNP H/F cooperate on G-quadruplex structures

(A) Prediction of G-quadruplex structures downstream of 5' ss of exons (n-1) and (n) of class S (red line), class I (blue line), or control (black line) exons. Results of the predictions are schematized under the graphs. (B) Splicing of alternative exons in MCF7 cells treated or not with TMPyP4, a drug that stabilizes G-quadruplexes (left panel), or treated with siDDX5/17 and/or siHNRNPH/F (as in Figure 2). (C) Western blot analysis of hnRNP H1 after immunoprecipitation (IP) of cell lysates from stable HA-DDX5 or FLAG-DDX17 MCF7 cell lines with HA (left) or FLAG (right) antibodies. Cells were induced or not with doxycycline (dox) prior to IP. Inputs (IN) correspond to 10% of total extract. (D) *In situ* Proximity-ligation assay (PLA) showing the interaction between hnRNP H1 and DDX5 in the nuclei (stained with DAPI) of MCF7 cells treated with control siRNA or siDDX5/17. (E) RT-qPCR analysis after RNA-IP using control IgG antibody or hnRNP H1 antibodies. MCF7 cells were cross-linked 48h after transfection with control siRNA (siCTRL) or siDDX5/17. Amplification was performed with primers (arrows) encompassing the 5' ss of class S exon (n) or class I exon (n-1) of the indicated genes. The results, calculated as the percentage of input RNA, were represented as mean values of at least three independent experiments ( $n \geq 3$ ) and normalized to the control sample (IP in the presence of control siRNA), which was arbitrarily set to  $1 \pm SD$  (paired Student's *t*-test : \*  $P < 0.05$ ). (F) Splicing rescue experiments. Stable MCF7 cells were transfected with siDDX5/DDX17 (3' UTR) and treated with doxycyclin to induce the expression of wild-type DDX5 or RNA helicase mutant DDX5-KA. One representative experiment is shown ( $n=3$ ). Quantification is shown as the fold change of psi relative to the control sample.

**Figure 4. DDX5/DDX17 and hnRNP H/F control the splicing of a subset of exons during in epithelial cells and in myoblasts.**

(A) Diagram showing the number of skipped (class S, red) and included (class I, blue) exons upon silencing of *DDX5* and *DDX17* in MCF10a cells. (B) RT-PCR analysis of the splicing of a subset of alternative exons in MCF10A cells treated with control siRNA, si*DDX5/17* and/or si*HNRNPH/F*. (C) Expression of *DDX5* and *DDX17* in TGF $\beta$ -treated MCF10A cells. Actin: loading control. (D) RT-PCR analysis of alternative exons in MCF10A cells treated with TGF $\beta$  (left panel) or si*DDX5/17* (right panel). Quantification is shown as the fold change of psi relative to the control sample. (E) Effect of hnRNP H/F on the splicing of TGF $\beta$ -induced and *DDX5/DDX17*-regulated exons. (F) Expression of *DDX5* and *DDX17* during a time-course differentiation of C2C12 cells. Actin: loading control. (G) Splicing of alternative exons was analyzed in C2C12 cells induced to differentiate (left panel) or treated with si*Ddx5/17* (right panel). The black line separates exons that are coregulated or not by hnRNP H/F. Quantification is as in D. \* the two upper bands were considered as included forms and compared to the lower skipped band.

**Figure 5. A negative feedback loop involving miRNAs regulates DDX5/DDX17 expression during differentiation.**

(A) Expression of *Ddx5* and *Ddx17* in C2C12 cells overexpressing pre-miR-1, pre-miR-206 or a control pre-miR. (B) Luciferase assays measuring in HeLa cells the effect of miR-1 or miR-206 on wild-type and mutated mouse *Ddx5* and *Ddx17* 3' UTR. Histograms represent a mean of at least 3 independent experiments  $\pm$  SD (unpaired Student's *t*-test : \*  $P < 0.05$  ; \*\*  $P < 0.01$  ; \*\*\*  $P < 0.001$ ). (C) Expression of miR-1 (top panel) and miR-206 (bottom panel) in C2C12 cells. Mature miRNA (yellow) or pri-miRNA (orange) were quantified 24 h after

differentiation of cells pre-treated with control siRNA or si*Ddx5/17*. Value were normalized to undifferentiated cells.  $N=3 \pm SD$  ; \*  $P < 0.05$  ; \*\*  $P < 0.01$  (paired *t*-test). **(D)** Chromatin immunoprecipitation experiment showing Ddx5 binding to mmu-*miR-1a-1*(top) and mmu-*miR-206*(bottom) promoters. **(E)** Expression of DDX5 and DDX17 in TGFβ-treated MCF10a cells overexpressing pre-miR-181b or a control pre-miRNA. **(F)** Luciferase assays measuring the effect of miR-181b on wild-type and mutated human *DDX5* 3' UTR in TGFβ-treated HMEC cells. Details are as in B. **(G)** Expression of mature miR-181b (light blue) and pri-miR-181b (dark blue) in MCF10a cells treated or not with TGFβ for 24 h. Details are as in C. **(H)** Chromatin immunoprecipitation experiment showing DDX5 binding to the miR-181b promoter in MCF10a cells after 2 h of TGFβ treatment. **(I)** RT-PCR analysis of splicing of exons in C2C12 cells ectopically expressing miR-1 or miR-206. **(J)** RT-PCR analysis of splicing of exons in MCF10a cells treated or not with TGFβ for 24 h.

**Figure 6. DDX5 and DDX17 control the expression of master transcriptional regulators of EMT and myogenesis.**

**(A)** RT-qPCR measurement of expression of *Myog* and *Mef2c* mRNAs in C2C12 mouse myoblast cells after silencing of *Myod* (si*Myod*) or *Ddx5* and *Ddx17* (si*Ddx5/17*), followed by incubation in growth (- Diff) or differentiation (Diff) medium. “si ctrl”: control siRNA. Right panel: ChIP experiment showing Ddx5 binding to *Myog* and *Mef2c* promoters in differentiating C2C12 cells. **(B)** Immunolabelling of C2C12 cells with an anti-myosin heavy chain (MHC) antibody. Scale bar : 20 μm. **(C)** RT-qPCR measurement of expression of *SNAI1* and *SNAI2* mRNAs in human MCF10a cells after silencing of *SMAD2* and *SMAD3* (si*SMAD2/3*) or *DDX5* and *DDX17* (si*DDX5/17*), followed or not by a 24 h treatment with TGFβ. “si ctrl”: control siRNA. Right panel: ChIP experiment showing DDX5 binding to

*SNAI1* and *SNAI2* promoters in MCF10a after 2 h of TGF $\beta$  treatment. Histograms in A, and C represent mean values of at least three independent experiments ( $n \geq 3$ ) normalized to the control sample  $\pm$  SD (paired Student's *t*-test : \*  $P < 0.05$  ; \*\*  $P < 0.01$ ). (D) Immunolabelling of MCF10a cell junctions with an anti-E-cadherin antibody. Scale bar : 20  $\mu$ m. (E) Orchestrated regulation of gene expression by DDX5 and DDX17 during cell differentiation. The top (orange) and bottom (green) schemes represent the differentiation of mouse C2C12 cells into myotubes and the TGF $\beta$ -induced EMT of MCF10a cells, respectively. In myoblasts and epithelial cells, DDX5 and DDX17 cooperate with hnRNP H/F to control the splicing of a specific subset of exons. Upon induction of differentiation, DDX5 and DDX17 coregulate MyoD- or SMAD-dependent transcriptional activity, directly controlling key effectors of differentiation, including miRNAs that in turn directly repress the expression of both helicases. DDX5 and DDX17 downregulation contributes to the cell-specific variations in the alternative splicing program and to the phenotypic changes occurring during differentiation, like the formation of lamellipodia.

## References

- Abdelhaleem, M. (2005). RNA helicases: regulators of differentiation. *Clinical biochemistry* 38, 499-503.
- Bittencourt, D., Dutertre, M., Sanchez, G., Barbier, J., Gratadou, L., and Auboeuf, D. (2008). Cotranscriptional splicing potentiates the mRNA production from a subset of estradiol-stimulated genes. *Mol Cell Biol* 28, 5811-5824.
- Bland, C.S., Wang, E.T., Vu, A., David, M.P., Castle, J.C., Johnson, J.M., Burge, C.B., and Cooper, T.A. (2010). Global regulation of alternative splicing during myogenic differentiation. *Nucleic Acids Res* 38, 7651-7664.
- Camats, M., Guil, S., Kokolo, M., and Bach-Elias, M. (2008). P68 RNA helicase (DDX5) alters activity of cis- and trans-acting factors of the alternative splicing of H-Ras. *PLoS ONE* 3, e2926.

- Cano, A., Perez-Moreno, M.A., Rodrigo, I., Locascio, A., Blanco, M.J., del Barrio, M.G., Portillo, F., and Nieto, M.A. (2000). The transcription factor snail controls epithelial-mesenchymal transitions by repressing E-cadherin expression. *Nat Cell Biol* 2, 76-83.
- 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., and 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.
- Chen, C.D., Kobayashi, R., and Helfman, D.M. (1999). Binding of hnRNP H to an exonic splicing silencer is involved in the regulation of alternative splicing of the rat beta-tropomyosin gene. *Gene Develop.* 13, 593-606.
- Dardenne, E., Pierredon, S., Driouch, K., Gratadou, L., Lacroix-Triki, M., Espinoza, M.P., Zonta, E., Germann, S., Mortada, H., Villemin, J.P., *et al.* (2012). Splicing switch of an epigenetic regulator by RNA helicases promotes tumor-cell invasiveness. *Nat Struct Mol Biol* 19, 1139-1146.
- Davis, B.N., Hilyard, A.C., Lagna, G., and Hata, A. (2008). SMAD proteins control DROSHA-mediated microRNA maturation. *Nature* 454, 56-61.
- Decorsiere, A., Cayrel, A., Vagner, S., and Millevoi, S. (2011). Essential role for the interaction between hnRNP H/F and a G quadruplex in maintaining p53 pre-mRNA 3'-end processing and function during DNA damage. *Genes Dev* 25, 220-225.
- Del Gatto-Konczak, F., Bourgeois, C.F., Le Guiner, C., Kister, L., Gesnel, M.C., Stevenin, J., and Breathnach, R. (2000). The RNA-binding protein TIA-1 is a novel mammalian splicing regulator acting through intron sequences adjacent to a 5' splice site. *Mol Cell Biol* 20, 6287-6299.
- Ebert, M.S., and Sharp, P.A. (2012). Roles for microRNAs in conferring robustness to biological processes. *Cell* 149, 515-524.
- Fattet, L., Ay, A.S., Bonneau, B., Jallades, L., Mikaelian, I., Treilleux, I., Gillet, G., Hesling, C., and Rimokh, R. (2013). TIF1gamma requires sumoylation to exert its repressive activity on TGFbeta signaling. *J Cell Sci* 126, 3713-3723.
- Forch, P., Puig, O., Kedersha, N., Martinez, C., Granneman, S., Seraphin, B., Anderson, P., and Valcarcel, J. (2000). The apoptosis-promoting factor TIA-1 is a regulator of alternative pre-mRNA splicing. *Mol Cell* 6, 1089-1098.
- Fukuda, T., Yamagata, K., Fujiyama, S., Matsumoto, T., Koshida, I., Yoshimura, K., Mihara, M., Naitou, M., Endoh, H., Nakamura, T., *et al.* (2007). DEAD-box RNA helicase subunits of the Drosha complex are required for processing of rRNA and a subset of microRNAs. *Nat Cell Biol* 9, 604-611.
- Fuller-Pace, F.V., and Moore, H.C. (2011). RNA helicases p68 and p72: multifunctional proteins with important implications for cancer development. *Future Oncol* 7, 239-251.
- Germann, S., Gratadou, L., Zonta, E., Dardenne, E., Gaudineau, B., Fougere, M., Samaan, S., Dutertre, M., Jauliac, S., and Auboeuf, D. (2012). Dual role of the ddx5/ddx17 RNA helicases in the control of the pro-migratory NFAT5 transcription factor. *Oncogene*.
- Huelga, S.C., Vu, A.Q., Arnold, J.D., Liang, T.Y., Liu, P.P., Yan, B.Y., Donohue, J.P., Shiue, L., Hoon, S., Brenner, S., *et al.* (2012). Integrative genome-wide analysis reveals cooperative regulation of alternative splicing by hnRNP proteins. *Cell reports* 1, 167-178.
- Janknecht, R. (2010). Multi-talented DEAD-box proteins and potential tumor promoters: p68 RNA helicase (DDX5) and its paralog, p72 RNA helicase (DDX17). *American journal of translational research* 2, 223-234.
- Jin, Y., Yang, Y., and Zhang, P. (2011). New insights into RNA secondary structure in the alternative splicing of pre-mRNAs. *RNA Biol* 8, 450-457.
- Kalsotra, A., and Cooper, T.A. (2011). Functional consequences of developmentally regulated alternative splicing. *Nat Rev Genet* 12, 715-729.

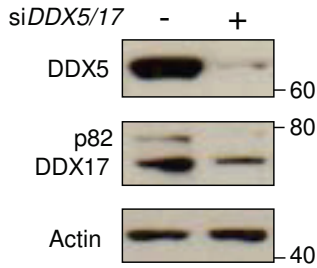


- Kar, A., Fushimi, K., Zhou, X., Ray, P., Shi, C., Chen, X., Liu, Z., Chen, S., and Wu, J.Y. (2011). RNA helicase p68 (DDX5) regulates tau exon 10 splicing by modulating a stem-loop structure at the 5' splice site. *Mol Cell Biol* *31*, 1812-1821.
- Katz, Y., Wang, E.T., Airoidi, E.M., and Burge, C.B. (2010). Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nature methods* *7*, 1009-1015.
- Lasko, P. (2013). The DEAD-box helicase Vasa: evidence for a multiplicity of functions in RNA processes and developmental biology. *Biochim Biophys Acta* *1829*, 810-816.
- Laurent, F.X., Sureau, A., Klein, A.F., Trouslard, F., Gasnier, E., Furling, D., and Marie, J. (2012). New function for the RNA helicase p68/DDX5 as a modifier of MBNL1 activity on expanded CUG repeats. *Nucleic Acids Res* *40*, 3159-3171.
- Lim, L.P., and Burge, C.B. (2001). A computational analysis of sequence features involved in recognition of short introns. *Proc Natl Acad Sci U S A* *98*, 11193-11198.
- Lin, C., Yang, L., Yang, J.J., Huang, Y., and Liu, Z.R. (2005). ATPase/helicase activities of p68 RNA helicase are required for pre-mRNA splicing but not for assembly of the spliceosome. *Mol Cell Biol* *25*, 7484-7493.
- Lin, J.C., and Tarn, W.Y. (2011). RBM4 down-regulates PTB and antagonizes its activity in muscle cell-specific alternative splicing. *J Cell Biol* *193*, 509-520.
- Linder, P., and Jankowsky, E. (2011). From unwinding to clamping - the DEAD box RNA helicase family. *Nat Rev Mol Cell Biol* *12*, 505-516.
- Mallinjoind, P., Villemin, J.P., Mortada, H., Polay-Espinoza, M., Desmet, F.O., Samaan, S., Chautard, E., Tranchevent, L.C., and Auboeuf, D. (2013). Endothelial, epithelial, and fibroblast cells exhibit specific splicing programs independently of their tissue of origin. *Genome Res*.
- Merkin, J., Russell, C., Chen, P., and Burge, C.B. (2012). Evolutionary dynamics of gene and isoform regulation in Mammalian tissues. *Science* *338*, 1593-1599.
- Paul, S., Dansithong, W., Jog, S.P., Holt, I., Mittal, S., Brook, J.D., Morris, G.E., Comai, L., and Reddy, S. (2011). Expanded CUG repeats Dysregulate RNA splicing by altering the stoichiometry of the muscleblind 1 complex. *J Biol Chem* *286*, 38427-38438.
- Paz, I., Akerman, M., Dror, I., Kosti, I., and Mandel-Gutfreund, Y. (2010). SFmap: a web server for motif analysis and prediction of splicing factor binding sites. *Nucleic Acids Res* *38*, W281-285.
- Rochlin, K., Yu, S., Roy, S., and Baylies, M.K. (2010). Myoblast fusion: when it takes more to make one. *Dev Biol* *341*, 66-83.
- Samaan, S., Tranchevent, L.C., Dardenne, E., Polay Espinoza, M., Zonta, E., Germann, S., Gratadou, L., Dutertre, M., and Auboeuf, D. (2013). The Ddx5 and Ddx17 RNA helicases are cornerstones in the complex regulatory array of steroid hormone-signaling pathways. *Nucleic Acids Res*.
- Samatanga, B., Dominguez, C., Jelesarov, I., and Allain, F.H. (2013). The high kinetic stability of a G-quadruplex limits hnRNP F qRRM3 binding to G-tract RNA. *Nucleic Acids Res* *41*, 2505-2516.
- Schor, I.E., Gomez Acuna, L.I., and Kornblihtt, A.R. (2013). Coupling between transcription and alternative splicing. *Cancer treatment and research* *158*, 1-24.
- Shapiro, I.M., Cheng, A.W., Flytzanis, N.C., Balsamo, M., Condeelis, J.S., Oktay, M.H., Burge, C.B., and Gertler, F.B. (2011). An EMT-driven alternative splicing program occurs in human breast cancer and modulates cellular phenotype. *PLoS Genet* *7*, e1002218.
- Thiery, J.P., Acloque, H., Huang, R.Y., and Nieto, M.A. (2009). Epithelial-mesenchymal transitions in development and disease. *Cell* *139*, 871-890.
- Wang, B., Hsu, S.H., Majumder, S., Kutay, H., Huang, W., Jacob, S.T., and Ghoshal, K. (2010a). TGFbeta-mediated upregulation of hepatic miR-181b promotes hepatocarcinogenesis by targeting TIMP3. *Oncogene* *29*, 1787-1797.

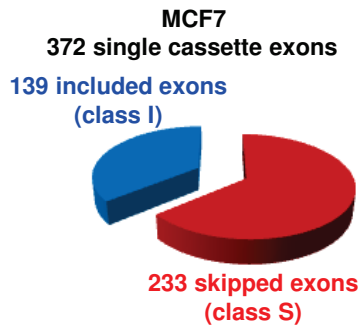
- Wang, D., Huang, J., and Hu, Z. (2012a). RNA helicase DDX5 regulates microRNA expression and contributes to cytoskeletal reorganization in basal breast cancer cells. *Molecular & cellular proteomics* : MCP *11*, M111 011932.
- Wang, E., Aslanzadeh, V., Papa, F., Zhu, H., de la Grange, P., and Cambi, F. (2012b). Global profiling of alternative splicing events and gene expression regulated by hnRNPH/F. *PLoS One* *7*, e51266.
- Wang, E., and Cambi, F. (2009). Heterogeneous nuclear ribonucleoproteins H and F regulate the proteolipid protein/DM20 ratio by recruiting U1 small nuclear ribonucleoprotein through a complex array of G runs. *J Biol Chem* *284*, 11194-11204.
- Wang, E.T., Sandberg, R., Luo, S., Khrebtkova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P., and Burge, C.B. (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature* *456*, 470-476.
- Wang, Z., Kayikci, M., Briese, M., Zarnack, K., Luscombe, N.M., Rot, G., Zupan, B., Curk, T., and Ule, J. (2010b). iCLIP predicts the dual splicing effects of TIA-RNA interactions. *PLoS Biol* *8*, e1000530.
- Warner, D.R., Bhattacharjee, V., Yin, X., Singh, S., Mukhopadhyay, P., Pisano, M.M., and Greene, R.M. (2004). Functional interaction between Smad, CREB binding protein, and p68 RNA helicase. *Biochem Biophys Res Commun* *324*, 70-76.
- Warzecha, C.C., Jiang, P., Amirikian, K., Dittmar, K.A., Lu, H., Shen, S., Guo, W., Xing, Y., and Carstens, R.P. (2010). An ESRP-regulated splicing programme is abrogated during the epithelial-mesenchymal transition. *EMBO J* *29*, 3286-3300.
- Williams, A.H., Liu, N., van Rooij, E., and Olson, E.N. (2009). MicroRNA control of muscle development and disease. *Curr Opin Cell Biol* *21*, 461-469.
- Xiao, S., Zhang, J.Y., Zheng, K.W., Hao, Y.H., and Tan, Z. (2013). Bioinformatic analysis reveals an evolutionary selection for DNA:RNA hybrid G-quadruplex structures as putative transcription regulatory elements in warm-blooded animals. *Nucleic Acids Res* *41*, 10379-10390.
- Xiao, X., Wang, Z., Jang, M., Nutiu, R., Wang, E.T., and Burge, C.B. (2009). Splice site strength-dependent activity and genetic buffering by poly-G runs. *Nat Struct Mol Biol* *16*, 1094-1100.
- Yang, L., Lin, C., and Liu, Z.R. (2006). P68 RNA helicase mediates PDGF-induced epithelial mesenchymal transition by displacing Axin from beta-catenin. *Cell* *127*, 139-155.
- Yilmaz, M., and Christofori, G. (2009). EMT, the cytoskeleton, and cancer cell invasion. *Cancer metastasis reviews* *28*, 15-33.
- Zhang, X.H., Leslie, C.S., and Chasin, L.A. (2005). Dichotomous splicing signals in exon flanks. *Genome Res* *15*, 768-779.
- Zonta, E., Bittencourt, D., Samaan, S., Germann, S., Dutertre, M., and Auboeuf, D. (2013). The RNA helicase DDX5/p68 is a key factor promoting c-fos expression at different levels from transcription to mRNA export. *Nucleic Acids Res* *41*, 554-564.

# Figure 1

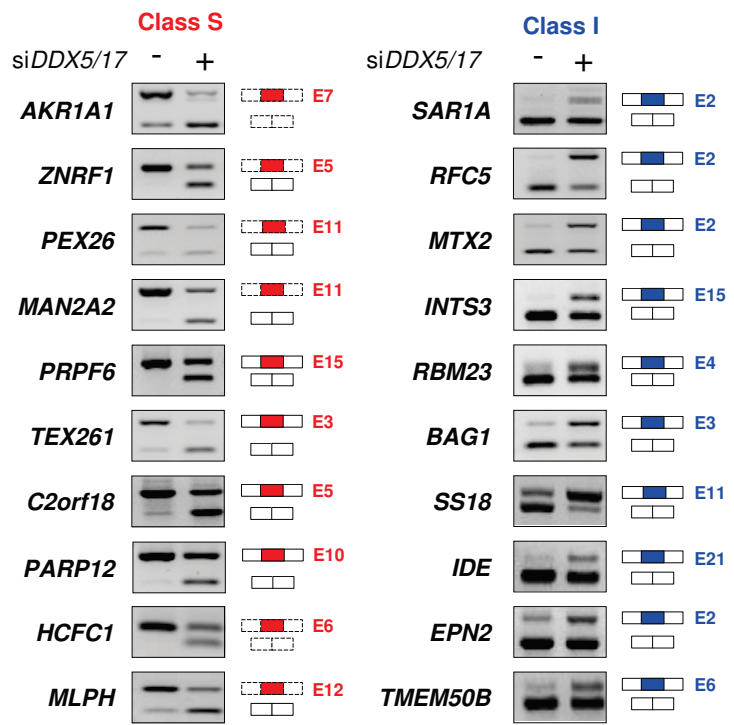
## A



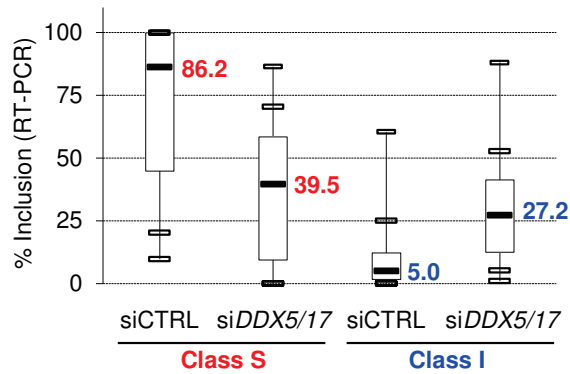
## B



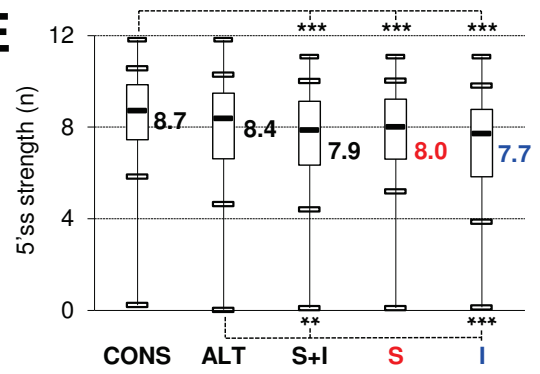
## C



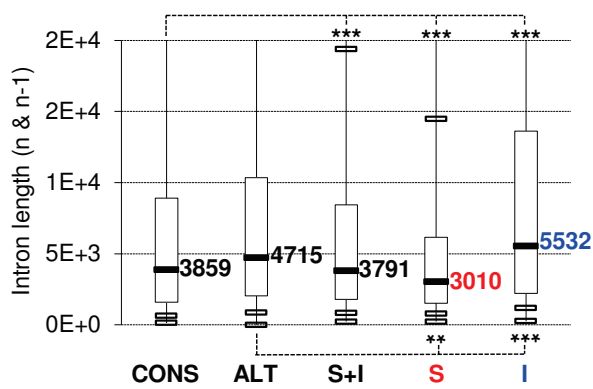
## D



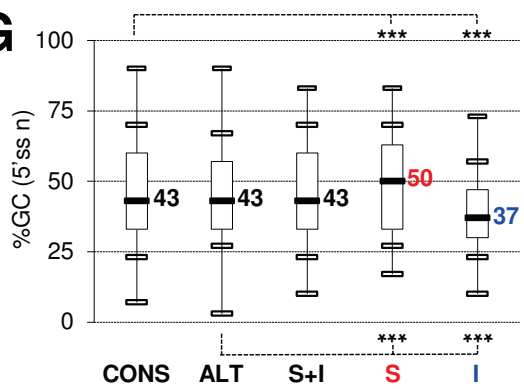
## E



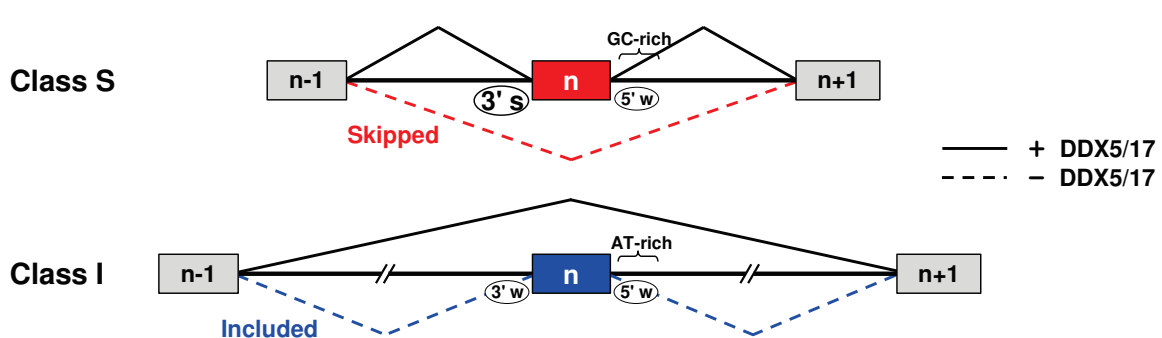
## F



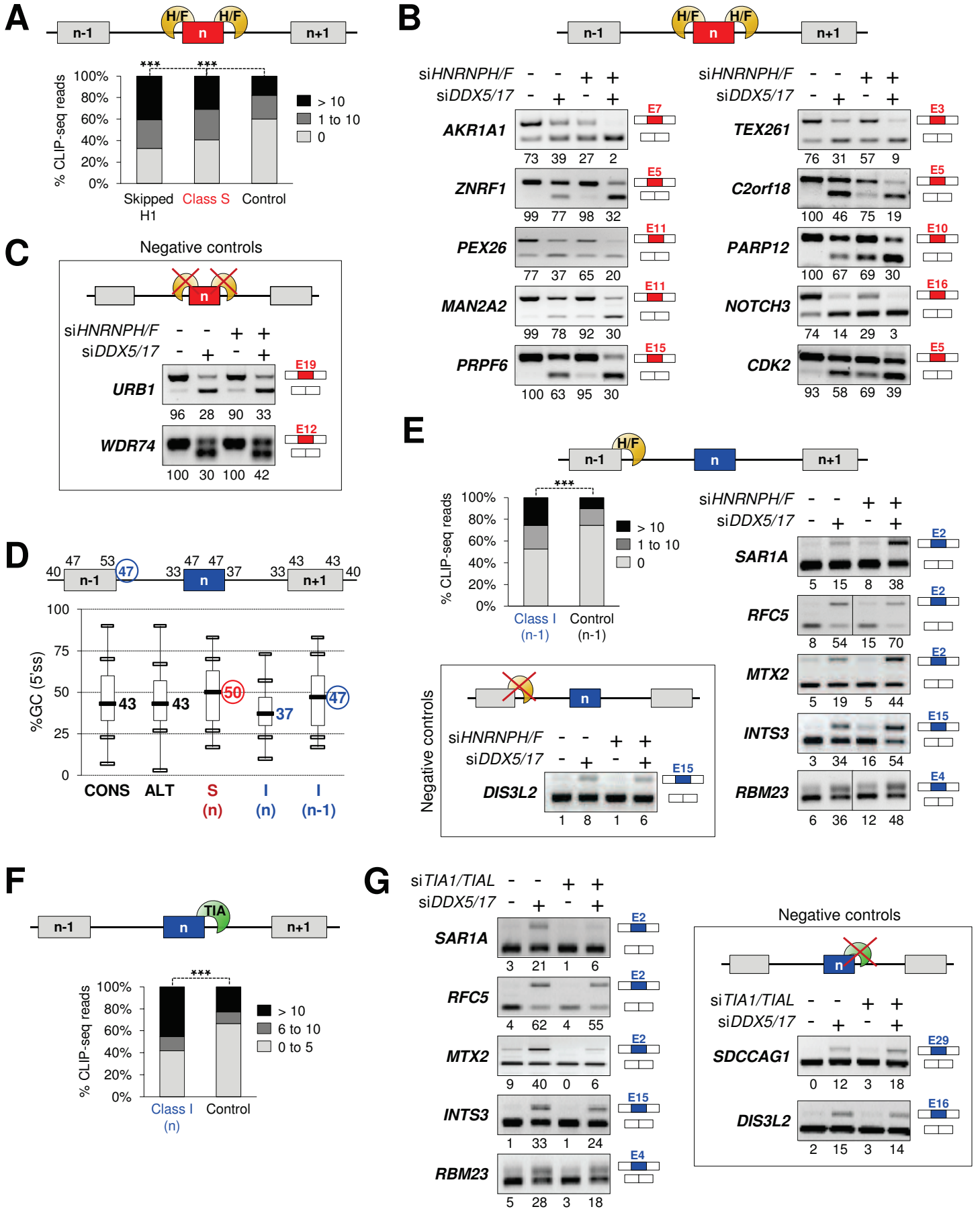
## G



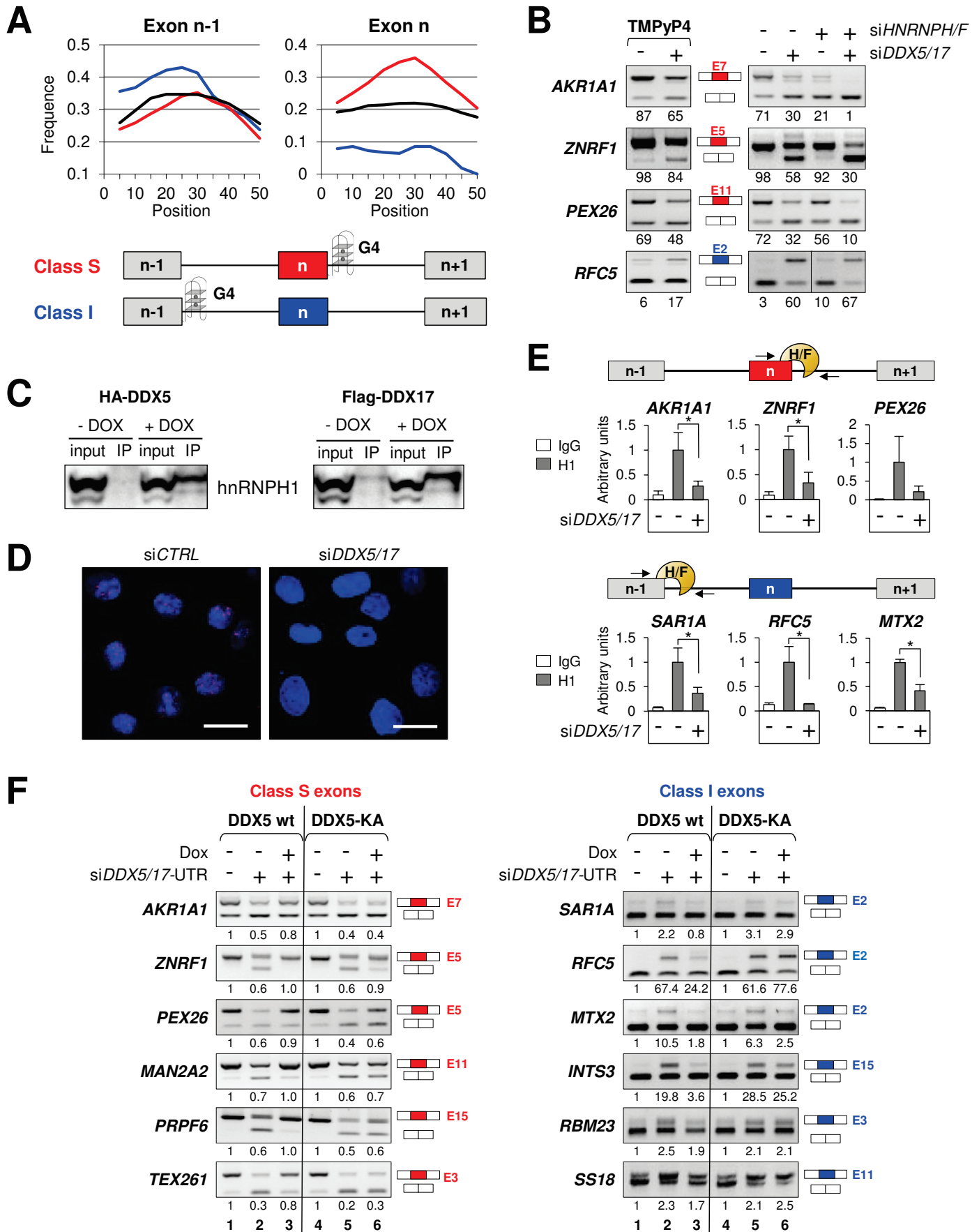
## H



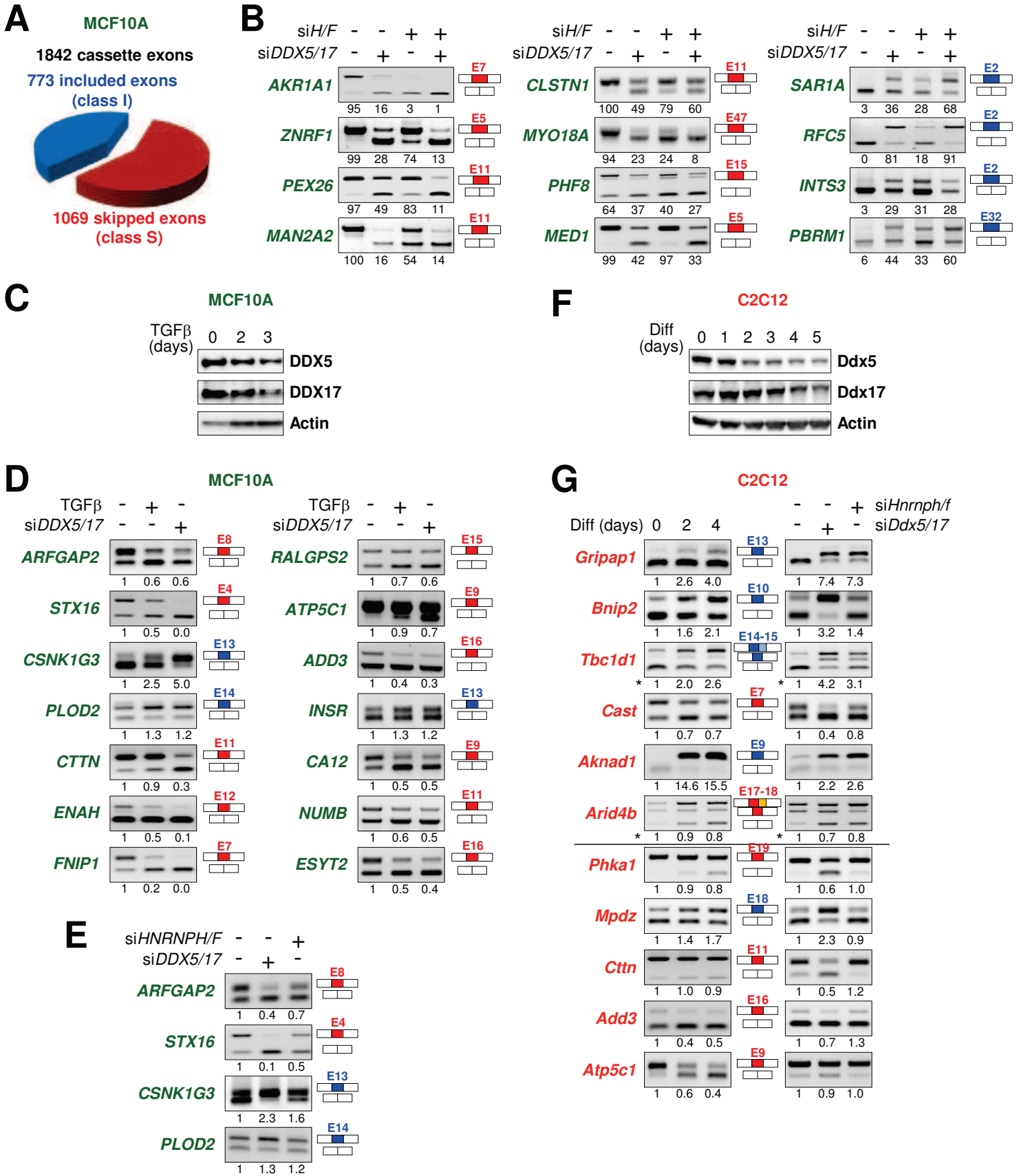
# Figure 2



# Figure 3

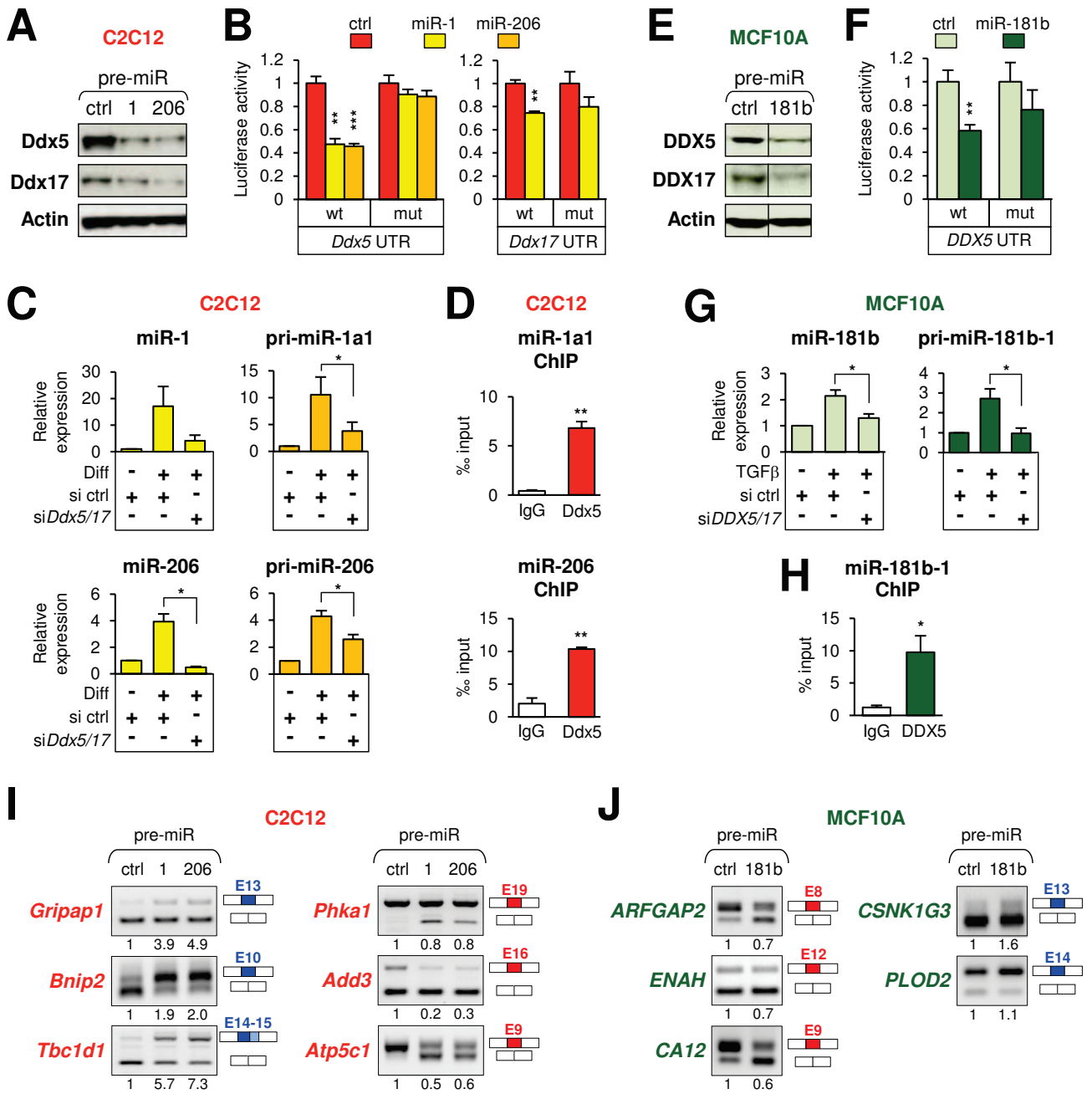


# Figure 4

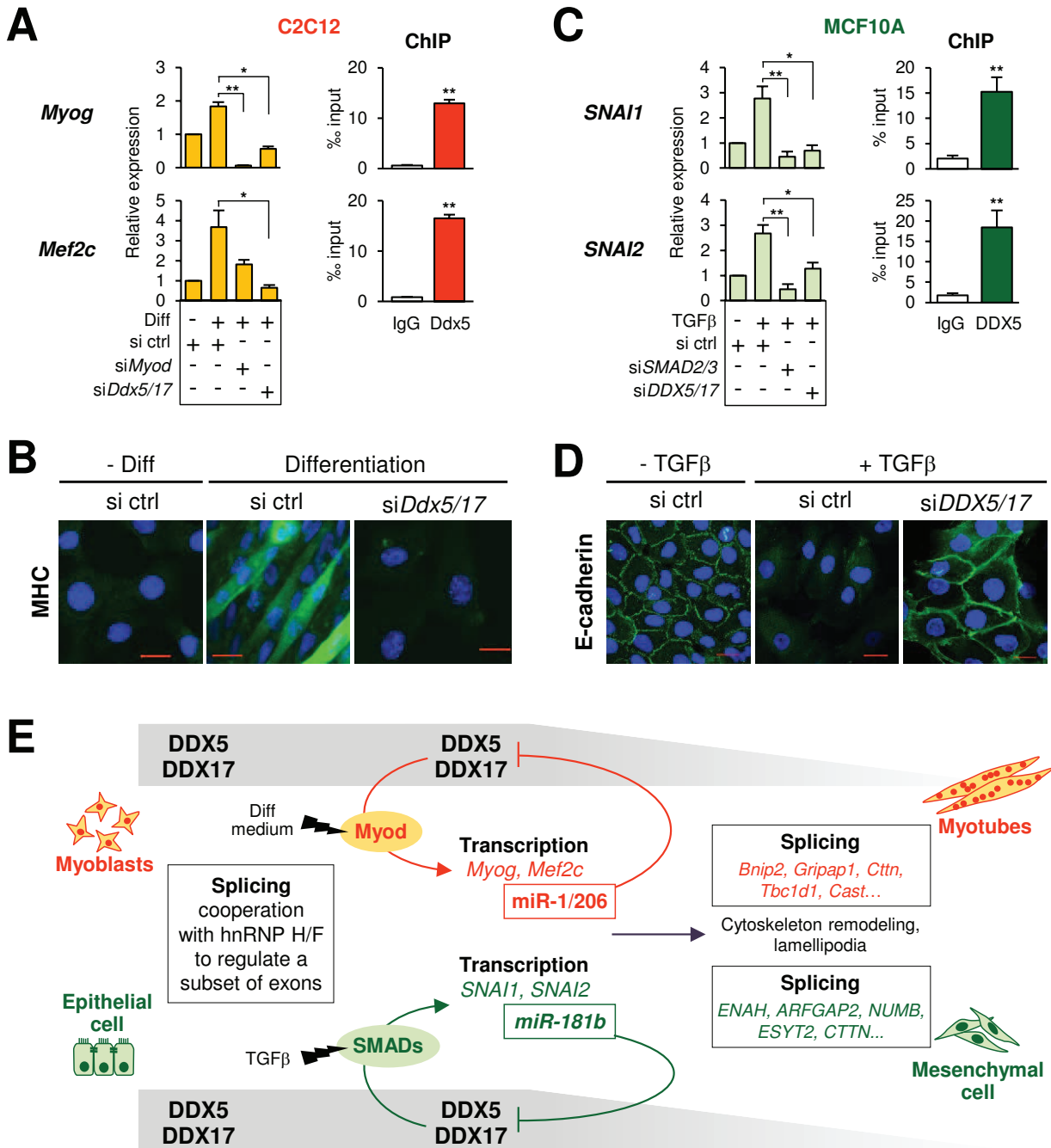




# Figure 5



**Figure 6**





## SUPPLEMENTAL INFORMATION

**Figure S1.** Regulation of alternative splicing of a subset of exons by a collaboration between DDX5/DDX17 and hnRNP H/F at 5' splice sites followed by GC-rich sequences.

**Figure S2.** Interactions between DDX5 and hnRNP H1 in epithelial MCF10A cells ; characteristics of DDX5/DDX17-regulated exons in epithelial MCF10A cells.

**Figure S3.** Interactions between DDX5 and hnRNP H1 in C2C12 cells.

**Figure S4.** Predicted miRNA binding sites in the 3' UTR of *DDX5* and *DDX17*.

**Figure S5.** Interactions between Ddx5 and MyoD in C2C12 cells ; regulation of miR-181b expression by SMAD proteins in MCF10A cells.

**Figure S6.** DDX5 and DDX17 regulate TGF $\beta$ -induced EMT by controlling the expression of master EMT transcription factors SNAIL1/2 and the resulting expression of epithelial and mesenchymal markers.

**Figure S7.** Formation of lamellipodia in DDX5/DDX17-depleted cells ; putative regulation of cytoskeleton organization by DDX5 and DDX17.

**Table S1.** List of Alternative Skipped Exon (ASE) regulated by DDX5/DDX17 in MCF7 cells.

**Table S2.** RNA-seq and CLIP seq data related to Figure 2.

**Table S3.** List of Alternative Skipped Exons (ASE) regulated by DDX5/DDX17 in MCF10A cells.

**Table S4.** List of siRNA and PCR primers used in this study.

**Supplementary movies** are viewable on a specially dedicated website localized on our bioinformatics platform at the following address :

<http://fasterdb.lyon.unicancer.fr/dardenne/>

**Movie S1.** Time-lapse videomicroscopy of MCF10A cells treated with control siRNA in the absence of TGF $\beta$ .

**Movie S2 :** Time-lapse videomicroscopy of MCF10A cells treated with control siRNA and with TGF $\beta$ , showing the formation of lamellipodia .

**Movie S3 :** Time-lapse videomicroscopy of MCF10A cells treated with *DDX5/17* siRNA in the absence of TGF $\beta$ , showing the formation of lamellipodia .

**Movie S4 :** Time-lapse videomicroscopy of C2C12 cells treated with control siRNA.

**Movie S5 :** Time-lapse videomicroscopy of C2C12 cells treated with *Ddx5/17* siRNA showing the formation of lamellipodia.

## Supplemental Figure Legends

**Figure S1.** (A) Plot representing the difference in percentage of inclusion of single cassette exons (N=121). X-axis and Y-axis represent the difference in the control siRNA and si*DDX5/17* conditions of the percentage of exon inclusion rate calculated with RT-PCR band intensities quantified by ImageJ software or from the Exon Arrays, respectively. Primers used for the RT-PCR validation of the 121 exons can be found in Table S4. (B) Boxplot representing the percentage of inclusion computed from Exon Arrays in siCTRL and si*DDX5/17* conditions. (C) Boxplot representing the percentage of inclusion computed from mRNAs from public libraries. (D) Boxplot representing the 3' splice site (3'ss) strength of class S (S) and class I (I) exons or the total of both classes (S+I), as well as control sets of constitutive (CONS) or alternative (ALT) exons. (E) Density of previously published intronic splicing enhancers (ISE Lim and ISE Zhang) and hnRNP H/F binding motifs within 50nt downstream of *DDX5/DDX17*-regulated class S exons, control alternative and constitutive exons. Normalized occurrence corresponds to the percentage of motif numbers in a window of 10nt divided by the total number of sequences in each category x10. (F) Left panel: RT-qPCR validation of the effect of si*HNRNPH/F* in MCF7 cells. The total *HNRNPF* mRNA level in depleted cells was normalized relative to its level in cells treated with control siRNA. Right panels: western-blots validating the effect of si*HNRNPH/F* and si*DDX5/17* siRNAs on the expression of hnRNP H1, *DDX5* and *DDX17*. Actin: loading control. (G) Western-blots validating the effect of si*TIA1/TIAL* and si*DDX5/17* siRNAs on the expression of *TIA1*, *TIAL*, *DDX5* and *DDX17*. Actin: loading control. (H) *DDX17* splicing rescue experiments. Stable MCF7 cells were transfected with si*DDX5/DDX17* (3' UTR) and treated with doxycyclin to induce the expression of wild-type (*DDX17*) or RNA helicase mutant (*DDX5-KR*). One

representative experiment is shown ( $n=3$ ). Quantification is shown as the fold change of psi relative to the control sample.

**Figure S2.** (A) Western-blot validating the effect of siDDX5/17 in MCF10A cells. (B) Detection of endogenous DDX5-hnRNP H1 complexes in MCF10A cells by *in situ* proximity-ligation assays. Cells treated with siDDX5/17 siRNA were used as negative control. Scale bar : 20  $\mu$ m. (C) Boxplot representing the 5' splice site (5'ss) strength of class S (S) and class I (I) exons regulated by DDX5/DDX17 in MCF10A cells, compared to a set of constitutive (CONS) control exons (as in Figure 1E). (\*\*\*)  $P < 0.001$ . (D) Boxplot representing the percentage of GC within the 30 nt downstream of the 5'ss of MCF10A class S exons (n) and class I exons (n-1). Calculated  $p$ -values did not reach statistical significance but the results show a similar trend compared to MCF7 cells (Figure 1G). (E) 100% stacked bar chart representing the percentage of hnRNP H1/F CLIP-seq reads downstream of MCF10A class S exons (n) and class I exons (n-1). Details are as in Figure 2. (\*\*\*)  $P < 0.001$ ,  $\chi^2$  test)

**Figure S3.** (A) Coimmunoprecipitation assays. Epithelial TGF $\beta$ -responsive HMEC cells were used instead of MCF10A cells that are very hard to transfect with plasmids. Cells were transfected with Flag-SMAD3 or Flag-SMAD4 plasmids and treated or not for 2 h with TGF $\beta$ . 10% of input were used as control and 100% of immunoprecipitation was loaded. DDX5 was detected by western-blot after Flag immunoprecipitation. (B) Western blot showing the expression of Ddx5, Ddx17 and Myod in C2C12 cells transfected with control or siDdx5/17 siRNAs. Actin was used as a loading control. (C) Coimmunoprecipitation between Ddx5 and hnRNP H1 in C2C12 cells. Endogenous Ddx5 was specifically detected in the fraction immunoprecipitated with HnRNP H1 and not with control IgG antibodies. (D) Detection of endogenous Ddx5-HnRNP H1 complexes in C2C12 cells by *in situ* proximity-ligation

assays. Cells treated with *siDdx5/17* or *siHnrnp/hf* siRNAs were used as negative controls of the interaction. Scale bar : 20  $\mu$ m.

**Figure S4.** (A) Predicted miRNA binding sites in the *DDX5* 3' UTR. Top panel : screen capture from the TargetScan website showing the predicted sites for miR-1, miR-206 and miR-181b in the *DDX5* 3' UTR. Note that these sites were predicted by other freely accessible softwares such as miRanda or PicTar (data not shown). Below are shown details about the predicted base-pairing between the different miRNAs and their respective target sites. As the effects of miR-1/206 and miR-181b were studied respectively in mouse C2C12 cells and human MCF10A cells, their base-pairing with the *DDX5* 3' UTR was shown within their respective species context. The identical nucleotides between highly related miR-1 and miR-206 are boxed in blue, indicating the perfect conservation at the level of the 5' end-located seed region, which is critical for the base-pairing with its mRNA target (boxed in green). Two binding sites for miR-181b were predicted within the 3' UTR of human *DDX5* transcripts, but the second site is less evolutionarily conserved and was not tested in our experiments. The 4-nucleotide deletions that were made within the *DDX5* UTR luciferase constructs to destroy the predicted miR-1/206 and miR-181b target sites in luciferase assays are shown at the bottom of each box. (B) Predicted miRNA binding sites in the *DDX17* 3' UTR. Details are as in A. No binding site was predicted for the miR-181 family.

**Figure S5.** (A) Detection of endogenous Ddx5-Myod complexes in C2C12 cells by *in situ* proximity ligation assays. Cells transfected with *Ddx5/17* or *Myod* siRNAs were used as negative controls. The western-blot under the *siMyod* panel demonstrates the efficiency of the depletion. (B) Validation by western-blotting of the effect of siRNAs against *SMAD2/3* (a mixture of 2 siRNAs). Vinculin was used as loading control. (C) Expression of mature miR-

181b was measured by RT-qPCR 24 h after the induction by TGF $\beta$  in MCF10a cells pre-treated with control or SMAD2/3 siRNA.

**Figure S6.** (A) DDX5/DDX17 regulate the expression of SNAIL1/2. Western-blot analysis of SNAIL1 and SNAIL2 (SLUG) expression after 48 h of DDX5/DDX17 depletion and 33 h of TGF $\beta$  treatment. (B) Chromatin immunoprecipitation showing the binding of SMAD4 to *SNAI1* and *SNAI2* promoter regions. Immunoprecipitation of SMAD4 proteins was performed after 2 h of TGF $\beta$  treatment. The amplicons analyzed by RT-qPCR correspond to a region of *SNAI1* and *SNAI2* promoters containing a SMAD binding element.  $N=3$  (biological replicates)  $\pm$  SD ; \*\*\* $P < 0.001$  (paired  $t$ -test). The SMAD4 ChIP on *SNAI2* promoter did not reach statistical significance, even though each of the independent biological replicates showed clear binding compared to the control IgG. (C) RT-qPCR analysis of the expression of epithelial (*OCLN*) and mesenchymal (*FAP*) markers during TGF $\beta$ -induced EMT of MCF10A cells, pre-treated with control or siDDX5/17 siRNA. Value were normalized to untreated cells.  $N=3$  (biological replicates)  $\pm$  SD ; \*  $P < 0.05$  (paired  $t$ -test). (D) Western blot analysis of E-Cadherin and Occludin expression. As expected TGF $\beta$  treatment induced a dramatic decrease of the expression of both proteins, but pre-treatment with DDX5/17 siRNA compromised this inhibition. (E) MCF10A cell immunostaining of epithelial junctions using Occludin and  $\beta$ -catenin antibodies following DDX5/DDX17 depletion and TGF $\beta$  treatment. Nuclei were stained with Hoescht.

**Figure S7.** (A) Formation of lamellipodia in DDX5/DDX17-depleted cells. Top: Phase-contrast images of MCF10A cells transfected with control or siDDX5/17 siRNAs and treated or not with TGF $\beta$ . Bottom: C2C12 cells transfected with control or siDdx5/17 siRNA. Scale bar : 50  $\mu$ m. Inset: magnified view of lamellipodia (white arrows) in TGF $\beta$ -treated MCF10A cells and

in both DDX5/DDX17-depleted cell lines (scale bar : 25  $\mu$ m). **(B)** Scheme of putative regulation of cytoskeleton organization by DDX5 and DDX17. The scheme represents the KEGG pathway map entitled “Regulation of actin cytoskeleton”, downloaded from the Kyoto Encyclopedia of Genes and Genomes website (<http://www.genome.jp/kegg/>). Genes that were regulated at the splicing level by DDX5/DDX17 during differentiation of C2C12 cells (orange boxes), during TGF $\beta$ -induced EMT of MCF10a cells (green boxes) or in both processes (green/orange boxes) were added to the scheme. Single headed arrows indicate a direct impact or a functional link between DDX5/DDX17-regulated genes and elements of the pathway ; double headed arrows indicate reported interactions. Some details about the functional links between the genes and cytoskeleton remodelling and/or lamellipodia formation are given in the table below.

## **Supplementary methods**

### **Antibodies and co-immunoprecipitation**

Primary antibodies used for western-blotting: DDX5 (ab10261, Abcam), DDX17 (ab24601, Abcam), HA (clone 3F10, Roche), Flag (F1804, Sigma), hnRNP H1 (A300-511A, Bethyl), Actin (I-19 sc-1616, Santa Cruz), TIA-1 (sc1751, Santa-Cruz), TIAL1 (TIAR) (sc1749, Santa-Cruz), SMAD3 (sc-9513, Santa Cruz), SMAD2 (sc-3122, Santa Cruz), SNAIL1 (sc-28199, Santa Cruz), SNAIL2/SLUG (sc-9585, Santa Cruz), E-cadherin (610181, BD), OCLN (71-1500, Life Technologies), MYOD (M3512, Dako), GAPDH (50150, Covalab) or actin (I-19 sc1616, Santa Cruz).

For co-IP in stable MCF-7 cell lines, cells were induced with doxycycline to express DDX5-HA or Flag-DDX17. For HMEC cells, cells were transfected for 48 h with expression vectors encoding Flag-SMAD3 and Flag-SMAD4 (a kind gift of P. Ten Dijke, Leiden, Netherlands),

and then treated with TGF $\beta$ -1 or not (SB-431542 treatment) for 2 h before lysis. In all experiments (including co-IP of endogenous proteins), cells were lysed in buffer containing 50 mM Tris-HCl pH 8, 150 mM NaCl, 1% NP40, protease and phosphatase inhibitors (Roche) and incubated for 30 minutes on ice. Cleared cell extracts were incubated overnight with 4-5  $\mu$ g of HA (clone 3F10, Roche), Flag (F1804, Sigma) or the indicated antibody, or control antibodies bound to protein G magnetic beads (Life Technologies). The protein-bound magnetic beads were washed three times with PBS 1X, resuspended in loading buffer before Western blot analysis. Inputs correspond to 5-10% of the cleared cell lysate, depending on the experiment.

### **Immunofluorescence**

Cells were fixed in 4% paraformaldehyde for 20 min. After 3 washes in 1x PBS, cells were permeabilized in 0.2% Triton X-100 for 30 min and left for 1 h in blocking solution (1x PBS, 15% serum, 0.1% Triton X-100). Primary antibodies : MHC (MY-32, Sigma), OCLN (71-1500, Life Technologies), E-cadherin (610181, BD) and  $\beta$ -catenin (Santa cruz). Proteins were revealed using a FITC-conjugated anti-mouse IgG (Sigma). Image acquisition was performed using a Zeiss LSM780 or a Leica SP5 confocal microscope (63x objective) and images were processed with the ImageJ program.

### ***In situ* proximity ligation assay**

MCF7, MCF10A or C2C12 cells were transfected for 48 h with either control or *DDX5/17* siRNA. In the case of C2C12 cells, growth medium was removed after 24 h and incubated with differentiation medium for 24 h. The *in situ* proximity ligation assay (PLA) was conducted using the Duolink II Kit (Olink Bioscience) as recommended by the manufacturer. Fixed cells were incubated with antibodies against Ddx5 and the tested interactant. Interactions were

revealed using secondary antibodies coupled to specific PLA DNA probes that hybridized and were enzymatically joined when located in close proximity. After rolling circle amplification, each interaction generated a fluorescent spot that was analyzed by confocal microscopy.

### **Videomicroscopy**

MCF10A or C2C12 cells were transfected with control or DDX5/17 siRNA (for 15 and 24 h, respectively). When indicated, MCF10A were also treated for 33 h with TGF $\beta$ . Culture plates were then placed in the incubation chamber (37°C and 5% CO<sub>2</sub>) of an AXIOVERT 200M microscope. Pictures were taken every 5 min for 24 h (MCF10A) or 9 h (C2C12) for each field and final stacked pictures were processed using ImageJ. Movies are viewable on a specially dedicated website localized on our bioinformatics platform at the following address :

<http://fasterdb.lyon.unicancer.fr/dardenne/>

### **Luciferase assays**

The following 3' UTR fragments were PCR-amplified from cDNA or genomic DNA and cloned in the psiCHECK-2 plasmid (Promega) downstream of the renilla luciferase gene : mDdx5 UTR (Figure 5B), fragment 9-311 (NM\_007840) ; mDdx17 UTR (Figure 5B) : fragment 1327-2222 (NM\_152806) ; hDDX5 UTR (Figure 5F) : fragment 852-1667 plus 37 additional nucleotides downstream of the end of 3' UTR (NM\_004396). Mutant derivatives were obtained by deletion of 4 nucleotides within miR-1/206 and miR-181b binding sites as indicated in Figure S4 by directed mutagenesis. The same plasmid also contains the firefly luciferase gene used for internal normalization of the data.

For the analysis of miR-1/206 effect, HeLa cells were seeded in 24-well plates and transfected for 24 h with 50 ng of luciferase reporter along with 10 nM pre-miR precursor (control or pre-miR-1 or pre-miR-206, Life Technologies), using Lipofectamine 2000 (Life Technologies)

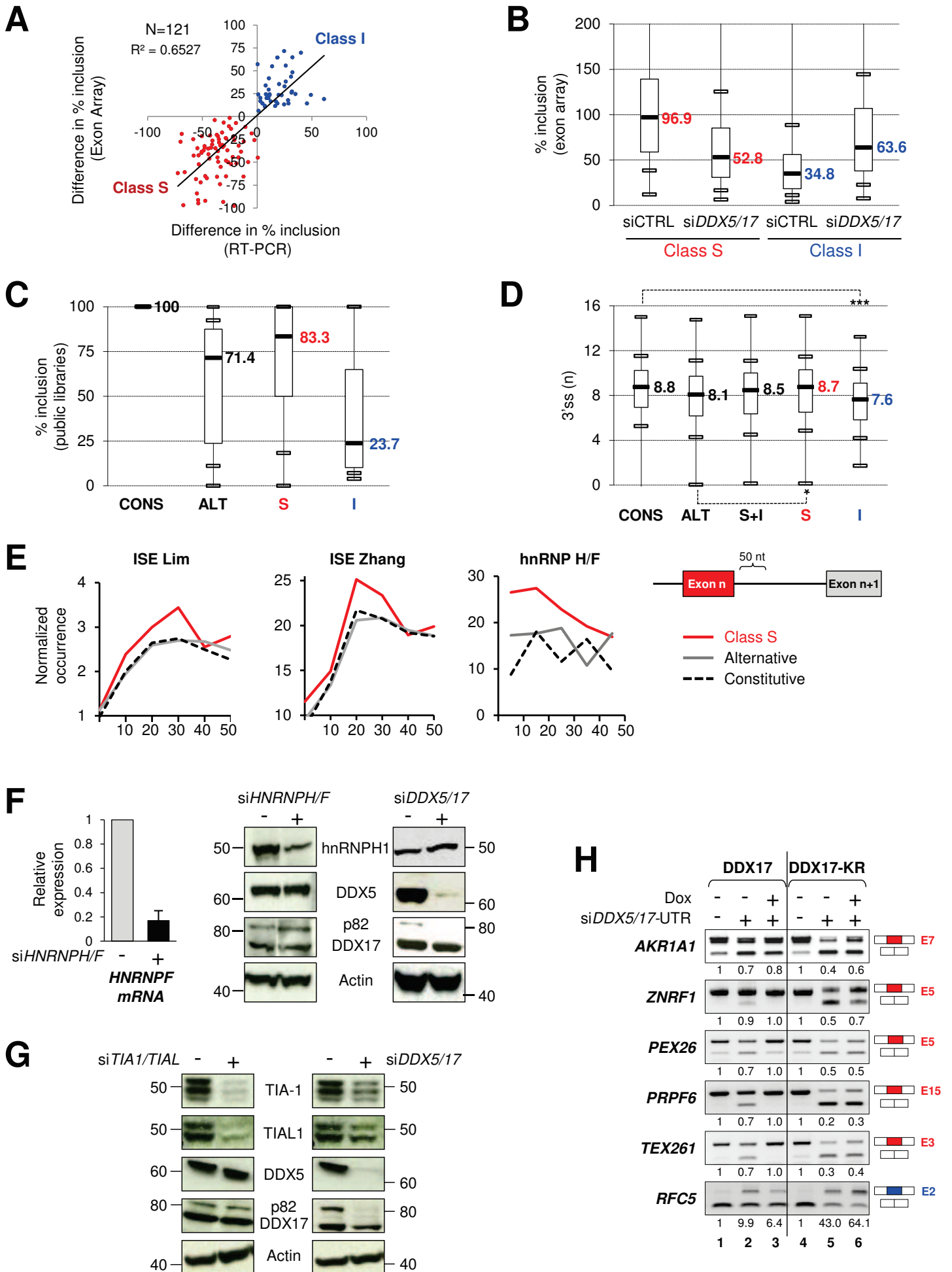


following manufacturer's instructions. For miR-181b, the protocol was the same except that HMEC cells were used and pre-treated for 24 h with TGF $\beta$  prior to transfection. 24 h after transfection, cells were lysed and luciferase assays were carried out using the Dual-Luciferase® Reporter Assay System (Promega), as recommended by the manufacturer. Data were expressed as the ratio of renilla (test) to firefly (internal control) luminescence, and normalized to the ratio obtained with the control pre-miR (arbitrarily set to 1).

### **Microarray analysis**

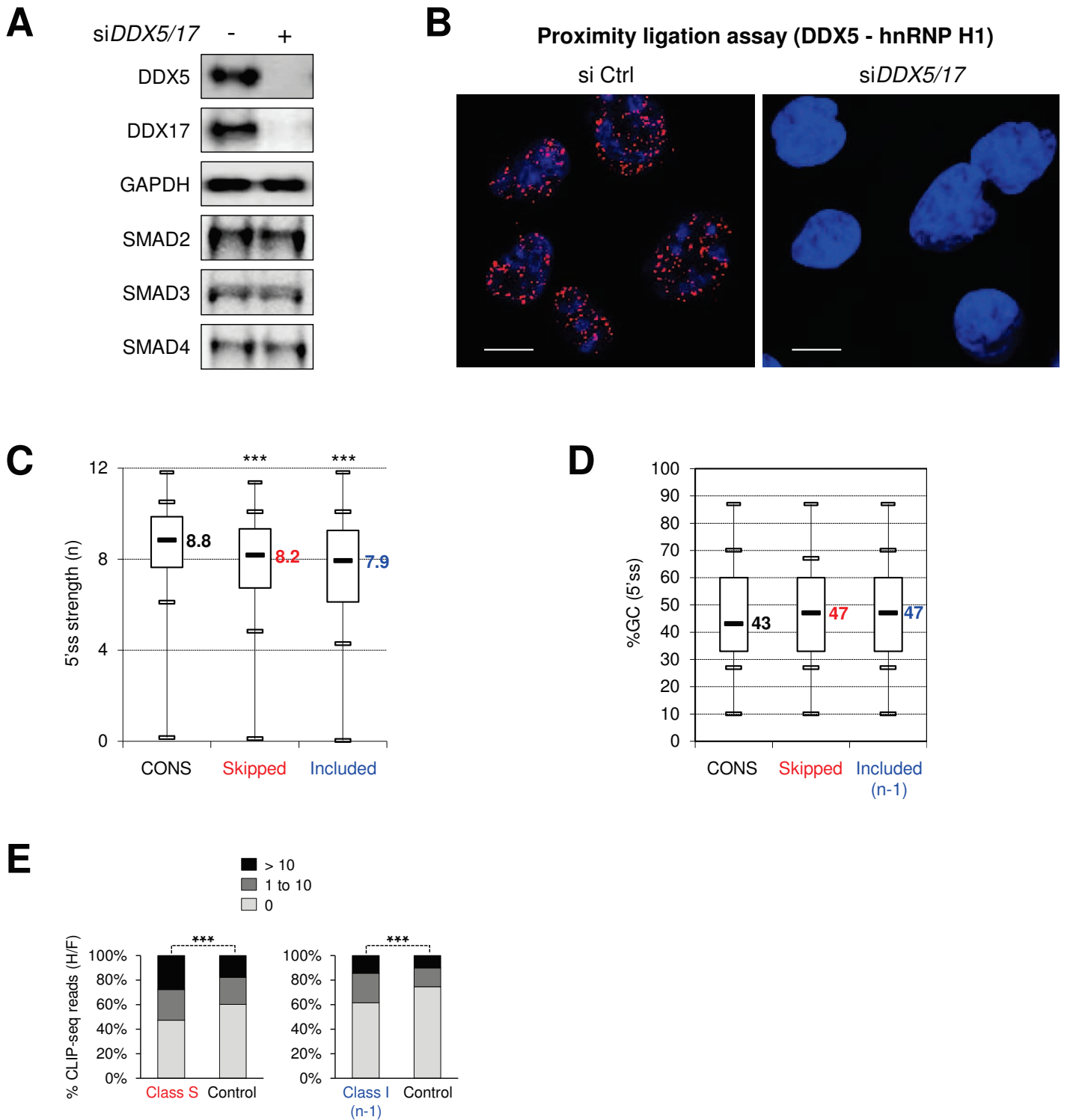
One  $\mu$ g of total RNA purified with TRIzol were labeled with Ambion reagents and hybridized to Affymetrix GeneChip Human Exon 1.0 ST arrays. Affymetrix exon array data were normalized using quantile normalization. The background correction was performed using antigenomic probes. Only probes targeting exons annotated from full-length cDNA sequences were used for analysis. Among these preselected probes, poor quality probes (e.g., probes labeled as “cross-hybridizing” probes by Affymetrix), and probes with intensity signals that were too low when compared to antigenomic background probes with the same GC content, were removed from the analysis. Only probes with a DABG p-value  $\leq 0.05$  in at least half of the chips were considered for further statistical analysis. Paired statistical analyses were performed using the Student's paired t-test on the splicing index to analyze the exon array data. The splicing index (SI) corresponds to a comparison of gene-normalized exon intensity values between the two analyzed experimental conditions. Data will be deposited in the Gene Expression Omnibus upon acceptance of the manuscript, but are available on request.

# Figure S1

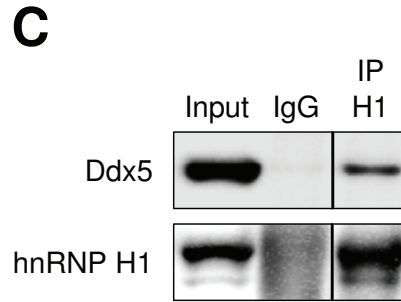
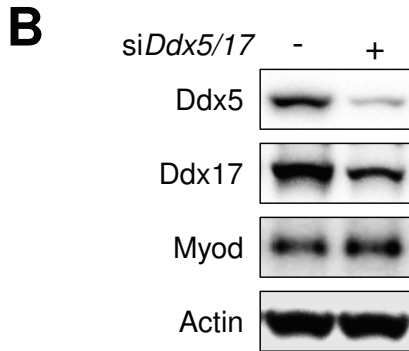
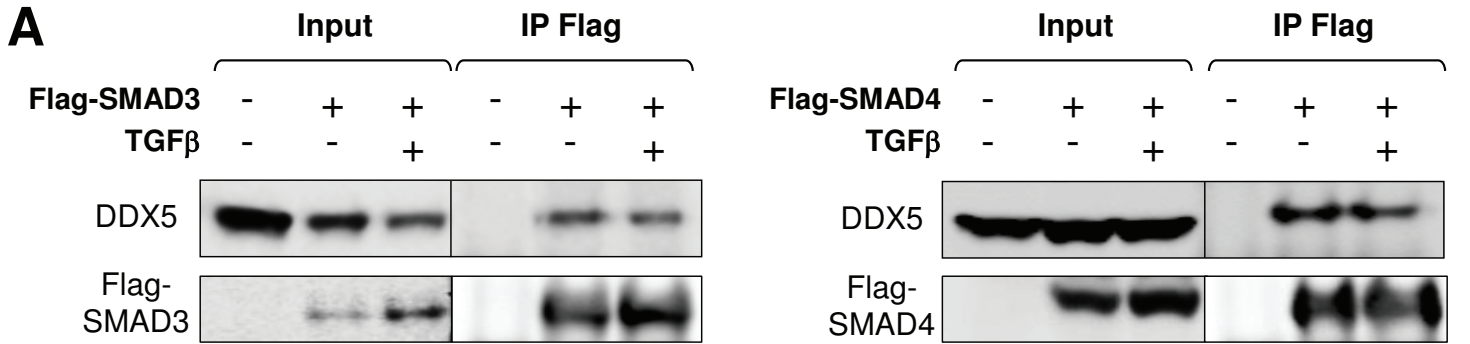


# Figure S2

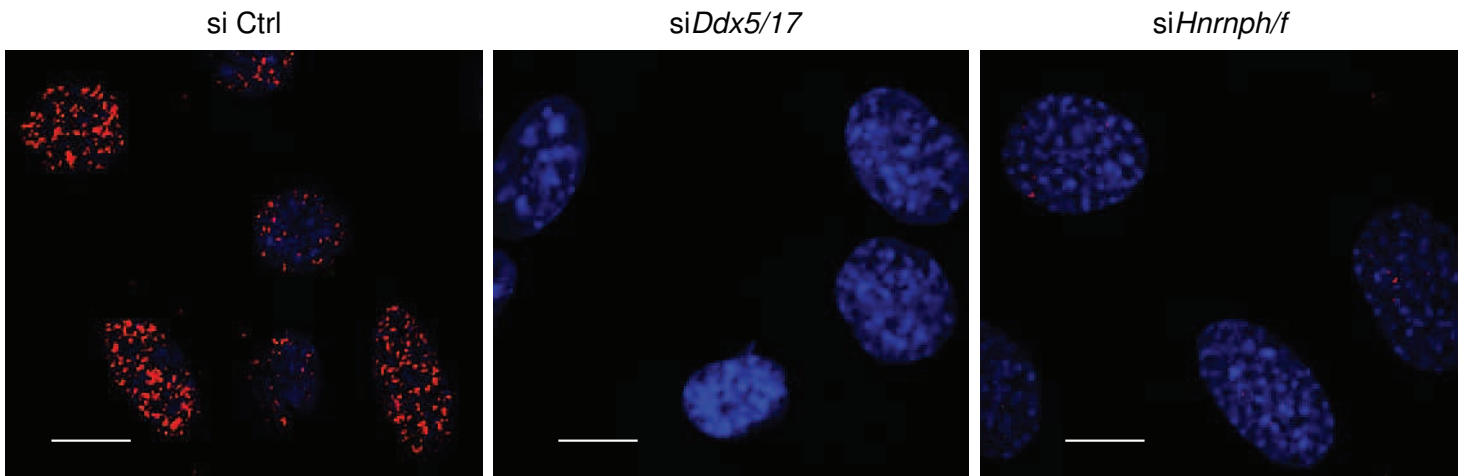
**MCF10a cells**



# Figure S3



**D** **C2C12** Proximity ligation assay (Ddx5 - hnRNP H1)





# Figure S5

**A**

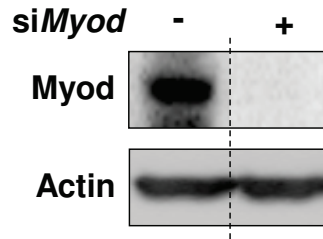
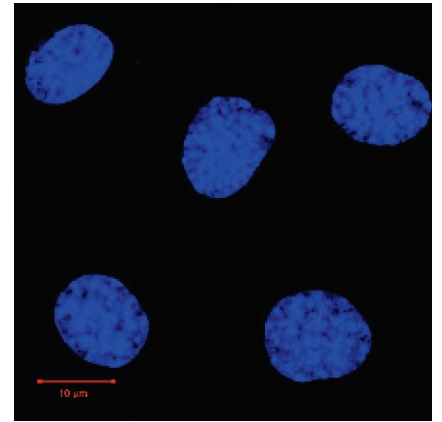
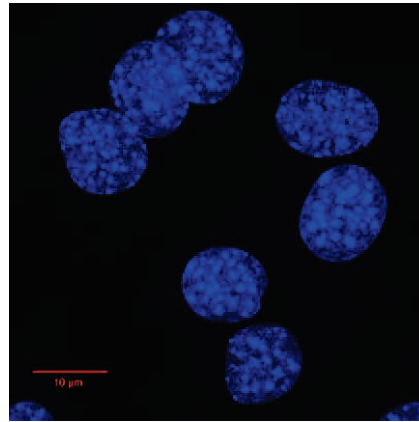
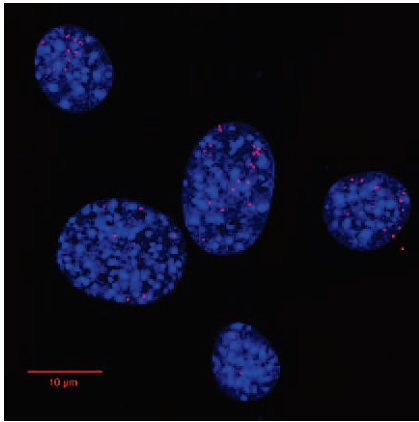
**C2C12 cells**

Proximity Ligation Assay Ddx5-Myod

si ctrl

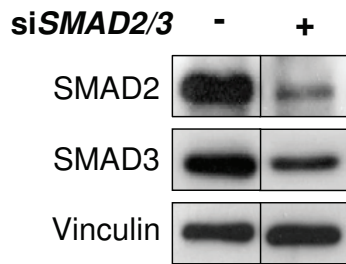
siMyod

siDdx5/17

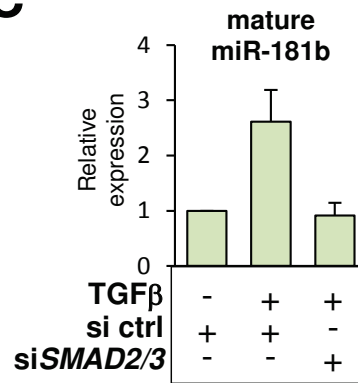


**B**

**MCF10A cells**



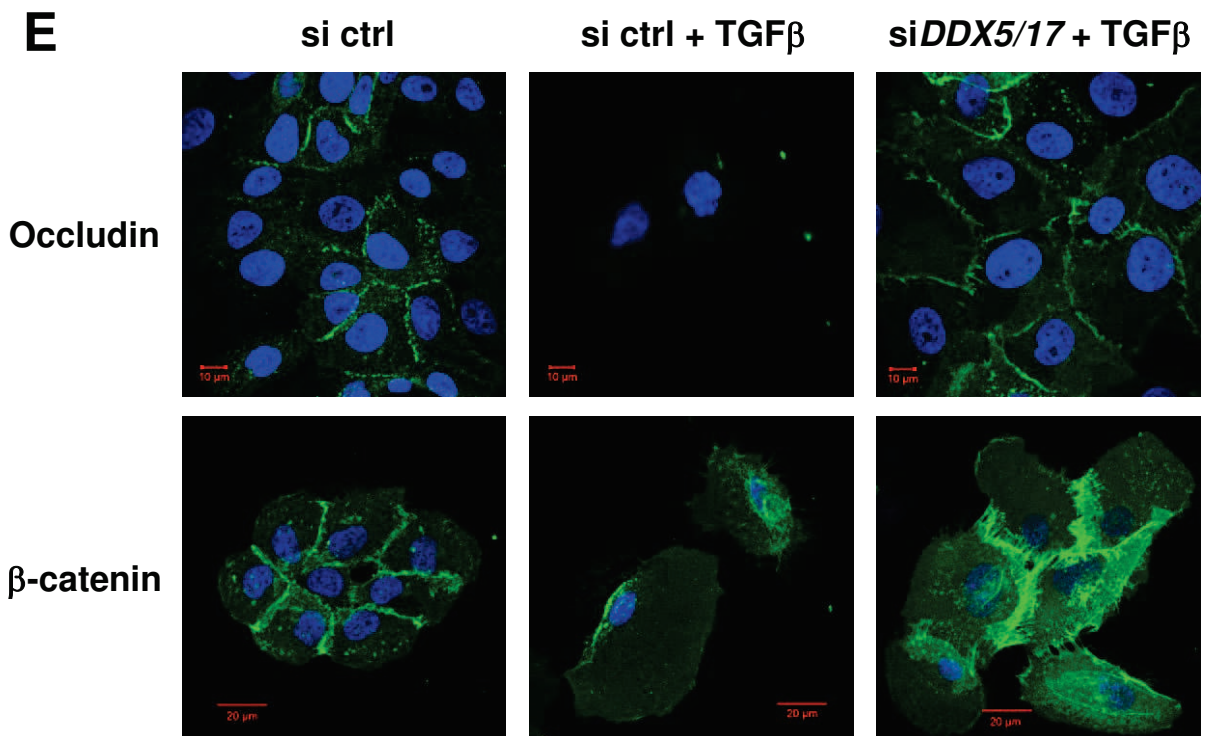
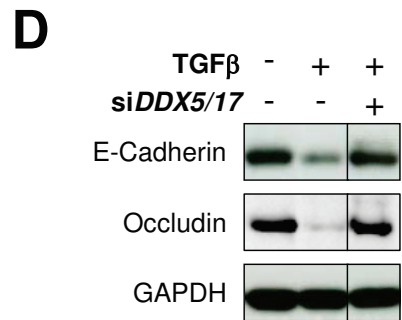
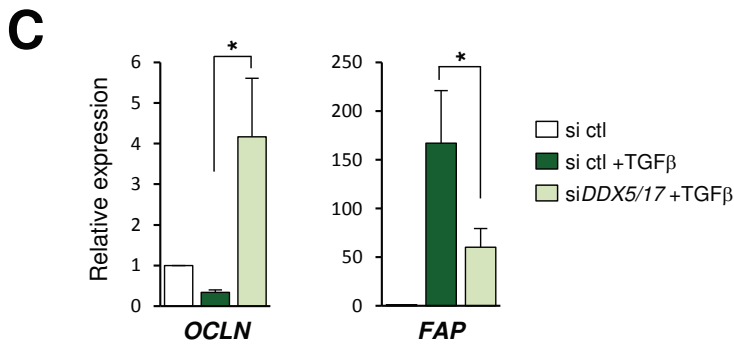
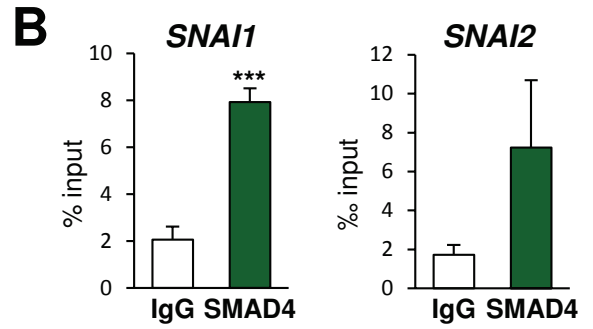
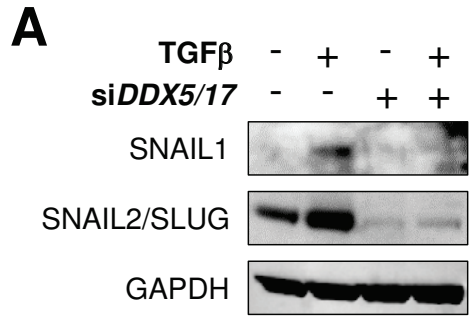
**C**





# Figure S6

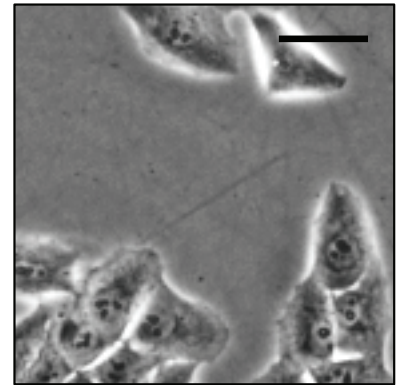
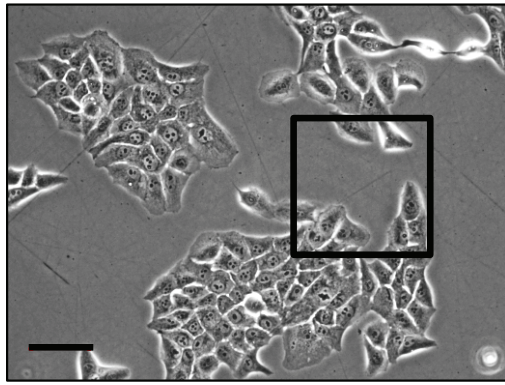
MCF10a cells



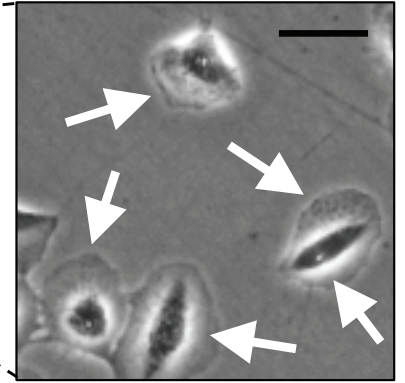
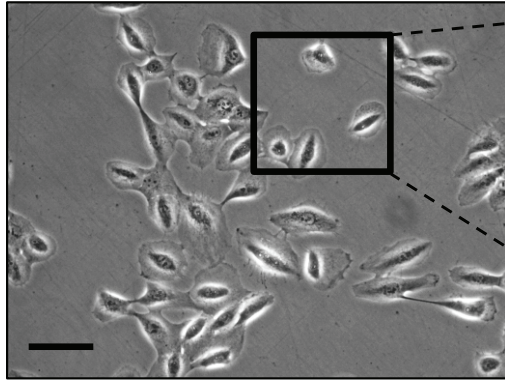
**Figure S7A**

**MCF10A**

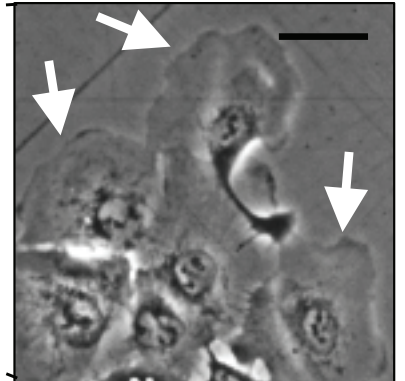
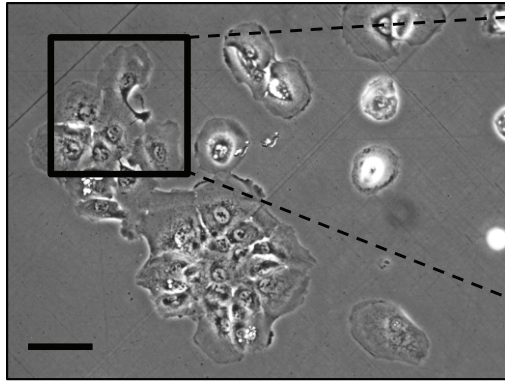
si ctrl



si ctrl  
+ TGF $\beta$

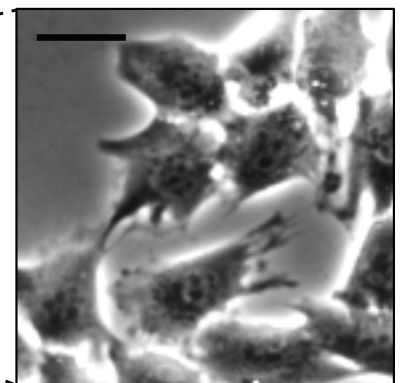
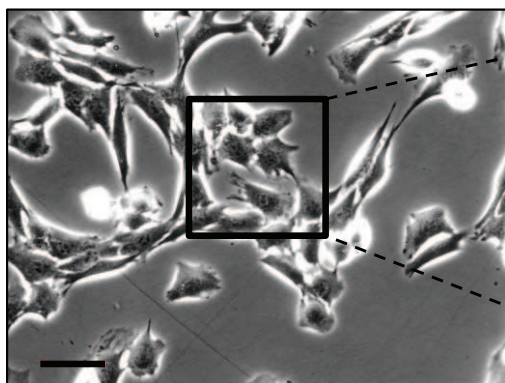


siDDX5/17

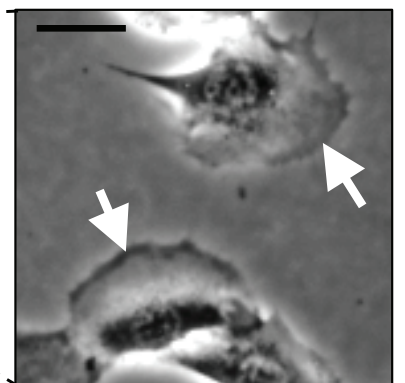
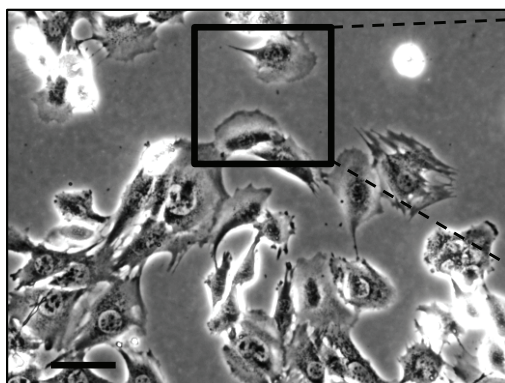


**C2C12**

si ctrl

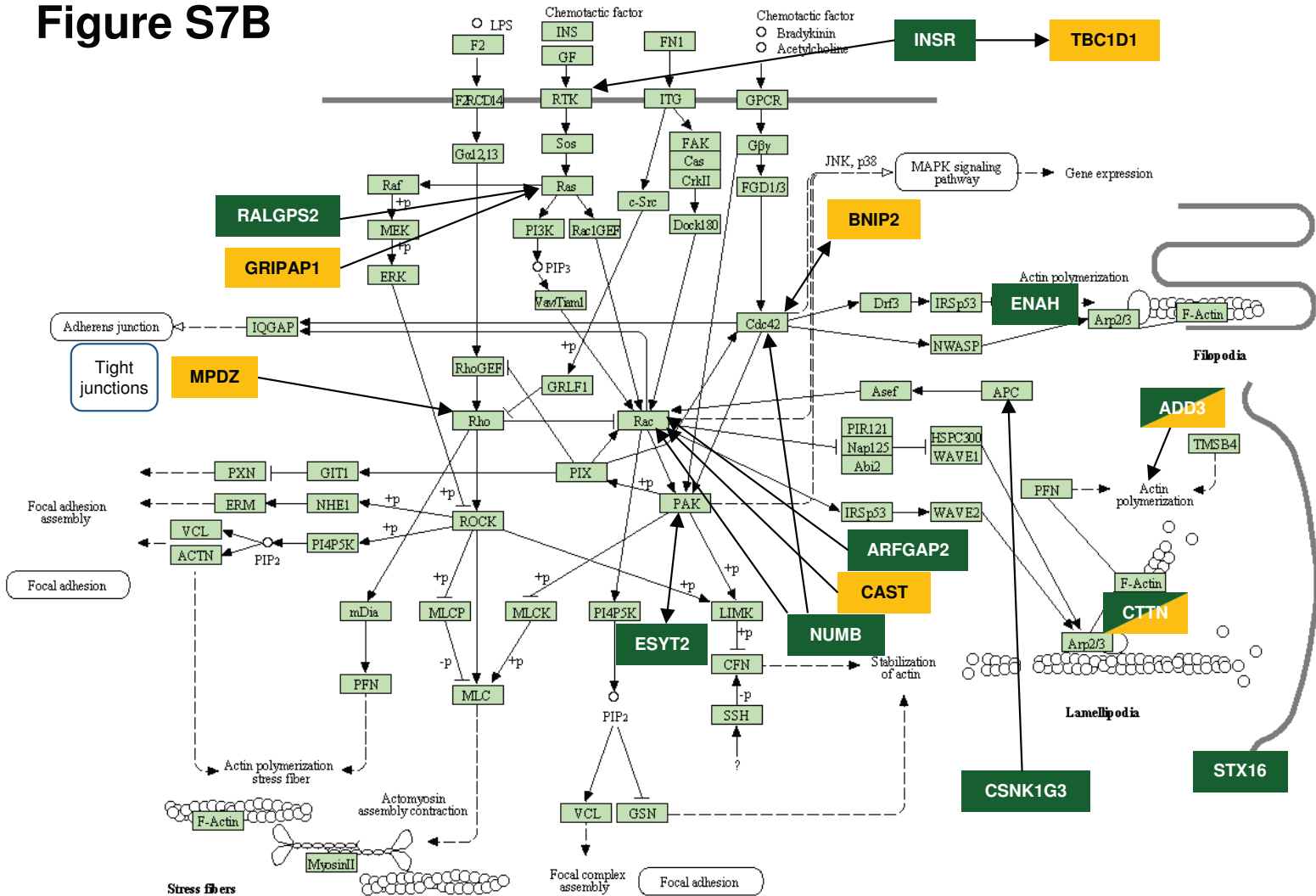


siDdx5/17





# Figure S7B



Protein	Description	Functions
<b>ADD3</b>	Adducin 3	Membrane-cytoskeleton-associated protein involved in the assembly of the spectrin-actin network ; crucial role in cell motility
<b>ARFGAP2</b>	GTPase activating protein for ARF1	ARF1 controls Rac1 signaling to regulate cell migration
<b>BNIP2</b>	BCL2/adenovirus E1B 19kDa interacting protein 2	Involved in cell elongation ; interacts with CDC42, GEFs and GAPs
<b>CAST</b>	Calpastatin, calpain inhibitor	Role in myoblast fusion ; calpains are involved in cell spreading, lamellipodia formatio (Rac1 signaling )
<b>CSNK1G3</b>	Casein kinase 1, gamma 3	Casein Kinase 1 proteins phosphorylate APC and β-catenin
<b>CTTN</b>	Cortactin	Contributes to the organization of the actin cytoskeleton and cell structure ; plays a role in the regulation of cell migration and in the invasiveness of cancer cells
<b>ESYT2</b>	Extended Synaptotagmin-2	Regulator of endocytosis ; Esy2 binding to PAK1 suppresses actin polymerization and inhibits the activation of PAK1 by the GTPases Cdc42 and Rac
<b>ENAH</b>	Enabled Homolog (Drosophila)	Actin-associated protein involved in several processes dependent on cytoskeleton remodeling and cell polarity such as lamellipodial and filopodial dynamics in migrating cells
<b>GRIPAP1</b>	Ras-GTPase guanine exchange factor (RasGEF)	Modulates Ras-GTPase activity
<b>INSR</b>	Insulin receptor	Receptor tyrosine kinase that activates several signaling pathways
<b>NUMB</b>	Numb	Regulator of actin-dependent asymmetric division, cell-cell adhesion and polarity, lamellipodia formation ; functions in a CDC42- and Rac1-dependent manner
<b>MPDZ</b>	Multiple PDZ domain protein	Localized in tight junctions ; interacts with RhoA-GEF
<b>RALGPS2</b>	Ral GEF with PH domain and SH3 binding motif 2	Modulates Ral GTPase function. RalA is an effector of Ras signaling and a mediator or membrane trafficking
<b>STX16</b>	Syntaxin 16	Membrane trafficking
<b>TBC1D1</b>	Rab GTPase-activating protein (Rab-GAP)	Involved in the trafficking and translocation of GLUT4-containing vesicles and insulin-stimulated glucose uptake

## II Discussion

Les hélicases ARN homologues à boîte DEAD DDX5 et DDX17 sont des protéines « multi-tâches », elles sont impliquées dans de nombreuses étapes de la régulation du métabolisme des ARNs dont la transcription, l'épissage et la dégradation des ARNs. Néanmoins ces différentes fonctions moléculaires ont jusqu'à maintenant été décrites dans des contextes cellulaires différents.

La différenciation cellulaire est un processus biologique durant lequel le transcriptome est sujet à de multiples niveaux de régulation. De plus en plus d'évidences montrent que tous ces niveaux sont liés et coordonnés entre eux. Néanmoins le ou les éléments coordonnateurs sont encore manquants dans la plupart des modèles de différenciation.

C'est dans ce contexte que nous montrons que DDX5 et DDX17 coordonnent de manière dynamique et directe la transcription, l'épissage et la dégradation des ARNs lors de la myogenèse et de la Transition Epithélio-Mésenchymateuse (TEM). Nous montrons dans un premier temps que ces deux protéines coopèrent avec le facteur d'épissage hnRNP H/F pour réguler l'épissage d'exons définis par un site 5' d'épissage faible situé au sein de régions formant des G-quadruplex. Cette coopération entre DDX5/DDX17 et hnRNP H/F participe à la mise en place du programme d'épissage des cellules non différenciées. Lors des processus de différenciation étudiés, on observe une baisse de l'expression de DDX5 et DDX17 et celle-ci contribue à la reprogrammation des programmes d'épissage cellulaires observée durant la différenciation. Nous montrons que la baisse d'expression de DDX5/DDX17 est médiée par des microARNs spécifiques de la différenciation dont l'expression est directement dépendante de DDX5/DDX17. Enfin, comme précédemment décrit, nous montrons que DDX5 et DDX17 sont des co-activateurs transcriptionnels de facteurs de transcription majeurs spécifiques des modèles de différenciation.

## A Mécanisme de régulation de l'épissage

Les hélicases ARN DDX5 et DDX17 ont jusqu'à maintenant été impliquées, à plusieurs reprises dans le contrôle de l'épissage alternatif. D'un point de vue mécanistique, leur activité hélicase apparaît toujours nécessaire à cette fonction et semble systématiquement impliquer le remodelage de structures secondaires.

Nous proposons ici un modèle mettant en évidence l'importance des structures secondaires formées par l'ARN dans la régulation de l'épissage alternatif. En effet, même si de plus en plus d'études montrent l'impact de ce type de structure sur l'épissage, les mécanismes sous-jacents sont encore mal compris. Ici, nous proposons un modèle dans lequel DDX5 et DDX17 coopèrent avec le facteur d'épissage hnRNP H/F afin de contrôler la reconnaissance de sites 5' d'épissage suivis par des structures en G-quadruplex. Celles-ci constituent des structures secondaires stables nécessitant d'être déstabilisées pour permettre à des facteurs de se fixer au niveau de leurs séquences cibles. Ainsi DDX5 et DDX17 pourraient, de part leur activité hélicase, assurer cette fonction. D'autre part, notre modèle renforce aussi l'importance du contexte et de la cinétique dans la régulation de l'épissage. Ainsi dans certains cas, DDX5 et DDX17 permettraient le recrutement de hnRNP H/F en aval d'exons présentant un site 5' au sein d'une séquence riche en GC, favorisant la reconnaissance de l'exon et entraînant donc son inclusion. Dans d'autres cas, mais par le même type de mécanisme, DDX5 et DDX17 favoriseraient le recrutement de hnRNP H/F au niveau d'un site H/F structuré en G-quadruplex situé cette fois en amont d'un exon faible à la fois pour son site 3' et 5' et normalement peu inclus. Dans ce cas, la déplétion de DDX5/DDX17 et/ou hnRNP H/F augmenterait la fenêtre de temps durant laquelle des facteurs activateurs de l'épissage de ces exons faibles peuvent être recrutés, entraînant ainsi l'inclusion des exons.

Ainsi dans notre publication, nous montrons un mécanisme de régulation de l'épissage commun à deux modèles de différenciation : la coordination avec le facteur d'épissage hnRNP H/F. Néanmoins ce modèle ne permet d'expliquer la régulation que d'un sous-programme d'épissage commun aux deux modèles. Des données préliminaires bioinformatiques non montrées nous indiquent un net enrichissement en motifs MBNL1 dans l'environnement des exons régulés lors de la myogenèse et

impactés par la déplétion de DDX5 et DDX17. Une partie de ces prédictions a été validée par RT-PCR, montrant une coordination entre ces deux facteurs dans la régulation de l'épissage alternatif lors de la différenciation myoblastique.

D'autre part, la fonction de hnRNP H/F dans l'épissage de cellules musculaires a été démontrée (Chen et al., 1999; Paul et al., 2006). Il a également été montré que MBNL1 et hnRNP H/F font parties du même complexe ribonucléoprotéiques dans un contexte musculaire pathologique (Paul et al., 2011). Ceci suggère une coordination physique et fonctionnelle entre MBNL1 et hnRNP H dans l'épissage musculaire. Ces données ainsi que nos résultats nous permettent d'imaginer une coordination de l'action des trois facteurs DDX5/DDX17, hnRNP H/F et MBNL1 dans ce processus.

## B Remodelage du cytosquelette

D'un point de vue phénotypique, la déplétion de DDX5 et DDX17 s'accompagne d'un remodelage profond du cytosquelette cellulaire et, plus particulièrement, de la formation de lamellipodes. De manière intéressante, ces projections membranaires ont été impliquées dans l'acquisition de la mobilité cellulaire qui est une des fonctions principales des cellules mésenchymateuses ainsi que d'une importance toute particulière lors des étapes de migration et de fusion de la myogenèse (Rochlin et al., 2010). Une partie des événements d'épissage que nous présentons co-régulés par DDX5/DDX17 et hnRNP H/F sont régulés lors de la différenciation et impactent des gènes impliqués dans le remodelage du cytosquelette. Ceci se révèle vrai pour les exons communs aux deux modèles mais aussi pour les exons spécifiques aux deux systèmes de différenciation suggérant un rôle général de DDX5/DDX17 dans la régulation des gènes de remodelage du cytosquelette.

Un projet en cours au sein de l'équipe renforce cette idée. En effet, un très grand nombre de jeux de données de puce exon ont été produits et analysés dans l'équipe, en condition de déplétion de DDX5 et DDX17 et ce, dans une variété de conditions cellulaires. Un certain nombre d'exons apparaissent systématiquement régulés de la même manière par DDX5 et DDX17. De manière particulièrement intéressante,

la plupart de ces événements affectent des gènes impliqués dans la dynamique du cytosquelette d'actine (adhésion cellulaire, jonction cellulaire ou encore endocytose). D'autre part, l'impact de DDX5 et DDX17 sur le remodelage du cytosquelette n'est pas uniquement lié à son activité de régulation de l'épissage. En effet, une étude a montrée que DDX5/DDX17 régulent l'expression du microARN miR-182, qui contrôle à son tour l'expression de gènes impliqués dans le remodelage du cytosquelette tels que la cofiline ou la profiline. Il a été montré que ceci impacte le phénotype migratoire des cellules (Wang et al., 2012a). Ces résultats suggèrent que DDX5 et DDX17 contrôlent un programme d'expression génique responsable de la dynamique du cytosquelette cellulaire.

## C Perspectives

Si le modèle cellulaire des C2C12 est très communément utilisé pour l'étude de la myogenèse, son utilisation pose néanmoins certains problèmes. Dans un premier temps, la morphologie des C2C12 en prolifération n'est pas celle de cellules myoblastiques mais plutôt celle de fibroblastes. On peut se demander quels changements en termes de facteurs d'adhésion ont pu provoquer une telle transformation morphologique. D'autre part, l'immortalisation qu'elles ont subie pour échapper à la sénescence dérègle leur machinerie de cycle cellulaire, entraînant potentiellement une multitude d'autres dérèglements. Cette divergence avec des cellules satellites primaires a été étudiée et a montré que près de 25% des gènes dérégulés dans les C2C12 ne l'étaient pas dans les cellules primaires (Cornelison, 1998). Ce modèle d'étude bien que très pratique présente donc des limitations évidentes en termes de pertinence physiologique. Il apparaît donc nécessaire de valider notre modèle dans, par exemple, des cellules myoblastiques primaires humaines.

D'autre part, il serait intéressant de regarder l'expression de DDX5/DDX17 dans d'autres modèles de différenciation. Est-elle aussi régulée au cours de la différenciation ? Si c'est le cas, les mécanismes de régulation de l'expression contrôlés par DDX5 et DDX17 sont-ils les mêmes ? En ce qui concerne l'épissage, observe-t-

on aussi une coopération avec hnRNP H/F ? Des données préliminaires, notamment lors de la différenciation neuronale, montrent que le profil d'expression de DDX5 et DDX17 suit le même modèle que lors de la myogenèse. Il serait donc tout particulièrement intéressant de déterminer les facteurs majeurs de ce processus dont l'activité et/ou l'expression seraient contrôlés par DDX5 et DDX17. Ceci renforcerait le rôle de grand orchestrateur de ces deux hélicases lors de la différenciation cellulaire. DDX5 et DDX17 ne sont pas les seules hélicases ARN à avoir été impliquées dans de multiples étapes de la régulation des gènes, notamment lors de processus de différenciation (Abdelhaleem, 2005) telles que DDX3X, DDX4 ou DHX9 (Linder and Jankowsky, 2011). Par exemple, l'hélicase à boîte DEAD DDX4 (ou Vasa) est également très conservée dans le règne animal, elle a été impliquée dans de nombreux processus biologiques tels que la gamétogénèse, la spécification des cellules germinales ou la biologie des cellules souches (Lasko, 2013). D'un point de vue moléculaire, DDX4 a été impliquée dans la régulation traductionnelle, la biogénèse des petits ARNs et la condensation des chromosomes. Il serait donc intéressant de déterminer si d'autres hélicases peuvent également assurer un rôle coordinateur dans des processus biologiques différents. En effet, on sait maintenant que tous les niveaux de régulation de l'expression génique sont coordonnés, mettant en évidence l'importance d'une intégration finement régulée. On a donc toutes les raisons de penser que ce type de protéines est beaucoup plus fréquent que ce qui est admis aujourd'hui.

Dans ce contexte, il est important de souligner que l'émergence des technologies à haut débit (puces ADN, séquençage, CLIP-seq...) associées à des analyses bioinformatiques poussées ont permis de mettre en évidence le rôle global ainsi que de dégager les grandes lignes des « règles de régulation » de ce type de protéines. Néanmoins celles-ci font parties, la plupart du temps, de large complexes ribonucléoprotéiques, il est donc difficile faire leur étude dans ce contexte et la plupart des études se limitent alors à étudier leur interaction avec quelques partenaires particuliers. Enfin, il est particulièrement difficile, à partir de larges jeux de données, de distinguer ce qui résulte en particulier d'une fonction moléculaire donnée, ce qui est une conséquence directe d'une conséquence indirecte...

Enfin un dernier élément n'est pas abordé dans la publication soumise mais a été observé au cours du projet. Il s'agit de la très forte induction de la rétention de l'intron 11 de DDX5 au cours de la différenciation. Cet intron très conservé au cours de l'évolution a été récemment montré contenir un microARN fonctionnel dont les cibles ne sont pas encore connues (Moore et al., 2011). La très forte induction d'expression de cet intron soulève donc de nombreuses questions. Ce miR est-il exprimé et fonctionnel au cours de la différenciation? Participe-t-il à la mise en place des programmes d'expression génique de la différenciation? D'autre part, serait-il un régulateur de l'expression de DDX5? Ce résultat constitue une donnée particulièrement intéressante à développer dans le cadre d'un nouveau projet.

4

## Résultats complémentaires





## I Préambule

La Dystrophie Myotonique de type 1 (DM1) est caractérisée par une expansion anormale de triplets CTG dans l'extrémité 3'UTR du gène DMPK, entraînant la séquestration des transcrits mutants dans des foci nucléaires. Ceux-ci entraînent la séquestration du facteur d'épissage MBNL1 induisant de multiples défauts d'épissage.

Récemment une étude a montré que MBNL1 et DDX5 interagissent à la fois dans des myoblastes humains contrôles et DM1 (Paul et al., 2011). D'autre part, une autre étude récente a montré que DDX5 colocalise avec des répétitions CUG dans un modèle cellulaire mimant la DM1 (Laurent et al., 2012). DDX5 favorise le recrutement de MBNL1 au niveau des répétitions pathologiques ainsi qu'au niveau d'une structure en tige-boucle de l'ARN pré-messager Tnnt2 dont l'épissage est altéré dans la DM1. Des mutations au niveau du core hélicase de DDX5 inhibent à la fois le recrutement de MBNL1 et la colocalisation de DDX5 au niveau des répétitions CUG (Laurent et al., 2012).

Les auteurs proposent alors un modèle dans lequel la fonction de DDX5 serait de remodeler les structures secondaires de l'ARN afin de faciliter le recrutement de MBNL1 au niveau de ces séquences régulatrices. Ainsi dans un contexte sain, en coordination avec DDX5, MBNL1 pourrait assurer sa fonction de facteur d'épissage. Dans un contexte pathologique de gain de fonction de l'ARN comme la DM1, DDX5 permettrait le recrutement de MBNL1 au niveau des répétitions CUG entraînant sa séquestration et sa perte de fonction (Laurent, 2011, Figure 34).

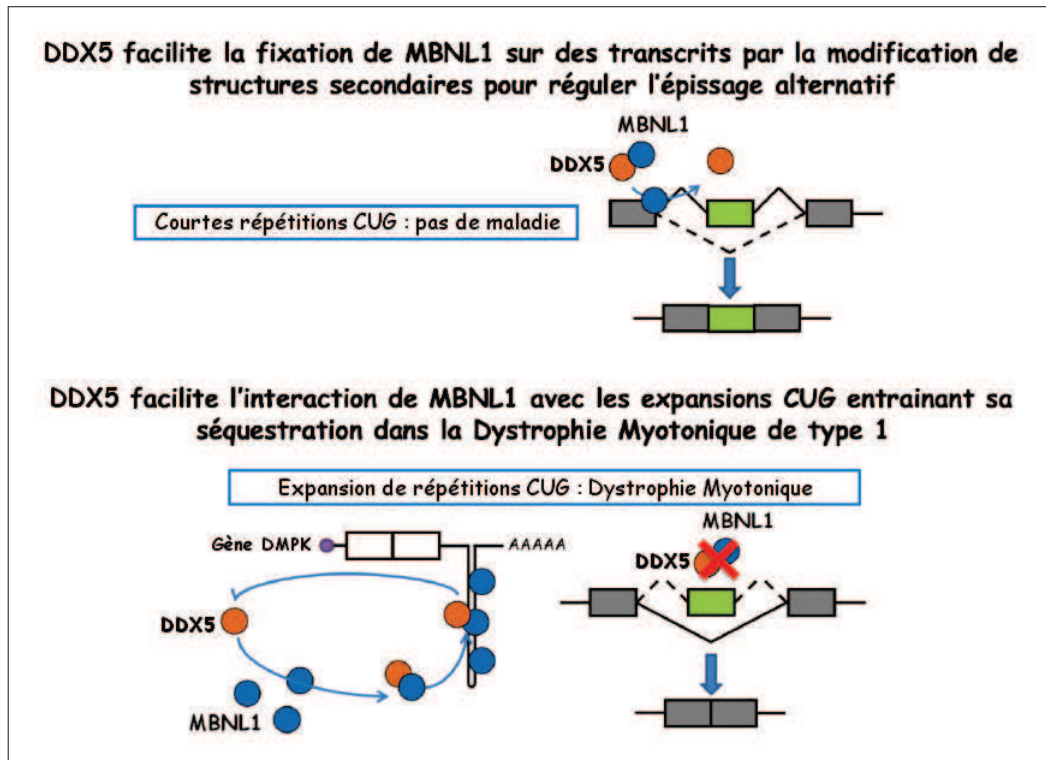


FIGURE 34: *Modèle proposé du rôle de DDX5 dans la DM1.* Dans les cellules saines, DDX5 forme un complexe avec MBNL1, ouvre localement la structure secondaire de l'ARN afin de faciliter son interaction avec ses sites régulateurs entrainant un épissage alternatif correct et la formation d'isoformes fonctionnelles. Dans un contexte DM1, le complexe DDX5/MBNL1 facilite la fixation de MBNL1 sur les paires de nucléotides G-C par l'ouverture locale des expansions CUG entrainant sa séquestration dans la pathologie. L'absence de complexe DDX5/MBNL1 au niveau des cibles d'épissage aboutit à une altération de l'épissage alternatif de nombreux transcrits avec une expression d'isoformes aberrantes à l'origine des symptômes cliniques (Adapté de [Laurent, 2011](#)).

## II Problématiques

DDX5 et DDX17, de par leurs nombreuses fonctions moléculaires, sont capables de réguler l'expression ainsi que l'épissage de nombreux gènes. Il est donc nécessaire dans un premier temps de déterminer l'impact de DDX5/DDX17 sur la fonction

et/ou l'expression de MBNL1. Ces deux protéines régulent-elles l'expression globale de MBNL1? MBNL1 est sujet à l'épissage alternatif, les différentes isoformes qui en résultent ne présentent pas les mêmes fonctions (Kino et al., 2004). DDX5/DDX17 régulent-elles ces épissages? Si oui, quel est l'impact fonctionnel pour MBNL1? C'est pourquoi déterminer précisément, dans un premier temps, l'impact de DDX5/DDX17 sur l'expression de MBNL1 apparaît particulièrement important dans un contexte biologique tel que la différenciation où les profils d'expression de DDX5 et DDX17 changent.

D'autre part, si l'interaction protéique entre DDX5 et MBNL1 a déjà été montrée, elle apparaît jusqu'à maintenant peu robuste (Paul et al., 2011; Laurent et al., 2012). Enfin, DDX5 et DDX17 ont été montrées nécessaires à la différenciation myogénique en tant que protéines co-activatrices de MyoD (Caretto et al., 2006 et notre manuscrit actuellement en révision dans *Molecular Cell*). D'autre part, nous montrons que, en coopérant avec le facteur d'épissage hnRNP H/F, DDX5 et DDX17 participent à la mise en place d'un programme d'épissage dans les cellules myoblastiques. Des défauts de différenciation sont observés dans la DM1 (Timchenko et al., 2001b; Timchenko et al., 2004). Dans ce contexte, nous avons cherché à déterminer quels étaient les événements d'épissage régulés au cours de la différenciation et qui sont dépendants de DDX5/DDX17.

### III DDX5/DDX17 et Mbnl1

#### A Régulation de l'expression de MBNL1 par DDX5/DDX17

Des expériences de déplétion de DDX5 et DDX17 montrent que ces deux protéines affectent l'expression globale de MBNL1 (Figure 35A). On observe une légère diminution des deux isoformes principales exprimées dans les C2C12 avec un ratio final au profit de la forme longue. Néanmoins il faut préciser que ce résultat n'est pas toujours reproductible et varie d'une transfection à l'autre.

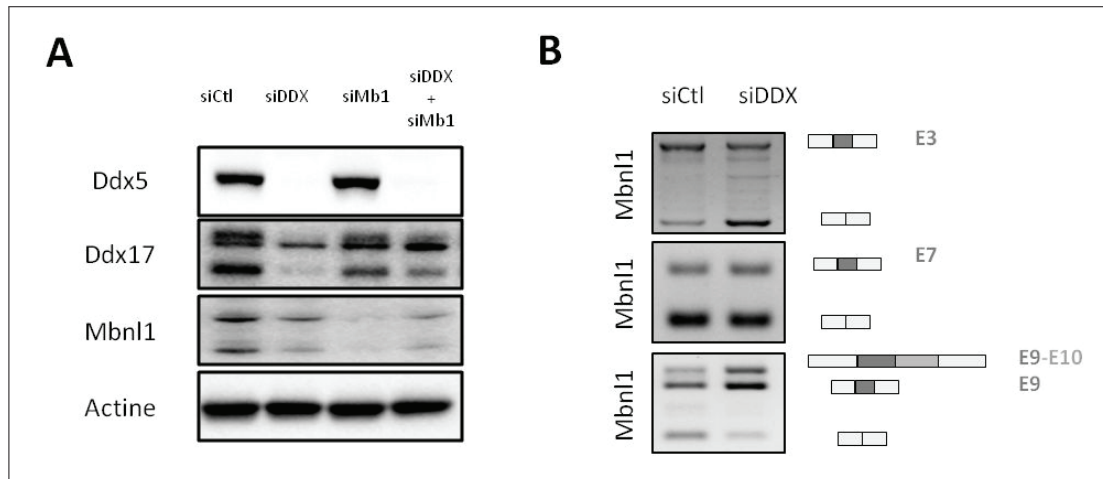


FIGURE 35: *DDX5/DDX17* régulent l'expression globale de *MBNL1* ainsi que son épissage alternatif. (A) La déplétion de *DDX5/DDX17* impacte l'expression protéique globale de *MBNL1* (Western Blot). (B) La déplétion de *DDX5/DDX17* impacte l'épissage alternatif des exons 3 et 9 de *MBNL1* (RT-PCR).

D'autre part, *MBNL1* est un gène sujet à l'épissage alternatif. En particulier, chez la souris, trois exons peuvent être alternativement épissés et impacter la fonction de ce facteur d'épissage : ce sont les exons 3, 7 et 9 chez la souris (Figure 35B).

Le domaine de liaison à l'ARN de *MBNL1* est constitué de 4 motifs à doigts de zinc (Znf) de type CCCH qui sont codés par les exons 1, 2 et 4. La région codée par l'exon 3 est localisée entre les motifs Znf2 et Znf3 et joue un rôle d'espaceur (ou « linker »). Sa présence favorise l'affinité de liaison de *MBNL1* aux ARN et est nécessaire à la fonction de facteur d'épissage de *MBNL1* (Tran et al., 2011). L'exon 7 code pour un signal de localisation nucléaire et l'exon 9 semble être impliqué dans l'oligomérisation de *MBNL1* (Tran et al., 2011). On observe des variations de l'inclusion de deux de ces exons lorsque l'on déplete les hélicases *DDX5* et *DDX17* (Figure 35B).

Ainsi l'exon 3 de *MBNL1* apparaît moins inclus en absence de *DDX5/DDX17*, suggérant une baisse d'affinité de *MBNL1* pour les motifs CUG et donc de son activité de facteur d'épissage en l'absence de *DDX5/DDX17*. Si l'exon 7 impliqué dans la localisation nucléaire de *MBNL1* n'est pas altéré, l'exon 9 est fortement inclus en l'absence de *DDX5* et *DDX17*. Ces résultats suggèrent qu'en l'absence de

DDX5 et DDX17, l'isoforme produite de MBNL1 a une moins bonne affinité pour ces ARNs cibles mais une meilleure capacité d'oligomérisation.

Le profil d'épissage de MBNL1 est également dérégulé dans la DM1 ([Dhaenens et al., 2008](#)). En effet, on observe une surexpression de l'isoforme foétale, incluant tous les exons alternatifs. Cette isoforme présente une meilleure affinité de liaison à ses motifs ARN, est localisée exclusivement dans le noyau et se dimérise plus facilement que les autres isoformes de MBNL1. Ainsi les défauts d'épissage de MBNL1 observés dans la DM1 pourraient contribuer à favoriser sa séquestration dans les foci nucléaires ([Tran et al., 2011](#)).

La déplétion de DDX5 et DDX17 reproduit en partie le profil d'expression pathologique de MBNL1 observée dans la DM1. On peut donc émettre l'hypothèse que la colocalisation de DDX5/DDX17 avec les foci DM1 pourrait avoir un impact à la fois sur les altérations d'expression et de fonction observées de MBNL1 dans la pathologie.

## B Localisation de MBNL1

L'exon portant le signal de localisation nucléaire de MBNL1 n'est pas altéré par un siARN ciblant DDX5 et DDX17. La localisation de ce facteur d'épissage ne devrait donc pas être affectée. Nous avons confirmé ce résultat par ImmunoFluorescence anti-MBNL1 dans les cellules C2C12 en présence et en absence de DDX5 et DDX17 ([Figure 36](#)).

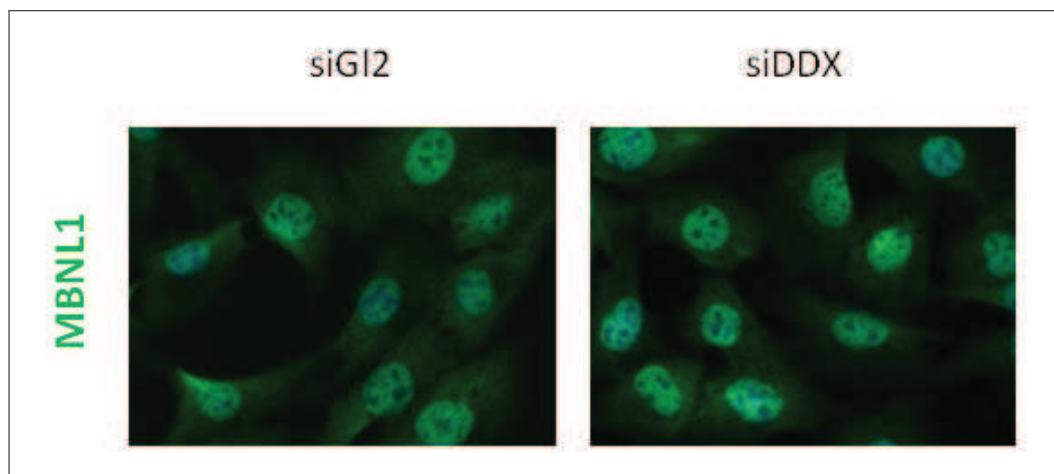


FIGURE 36: *La déplétion de DDX5/DDX17 n'impacte pas la localisation de MBNL1. Immunofluorescence MBNL1 en condition de déplétion de DDX5 et DDX17.*

Tout comme dans les cellules contrôles, MBNL1 est observée à la fois dans le cytoplasme et dans le noyau lorsque l'on déplete DDX5 et DDX17.

Il est à noter que l'on n'observe pas ou très peu de différence d'expression de MBNL1 dans ces conditions. Ce résultat peut en partie être expliqué par le fait que la diminution d'expression observée par western blot est faible et qu'elle traduit l'expression moyenne de la protéine au sein d'une population cellulaire, à la différence de l'immunofluorescence.

### C Interaction DDX5 entre MBNL1

L'interaction entre DDX5 et MBNL1 a été observée dans le cadre de deux précédentes études (Paul et al., 2011; Laurent et al., 2012). Néanmoins cette interaction apparaît être faible et variable en fonction de la technique utilisée. Cette faiblesse d'interaction peut, en partie, être expliquée par la qualité des anticorps anti-MBNL1. En effet, les anticorps disponibles sont peu réactifs et mal adaptés à des techniques comme la co-immunoprécipitation qui nécessite de grandes quantités de cellules et d'anticorps.

C'est pourquoi la technique du In situ PLA (Proximity Ligation Assay) apparaît tout

particulièrement adaptée à notre problématique. Cette technique permet de détecter une proximité entre molécules, et donc par « estimation », une interaction directe. Ainsi peu de matériel cellulaire est nécessaire pour révéler même des interactions faibles. On estime que le PLA génère un signal positif lorsque deux protéines se situent à moins de 40nm l'une de l'autre.

L'interaction entre DDX5 et MBNL1 a ainsi été confirmée par PLA : on obtient l'essentiel des signaux positifs dans le noyau mais on observe aussi quelques spots dans le cytoplasme. En revanche, on ne détecte plus de signal lorsque l'on traite les cellules avec un siARN ciblant DDX5 et DDX17 (Figure 37).

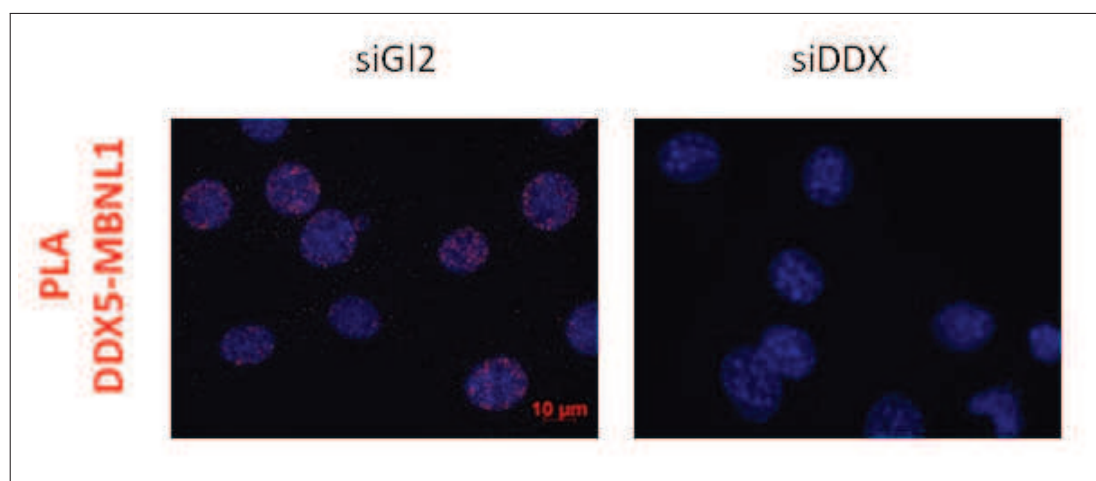


FIGURE 37: *Interaction de DDX5 et MBNL1 révélée par Proximity Ligation Assay (In situ PLA).*

Ces résultats montrent que DDX5 et MBNL1 interagissent à la fois dans le noyau et le cytoplasme de cellules myoblastiques.

## D Expression de MBNL1 durant la différenciation myogénique

L'expression de DDX5 diminue au cours de la différenciation myogénique. Nous proposons que cette baisse d'expression contribue à la reprogrammation du



programme d'épissage observée durant la myogénèse.

Dans la mesure où la déplétion de DDX5/DDX17 impacte l'expression de MBNL1 et dans l'hypothèse d'une coopération entre DDX5 et MBNL1, il était alors important de déterminer le profil d'expression de MBNL1 au cours de la différenciation.

A la différence de DDX5 et DDX17, on n'observe une diminution d'expression d'aucune des isoformes de MBNL1 (Figure 38).

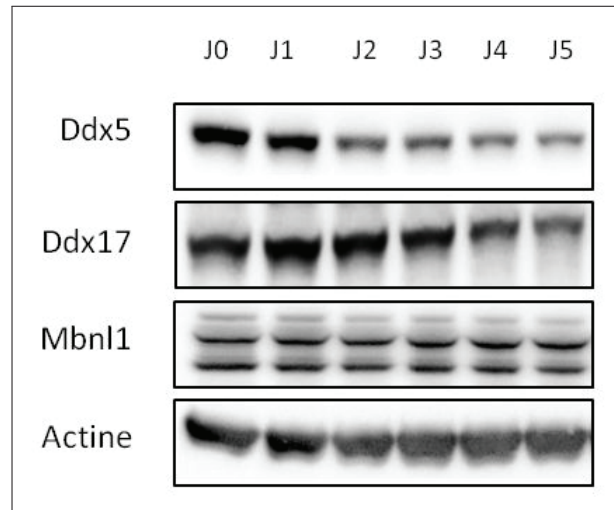


FIGURE 38: *L'expression protéique de DDX5 diminue au cours de la différenciation myogénique mais pas celle de MBNL1.*

La diminution d'expression de DDX5 au cours de la différenciation myogénique apparaît ne pas avoir d'impact sur celle de MBNL1.

## IV Rôle de DDX5/DDX17 dans les défauts d'épissage observés dans la DM1 ?

### A Démarche

Notre deuxième objectif était de déterminer quels sont les événements d'épissage régulés au cours de la différenciation qui sont dépendants de DDX5/DDX17. Parmi ceux-ci, nous voulions en particulier nous focaliser sur ceux qui sont importants

pour le phénotype musculaire et qui pourraient être dérégulés de la même manière chez les patients DM1, reflétant potentiellement un lien avec l'activité de DDX5 et DDX17.

Pour ce faire, nous avons donc utilisé de multiples jeux de données de puces « tout exon », produits au sein de l'équipe mais aussi publiquement disponibles. Dans la mesure où nous disposons de la lignée myoblastique murine C2C12, le jeu de données utilisé pour générer les « listes de référence » du projet a été obtenu au sein du laboratoire. Suite à la transfection de cellules C2C12 avec un siARN contrôle ou ciblant à la fois DDX5 et DDX17, les cellules ont été traitées ou non pour entrer en différenciation. On obtient alors des listes d'événements d'épissage régulés par la déplétion de DDX5 et DDX17 dans des cellules en prolifération (non différenciées), ainsi que des événements d'épissage régulés au cours de la différenciation et de façon DDX5/DDX17-dépendante.

Afin de se focaliser sur les événements d'épissage les plus susceptibles d'être impliqués dans le phénotype musculaire, nous avons ensuite uniquement sélectionné les événements qui s'avèrent enrichi dans le tissu musculaire par rapports aux autres tissus. Pour cela, nous avons utilisé des jeux de données publiques de puces comportant des dizaines de tissus dont le muscle, à la fois chez la souris et l'homme. Le fait de travailler sur deux espèces différentes a nécessité une étape d'annotation des exons conservés entre espèces. Celle-ci a été réalisée à partir des gènes orthologues tels que définis par Ensembl puis les alignements entre exons ont été réalisés à l'aide du logiciel Mégablast. N'ont finalement été conservés que les exons orthologues uniques entre souris et homme et présentant une identité de séquence supérieure à 80%.

Enfin, divers jeux de données de puces exon obtenus à partir de patients DM1 ont été analysés afin d'établir des listes à large échelle d'événements dérégulés dans la pathologie.

Nous avons donc cherché à déterminer les événements d'épissage, dont la régulation au cours de la différenciation est dépendante de DDX5/DDX17, et qui pourraient refléter une altération du processus de différenciation des cellules musculaires chez les patients DM1. Le modèle proposé dans notre publication présente DDX5 et

DDX17 comme nécessaires, dans un premier temps, à l'épissage dans les cellules myoblastiques puis, dans un deuxième temps, leur baisse d'expression contribuerait à la mise en place du programme d'épissage de cellules musculaires. C'est pourquoi nous avons défini deux groupes d'exons régulés au cours de la différenciation et dépendants de DDX5/DDX17.

La validation de nos prédictions a été réalisée dans des échantillons C2C12 (+/- différenciation, +/- DDX5/DDX17) ainsi que dans des échantillons de lignées DM1 immortalisées (+/- différenciation) amicalement fournies par Denis Furling de l'Institut de Myologie à Paris.

## **B 1ère catégorie d'exons**

Nous avons validé une série d'événements d'épissage régulés au cours de la différenciation et inversement régulés en l'absence de DDX5/DDX17 et dans des échantillons DM1 ([Tableau 6](#), [Figure 39](#)).

Ont été sélectionnés les événements régulés durant la différenciation dans au moins l'un des deux modèles (myoblastes C2C12 et myoblastes humains) mais dérégulés à la fois par la déplétion de DDX5/DDX17 et dans la DM1.

IV. RÔLE DE DDX5/DDX17 DANS LES DÉFAUTS D'ÉPISSAGE OBSERVÉS  
DANS LA DM1 ? 4. RÉSULTATS COMPLÉMENTAIRES

Protéine	Description	Fonction
<i>ABLIM1</i>	Actin Binding LIM Protein 1	Protéine cytosquelettique LIM, joue un rôle majeur dans la régulation des voies de développement cellulaire. Joue potentiellement un rôle dans le développement de la rétine.
<i>BIN1</i>	Bridging Integrator 1	Protéine impliquée dans les invaginations tubulaires des membranes, est requise pour la biogénèse des tubules musculaires T.
<i>FEZ2</i>	Fasciculation And Elongation Protein Zeta 2	Protéine nécessaire à l'empaquetage et l'élongation des faisceaux d'axones.
<i>ITGB1</i>	Intégrine, Beta 1	Protéine participant à l'innervation et la vascularisation du muscle squelettique.
<i>MSI2</i>	Musashi RNA-Binding Protein 2	Protéine qui se lie à l'ARN régulant l'expression d'une série d'ARNm cibles au niveau traductionnelle. Joue potentiellement un rôle dans la prolifération et la maintenance des cellules souches dans le système nerveux central.
<i>MYOM1</i>	Myoméline	Protéine musculaire se liant à la myosine et la titine, constitue l'un des composants majeurs des fibres musculaires formant les bandes M, impliquée dans l'élasticité des bandes M.
<i>NUMB</i>	Numb Homolog (Drosophila)	Protéine régulatrice de la division asymétrique, de l'adhésion cellulaire et de la polarité. Joue un rôle dans la neurogénèse.
<i>PPP2R5C</i>	Protéine Phosphatase 2, sous-unité B', Gamma	Une des quatre Ser/Thr phosphatases majeures, impliquée dans le contrôle négatif de la croissance cellulaire et la division.

TABLE 6: *Gènes dont l'épissage est régulé au cours de la différenciation et inversement régulés lors de la déplétion de DDX5/DDX17 et dans la DM1.*

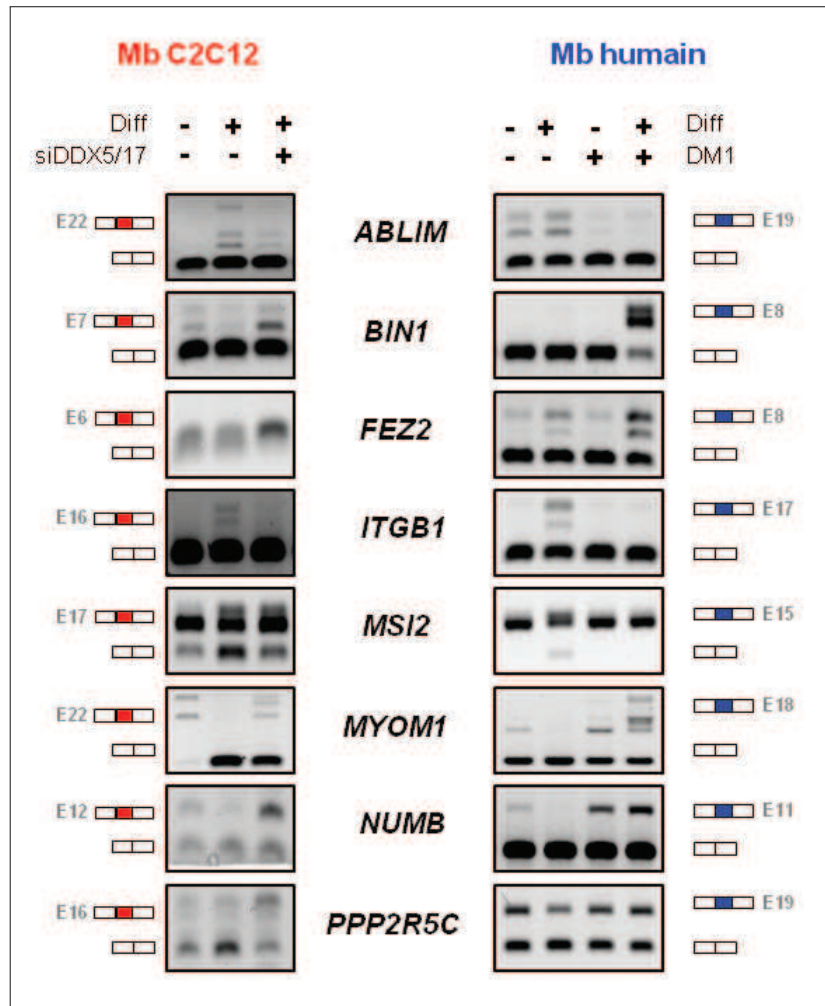


FIGURE 39: *Événements d'épissage régulés au cours de la différenciation et inversement régulés lors de la déplétion de DDX5/DDX17 et dans la DM1. Mb : Myoblastes.*

Parmi ces évènements, certains gènes sont particulièrement intéressants de par leur fonction :

- **ABLIM1** : Ce gène code pour une protéine LIM cytosquelettique, capable de lier les filaments d'actine. Les protéines LIM sont connues comme des régulateurs du développement embryonnaire (Hobert and Westphal, 2000). En particulier, ABLIM1 a été impliqué dans le développement de la rétine (Roof et al., 1997). Les patients atteints de DM1 présentent souvent des cataractes ainsi qu'une dégénérescence de la rétine. Ce gène apparaît donc fonctionnellement intéressant.
- **BIN1** : Ce gène code pour une protéine nécessaire à la biogénèse des tubules

T musculaires, nécessaires au couplage excitation-contraction dans les cellules musculaires. En particulier, un variant d'épissage (autre que celui dépendant de DDX5/DDX17) a été récemment identifié comme participant activement au phénotype de la DM1 (Fugier et al., 2011). Les patients présentent un saut de l'exon 11 qui est responsable d'une forme inactive de la protéine, directement associée à une faiblesse musculaire, qui constitue le principal symptôme de la DM1.

- **ITGB1** : Ce gène code pour l'intégrine *beta-1* qui est impliquée dans la vascularisation et l'innervation du muscle squelettique (Schwander et al., 2004). De plus, le variant d'épissage contenant l'exon régulé par DDX5/DDX17 a été identifié comme spécifique du muscle (Belkin et al., 1996). Cet exon, lorsqu'il est inclus, confère à la protéine des propriétés d'attache à la matrice cytosquelettique, capacité nécessaire pour assurer la fonction de contraction des cellules musculaire.
- **MYOM1** : Ce gène code pour une protéine structurale des bandes M des fibres musculaires impliquées dans l'élasticité musculaire (Speel et al., 1998 ; Pinotsis et al., 2012). Ce variant a récemment été montré comme étant dérégulé dans la DM1 de manière dépendante à MBNL1 (Koebis et al., 2011).
- **NUMB** : Ce gène est impliqué dans le contrôle de la division asymétrique et dans la détermination cellulaire, en particulier neuronale (Casanova, 2007). L'exon régulé par DDX5/DDX17 est spécifique des cellules en prolifération non différenciées, le saut d'exon étant nécessaire à l'entrée en différenciation (Yan, 2010 ; Bani-Yaghoub et al., 2007). Les patients atteints de DM1 présentent des troubles neurologiques importants et de plus en plus d'études montrent une dérégulation importante des programmes d'expression dans le cerveau des patients DM1.

## C 2ème catégorie d'exons

Nous avons montré que l'expression de DDX5/DDX17 diminue au cours de la différenciation et que cette diminution contribue probablement aux changements dans le programme d'épissage exprimé par les cellules au cours de la myogenèse.

IV. RÔLE DE DDX5/DDX17 DANS LES DÉFAUTS D'ÉPISSAGE OBSERVÉS  
DANS LA DM1 ? 4. RÉSULTATS COMPLÉMENTAIRES

Nous avons donc cherché à identifier les événements d'épissage régulés de la même façon au cours de la différenciation et suite à l'inactivation de l'expression de DDX5/DDX17 dans des cellules non différenciées. Ce type d'évènement d'épissage correspondrait, dans notre modèle, à des événements survenant plus tardivement dans la différenciation et résultant de la baisse d'expression de DDX5/DDX17 induite par les miRNAs miR-1/206.

Ont été sélectionnés les événements régulés durant la différenciation dans au moins l'un des deux modèles (myoblastes C2C12 et myoblastes humains) mais régulés par la déplétion de DDX5/DDX17 et inversement dans la DM1 (Tableau 7, Figure 40).

Protéine	Description	Fonction
<i>EPB41L3</i>	Erythrocyte membrane protein band 4.1-like 3	Protéine régulatrice de la croissance cellulaire dans la péthogénèse des méningiomes.
<i>KIAA1797</i>	Focadhésine	Potentiel suppresseur de tumeurs dans les gliomes.
<i>KIF1B</i>	Kinesin family member 1B	Protéine moteur pour le transport antérograde de la mitochondrie et des vésicules synaptiques. Sa fonction a été essentiellement étudiée dans les cellules neuronales.
<i>MKMK1</i>	MAP kinase interacting serine/threonine kinase 1	Protéine jouant potentiellement un rôle dans la réponse au stress environnemental et les cytokines.
<i>PPP3CA</i>	Protéine phosphatase 3, catalytic subunit, alpha isozyme	Sous-unité catalytique de la calcineurine, protéine régulatrice de la différenciation myogénique (par activation de l'expression de MyoG et MEF2A). La calcineurine est également un senseur du calcium intra-cellulaire.
<i>TJP1</i>	Tight junction protein 1	Protéine impliquée dans la transduction du signal nécessaire à l'assemblage des jonctions serrées. Joue un rôle dans la migration cellulaire.
<i>TRIM55</i>	Tripartite motif containing 55	Protéine jouant potentiellement un rôle dans l'expression génique et le taux de renouvellement protéique dans les cellules musculaires.

TABLE 7: *Gènes dont l'épissage est régulé au cours de la différenciation ainsi que lors de la déplétion de DDX5/DDX17 et inversement régulés dans la DM1.*

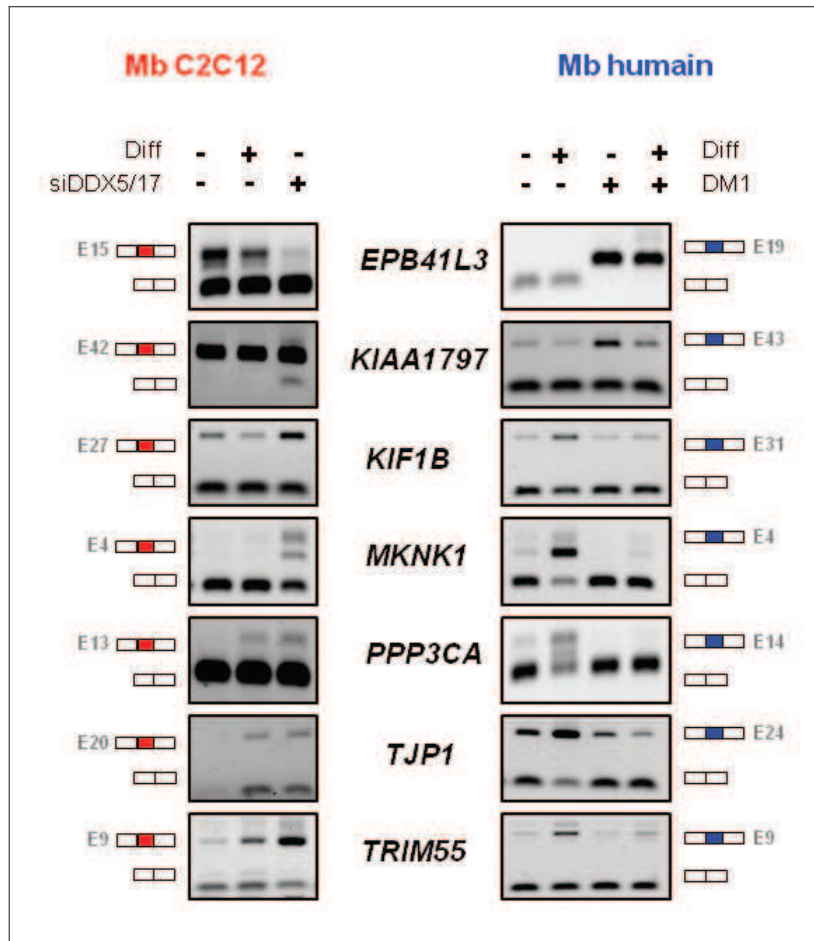


FIGURE 40: *Événements d'épissage régulés au cours de la différenciation, de la même manière lors de la déplétion de DDX5/DDX17 et inversement dans la DM1. Mb : Myoblastes.*

## D Perspectives

Ces résultats sont préliminaires et certains aspects doivent, en particulier, être abordés dans la suite du projet.

Dans l'hypothèse d'une coopération entre DDX5 et MBNL1, la prochaine étape est de déterminer si les événements d'épissage DDX5/DDX17-dépendants sélectionnés précédemment sont régulés de la même manière par MBNL1. Des résultats préliminaires, dans notre modèle cellulaire, nous indiquent néanmoins que peu de ces événements apparaissent également dépendants de MBNL1. Ces résultats peuvent avoir plusieurs explications. D'une part, notre démarche n'a abouti qu'à un faible



nombre d'événements à tester. Il n'est donc pas possible de conclure quand à un mécanisme général de régulation. D'autre part, il existe une redondance fonctionnelle entre protéines MBNL, en particulier entre MBNL1 et MBNL2, qui présentent le même profil d'expression (Carpentier et al., 2014). Il est donc nécessaire d'effectuer une double déplétion à la fois de MBNL1 et MBNL2 afin d'observer les effets sur l'épissage des exons candidats.

Le faible nombre d'événements sélectionnés par notre démarche nous montre que celle-ci est trop stringente. De nombreux événements potentiellement intéressants ont été éliminés lors des différentes étapes de sélection. Un exemple de ce type d'événements d'épissage perdu est le « switch » d'épissage de la PKM (Pyruvate Kinase Muscle), enzyme du métabolisme du glucose. En l'absence de DDX5 et DDX17, l'isoforme fœtale PKM2 est réexprimée. Très récemment une étude a identifié cet événement d'épissage comme dérégulé dans la DM1 entraînant des défauts dans le métabolisme énergétique (Gao and Cooper, 2013b).

D'autre part, on sait que le modèle cellulaire C2C12 ne reproduit pas à l'identique la myogenèse se produisant dans les myoblastes primaires humains (Cornelison, 1998). L'une des difficultés observées lors de l'étape de validation a été le fait que certains événements régulés au cours de la myogenèse dans les C2C12 ne l'étaient pas dans celle des myoblastes humains (et vice-versa). Ce constat nous a d'ailleurs amené à conserver les événements régulés durant la différenciation à la fois de l'un ou l'autre modèle.

Il est donc nécessaire de faire des efforts sur notre « workflow » d'analyse de puces « tout exon » afin d'enrichir au maximum nos listes d'exon candidats, tout en gardant conscience de la limitation de certains de nos jeux de données. Ainsi l'objectif serait d'avoir le maximum d'événements fiables à tester dans un premier temps dans notre modèle cellulaire (dans lequel il nous est possible de tester la déplétion d'une série de facteurs d'épissage tels que MBNL1 +/- MBNL2, hnRNP H...) puis dans un deuxième temps, dans un modèle DM1 de différenciation.

Les candidats validés dans ces différentes conditions, apparaissant particulièrement intéressants fonctionnellement, pourront faire l'objet d'une étude fonctionnelle préliminaire dans notre laboratoire, à l'aide d'approches visant à comprendre

la fonction des exons impliqués dans les produits protéiques correspondant. Par exemple, l'approche oligoswitch nous permet de forcer l'expression d'un variant en particulier. A la différence des siARNs qui peuvent entraîner un déficit d'expression, ces oligonucléotides antisens ciblent les jonctions intron-exon ou les séquences cis-régulatrices afin de forcer l'épissage d'un pré-ARNm.

Un autre aspect à exploiter serait de déterminer le profil d'expression chez les patients DM1. En effet, le facteur d'épissage CUGBP1 est hyperphosphorylé chez les patients DM1, rendant ce facteur plus stable et entraînant une augmentation de son expression. Cette hyperphosphorylation est médiée par PKC (Protéine Kinase C) par un mécanisme encore inconnu (Wang et al., 2007). On sait que DDX5 est aussi une cible de PKC et qu'elle peut être phosphorylée en S557 (Buelt et al., 1994). Si l'impact de cette modification post-traductionnelle n'est pas encore connue, il apparaît intéressant de déterminer si elle survient chez les patients DM1 et si elle a un impact sur l'expression et/ou l'activité de DDX5. Si c'est le cas, l'impact sur les programmes d'expression géniques peut être très important et apporte d'autant plus de relevance à l'étude de ces deux protéines dans le contexte de la DM1.

De manière générale, il est particulièrement important de déterminer l'expression de DDX5 chez les patients DM1 et d'autant plus durant la myogenèse. Ainsi si le profil d'expression n'est pas le même que lors de la myogenèse normale, ceci peut contribuer aux défauts de différenciation observés dans la DM1.

Enfin, nous nous sommes focalisés sur l'impact de DDX5 et DDX17 sur l'épissage dans la pathologie de la DM1. On sait aujourd'hui que celle-ci présente également une dérégulation massive de l'expression des micro-ARNs (Kalsotra et al., 2014). Les protéines DDX5 et DDX17 ont été impliqués dans la biogénèse des miARNs, il serait donc intéressant d'étudier leur rôle dans l'altération de la biogénèse des miARNs observée dans la DM1. En particulier, une étude récente a montré que le processing de miR-1 est dérégulé chez les patients DM1 (Rau et al., 2011). Dans un contexte sain, MBNL1 se fixe sur la boucle que forme le pré-miARN empêchant un facteur inhibiteur de se fixer. Dans la DM1, suite à sa perte de fonction, MBNL1 n'est plus en mesure d'assurer ce rôle entraînant une chute de l'expression du miR-1 mature. Même si l'on ne montre, dans notre modèle, qu'un rôle strictement transcriptionnel

*IV. RÔLE DE DDX5/DDX17 DANS LES DÉFAUTS D'ÉPISSAGE OBSERVÉS  
DANS LA DM1 ?* *4. RÉSULTATS COMPLÉMENTAIRES*

---

de DDX5/DDX17 sur l'expression de miR-1, on ne peut pas exclure un rôle dans la maturation d'autres miARNs.

5

## Annexes



## I Préambule

Comme précédemment décrit dans l'introduction, le développement des technologies à haut débit a permis, de manière fiable, l'analyse à large échelle de l'épissage alternatif. Parmi les différentes méthodes qui sont couramment utilisées, les puces à ADN « tout exon » apparaissent être une méthode performante et à moindre coût (Xu et al., 2011 ; Raghavachari et al., 2012).

Mon premier objectif dans l'équipe en tant que doctorante en bioinformatique a été de développer un système d'analyse dédié aux puces « tout exon » de la technologie Affymetrix. En particulier, ce système devait s'inscrire dans la continuité du projet « Fasterdb » précédemment réalisé dans l'équipe. Celui-ci a constitué en une réannotation de l'intégralité des transcriptomes codants de l'homme et la souris pour en analyser, en particulier, les événements d'épissage.

Par ce travail, j'ai été amené à participer à divers projets dans l'équipe dont certains ont donné lieu à des publications.

Après une présentation de la spécificité des puces « tout exon », je présenterai le système d'analyse mis en place. Enfin, les publications sur lesquelles je suis citée sont jointes avec une description courte de ma contribution pour chacune d'elle.

## II Les puces ADN « tout exon » : généralités et méthodologie

### A Généralités

Les puces « tout exon » de la technologie Affymetrix sont dédiées à l'analyse du transcriptome humain (GeneChip Human Exon 1.0 ST Array) ou murin (GeneChip Mouse Exon 1.0 ST Array).

Les puces sont constituées de « features » ciblant des séquences spécifiques du génome. Elles peuvent regrouper jusqu'à plusieurs millions de sondes (communément appelées « probes »), oligonucléotides simples brins de 25 nucléotides.

Les puces de la génération antérieure (puces dites « classiques » ou 3') permettent l'étude du transcriptome à l'échelle du gène uniquement. Les sondes de ce type de puces ciblent exclusivement les quelques centaines de bases à l'extrémité 3' d'un gène. Cette approche considère que celle-ci est clairement définie et connue pour chaque gène autrement dit que tous les transcrits issus d'un même gène ont une même extrémité 3' ; le gène est alors considéré comme une unité entière et cohérente. Ce type de puce impose donc un biais important : les transcrits comportant des extrémités 3' alternatives en amont de la séquence ciblée par les sondes ne seront pas traités. Par conséquent, ces puces ne permettent pas une mesure « complète » de l'expression des gènes. D'autre part, ne ciblant qu'une portion faible des gènes, ces puces ne sont évidemment pas adaptées à l'étude des événements d'épissage alternatif. Les puces « tout exon » possèdent des sondes ciblant le gène dans toute sa longueur afin d'hybrider tous les exons constituant le gène. Il est alors possible d'analyser l'expression des différents isoformes d'un gène. On est capable de mesurer non plus seulement une expression quantitative du gène mais aussi une expression qualitative. Les puces « tout exon » Affymetrix permettent d'interroger les transcrits d'environ 20 000 gènes bien caractérisés (issus de données RefSeq) ainsi qu'un certain nombre de gènes prédits de manière bioinformatique. Les algorithmes de conception de sondes utilisent une vaste variété de collections d'exons et transcrits identifiés ou prédits telles que celles des banques Ensembl et GenScan.

Les puces exon contiennent approximativement 5.4 millions de sondes constituant 1.4 millions d'ensembles de sondes (probeset ou PS) correspondant à des exons ou des portions d'exon. Cela implique des différences importantes entre l'organisation des puces exons et celle des puces classiques. La contrainte majeure des puces exon est le nombre très important de sondes nécessaires à son utilisation. Le nombre de sondes ciblant une « unité » de base à analyser (gène pour les puces classiques et exon pour les puces exons) est en moyenne de onze pour les puces classiques ; celui-ci est de quatre pour les puces exons.

Dans les puces ADN classiques, chacune de ces sondes (dites « Perfect Match » ou PM) est associée à une sonde adjacente contenant un nucléotide différent en position centrale (sonde dite « Mis-Match » ou MM). Les conditions d'hybridation sont

déterminées pour maximiser la liaison avec les sondes PM et minimiser celle avec les sondes MM. Ces sondes MM permettent alors de mesurer l'affinité de leur sonde PM correspondante, elles permettent le calcul du bruit de fond de la puce puisque ces sondes ne sont pas censées s'hybrider ou très peu. Dans la plupart des puces classique, une paire de 11 sondes PM/MM est utilisée pour cibler un gène. Les 22 sondes constituent alors un probeset ciblant théoriquement un seul gène (Figure 41).

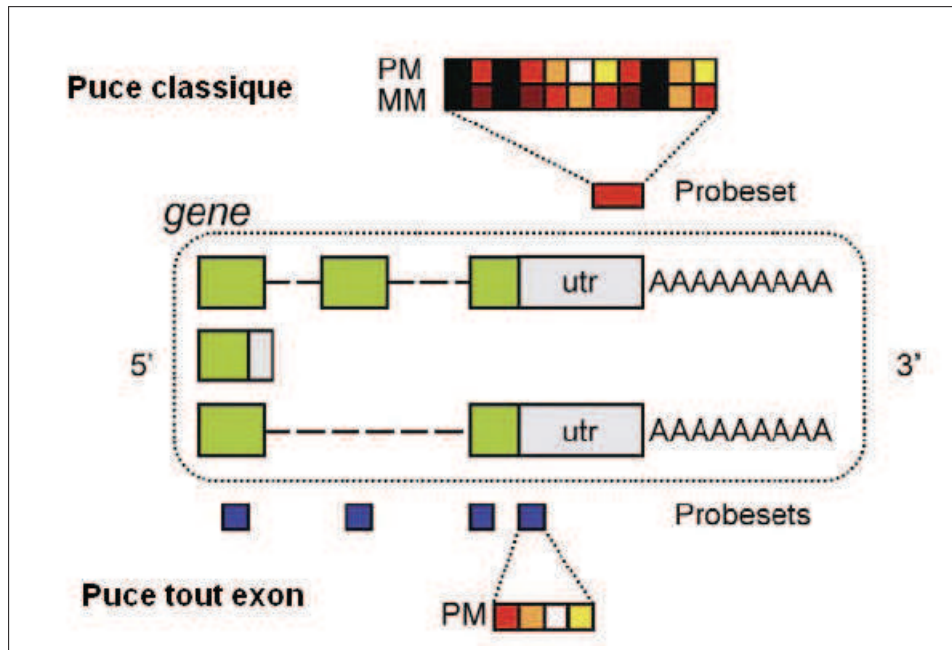


FIGURE 41: *Différences de conception entre les puces classiques et les puces « tout exon ».* La puce exon n'est constituée que de sondes PM, ces PS d'en moyenne 4 sondes sont répartis sur toute la longueur du gène.

Dans la puce « tout exon », les sondes MM sont remplacées par un lot distinct de plusieurs centaines de milliers de sondes. Elles sont utilisées pour corriger les intensités d'hybridation des sondes de la puce. Cette étape est d'autant plus importante qu'à la différence des puces classiques, les sondes de la puce « tout exon » ne présentent pas les mêmes propriétés physicochimiques. En effet, dans les puces classiques les sondes sont conçues pour toutes avoir le même contenu en GC : leur capacité d'hybridation n'est donc pas biaisée par leur contenu nucléotidique mais est seulement dépendante de leur complémentarité à la séquence cible. Les contraintes de conception sont autres pour la puce exon : un exon peut être de petite taille



(quelques dizaines de nucléotides seulement) et le choix des sondes pouvant le cibler est alors très limité. Dans cette configuration, il n'est pas possible de produire plus de 5 millions de sondes ayant exactement le même contenu en GC. Les intensités mesurées avec les puces exons ne sont donc pas seulement dépendantes d'une différence d'expression des transcrits ciblés mais aussi des propriétés intrinsèques des sondes. C'est pourquoi il est particulièrement important de corriger les intensités des puces exon avant toute analyse. Les sondes contrôles utilisées sont appelées anti-génomiques ; elles sont représentatives de tous les contenus et combinaisons en GC possibles.

## B Annotations

La qualité des sondes des puces « tout exon » est très variable avec une incidence directe sur la fiabilité des résultats. D'autre part, le nombre de sondes par exon peut être faible. Une valeur aberrante donnée par une sonde de « mauvaise qualité » peut donc induire l'élimination abusive d'un exon régulé (faux négatif) ou au contraire la sélection erronée d'un exon (faux positif). C'est pourquoi nous avons choisi dans notre système d'analyse d'annoter les sondes selon différents critères en définissant des sondes de « bonne » ou « mauvaise » qualité. Selon le degré de stringence voulue sur une analyse, il est alors possible de jouer sur ces critères.

Les différents critères utilisés sont les suivants :

- le contenu en GC de la sonde
- le nombre d'occurrences génomiques de la sonde
- la réalisation d'un test statistique pour déterminer si une sonde est dans le bruit de fond ou non
- la cible génomique de la sonde

Le contenu en GC d'une sonde influence directement ses propriétés d'hybridation. Celui-ci est très variable dans une puce « tout exon » : il varie de 3 à 25 parmi les sondes exploitables c'est-à-dire non contrôles. Nous avons décidé d'éliminer des algorithmes d'analyse les sondes avec plus de 18 GC car on les estime trop imprévisibles quant à leur capacité d'hybridation.

L'alignement des sondes de la puce a été réalisé contre l'intégralité des chromosomes. Près de 400.000 sondes s'hybrident plus d'une fois sur le génome. Elles sont alors annotées afin de pouvoir les exclure ou non des algorithmes d'analyse par la suite en fonction de la stringence souhaitée.

On applique un test statistique de type « DABG » (Detection Above BackGround) pour déterminer l'expression d'une sonde. Une p-valeur est associée à chaque intensité de sonde en déterminant le nombre de sondes contrôles ayant une intensité plus importante que la sonde avant soustraction du bruit de fond. Autrement dit, plus l'intensité de la sonde est au-dessus du bruit de fond, plus le nombre de sondes contrôles ayant une intensité supérieure est petit et donc plus la p-valeur est faible. On fixe le seuil d'expression d'une sonde une p-valeur de 0.05. On élimine de l'analyse les sondes non exprimées dans les deux conditions de l'expérience (mais on conserve les sondes exprimées dans une seule).

Enfin, le projet Fasterdb réalisé dans l'équipe a permis d'établir un répertoire de tous les événements d'épissage trouvés pour les gènes codants humains et murins. En effet, à partir des séquences Ensembl d'un gène et des transcrits qui lui sont associés, les gènes sont redéfinis (le début et leur fin ainsi que leur structure en exons et introns) en fonction de critères déterminés par l'équipe. Cet algorithme permet donc un recensement des événements d'épissage observés pour tous les gènes analysés. C'est cette annotation qui permet d'associer une sonde à un exon particulier d'un gène donné. Cette étape est particulièrement importante car elle permet de redéfinir les exons en « entités » cohérentes (réagissant de la même manière) pour une analyse des puces la plus proche possible de la réalité biologique. D'autre part, connaître quel est le type d'épissage régulé est essentiel lors de l'étape de validation expérimentale.

## C Système d'analyse mis en place

Un jeu classique de puces est constitué de deux conditions à comparer (par exemple : pathologique versus contrôle, traitement versus contrôle, déplétion de facteur versus contrôle...). Pour chaque condition, on a plusieurs réplicats expérimentaux (en général trois ou quatre). Les ARN totaux des cellules sont extraits, rétro-transcrits en

ADNc puis fragmentés et marqués à la biotine. Après dépôt sur les puces, hybridation et marquage des échantillons, les puces sont scannées et un fichier brut « CEL » est alors généré pour chaque puce. Celui-ci contient pour chaque sonde son intensité brute non corrigée et non traitée sous la forme d'une valeur numérique (transposition de l'intensité de fluorescence correspondant à l'intensité d'appariement de la sonde). Ce sont ces fichiers bruts qui sont traités par le système d'analyse.

## 1 Prétraitement

On ne peut analyser directement les données brutes, notamment parce que la variabilité induite par la technique peut être aussi importante que celle du système biologique. Il est donc nécessaire de minorer les variations artefactuelles à la fois entre puces (variabilité inter-puce) et au sein d'une même puce (variabilité intra-puce). C'est le rôle du prétraitement. Le prétraitement est constitué de plusieurs étapes :

- la normalisation des données brutes
- la correction du bruit de fond
- l'ajout d'un « facteur de stabilité »
- la transformation en logarithme 2

Il existe de nombreuses manières de normaliser les données entre puces. L'une des méthodes les plus courantes est la normalisation par quantile. C'est une méthode globale de normalisation c'est-à-dire qu'elle utilise les intensités de l'ensemble des sondes de la puce. L'hypothèse de base à l'analyse des puces ADN est que la majorité des gènes ont leur expression inchangée, on cherche à révéler ceux pour lesquels elle est différentielle dans une condition comparée à une autre. Pour ce faire, on doit appliquer une normalisation robuste forçant toutes les puces de l'expérience à avoir la même distribution.

Comme précédemment décrit, la correction du bruit de fond, résultant de l'intensité des sondes appariées de manière non spécifique, est l'un des enjeux majeurs dans le traitement des puces exon. De part la variabilité de qualité des sondes utilisées, il est nécessaire de corriger toute variation non spécifique. Cette correction est, dans notre

cas, réalisée à l'aide des sondes contrôles anti-génomiques et du logiciel Affymetrix APT (Affymetrix Power Tools).

L'ajout d'un « facteur de stabilité » arbitraire (ici : une valeur numérique 4) permet de minimiser les variations de régulation extrêmes (lorsqu'un gène ou un exon n'est pas exprimé dans l'une des deux conditions, le rapport alors calculé avec cette valeur quasi-nulle tend artificiellement vers l'infini).

Enfin, on applique une transformation logarithmique en base 2 aux intensités. Cette transformation permet de rendre les données symétriques et d'obtenir une distribution proche de la loi normale, critère nécessaire à l'application de tests statistiques de type paramétriques réalisés par la suite.

## 2 Traitement

Après le prétraitement des données, le processus classique d'analyse des puces se poursuit avec l'étape de sommation des intensités. Les sondes définissant un PS (Probeset) servent sans sélection à déterminer le niveau d'expression de l'exon. Il en est de même pour les PS définissant un MPS (Metaprobeset ou gène annoté) déterminant le niveau d'expression du gène. Le problème de ce genre de démarche est la difficulté de « rassembler » des données de variabilité différente due à l'affinité variable des sondes et expérimentale d'un réplicat à l'autre. Il existe plusieurs méthodes complexes de sommation des données appliquées aux puces « tout exon » ; les plus courantes sont RMA et PLIER. Ces algorithmes associent pour chaque puce (c'est-à-dire pour chaque réplicat de chaque condition) une valeur d'expression pour un gène et ou un exon donné. C'est seulement une fois les niveaux d'expression des exons et gènes calculées que la sélection se produit (élimination des valeurs aberrantes ou « outliers »). Celle-ci comporte de nombreux critères mais qui ne s'appliquent qu'à l'échelle de l'exon. Ces algorithmes ne sont pas parfaits : le taux de faux positifs est élevé. D'autre part, le fait que la sélection se fasse en aval du calcul des niveaux d'expression implique des critères de sélection très stringents par la suite pour ne pas traiter des valeurs aberrantes et donc peut impliquer un taux de faux négatifs important.

Notre démarche est de simplifier au maximum le problème de complexité du calcul du niveau d'expression d'un gène. Comme décrit précédemment, l'étape de sélection permet d'éliminer les sondes au comportement potentiellement aberrant. On traite la variabilité expérimentale en ne considérant plus les intensités des réplicats séparément mais en traitant plutôt le rapport de celles-ci. En effet, dans la mesure où un réplicat expérimental d'une condition est réalisé en même temps que le réplicat correspondant dans l'autre condition, le biais expérimental est le même pour un réplicat donné. L'utilisation du rapport (« Fold Change » ou FC) des intensités d'un même réplicat permet alors de minorer la variabilité expérimentale. Pour l'intégralité de l'analyse, on ne considère plus les intensités seules mais uniquement les rapports d'intensités.

#### **a Niveau de régulation globale du gène**

Pour déterminer le niveau de régulation globale d'un gène, on doit dans un premier temps éliminer les sondes au comportement différent pour ne garder que les sondes qui se comportent toutes de la même manière dans une condition donnée. On espère ainsi éliminer les variations dues à des événements d'épissage et ne conserver les sondes qui ne ciblent que les exons « constitutifs ». C'est l'étape de lissage : on élimine du calcul les sondes dont la différence d'intensité avec la moyenne de toutes les intensités des sondes est supérieure à l'écart-type.

On calcule ensuite la médiane des FCs des sondes sur cet ensemble de valeurs lissées ; une médiane en comparaison à une moyenne permet d'avoir une valeur moins influencée par les valeurs extrêmes du lot. Enfin on réalise un test de Student sur les FCs obtenus pour un gène. Il s'agit d'un test statistique permettant de comparer la moyenne de deux échantillons ou (et c'est le cas ici) la moyenne d'un échantillon à une valeur fixée. Ici, on compare donc notre échantillon de 10 à 40 médianes à une valeur fixée de 1. On obtient alors la signification statistique de la régulation du gène dans une condition par rapport à une autre. Une médiane globale permet d'obtenir le niveau de régulation du gène. Enfin plus on effectue un même test statistique plus on augmente la probabilité d'obtenir une p-valeur significative de manière aléatoire.

Dans le cas de la régulation des gènes, on effectue près de 20.000 tests statistiques, il est donc nécessaire de corriger ce biais à l'aide de fonction de correction des p-valeurs. Dans notre cas, nous utilisons une correction de Benjamini-Hochberg.

Si le FC du gène est supérieure à 1,5 (seuil arbitraire communément utilisé dans l'analyse de puce) et que la p-valeur corrigée est significative, on décide que le gène est régulé dans une condition par rapport à l'autre (Figure 42).

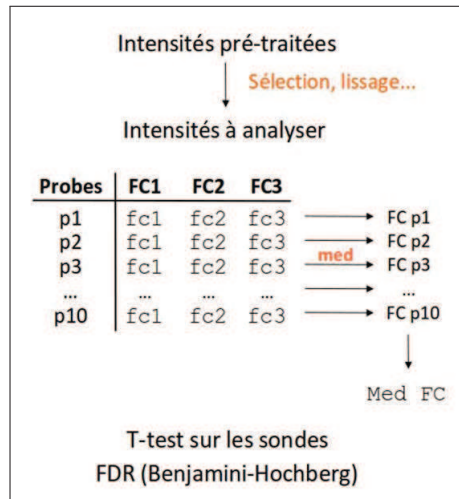


FIGURE 42: *Workflow d'analyse au niveau du gène.*

### b Niveau de régulation de l'exon

L'épissage alternatif est par définition l'utilisation différentielle des exons d'un gène. Il ne faut cependant pas oublier de prendre en compte le niveau d'expression du gène lui-même. En effet, si un gène est deux fois plus exprimé dans une condition par rapport à une autre alors la comparaison directe des intensités de ces exons portera à croire que ces exons sont deux fois plus inclus dans la première condition que la deuxième. Il est donc nécessaire de « normaliser » les intensités par l'expression global du gène. On appelle ces valeurs d'intensités normalisées les NI (Normalized Intensity).

Le rapport des deux intensités normalisées permet alors d'obtenir la magnitude d'expression différentielle de la sonde dans une condition par rapport à une autre. Si cette valeur dépasse un certain seuil, on peut estimer qu'une sonde est différentiellement exprimée. Cette valeur est appelée « Splicing Index » ou SI :

$$NI_A = \frac{\text{Intensité Condition}_A}{\text{Expression du gène dans Condition}_A} \text{ et } NI_B = \frac{\text{Intensité Condition}_B}{\text{Expression du gène dans Condition}_B}$$

$$SI = \frac{NI \text{ Condition}_A}{NI \text{ Condition}_B}$$

C'est à partir de ces valeurs que l'on calcul le SI d'une entité. Afin de ne traiter que les entités exprimées dans l'expérience, on ne traite que les sondes présentant une p-valeur de « DABG » inférieur à 0.05 dans au moins une condition c'est-à-dire les entités au dessus du bruit de fond. On cherche à mettre en évidence des entités régulées avec le plus de confiance possible. Il est important de toujours associer un test statistique à une valeur de SI afin d'avoir à la fois la signifiante statistique et le niveau de la régulation étudiée (Figure 43).

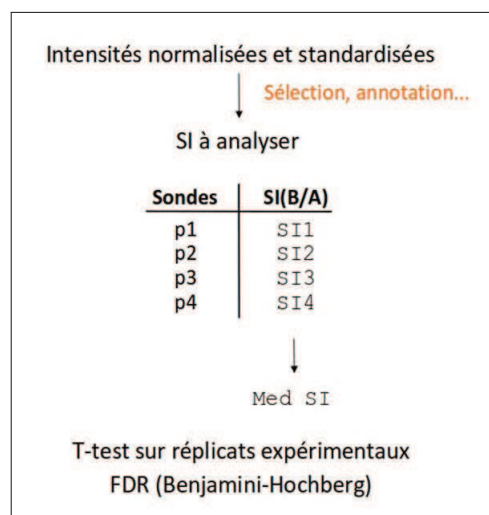


FIGURE 43: *Workflow d'analyse au niveau de l'exon.*

Enfin on applique de nouveau une correction pour les tests multiples. La p-valeur obtenue doit être significative (inférieure à 0.05) et la médiane des SI supérieure à 1.5 pour qu'une entité soit considérée comme différentiellement exprimée.

### III Articles

#### A Publication dans NSMB

Dans cet article publié en 2012, nous montrons que DDX5 et DDX17 régule l'épissage alternatif de nombreux facteurs d'interaction à la chromatine, parmi lesquels la macroH2A1. Les exons régulés par DDX5 et DDX17 de cette histone sont les exons mutuellement exclusifs 6 et 7 codant pour deux isoformes protéiques différentes. Celles-ci ont des effets opposés sur le phénotype migratoire et invasif des cellules 4T1. Elles contrôlent positivement ou négativement des gènes tel que SOD3, qui sont impliqués dans le contrôle du métabolisme redox.

Le jeu de données de puce exon utilisé pour ce projet est le premier sur lequel le système d'analyse développé à été appliqué. De nombreux allers-retours entre algorithmes d'analyse (et en particulier critères de sélection) et les validations par RT-PCR ont été effectués afin d'optimiser les résultats obtenus. Il était particulièrement important d'avoir les listes les plus fiables possibles dans la mesure où ce sont les recherches en enrichissement de fonctions qui ont orienté l'hypothèse biologique et l'analyse fonctionnelle du gène candidat (rôle de SOD3 dans le métabolisme redox).



# Splicing switch of an epigenetic regulator by RNA helicases promotes tumor-cell invasiveness

Etienne Dardenne<sup>1,5</sup>, Sandra Pierredon<sup>2,3,5</sup>, Keltouma Driouch<sup>4</sup>, Lise Gratadou<sup>1</sup>, Magali Lacroix-Triki<sup>2,3</sup>, Micaela Polay Espinoza<sup>1</sup>, Eleonora Zonta<sup>1</sup>, Sophie Germann<sup>1</sup>, Hussein Mortada<sup>1</sup>, Jean-Philippe Villemain<sup>1</sup>, Martin Dutertre<sup>1</sup>, Rosette Lidereau<sup>4</sup>, Stéphan Wagner<sup>2,3</sup> & Didier Auboeuf<sup>1</sup>

Both epigenetic and splicing regulation contribute to tumor progression, but the potential links between these two levels of gene-expression regulation in pathogenesis are not well understood. Here, we report that the mouse and human RNA helicases Ddx17 and Ddx5 contribute to tumor-cell invasiveness by regulating alternative splicing of several DNA- and chromatin-binding factors, including the macroH2A1 histone. We show that *macroH2A1* splicing isoforms differentially regulate the transcription of a set of genes involved in redox metabolism. In particular, the *SOD3* gene that encodes the extracellular superoxide dismutase and plays a part in cell migration is regulated in an opposite manner by *macroH2A1* splicing isoforms. These findings reveal a new regulatory pathway in which splicing factors control the expression of histone variant isoforms that in turn drive a transcription program to switch tumor cells to an invasive phenotype.

Modifications of gene-expression programs resulting from alterations of epigenetic, transcription and pre-mRNA-splicing regulation largely contribute to the development of various cancers. At the epigenetic level, a plethora of chromatin alterations affecting canonical histone-tail modifications appear to be involved in tumor malignancy<sup>1–3</sup>. There is also an increasing interest in analyzing the role in cancer of histone variants of the H2A family, including H2A.Z, H2AX, macroH2A1 and macroH2A2. Indeed, the exchange of canonical histones for histone variants is one of the most extensive epigenetic regulation events that affects genomic functions such as cell-cycle control, response to DNA damage, heterochromatin silencing and transcriptional regulation<sup>3–8</sup>. Changes in the expression of H2A.Z, H2AX or macroH2A1 histones in cancer can affect DNA repair, thereby leading to alterations of chromosome segregation or cell proliferation<sup>3–7</sup>. Understanding the role in physiological or pathological states of histone variants, whose diversity can also be increased by alternative splicing, remains a major issue.

Numerous studies have also reported changes in the expression of splicing variants in cancers<sup>9–12</sup>. Alternative splicing is the main mechanism responsible for increasing the diversity of the proteome. Indeed, 90% of human genes generate several splicing variants, producing protein isoforms that can have different and sometimes opposing biological activities<sup>11–14</sup>. Differential selection of alternatively spliced exons relies on the recognition of the splicing sites by the spliceosome and on splicing-regulatory sequences, located within exons and introns, that are recognized by splicing factors<sup>13–15</sup>. In a way akin to how transcriptional programs are controlled by transcription factors, splicing

factors control the splicing programs and thus the alternative splicing of gene networks. Although many differences at the splicing level have been described for cancer cells compared to normal cells, one major challenge is to better understand how splicing programs and splicing factors may contribute to specific cancer-associated phenotypes, including cellular invasion and migration. Indeed, deciphering the genetic programs involved in cancer progression will allow researchers to better understand how cancer cells can evolve and form metastases, which are the main cause of death in cancer. Another major perspective is to determine whether and how cross-talk between the different layers of gene-expression regulation, such as epigenetic and splicing programs, contribute to tumor progression.

The 4T1 mouse model of tumor progression comprises four syngeneic tumor cell lines that can give rise to primary tumors with a spectrum of metastatic phenotypes when implanted into mouse mammary fat pads<sup>10,16,17</sup>. Using splicing-sensitive microarrays to analyze the 4T1 mouse model, we and others have reported that several alternative-splicing events are associated with the ability of primary tumors to disseminate<sup>10,17</sup>. Many of the splicing variants differentially expressed between primary tumors that do or do not develop into metastases are also expressed in normal tissues and have been conserved during evolution, which suggests that they are under the control of conserved splicing regulators<sup>10</sup>.

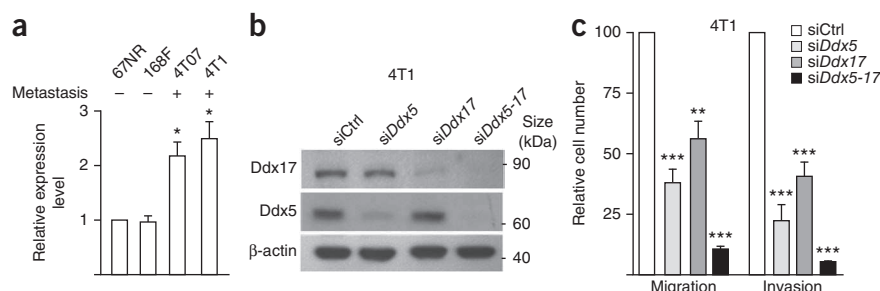
To characterize the role of splicing factors in tumor progression, we focus on the Ddx17 RNA helicase (also known as p72) that we observed to be overexpressed in invasive tumor cells and that we show to contribute to tumor-cell migration and invasion together

<sup>1</sup>Institut National de la Santé et de la Recherche Médicale, U1052, Centre de Recherche en Cancérologie de Lyon, Lyon, France. <sup>2</sup>Institut National de la Santé et de la Recherche Médicale, U563, Institut Claudius Regaud, Toulouse, France. <sup>3</sup>Institut National de la Santé et de la Recherche Médicale, U981, Institut de Cancérologie Gustave Roussy, Villejuif, France. <sup>4</sup>Laboratory of Oncogenetics, Institut Curie–Hôpital René Huguenin, St-Cloud, France. <sup>5</sup>These authors contributed equally to this work. Correspondence should be addressed to S.V. (stephan.wagner@inserm.fr) or D.A. (didier.auboeuf@inserm.fr).

Received 9 November 2011; accepted 24 August 2012; published online 30 September 2012; doi:10.1038/nsmb.2390

**Figure 1** The Ddx5 and Ddx17 RNA helicases contribute to cell migration and invasion.

(a) Relative expression level of *Ddx17* mRNA measured by RT-qPCR in mouse primary tumors with different metastatic capabilities, as indicated. (b) Western blot analysis of Ddx5, Ddx17 and  $\beta$ -actin, 48 h after the transfection of mouse 4T1 cells with a control siRNA (siCtrl) or siRNAs targeting *Ddx5* (si*Ddx5*), *Ddx17* (si*Ddx17*) or both *Ddx5* and *Ddx17* (si*Ddx5-17*). (c) Relative number of migrating and invasive 4T1 cells after transfection with siCtrl, si*Ddx5*, si*Ddx17* or si*Ddx5-17*. Data are shown as mean number of migrating and invasive cells transfected with each siRNA relative to those of cells transfected with the control siRNA. In a and c, histograms represent the average of at least three independent experiments. Error bars, s.e.m.; \* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$  (*t* test).



with its paralog Ddx5 (p68). Ddx17 and Ddx5 are transcriptional co-regulators of the estrogen receptor and the p53 transcriptional factor<sup>18–21</sup> and are also splicing regulators known to control alternative splicing of the genes encoding CD44, H-ras and tau<sup>18,19,22–24</sup>. Our genome-wide analysis at the exon level revealed that Ddx5 and Ddx17 regulate an alternative-splicing network that affects several DNA- and chromatin-binding factors, including the macroH2A1 histone gene (*mH2A1*; also called *H2AFY*). We demonstrate that the *mH2A1* splicing isoforms are differentially expressed in primary tumors that either give rise to metastases or do not. Furthermore, these isoforms differentially regulate breast cancer–cell invasiveness through the transcriptional regulation of genes involved in redox metabolism, including *SOD3*. Thus, our work reveals a new cascade of altered gene-expression events participating in tumor progression, wherein splicing factors (Ddx5 and Ddx17) control the alternative splicing of an epigenetic factor (*mH2A1*), leading in turn to transcriptional alterations of genes involved in a specific cellular program.

## RESULTS

### Ddx5 and Ddx17 contribute to cell migration and invasion

While analyzing the transcriptome of the 4T1 mouse model of mammary tumor progression<sup>10</sup>, we observed that Ddx17 was overexpressed in primary tumors that are more likely to give rise to metastases (4T1 cells) or micrometastases (4T07 cells) as compared to tumors that are not able to disseminate (67NR cells) or that disseminate to the lymph nodes (168FARN cells). We validated this result by reverse-transcription quantitative PCR (RT-qPCR) on RNAs extracted from mouse tumors (Fig. 1a) and also showed that Ddx17 was overexpressed at both the RNA and the protein levels in the invasive 4T1 cell line as compared to the noninvasive 67NR cell line (Supplementary Fig. 1a). To test the potential role of Ddx17 in tumor-cell invasiveness, we knocked down its expression by using a specific siRNA (Fig. 1b). Because Ddx5 and Ddx17 often have redundant functions<sup>18,19</sup> and although Ddx5 did not seem to be misregulated (Supplementary Fig. 1a), we also used siRNAs against either Ddx5 or against a conserved region that targeted both Ddx5 and Ddx17 (Fig. 1b). Both cell migration and invasion were inhibited by Ddx5 or Ddx17 depletion, and this inhibition was enhanced when both Ddx5 and Ddx17 were depleted (Fig. 1c). Similar effects were obtained with either the single siRNA targeting both Ddx5 and Ddx17 or a mixture of siRNAs separately targeting Ddx5 and Ddx17 in 4T1 cells, whereas depletion of Ddx5 and Ddx17 in the noninvasive 67NR cells had no effect (Supplementary Fig. 1b,c). These results demonstrate that Ddx5 and Ddx17 have a role in cell invasiveness because depletion of Ddx5 and/or Ddx17 neither increased cell apoptosis nor significantly decreased cell-cycle progression (Supplementary Fig. 1d,e).

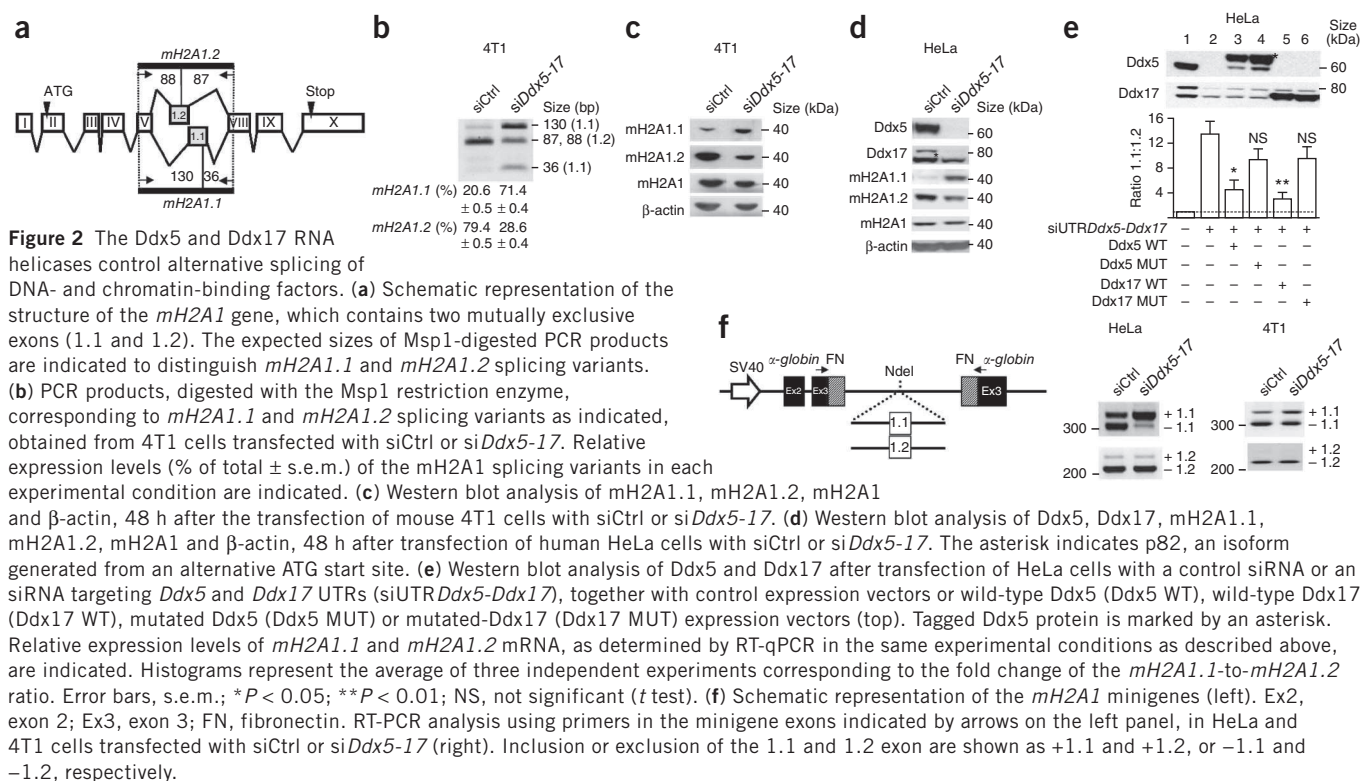
To understand how Ddx5 and Ddx17 regulate cell migration and invasion, we used Affymetrix exon arrays to analyze the transcriptome of 4T1 cells at the exon level, following siRNA-mediated depletion of Ddx5 and Ddx17. There were 1,136 genes predicted to be affected at the global gene-expression level (Supplementary Table 1) and 1,325 genes predicted to be affected at the exon level (Supplementary Table 2). The effect of Ddx5 and Ddx17 depletion was validated by RT-PCR for 31 out of 42 splicing variants (74% validation rate) (Supplementary Fig. 2).

Gene Ontology analysis through the Panther website<sup>25</sup> indicated that genes regulated at the splicing level as compared to genes regulated at the global level were enriched in ‘binding activity’ ( $P = 2.8 \times 10^{-13}$ ), in particular ‘nucleic acid binding’ ( $P = 2.9 \times 10^{-11}$ ), ‘RNA binding’ ( $P = 9 \times 10^{-8}$ ), ‘chromatin binding’ ( $P = 1.9 \times 10^{-5}$ ) and ‘DNA binding’ ( $P = 5 \times 10^{-4}$ ). This observation suggests that the set of alternative exons that are co-regulated by Ddx5 and Ddx17 are enriched in functionally related genes and that Ddx5 and Ddx17 regulate the alternative splicing of DNA- and chromatin-binding factors that could in turn affect the transcription of genes involved in tumor progression.

### Ddx5 and Ddx17 control *mH2A1* alternative splicing

To test this hypothesis, we focused on the *mH2A1* histone gene because recent reports have described a role for it in cancer<sup>4,7,8</sup>. The mouse *mH2A1* gene generates two splicing isoforms (*mH2A1.1* and *mH2A1.2*) through the use of two mutually exclusive exons (Fig. 2a). As predicted by the exon-array analysis, depletion of Ddx5 and Ddx17 increased the level of the *mH2A1.1* isoform and decreased that of the *mH2A1.2* isoform but had no effect on the global *mH2A1* gene expression level (Fig. 2b and Supplementary Fig. 3a,b). These results were confirmed at the protein level (Fig. 2c) and were extended in human HeLa cells (Fig. 2d and Supplementary Fig. 4).

To perform rescue experiments, siRNAs targeting the untranslated regions (UTRs) of the endogenous Ddx5 and Ddx17 mRNAs were designed. Transfection of these siRNAs led to an increase in the ratio of *mH2A1.1* to *mH2A1.2* (Fig. 2e, lane 2 compared to lane 1). Furthermore, cell transfection with either Ddx5 or Ddx17 expression vectors (to produce mRNAs without UTRs that are therefore refractory to the UTR-targeting siRNAs) rescued to a large extent the splicing pattern generated by endogenous Ddx5 and Ddx17 protein depletion (Fig. 2e, lane 2 compared to lane 3 and to lane 5). Even though overexpression of Ddx5 or Ddx17 did not fully rescue the splicing pattern induced by Ddx5 and Ddx17 depletion, our data suggest that Ddx5 and Ddx17 have similar functions, as depletion of either Ddx5 or Ddx17 affected *mH2A1* splicing (Supplementary Fig. 4c). Notably, overexpressing mutated Ddx5 or Ddx17 proteins



that lacked the RNA helicase activity did not significantly rescue the *mH2A1* splicing pattern (Fig. 2e, lane 4 compared to lane 3 and lane 6 compared to lane 5).

To directly assay the effect of Ddx5 and Ddx17 on *mH2A1* alternative splicing, exons 1.1 or 1.2 and ~200 nucleotides of their surrounding intronic sequences were cloned into a reporter minigene (Fig. 2f). As expected, Ddx5 and Ddx17 depletion increased exon 1.1 inclusion in HeLa and 4T1 cells transfected with the minigene. Ddx5 and Ddx17 depletion did not increase exon 1.2 skipping (Fig. 2f), which suggested that sequences not included in the minigene are required to regulate exon 1.2. Alternatively, Ddx5 and Ddx17 may regulate exon 1.1 splicing, which in turn could affect exon 1.2 splicing. These results demonstrate that Ddx5 and Ddx17 regulate *mH2A1* alternative splicing and that the RNA helicase activity of Ddx5 and Ddx17 is required for their effects on splicing.

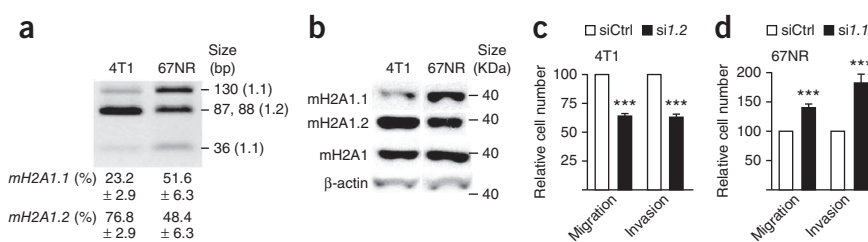
### *mH2A1* alternative splicing participates in cell invasiveness

We next investigated a possible role of *mH2A1* splicing isoforms in tumor-cell invasiveness. In this context, we observed that the

*mH2A1.1* isoform was expressed at a higher level in the noninvasive 67NR cells than in the invasive 4T1 cells, whereas the *mH2A1.2* isoform was expressed at a higher level in 4T1 cells, both at the RNA and protein levels (Fig. 3a,b). Whereas 67NR cells expressed similar levels of *mH2A1.1* and *mH2A1.2* isoforms, 4T1 cells expressed much less *mH2A1.1* than *mH2A1.2* (Fig. 3a,b and Supplementary Fig. 3c,d). This result was confirmed in tumor samples, as a lower ratio of *mH2A1.1* to *mH2A1.2* was observed in mouse primary tumors that gave rise to metastasis (Supplementary Fig. 3e).

Supporting the hypothesis that the *mH2A1* splicing isoforms have different roles in cell migration and invasion, specific depletion of the *mH2A1.2* isoform in the invasive 4T1 cells (Supplementary Fig. 3c) inhibited cell migration and invasion (Fig. 3c), whereas the depletion of the *mH2A1.1* isoform in the noninvasive 67NR cells (Supplementary Fig. 3c) increased cell migration and invasion (Fig. 3d). These results, which were confirmed by using different siRNAs (Supplementary Fig. 5a), demonstrate that the *mH2A1.2* isoform favored breast cancer-cell invasiveness, whereas the *mH2A1.1* isoform had the opposite effect. Transfection of 4T1 cells with siRNA targeting the *mH2A1.1*

**Figure 3** Alternative splicing of the macroH2A1 histone variant is involved in tumor progression. (a) PCR products, digested with the Msp1 restriction enzyme, corresponding to *mH2A1.1* and *mH2A1.2* splicing variants. Template is RNA isolated from 67NR or 4T1 cells. Relative expression levels (% of total ± s.e.m.) of the *mH2A1* splicing variants in each experimental condition are indicated. (b) Western blot analysis of *mH2A1.1*, *mH2A1.2*, *mH2A1* and β-actin, expressed in 4T1 or 67NR cells. (c) Relative number of migrating and invasive 4T1 cells after transfection with siCtrl or an siRNA targeting *mH2A1.2* (si1.2). (d) Relative number of migrating and invasive 67NR cells after transfection with siCtrl or a siRNA targeting *mH2A1.1* (si1.1). In c and d, histograms represent the average of at least three independent experiments. Error bars, s.e.m.; \*\*\**P* < 0.001 (*t* test).



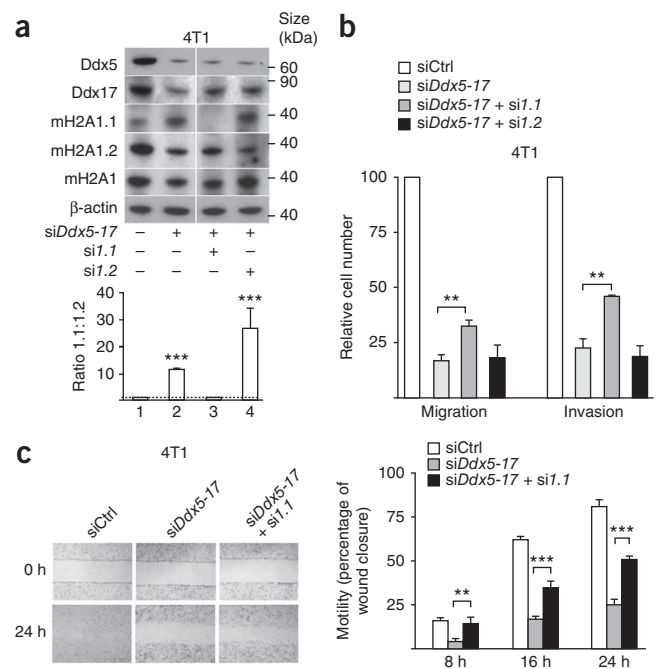
**Figure 4** Alternative splicing of the macroH2A1 histone variant contributes to the Ddx5-Ddx17 effect on cell migration and invasion. (a) Western blot analysis of Ddx5, Ddx17, mH2A1.1, mH2A1.2, mH2A1 and  $\beta$ -actin, 48 h after transfection of mouse 4T1 cells with siCtrl, siDdx5-17, si1.1 or si1.2 isoforms (top). Relative expression levels of mH2A1.1 and mH2A1.2 measured by RT-qPCR in the same experimental conditions as described above are indicated (bottom). Histograms represent the average of three independent experiments corresponding to the fold change of the mH2A1.1 to mH2A1.2 mRNA ratio. (b) Relative number of migrating and invasive 4T1 cells after transfection with siCtrl, siDdx5-17 and either si1.1 or si1.2. (c) Motility of 4T1 cells transfected with siCtrl or with siDdx5-17 alone or together with si1.1, measured by the wound healing assay at 8, 16 or 24 h. Histograms represent the average of at least three independent experiments. Error bars, s.e.m.; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$  ( $t$  test).

isoform had no effect, as the mH2A1.1 isoform is scarcely expressed in 4T1 cells, and depletion of the mH2A1.2 isoform had no effect in the noninvasive 67NR cells (Supplementary Fig. 5b).

To further test whether mH2A1 alternative-splicing regulation contributes to the effect of Ddx5 and Ddx17 on tumor cell invasiveness, Ddx5- and Ddx17-depleted 4T1 cells were transfected with siRNAs that specifically targeted each of the mH2A1 splicing variants (Fig. 4a). Notably, depletion of the mH2A1.1 isoform but not the mH2A1.2 isoform reduced the inhibitory effect of Ddx5 and Ddx17 depletion on cell invasion and migration (Fig. 4b). This result was further confirmed by the depletion of the mH2A1.1 isoform in Ddx5- and Ddx17-depleted cells, which reduced the inhibitory effect of Ddx5 and Ddx17 depletion on cell motility (Fig. 4c, siDdx5-17 compared to siDdx5-17 + si1.1). These results support a model in which mH2A1 alternative splicing mediates, at least in part, the inhibition by Ddx5 and Ddx17 depletion of tumor-cell invasiveness.

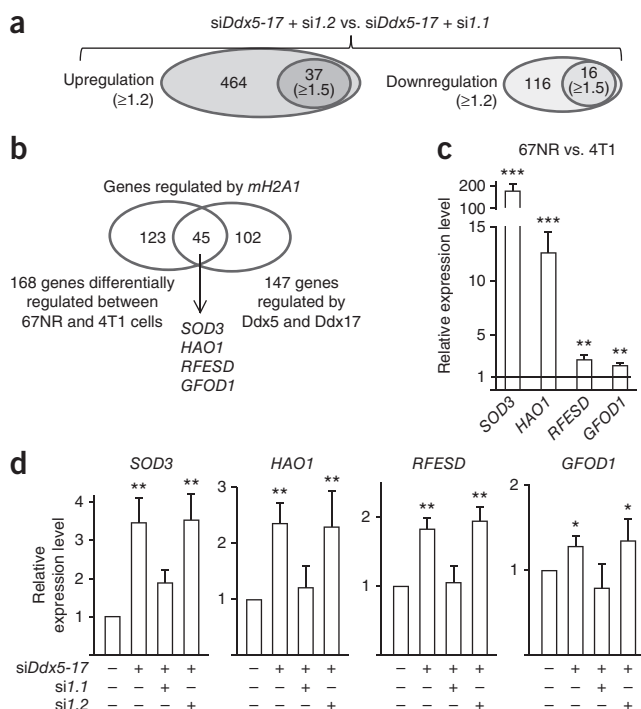
### mH2A1 splicing isoforms control redox metabolism genes

We next performed a high-throughput gene-expression analysis to identify genes whose expression might be differentially regulated by the mH2A1 splicing isoforms and thus could be involved in the



Ddx5- and Ddx17-dependent regulation of the invasive phenotype. For this purpose, we compared the transcriptome of 4T1 cells that had been depleted of either mH2A1.1 or mH2A1.2 in the context of Ddx5 and Ddx17 depletion. Manipulating mH2A1 isoform expression had only weak effects, with changes in the expression levels of only 53 genes by more than 1.5 fold (Fig. 5a and Supplementary Table 3). This was anticipated because several reports have indicated that mH2A1 has only a slight impact on steady-state levels of gene expression<sup>26–29</sup>. Reducing the threshold to 1.2-fold revealed 633 genes that were differentially regulated when comparing mH2A1.2- to mH2A1.1-depleted cells (Fig. 5a and Supplementary Table 3). We validated by RT-qPCR 17 of the 21 tested genes that had fold changes ranging from 1.2 to 2.6 (Supplementary Fig. 6a and Supplementary Table 3). Among the 633 genes, 501 genes were upregulated and 132 genes were downregulated in mH2A1.2-depleted cells compared to mH2A1.1-depleted cells, which suggests that the mH2A1.1 and mH2A1.2 isoforms have a differential impact on gene transcriptional activity.

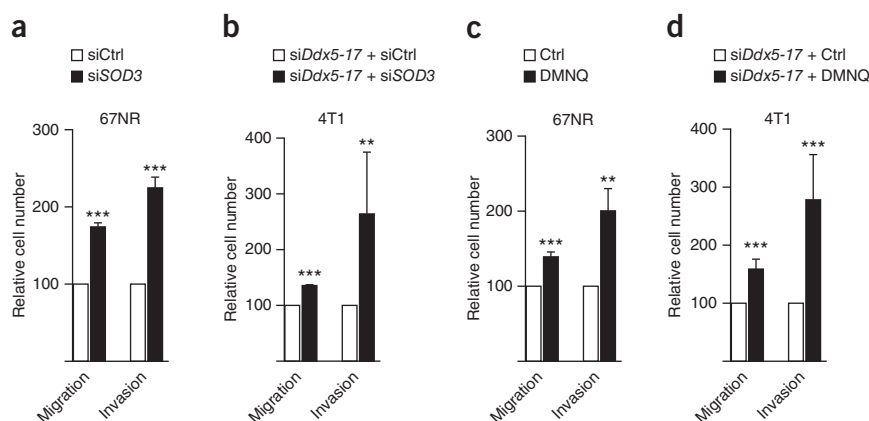
Of the genes that were differentially regulated by mH2A1 splicing isoforms, 45 were also affected by Ddx5 and Ddx17 depletion and were also differentially expressed in 67NR cells as compared to 4T1 cells (Fig. 5b), and more than one-third of these are involved in redox metabolism (Supplementary Table 3). Notably, metastatic cells may come from cancer cells that have avoided oxidative



**Figure 5** Regulation of mH2A1 splicing isoforms by Ddx5 and Ddx17 affects regulation of genes involved in redox metabolism. (a) Number of genes regulated at the global expression level, as determined by comparing 4T1 cells transfected with siDdx5-17 and with si1.2 or si1.1. (b) Venn diagram corresponding to mH2A1-regulated genes that were regulated by Ddx5 and Ddx17 and differentially expressed in 67NR as compared to 4T1 cells. (c) Relative expression level of SOD3, HAO1, RFESD and GFOD1 mRNAs, as determined by RT-qPCR, comparing nonmetastatic 67NR cells with metastatic 4T1 cells. (d) Relative expression level of SOD3, HAO1, RFESD and GFOD1 mRNAs measured by RT-qPCR in 4T1 cells transfected with siCtrl or siDdx5-17 alone or in combination with either si1.1 or si1.2. Histograms represent the average of at least three independent experiments. Error bars, s.e.m.; \* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$  ( $t$  test).



**Figure 6** Ddx5 and Ddx17-regulated *SOD3* expression is involved in cell migration and invasion. (a) Relative number of migrating and invasive 67NR cells after transfection with siCtrl or siRNA against *SOD3* (si*SOD3*), as determined by comparing with the number of cells transfected with siCtrl that migrated and invaded. (b) Relative number of migrating and invasive 4T1 cells after transfection with si*Ddx5-17* alone or in combination with si*SOD3*. (c) Relative number of migrating and invasive 67NR cells in the absence or presence of DMNQ. (d) Relative number of migrating and invasive 4T1 cells in the absence or presence of DMNQ. Histograms represent the average of at least three independent experiments. Error bars, s.e.m.; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$  (*t* test).

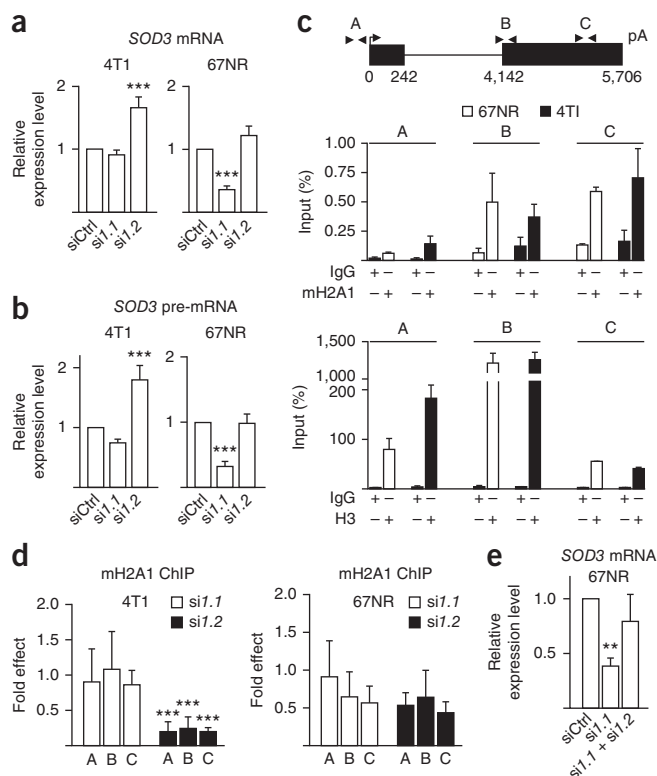


damage from excess reactive oxygen species (ROS) in the primary tumor site<sup>30,31</sup>. We therefore focused on four genes involved in redox metabolism: the extracellular superoxide dismutase (*SOD3*), hydroxyacid oxidase 1 (*HAO1*), Rieske (Fe-S) domain containing (*RFESD*) and glucose-fructose oxidoreductase domain containing 1 (*GFOD1*). We observed by RT-qPCR that 67NR cells that expressed a high *mH2A1.1* to *mH2A1.2* ratio as compared to metastatic 4T1 cells also expressed higher levels of *SOD3*, *HAO1*, *RFESD* and *GFOD1* mRNAs than metastatic 4T1 cells (Fig. 5c). Furthermore, Ddx5 and Ddx17 depletion increased *SOD3*, *HAO1*, *RFESD* and *GFOD1* mRNA expression (Fig. 5d). Finally, depletion of *mH2A1.1*, but not of *mH2A1.2*, abrogated the increase mediated by Ddx5 and Ddx17 depletion (Fig. 5d). Altogether, these data show that *mH2A1* splicing isoforms mediate a part of Ddx5- and Ddx17-induced regulation of gene expression in 4T1 cells, including that of genes involved in redox metabolism.

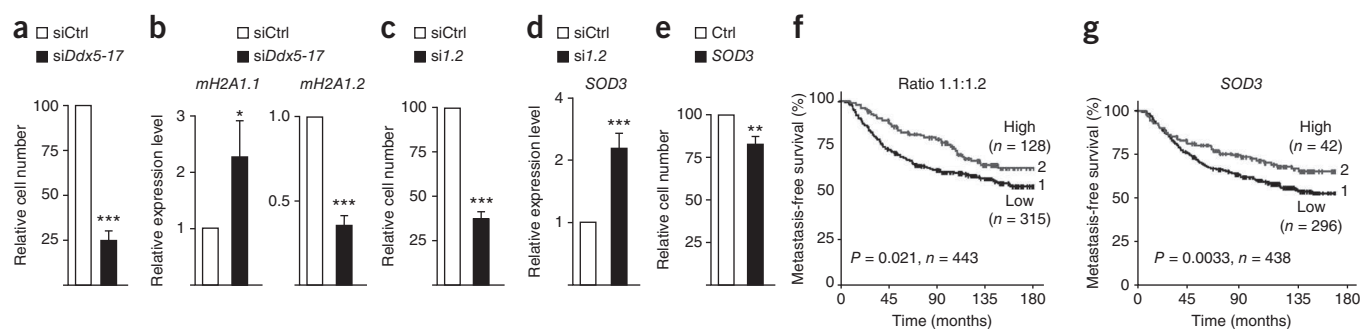
### *SOD3* regulation by *mH2A1* contributes to cell invasiveness

To directly examine the impact on tumor-cell invasiveness of Ddx5- and Ddx17-mediated regulation of *mH2A1* alternative splicing and the downstream *mH2A1*-dependent transcriptional regulation, we focused on *SOD3*, which converts extracellular superoxide radicals ( $[O_2]^{*-}$ ) into hydrogen peroxide ( $H_2O_2$ )<sup>32–35</sup> and is known to be involved in invasion<sup>32–35</sup>. As shown in Figure 6a, *SOD3* depletion (Supplementary Fig. 5d) indeed led to an increased invasiveness of the noninvasive 67NR cells. Notably, transfection of Ddx5- and Ddx17-depleted 4T1 cells with an siRNA targeting *SOD3* increased cell migration and invasion (Fig. 6b). This result suggests that Ddx5 and Ddx17 depletion increased the *mH2A1.1* isoform level (Fig. 2), which in turn increased *SOD3* expression level (Fig. 5d) and thereby reduced cell migration and invasion (Fig. 6b). Further supporting this model, directly adding *SOD3* protein to 4T1 cell-culture medium inhibited cell migration (Supplementary Fig. 5e), whereas adding 2,3-dimethoxy-1,4-naphthoquinone (DMNQ), a redox-cycling agent that induces superoxide anion formation and mimics *SOD3* depletion, increased the migration and invasion capability of 67NR cells and Ddx5- and Ddx17-depleted 4T1 cells (Fig. 6c,d).

The differential effects of the *mH2A1* splicing variants on the gene-expression level could be explained if the two isoforms act on different genes or if they act on the same genes but have opposite effects on gene transcription activity. Consistent with the second hypothesis, *mH2A1.2* depletion in metastatic 4T1 cells led to an increasing amount of *SOD3* mRNA, which suggested that *mH2A1.2* repressed *SOD3* expression (Fig. 7a). Conversely, depletion of *mH2A1.1* in non-metastatic 67NR cells decreased the amount of *SOD3* mRNA, which



**Figure 7** *mH2A1* associates with the *SOD3* gene, and its splicing variants regulate *SOD3* transcription activity in an opposite manner. (a) Relative *SOD3* mRNA expression level measured by RT-qPCR in 4T1 or 67NR cells transfected with siCtrl, si*I.1* or si*I.2*. (b) Relative *SOD3* pre-mRNA expression level measured by RT-qPCR, in 4T1 or 67NR cells transfected with siCtrl, si*I.1*, or si*I.2*. (c) Schematic representation of the *SOD3* gene structure and localization of the primers used to analyze *mH2A1* incorporation (upper panel). qPCR analysis with primers targeting the *SOD3* gene after ChIP using control antibodies (against IgG) or antibodies against *mH2A1* (middle panel) or histone H3 (lower panel). (d) Effect of *mH2A1.1* depletion (si*I.1*) or *mH2A1.2* depletion (si*I.2*) on the incorporation of *mH2A1* in the *SOD3* gene in 4T1 cells or in 67NR cells. Primers A, B, and C were as shown in c. (e) Relative *SOD3* mRNA expression level measured by RT-qPCR in 67NR cells transfected with siCtrl, si*I.1* alone or in combination with si*I.2*. Histograms represent the average of at least three independent experiments. Error bars, s.e.m.; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$  (*t* test).



**Figure 8** mH2A1 alternative splicing controlled by Ddx5 and Ddx17 in MDA-MB-231 cells. (a) Relative number of migrating MDA-MB-231 cells after transfection with siCtrl or siDdx5-17. (b) Relative expression level of *mH2A1.1* and *mH2A1.2* mRNAs measured by RT-qPCR in MDA-MB-231 cells transfected with siCtrl or siDdx5-17. (c) Relative number of migrating MDA-MB-231 cells after transfection with siCtrl or siI.2. (d) Relative expression level of *SOD3* mRNA measured by RT-qPCR in MDA-MB-231 cells transfected with siCtrl or siI.2. (e) Relative number of migrating MDA-MB-231 cells after addition of recombinant SOD3 protein to the cell-culture medium. Histograms represent the average of at least three independent experiments. Error bars, s.e.m.; \* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$  ( $t$  test). (f) Kaplan-Meier curves for 180-month outcome in patients with breast cancer ( $n = 443$ ) based on the *mH2A1.1*-to-*mH2A1.2* splicing ratio. (g) Kaplan-Meier curves for 180-month outcome in patients with breast cancer ( $n = 438$ ) based on *SOD3* mRNA expression level.

suggested that mH2A1.1 stimulated SOD3 expression (Fig. 7a). These results were confirmed by the *SOD3* pre-mRNA levels (Fig. 7b), so we concluded that mH2A1 isoforms had opposite effects on *SOD3* gene transcription activity.

We next tested whether the mH2A1 histone was incorporated into the *SOD3* gene in 67NR and 4T1 cells. Chromatin immunoprecipitation (ChIP) was performed with an antibody that recognizes both mH2A1 isoforms, as isoform-specific antibodies did not give any signal above background (data not shown). PCR analysis was performed by using primers at several locations along the *SOD3* gene (Fig. 7c). As shown in Figure 7c, mH2A1 was detected all along the *SOD3* gene, both in 67NR and 4T1 cells. A slightly higher level of mH2A1 was detected on the *SOD3* promoter in 4T1 as compared to 67NR cells (Fig. 7c), but there was also more nucleosome in the *SOD3* promoter in 4T1 as compared to 67NR cells, as demonstrated by H3 histone ChIP (Fig. 7c). Moreover, mH2A1.2 depletion but not mH2A1.1 depletion strongly decreased the mH2A1 histone level incorporated in the *SOD3* gene in 4T1 cells (Fig. 7d). Thus, the mH2A1.2 isoform, which is the main mH2A1 isoform expressed in 4T1 cells, is incorporated in the *SOD3* gene (Fig. 7d) and contributes to *SOD3* gene repression in these metastatic cells (Fig. 7a).

In contrast, in the nonmetastatic 67NR cells that expressed a similar level of the mH2A1.1 and mH2A1.2 isoforms, depletion of either the mH2A1.1 or the mH2A1.2 isoform induced a similar, statistically insignificant decrease in the level of the mH2A1 incorporated into the *SOD3* gene (Fig. 7d). Because mH2A1.1 depletion in 67NR cells decreased *SOD3* mRNA and pre-mRNA levels and mH2A1.2 depletion had opposite effects in 4T1 cells (Fig. 7a,b), we hypothesized that the mH2A1.2 isoform replaced the mH2A1.1 isoform on the *SOD3* gene following mH2A1.1 depletion in 67NR cells. This hypothesis was supported by the fact that *SOD3* gene repression in the 67NR cells that resulted from mH2A1.1 depletion was reversed by the co-depletion of the mH2A1.2 isoform (Fig. 7e). Collectively, these results suggest that both competition between the mH2A1.1 and mH2A1.2 isoforms for their incorporation into the *SOD3* gene and a higher mH2A1.2-to-mH2A1.1 ratio contribute to breast cancer-cell invasiveness through *SOD3* gene repression.

To further challenge this model, the human invasive MDA-MB-231 breast cancer-cell line was transfected with an siRNA that targeted both Ddx5 and Ddx17, which inhibited cell migration (Fig. 8a) and decreased the mH2A1.2 isoform and increased the mH2A1.1

isoform (Fig. 8b), as had been observed in the 4T1 mouse cell line. As expected, depletion of the mH2A1.2 isoform in MDA-MB-231 cells also inhibited cell migration (Fig. 8c) and increased the *SOD3* mRNA expression level (Fig. 8d). Therefore, the mH2A1.2 isoform may favor cancer-cell migration by repressing *SOD3* expression, as addition of SOD3 in MDA-MB-231 cell-culture medium inhibited cell migration (Fig. 8e).

Further supporting the physiopathological significance of our findings, analyses of a cohort of 443 patients with primary breast tumors with a well-documented clinical follow-up (Supplementary Table 5) indicated that low mH2A1.1-to-mH2A1.2 expression ratio was associated with shorter metastasis-free survival (Fig. 8f). Notably, a high level of mH2A1.2 isoform expression was associated with shorter metastasis-free survival, which represents a prognostic factor for disease severity that is independent of other known markers (Supplementary Fig. 7). Notably, the mH2A1.2 splicing isoform that represses *SOD3* expression in 4T1 and MDA-MB-231 cells (Figs. 7a and 8d) correlated negatively with the expression level of *SOD3* ( $P = 0.00019$ ; Spearman rank correlation test) in the primary breast-tumor collection. Finally, a high expression level of *SOD3* was associated with longer metastasis-free survival (Fig. 8g), in agreement with the inhibitory effect of SOD3 on *in vitro* migration (Figs. 6 and 8e).

## DISCUSSION

RNA helicases, which comprise more than 60 enzymes that use the energy of ATP hydrolysis to remodel RNA or RNA-protein complexes, are thought to be required at all stages of cellular RNA metabolism<sup>18,19,36</sup>. However, their target mRNAs and their cellular functions are poorly defined. In this context, whereas Ddx5 and Ddx17 have been shown to play a part in the splicing process, their effect on alternative splicing has been reported only for a few cases<sup>22–24</sup>. To the best of our knowledge, we provide the first evidence that RNA helicases can have a widespread impact on alternative splicing (Supplementary Table 2 and Supplementary Fig. 2). Notably, we observed that the 5' splicing sites of exons that are skipped upon Ddx5 and Ddx17 depletion are within GC-rich regions as compared to the 5' splicing sites of exons that are included following depletion (Supplementary Fig. 2b). Because we show that the helicase activity of Ddx5 and Ddx17 is critical for their ability to regulate mH2A1 splicing (Fig. 2e), this result suggests that these proteins regulate alternative splicing by affecting RNA secondary structure<sup>24</sup>.

Ddx5 and Ddx17 have been reported to have an impact on cellular phenotypes, owing to their functions as transcriptional coregulators<sup>18,19,37</sup>. For example, we recently showed that Ddx5 and Ddx17 are transcriptional coactivators of the NFAT5 promigratory transcription factor and that they participate in controlling the transcriptional activity of genes involved in cell migration and invasion, including *S100A4* (ref. 38). We now show that the splicing activity of Ddx5 and Ddx17 contributes to their role in cell migration and invasion, by modulating the splicing of the *mH2A1* gene, whose product in turn regulates genes involved in cell migration and invasion, such as *SOD3*. As NFAT5 and mH2A1 do not control the expression of the same genes (**Supplementary Fig. 8**), our data demonstrate that both the transcriptional and the splicing activities of Ddx5 and Ddx17 influence tumor progression by affecting different gene-expression programs. Notably, a recent report also demonstrates that Ddx5 affects the expression of another set of genes that are also involved in tumor progression (for example, cofilin and profilin) by modulating the expression of micro RNAs (miRNAs)<sup>39</sup>. In sum, these results demonstrate that the different activities of Ddx5 and Ddx17 in transcription, splicing and miRNA processing affect different gene-expression programs, which collectively contribute to their effects on tumor progression.

In agreement with a proposed role in cancer, several reports have shown that Ddx5 or Ddx17 are overexpressed in different tumor types<sup>18,19,40–42</sup>. Although we found that Ddx17 was overexpressed in the metastatic 4T1 cells and tumors (**Fig. 1a** and **Supplementary Fig. 1**), a high level of *Ddx17* (and/or *Ddx5*) mRNA did not correlate with a poor metastasis-free survival rate in the analyzed set of human tumors (data not shown). However, analyzing Ddx5 and Ddx17 expression in human tumors is very challenging because Ddx5 or Ddx17 alterations might occur at the protein level or at a post-translational modification level, as suggested by several reports<sup>43–47</sup>.

Our work also revealed a role for *mH2A1* histone pre-mRNA alternative splicing in breast cancer progression. Notably, it has been previously shown that the mH2A1.1 isoform inhibits lung cancer–cell proliferation and that patients with low mH2A1.1 levels in lung tumor samples are more likely to relapse than those with strong nuclear mH2A1.1 staining<sup>7,8</sup>. Similar results were recently reported in colon cancer<sup>48</sup>. Notably, the proportion of mH2A1.1 isoform (as a percentage of total mH2A1) decreased in tumors relative to normal samples in several cancer types including breast cancer<sup>8</sup>. In the tumor samples that we analyzed, we observed that the mH2A1 global level strongly correlates with the level of mH2A1.2 isoform (**Supplementary Fig. 7b**), which we showed favors cancer-cell invasiveness (**Fig. 3**). On the basis of the data obtained from cancers in lung<sup>7</sup>, colon<sup>48</sup> and breast (this study), one model is that the decrease in the mH2A1.1 isoform–expression level in tumor cells as compared to normal cells could favor cell proliferation, and overproduction of the mH2A1.2 isoform in tumor samples could further increase cell aggressiveness. Alteration of mH2A1 alternative splicing during tumor initiation and progression could be mediated by several splicing factors, as the QKI splicing factor, which is downregulated in cancer compared to normal cells, has also been shown to affect mH2A1 splicing<sup>8</sup>.

Notably, a recent study demonstrated that the loss of mH2A1 histone is positively correlated with an increasing malignant phenotype of melanoma cells in culture and human tissue samples and that the suppressive effects of mH2A1 on melanoma progression is mediated through the regulation of CDK8 (ref. 4). Because we did not observe a general downregulation of the *mH2A1* gene associated with breast cancer aggressiveness (data not shown) and because we did not find any change in *CDK8* gene expression in the 4T1 breast cancer

model of tumor progression (**Supplementary Fig. 6b**), these results suggest that mH2A1 expression might be altered in different ways depending on tumor types, which may result in alteration of different gene-expression programs.

In addition, the role of mH2A1 in cancer may also rely on post-translational modifications. In this context, it is important to underscore the fact that the macrodomain of the mH2A1.1 splicing isoform, but not that of the mH2A1.2 isoform, binds ADP-ribose and related NAD metabolites. It has been proposed that mH2A1.1 but not mH2A1.2 recognizes ADP-ribosylated chromatin components to promote rearrangement of chromatin loops<sup>28,29,49</sup>. This specific feature of the mH2A1.1 splicing isoform is particularly relevant with respect to our observation that mH2A1.1 and mH2A1.2 differentially regulate genes involved in redox metabolism. Indeed, ADP-ribosylation is a consequence of PARP activation under stress conditions, particularly in the context of redox metabolism alteration and exposure to ROS<sup>30,50–52</sup>. In addition, NAD metabolites have a central role in redox and ROS metabolism, and increasing evidence indicates that redox metabolism in turn affects epigenetic regulation<sup>30,50–52</sup>. Therefore, nonmetastatic 67NR cells that express a higher mH2A1.1-to-mH2A1.2 ratio than metastatic 4T1 cells might be able to adapt their genetic programs to redox metabolism by expressing genes such as *SOD3*, which would in turn affect cell invasiveness. Indeed, the *SOD3* gene codes for the extracellular superoxide dismutase that converts extracellular [O<sub>2</sub>]<sup>•−</sup> into H<sub>2</sub>O<sub>2</sub>. Notably, it has been recently shown that the overexpression of *SOD3* inhibits invasion by breast cancer cells<sup>32–35</sup>, which is consistent with the findings of this study. Indeed, by converting extracellular [O<sub>2</sub>]<sup>•−</sup> into H<sub>2</sub>O<sub>2</sub>, *SOD3* attenuates oxidative fragmentation of the extracellular-matrix components such as heparan sulfate, thereby decreasing invasion and migration, and it may prevent other effects of ROS on cell motility<sup>30,32–35,53–56</sup>. In support of a role for *SOD3* in tumors, we report that *SOD3* is an indicative factor for good prognosis (**Fig. 8g**). Notably, as well, it has been recently shown that the mH2A1.2 isoform but not the mH2A1.1 isoform can interact with and regulate HER2, which is a major player in breast cancer initiation and progression<sup>57</sup>. Therefore, further experiments are required to decipher the precise mechanisms by which mH2A1.1 and mH2A1.2 differentially contribute to cancer.

In conclusion, this study reveals a role of tumor-associated alternative splicing of an epigenetic factor in tumor progression, which leads in turn to transcriptional alterations of genes involved in a specific cellular program. Because epigenetic factors can also have an impact on alternative splicing (reviewed in ref. 58), it is likely that cross-talk between splicing and epigenetic networks may contribute to the control of cell phenotypes during both physiological and pathological states.

## METHODS

Methods and any associated references are available in the [online version of the paper](#).

**Accession codes.** Microarray data have been deposited in the Gene Expression Omnibus database under accession code [GSE40737](#).

*Note: Supplementary information is available in the [online version of the paper](#).*

## ACKNOWLEDGMENTS

We thank M. Buschbeck and V.A. Raker for critically reading the manuscript and L. Mhamdi for technical assistance. We thank C. Vanbelle and C. Boucharon from CeCILE - SFR Sante Lyon-Est for technical assistance and helpful discussion. This work was supported by the Institut National du Cancer (2008-1-PL BIO 01 to D.A. and S.V.), Agence Nationale de la Recherche (BLAN07-3\_186592 to D.A. and S.V.), Association d'Aide à la Recherche Cancérologique de Saint-Cloud (to R.L.)



and Fondation Recherche Médicale (Equipe FRM DEQ20110421278). S.G. was supported by Association pour la Recherche sur le Cancer; L.G. and S.P. by Agence Nationale de la Recherche; E.Z., H.M. and J.-P.V. by FRM; E.D. by Ligue Nationale Contre le Cancer; and M.P.E. by Association Française contre les Myopathies.

#### AUTHOR CONTRIBUTIONS

E.D., S.P., M.D., R.L., S.V. and D.A. designed the experiments; S.V. and D.A. supervised the project; E.D., S.P., K.D., L.G., M.L.-T., E.Z. and S.G. conducted the experiments; M.P.E., H.M. and J.-P.V. performed all bioinformatics analysis. All authors performed the data analysis and wrote the manuscript.

#### COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at <http://www.nature.com/doi/10.1038/nsmb.2390>.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

- Füllgrabe, J., Kavanagh, E. & Joseph, B. Histone onco-modifications. *Oncogene* **30**, 3391–3403 (2011).
- Portela, A. & Esteller, M. Epigenetic modifications and human disease. *Nat. Biotechnol.* **28**, 1057–1068 (2010).
- Dalvai, M. & Bystrycky, K. The role of histone modifications and variants in regulating gene expression in breast cancer. *J. Mammary Gland Biol. Neoplasia* **15**, 19–33 (2010).
- Kapoor, A. *et al.* The histone variant macroH2A suppresses melanoma progression through regulation of CDK8. *Nature* **468**, 1105–1109 (2010).
- Srivastava, N., Gochhait, S., de Boer, P. & Bamezai, R.N. Role of H2AX in DNA damage response and human cancers. *Mutat. Res.* **681**, 180–188 (2009).
- Valdés-Mora, F. *et al.* Acetylation of H2A.Z is a key epigenetic modification associated with gene deregulation and epigenetic remodeling in cancer. *Genome Res.* **22**, 307–321 (2012).
- Sporn, J.C. *et al.* Histone macroH2A isoforms predict the risk of lung cancer recurrence. *Oncogene* **28**, 3423–3428 (2009).
- Novikov, L. *et al.* QKI-mediated alternative splicing of the histone variant macroH2A1 regulates cancer cell proliferation. *Mol. Cell. Biol.* **31**, 4244–4255 (2011).
- Venables, J.P. *et al.* Cancer-associated regulation of alternative splicing. *Nat. Struct. Mol. Biol.* **16**, 670–676 (2009).
- Dutertre, M. *et al.* Exon-based clustering of murine breast tumor transcriptomes reveals alternative exons whose expression is associated with metastasis. *Cancer Res.* **70**, 896–905 (2010).
- David, C.J. & Manley, J.L. Alternative pre-mRNA splicing regulation in cancer: pathways and programs unhinged. *Genes Dev.* **24**, 2343–2364 (2010).
- Dutertre, M., Vagner, S. & Auboeuf, D. Alternative splicing and breast cancer. *RNA Biol.* **7**, 403–411 (2010).
- Stamm, S. *et al.* Function of alternative splicing. *Gene* **344**, 1–20 (2005).
- Pan, Q., Shai, O., Lee, L.J., Frey, B.J. & Blencowe, B.J. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.* **40**, 1413–1415 (2008).
- Witten, J.T. & Ule, J. Understanding splicing regulation through RNA splicing maps. *Trends Genet.* **27**, 89–97 (2011).
- Aslakson, C.J. & Miller, F.R. Selective events in the metastatic process defined by analysis of the sequential dissemination of subpopulations of a mouse mammary tumor. *Cancer Res.* **52**, 1399–1405 (1992).
- Bemmo, A. *et al.* Exon-level transcriptome profiling in murine breast cancer reveals splicing changes specific to tumors with different metastatic abilities. *PLoS ONE* **5**, e11981 (2010).
- Fuller-Pace, F.V. & Moore, H.C. RNA helicases p68 and p72: multifunctional proteins with important implications for cancer development. *Future Oncol.* **7**, 239–251 (2011).
- Janknecht, R. Multi-talented DEAD-box proteins and potential tumor promoters: p68 RNA helicase (DDX5) and its paralogs, p72 RNA helicase (DDX17). *Am. J. Transl. Res.* **2**, 223–234 (2010).
- Bates, G.J. *et al.* The DEAD box protein p68: a novel transcriptional coactivator of the p53 tumour suppressor. *EMBO J.* **24**, 543–553 (2005).
- Endoh, H. *et al.* Purification and identification of p68 RNA helicase acting as a transcriptional coactivator specific for the activation function 1 of human estrogen receptor alpha. *Mol. Cell. Biol.* **19**, 5363–5372 (1999).
- Höning, A., Auboeuf, D., Parker, M.M., O'Malley, B.W. & Berget, S.M. Regulation of alternative splicing by the ATP-dependent DEAD-box RNA helicase p72. *Mol. Cell. Biol.* **22**, 5698–5707 (2002).
- Guil, S. *et al.* Roles of hnRNP A1, SR proteins, and p68 helicase in c-H-ras alternative splicing regulation. *Mol. Cell. Biol.* **23**, 2927–2941 (2003).
- Kar, A. *et al.* RNA Helicase p68 (DDX5) regulates tau exon 10 splicing by modulating a stem-loop structure at the 5' splice site. *Mol. Cell. Biol.* **31**, 1812–1821 (2011).
- Thomas, P.D. *et al.* PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res.* **13**, 2129–2141 (2003).
- Gamble, M.J., Frizzell, K.M., Yang, C., Krishnakumar, R. & Kraus, W.L. The histone variant macroH2A1 marks repressed autosomal chromatin, but protects a subset of its target genes from silencing. *Genes Dev.* **24**, 21–32 (2010).
- Buschbeck, M. *et al.* The histone variant macroH2A is an epigenetic regulator of key developmental genes. *Nat. Struct. Mol. Biol.* **16**, 1074–1079 (2009).
- Timinszky, G. *et al.* A macrodomain-containing histone rearranges chromatin upon sensing PARP1 activation. *Nat. Struct. Mol. Biol.* **16**, 923–929 (2009).
- Gamble, M.J. & Kraus, W.L. Multiple facets of the unique histone variant macroH2A: from genomics to cell biology. *Cell Cycle* **9**, 2568–2574 (2010).
- Grek, C.L. & Tew, K.D. Redox metabolism and malignancy. *Curr. Opin. Pharmacol.* **10**, 362–368 (2010).
- Pani, G., Galeotti, T. & Chiarugi, P. Metastasis: cancer cell's escape from oxidative stress. *Cancer Metastasis Rev.* **29**, 351–378 (2010).
- Chaiswing, L. & Oberley, T.D. Extracellular/microenvironmental redox state. *Antioxid. Redox Signal.* **13**, 449–465 (2010).
- Mori, K., Shibamura, M. & Nose, K. Invasive potential induced under long-term oxidative stress in mammary epithelial cells. *Cancer Res.* **64**, 7464–7472 (2004).
- Teoh, M.L., Fitzgerald, M.P., Oberley, L.W. & Domann, F.E. Overexpression of extracellular superoxide dismutase attenuates heparanase expression and inhibits breast carcinoma cell growth and invasion. *Cancer Res.* **69**, 6355–6363 (2009).
- Svensk, A.M., Soini, Y., Paakko, P., Hiravikoski, P. & Kinnula, V.L. Differential expression of superoxide dismutases in lung cancer. *Am. J. Clin. Pathol.* **122**, 395–404 (2004).
- Abdelhaleem, M. Helicases: an overview. *Methods Mol. Biol.* **587**, 1–12 (2010).
- Yang, L., Lin, C. & Liu, Z.R. P68 RNA helicase mediates PDGF-induced epithelial mesenchymal transition by displacing Axin from  $\beta$ -catenin. *Cell* **127**, 139–155 (2006).
- Germann, S. *et al.* Dual role of the ddx5/ddx17 RNA helicases in the control of the pro-migratory NFAT5 transcription factor. *Oncogene* **618**, 1038–1048 (2012).
- Wang, D. *et al.* RNA helicase DDX5 regulates microRNA expression and contributes to cytoskeletal reorganization in basal breast cancer cells. *Mol. Cell Proteomics* **11**, M111.011932 (2012).
- Shin, S., Rossow, K.L., Grande, J.P. & Janknecht, R. Involvement of RNA helicases p68 and p72 in colon cancer. *Cancer Res.* **67**, 7572–7578 (2007).
- Mooney, S.M., Grande, J.P., Salisbury, J.L. & Janknecht, R. Sumoylation of p68 and p72 RNA helicases affects protein stability and transactivation potential. *Biochemistry* **49**, 1–10 (2010).
- Wortham, N.C. *et al.* The DEAD-box protein p72 regulates ER $\alpha$ -oestrogen-dependent transcription and cell growth, and is associated with improved survival in ER $\alpha$ -positive breast cancer. *Oncogene* **28**, 4053–4064 (2009).
- Yang, L., Lin, C. & Liu, Z.R. Phosphorylations of DEAD box p68 RNA helicase are associated with cancer development and cell proliferation. *Mol. Cancer Res.* **3**, 355–363 (2005).
- Carter, C.L., Lin, C., Liu, C.Y., Yang, L. & Liu, Z.R. Phosphorylated p68 RNA helicase activates Snail1 transcription by promoting HDAC1 dissociation from the Snail1 promoter. *Oncogene* **29**, 5427–5436 (2010).
- Causevic, M. *et al.* Overexpression and poly-ubiquitylation of the DEAD-box RNA helicase p68 in colorectal tumours. *Oncogene* **20**, 7734–7743 (2001).
- Jacobs, A.M. *et al.* SUMO modification of the DEAD box protein p68 modulates its transcriptional activity and promotes its interaction with HDAC1. *Oncogene* **26**, 5866–5876 (2007).
- Mooney, S.M., Goel, A., D'Assoro, A.B., Salisbury, J.L. & Janknecht, R. Pleiotropic effects of p300-mediated acetylation on p68 and p72 RNA helicase. *J. Biol. Chem.* **285**, 30443–30452 (2010).
- Sporn, J.C. & Jung, B. Differential regulation and predictive potential of MacroH2A1 isoforms in colon cancer. *Am. J. Pathol.* **180**, 2516–2526 (2012).
- Kustatscher, G., Hothorn, M., Pugieux, C., Scheffzek, K. & Ladurner, A.G. Splicing regulates NAD metabolite binding to histone macroH2A. *Nat. Struct. Mol. Biol.* **12**, 624–625 (2005).
- Krishnakumar, R. & Kraus, W.L. The PARP side of the nucleus: molecular actions, physiological outcomes, and clinical targets. *Mol. Cell* **39**, 8–24 (2010).
- Giansanti, V., Dona, F., Tillhon, M. & Scovassi, A.I. PARP inhibitors: new tools to protect from inflammation. *Biochem. Pharmacol.* **80**, 1869–1877 (2010).
- Cyr, A.R. & Domann, F.E. The redox basis of epigenetic modifications: from mechanisms to functional consequences. *Antioxid. Redox Signal.* **15**, 551–589 (2011).
- Acharya, A., Das, I., Chandhok, D. & Saha, T. Redox regulation in cancer: a double-edged sword with therapeutic potential. *Oxid. Med. Cell. Longev.* **3**, 23–34 (2010).
- Weaver, A.M. Regulation of cancer invasion by reactive oxygen species and Tks family scaffold proteins. *Sci. Signal.* **2**, pe56 (2009).
- Svineng, G., Ravuri, C., Rikardsen, O., Huseby, N.E. & Winberg, J.O. The role of reactive oxygen species in integrin and matrix metalloproteinase expression and function. *Connect. Tissue Res.* **49**, 197–202 (2008).
- Radisky, D.C. *et al.* Rac1b and reactive oxygen species mediate MMP-3-induced EMT and genomic instability. *Nature* **436**, 123–127 (2005).
- Li, X. *et al.* The atypical histone macroH2A1.2 interacts with HER-2 in cancer cells. *J. Biol. Chem.* **287**, 23171–23183 (2012).
- Luco, R.F., Allo, M., Schor, I.E., Kornblitt, A.R. & Misteli, T. Epigenetics in alternative pre-mRNA splicing. *Cell* **144**, 16–26 (2011).



## ONLINE METHODS

**Cell culture, treatment and transfection.** 67NR, 168FARN, 4T07 and 4T1 cells were kindly provided by F. Miller (Michigan Cancer Foundation, Detroit, Michigan, USA). Cells were cultured in DMEM supplemented with 10% FBS at 37 °C. Transient transfection of siRNAs (25 nM) was performed by using Lipofectamine RNAiMAX (Invitrogen) following the manufacturer's instructions. Cells were harvested 48 h after transfection. For rescue experiments, HeLa cells were transfected with siRNA targeting the *Ddx5* and *Ddx17* UTR regions. Twenty-four hours later, wild-type and either mutated *Ddx5* or *Ddx17* expression vectors were transfected by using Lipofectamine 2000 (Invitrogen). Cells were harvested 48 h later. siRNA sequences are provided in **Supplementary Table 4**. si*Ddx5-17* is an siRNA that targets a conserved sequence that simultaneously affects *Ddx5* and *Ddx17* (and not a mixture of si*Ddx5* and si*Ddx17*) (**Supplementary Table 4**).

**Minigene assay and PCR digestion.** *hH2AFY* exon 6, with 200 bp and 174 bp of flanking introns, was amplified by PCR with the primers 5'-GGAATTCATATGGAATTCCTAGTTTGTCTGC-3' and 5'-GGAATTCATATGGAATTCACACTTGGAAATGGC-3' that contained the NdeI restriction site. *hH2AFY* exon 7, with 106 bp and 212 bp of flanking introns, was amplified by PCR with the primers 5'-GGAATTCATATGGAATTCATGTGCCTGTGCGT-3' and 5'-GGAATTCATATGGAATTCATTTGTGAGCTGC-3' that contained the NdeI restriction sites. NdeI-digested PCR products were cloned into pTBminigene. Primers used for minigene assay are described in **Supplementary Table 4**.

**RNA preparation and RT-qPCR.** Total RNA was extracted by using TRIzol (Invitrogen). One microliter of Glycoblue (Ambion) was added before RNA precipitation with isopropanol. RNA (1 µg) was treated with DNase I (DNAfree, Ambion) for 30 min at 37 °C and reverse-transcribed (RT) by using SuperScript II and random primers (Invitrogen). Before PCR, all RT reaction mixtures were diluted to contain 2.5 ng/µl of initial RNA. PCR reactions were performed using 5 µl of the diluted cDNAs and GoTaq polymerase (Promega). Quantitative PCR (qPCR) was performed using 2.5 µl of the diluted cDNAs and SYBR Green I Master mix (Roche). Primer sequences are provided in **Supplementary Table 4**. The relative RNA levels were determined on the basis of the threshold cycle for each PCR product (Ct).

**ChIP.** Cells were cross-linked with 1% formaldehyde for 10 min and lysed in buffer containing 50 mM Tris-HCl, pH 7.5, 150 mM KCl, 5 mM EDTA, 1% NP40, 0.1% SDS, 0.5% sodium deoxycholate, 50 mM sodium fluoride and phosphatase and protease inhibitors. Chromatin was sheared by sonication with a Bioruptor (Diagenode) to generate ~200-bp DNA fragments and incubated with antibodies against mH2A1 (5 µg, Abcam cat. no. 37264), H3 (1 µg, Abcam cat. no. 1791) or control immunoglobulins (Santa Cruz) and magnetic beads overnight at 4 °C. Beads were washed five times with buffer containing 50 mM Tris-HCl, pH 7.5, 150 mM KCl, 1% NP40 and 0.25% sodium deoxycholate, and two times with TE buffer<sup>59</sup>. Antibody-bound chromatin was reverse cross-linked overnight at 65 °C and treated with proteinase K (Qiagen) before DNA purification on column (Qiagen). Purified DNA was diluted three times in water and analyzed by qPCR.

**Western blot analysis.** Protein extracts were obtained by using NP-40 buffer (50 mM Tris-HCl, pH 8, 0.4 M NaCl, 5 mM EDTA, pH 8, 1% NP40, 0.2% SDS and 1 mM DTT) and protease inhibitors (Sigma). Ten micrograms of total proteins were separated by SDS-PAGE, and western blotting was subsequently performed with antibodies against *Ddx5* (PAb204, Upstate/Millipore, 1:1,000), *Ddx17* (ab24601, Abcam, 1:250), mH2A1.1 (4160, Cell Signaling, 1:1,000), mH2A1.2 (4287, Cell Signaling, 1:1,000), mH2A1 (ab37264, Abcam, 1:1,000) and  $\beta$ -actin (Mab1501, Millipore, 1:1,000).

**Cell-migration, invasion and wound-healing assays.** For cell-migration assay, cells ( $4 \times 10^4$  for 67NR and  $2 \times 10^4$  for 4T1) were plated, in serum-free medium 48 h after transfection, in the upper chamber of 8-µm pore-size ThinCerts.

The bottom chamber was filled with complete medium. After 24 h, cells at the top side of the filter were removed with a cotton swab. The remaining cells at the bottom side were washed with PBS, fixed with 10% TCA and stained with Amido black. Cells were counted from three randomized fields per well under a Nikon inverted microscope. A cell-invasion assay was performed in similar conditions, except that wells were pre-coated with matrigel at a 1:3 dilution for 4T1 or a 1:12 dilution for 67NR (BD Biosciences). For wound-healing assays, 2 d after transfection, a wound space of a confluent cell monolayer was produced by scratching the plate with a p200 pipette tip. The percentage of migration was defined at different times (8, 16, or 24 h) by measuring the lengthwise migration three times with the program ImageJ. Mitomycin C was added 1 h before wounding and maintained during migration. Recombinant SOD protein (100 units; S9697, Sigma) were added after cells were plated in the upper chamber.

**Affymetrix exon array.** One microgram of total RNA purified with TRIzol was labeled with Ambion reagents and hybridized to Affymetrix GeneChip Mouse Exon 1.0 ST arrays. Affymetrix exon-array data were normalized with quantile normalization. The background correction was performed by using antigenomic probes. Only probes targeting exons annotated from full-length cDNA were used for analysis. Among these preselected probes, poor-quality probes (for example, probes labeled as 'cross-hybridizing' probes by Affymetrix) and probes with intensity signals that were too low when compared to antigenomic background probes with the same GC content were removed from the analysis. Only probes with a DABG *P* value of  $\leq 0.05$  in at least half of the chips were considered for further statistical analysis. Paired statistical analyses were performed by using the Student's paired *t*-test on the splicing index to analyze the exon-array data. The splicing index corresponds to a comparison of gene-normalized exon intensity values between the two analyzed experimental conditions.

**Clinical samples, quantitative PCR and statistical analysis.** Samples of 443 breast cancer tumors excised from women at the Institut Curie/Hôpital René Huguenin (Saint-Cloud, France) from 1978 to 2008 were analyzed. This study was approved by the local ethical committee (Breast Group of Hôpital René Huguenin), which provided informed consent. Standard prognostic factors are reported in **Supplementary Figure 4**. Immediately after surgery, the tumor samples were flash frozen and stored in liquid nitrogen until RNA extraction. The samples were examined histologically for the presence of at least 60% tumor cells. The patients (mean age, 61.7 yrs; range, 35–91 yrs) all met the criteria of having primary unilateral breast carcinoma for which complete clinical, histological and biological data were available; 378 patients received adjuvant therapy, consisting of chemotherapy alone in 90 cases, hormone therapy alone in 175 cases and both treatments in 93 cases. The median follow-up was 8.9 years (range 0.5–29.0). A total of 168 patients relapsed at distant sites.

All RT-qPCR reactions were performed by using an ABI Prism 7700 Sequence Detection System and the SYBR Green PCR Core Reagents kit (PerkinElmer Applied Biosystems). TATA-box-binding protein (TBP) transcripts were used as an endogenous RNA control, and each sample was normalized on the basis of its TBP content. To determine whether the mH2A1 splicing-variant ratio was associated with patient clinical outcome, metastasis-free survival distribution was estimated by the Kaplan-Meier method. To select cutoff expression levels to classify each patient in one of two risk groups, sensitivity and specificity was explored by using a receiver-operating curve (ROC) analysis. Briefly, the area under the ROC curve (with 95% confidence interval) was calculated, and a test for the null hypothesis that the area under the curve was 50% was performed. The ROC analysis provided the threshold expression value to balance sensitivity and specificity for the detection of life-threatening cancer, and this cut point was used in the Kaplan-Meier analysis. Kaplan-Meier curves estimate metastasis-free survival from 0 to 180 months after breast cancer diagnosis. The Spearman rank correlation test provided correlation between continuous variables.

59. Bittencourt, D. *et al.* Cotranscriptional splicing potentiates the mRNA production from a subset of estradiol-stimulated genes. *Mol. Cell. Biol.* **28**, 5811–5824 (2008).

## B Publication dans NAR

Dans cet article publié l'année dernière, nous montrons que les hélicases ARN DDX5 et DDX17 régulent un large nombre de gènes cibles des voies de signalisation de réponse aux hormones, à la fois au niveau transcriptionnel et au niveau de l'épissage alternatif. D'autre part, DDX5 et DDX17 apparaissent aussi être des régulateurs de l'expression des récepteurs nucléaires dont elles sont co-activateurs.

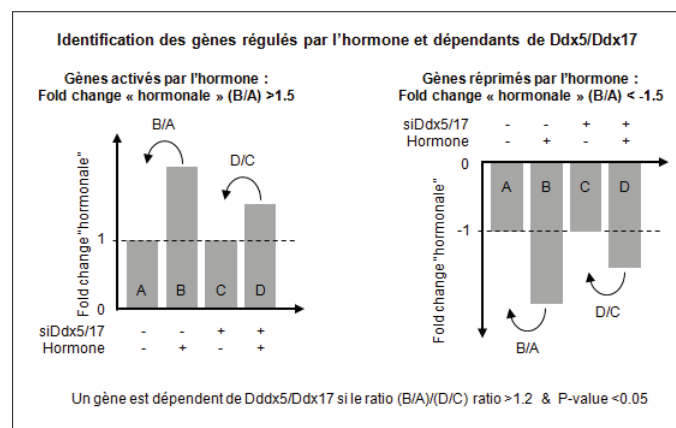
D'un point de vue bioinformatique, l'objectif de ce projet était de déterminer à large échelle l'impact de la déplétion des facteurs DDX5 et DDX17 lors d'un processus biologique complexe : l'activation des voies de réponse aux hormones dans un modèle de cancer du sein (voie ER $\alpha$ ) et du cancer de la prostate (voie AR).

Quatre conditions différentes de puces, avec quatre réplicats expérimentaux à chaque fois, devaient donc être analysé :

- condition A : sans traitement, transfection siARN contrôle
- condition B : avec traitement, transfection siARN contrôle
- condition C : sans traitement, transfection siARN contre DDX5 et DDX17
- condition D : avec traitement, transfection siARN contre DDX5 et DDX17

Nous avons donc mis au point un algorithme dédié permettant le pré-traitement et l'analyse de quatre conditions à la fois. Cet algorithme permet de déterminer les gènes et événements d'épissage régulés par un traitement aux hormones et dépendant des hélicases DDX5/DDX17.

La démarche général de cette algorithme est présenté ci-dessous :



# The Ddx5 and Ddx17 RNA helicases are cornerstones in the complex regulatory array of steroid hormone-signaling pathways

Samaan Samaan<sup>1,2,3,4</sup>, Léon-Charles Tranchevent<sup>2,3,4,5</sup>, Etienne Dardenne<sup>2,3,4,5</sup>, Micaela Polay Espinoza<sup>2,3,4,5</sup>, Eleonora Zonta<sup>1,2,3,4</sup>, Sophie Germann<sup>2,3,4,5</sup>, Lise Gratadou<sup>2,3,4,5</sup>, Martin Dutertre<sup>2,3,4,5</sup> and Didier Auboeuf<sup>2,3,4,5,\*</sup>

<sup>1</sup>Université de Paris Diderot-Paris 7, F-75013 Paris, France, <sup>2</sup>Inserm U1052, F-69008 Lyon, France, <sup>3</sup>CNRS UMR5286, F-69008 Lyon, France, <sup>4</sup>Centre de Recherche en Cancérologie de Lyon, 69008 Lyon, France and <sup>5</sup>Université de Lyon 1, F-69100 Villeurbanne, France

Received August 30, 2013; Revised October 25, 2013; Accepted November 4, 2013

## ABSTRACT

**Estrogen and androgen receptors (ER and AR) play key roles in breast and prostate cancers, respectively, where they regulate the transcription of large arrays of genes. The activities of ER and AR are controlled by large networks of protein kinases and transcriptional coregulators, including Ddx5 and its highly related paralog Ddx17. The Ddx5 and Ddx17 RNA helicases are also splicing regulators. Here, we report that Ddx5 and Ddx17 are master regulators of the estrogen- and androgen-signaling pathways by controlling transcription and splicing both upstream and downstream of the receptors. First, Ddx5 and Ddx17 are required downstream of ER and AR for the transcriptional and splicing regulation of a large number of steroid hormone target genes. Second, Ddx5 and Ddx17 act upstream of ER and AR by controlling the expression, at the splicing level, of several key regulators of ER and AR activities. Of particular interest, we demonstrate that Ddx5 and Ddx17 control alternative splicing of the GSK3 $\beta$  kinase, which impacts on both ER and AR protein stability. We also provide a freely available online resource which gives information regarding splicing variants of genes involved in the estrogen- and androgen-signaling pathways.**

## INTRODUCTION

The sex steroid hormones, estrogen and testosterone, influence normal physiology, reproduction and behavior. Their biological functions are mediated through cognate nuclear receptors that govern gene expression in hormone-

sensitive tissues. Many lines of evidence have implicated steroid hormones as etiologic factors in the origin and progression of various malignancies (1,2). Perturbation of the estrogen-signaling pathway is associated with two-thirds of breast cancers that express the estrogen receptor alpha (ER $\alpha$ ), which is considered as a good prognosis marker. The androgen receptor (AR) is activated by the binding of testosterone or its physiologically active metabolite, 5 $\alpha$ -dihydrotestosterone (DHT) and is involved in prostate cancer initiation and metastasis.

ER $\alpha$  and AR are members of the large superfamily of nuclear receptors and act as ligand-activated transcription factors. The canonical model of steroid receptor action implies a ligand-specific conformational change triggering its phosphorylation, homodimerization and binding to hormone responsive elements located in promoters or regulatory regions of target genes (3). Activation of the estrogen and androgen pathways requires the concerted action of a plethora of factors. Some of them are involved in posttranslational modifications of the hormone receptors, impacting for example on their subcellular localization or stability (4–6). Other factors, the so-called transcriptional coregulators, are recruited by hormone receptors on target promoters to mediate their effects on transcription (7,8). For both ER $\alpha$  and AR, dozens of coregulators, including coactivators and corepressors, have been identified. However, most of them have been studied in the context of a few target genes, and therefore it is currently unknown whether they contribute to the hundreds of gene regulations induced by hormones.

Among these, the DEAD-box RNA helicase Ddx5 (p68) and its highly related paralog, Ddx17 (p72), are transcriptional coregulators of ER $\alpha$  and AR (9,10). Ddx5 and Ddx17 interact directly with ER $\alpha$  and AR and were shown in a few cases to be recruited to target promoters (11,12) where they might modulate RNA polymerase II recruitment. However, the extent of ER $\alpha$  and AR endogenous

\*To whom correspondence should be addressed. Tel: +334 26 55 67 46; Fax: +334 78 78 27; Email: didier.auboeuf@inserm.fr

target genes on which Ddx5/Ddx17 act as transcriptional coregulators is not known. In addition to being transcriptional coregulators, Ddx5 and Ddx17 are *bona fide* components of the splicing machinery, the spliceosome (13), and play a role in the regulation of alternative splicing that leads to the production from the same gene of several splicing variants coding for different protein isoforms with different and sometimes opposite biological activities (14–17). Alternative splicing is the rule, not the exception, as 90% of human genes produce several splicing variants. Alternative splicing is the main mechanism increasing the diversity of the proteome coded by a limited number of genes (18). In this context, the Ddx5 and Ddx17 multifunctional proteins could coordinate transcription and splicing allowing the production of the proper isoform from hormone target genes as previously suggested by using minigene reporter assay (19). However, whether Ddx5 and Ddx17 regulate splicing of endogenous hormone target genes is not known.

In this work, we report using large-scale approaches for first time that Ddx5 and Ddx17 are master regulators of the estrogen and androgen-signaling pathways. Indeed, these proteins are not only required for regulating the expression of a large number of endogenous estrogen- and androgen-target genes both at the transcriptional and splicing level but remarkably, they are also acting upstream of the estrogen and androgen receptors by controlling the expression, at the splicing level, of several key regulators of the hormone-signaling pathways.

## MATERIALS AND METHODS

### Cell culture and stable cell lines

MCF-7 cells were grown in DMEM and LNCaP cells in RPMI-1640 (Gibco/LifeTechnologies). Both mediums were supplemented with 10% fetal bovine serum, 1% glutamine and 1% penicillin/streptomycin. Cells were maintained at 37°C in a humidified atmosphere of 95% air and 5% CO<sub>2</sub>. Wild-type Ddx5-HA and Ddx17-HA and mutated Ddx5-HA K144A and Ddx17-HA K142R were cloned into pTRE2-hyg vectors to generate inducible MCF-7/Tet-On stable cell lines (Clontech). Resistant clones were selected with hygromycin (300 mg/ml, Clontech) and protein expression was checked after Doxycycline treatment (1 µg/ml) for 48 h.

### Cell transfection and treatment

A total of  $3 \times 10^6$  cells were transiently transfected with 26.6 nM siRNAs (Supplementary Table S3) using Lipofectamine RNAiMax (LifeTechnologies). Twenty-four hours before treatment, MCF-7 and LNCaP cells, were cultured in red phenol-free medium supplemented with 2% charcoal-treated FBS. Cells were treated with E2 (10 nM; Sigma) for 1 or 10 h or DHT (10 nM; Sigma) for 24 h. Control cells received equal volumes of vehicle (ethanol).

### RNA preparation, RT-PCR and RT-qPCR

Total RNAs were prepared using TRIpure Isolation Reagent (Roche), and 1 µl of Glycoblue (Ambion) was

added before RNA precipitation. Nuclear fractionation was performed as previously described (20). Reverse transcription (RT) was performed with 1–2 µg of total RNA using M-MLV Reverse Transcriptase (LifeTechnologies) and random primers. The RT reactions were diluted and used either for PCR analysis using GoTaq Flexi DNA Polymerase (Promega) or in qPCR using SYBR Green I Master Mix (Roche) on a Roche LightCycler 480 II. Primer sequences are provided in Supplementary Table S3. The relative RNA levels were determined on the basis of the threshold cycle ( $C_t$ ) for each qPCR product and normalized to 18S ribosomal RNA levels.

### Affymetrix exon array

One microgram of total RNA was processed with the GeneChip WT Sense Target Labeling kit and hybridized to GeneChip Human Exon 1.0 ST arrays. Affymetrix exon-array data were normalized with quantile normalization. Antigenomic probes were used to perform the background correction. Only probes targeting exons annotated from full-length cDNA were retained for analysis. Cross hybridizing probes and probes with lower signals intensity than anti-genomic background probes showing the same GC content were removed. Only probes with a DABG  $P$ -value  $\leq 0.05$  in at least half of the arrays were considered for further statistical analysis. Arrays were performed in four independent replicates. The strategies adopted to identify E2-, DHT-regulated and Ddx5/17-dependent genes are described in Supplementary Figure S1. Briefly, the median intensity of all constitutive exonic probes was calculated for each gene in each sample, and the experimental samples and control groups were compared using a Student's paired  $t$ -test. Paired statistical analyses were performed using the Student's paired  $t$  test on the splicing index (SI) to analyze the Exon Array data at the exon level (SI > 1.5,  $P < 0.05$ ). The SI corresponds to a comparison of gene-normalized exon intensity values between the two analyzed experimental conditions.

### Western blot analysis

Total protein extracts were obtained using NP-40 buffer (50 mM Tris-HCl pH 8, 0.4 M NaCl, 5 mM EDTA pH 8, 1% NP40, 0.2% SDS, 1 mM DTT) supplemented with Protease and Phosphatase Inhibitors (Roche). An amount of 20 µg of proteins were separated by NuPAGE<sup>®</sup> Novex 3–8% Tris-Acetate or 12% bis-tris-Acetate Gels (LifeTechnologies). Membranes were incubated with specific primary antibodies against Ddx5 (ab10261, Abcam), Ddx17 (ab24601 Abcam), SMRT (H-300, sc-20778 Santa Cruz), AR (441, sc-7305), ER $\alpha$  (F-10, sc-8002 Santa Cruz), GSK3 $\beta$  (27C10, Cell signaling) and  $\beta$ -Actin (I-19, sc-1616 Santa Cruz).

### Chromatin-immunoprecipitation

ER $\alpha$  chromatin-immunoprecipitation (ChIP) assay was performed in MCF-7 cells using anti-ER $\alpha$  antibody (HC-20, sc-543 Santa Cruz). Ddx5-HA ChIP assay was performed in inducible MCF-7 cell line as previously described (21) using anti-HA antibody (3F10, Roche).



MCF-7 or MCF-7 inducible cell lines were maintained in red phenol-free medium supplemented with 2% charcoal-treated FBS. Cells were treated with E2 (10 nM; Sigma) for 1 h before fixation. Primers used are detailed in [Supplementary Table S3](#).

## RESULTS

### Ddx5 and Ddx17 are master regulators of the estrogen-signaling pathway

In order to assess the extent to which Ddx5 and Ddx17 participate in gene expression regulation in response to estradiol (E2), ER-positive MCF-7 breast cancer cells were treated with E2 or vehicle for 10 h after being transfected with a control siRNA (siCTRL) or an siRNA (siDdx5/17) targeting a conserved region shared by Ddx5 and Ddx17 ([Figure 1A](#)). Whole-transcriptome analysis was performed using Affymetrix Exon Arrays, which allowed us to assess expression at the global gene level as well as at the exon level. The median intensity of exonic probes of each gene was computed in each sample from four independent experiments. Using cutoffs of 1.5 for fold change and 0.05 for *P*-value, 354 (67%) and 173 (33%) genes were predicted to be activated and repressed by E2, respectively ([Figure 1B](#)). Comparative analysis of control and Ddx5/17-depleted cells ([Supplementary Table S1](#)) revealed that Ddx5 and Ddx17 are required for the regulation of 186 (53%) of E2-activated and 98 (57%) of E2-repressed genes ([Figure 1B](#) and [Supplementary Table S1](#)) as validated by RT-qPCR for a large number of cases ([Figure 1C](#)). Similar results were observed in a subset of selected genes using another set of siRNAs targeting Ddx5 and Ddx17 ([Figure 1D](#) and [Supplementary Figure S2A](#)).

The observed effects likely occurred at the transcriptional level as Ddx5/17 depletion abrogated E2-mediated effects at the pre-mRNA level and did not significantly affect the half-life of the tested E2-regulated mRNAs ([Figure 1E](#) and [Supplementary Figure S2C](#)). Furthermore, ER $\alpha$  and Ddx5 were detected on a subset of E2-target genes and their recruitment was increased after E2 treatment as demonstrated by ChIP assay ([Figure 1F](#)). These results are in agreement with previous reports (9,11,22,23) indicating that Ddx5 and Ddx17 are *bona fide* ER $\alpha$  transcriptional coregulators and are therefore required for mediating E2 effects on transcription. However, we observed that Ddx5/17 depletion also decreased ER $\alpha$  protein (but not mRNA) levels ([Figure 1G](#)), suggesting that Ddx5 and Ddx17 may also contribute to the estrogen-signaling pathway by another mechanism (see below). Of note, Ddx5/17 depletion did not affect other signaling pathways ([Supplementary Figure S2D](#)).

In addition to their role in transcription, Ddx5 and Ddx17 have been shown to play a role in alternative splicing regulation (14–17). The analysis of Exon Array data revealed that Ddx5 and Ddx17 silencing changed the splicing pattern of 65 genes among the E2-regulated genes ([Figure 2A](#) and [Supplementary Table S1](#)), as validated by RT-PCR ([Figure 2B](#)). Therefore, Ddx5 and Ddx17 are not only required for the E2-mediated effects on

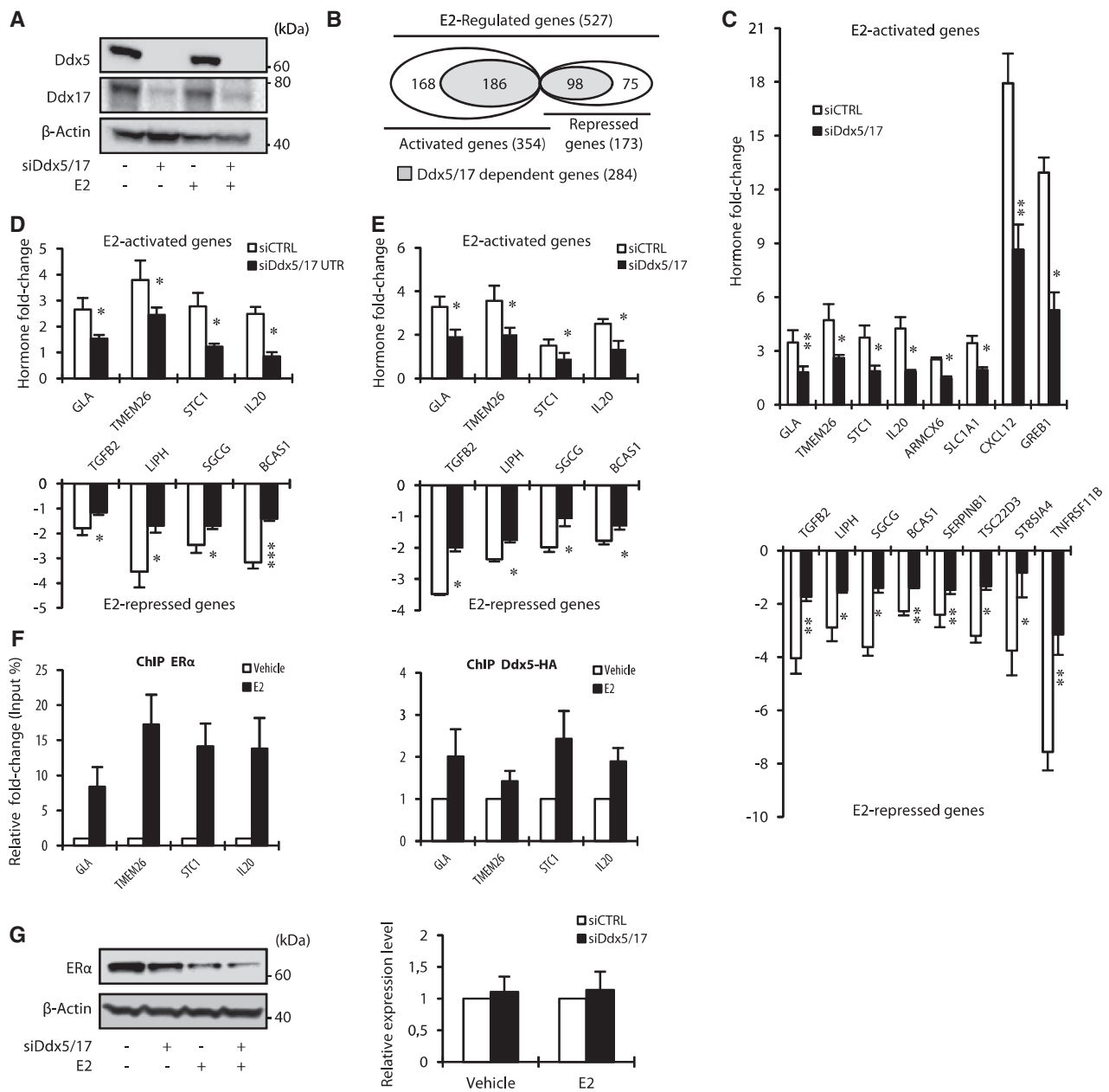
transcription, but also for the production of specific isoforms from, at least, a subset of endogenous E2-regulated genes. However, as most of the genes regulated at the splicing level by Ddx5 and Ddx17 were not regulated at the transcriptional level by E2 treatment ([Figure 2A](#)), this suggested that Ddx5/17 effects on splicing and E2 effects on transcription were mainly independent.

### Ddx5 and Ddx17 regulate splicing of several genes involved in the estrogen-signaling pathway

Among 193 genes encoding for mediators or regulators of the estrogen-signaling pathway, 26 were predicted to be regulated by Ddx5/17 at the splicing level ([Figure 2C](#) and [Supplementary Figure S3A](#)). For example, several genes coding for protein kinases involved in ER $\alpha$  phosphorylation, including CDK2, p38 MAPK (ERK2) and GSK3 $\beta$ , were affected at the splicing level when Ddx5 and Ddx17 were knocked down ([Figure 2D](#) and [Supplementary Figure S3A](#)). Likewise, several ER $\alpha$  transcriptional coregulators were affected at the splicing level by Ddx5/17 silencing ([Figure 2D](#) and [Supplementary Figure S3A](#)). This included transcriptional coactivators like CREB-binding protein (CBP) and the Mediator Subunit 1 (MED1) that plays a crucial role to anchor liganded ER $\alpha$  to the transcriptional pre-initiation complex (24), as well as several corepressors such as the nuclear corepressor 1 (N-CoR1) and its paralog N-COR2, also known as SMRT (Silencing Mediator of Retinoic acid and Thyroid hormone receptor) (25). The identity of some splicing variants was verified by sequencing ([Supplementary Figure S4](#)).

To further confirm the effects of Ddx5 and Ddx17 on splicing, inducible stable MCF-7 cell lines expressing wild type or helicase-mutated Ddx5 were transfected with a control siRNA or a mixture of siRNAs targeting *DDX5* and *DDX17* UTRs. The re-expression of wild-type, but not mutated Ddx5, partially rescued the effect of the siDdx5/17 UTR on the splicing pattern ([Figure 2E](#), left panel). Similar results were obtained in inducible stable MCF-7 cell lines expressing wild-type Ddx17 or the mutant K142R lacking the helicase activity ([Figure 2E](#), right panel). This observation raised the possibility that Ddx5 and Ddx17 contributed to the ER $\alpha$ -signaling pathway not only as ER $\alpha$  transcriptional coregulators but also by regulating the splicing pattern of key actors of the pathway.

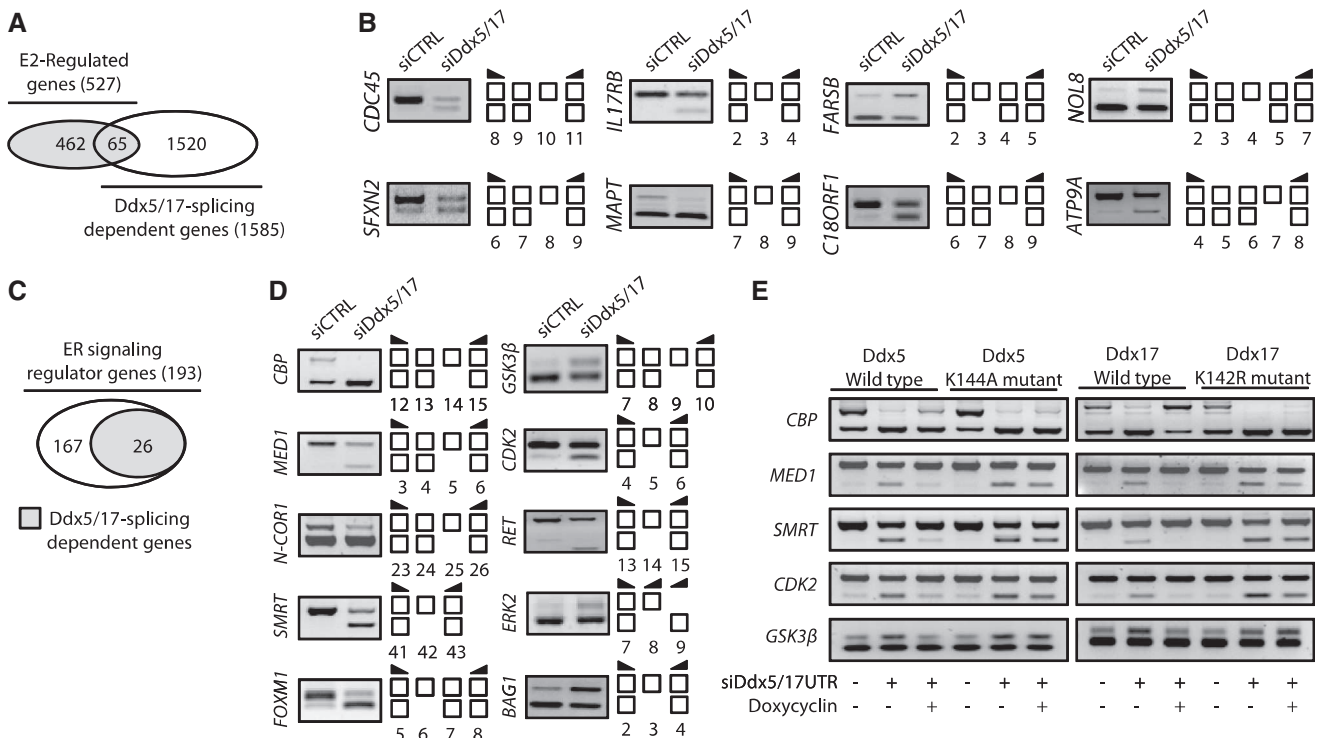
To address the functional relevance of splicing regulation of key actors of the ER $\alpha$ -signaling pathway by Ddx5 and Ddx17, we focused on the SMRT transcriptional corepressor. As shown in [Figure 3A](#), RT-qPCR analysis confirmed the RT-PCR data ([Figure 2D](#)) and showed that Ddx5/17 depletion induced exon 42 skipping (SMRT $\Delta$ E42). Interestingly, exon 42 skipping is predicted to introduce a premature stop codon in exon 43 ([Figure 3B](#), upper panel), suggesting that SMRT $\Delta$ E42 transcripts might be unproductive (that is, not translated) and regulated by the nonsense-mediated mRNA decay (NMD) pathway. Supporting this hypothesis, incubation of MCF-7 cells with the translation inhibitor cycloheximide, that blocks mRNA degradation by the NMD pathway, increased the ratio of SMRT $\Delta$ E42



**Figure 1.** Ddx5 and Ddx17 are master regulators of the estrogen-signaling pathway. (A) Western blot analysis of Ddx5, Ddx17 and  $\beta$ -Actin as loading control, using total protein extract from MCF-7 cells transfected with either a control siRNA or siDdx5/17 in the presence or absence of estradiol (E2) for 10 h. (B) *In silico* prediction of Exon array analysis of genes regulated at the global expression level by E2 and dependent of Ddx5 and Ddx17 (in gray). (C) Hormone fold change induced by E2 treatment for 10 h of the expression level of a subset of mRNAs as determined by RT-qPCR in MCF-7 cells transfected with siCTRL or siDdx5/17 (genes upregulated by E2 in the upper panel and genes downregulated by E2 in the lower panel). (D) Hormone fold change induced by E2 treatment for 10 h of the expression level of a subset of mRNAs as determined by RT-qPCR in MCF-7 cells transfected with siCTRL or a mixture of siRNAs-targeting Ddx5 and Ddx17 UTRs (siDdx5/17 UTR). (E) Hormone fold change as assessed by RT-qPCR measuring the unspliced pre-mRNAs of E2-regulated genes in the nuclear extracts. MCF-7 cells were transiently transfected with Control siRNA or siDdx5/17 then treated with E2 or vehicle for 1 h. (F) qPCR analysis using primers spanning promoter regions of a subset of E2-regulated genes on genomic DNA in a ChIP assay using MCF-7 for ER $\alpha$  (left panel) and MCF-7 stable cell lines expressing Ddx5-HA (right panel). (G) Western blot analysis of ER $\alpha$  and  $\beta$ -Actin as loading control using MCF-7 cells transfected with either a control siRNA or siDdx5/17 in the presence or absence of estradiol (E2) for 10 h (left panel). Relative expression level of ER $\alpha$  mRNA as assessed by RT-qPCR analysis (right panel). Histograms represent the average of at least three independent experiments. Error bars, s.e.m.; \* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$  (*t*-test).

transcripts to transcripts containing exon 42 (SMRT + E42) (Figure 3B, lower panel). In addition and as expected from the increased NMD-degraded SMRT $\Delta$ E42 production, Ddx5/17 depletion induced a

decrease of SMRT protein expression levels (Figure 3C, left panel). Finally, MCF-7 cells transfected with siRNA targeting SMRT mRNA showed a significantly reduced E2 effect on a subset of downregulated transcripts



**Figure 2.** Ddx5 and Ddx17 regulate splicing of a subset of E2-target genes and genes involved in estrogen-signaling pathway. (A) *in silico* prediction of Exon array analysis. Venn diagram representing E2-regulated genes and genes that were regulated by Ddx5 and Ddx17 at the splicing level. (B) Validation by RT-PCR of splicing variants produced from E2-target genes and regulated by Ddx5 and Ddx17. Exons are represented by white boxes and primers used for PCR by black arrows. (C) Diagram representing ER-signaling regulators genes that were regulated by Ddx5 and Ddx17 at the splicing level. (D) Validation by RT-PCR of splicing events regulated by Ddx5/17. MCF-7 cells were transfected for 48 h with a Control siRNA (siCTRL) or siDdx5/17, then treated for 10 h with E2 (10 nM). Depletion of Ddx5 and Ddx17 induced both exon inclusion and exon skipping. (E) Rescue experiments in inducible stable cell lines. Re-expression of wild-type or mutated Ddx5 (left panel) and Ddx17 (right panel) forms in MCF-7 inducible stable cell lines after transfection with siRNAs targeting *DDX5* and *DDX17* UTRs was induced by Doxycyclin (1  $\mu$ g/ml) treatment.

compared with control cells (Figure 3C, right panel and 3D). Altogether, these data indicated that the role of Ddx5 and Ddx17 in the regulation of the estrogen-signaling pathway is not limited to their function as ER transcriptional coregulators on estrogen-regulated genes, but also pointed to a previously undocumented regulation of coregulated splicing events directly linked to this pathway.

### Ddx5 and Ddx17 play a major role in the androgen-signaling pathway

As for ER $\alpha$ , Ddx5 and Ddx17 are transcriptional coregulators of AR. To study the genome-wide contribution of Ddx5 and Ddx17 to the androgen-signaling pathway, we applied the same experimental strategy using the human LNCaP prostate cancer cell line that was treated with DHT for 24 h (Figure 4A). Affymetrix Exon array analysis predicted that out of a total of 1573 DHT-regulated genes, 1319 (84%) were activated and 254 (16%) repressed (Figure 4B). *In silico* analysis predicted that 868 (66%) of DHT-activated and 103 (40%) of DHT-repressed genes were regulated in a Ddx5/17-dependent manner (Figure 4B and Supplementary Table S2), as validated by RT-qPCR (Figure 4C).

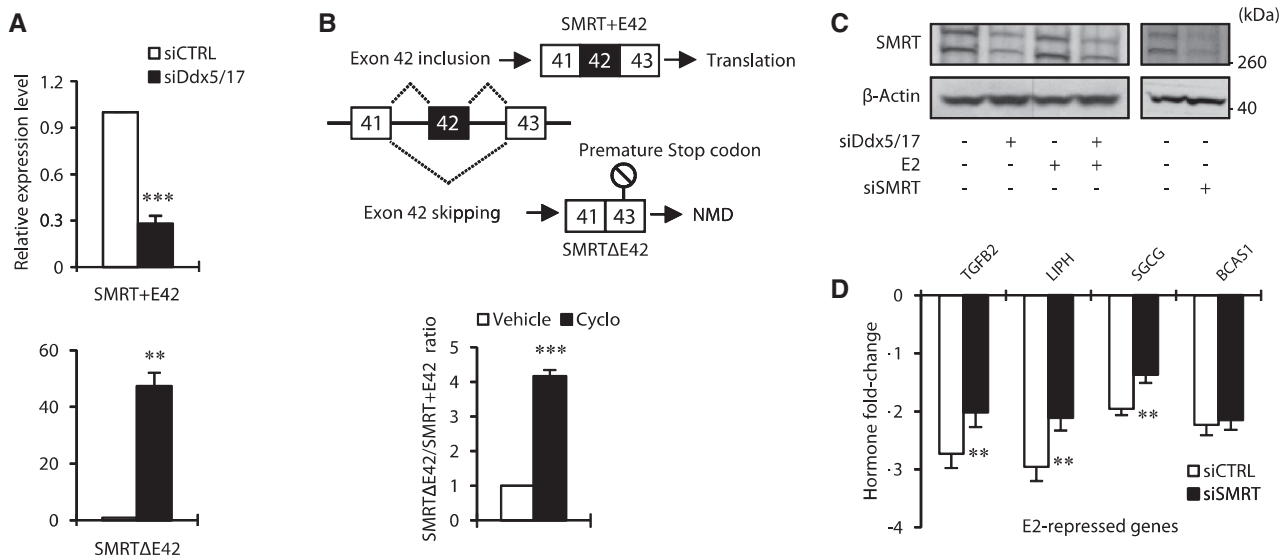
When analyzing Exon Array data at the exon level, we identified 371 DHT-regulated genes (among 1573) whose

splicing was affected by Ddx5/17 depletion (Figure 4D and Supplementary Table S2), as validated by RT-PCR (Figure 4E). Finally, as shown in Figure 4F, several (24 out of 144) Ddx5/17-regulated splicing variants are produced from genes involved in the regulation of the androgen-signaling pathway (Figure 4G and Supplementary Figure S3B).

Of particular interest, and similarly to our observation on ER $\alpha$ , Ddx5/17 depletion in LNCaP cells, reduced AR protein (but not mRNA) level (Figure 4H). Therefore, Ddx5 and Ddx17 likely contribute to both estrogen and androgen-signaling pathways through several mechanisms.

### Ddx5 and Ddx17 stabilize ER $\alpha$ and AR by modulating GSK3 $\beta$ splicing

Comparing the data obtained in both MCF-7 and LNCaP cells, we observed that several genes regulated at the splicing level by Ddx5/17 are involved in both ER $\alpha$  and AR-signaling pathways (e.g. CBP, MED1, N-COR1, SMRT and GSK3 $\beta$ ) (Figure 2D, 4G and Supplementary Figure S3). Interestingly, it has been shown that the GSK3 $\beta$  kinase phosphorylates and stabilizes both ER $\alpha$  (5,26) and AR (6,27). Two GSK3 $\beta$  isoforms, resulting from exon 9 alternative splicing, have been reported in rodents (28,29) and human (30). The conventional



**Figure 3.** Ddx5 and Ddx17 control SMRT protein expression level. (A) RT-qPCR analysis measuring SMRT splicing variant expression level in MCF-7 cells transfected for 48 h with a control siRNA (siCTRL) or siDdx5/17. Specific primers were designed to amplify transcripts containing or not exon 42, SMRT + E42 and SMRTΔE42, respectively. (B) Upper panel, skipping of exon 42 results in a premature stop codon in exon 43. Lower panel, RT-qPCR analysis measuring SMRT splicing variant ratio in MCF-7 cells after 6 h of cycloheximide treatment (10 μg/ml). (C) Western blot analysis of SMRT protein expression level and β-Actin as a loading control in MCF-7 cells transfected with a control siRNA, siDdx5/17 or siRNA-targeting SMRT mRNA. (D) Hormone fold change induced by E2 treatment for 10 h of the expression level of a subset of mRNAs as determined by RT-qPCR in MCF-7 cells transfected with siCTRL or siSMRT. Histograms represent the average of at least three independent experiments. Error bars represent s.e.m.; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$  ( $t$  test).

isoform GSK3β1 that regulates ERα and AR protein stability does not contain exon 9 in contrast to the GSK3β2 isoform (Figure 5A, upper panel). Interestingly, exon 9 encodes for a 13 amino acid peptide within the kinase catalytic domain, which reduces the kinase activity of the GSK3β2 isoform compared with the canonical GSK3β1 isoform (28,30,31).

We first validated the effect of Ddx5/17 depletion on GSK3β exon 9 splicing by RT-qPCR analysis (bottom panel of Figure 5A). This confirmed that in the absence of Ddx5/17, exon 9 inclusion augments, indicating that the production of GSK3β2 splicing variant is increased while the GSK3β1 splicing variant is decreased in MCF-7 cells. Accordingly, a western blot analysis showed that Ddx5/17 silencing resulted in the appearance of a slower migrating protein band detected by the GSK3β antibody (Figure 5B), consistent with the detection of the GSK3β2 isoform (32), while the GSK3β1 isoform decreased.

As previously reported (5,26,33), GSK3β depletion and pharmacological inhibition resulted in ERα protein level decrease (Figure 5C). Remarkably, Ddx5/17 depletion had the same effect (Figure 5C) and this effect was further enhanced in MCF-7 cells transfected with specific siRNAs (Figure 5D, upper panel) targeting the GSK3β1 but not the GSK3β2 isoform (Figure 5D, bottom panel). Collectively these results support a model in which Ddx5/17 silencing reduces the production of the canonical GSK3β1 isoform, resulting in a decreased ERα protein level. Similar results were obtained in LNCaP cells: Ddx5/17 depletion decreased the production of the conventional GSK3β1 isoform (Figure 5E, upper panel) as well as AR protein level, as did GSK3β depletion or

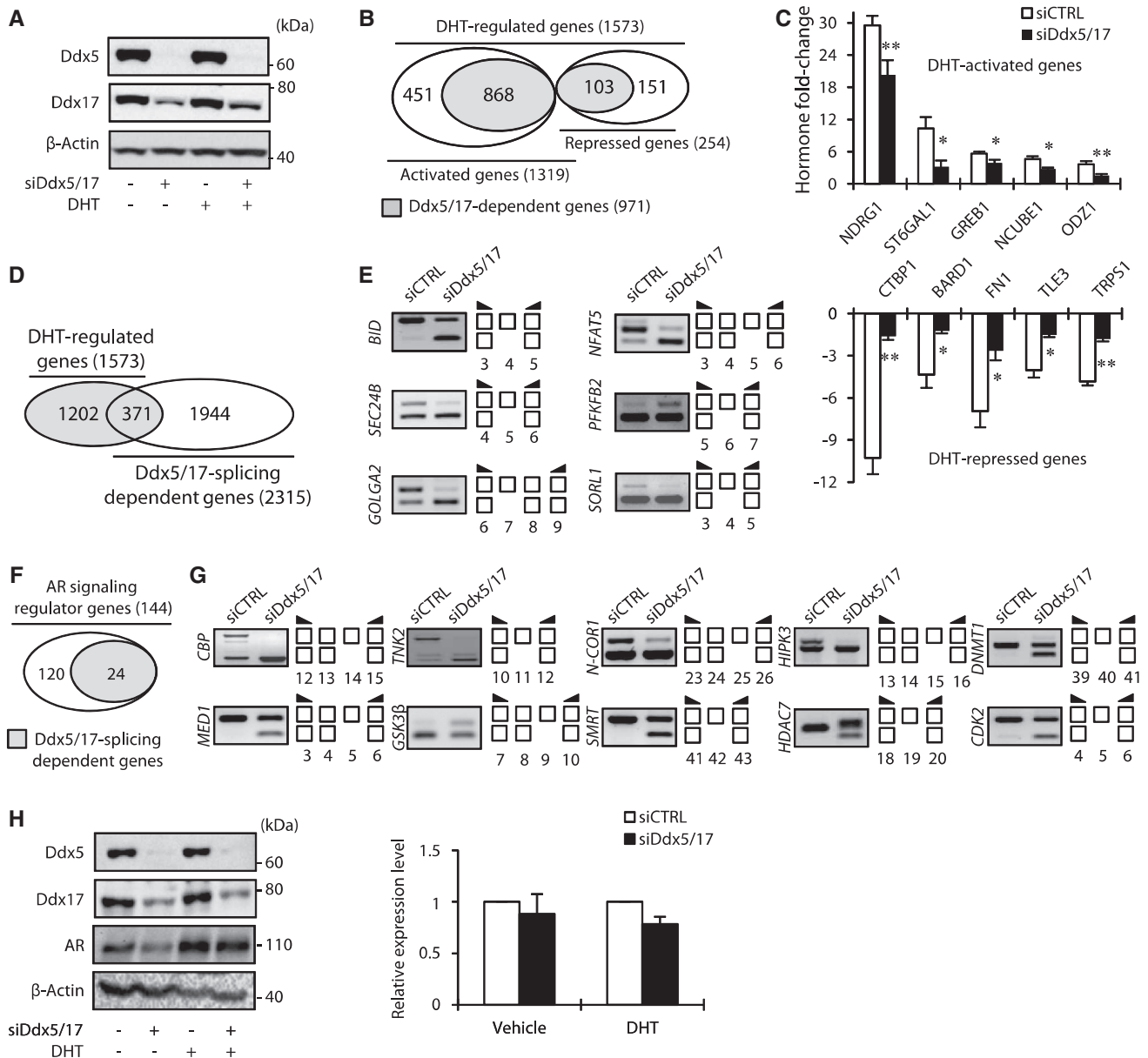
pharmacological inhibition (Figure 5E, bottom panel), as previously reported (6,27,34,35).

Therefore, Ddx5/17 directly participate to the estrogen- and androgen-signaling pathways as ERα and AR transcriptional coregulators but also indirectly by regulating alternative splicing of other transcriptional coregulators or other actors of the signaling pathways (Figure 5F, and see 'Discussion' section).

## DISCUSSION

Activation of the estrogen and androgen-signaling pathways requires the concerted action of a plethora of factors that either affect post-translational modifications of ERα and AR or are recruited on hormone target genes to mediate their transcriptional effects. In this work, we report that the Ddx5 and Ddx17 RNA helicases, that have been shown to be ERα and AR transcriptional coregulators (see 'Introduction' section), are master regulators of the estrogen and androgen-signaling pathways by controlling transcription and splicing both upstream and downstream of receptors (Figure 5F). Indeed, Ddx5 and Ddx17 regulate the expression of a large number of estrogen- and androgen-target genes both at the transcriptional and splicing levels (Figures 1–4). Therefore, Ddx5 and Ddx17 intervene downstream of the signaling pathways for the transcriptional and splicing regulation of hormone-targeted genes. This result observed on a large scale on endogenous hormone target genes supports a model where at least some transcriptional coregulators are not only involved in the quantitative regulation of endogenous gene expression but also in



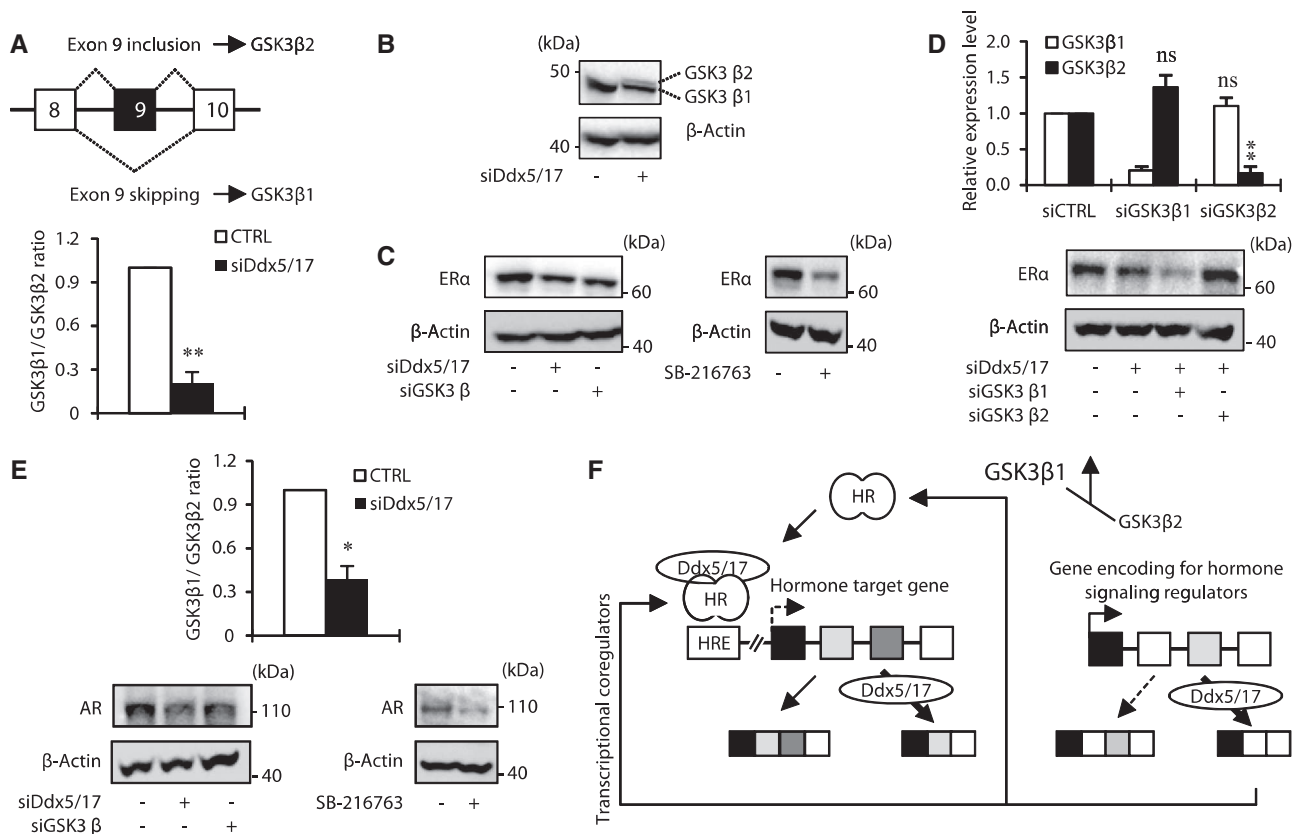


**Figure 4.** Ddx5 and Ddx17 play a major role in the androgen-signaling pathway. (A) Western blot analysis of Ddx5, Ddx17 and  $\beta$ -Actin as loading control, using total protein extract from LNCaP cells transfected with either a control siRNA (siCTRL) or siDdx5/17 in the presence or absence of DHT for 24 h. (B) *In silico* prediction of exon array analysis of genes regulated at the global expression level by DHT and dependent of Ddx5 and Ddx17 (in gray). (C) Hormone fold change induced by DHT treatment for 24 h of the expression level of a subset of mRNAs as determined by RT-qPCR in LNCaP cells transfected with siCTRL or siDdx5/17 (genes upregulated by DHT in the upper panel and genes downregulated by DHT in the lower panel). (D) *In silico* prediction of Exon array analysis. Venn diagram representing DHT-regulated genes and genes that were regulated by Ddx5 and Ddx17 at the splicing level. (E) Validation by RT-PCR of splicing variants produced from DHT-target genes and regulated by Ddx5 and Ddx17. LNCaP cells were transfected with siCTRL or siDdx5/17 then treated by DHT (10 nM) for 24 h. (F) Diagram representing AR-signaling regulator genes that were regulated by Ddx5 and Ddx17 at the splicing level. (G) Validation by RT-PCR of splicing events regulated by Ddx5/Ddx17 in the same experimental condition than F panel. (H) Western blot analysis of Ddx5, Ddx17, AR proteins level.  $\beta$ -Actin was used as loading control (left panel). LNCaP cells were transfected with either a control siRNA or siDdx5/17 in the presence or absence of DHT for 24 h. RT-qPCR analysis of the relative expression level of AR mRNA (right panel). Histograms represent the average of at least three independent experiments. Error bars represent s.e.m.; \* $P < 0.05$ ; \*\* $P < 0.01$  (*t*-test).

their qualitative regulation by allowing the production of specific splicing variants in response to stimuli.

Furthermore, we show that Ddx5 and Ddx17 act upstream of the estrogen and androgen-signaling pathways by controlling the expression, at the splicing

level, of several key regulators of the hormone-signaling pathways. (Figures 1-4 and 5F). Of particular interest, we showed that Ddx5 and Ddx17 controlled the expression level of the SMRT transcriptional coregulator. Indeed, Ddx5/17 depletion favored the production of a SMRT



**Figure 5.** Ddx5 and Ddx17 stabilize ER $\alpha$  and AR protein by modulating GSK3 $\beta$  splicing. (A) Schematic representation of GSK3 $\beta$  primary transcripts (upper panel). RT-qPCR analysis measuring the GSK3 $\beta$ 1 and GSK3 $\beta$ 2 ratio in MCF-7 cells transfected with a control siRNA or siDdx5/17. (B) Western blot analysis of GSK3 $\beta$  and  $\beta$ -actin as a loading control in MCF-7 cells transfected with a control siRNA or siDdx5/17. (C) Western blot analysis of ER $\alpha$  protein level and  $\beta$ -actin as a loading control in MCF-7 cells transfected with a control siRNA, siDdx5/17 or siGSK3 $\beta$  (left panel) or treated with 20  $\mu$ M of the SB-216673 GSK3 $\beta$  inhibitor for 24 h (right panel). (D) RT-qPCR analysis measuring GSK3 $\beta$ 1 and GSK3 $\beta$ 2 splicing variant expression level in MCF-7 cells transfected with specific siRNAs targeting each isoform (i.e. siGSK3 $\beta$ 1 and siGSK3 $\beta$ 2, upper panel). Western blot analysis of ER $\alpha$  protein level and  $\beta$ -Actin as a loading control in MCF-7 cells transfected with a control siRNA, siDdx5/17 and siGSK3 $\beta$ 1 or siGSK3 $\beta$ 2 (lower panel). (E) RT-qPCR analysis measuring the GSK3 $\beta$ 1 and GSK3 $\beta$ 2 ratio in LNCaP cells transfected with a control siRNA or siDdx5/17 (upper panel). Western blot analysis of AR protein level and  $\beta$ -actin as a loading control in LNCaP cells transfected with a control siRNA, siDdx5/17 or siGSK3 $\beta$  or treated with 20  $\mu$ M of the SB-216673 GSK3 $\beta$  inhibitor for 24 h (lower panels). (F) Ddx5 and Ddx17 that are recruited by Steroid Hormone Receptors (HR) on Hormone Responsive Elements (HRE) act as transcriptional coregulators and regulate alternative splicing of a least a subset of hormone-regulated genes (left panel). Meanwhile, Ddx5 and Ddx17 control the expression at the splicing level of key regulators of the steroid hormone-signaling pathways. Effectors of the steroid hormone-signaling pathways regulated at the splicing level by Ddx5/17 can either be transcriptional coregulators (e.g. SMRT) or proteins modulating HR post-translational modifications, like GSK3 $\beta$  that controls AR and ER $\alpha$  protein expression levels.

splicing variant that is degraded by the NMD pathway, leading to the decrease of SMRT protein level (Figure 3). Furthermore, we demonstrated that Ddx5 and Ddx17 regulated alternative splicing of GSK3 $\beta$ , thereby controlling AR and ER $\alpha$  protein levels (Figure 5). Thus, our data show that Ddx5/17 control steroid hormone-signaling pathways on a large scale, by acting both upstream and downstream of hormone receptors.

One interesting remaining question is whether Ddx5 and Ddx17 effects on transcription and splicing are somehow connected. At the transcriptional level, it has been proposed that Ddx5 and Ddx17 are recruited at the promoter level by transcriptional factors and serve as a bridge between transcription factors and other transcriptional coregulators, in particular, histone modifiers CBP (36,37) and HDACs (38). The effects of Ddx5 and Ddx17 in splicing are less characterized.

It has been proposed that, these proteins may help the binding of splicing regulators on their target pre-mRNAs owing to their RNA helicase activity (14,15,39). Interestingly, Ddx5 and Ddx17 interact with RNA polymerase II (RNAPII) and could affect transcription elongation (40,41). Because transcription elongation can have an impact on alternative splicing regulation (42), Ddx5 and Ddx17 could impact on splicing owing to their effects on transcription elongation. However, as most of the genes regulated at the splicing level by Ddx5 and Ddx17 were not regulated at the transcriptional level by E2 or DHT treatment (Figures 2A and 4D), this suggested that Ddx5/17 effects on splicing and E2 or DHT effects on transcription were mainly independent. However, this does not exclude a link between transcription and splicing. For example, Ddx5 and Ddx17 may have to be first recruited by the transcriptional

machinery before being loaded onto the nascent RNA molecules to impact on splicing; the recruitment by the transcriptional machinery may not necessarily be measurable by an effect on gene transcription activity.

Even though further experiments are required to better understand the mechanisms by which Ddx5 and Ddx17 affect transcription and splicing, their broad effects on the estrogen and androgen-signaling pathways has several physio-pathological consequences. First, the misregulation of Ddx5 and/or Ddx17 expression level and/or activity owing to posttranslational modifications that was reported in many cases of breast and prostate cancer (10,23,43–45), may have a broad impact on the estrogen- and androgen-signaling pathways that play a major role in the etiology and progression of those cancers. For example, the *Ddx5* gene is fused in frame to the *ETV4* gene in prostate tumors (43). Based on our data, it can be anticipated that such a translocation may have a broad impact on the androgen-signaling pathway in these tumors.

Second, we observed that many of the Ddx5/17 splicing-regulated genes have been reported to play a direct role in tamoxifen resistance, one of the major endocrine therapies of breast cancer. This includes several transcriptional coregulators such as SMRT, MED1 and NCoR1 involved in resistance to tamoxifen (46–48) and several kinases such as GSK3 $\beta$ , CDK2 and RET which participate, directly or indirectly, to ER $\alpha$  phosphorylation and thus, the hormone-independent activation of ER and tamoxifen resistance (49–51). Interestingly as well, Ddx5 expression correlates with that of HER-2/neu epidermal growth factor receptor (23,45). It has been reported that overexpression of HER2/neu is associated with increased ER phosphorylation on ser-118 residue (52) which could play a role in resistance to tamoxifen by permanent ER $\alpha$  activation. Interestingly, even though further experiments are required because of potential contradictory reported data, high Ddx5 expression level could be associated with higher tumor grade or poor prognosis in ER-positive breast cancer patients (23,53). Furthermore, a recent report indicate that, Ddx5 and Ddx17 proteins level change significantly according to breast cancer subtype (45) and is highly dependent on certain miRNAs abundance and regulation, in particular miR-206 that has been reported to play a major role in breast cancer and estrogen pathway (54). Therefore, changes in the expression of Ddx5 and Ddx17, in particular under the control of altered miRNAs in breast cancer, could play a role in resistance to tamoxifen, in particular owing to their effects on splicing of genes involved in resistance (see above). It would be particularly interesting to test whether resistance to endocrine therapy is associated with mis-regulation of splicing of genes involved in steroid hormone-signaling pathways.

While many researches have been concentrated on the canonical GSK3 $\beta$ 1, ubiquitously distributed in organs, minor attention was attributed to GSK3 $\beta$ 2 isoform, whose substrate preference and physiological significance remain unclear. In GSK3 $\beta$ 2 long isoform, exon 9 translation produces a 13 amino acid insert of unknown function in an external loop near the catalytic domain (28–30)

which was associated with reduced kinase activity (30,31). Changes in the balance of GSK3 $\beta$ 2/GSK3 $\beta$ 1 isoforms due to alternative splicing regulation could have a broad impact in cancer initiation or progression, not only owing to an effect on ER and AR protein level, as reported in this work but also owing to the key role of GSK3 $\beta$  in the  $\beta$ -catenin homeostasis and its cross-talk with WNT-signaling pathway.

Because there is increasing evidence that the interplay between transcription and splicing participates in signaling pathway regulation and outcomes, we created a freely available website, named SSAS-DB (<http://fasterdb.lyon.unicancer.fr/ssas-db/home.pl>) that provides genome-wide information on the impact of estrogen, androgen and Ddx5/17 on gene expression in breast and prostate cancer cell lines, at both whole-gene (transcriptional) and exon- (alternative splicing) levels. This website is a valuable tool to analyze the splicing variants of hormone-regulated genes or genes involved in hormone-signaling pathways (see Supplementary Figure S5 for further information on SSAS-DB).

In conclusion, our work provides a new understanding of steroid hormone-signaling pathways owing to the identification of two related proteins, Ddx5 and Ddx17, that not only mediate the regulation at both transcriptional and splicing levels of a large subset of steroid hormone target genes, but also control a set of splicing events in genes that, in turn, control steroid receptor abundance and activity. Thus, we propose that Ddx5 and Ddx17 are master regulators of estrogen and androgen-signaling pathways by acting both upstream and downstream of receptors.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online, including [55].

## ACKNOWLEDGEMENTS

The authors thank Mark McCarron for helpful reading of the paper and Marie-Pierre Lambert for drawing figures.

## FUNDING

The Institut National du Cancer (INCa); Agence Nationale de la Recherche (ANR); Fondation Recherche Médicale (FRM); Ligue Nationale Contre le Cancer (to S.S. and E.D.); Association Française contre les Myopathies (to M.P.E.), FRM (to E.Z.) and Association pour la Recherche sur le Cancer (to S.G.). Funding for open access charge: Institut National du Cancer (INCa) Agence Nationale de la Recherche (ANR) Fondation Recherche Médicale (FRM) Ligue Nationale Contre le Cancer (LNCC) Association Française contre les Myopathies (AFM) Association pour la Recherche sur le Cancer (ARC)

*Conflict of interest statement.* None declared.

## REFERENCES

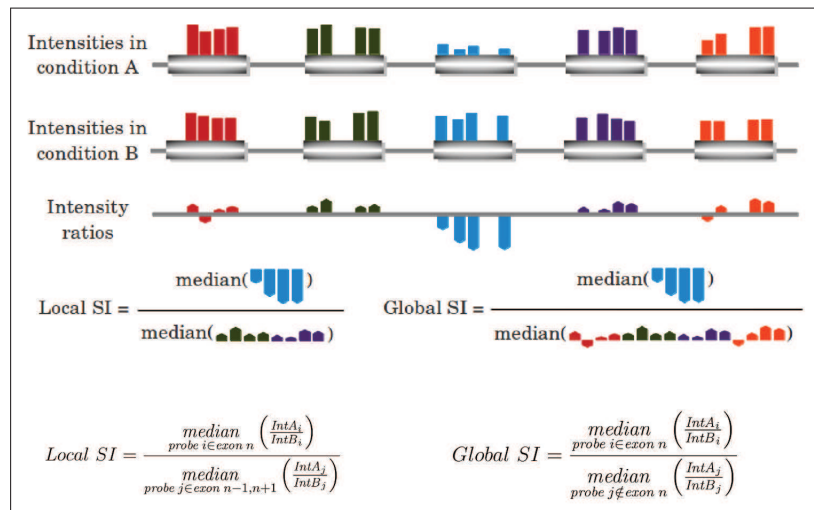
- Madhunapantula, S.V., Mosca, P. and Robertson, G.P. (2010) Steroid hormones drive cancer development. *Cancer Biol. Ther.*, **10**, 765–766.
- Green, S.M., Mostaghel, E.A. and Nelson, P.S. (2012) Androgen action and metabolism in prostate cancer. *Mol. Cell Endocrinol.*, **360**, 3–13.
- Kato, S., Yokoyama, A. and Fujiki, R. (2011) Nuclear receptor coregulators merge transcriptional coregulation with epigenetic regulation. *Trends Biochem. Sci.*, **36**, 272–281.
- Le Romancer, M., Poulard, C., Cohen, P., Sentis, S., Renoir, J.M. and Corbo, L. (2011) Cracking the estrogen receptor's posttranslational code in breast tumors. *Endocr. Rev.*, **32**, 597–622.
- Grisouard, J., Medunjanin, S., Hermani, A., Shukla, A. and Mayer, D. (2007) Glycogen synthase kinase-3 protects estrogen receptor alpha from proteasomal degradation and is required for full transcriptional activity of the receptor. *Mol. Endocrinol.*, **21**, 2427–2439.
- Schutz, S.V., Cronauer, M.V. and Rinnab, L. (2010) Inhibition of glycogen synthase kinase-3beta promotes nuclear export of the androgen receptor through a CRM1-dependent mechanism in prostate cancer cell lines. *J. Cell. Biochem.*, **109**, 1192–1200.
- O'Malley, B.W., Malovannaya, A. and Qin, J. (2012) Minireview: nuclear receptor and coregulator proteomics—2012 and beyond. *Mol. Endocrinol.*, **26**, 1646–1650.
- Rosenfeld, M.G., Lunyak, V.V. and Glass, C.K. (2006) Sensors and signals: a coactivator/corepressor/epigenetic code for integrating signal-dependent programs of transcriptional response. *Genes Dev.*, **20**, 1405–1428.
- Watanabe, M., Yanagisawa, J., Kitagawa, H., Takeyama, K., Ogawa, S., Arai, Y., Suzawa, M., Kobayashi, Y., Yano, T., Yoshikawa, H. et al. (2001) A subfamily of RNA-binding DEAD-box proteins acts as an estrogen receptor alpha coactivator through the N-terminal activation domain (AF-1) with an RNA coactivator, SRA. *EMBO J.*, **20**, 1341–1352.
- Clark, E.L., Coulson, A., Dalgliesh, C., Rajan, P., Nicol, S.M., Fleming, S., Heer, R., Gaughan, L., Leung, H.Y., Elliott, D.J. et al. (2008) The RNA helicase p68 is a novel androgen receptor coactivator involved in splicing and is overexpressed in prostate cancer. *Cancer Res.*, **68**, 7938–7946.
- Metivier, R., Penot, G., Hubner, M.R., Reid, G., Brand, H., Kos, M. and Gannon, F. (2003) Estrogen receptor-alpha directs ordered, cyclical, and combinatorial recruitment of cofactors on a natural target promoter. *Cell*, **115**, 751–763.
- Wong, H.Y., Demmers, J.A., Bezstarosti, K., Grootegoed, J.A. and Brinkmann, A.O. (2009) DNA dependent recruitment of DDX17 and other interacting proteins by the human androgen receptor. *Biochim. Biophys. Acta*, **1794**, 193–198.
- Liu, Z.R. (2002) p68 RNA helicase is an essential human splicing factor that acts at the U1 snRNA-5' splice site duplex. *Mol. Cell Biol.*, **22**, 5443–5450.
- Kar, A., Fushimi, K., Zhou, X., Ray, P., Shi, C., Chen, X., Liu, Z., Chen, S. and Wu, J.Y. (2011) RNA helicase p68 (DDX5) regulates tau exon 10 splicing by modulating a stem-loop structure at the 5' splice site. *Mol. Cell Biol.*, **31**, 1812–1821.
- Guil, S., Gattoni, R., Carrascal, M., Abian, J., Stevenin, J. and Bach-Elias, M. (2003) Roles of hnRNP A1, SR proteins, and p68 helicase in c-H-ras alternative splicing regulation. *Mol. Cell Biol.*, **23**, 2927–2941.
- Germann, S., Gratadou, L., Zonta, E., Dardenne, E., Gaudineau, B., Fougere, M., Samaan, S., Dutertre, M., Jauliac, S. and Auboeuf, D. (2012) Dual role of the ddx5/ddx17 RNA helicases in the control of the pro-migratory NFAT5 transcription factor. *Oncogene*, **31**, 4536–4549.
- Dardenne, E., Pierredon, S., Driouch, K., Gratadou, L., Lacroix-Triki, M., Espinoza, M.P., Zonta, E., Germann, S., Mortada, H., Villemin, J.P. et al. (2012) Splicing switch of an epigenetic regulator by RNA helicases promotes tumor-cell invasiveness. *Nat. Struct. Mol. Biol.*, **19**, 1139–1146.
- Kelemen, O., Convertini, P., Zhang, Z., Wen, Y., Shen, M., Falaleeva, M. and Stamm, S. (2013) Function of alternative splicing. *Gene*, **514**, 1–30.
- Auboeuf, D., Honig, A., Berget, S.M. and O'Malley, B.W. (2002) Coordinate regulation of transcription and splicing by steroid receptor coregulators. *Science*, **298**, 416–419.
- Barbier, J., Dutertre, M., Bittencourt, D., Sanchez, G., Gratadou, L., de la Grange, P. and Auboeuf, D. (2007) Regulation of H-ras splice variant expression by cross talk between the p53 and nonsense-mediated mRNA decay pathways. *Mol. Cell Biol.*, **27**, 7315–7333.
- Zonta, E., Bittencourt, D., Samaan, S., Germann, S., Dutertre, M. and Auboeuf, D. (2013) The RNA helicase DDX5/p68 is a key factor promoting c-fos expression at different levels from transcription to mRNA export. *Nucleic Acids Res.*, **41**, 554–564.
- Endoh, H., Maruyama, K., Masuhiro, Y., Kobayashi, Y., Goto, M., Tai, H., Yanagisawa, J., Metzger, D., Hashimoto, S. and Kato, S. (1999) Purification and identification of p68 RNA helicase acting as a transcriptional coactivator specific for the activation function 1 of human estrogen receptor alpha. *Mol. Cell Biol.*, **19**, 5363–5372.
- Wortham, N.C., Ahamed, E., Nicol, S.M., Thomas, R.S., Periyasamy, M., Jiang, J., Ochocka, A.M., Shousha, S., Huson, L., Bray, S.E. et al. (2009) The DEAD-box protein p72 regulates ERalpha/oestrogen-dependent transcription and cell growth, and is associated with improved survival in ERalpha-positive breast cancer. *Oncogene*, **28**, 4053–4064.
- Zhang, X., Krutchinsky, A., Fukuda, A., Chen, W., Yamamura, S., Chait, B.T. and Roeder, R.G. (2005) MED1/TRAP220 exists predominantly in a TRAP/ Mediator subpopulation enriched in RNA polymerase II and is required for ER-mediated transcription. *Mol. Cell*, **19**, 89–100.
- Varlakhanova, N., Snyder, C., Jose, S., Hahn, J.B. and Privalsky, M.L. (2010) Estrogen receptors recruit SMRT and N-CoR corepressors through newly recognized contacts between the corepressor N terminus and the receptor DNA binding domain. *Mol. Cell Biol.*, **30**, 1434–1445.
- Medunjanin, S., Hermani, A., De Servi, B., Grisouard, J., Rincke, G. and Mayer, D. (2005) Glycogen synthase kinase-3 interacts with and phosphorylates estrogen receptor alpha and is involved in the regulation of receptor activity. *J. Biol. Chem.*, **280**, 33006–33014.
- Schutz, S.V., Schrader, A.J., Zengerling, F., Genze, F., Cronauer, M.V. and Schrader, M. (2011) Inhibition of glycogen synthase kinase-3beta counteracts ligand-independent activity of the androgen receptor in castration resistant prostate cancer. *PLoS One*, **6**, e25341.
- Wood-Kaczmar, A., Kraus, M., Ishiguro, K., Philpott, K.L. and Gordon-Weeks, P.R. (2009) An alternatively spliced form of glycogen synthase kinase-3beta is targeted to growing neurites and growth cones. *Mol. Cell Neurosci.*, **42**, 184–194.
- Mukai, F., Ishiguro, K., Sano, Y. and Fujita, S.C. (2002) Alternative splicing isoform of tau protein kinase 1/glycogen synthase kinase 3beta. *J. Neurochem.*, **81**, 1073–1083.
- Schaffer, B., Wiedau-Pazos, M. and Geschwind, D.H. (2003) Gene structure and alternative splicing of glycogen synthase kinase 3 beta (GSK-3beta) in neural and non-neural tissues. *Gene*, **302**, 73–81.
- Castano, Z., Gordon-Weeks, P.R. and Kypta, R.M. (2010) The neuron-specific isoform of glycogen synthase kinase-3beta is required for axon growth. *J. Neurochem.*, **113**, 117–130.
- Kwok, J.B., Hallupp, M., Loy, C.T., Chan, D.K., Woo, J., Mellick, G.D., Buchanan, D.D., Silburn, P.A., Halliday, G.M. and Schofield, P.R. (2005) GSK3B polymorphisms alter transcription and splicing in Parkinson's disease. *Ann. Neurol.*, **58**, 829–839.
- Kim, H.M., Kim, C.S., Lee, J.H., Jang, S.J., Hwang, J.J., Ro, S. and Choi, J. (2013) CG0009, a novel glycogen synthase kinase 3 inhibitor, induces cell death through cyclin D1 depletion in breast cancer cells. *PLoS One*, **8**, e60383.
- Liao, X., Thrasher, J.B., Holzbeierlein, J., Stanley, S. and Li, B. (2004) Glycogen synthase kinase-3beta activity is required for androgen-stimulated gene expression in prostate cancer. *Endocrinology*, **145**, 2941–2949.



35. Mazor, M., Kawano, Y., Zhu, H., Waxman, J. and Kypta, R.M. (2004) Inhibition of glycogen synthase kinase-3 represses androgen receptor activity and prostate cancer cell growth. *Oncogene*, **23**, 7882–7892.
36. Rossow, K.L. and Janknecht, R. (2003) Synergism between p68 RNA helicase and the transcriptional coactivators CBP and p300. *Oncogene*, **22**, 151–156.
37. Shin, S. and Janknecht, R. (2007) Concerted activation of the Mdm2 promoter by p72 RNA helicase and the coactivators p300 and P/CAF. *J. Cell Biochem.*, **101**, 1252–1265.
38. Wilson, B.J., Bates, G.J., Nicol, S.M., Gregory, D.J., Perkins, N.D. and Fuller-Pace, F.V. (2004) The p68 and p72 DEAD box RNA helicases interact with HDAC1 and repress transcription in a promoter-specific manner. *BMC Mol. Biol.*, **5**, 11.
39. Laurent, F.X., Sureau, A., Klein, A.F., Trouslard, F., Gasnier, E., Furling, D. and Marie, J. (2012) New function for the RNA helicase p68/DDX5 as a modifier of MBNL1 activity on expanded CUG repeats. *Nucleic Acids Res.*, **40**, 3159–3171.
40. Clark, E.L., Fuller-Pace, F.V., Elliott, D.J. and Robson, C.N. (2008) Coupling transcription to RNA processing via the p68 DEAD box RNA helicase androgen receptor co-activator in prostate cancer. *Biochem. Soc. Trans.*, **36**, 546–547.
41. Clark, E.L., Hadjimichael, C., Temperley, R., Barnard, A., Fuller-Pace, F.V. and Robson, C.N. (2013) p68/Ddx5 supports beta-catenin & RNAP II during androgen receptor mediated transcription in prostate cancer. *PLoS One*, **8**, e54150.
42. Kadener, S., Fededa, J.P., Rosbash, M. and Kornblihtt, A.R. (2002) Regulation of alternative splicing by a transcriptional enhancer through RNA pol II elongation. *Proc. Natl Acad. Sci. USA*, **99**, 8185–8190.
43. Han, B., Mehra, R., Dhanasekaran, S.M., Yu, J., Menon, A., Lonigro, R.J., Wang, X., Gong, Y., Wang, L., Shankar, S. *et al.* (2008) A fluorescence in situ hybridization screen for E26 transformation-specific aberrations: identification of DDX5-ETV4 fusion protein in prostate cancer. *Cancer Res.*, **68**, 7629–7637.
44. Janknecht, R. (2010) Multi-talented DEAD-box proteins and potential tumor promoters: p68 RNA helicase (DDX5) and its paralog, p72 RNA helicase (DDX17). *Am. J. Transl. Res.*, **2**, 223–234.
45. Wang, D., Huang, J. and Hu, Z. (2012) RNA helicase DDX5 regulates microRNA expression and contributes to cytoskeletal reorganization in basal breast cancer cells. *Mol. Cell. Proteom.*, **11**, M111 011932.
46. Graham, J.D., Bain, D.L., Richer, J.K., Jackson, T.A., Tung, L. and Horwitz, K.B. (2000) Thoughts on tamoxifen resistant breast cancer: Are coregulators the answer or just a red herring? *J. Steroid Biochem. Mol. Biol.*, **74**, 255–259.
47. Nagalingam, A., Tighiouart, M., Ryden, L., Joseph, L., Landberg, G., Saxena, N.K. and Sharma, D. (2012) Med1 plays a critical role in the development of tamoxifen resistance. *Carcinogenesis*, **33**, 918–930.
48. van Agthoven, T., Sieuwerts, A.M., Veldscholte, J., Meijer-van Gelder, M.E., Smid, M., Brinkman, A., den Dekker, A.T., Leroy, I.M., van Ijcken, W.F., Sleijfer, S. *et al.* (2009) CITED2 and NCOR2 in anti-oestrogen resistance and progression of breast cancer. *Br. J. Cancer*, **101**, 1824–1832.
49. Caldon, C.E., Sergio, C.M., Kang, J., Muthukaruppan, A., Boersma, M.N., Stone, A., Barraclough, J., Lee, C.S., Black, M.A., Miller, L.D. *et al.* (2012) Cyclin E2 overexpression is associated with endocrine resistance but not insensitivity to CDK2 inhibition in human breast cancer cells. *Mol. Cancer Therapeut.*, **11**, 1488–1499.
50. Oyama, M., Nagashima, T., Suzuki, T., Kozuka-Hata, H., Yumoto, N., Shiraishi, Y., Ikeda, K., Kuroki, Y., Gotoh, N., Ishida, T. *et al.* (2011) Integrated quantitative analysis of the phosphoproteome and transcriptome in tamoxifen-resistant breast cancer. *J. Biol. Chem.*, **286**, 818–829.
51. Plaza-Menacho, I., Morandi, A., Robertson, D., Pancholi, S., Drury, S., Dowsett, M., Martin, L.A. and Isacke, C.M. (2010) Targeting the receptor tyrosine kinase RET sensitizes breast cancer cells to tamoxifen treatment and reveals a role for RET in endocrine resistance. *Oncogene*, **29**, 4648–4657.
52. Yamashita, H., Nishio, M., Toyama, T., Sugiura, H., Kondo, N., Kobayashi, S., Fujii, Y. and Iwase, H. (2008) Low phosphorylation of estrogen receptor alpha (ERalpha) serine 118 and high phosphorylation of ERalpha serine 167 improve survival in ER-positive breast cancer. *Endocrine-related Cancer*, **15**, 755–763.
53. Dutertre, M., Grataudou, L., Dardenne, E., Germann, S., Samaan, S., Lidereau, R., Driouch, K., de la Grange, P. and Auboeuf, D. Estrogen regulation and physiopathologic significance of alternative promoters in breast cancer. *Cancer Res.*, **70**, 3760–3770.
54. Kondo, N., Toyama, T., Sugiura, H., Fujii, Y. and Yamashita, H. (2008) miR-206 Expression is down-regulated in estrogen receptor alpha-positive human breast cancer. *Cancer Res.*, **68**, 5004–5008.

### C Publication dans Genome research

Dans cette publication disponible en ligne et très prochainement publiée dans le journal, nous montrons que les trois types cellulaires majeurs (fibroblaste, épithélial et endothélial) présentent un programme d'épissage spécifique, de manière indépendante à leur origine tissulaire. D'autre part, ces programmes apparaissent contrôlés par de nombreux facteurs d'épissage identifiés dans le cadre de ce travail. Ce projet comporte de nombreux aspects bioinformatiques. Je me suis occupé en particulier des analyses de très larges jeux de données de puces exon (provenant du projet ENCODE et de différents laboratoires). La difficulté rencontrée ici était la détection des exons cassettes présentant des variations faibles entre type cellulaires. Ce phénomène est d'autant plus important dans les gènes ne présentant pas un seul événement d'épissage. C'est pour répondre à cette problématique que nous mis au point le Splicing Index local, algorithme utilisé dans ce projet, dont la démarche est la suivante :



# Endothelial, epithelial, and fibroblast cells exhibit specific splicing programs independently of their tissue of origin

5 Pierre Mallinjou, <sup>1,2,3,4,5</sup> Jean-Philippe Villemin, <sup>1,2,3,4,5</sup> Hussein Mortada, <sup>1,2,3,4,5</sup>  
 Micaela Polay-Espinoza, <sup>1,2,3,4</sup> François-Olivier Desmet, <sup>1,2,3,4</sup> Samaan Samaan, <sup>1,2,3,4</sup>  
 Emilie Chautard, <sup>1,2,3,4</sup> Léon-Charles Tranchevent, <sup>1,2,3,4</sup> and Didier Auboeuf <sup>1,2,3,4,6</sup>

<sup>1</sup>Inserm UMR-S1052, Centre de Recherche en Cancérologie de Lyon, 69008 Lyon, France; <sup>2</sup>CNRS UMR5286, Centre de Recherche en Cancérologie de Lyon, 69008 Lyon, France; <sup>3</sup>Université de Lyon, 69007 Lyon, France; <sup>4</sup>Centre Léon Bérard, 69008 Lyon, France

**AU1** 10 Alternative splicing is the main mechanism of increasing the proteome diversity coded by a limited number of genes. It is well established that different tissues or organs express different splicing variants. However, organs are composed of common major cell types, including fibroblasts, epithelial, and endothelial cells. By analyzing large-scale data sets generated by the The ENCODE Project Consortium and after extensive RT-PCR validation, we demonstrate that each of the three major cell types expresses a specific splicing program independently of its organ origin. Furthermore, by analyzing splicing factor expression across samples, publicly available splicing factor binding site data sets (CLIP-seq), and exon array data sets after splicing factor depletion, we identified several splicing factors, including ESRP1 and 2, MBNLI, NOVA1, PTBPI, and RBFOX2, that contribute to establishing these cell type-specific splicing programs. All of the analyzed data sets are freely available in a user-friendly web interface named FasterDB, which describes all known splicing variants of human and mouse genes and their splicing patterns across several dozens of normal and cancer cells as well as across tissues. Information regarding splicing factors that potentially contribute to individual exon regulation is also provided via a dedicated CLIP-seq and exon array data visualization interface. To the best of our knowledge, FasterDB is the first database integrating such a variety of large-scale data sets to enable functional genomics analyses at exon-level resolution.

[Supplemental material is available for this article.]

25 Human genes are an assemblage of exons that can be differentially selected during splicing. Alternative splicing, which can produce splicing variants with different exonic content from a single gene, is the rule rather than an exception, as 95% of human genes generate several splicing variants (Kim et al. 2008; Hallegger et al. 2010; Kalsotra and Cooper 2011; Blencowe 2012; Kelemen et al. 2013). Alternative splicing relies on the combinatory action of splicing factors (e.g., SR and hnRNP proteins) that bind to exonic or intronic splicing regulatory sequences to either strengthen or inhibit splice site recognition by the splicing machinery, therefore enhancing or repressing the inclusion of alternative exons (Barash et al. 2010; Goren et al. 2010; Witten and Ule 2011). Similar to how transcription factors control transcriptional programs by directing the expression of gene networks, splicing factors control splicing programs by regulating alternative splicing of co-regulated exons (Hartmann and Valcarcel 2009; Barash et al. 2010; Goren et al. 2010; Witten and Ule 2011). Alternative splicing is the main mechanism used to increase the proteome diversity coded by a limited number of genes, as the majority of alternative exons contains coding sequences (Kim et al. 2008; Hallegger et al. 2010; Kalsotra and Cooper 2011; Blencowe 2012; Kelemen et al. 2013). Because of the diversity generated by alternative splicing and the complexity of its regulation, functional genomics at exon-level

resolution requires the development of new integrative bioinformatics approaches.

Functional genomics at exon-level resolution is necessary to better understand tissue-specific functions. Indeed, it is well established that different tissues (or organs) express different splicing variants (Bland et al. 2010; de la Grange et al. 2010; Hartmann et al. 2011; Llorian and Smith 2011; Barbosa-Morais et al. 2012; Merkin et al. 2012). The development of new technologies like splicing-sensitive microarrays and massive RNA sequencing fully establish that different tissues express different splicing programs as a consequence of the combinatorial actions of more-or-less tissue-specific splicing factors (Pan et al. 2008; Wang et al. 2008; Merkin et al. 2012). However, most organs are composed of common cell types, such as fibroblast and epithelial cells, which perform specific functions. Epithelial cells are tightly connected cells arranged in monolayer with several functions, such as protection, diffusion, secretion, absorption, and excretion, and establishing boundaries between compartments. Fibroblasts comprise the structural framework of tissues and synthesize the extracellular matrix, a supportive framework for epithelial cells. Unlike epithelial cells, fibroblasts can migrate as individual cells. Another important cell type is represented by endothelial cells that compose the endothelium, the thin layer of cells that lines the interior surface of blood vessels that supply tissues and organs with blood.

<sup>5</sup>These authors contributed equally to this work.

<sup>6</sup>Corresponding author

E-mail [didier.auboeuf@inserm.fr](mailto:didier.auboeuf@inserm.fr)

Article published online before print. Article, supplemental material, and publication date are at <http://www.genome.org/cgi/doi/10.1101/gr.162933.113>.

© 2014 Mallinjou et al. This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see <http://genome.cshlp.org/site/misc/terms.xhtml>). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 3.0 Unported), as described at <http://creativecommons.org/licenses/by-nc/3.0/>.

Recent large-scale analyses suggest that splicing programs may contribute to establishing cell type-specific functions. Indeed, the epithelial-to-mesenchymal transition (EMT) that corresponds to the *trans*-differentiation of epithelial to mesenchymal (fibroblast-like) cells relies not only on transcriptional programs but also on extensive changes in alternative splicing (Warzecha et al. 2009a, 2010; Shapiro et al. 2011; Venables et al. 2013). Splicing factors, including ESRP1 and 2 and RBFOX2 (RBM9), have been reported to play a key role in EMT (Warzecha et al. 2009a, 2010; Shapiro et al. 2011; Dittmar et al. 2012; Venables et al. 2013). Although a recent report indicates that several dozen genes are differentially spliced when comparing normal fibroblasts to epithelial cells isolated from colon and ovarian tissues (Venables et al. 2013), it is currently not known whether the major common cell types, namely fibroblast, epithelial, and endothelial cells, express specific splicing programs independently of their organ origin.

To address this question, we analyzed an ENCODE data set based on exon arrays performed on RNAs prepared from several dozen normal fibroblast, epithelial, and endothelial cells isolated from different tissues (Thurman et al. 2012). After extensive RT-PCR validation, we show that each major cell type expresses a specific splicing program independently of their tissue origin, suggesting a role of alternative splicing in establishing specific functions of common cell types shared by many tissues. Additionally, by analyzing several publicly available large-scale data sets related to these factors, we identified a set of splicing factors that coordinate these splicing programs. In order to provide the community with full support for alternative splicing analysis, we have made all of the analyzed data sets freely available through a user-friendly web interface named FasterDB, which describes all the known splicing variants of human and mouse genes as well as their splicing patterns across several dozens of normal cells, cancer cells, and tissues. Furthermore, FasterDB integrates several kinds of publicly available data sets, namely, splicing factor binding sites (CLIP-seq) and splicing factor depletion followed by exon-array analysis, to provide users with information regarding the splicing factors that potentially contribute to regulating each exon. To the best of our knowledge, FasterDB is the first database to integrate such a variety of data sets, thereby enabling functional genomics at exon-level resolution.

## Results

### Identification of cell type-specific splicing programs independent of tissue origin

While a splicing program switch has been shown to occur in a few models of *trans*-differentiation of epithelial to mesenchymal (fibroblast-like) cells (Warzecha et al. 2009a, 2010; Shapiro et al. 2011; Dittmar et al. 2012; Venables et al. 2013), we tested whether epithelial cells and fibroblasts express a different splicing program independently of their tissue of origin. For this purpose, we analyzed an ENCODE data set corresponding to 18 and 12 normal fibroblast and epithelial cells, respectively, isolated from different organs (Supplemental Table S1). As shown in Figure 1A, fibroblast and epithelial cells express a transcriptome that differs both quantitatively (at the gene level) and qualitatively (at the exon level) (Supplemental Table S2). Focusing on variations at the exon level and using annotations based on known publicly available transcripts (see below), we observed that ~20% of the cases corresponded to alternative first exons (AFE) and alternative last exons (ALE), while 12% of the cases corresponded to alternatively

spliced exons (ASE) (Fig. 1A). About half of the cases were not annotated (NA), indicating potentially novel alternative exons. Remarkably, using the splicing index (SI), which represents the inclusion rate of the differentially spliced exons, fibroblasts and epithelial cells were clustered independently of their organ origin (Fig. 1B; Supplemental Fig. S1). Similar results were obtained by principal component analyses, which showed a clear clustering of the variables (cell lines) depending on their type (epithelial or fibroblast; Supplemental Fig. S3).

As shown in Figure 1C (see also Supplemental Fig. S1), a large set of ASE events was validated by RT-PCR. Further demonstrating that a set of exons can be differentially spliced in major cell types even if they have the same tissue origin, the splicing pattern observed in HMEC cells, which are epithelial cells from the mammary gland, was different from the splicing pattern observed in HMF cells, which are fibroblast cells also from the mammary gland. Additionally, the splicing pattern observed in HMEC cells was similar to the splicing pattern observed in other epithelial cells from different tissues, and the splicing pattern observed in HMF cells was similar to the splicing pattern observed in other fibroblast cells from different tissues (Fig. 1C).

Endothelial cells are also a general cell type present in all organs, as they form blood vessels that supply organs with blood. Because of some functional similarities, endothelial cells are often compared to epithelial cells although they have different embryonic origins. It is not known whether endothelial cells express a specific splicing program or not. As shown in Figure 2, A and B, endothelial cells differ from epithelial cells in both transcriptional and splicing programs (Supplemental Table S3); this was additionally validated by RT-PCR (Fig. 2C; Supplemental Fig. S2). Likewise, endothelial cells differ from fibroblasts in their transcriptional and splicing programs (Supplemental Table S4).

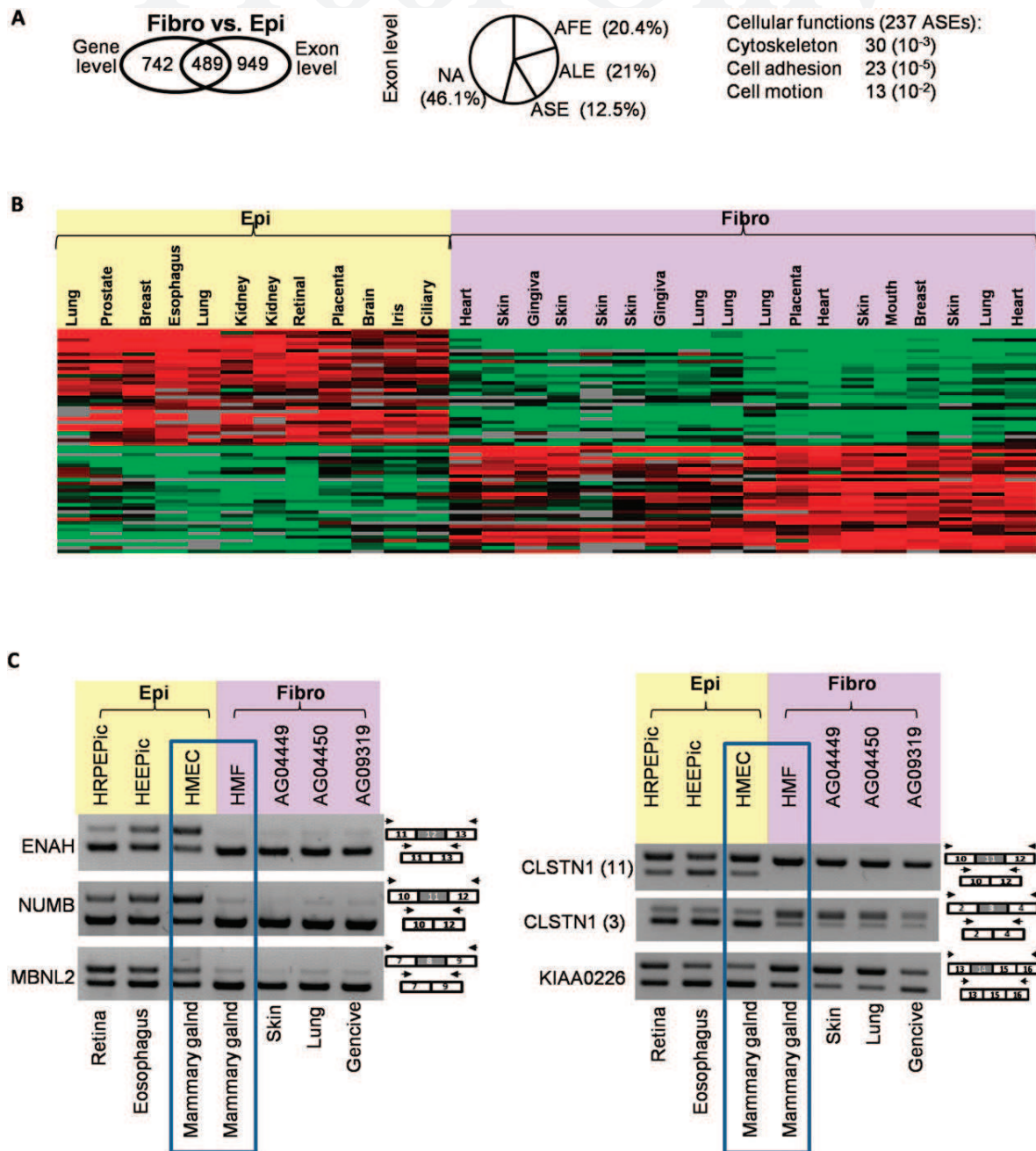
A gene ontology (GO) annotation analysis revealed that genes regulated at the splicing level (i.e., ASE) may contribute to cell type-specific cellular programs, as they were mainly involved in cytoskeleton, cell adhesion, and motion, which are the main features distinguishing these cell types (Figs. 1A, 2A).

To further challenge a model in which each of the three major cell types (fibroblast, epithelial, and endothelial) is characterized by specific splicing programs independently from their tissue of origin, we compared each cell type to the other two (Supplemental Tables S5–S7). Remarkably, we identified a set of alternative cassette exons whose splicing index allowed a perfect clustering of each major cell type (Fig. 3A; Supplemental Fig. S4). Thus, we identified a set of exons that are differentially spliced across three major cell types independently of their tissue origin, as validated by RT-PCR (Fig. 3B; Supplemental Table S1). Similar results were obtained using alternative first or last exons (AFE or ALE, respectively), demonstrating that major cell types can differ by the exonic content (ASE, AFE, and ALE) of the transcripts they express (Supplemental Figs. S5, S6).

### Cell type-specific splicing programs are controlled by balanced expression of antagonist splicing factors

By focusing on splicing factors with a marked cell type-specific expression pattern (e.g., up-regulated in only one cell type and down-regulated in the other two; see Supplemental Table S8), we identified a set of splicing factors whose expression level allowed us to classify each cell type (Fig. 4A; Supplemental Fig. S7). In particular, ESRP1 had a higher level of expression in epithelial cells than in other cell types, as confirmed by RT-qPCR (Fig. 4A,B;





**Figure 1.** Epithelial- and fibroblast-specific splicing variants. (A) Transcriptome analysis of fibroblasts (fibro) as compared to epithelial (epi) cells at both gene and exon levels. The number of genes differentially expressed at the gene and/or exon level when comparing both cell types is shown in the *left* panel. The *middle* panel indicates the classification of events corresponding to exon level variations. Differentially expressed exons were classified according to their annotation using publicly available transcripts: alternative first exon (AFE), alternative last exon (ALE), and alternative skipped exon (ASE). Not annotated (NA) corresponds to exons that do not correspond to any of the above-mentioned categories. The cellular functions of genes differentially spliced when comparing fibroblasts to epithelial cells are indicated in the *right* panel. (B) Heatmap presentation of the splicing index (SI) values for exons differentially spliced when comparing fibroblast to epithelial cells. Each line corresponds to a regulated exon, while each column corresponds to a specific cell. Green boxes ( $-1.5 < SI < 0$ ) correspond to a low inclusion level in the cell as compared to all the others; red boxes ( $0 < SI < 1.5$ ), high inclusion level; black boxes ( $SI = 0$ ), no difference for exon inclusion between cells; and gray boxes, missing values. Exons were computationally split into several groups depending on their inclusion rate that correlates with the two major cell types. (C) RT-PCR validations using RNAs from fibroblasts and epithelial cells, as indicated.

Supplemental Table S1), as expected (Warzecha et al. 2009a, 2010; Dittmar et al. 2012). Interestingly, PTBP1 (PTB) and MBNL1, which have been suggested to play a role in angiogenesis (Pascual et al. 2006; Pen et al. 2007; Masuda et al. 2009), were enriched in

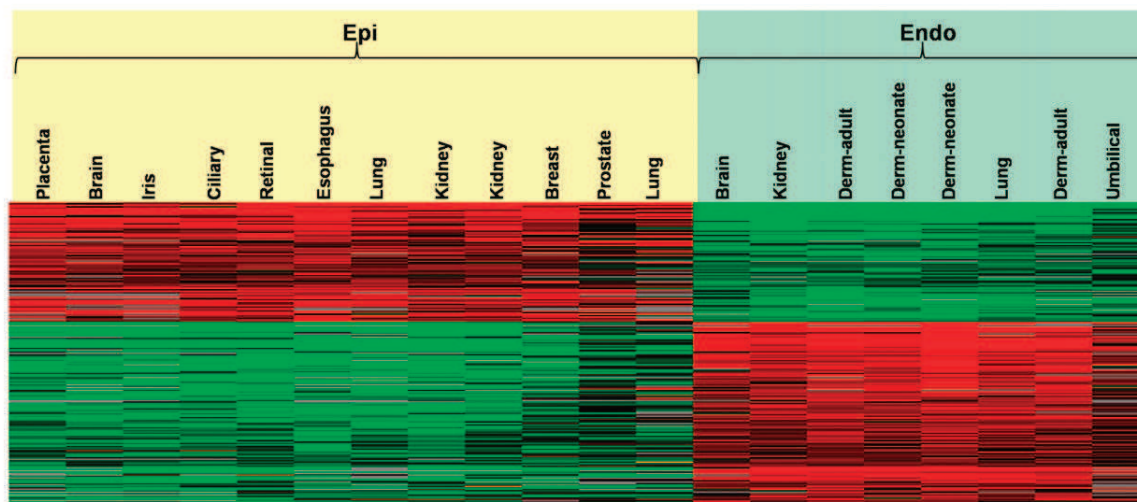
endothelial cells (Fig. 4A,B; Supplemental Table S1). Finally, RBFOX2 (RBM9) and NOVA1 were enriched in fibroblasts (Fig. 4A,B). Interestingly, in addition to sharing many common target exons with NOVA1, RBFOX2 is up-regulated during EMT and plays

195

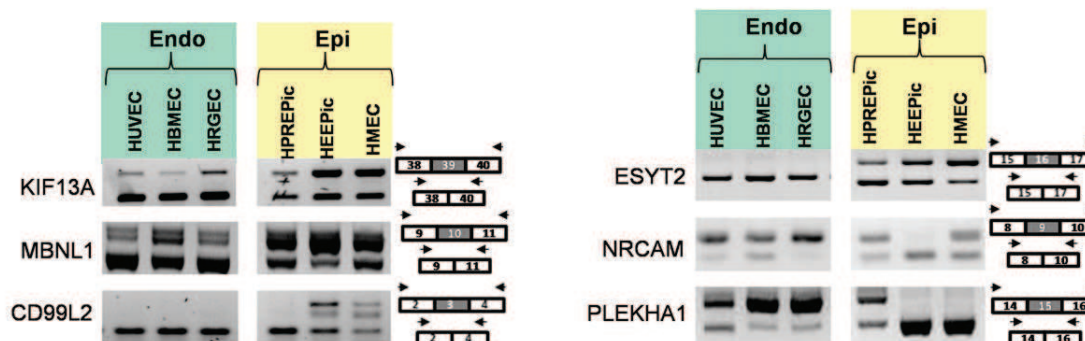
A



B



C



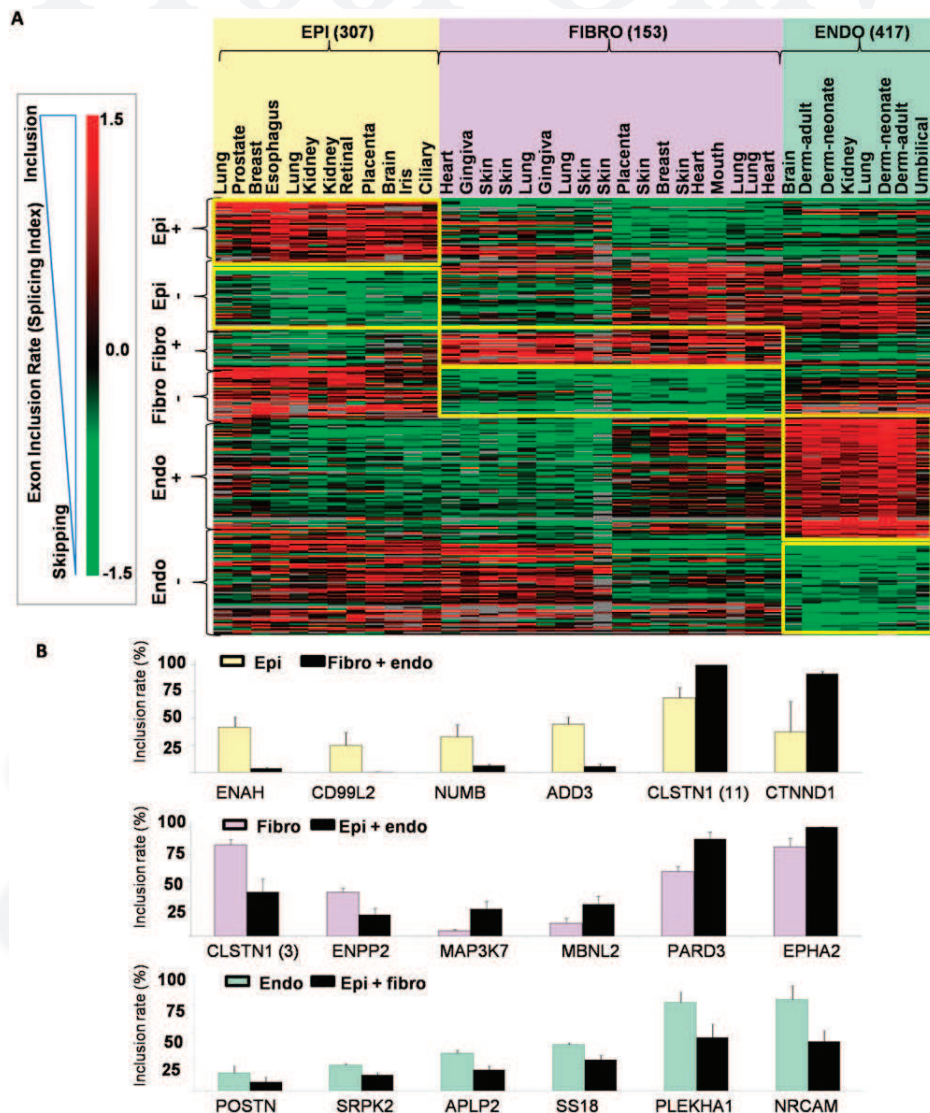
**Figure 2.** Epithelial- and endothelial-specific splicing variants. (A) Same as in Figure 1A but comparing endothelial (Endo) to epithelial (Epi) cells. (B) Same as in Figure 1B but comparing endothelial (Endo) to epithelial (Epi) cells. (C) RT-PCR validations using RNAs from endothelial and epithelial cells, as indicated.

a critical role in this process (Zhang et al. 2008; Venables et al. 2013). These results were confirmed by analyzing the expression level of these splicing factors using RNA-seq and gene expression array data sets (Supplemental Fig. S8).

To go a step further, we focused on the PTBP1, ESRP1, and RBFOX2 splicing factors, since their expression levels allowed each major cell type to be clustered (Supplemental Fig. S7B) and since large-scale data sets corresponding to these factors are publicly available (Warzecha et al. 2009b; Xiao et al. 2009; Xue et al. 2009; Yeo et al. 2009; Katz et al. 2010; Llorian et al. 2010; Huelga et al. 2012). As shown in Figure 4C, ESRP1 and 2 expression levels positively and negatively correlated with the inclusion rate (e.g., SI) of a set of included (EPI+) and excluded (EPI-) exons, respectively.

Inversely, ESRP1 and 2 expression levels negatively and positively correlated with the inclusion rate of a set of included (FIBRO+) and excluded (FIBRO-) exons, respectively. Meanwhile, RBFOX2 expression levels positively and negatively correlated with the inclusion rate of a set of included (FIBRO+) and excluded (FIBRO-) exons, respectively. Finally, PTBP1 expression levels also nicely correlated with the inclusion rate of different exons enriched in endothelial cells.

To further explore the contribution of these splicing factors in controlling cell type-specific splicing programs, we analyzed publicly available data sets of crosslinking/immunoprecipitation-sequencing (CLIP-seq) and/or exon array or RNA-seq analyses after RNAi for PTBP1, ESRP1, and RBFOX2 in human cell lines (Warzecha



**Figure 3.** Cell type-specific splicing programs. (A) Heatmap presentation of the SI values for exons differentially spliced across fibroblast, endothelial, and epithelial cells. Exons were computationally split into several groups depending on their inclusion rate in the three major cell types. Every brace corresponds to a group (names are indicated on the left); their exons are surrounded by a yellow rectangle on the heatmap. The different categories of exons are indicated as follows: epithelial-included (EPI+), epithelial-skipped (EPI-), endothelial-included (ENDO+), endothelial-skipped (ENDO-), fibroblast-included (FIBRO+), and fibroblast-skipped (FIBRO-) exons. (B) Inclusion rate of selected exons measured after RT-PCR using cells from different origins. The gene symbol and exon position are indicated. Inclusion rates are given for epithelial cells compared to both fibroblast and endothelial cells (upper panel), fibroblasts compared to both epithelial and endothelial cells (middle panels), and endothelial cells compared to both fibroblast and epithelial cells (lower panel).

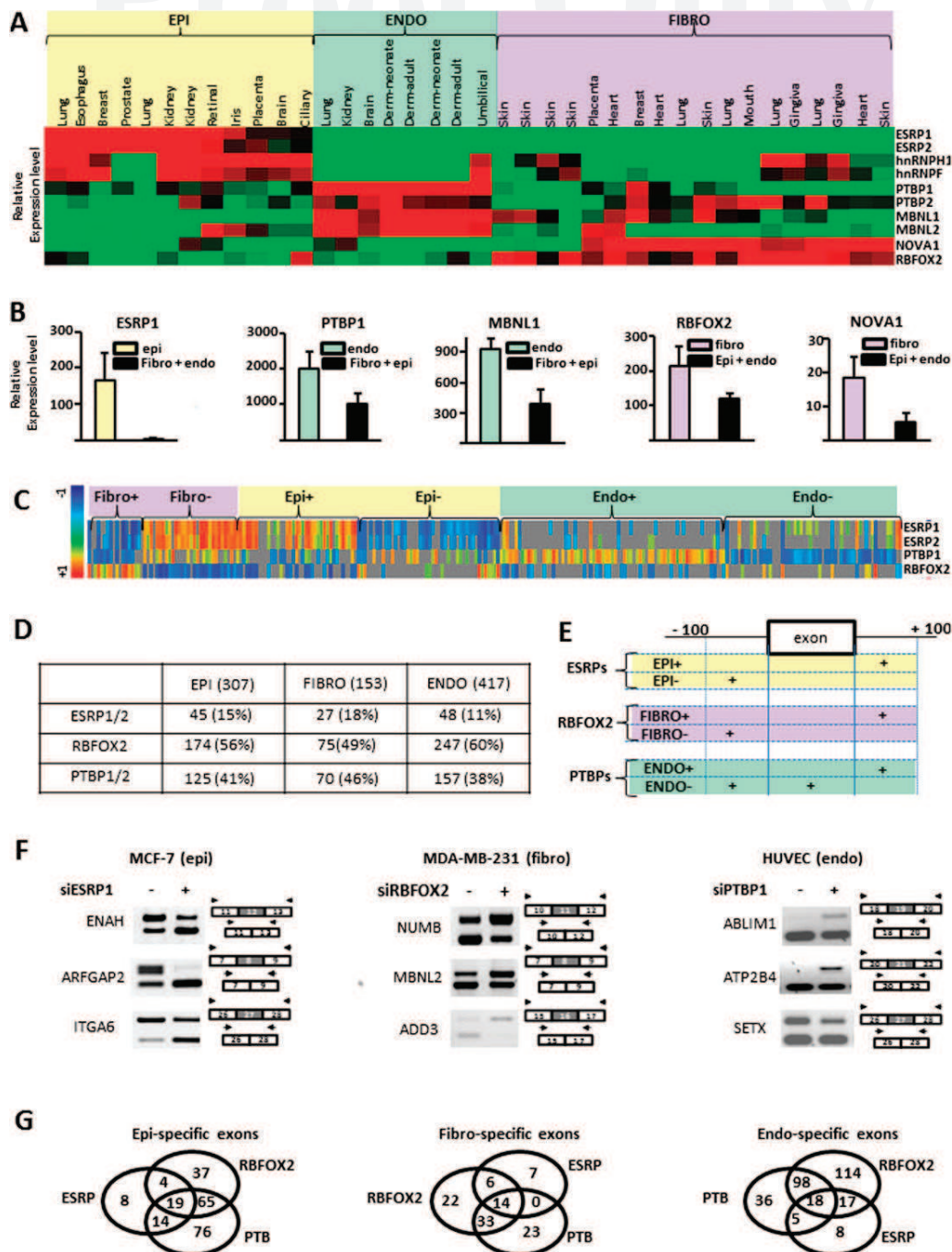
225 et al. 2009b; Xiao et al. 2009; Xue et al. 2009; Yeo et al. 2009; Katz  
 et al. 2010; Llorian et al. 2010; Huelga et al. 2012). These analyses  
 (Fig. 4D; Supplemental Table S8; see Supplemental Fig. S7 for the  
 strategy used) and the RT-PCR validations (Fig. 4F; Supplemental  
 Fig. S7) demonstrated that ESRP1/2, PTBP1/2, and RBFOX2 regulate  
 230 a large subset within the cell type-specific alternative exons.

Additionally, ESRP1/2 binding sites were enriched in introns  
 downstream from epithelial-included exons and upstream of epithelial-  
 excluded exons (Fig. 4E; Supplemental Fig. S9; Supplemental  
 Table S8), as expected from the previously reported ESRP  
 235 splicing code (Warzecha et al. 2009a, 2010; Shapiro et al. 2011;  
 Dittmar et al. 2012). Likewise, RBFOX2 and PTBP1 binding sites

were differentially enriched when comparing fibroblast-included  
 and -excluded exons or endothelial-included and -excluded exons,  
 respectively (Fig. 4E; Supplemental Fig. S9; Supplemental Table S8),  
 in agreement with the previously reported RBFOX2 and PTBP1  
 240 splicing codes (Xue et al. 2009; Llorian et al. 2010).

Remarkably, a large number of fibroblast-, epithelial-, or  
 endothelial-specific exons were predicted to be targeted by at least  
 two splicing factors (Fig. 4G; Supplemental Table S8). Strikingly,  
 as illustrated here for a few genes, some ESRP-regulated exons were  
 245 regulated inversely by RBFOX2, while some RBFOX2-regulated  
 exons were regulated inversely by PTBP1/2 (Supplemental Fig. S7).  
 These data suggest that cell type-specific splicing programs may be





**Figure 4.** Cell type-specific expression of splicing factors. (A) Heatmap of splicing factor expression level. Each line represents a splicing factor, while each column represents a specific cell. The color of the square corresponds to the variation of the expression level of the splicing factor in each specific cell as compared to the others (green, less expressed in the cell; red, more expressed in the cell; and black, no difference). (B) RT-qPCR analysis of the expression level of ESRP1, PTBP1, MBNL1, RBFOX2, and NOVA1 in a collection of fibroblasts, epithelial and endothelial cells. (C) Spearman correlations between splicing factor expression level and the inclusion rate of epithelial-included (EPI+), epithelial-skipped (EPI-), endothelial-included (ENDO+), endothelial-skipped (ENDO-), fibroblast-included (FIBRO+), and fibroblast-skipped (FIBRO-) exons. Warm colors indicate positive correlation (e.g., a high exon inclusion level that correlates with a high splicing factor expression level), whereas cold colors indicate negative correlation (e.g., a low exon inclusion level that correlates with a high splicing factor expression level). Gray boxes indicate correlations that were discarded because of values that were not statistically significant or insufficient data available to compute correlations. (D) Summary table of epithelial-, fibroblast-, and endothelial-specific exons predicted to be regulated by the ESRP1, PTBP1, and RBFOX2 splicing factors using RNA-seq, exon array, and CLIP-seq data sets (see Supplemental Fig. S4 for more information). The number and percentage of exons predicted to be regulated by each splicing factor in each category are indicated. (E) Schematic representation of splicing factor binding site enrichment in several sets of exons differentially regulated across epithelial, endothelial, or fibroblast cells. Columns define regions in which the binding site searches were done. (F) RT-PCR analyses of the effect of depleting ESRP1, RBFOX2, or PTBP1 on alternative splicing of selected genes in the MCF-7 epithelial cell line, the MDA-MB-231 fibroblast-like cell line, or the HUVEC endothelial cell line, respectively. (G) Venn diagram representing the number of epithelial-, fibroblast-, and endothelial-specific exons predicted to be regulated by ESRP1, PTBP1, and/or RBFOX2.

controlled by a balanced expression of antagonist splicing factors. In other words, the splicing signature of a specific cell type likely depends on the combinatorial effects of both up- and down-regulated splicing factors, rather than by only up-regulated factors.

#### FasterDB: An integrative bioinformatics platform dedicated to alternative splicing analysis

In order to provide full support for performing functional genomics at exon-level resolution, the data sets presented above were organized in a freely available and user-friendly web interface. This novel database, named FasterDB (<http://fasterdb.lyon.unicancer.fr/>), provides researchers with information regarding the splicing variants generated for their genes of interest. Supplemental Figures S10–S21 demonstrate the use of FasterDB on the *ENAH* gene, which codes for the MENA protein that modulates cell adhesion and migration (Di Modugno et al. 2012) and which has an exon (exon 12) that is specifically included in epithelial cells (Figs. 1C, 3B). Thus, once the *ENAH* gene symbol is entered into the FasterDB search engine, FasterDB provides the description of the human and mouse genes as well as all known gene transcripts reported in public databases (Supplemental Fig. S10). Information regarding alternative use of exons, various features of exons and introns, exon conservation, UTRs, and miRNA binding site prediction can be obtained as well as (Supplemental Figs. S11–S16).

FasterDB also provides information on the expression and splicing pattern of all protein-coding genes across a collection of 73 human cell lines (both normal and cancerous), 11 normal human tissues, nine mouse cell lines, and 11 normal mouse tissues corresponding to Affymetrix exon array data sets generated by the The ENCODE Project Consortium and Affymetrix (Thurman et al. 2012). Clicking on “Expression” in the main toolbar (Supplemental Fig. S10B) reveals the relative level of gene expression across these panels (Supplemental Fig. S17) and provides the inclusion rate of any selected exon across the sample collection (Supplemental Fig. S18A), which can be helpful for selecting the adequate cellular model for functional studies of alternative splicing variants. To help with data mining, users can click on each cell line for a direct link to a dedicated visualization interface, named ELEXIR (Supplemental Figs. S18, S19).

Finally, FasterDB provides some clues about the splicing factors that might be involved in splicing regulation of their favorite gene. Indeed, the “Splicing factors” button in the main tool bar (Supplemental Fig. S10B) allows users to select different splicing factors and to access to different kinds of information based on publicly available data sets (Hung et al. 2008; Warzecha et al. 2009b; Xiao et al. 2009, 2012; Xue et al. 2009; Yeo et al. 2009; Katz et al. 2010; Llorian et al. 2010; Wang et al. 2010; Grellscheid et al. 2011; Lebedeva et al. 2011; Mukherjee et al. 2011; Huelga et al. 2012; Lagier-Tourenne et al. 2012; Zarnack et al. 2013). For example, selecting hnRNPH/F on the “Splicing factors” screen (Supplemental Fig. S20) gives access to predicted binding motifs (Fig. 5A) as well as to a dedicated CLIP-seq data visualization interface (Supplemental Fig. S21) based on the data set generated by Huelga and Katz (Katz et al. 2010). From there, it is possible to zoom in and view the *in cellulo* binding sites of the selected splicing factor in the vicinity of the selected exon. For example, two hnRNPF binding sites were identified around *ENAH* exon 12, suggesting that hnRNPH/F might regulate *ENAH* splicing (Fig. 5B). To further challenge this possibility, the “Exon Arrays” button (Fig. 5A) gives access to Affymetrix exon array data obtained after hnRNPH/F depletion by Xiao and collaborators (Xiao et al. 2009). As shown in

Figure 5C (upper panel), *ENAH* exon 12 probes appear red when comparing hnRNPH/F-depleted cells to control cells. This suggests that hnRNPH/F depletion favors exon inclusion, as validated by RT-PCR (Fig. 5C). Meanwhile, analysis of the ESRP data set generated by Warzecha et al. (2009b) predicts that ESRP depletion induces skipping of this exon (Fig. 5C, lower panel), as validated by RT-PCR (Fig. 4F). In sum, FasterDB will provide support for performing functional genomics at exon-level resolution by integrating publicly available large-scale data sets.

#### Discussion

It is well established that different tissues or organs (e.g., liver or kidney) express different splicing variants (Pan et al. 2008; Wang et al. 2008; Bland et al. 2010; de la Grange et al. 2010; Hartmann et al. 2011; Llorian and Smith 2011; Barbosa-Morais et al. 2012; Merkin et al. 2012). However, all organs are composed of common major cell types like fibroblast, epithelial, and endothelial cells. In this report, we demonstrate that each major cell type expresses a specific splicing program independently of their organ origin (Figs. 1–3). It will be interesting to next determine whether a major cell type (e.g., endothelial cells) isolated from different tissues also expresses tissue-specific splicing programs. Looking at each column representing one specific cell type from a given organ in the clustering analyses (Figs. 1, 2; Supplemental Figs. S1–S3), the splicing pattern seems likely to integrate at least two levels of specificity, of cell type and tissue origin (Supplemental Fig. S22). Our observation has several consequences. For example, some tissue-specific splicing variants previously identified by comparing different tissues could in fact reflect the different relative proportions between common cell types in those tissues, as illustrated in Supplemental Figure S22. This could be particularly relevant for tumors that are often compared to normal control tissues to identify cancer-associated splicing variants. As many cancer cells derive from epithelial cells, some previously reported cancer-associated splicing variants could reflect epithelial cell enrichment in tumors as compared to normal tissues.

Remarkably, it has been previously shown that clustering cells based on global gene expression level also reveals that most cell lines cluster together rather than with their tissues of origin (Lukk et al. 2010). As similar results were obtained using alternative spliced exons (ASE) (Fig. 3) and alternative first or last exons (AFE or ALE, respectively) (Supplemental Figs. S5, S6), this demonstrates that major cell types not only differ by the set of genes they expressed but also by the exonic content (e.g., ASE, AFE, and ALE) of the transcripts produced by the expressed genes. Therefore, our data imply that understanding the function of cell types not only requires the characterization of expressed genes but also of splicing variants generated by these expressed genes. In this context, it is interesting to underscore that genes that are differentially spliced in fibroblasts, epithelial, and endothelial cells are often involved in cell–cell or cell–substrate interactions (Supplemental Fig. S23). This suggests that alternative splicing may play a role in modeling interactions with neighboring cells and with the microenvironment.

We also identified three splicing factors, ESRP1, PTBP1, and RBFOX2, which are likely to play a major role in setting up these cell type-specific splicing programs (Fig. 4). Obviously, other splicing factors may contribute as well (Supplemental Table S8). For example, the highly related hnRNPH1 and hnRNPF splicing factors, which are in the same family as ESRP1 and are also involved in EMT (Warzecha et al. 2009a, 2010; Shapiro et al. 2011; Dittmar et al. 2012), were shown to also have an enriched ex-



**Figure 5.** Dedicated CLIP-seq data visualization web interface. (A) Localization of predicted hnRNP/F binding motifs in the vicinity of the *ENAH* exon 12 and their link to splicing factor data sets. (B) Reads from a CLIP-seq experiment corresponding to hnRNP/F binding sites in the vicinity of the *ENAH* exon 12. (C) Visualization of exon array probe intensities in hnRNP/F- or ESRP1-depleted cells as compared to control cells. RT-PCR analysis of the *ENAH* exon 12 inclusion rate in hnRNP/F-depleted or control cells.

pression in epithelial cells (Fig. 4A; Supplemental Table S8). Of note, although NOVA1 plays a role in establishing a neuronal splicing program, we observed that NOVA1 expression was enriched in fibroblasts, similar to RBFOX2 (Supplemental Table S8). Interestingly, it has been shown that the targets of RBFOX2 and of NOVA1 often overlap (Zhang et al. 2008). Conducting splicing analysis in a cell type-specific manner might allow new cellular functions of splicing factors to be identified. In this context, our data suggest a role of the PTBP1 and MBNL1 splicing factors in endothelial cell biology (Supplemental Table S8). Interestingly, PTBP1 and MBNL1 have been suggested to play a role in angiogenesis (Pascual et al. 2006; Pen et al. 2007; Masuda et al. 2009).

Functional genomics now needs to be performed at an exonic resolution level, given the prevalence and importance of alternative splicing. This will require new integrative bioinformatic approaches to be developed that can handle the diversity of alternative exons and the complexity of alternative splicing regulation. In this setting, we have made the analyzed data sets used in this report freely available through the user friendly web interface of FasterDB, which describes all known splicing variants of human and mouse genes and their splicing pattern across several dozen normal and cancer cells and tissues, as well as information regarding which splicing factors contribute to individual exon regulation. The aim of this database is to help researchers identify the different splicing variants of their favorite genes as well as the



395 tissue, cell type, or cell line in which they are expressed, in order to  
facilitate further functional and/or mechanistic studies. For exam-  
400 ple, when researchers identify a splicing variant in physio-  
pathological conditions, FasterDB will be useful in deciding which  
cellular model might be suitable for functional analysis.

The second aim of FasterDB is to help researchers charac-  
405 terize the splicing factors regulating the identified splicing vari-  
ants. To the best of our knowledge, FasterDB is the first database  
integrating large-scale data sets focused on splicing, including  
CLIP-seq and splicing-sensitive microarray data sets (Hung et al.  
2008; Warzecha et al. 2009b; Xiao et al. 2009, 2012; Xue et al.  
2009; Yeo et al. 2009; Katz et al. 2010; Llorian et al. 2010; Wang  
410 et al. 2010; Grellscheid et al. 2011; Lebedeva et al. 2011; Mukherjee  
et al. 2011; Harrow et al. 2012; Huelga et al. 2012; Lagier-Tourenne  
et al. 2012; Zarnack et al. 2013). Our next goal will be to include  
415 other data sets, such as RNA-seq data sets. We also aim to label the  
protein domains coded by each alternatively spliced exon in order  
to help users predict potential functional consequences resulting  
from alternative splicing.

## Methods

### 415 FasterDB core database

Human and mouse exons were collected from Ensembl (release 60,  
assemblies GRCh37 and NCBI m37) (Flicek et al. 2013) and aligned  
against the NCBI transcript database using Mega Blast (v 2.2.25).  
420 These exons were aligned against genomic sequences to define  
their chromosomal coordinates and then clustered by genomic  
position to define seven major events (e.g., alternative first exon,  
alternative last exon, alternative 3' splice site, alternative 5' splice  
site, intron retention, exon deletion, and exon skipping). Scores  
were computed using MaxEntScan for each splice site. For each  
425 gene, a nonredundant repertory of untranslated regions was  
established using all corresponding transcripts. UTRs were more  
fully characterized by describing the motifs found in their se-  
quences using PatSearch Tool (Grillo et al. 2003). miRNA binding  
sites were predicted using PITA, miRanda, and PicTar (John et al.  
430 2004; Lewis et al. 2005; Kertesz et al. 2007). Conserved exons be-  
tween human and mouse were identified by aligning each human  
exon against the exons of its orthologous mouse gene as provided  
by Ensembl. The "in silico PCR" tool is based on a multialignment  
of the transcript exons performed with ClustalW. FasterDB is built  
435 in Perl (v5.14.2) and runs on a Ubuntu server (v12.04.1) that hosts  
Apache (v2.2.16) and MySQL (v5.1.49) servers. More information  
is available in Supplemental Material and can be downloaded from  
<http://fasterdb.lyon.unicancer.fr>.

### Exon array data set analyses

440 Cell lines and tissues expression data were downloaded from the  
GEO and Affymetrix websites and corresponded to Affymetrix  
Exon Arrays data sets as listed in Supplemental Table S10. The  
preprocessing pipeline is described in more detail in the Supple-  
445 mental Material (Part 2, Section 2.2, page 9). Briefly, low-quality  
and cross-hybridizing probes were removed, and signals were  
summarized at the gene level by using the median over all  
remaining probes. Three different ways of computing the inclusion/  
exclusion rates of a given exon were computed by measuring the  
450 ratio of (1) its expression level versus the gene expression level  
(NI); (2) its NI in the condition of interest relative to the NI in the  
control condition (global SI); or (3) its expression level versus the  
expression level of the flanked exons (local SI). Computation of  
the global and local SI is described in more detail in the Supplemental

Material (Part 2, section 2.2, page 10). Identification of cell type-  
specific alternative exons was performed after computing local SI and  
455 selecting exons with local SI above 1.45 and with a *P*-value <0.05.  
Heatmaps were generated using the multiexperiment viewer appli-  
cation of the TM4 package. Hierarchical clustering of Mev4 was  
also used to cluster cell lines that have similar regulation. Input  
460 data contains the SI of each of the regulated ASEs in each of the  
analyzed cell lines. The SI was computed for each cell line in  
comparison with all the different types of cell lines. Functional  
enrichment and KEGG pathway mapping were done using DAVID  
(Huang da et al. 2009).

### Elexir web interface

465 Elexir is a web application that allows users to choose an experi-  
ment within a set of stored exon array experiments. This interface  
was developed to allow the end user to easily browse and query the  
expression levels of one or more genes between conditions, with  
possible replicates for each condition. Different test conditions can  
470 be chosen for each experiment. A condition can be defined as a cell  
line, a tissue type, or a treatment. Paired or unpaired analyses can  
be done depending on sample relationships used for test and  
control conditions. The intensity report displays a schematic graph  
of the gene being analyzed with corresponding probes for each  
475 exon. The height of each bar represents the normalized probe in-  
tensity (in log<sub>2</sub>). The color reflects the ratio of probe intensity  
between test and control conditions. Green and red indicate that  
the probe intensity in the test condition is lower and higher, re-  
spectively, than the control condition. Information for each probe  
480 is given in the underlying descriptive table or can be easily ob-  
tained by clicking on the bar.

### Splicing factor analyses

The regulation of cell type-specific ASEs by specific splicing factors  
was tested using three different sources: microarray data, literature  
485 RNA-seq, and CLIP-seq data sets. Chromosomal coordinates of  
exons regulated by ESRP1, PTBP1, and/or RBFOX2 identified by  
RNA-seq were retrieved from previously published work (Warzecha  
et al. 2009b; Xiao et al. 2009; Xue et al. 2009; Yeo et al. 2009; Katz  
490 et al. 2010; Llorian et al. 2010; Huelga et al. 2012). ASEs predicted  
to be regulated by a splicing factor using exon array or RNA-seq  
data sets were compared to ASEs predicted to be cell type-specific.  
For each cell type, we paid attention to the ASE regulation sense.  
For example, PTBP1 is up-regulated, while ESRP1 and RBFOX2 are  
495 down-regulated, in endothelial cells; thus, an endothelial ASE is  
predicted to be regulated by one of these factors if it is skipped  
upon PTBP1 depletion or included by ESRP1 or RBFOX2 depletion.  
To be considered as confident, CLIP hits must have been detected  
in the exon or within 100 nt upstream of or downstream from  
500 the exon.

Splicing factor binding sites were searched with the PatSearch  
Tool (Grillo et al. 2003) for the genomic sequences using previously  
defined splicing factor binding motifs (see Supplemental Table S8).  
A set of 1000 randomly selected alternative exons, and a set of 1000  
505 randomly selected constitutive exons, were used as controls. Four  
regions were defined: 100 nt upstream of and downstream from  
the exon, the first 60 nt of the exon and the last 60 nt of the exon.  
For each region, a sliding window of a specific length was con-  
sidered, and the enrichment score of the splicing factor was com-  
450 puted at each position as ( $\Sigma$  number of factor motifs at position X /  
total number of analyzed sequences)  $\times$  100. The total number of  
binding sites found in each region of each exon group was used to  
compute the standard deviation value.

## Experimental validation

515 RNA from different cell types were purchased as indicated in Supplemental Table S1. Primers used for RT-PCR are described in Supplemental Table S9. RT-qPCR analyses were performed using primers described in Supplemental Table S9. Epithelial MCF-7, fibroblast-like MDA-MB-231 cells, and endothelial HUVEC cells were transfected using RNAiMax (Invitrogen) with control siRNAs or siRNAs against ESRP1 and 2, RBFOX2, or PTBP1 and 2 (Supplemental Table S9) 48 h before RNA extraction. Electrophoretic gels were analyzed with ImageJ software.

## Data access

525 The FasterDB database is available at <http://fasterdb.lyon.unicancer.fr/>.

## Acknowledgments

We thank the researchers who deposited their data sets in public libraries. We hope that our website will be welcomed by both users and the researchers who generated the data sets and that we did not forget to acknowledge any of those researchers in Supplemental Table S10 and on the FasterDB website. This work was supported by the Fondation pour la Recherche Médicale (FRM), INSERM Plan Cancer 2009-2013, Institut National du Cancer (INCa), Agence Nationale de la Recherche (ANR), and Association Française Contre les Myopathies (AFM). P.M. was supported by Lyon Science Transfert, J.-P.V. and H.M. by FRM, F.-O.D. by the Association pour la Recherche sur le Cancer, M.P.-E. by AFM, S.S. by the Ligue Nationale Contre le Cancer, L.-C.T. and E.C. by INSERM “Plan Cancer 2009-2013.”

## References

- Barash Y, Calarco JA, Gao W, Pan Q, Wang X, Shai O, Blencowe BJ, Frey BJ. 2010. Deciphering the splicing code. *Nature* **465**: 53–59.
- Barbosa-Morais NL, Irimia M, Pan Q, Xiong HY, Gueroussov S, Lee LJ, Slobodeniuc V, Kutter C, Watt S, Colak R, et al. 2012. The evolutionary landscape of alternative splicing in vertebrate species. *Science* **338**: 1587–1593.
- Bland CS, Wang ET, Vu A, David MP, Castle JC, Johnson JM, Burge CB, Cooper TA. 2010. Global regulation of alternative splicing during myogenic differentiation. *Nucleic Acids Res* **38**: 7651–7664.
- 550 Blencowe BJ. 2012. An exon-centric perspective. *Biochem Cell Biol* **90**: 603–612.
- de la Grange P, Gratadou L, Delord M, Dutertre M, Auboeuf D. 2010. Splicing factor and exon profiling across human tissues. *Nucleic Acids Res* **38**: 2825–2838.
- 555 Di Modugno F, Iapicca P, Boudreau A, Mottolose M, Terrenato I, Perracchio L, Carstens RP, Santoni A, Bissell MJ, Nistico P. 2012. Splicing program of human MENA produces a previously undescribed isoform associated with invasive, mesenchymal-like breast tumors. *Proc Natl Acad Sci* **109**: 19280–19285.
- 560 Dittmar KA, Jiang P, Park JW, Amirikian K, Wan J, Shen S, Xing Y, Carstens RP. 2012. Genome-wide determination of a broad ESRP-regulated posttranscriptional network by high-throughput sequencing. *Mol Cell Biol* **32**: 1468–1482.
- Flicek P, Ahmed I, Amode MR, Barrell D, Beal K, Brent S, Carvalho-Silva D, Clapham P, Coates G, Fairley S, et al. 2013. Ensembl 2013. *Nucleic Acids Res* **41**: D48–D55.
- Gertler F, Condeelis J. 2011. Metastasis: Tumor cells becoming MENAcing. *Trends Cell Biol* **21**: 81–90.
- AU3** Goren A, Kim E, Amit M, Vaknin K, Kfir N, Ram O, Ast G. 2010. Overlapping splicing regulatory motifs—combinatorial effects on splicing. *Nucleic Acids Res* **38**: 3318–3327.
- 570 Grellscheid S, Dalglish C, Storbeck M, Best A, Liu Y, Jakubik M, Mende Y, Ehrmann I, Curk T, Rossbach K, et al. 2011. Identification of evolutionarily conserved exons as regulated targets for the splicing activator tra2 $\beta$  in development. *PLoS Genet* **7**: e1002390.
- 575 Grillo G, Licciulli F, Liuni S, Sbsa E, Pesole G. 2003. PatSearch: A program for the detection of patterns and structural motifs in nucleotide sequences. *Nucleic Acids Res* **31**: 3608–3612.
- Halleger M, Llorian M, Smith CW. 2010. Alternative splicing: Global insights. *FEBS J* **277**: 856–866.
- Hartmann B, Valcarcel J. 2009. Decrypting the genome’s alternative messages. *Curr Opin Cell Biol* **21**: 377–386.
- Hartmann B, Castelo R, Minana B, Peden E, Blanchette M, Rio DC, Singh R, Valcarcel J. 2011. Distinct regulatory programs establish widespread sex-specific alternative splicing in *Drosophila melanogaster*. *RNA* **17**: 453–468.
- Huang da W, Sherman BT, Lempicki RA. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**: 44–57.
- 590 Huelga SC, Vu AQ, Arnold JD, Liang TY, Liu PP, Yan BY, Donohue JP, Shiue L, Hoon S, Brenner S, et al. 2012. Integrative genome-wide analysis reveals cooperative regulation of alternative splicing by hnRNP proteins. *Cell Rep* **1**: 167–178.
- Hung LH, Heiner M, Hui J, Schreiner S, Benes V, Bindereif A. 2008. Diverse roles of hnRNP L in mammalian mRNA processing: A combined microarray and RNAi analysis. *RNA* **14**: 284–296.
- John B, Enright AJ, Aravin A, Tuschl T, Sander C, Marks DS. 2004. Human MicroRNA targets. *PLoS Biol* **2**: e363.
- Kalsotra A, Cooper TA. 2011. Functional consequences of developmentally regulated alternative splicing. *Nat Rev Genet* **12**: 715–729.
- 600 Katz Y, Wang ET, Airolidi EM, Burge CB. 2010. Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat Methods* **7**: 1009–1015.
- Kelemen O, Convertini P, Zhang Z, Wen Y, Shen M, Falaleeva M, Stamm S. 2013. Function of alternative splicing. *Gene* **514**: 1–30.
- Kertesz M, Iovino N, Unnerstall U, Gaul U, Segal E. 2007. The role of site accessibility in microRNA target recognition. *Nat Genet* **39**: 1278–1284.
- Kim E, Goren A, Ast G. 2008. Alternative splicing and disease. *RNA Biol* **5**: 17–19.
- 610 Lagier-Tourenne C, Polymenidou M, Hutt KR, Vu AQ, Baughn M, Huelga SC, Clutario KM, Ling SC, Liang TY, Mazur C, et al. 2012. Divergent roles of ALS-linked proteins FUS/TLS and TDP-43 intersect in processing long pre-mRNAs. *Nat Neurosci* **15**: 1488–1497.
- Lebedeva S, Jens M, Theil K, Schwanhauser B, Selbach M, Landthaler M, Rajewsky N. 2011. Transcriptome-wide analysis of regulatory interactions of the RNA-binding protein HuR. *Mol Cell* **43**: 340–352.
- 615 Lewis BP, Burge CB, Bartel DP. 2005. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**: 15–20.
- Llorian M, Smith CW. 2011. Decoding muscle alternative splicing. *Curr Opin Genet Dev* **21**: 380–387.
- Llorian M, Schwartz S, Clark TA, Hollander D, Tan LY, Spellman R, Gordon A, Schweitzer AC, de la Grange P, Ast G, et al. 2010. Position-dependent alternative splicing activity revealed by global profiling of alternative splicing events regulated by PTB. *Nat Struct Mol Biol* **17**: 1114–1123.
- 625 Lukk M, Kapushesky M, Nikkilä J, Parkinson H, Goncalves A, Huber W, Ukkonen E, Brazma A. 2010. A global map of human gene expression. *Nat Biotechnol* **4**: 322–324.
- Masuda K, Abdelmohsen K, Gorospe M. 2009. RNA-binding proteins implicated in the hypoxic response. *J Cell Mol Med* **13**: 2759–2769.
- 630 Merkin J, Russell C, Chen P, Burge CB. 2012. Evolutionary dynamics of gene and isoform regulation in Mammalian tissues. *Science* **338**: 1593–1599.
- Mukherjee N, Corcoran DL, Nusbaum JD, Reid DW, Georgiev S, Hafner M, Asciano M Jr, Tuschl T, Ohler U, Keene JD. 2011. Integrative regulatory mapping indicates that the RNA-binding protein HuR couples pre-mRNA processing and mRNA stability. *Mol Cell* **43**: 327–339.
- 635 Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. 2008. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet* **40**: 1413–1415.
- Pascual M, Vicente M, Monferrer L, Artero R. 2006. The Muscleblind family of proteins: An emerging class of regulators of developmentally programmed alternative splicing. *Differentiation* **74**: 65–80.
- 640 Pen A, Moreno MJ, Martin J, Stanimirovic DB. 2007. Molecular markers of extracellular matrix remodeling in glioblastoma vessels: Microarray study of laser-captured glioblastoma vessels. *Glia* **55**: 559–572.
- Shapiro IM, Cheng AW, Flytzanis NC, Balsamo M, Condeelis JS, Oktay MH, Burge CB, Gertler FB. 2011. An EMT-driven alternative splicing program occurs in human breast cancer and modulates cellular phenotype. *PLoS Genet* **7**: e1002218.
- 645 Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, Sheffield NC, Stergachis AB, Wang H, Vernot B, et al. 2012. The accessible chromatin landscape of the human genome. *Nature* **489**: 75–82.
- 655 Venables JP, Brosseau JP, Gadea G, Klinck R, Prinos P, Beaulieu JF, Lapointe E, Durand M, Thibault P, Tremblay K, et al. 2013. RBFOX2 is an important regulator of mesenchymal tissue-specific splicing in both normal and cancer tissues. *Mol Cell Biol* **33**: 396–405.



- 660 Wang ET, Sandberg R, Luo S, Khrebtkova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, Burge CB. 2008. Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**: 470–476.
- 665 Wang Z, Kayikci M, Briese M, Zarnack K, Luscombe NM, Rot G, Zupan B, Curk T, Ule J. 2010. iCLIP predicts the dual splicing effects of TIA-RNA interactions. *PLoS Biol* **8**: e1000530.
- Warzecha CC, Sato TK, Nabet B, Hogenesch JB, Carstens RP. 2009a. ESRP1 and ESRP2 are epithelial cell-type-specific regulators of FGFR2 splicing. *Mol Cell* **33**: 591–601.
- 670 Warzecha CC, Shen S, Xing Y, Carstens RP. 2009b. The epithelial splicing factors ESRP1 and ESRP2 positively and negatively regulate diverse types of alternative splicing events. *RNA Biol* **6**: 546–562.
- Warzecha CC, Jiang P, Amirikian K, Dittmar KA, Lu H, Shen S, Guo W, Xing Y, Carstens RP. 2010. An ESRP-regulated splicing programme is abrogated during the epithelial-mesenchymal transition. *EMBO J* **29**: 3286–3300.
- 675 Witten JT, Ule J. 2011. Understanding splicing regulation through RNA splicing maps. *Trends Genet* **27**: 89–97.
- Xiao X, Wang Z, Jang M, Nutiu R, Wang ET, Burge CB. 2009. Splice site strength-dependent activity and genetic buffering by poly-G runs. *Nat Struct Mol Biol* **16**: 1094–1100.
- Xiao R, Tang P, Yang B, Huang J, Zhou Y, Shao C, Li H, Sun H, Zhang Y, Fu XD. 2012. Nuclear matrix factor hnRNP U/SAF-A exerts a global control of alternative splicing by regulating U2 snRNP maturation. *Mol Cell* **45**: 656–668.
- Xue Y, Zhou Y, Wu T, Zhu T, Ji X, Kwon YS, Zhang C, Yeo G, Black DL, Sun H, et al. 2009. Genome-wide analysis of PTB-RNA interactions reveals a strategy used by the general splicing repressor to modulate exon inclusion or skipping. *Mol Cell* **36**: 996–1006.
- 685 Yeo GW, Coufal NG, Liang TY, Peng GE, Fu XD, Gage FH. 2009. An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells. *Nat Struct Mol Biol* **16**: 130–137.
- 690 Zarnack K, Konig J, Tajnik M, Martincorena I, Eustermann S, Stevant I, Reyes A, Anders S, Luscombe NM, Ule J. 2013. Direct competition between hnRNP C and U2AF65 protects the transcriptome from the exonization of *Alu* elements. *Cell* **152**: 453–466.
- 695 Zhang C, Zhang Z, Castle J, Sun S, Johnson J, Krainer AR, Zhang MQ. 2008. Defining the regulatory network of the tissue-specific splicing factors Fox-1 and Fox-2. *Genes Dev* **22**: 2550–2563.

Received July 3, 2013; accepted in revised form December 2, 2013.

# Bibliographie

- Abdelhaleem, M. (2005). Rna helicases : regulators of differentiation. *Clin Biochem*, 38(6) :499–503.
- Albini, S. and Puri, P. L. (2010). Swi/snf complexes, chromatin remodeling and skeletal myogenesis : it's time to exchange! *Exp Cell Res*, 316(18) :3073–3080.
- Alló, M., Schor, I. E., Muñoz, M. J., de la Mata, M., Agirre, E., Valcárcel, J., Eyraas, E., and Kornblihtt, A. R. (2010). Chromatin and alternative splicing. *Cold Spring Harb Symp Quant Biol*, 75 :103–111.
- Anantharaman, V., Koonin, E. V., and Aravind, L. (2002). Comparative genomics and evolution of proteins involved in rna metabolism. *Nucleic Acids Res*, 30(7) :1427–1464.
- Andreadis, A. (2006). Misregulation of tau alternative splicing in neurodegeneration and dementia. *Prog Mol Subcell Biol*, 44 :89–107.
- Andreadis, A., Gallego, M. E., and Nadal-Ginard, B. (1987). Generation of protein isoform diversity by alternative splicing : mechanistic and biological implications. *Annu Rev Cell Biol*, 3 :207–242.
- Angear, N., Jacquette, A., Gargiulo, M., Radvanyi, H., Moutier, S., Eymard, B., and Héron, D. (2011). A new window on neurocognitive dysfunction in the childhood form of myotonic dystrophy type 1 (dm1). *Neuromuscul Disord*, 21(7) :468–476.
- Antonini, G., Soscia, F., Giubilei, F., De Carolis, A., Gagnani, F., Morino, S., Ruberto, A., and Tatarelli, R. (2006). Health-related quality of life in myotonic dystrophy type 1 and its relationship with cognitive and emotional functioning. *J Rehabil Med*, 38(3) :181–185.
- Arsenault, M.-E., Prévost, C., Lescault, A., Laberge, C., Puymirat, J., and Mathieu,

- J. (2006). Clinical characteristics of myotonic dystrophy type 1 patients with small ctg expansions. *Neurology*, 66(8) :1248–1250.
- Aziz, A., Liu, Q.-C., and Dilworth, F. J. (2010). Regulating a master regulator : establishing tissue-specific gene expression in skeletal muscle. *Epigenetics*, 5(8) :691–695.
- Baltz, A. G., Munschauer, M., Schwanhäusser, B., Vasile, A., Murakawa, Y., Schueler, M., Youngs, N., Penfold-Brown, D., Drew, K., Milek, M., Wyler, E., Bonneau, R., Selbach, M., Dieterich, C., and Landthaler, M. (2012). The mrna-bound proteome and its global occupancy profile on protein-coding transcripts. *Mol Cell*, 46(5) :674–690.
- Bani-Yaghoub, M., Kubu, C. J., Cowling, R., Rochira, J., Nikopoulos, G. N., Bellum, S., and Verdi, J. M. (2007). A switch in numb isoforms is a critical step in cortical development. *Dev Dyn*, 236(3) :696–705.
- Baraniak, A. P., Lasda, E. L., Wagner, E. J., and Garcia-Blanco, M. A. (2003). A stem structure in fibroblast growth factor receptor 2 transcripts mediates cell-type-specific splicing by approximating intronic control elements. *Mol Cell Biol*, 23(24) :9327–9337.
- Barash, Y., Calarco, J. A., Gao, W., Pan, Q., Wang, X., Shai, O., Blencowe, B. J., and Frey, B. J. (2010). Deciphering the splicing code. *Nature*, 465(7294) :53–59.
- Barta, I. and Iggo, R. (1995). Autoregulation of expression of the yeast dbp2p 'dead-box' protein is mediated by sequences in the conserved dbp2 intron. *EMBO J*, 14(15) :3800–3808.
- Bates, G. J., Nicol, S. M., Wilson, B. J., Jacobs, A.-M. F., Bourdon, J.-C., Wardrop, J., Gregory, D. J., Lane, D. P., Perkins, N. D., and Fuller-Pace, F. V. (2005). The dead box protein p68 : a novel transcriptional coactivator of the p53 tumour suppressor. *EMBO J*, 24(3) :543–553.
- Belkin, A. M., Zhidkova, N. I., Balzac, F., Altruda, F., Tomatis, D., Maier, A., Tarone, G., Koteliensky, V. E., and Burrridge, K. (1996). Beta 1d integrin displaces

- the beta 1a isoform in striated muscles : localization at junctional structures and signaling potential in nonmuscle cells. *J Cell Biol*, 132(1-2) :211–226.
- Bellini, M., Biagi, S., Stasi, C., Costa, F., Mumolo, M. G., Ricchiuti, A., and Marchi, S. (2006). Gastrointestinal manifestations in myotonic muscular dystrophy. *World J Gastroenterol*, 12(12) :1821–1828.
- Benezra, R., Davis, R. L., Lockshon, D., Turner, D. L., and Weintraub, H. (1990). The protein id : a negative regulator of helix-loop-helix dna binding proteins. *Cell*, 61(1) :49–59.
- Bentzinger, C. F., Wang, Y. X., and Rudnicki, M. A. (2012). Building muscle : molecular regulation of myogenesis. *Cold Spring Harb Perspect Biol*, 4(2).
- Berget, S. M., Moore, C., and Sharp, P. A. (1977). Spliced segments at the 5' terminus of adenovirus 2 late mrna. *Proc Natl Acad Sci U S A*, 74(8) :3171–3175.
- Bharathy, N. and Taneja, R. (2012). Methylation muscles into transcription factor silencing. *Transcription*, 3(5) :215–220.
- Bird, T. D. (1993-2013). Myotonic dystrophy type 1.
- Black, B. L. and Olson, E. N. (1998). Transcriptional control of muscle development by myocyte enhancer factor-2 (mef2) proteins. *Annu Rev Cell Dev Biol*, 14 :167–196.
- Blais, A., Tsikitis, M., Acosta-Alvear, D., Sharan, R., Kluger, Y., and Dynlacht, B. D. (2005). An initial blueprint for myogenic differentiation. *Genes Dev*, 19(5) :553–569.
- Bland, C. S., Wang, E. T., Vu, A., David, M. P., Castle, J. C., Johnson, J. M., Burge, C. B., and Cooper, T. A. (2010). Global regulation of alternative splicing during myogenic differentiation. *Nucleic Acids Res*, 38(21) :7651–7664.
- Blau, H. M., Pavlath, G. K., Hardeman, E. C., Chiu, C. P., Silberstein, L., Webster, S. G., Miller, S. C., and Webster, C. (1985). Plasticity of the differentiated state. *Science*, 230(4727) :758–766.

- Blencowe, B. J., Ahmad, S., and Lee, L. J. (2009). Current-generation high-throughput sequencing : deepening insights into mammalian transcriptomes. *Genes Dev*, 23(12) :1379–1386.
- Bond, A. T., Mangus, D. A., He, F., and Jacobson, A. (2001). Absence of dbp2p alters both nonsense-mediated mrna decay and rna processing. *Mol Cell Biol*, 21(21) :7366–7379.
- Botta, A., Vallo, L., Rinaldi, F., Bonifazi, E., Amati, F., Biancolella, M., Gambardella, S., Mancinelli, E., Angelini, C., Meola, G., and Novelli, G. (2007). Gene expression analysis in myotonic dystrophy : indications for a common molecular pathogenic pathway in dm1 and dm2. *Gene Expr*, 13(6) :339–351.
- Boutz, P. L., Chawla, G., Stoilov, P., and Black, D. L. (2007). Micrnas regulate the expression of the alternative splicing factor nptb during muscle development. *Genes Dev*, 21(1) :71–84.
- Braunschweig, U., Gueroussov, S., Plocik, A. M., Graveley, B. R., and Blencowe, B. J. (2013). Dynamic integration of splicing within gene regulatory pathways. *Cell*, 152(6) :1252–1269.
- Breitbart, R. E., Nguyen, H. T., Medford, R. M., Destree, A. T., Mahdavi, V., and Nadal-Ginard, B. (1985). Intricate combinatorial patterns of exon splicing generate multiple regulated troponin t isoforms from a single gene. *Cell*, 41(1) :67–82.
- Brow, D. A. (2002). Allosteric cascade of spliceosome activation. *Annu Rev Genet*, 36 :333–360.
- Buelt, M. K., Glidden, B. J., and Storm, D. R. (1994). Regulation of p68 rna helicase by calmodulin and protein kinase c. *J Biol Chem*, 269(47) :29367–29370.
- Buj-Bello, A., Furling, D., Tronchère, H., Laporte, J., Lerouge, T., Butler-Browne, G. S., and Mandel, J.-L. (2002). Muscle-specific alternative splicing of myotubularin-related 1 gene is impaired in dm1 muscle cells. *Hum Mol Genet*, 11(19) :2297–2307.

- Camats, M., Guil, S., Kokolo, M., and Bach-Elias, M. (2008). P68 rna helicase (ddx5) alters activity of cis- and trans-acting factors of the alternative splicing of h-ras. *PLoS One*, 3(8) :e2926.
- Cao, Y., Kumar, R. M., Penn, B. H., Berkes, C. A., Kooperberg, C., Boyer, L. A., Young, R. A., and Tapscott, S. J. (2006). Global and gene-specific analyses show distinct roles for myod and myog at a common set of promoters. *EMBO J*, 25(3) :502–511.
- Cao, Y., Yao, Z., Sarkar, D., Lawrence, M., Sanchez, G. J., Parker, M. H., MacQuarrie, K. L., Davison, J., Morgan, M. T., Ruzzo, W. L., Gentleman, R. C., and Tapscott, S. J. (2010). Genome-wide myod binding in skeletal muscle cells : a potential for broad cellular reprogramming. *Dev Cell*, 18(4) :662–674.
- Caretti, G., Lei, E. P., and Sartorelli, V. (2007). The dead-box p68/p72 proteins and the noncoding rna steroid receptor activator sra : eclectic regulators of disparate biological functions. *Cell Cycle*, 6(10) :1172–1176.
- Caretti, G., Schiltz, R. L., Dilworth, F. J., Padova, M. D., Zhao, P., Ogryzko, V., Fuller-Pace, F. V., Hoffman, E. P., Tapscott, S. J., and Sartorelli, V. (2006). The rna helicases p68/p72 and the noncoding rna sra are coregulators of myod and skeletal muscle differentiation. *Dev Cell*, 11(4) :547–560.
- Carosella, E. D., Moreau, P., Le Maout, J., Le Discorde, M., Dausset, J., and Rouas-Freiss, N. (2003). Hla-g molecules : from maternal-fetal tolerance to tissue acceptance. *Adv Immunol*, 81 :199–252.
- Carpentier, C., Ghanem, D., Fernandez-Gomez, F.-J., Jumeau, F., Philippe, J.-V., Freyermuth, F., Labudeck, A., Eddarkaoui, S., Dhaenens, C.-M., Holt, I., Behm-Ansmant, I., Marmier-Gourrier, N., Branlant, C., Charlet-Berguerand, N., Marie, J., Schraen-Maschke, S., Buée, L., Sergeant, N., and Caillet-Boudin, M.-L. (2014). Tau exon 2 responsive elements deregulated in myotonic dystrophy type i are proximal to exon 2 and synergistically regulated by mbnl1 and mbnl2. *Biochim Biophys Acta*.

- Casanova, J. E. (2007). Partitioning numb. *EMBO Rep*, 8(3) :233–235.
- Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B. M., Strein, C., Davey, N. E., Humphreys, D. T., Preiss, T., Steinmetz, L. M., Krijgsveld, J., and Hentze, M. W. (2012). Insights into rna biology from an atlas of mammalian mrna-binding proteins. *Cell*, 149(6) :1393–1406.
- Castle, J. C., Zhang, C., Shah, J. K., Kulkarni, A. V., Kalsotra, A., Cooper, T. A., and Johnson, J. M. (2008). Expression of 24,426 human alternative splicing events and predicted cis regulation in 48 tissues and cell lines. *Nat Genet*, 40(12) :1416–1425.
- Causevic, M., Hislop, R. G., Kernohan, N. M., Carey, F. A., Kay, R. A., Steele, R. J., and Fuller-Pace, F. V. (2001). Overexpression and poly-ubiquitylation of the dead-box rna helicase p68 in colorectal tumours. *Oncogene*, 20(53) :7734–7743.
- Chabot, B. (1996). Directing alternative splicing : cast and scenarios. *Trends Genet*, 12(11) :472–478.
- Charlet-B, N., Savkur, R. S., Singh, G., Philips, A. V., Grice, E. A., and Cooper, T. A. (2002). Loss of the muscle-specific chloride channel in type 1 myotonic dystrophy due to misregulated alternative splicing. *Mol Cell*, 10(1) :45–53.
- Chen, C. D., Kobayashi, R., and Helfman, D. M. (1999). Binding of hnrnp h to an exonic splicing silencer is involved in the regulation of alternative splicing of the rat beta-tropomyosin gene. *Genes Dev*, 13(5) :593–606.
- Chen, J.-F., Mandel, E. M., Thomson, J. M., Wu, Q., Callis, T. E., Hammond, S. M., Conlon, F. L., and Wang, D.-Z. (2006). The role of microrna-1 and microrna-133 in skeletal muscle proliferation and differentiation. *Nat Genet*, 38(2) :228–233.
- Chen, J.-F., Tao, Y., Li, J., Deng, Z., Yan, Z., Xiao, X., and Wang, D.-Z. (2010). microrna-1 and microrna-206 regulate skeletal muscle satellite cell proliferation and differentiation by repressing pax7. *J Cell Biol*, 190(5) :867–879.



- Chow, L. T., Gelin, R. E., Broker, T. R., and Roberts, R. J. (1977). An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger rna. *Cell*, 12(1) :1–8.
- Clark, E. L., Fuller-Pace, F. V., Elliott, D. J., and Robson, C. N. (2008). Coupling transcription to rna processing via the p68 dead box rna helicase androgen receptor co-activator in prostate cancer. *Biochem Soc Trans*, 36(Pt 3) :546–547.
- Cline, T. W. (1984). Autoregulatory functioning of a drosophila gene product that establishes and maintains the sexually determined state. *Genetics*, 107(2) :231–277.
- Cordin, O., Hahn, D., and Beggs, J. D. (2012). Structure, function and regulation of spliceosomal rna helicases. *Curr Opin Cell Biol*, 24(3) :431–438.
- Cornelison, D. D. W. (1998). *Gene expression in wild-type and MyoD-null satellite cells : regulation of activation, proliferation, and myogenesis*. PhD thesis, California Institute of Technology.
- Cornelison, D. D. W. (2008). Context matters : in vivo and in vitro influences on muscle satellite cell activity. *J Cell Biochem*, 105(3) :663–669.
- Coutinho-Mansfield, G. C., Xue, Y., Zhang, Y., and Fu, X.-D. (2007). Ptb/nptb switch : a post-transcriptional mechanism for programming neuronal differentiation. *Genes Dev*, 21(13) :1573–1577.
- Crawford, J. B. and Patton, J. G. (2006). Activation of alpha-tropomyosin exon 2 is regulated by the sr protein 9g8 and heterogeneous nuclear ribonucleoproteins h and f. *Mol Cell Biol*, 26(23) :8791–8802.
- Cremisi, F. (2013). Micrnas and cell fate in cortical and retinal development. *Front Cell Neurosci*, 7 :141.
- Cusack, B. P., Arndt, P. F., Duret, L., and Roest Crolius, H. (2011). Preventing dangerous nonsense : selection for robustness to transcriptional error in human genes. *PLoS Genet*, 7(10) :e1002276.

- Damianov, A. and Black, D. L. (2010). Autoregulation of fox protein expression to produce dominant negative splicing factors. *RNA*, 16(2) :405–416.
- Dardenne, E., Pierredon, S., Driouch, K., Gratadou, L., Lacroix-Triki, M., Espinoza, M. P., Zonta, E., Germann, S., Mortada, H., Villemin, J.-P., Dutertre, M., Lidereau, R., Vagner, S., and Auboeuf, D. (2012). Splicing switch of an epigenetic regulator by rna helicases promotes tumor-cell invasiveness. *Nat Struct Mol Biol*, 19(11) :1139–1146.
- Darnell, R. B. (2010). Hits-clip : panoramic views of protein-rna regulation in living cells. *Wiley Interdiscip Rev RNA*, 1(2) :266–286.
- Das, R., Yu, J., Zhang, Z., Gygi, M. P., Krainer, A. R., Gygi, S. P., and Reed, R. (2007). Srp proteins function in coupling rna polymerase ii transcription to pre-mrna splicing. *Mol Cell*, 26(6) :867–881.
- Dauvilliers, Y. A. and Laperche, L. (2012). Myotonic dystrophy type 1, daytime sleepiness and rem sleep dysregulation. *Sleep Med Rev*, 16(6) :539–545.
- Davis, B. M., McCurrach, M. E., Taneja, K. L., Singer, R. H., and Housman, D. E. (1997). Expansion of a cug trinucleotide repeat in the 3' untranslated region of myotonic dystrophy protein kinase transcripts results in nuclear retention of transcripts. *Proc Natl Acad Sci U S A*, 94(14) :7388–7393.
- Davis, B. N., Hilyard, A. C., Nguyen, P. H., Lagna, G., and Hata, A. (2010). Smad proteins bind a conserved rna sequence to promote microRNA maturation by drosha. *Mol Cell*, 39(3) :373–384.
- de Haro, M., Al-Ramahi, I., De Gouyon, B., Ukani, L., Rosa, A., Faustino, N. A., Ashizawa, T., Cooper, T. A., and Botas, J. (2006). Mbnl1 and cugbp1 modify expanded cug-induced toxicity in a drosophila model of myotonic dystrophy type 1. *Hum Mol Genet*, 15(13) :2138–2145.
- de la Mata, M. and Kornblihtt, A. R. (2006). Rna polymerase ii c-terminal domain mediates regulation of alternative splicing by srp20. *Nat Struct Mol Biol*, 13(11) :973–980.

- Dembowski, J. A. and Grabowski, P. J. (2009). The cugbp2 splicing factor regulates an ensemble of branchpoints from perimeter binding sites with implications for autoregulation. *PLoS Genet*, 5(8) :e1000595.
- Dhaenens, C. M., Schraen-Maschke, S., Tran, H., Vingtdeux, V., Ghanem, D., Leroy, O., Delplanque, J., Vanbrussel, E., Delacourte, A., Vermersch, P., Maurage, C. A., Gruffat, H., Sergeant, A., Mahadevan, M. S., Ishiura, S., Buée, L., Cooper, T. A., Caillet-Boudin, M. L., Charlet-Berguerand, N., Sablonnière, B., and Sergeant, N. (2008). Overexpression of mbnl1 fetal isoforms and modified splicing of tau in the dm1 brain : two individual consequences of cug trinucleotide repeats. *Exp Neurol*, 210(2) :467–478.
- Dmitriev, P., Barat, A., Polesskaya, A., O’Connell, M. J., Robert, T., Dessen, P., Walsh, T. A., Lazar, V., Turki, A., Carnac, G., Laoudj-Chenivresse, D., Lipinski, M., and Vassetzky, Y. S. (2013). Simultaneous mirna and mrna transcriptome profiling of human myoblasts reveals a novel set of myogenic differentiation-associated mirnas and their target genes. *BMC Genomics*, 14 :265.
- Donahue, C. P., Muratore, C., Wu, J. Y., Kosik, K. S., and Wolfe, M. S. (2006). Stabilization of the tau exon 10 stem loop alters pre-mrna splicing. *J Biol Chem*, 281(33) :23302–23306.
- Du, H., Cline, M. S., Osborne, R. J., Tuttle, D. L., Clark, T. A., Donohue, J. P., Hall, M. P., Shiue, L., Swanson, M. S., Thornton, C. A., and Ares, Jr, M. (2010). Aberrant alternative splicing and extracellular matrix gene expression in mouse models of myotonic dystrophy. *Nat Struct Mol Biol*, 17(2) :187–193.
- Dutertre, M., Sanchez, G., De Cian, M.-C., Barbier, J., Dardenne, E., Gratadou, L., Dujardin, G., Le Jossic-Corcus, C., Corcos, L., and Auboeuf, D. (2010). Cotranscriptional exon skipping in the genotoxic stress response. *Nat Struct Mol Biol*, 17(11) :1358–1366.
- Ebralidze, A., Wang, Y., Petkova, V., Ebralidse, K., and Junghans, R. P. (2004). Rna leaching of transcription factors disrupts transcription in myotonic dystrophy. *Science*, 303(5656) :383–387.

- Ekström, A.-B., Hakenäs-Plate, L., Samuelsson, L., Tulinius, M., and Wentz, E. (2008). Autism spectrum conditions in myotonic dystrophy type 1 : a study on 57 individuals with congenital and childhood forms. *Am J Med Genet B Neuropsychiatr Genet*, 147B(6) :918–926.
- Estes, P. A., Cooke, N. E., and Liebhaber, S. A. (1992). A native rna secondary structure controls alternative splice-site selection and generates two human growth hormone isoforms. *J Biol Chem*, 267(21) :14902–14908.
- Fairman-Williams, M. E., Guenther, U.-P., and Jankowsky, E. (2010). Sfl and sf2 helicases : family matters. *Curr Opin Struct Biol*, 20(3) :313–324.
- Fardaei, M., Rogers, M. T., Thorpe, H. M., Larkin, K., Hamshere, M. G., Harper, P. S., and Brook, J. D. (2002). Three proteins, mbnl, mbll and mbxl, co-localize in vivo with nuclear foci of expanded-repeat transcripts in dm1 and dm2 cells. *Hum Mol Genet*, 11(7) :805–814.
- Feng, L., Xie, Y., Zhang, H., and Wu, Y. (2012). mir-107 targets cyclin-dependent kinase 6 expression, induces cell cycle g1 arrest and inhibits invasion in gastric cancer cells. *Med Oncol*, 29(2) :856–863.
- Feng, Y., Cao, J.-H., Li, X.-Y., and Zhao, S.-H. (2011). Inhibition of mir-214 expression represses proliferation and differentiation of c2c12 myoblasts. *Cell Biochem Funct*, 29(5) :378–383.
- Figeac, N., Daczewska, M., Marcelle, C., and Jagla, K. (2007). Muscle stem cells and model systems for their investigation. *Dev Dyn*, 236(12) :3332–3342.
- Fugier, C., Klein, A. F., Hammer, C., Vassilopoulos, S., Ivarsson, Y., Toussaint, A., Tosch, V., Vignaud, A., Ferry, A., Messaddeq, N., Kokunai, Y., Tsuburaya, R., de la Grange, P., Dembele, D., Francois, V., Precigout, G., Boulade-Ladame, C., Hummel, M.-C., Lopez de Munain, A., Sergeant, N., Laquerrière, A., Thibault, C., Deryckere, F., Auboeuf, D., Garcia, L., Zimmermann, P., Udd, B., Schoser, B., Takahashi, M. P., Nishino, I., Bassez, G., Laporte, J., Furling, D., and Charlet-Berguerand, N. (2011). Misregulated alternative splicing of bin1 is associated

- with t tubule alterations and muscle weakness in myotonic dystrophy. *Nat Med*, 17(6) :720–725.
- Fuller-Pace, F. V. (2013). The dead box proteins ddx5 (p68) and ddx17 (p72) : multi-tasking transcriptional regulators. *Biochim Biophys Acta*, 1829(8) :756–763.
- Fuller-Pace, F. V. and Moore, H. C. (2011). Rna helicases p68 and p72 : multifunctional proteins with important implications for cancer development. *Future Oncol*, 7(2) :239–251.
- Furling, D., Lemieux, D., Taneja, K., and Puymirat, J. (2001). Decreased levels of myotonic dystrophy protein kinase (dmpk) and delayed differentiation in human myotonic dystrophy myoblasts. *Neuromuscul Disord*, 11(8) :728–735.
- Gabut, M., Chaudhry, S., and Blencowe, B. J. (2008). Snapshot : The splicing regulatory machinery. *Cell*, 133(1) :192.e1.
- Gabut, M., Samavarchi-Tehrani, P., Wang, X., Slobodeniuc, V., O’Hanlon, D., Sung, H.-K., Alvarez, M., Talukder, S., Pan, Q., Mazzoni, E. O., Nedelec, S., Wichterle, H., Woltjen, K., Hughes, T. R., Zandstra, P. W., Nagy, A., Wrana, J. L., and Blencowe, B. J. (2011). An alternative splicing switch regulates embryonic stem cell pluripotency and reprogramming. *Cell*, 147(1) :132–146.
- Gao, Z. and Cooper, T. A. (2013a). Antisense oligonucleotides : rising stars in eliminating rna toxicity in myotonic dystrophy. *Hum Gene Ther*, 24(5) :499–507.
- Gao, Z. and Cooper, T. A. (2013b). Reexpression of pyruvate kinase m2 in type 1 myofibers correlates with altered glucose metabolism in myotonic dystrophy. *Proc Natl Acad Sci U S A*, 110(33) :13570–13575.
- Gates, D. P., Coonrod, L. A., and Berglund, J. A. (2011). Autoregulated splicing of muscleblind-like 1 (mbnl1) pre-mrna. *J Biol Chem*, 286(39) :34224–34233.
- Gaul, C., Schmidt, T., Windisch, G., Wieser, T., Müller, T., Vielhaber, S., Zierz, S., and Leplow, B. (2006). Subtle cognitive dysfunction in adult onset myotonic dystrophy type 1 (dm1) and type 2 (dm2). *Neurology*, 67(2) :350–352.

- Geißler, V., Altmeyer, S., Stein, B., Uhlmann-Schiffler, H., and Stahl, H. (2013). The rna helicase ddx5/p68 binds to hupf3 and enhances nmd of ddx17/p72 and smg5 mrna. *Nucleic Acids Res*, 41(16) :7875–7888.
- Germann, S., Gratadou, L., Zonta, E., Dardenne, E., Gaudineau, B., Fougère, M., Samaan, S., Dutertre, M., Jauliac, S., and Auboeuf, D. (2012). Dual role of the ddx5/ddx17 rna helicases in the control of the pro-migratory nfat5 transcription factor. *Oncogene*, 31(42) :4536–4549.
- Goers, E. S., Purcell, J., Voelker, R. B., Gates, D. P., and Berglund, J. A. (2010). Mbnl1 binds gc motifs embedded in pyrimidines to regulate alternative splicing. *Nucleic Acids Res*, 38(7) :2467–2484.
- Gomes, A. V., Venkatraman, G., Davis, J. P., Tikunova, S. B., Engel, P., Solaro, R. J., and Potter, J. D. (2004). Cardiac troponin t isoforms affect the ca(2+) sensitivity of force development in the presence of slow skeletal troponin i : insights into the role of troponin t isoforms in the fetal heart. *J Biol Chem*, 279(48) :49579–49587.
- Gomes-Pereira, M. and Monckton, D. G. (2006). Chemical modifiers of unstable expanded simple sequence repeats : what goes up, could come down. *Mutat Res*, 598(1-2) :15–34.
- Gonçalves, M. A. F. V., Janssen, J. M., Nguyen, Q. G., Athanasopoulos, T., Hauschka, S. D., Dickson, G., and de Vries, A. A. F. (2011). Transcription factor rational design improves directed differentiation of human mesenchymal stem cells into skeletal myocytes. *Mol Ther*, 19(7) :1331–1341.
- Goo, Y.-H. and Cooper, T. A. (2009). Cugbp2 directly interacts with u2 17s snrnp components and promotes u2 snrna binding to cardiac troponin t pre-mrna. *Nucleic Acids Res*, 37(13) :4275–4286.
- Gooding, C. and Smith, C. W. J. (2008). Tropomyosin exons as models for alternative splicing. *Adv Exp Med Biol*, 644 :27–42.

- Grabowski, P. (2011). Alternative splicing takes shape during neuronal development. *Curr Opin Genet Dev*, 21(4) :388–394.
- Graveley, B. R. (2000). Sorting out the complexity of sr protein functions. *RNA*, 6(9) :1197–1211.
- Graveley, B. R., Hertel, K. J., and Maniatis, T. (1998). A systematic analysis of the factors that determine the strength of pre-mrna splicing enhancers. *EMBO J*, 17(22) :6747–6756.
- Gregory, R. I., Yan, K.-P., Amuthan, G., Chendrimada, T., Doratotaj, B., Cooch, N., and Shiekhattar, R. (2004). The microprocessor complex mediates the genesis of micrnas. *Nature*, 432(7014) :235–240.
- Hallegger, M., Llorian, M., and Smith, C. W. J. (2010). Alternative splicing : global insights. *FEBS J*, 277(4) :856–866.
- Han, H., Irimia, M., Ross, P. J., Sung, H.-K., Alipanahi, B., David, L., Golipour, A., Gabut, M., Michael, I. P., Nachman, E. N., Wang, E., Trcka, D., Thompson, T., O’Hanlon, D., Slobodeniuc, V., Barbosa-Morais, N. L., Burge, C. B., Moffat, J., Frey, B. J., Nagy, A., Ellis, J., Wrana, J. L., and Blencowe, B. J. (2013). Mbnl proteins repress es-cell-specific alternative splicing and reprogramming. *Nature*, 498(7453) :241–245.
- Harper, P. S. and Harper, P. (2001). *Myotonic Dystrophy*. Bailliere Tindall.
- Haugen, P., Simon, D. M., and Bhattacharya, D. (2005). The natural history of group i introns. *Trends Genet*, 21(2) :111–119.
- Hegele, A., Kamburov, A., Grossmann, A., Sourlis, C., Wowro, S., Weimann, M., Will, C. L., Pena, V., Lührmann, R., and Stelzl, U. (2012). Dynamic protein-protein interaction wiring of the human spliceosome. *Mol Cell*, 45(4) :567–580.
- Henn, A., Bradley, M. J., and De La Cruz, E. M. (2012). Atp utilization and rna conformational rearrangement by dead-box proteins. *Annu Rev Biophys*, 41 :247–267.



- Hernández-Hernández, O., Guiraud-Dogan, C., Sicot, G., Huguet, A., Luilier, S., Steidl, E., Saenger, S., Marciniak, E., Obriot, H., Chevarin, C., Nicole, A., Revillod, L., Charizanis, K., Lee, K.-Y., Suzuki, Y., Kimura, T., Matsuura, T., Cisneros, B., Swanson, M. S., Trovero, F., Buisson, B., Bizot, J.-C., Hamon, M., Humez, S., Bassez, G., Metzger, F., Buée, L., Munnich, A., Sergeant, N., Gourdon, G., and Gomes-Pereira, M. (2013). Myotonic dystrophy ctg expansion affects synaptic vesicle proteins, neurotransmission and mouse behaviour. *Brain*, 136(Pt 3) :957–970.
- Hirai, H., Verma, M., Watanabe, S., Tastad, C., Asakura, Y., and Asakura, A. (2010). Myod regulates apoptosis of myoblasts through microRNA-mediated down-regulation of pax3. *J Cell Biol*, 191(2) :347–365.
- Ho, T. H., Bundman, D., Armstrong, D. L., and Cooper, T. A. (2005). Transgenic mice expressing cug-bp1 reproduce splicing mis-regulation observed in myotonic dystrophy. *Hum Mol Genet*, 14(11) :1539–1547.
- Hobert, O. and Westphal, H. (2000). Functions of lim-homeobox genes. *Trends Genet*, 16(2) :75–83.
- Hodges, C., Bintu, L., Lubkowska, L., Kashlev, M., and Bustamante, C. (2009). Nucleosomal fluctuations govern the transcription dynamics of rna polymerase ii. *Science*, 325(5940) :626–628.
- Hong, S., Noh, H., Chen, H., Padia, R., Pan, Z. K., Su, S.-B., Jing, Q., Ding, H.-F., and Huang, S. (2013). Signaling by p38 mapk stimulates nuclear localization of the microprocessor component p68 for processing of selected primary micrnas. *Sci Signal*, 6(266) :ra16.
- Hönig, A., Auboeuf, D., Parker, M. M., O'Malley, B. W., and Berget, S. M. (2002). Regulation of alternative splicing by the atp-dependent dead-box rna helicase p72. *Mol Cell Biol*, 22(16) :5698–5707.
- Hoskins, A. A., Friedman, L. J., Gallagher, S. S., Crawford, D. J., Anderson, E. G., Wombacher, R., Ramirez, N., Cornish, V. W., Gelles, J., and Moore, M. J. (2011).

- Ordered and dynamic assembly of single spliceosomes. *Science*, 331(6022) :1289–1295.
- Hoskins, A. A. and Moore, M. J. (2012). The spliceosome : a flexible, reversible macromolecular machine. *Trends Biochem Sci*, 37(5) :179–188.
- House, A. E. and Lynch, K. W. (2006). An exonic splicing silencer represses spliceosome assembly after atp-dependent exon recognition. *Nat Struct Mol Biol*, 13(10) :937–944.
- Huang, M.-B., Xu, H., Xie, S.-J., Zhou, H., and Qu, L.-H. (2011). Insulin-like growth factor-1 receptor is regulated by microrna-133 during skeletal myogenesis. *PLoS One*, 6(12) :e29173.
- Huang, Y., Li, W., Yao, X., Lin, Q.-J., Yin, J.-W., Liang, Y., Heiner, M., Tian, B., Hui, J., and Wang, G. (2012). Mediator complex regulates alternative mrna processing via the med23 subunit. *Mol Cell*, 45(4) :459–469.
- Huelga, S. C., Vu, A. Q., Arnold, J. D., Liang, T. Y., Liu, P. P., Yan, B. Y., Donohue, J. P., Shiue, L., Hoon, S., Brenner, S., Ares, Jr, M., and Yeo, G. W. (2012). Integrative genome-wide analysis reveals cooperative regulation of alternative splicing by hnrnp proteins. *Cell Rep*, 1(2) :167–178.
- Huichalaf, C., Sakai, K., Jin, B., Jones, K., Wang, G.-L., Schoser, B., Schneider-Gold, C., Sarkar, P., Pereira-Smith, O. M., Timchenko, N., and Timchenko, L. (2010). Expansion of cug rna repeats causes stress and inhibition of translation in myotonic dystrophy 1 (dm1) cells. *FASEB J*, 24(10) :3706–3719.
- Iaizzo, P. A., Franke, C., Hatt, H., Spittelmeister, W., Ricker, K., Rüdell, R., and Lehmann-Horn, F. (1991). Altered sodium channel behaviour causes myotonia in dominantly inherited myotonia congenita. *Neuromuscul Disord*, 1(1) :47–53.
- Ip, F. C., Chung, S. S., Fu, W. Y., and Ip, N. Y. (2000). Developmental and tissue-specific expression of dead box protein p72. *Neuroreport*, 11(3) :457–462.

- Ip, J. Y., Schmidt, D., Pan, Q., Ramani, A. K., Fraser, A. G., Odom, D. T., and Blencowe, B. J. (2011). Global impact of rna polymerase ii elongation inhibition on alternative splicing regulation. *Genome Res*, 21(3) :390–401.
- Irimia, M. and Blencowe, B. J. (2012). Alternative splicing : decoding an expansive regulatory layer. *Curr Opin Cell Biol*, 24(3) :323–332.
- Ishizuka, A., Siomi, M. C., and Siomi, H. (2002). A drosophila fragile x protein interacts with components of rnai and ribosomal proteins. *Genes Dev*, 16(19) :2497–2508.
- Izquierdo, J. M., Majós, N., Bonnal, S., Martínez, C., Castelo, R., Guigó, R., Bilbao, D., and Valcárcel, J. (2005). Regulation of fas alternative splicing by antagonistic effects of tia-1 and ptb on exon definition. *Mol Cell*, 19(4) :475–484.
- Jacobs, A.-M. F., Nicol, S. M., Hislop, R. G., Jaffray, E. G., Hay, R. T., and Fuller-Pace, F. V. (2007). Sumo modification of the dead box protein p68 modulates its transcriptional activity and promotes its interaction with hdac1. *Oncogene*, 26(40) :5866–5876.
- Jalal, C., Uhlmann-Schiffler, H., and Stahl, H. (2007). Redundant role of dead box proteins p68 (ddx5) and p72/p82 (ddx17) in ribosome biogenesis and cell proliferation. *Nucleic Acids Res*, 35(11) :3590–3601.
- Janknecht, R. (2010). Multi-talented dead-box proteins and potential tumor promoters : p68 rna helicase (ddx5) and its paralog, p72 rna helicase (ddx17). *Am J Transl Res*, 2(3) :223–234.
- Jankowsky, E. (2010). *RNA Helicases (RSC Biomolecular Sciences)*. Royal Society of Chemistry.
- Jankowsky, E. (2011). Rna helicases at work : binding and rearranging. *Trends Biochem Sci*, 36(1) :19–29.
- Jarmoskaite, I. and Russell, R. (2011). Dead-box proteins as rna helicases and chaperones. *Wiley Interdiscip Rev RNA*, 2(1) :135–152.

- Jelen, N., Ule, J., Zivin, M., and Darnell, R. B. (2007). Evolution of nova-dependent splicing regulation in the brain. *PLoS Genet*, 3(10) :1838–1847.
- Jensen, E. D., Niu, L., Caretti, G., Nicol, S. M., Teplyuk, N., Stein, G. S., Sartorelli, V., van Wijnen, A. J., Fuller-Pace, F. V., and Westendorf, J. J. (2008). p68 (ddx5) interacts with runx2 and regulates osteoblast differentiation. *J Cell Biochem*, 103(5) :1438–1451.
- Jiang, H., Mankodi, A., Swanson, M. S., Moxley, R. T., and Thornton, C. A. (2004). Myotonic dystrophy type 1 is associated with nuclear foci of mutant rna, sequestration of muscleblind proteins and deregulated alternative splicing in neurons. *Hum Mol Genet*, 13(24) :3079–3088.
- Jin, Y., Yang, Y., and Zhang, P. (2011). New insights into rna secondary structure in the alternative splicing of pre-mrnas. *RNA Biol*, 8(3) :450–457.
- Jones, S. (2004). An overview of the basic helix-loop-helix proteins. *Genome Biol*, 5(6) :226.
- Kahlina, K., Goren, I., Pfeilschifter, J., and Frank, S. (2004). p68 dead box rna helicase expression in keratinocytes. regulation, nucleolar localization, and functional connection to proliferation and vascular endothelial growth factor gene expression. *J Biol Chem*, 279(43) :44872–44882.
- Kaliman, P. and Llagostera, E. (2008). Myotonic dystrophy protein kinase (dmpk) and its role in the pathogenesis of myotonic dystrophy 1. *Cell Signal*, 20(11) :1935–1941.
- Kalkman, J. S., Schillings, M. L., van der Werf, S. P., Padberg, G. W., Zwarts, M. J., van Engelen, B. G. M., and Bleijenberg, G. (2005). Experienced fatigue in facioscapulohumeral dystrophy, myotonic dystrophy, and hmsn-i. *J Neurol Neurosurg Psychiatry*, 76(10) :1406–1409.
- Kalsotra, A., Singh, R. K., Gurha, P., Ward, A. J., Creighton, C. J., and Cooper, T. A. (2014). The mef2 transcription network is disrupted in myotonic dystrophy heart tissue, dramatically altering mirna and mrna expression. *Cell Rep*.

- Kalsotra, A., Wang, K., Li, P.-F., and Cooper, T. A. (2010). Micrnas coordinate an alternative splicing network during mouse postnatal heart development. *Genes Dev*, 24(7) :653–658.
- Kalsotra, A., Xiao, X., Ward, A. J., Castle, J. C., Johnson, J. M., Burge, C. B., and Cooper, T. A. (2008). A postnatal switch of celf and mbnl proteins reprograms alternative splicing in the developing heart. *Proc Natl Acad Sci U S A*, 105(51) :20333–20338.
- Kanadia, R. N., Johnstone, K. A., Mankodi, A., Lungu, C., Thornton, C. A., Esson, D., Timmers, A. M., Hauswirth, W. W., and Swanson, M. S. (2003). A muscleblind knockout model for myotonic dystrophy. *Science*, 302(5652) :1978–1980.
- Kanadia, R. N., Shin, J., Yuan, Y., Beattie, S. G., Wheeler, T. M., Thornton, C. A., and Swanson, M. S. (2006). Reversal of rna missplicing and myotonia after muscleblind overexpression in a mouse poly(cug) model for myotonic dystrophy. *Proc Natl Acad Sci U S A*, 103(31) :11748–11753.
- Kar, A., Fushimi, K., Zhou, X., Ray, P., Shi, C., Chen, X., Liu, Z., Chen, S., and Wu, J. Y. (2011). Rna helicase p68 (ddx5) regulates tau exon 10 splicing by modulating a stem-loop structure at the 5' splice site. *Mol Cell Biol*, 31(9) :1812–1821.
- Kassar-Duchossoy, L., Gayraud-Morel, B., Gomès, D., Rocancourt, D., Buckingham, M., Shinin, V., and Tajbakhsh, S. (2004). Mrf4 determines skeletal muscle identity in myf5 :myod double-mutant mice. *Nature*, 431(7007) :466–471.
- Kim, D.-H., Langlois, M.-A., Lee, K.-B., Riggs, A. D., Puymirat, J., and Rossi, J. J. (2005). Hnrnp h inhibits nuclear export of mrna containing expanded cug repeats and a distal branch point sequence. *Nucleic Acids Res*, 33(12) :3866–3874.
- Kim, E., Goren, A., and Ast, G. (2008). Alternative splicing : current perspectives. *Bioessays*, 30(1) :38–47.
- Kim, H. K., Lee, Y. S., Sivaprasad, U., Malhotra, A., and Dutta, A. (2006). Muscle-specific microrna mir-206 promotes muscle differentiation. *J Cell Biol*, 174(5) :677–687.

- Kimura, T., Nakamori, M., Lueck, J. D., Pouliquin, P., Aoike, F., Fujimura, H., Dirksen, R. T., Takahashi, M. P., Dulhunty, A. F., and Sakoda, S. (2005). Altered mrna splicing of the skeletal muscle ryanodine receptor and sarcoplasmic/endoplasmic reticulum ca<sup>2+</sup>-atpase in myotonic dystrophy type 1. *Hum Mol Genet*, 14(15) :2189–2200.
- Kino, Y., Mori, D., Oma, Y., Takeshita, Y., Sasagawa, N., and Ishiura, S. (2004). Muscleblind protein, mbnl1/exp, binds specifically to chg repeats. *Hum Mol Genet*, 13(5) :495–507.
- Kino, Y., Washizu, C., Oma, Y., Onishi, H., Nezu, Y., Sasagawa, N., Nukina, N., and Ishiura, S. (2009). Mbnl and celf proteins regulate alternative splicing of the skeletal muscle chloride channel clcn1. *Nucleic Acids Res*, 37(19) :6477–6490.
- Koebis, M., Ohsawa, N., Kino, Y., Sasagawa, N., Nishino, I., and Ishiura, S. (2011). Alternative splicing of myomesin 1 gene is aberrantly regulated in myotonic dystrophy type 1. *Genes Cells*, 16(9) :961–972.
- Kornblihtt, A. R., Schor, I. E., Alló, M., Dujardin, G., Petrillo, E., and Muñoz, M. J. (2013). Alternative splicing : a pivotal step between eukaryotic transcription and translation. *Nat Rev Mol Cell Biol*, 14(3) :153–165.
- Koshelev, M., Sarma, S., Price, R. E., Wehrens, X. H. T., and Cooper, T. A. (2010). Heart-specific overexpression of cugbp1 reproduces functional and molecular abnormalities of myotonic dystrophy type 1. *Hum Mol Genet*, 19(6) :1066–1075.
- Kreahling, J. M. and Graveley, B. R. (2005). The istem, a long-range rna secondary structure element required for efficient exon inclusion in the drosophila dscam pre-mrna. *Mol Cell Biol*, 25(23) :10251–10260.
- Krol, J., Fiszler, A., Mykowska, A., Sobczak, K., de Mezer, M., and Krzyzosiak, W. J. (2007). Ribonuclease dicer cleaves triplet repeat hairpins into shorter repeats that silence specific targets. *Mol Cell*, 25(4) :575–586.
- Kuroyanagi, H. (2009). Fox-1 family of rna-binding proteins. *Cell Mol Life Sci*, 66(24) :3895–3907.

- Laberge, L., Bégin, P., Dauvilliers, Y., Beaudry, M., Laforte, M., Jean, S., and Mathieu, J. (2009). A polysomnographic study of daytime sleepiness in myotonic dystrophy type 1. *J Neurol Neurosurg Psychiatry*, 80(6) :642–646.
- Ladd, A. N., Charlet, N., and Cooper, T. A. (2001). The celf family of rna binding proteins is implicated in cell-specific and developmentally regulated alternative splicing. *Mol Cell Biol*, 21(4) :1285–1296.
- Lamm, G. M., Nicol, S. M., Fuller-Pace, F. V., and Lamond, A. I. (1996). p72 : a human nuclear dead box protein highly related to p68. *Nucleic Acids Res*, 24(19) :3739–3747.
- Lane, D. P. and Hoeffler, W. K. (1980). Sv40 large t shares an antigenic determinant with a cellular protein of molecular weight 68,000. *Nature*, 288(5787) :167–170.
- Lasko, P. (2013). The dead-box helicase vasa : evidence for a multiplicity of functions in rna processes and developmental biology. *Biochim Biophys Acta*, 1829(8) :810–816.
- Laurent, F.-X. (2011). *Une nouvelle fonction pour la DEAD-box ARN hélicase p68/DDX5 dans la Dystrophie Myotonique de type 1*. PhD thesis, Université de Paris-Sud XI - Orsay.
- Laurent, F.-X., Sureau, A., Klein, A. F., Trouslard, F., Gasnier, E., Furling, D., and Marie, J. (2012). New function for the rna helicase p68/ddx5 as a modifier of mbnl1 activity on expanded cug repeats. *Nucleic Acids Res*, 40(7) :3159–3171.
- Le Grand, F. and Rudnicki, M. A. (2007). Skeletal muscle satellite cells and adult myogenesis. *Curr Opin Cell Biol*, 19(6) :628–633.
- Lee, C.-G. (2002). Rh70, a bidirectional rna helicase, co-purifies with u1snrnp. *J Biol Chem*, 277(42) :39679–39683.
- Lee, J.-A., Tang, Z.-Z., and Black, D. L. (2009). An inducible change in fox-1/a2bp1 splicing modulates the alternative splicing of downstream neuronal target exons. *Genes Dev*, 23(19) :2284–2293.



- Lee, J. E. and Cooper, T. A. (2009). Pathogenic mechanisms of myotonic dystrophy. *Biochem Soc Trans*, 37(Pt 6) :1281–1286.
- Lemerrier, C., To, R. Q., Carrasco, R. A., and Konieczny, S. F. (1998). The basic helix-loop-helix transcription factor mist1 functions as a transcriptional repressor of myod. *EMBO J*, 17(5) :1412–1422.
- Leroy, O., Dhaenens, C.-M., Schraen-Maschke, S., Belarbi, K., Delacourte, A., Andreadis, A., Sablonnière, B., Buée, L., Sergeant, N., and Caillet-Boudin, M.-L. (2006). Etr-3 represses tau exons 2/3 inclusion, a splicing event abnormally enhanced in myotonic dystrophy type i. *J Neurosci Res*, 84(4) :852–859.
- Lewis, B. P., Shih, I.-h., Jones-Rhoades, M. W., Bartel, D. P., and Burge, C. B. (2003). Prediction of mammalian microRNA targets. *Cell*, 115(7) :787–798.
- Lin, C., Yang, L., Yang, J. J., Huang, Y., and Liu, Z.-R. (2005). Atpase/helicase activities of p68 rna helicase are required for pre-mrna splicing but not for assembly of the spliceosome. *Mol Cell Biol*, 25(17) :7484–7493.
- Lin, S., Coutinho-Mansfield, G., Wang, D., Pandit, S., and Fu, X.-D. (2008). The splicing factor sc35 has an active role in transcriptional elongation. *Nat Struct Mol Biol*, 15(8) :819–826.
- Lin, S., Tian, L., Shen, H., Gu, Y., Li, J.-L., Chen, Z., Sun, X., You, M. J., and Wu, L. (2013). Ddx5 is a positive regulator of oncogenic notch1 signaling in t cell acute lymphoblastic leukemia. *Oncogene*, 32(40) :4845–4853.
- Lin, X., Miller, J. W., Mankodi, A., Kanadia, R. N., Yuan, Y., Moxley, R. T., Swanson, M. S., and Thornton, C. A. (2006). Failure of mbnl1-dependent post-natal splicing transitions in myotonic dystrophy. *Hum Mol Genet*, 15(13) :2087–2097.
- Linder, P. and Jankowsky, E. (2011). From unwinding to clamping - the dead box rna helicase family. *Nat Rev Mol Cell Biol*, 12(8) :505–516.

- Listerman, I., Sapra, A. K., and Neugebauer, K. M. (2006). Cotranscriptional coupling of splicing factor recruitment and precursor messenger rna splicing in mammalian cells. *Nat Struct Mol Biol*, 13(9) :815–822.
- Liu, F., Putnam, A., and Jankowsky, E. (2008). Atp hydrolysis is required for dead-box protein recycling but not for duplex unwinding. *Proc Natl Acad Sci U S A*, 105(51) :20209–20214.
- Liu, H. X., Cartegni, L., Zhang, M. Q., and Krainer, A. R. (2001). A mechanism for exon skipping caused by nonsense or missense mutations in *brca1* and other genes. *Nat Genet*, 27(1) :55–58.
- Liu, J., Wu, C.-P., Lu, B.-F., and Jiang, J.-T. (2013). Mechanism of t cell regulation by micrnas. *Cancer Biol Med*, 10(3) :131–137.
- Liu, W., Zhou, Y., Hu, Z., Sun, T., Denise, A., Fu, X.-D., and Zhang, Y. (2010). Regulation of splicing enhancer activities by rna secondary structures. *FEBS Lett*, 584(21) :4401–4407.
- Liu, Z.-R. (2002). p68 rna helicase is an essential human splicing factor that acts at the u1 snrna-5' splice site duplex. *Mol Cell Biol*, 22(15) :5443–5450.
- Llorian, M. and Smith, C. W. J. (2011). Decoding muscle alternative splicing. *Curr Opin Genet Dev*, 21(4) :380–387.
- López Castel, A., Cleary, J. D., and Pearson, C. E. (2010). Repeat instability as the basis for human diseases and as a potential target for therapy. *Nat Rev Mol Cell Biol*, 11(3) :165–170.
- Lorsch, J. R. and Herschlag, D. (1998). The dead box protein eif4a. 1. a minimal kinetic and thermodynamic framework reveals coupled binding of rna and nucleotide. *Biochemistry*, 37(8) :2180–2193.
- Lu, J., Webb, R., Richardson, J. A., and Olson, E. N. (1999). Myor : a muscle-restricted basic helix-loop-helix transcription factor that antagonizes the actions of myod. *Proc Natl Acad Sci U S A*, 96(2) :552–557.

- Luco, R. F., Allo, M., Schor, I. E., Kornblihtt, A. R., and Misteli, T. (2011). Epigenetics in alternative pre-mrna splicing. *Cell*, 144(1) :16–26.
- Luco, R. F., Pan, Q., Tominaga, K., Blencowe, B. J., Pereira-Smith, O. M., and Misteli, T. (2010). Regulation of alternative splicing by histone modifications. *Science*, 327(5968) :996–1000.
- Machuca-Tzili, L., Brook, D., and Hilton-Jones, D. (2005). Clinical and molecular aspects of the myotonic dystrophies : a review. *Muscle Nerve*, 32(1) :1–18.
- Machuca-Tzili, L. E., Buxton, S., Thorpe, A., Timson, C. M., Wigmore, P., Luther, P. K., and Brook, J. D. (2011). Zebrafish deficient for muscleblind-like 2 exhibit features of myotonic dystrophy. *Dis Model Mech*, 4(3) :381–392.
- Mahadevan, M. S., Yadava, R. S., Yu, Q., Balijepalli, S., Frenzel-McCardell, C. D., Bourne, T. D., and Phillips, L. H. (2006). Reversible model of rna toxicity and cardiac conduction defects in myotonic dystrophy. *Nat Genet*, 38(9) :1066–1070.
- Mankodi, A., Takahashi, M. P., Jiang, H., Beck, C. L., Bowers, W. J., Moxley, R. T., Cannon, S. C., and Thornton, C. A. (2002). Expanded cug repeats trigger aberrant splicing of clc-1 chloride channel pre-mrna and hyperexcitability of skeletal muscle in myotonic dystrophy. *Mol Cell*, 10(1) :35–44.
- Manley, J. L. and Krainer, A. R. (2010). A rational nomenclature for serine/arginine-rich protein splicing factors (sr proteins). *Genes Dev*, 24(11) :1073–1074.
- Martorell, L., Monckton, D. G., Gamez, J., Johnson, K. J., Gich, I., Lopez de Munain, A., and Baiget, M. (1998). Progression of somatic ctg repeat length heterogeneity in the blood cells of myotonic dystrophy patients. *Hum Mol Genet*, 7(2) :307–312.
- Martorell, L., Monckton, D. G., Sanchez, A., Lopez De Munain, A., and Baiget, M. (2001). Frequency and stability of the myotonic dystrophy type 1 premutation. *Neurology*, 56(3) :328–335.
- Massari, M. E. and Murre, C. (2000). Helix-loop-helix proteins : regulators of transcription in eucaryotic organisms. *Mol Cell Biol*, 20(2) :429–440.

- Masuda, A., Andersen, H. S., Doktor, T. K., Okamoto, T., Ito, M., Andresen, B. S., and Ohno, K. (2012). Cugbp1 and mbnl1 preferentially bind to 3' utrs and facilitate mrna decay. *Sci Rep*, 2 :209.
- Matlin, A. J., Clark, F., and Smith, C. W. J. (2005). Understanding alternative splicing : towards a cellular code. *Nat Rev Mol Cell Biol*, 6(5) :386–398.
- Matynia, A., Ng, C. H., Dansithong, W., Chiang, A., Silva, A. J., and Reddy, S. (2010). Muscleblind1, but not dmpk or six5, contributes to a complex phenotype of muscular and motivational deficits in mouse models of myotonic dystrophy. *PLoS One*, 5(3) :e9857.
- Mazurek, A., Luo, W., Krasnitz, A., Hicks, J., Powers, R. S., and Stillman, B. (2012). Ddx5 regulates dna replication and is required for cell proliferation in a subset of breast cancer cells. *Cancer Discov*, 2(9) :812–825.
- McKinsey, T. A., Zhang, C. L., and Olson, E. N. (2001). Control of muscle development by dueling hats and hdacs. *Curr Opin Genet Dev*, 11(5) :497–504.
- Mercer, T. R. and Mattick, J. S. (2013). Understanding the regulatory and transcriptional complexity of the genome through structure. *Genome Res*, 23(7) :1081–1088.
- Michalowski, S., Miller, J. W., Urbinati, C. R., Paliouras, M., Swanson, M. S., and Griffith, J. (1999). Visualization of double-stranded rnas from the myotonic dystrophy protein kinase gene and interactions with cug-binding protein. *Nucleic Acids Res*, 27(17) :3534–3542.
- Miki, T., Bottaro, D. P., Fleming, T. P., Smith, C. L., Burgess, W. H., Chan, A. M., and Aaronson, S. A. (1992). Determination of ligand-binding specificity by alternative splicing : two distinct growth factor receptors encoded by a single gene. *Proc Natl Acad Sci U S A*, 89(1) :246–250.
- Miller, J. W., Urbinati, C. R., Teng-Umnuay, P., Stenberg, M. G., Byrne, B. J., Thornton, C. A., and Swanson, M. S. (2000). Recruitment of human muscleblind

- proteins to (cug)(n) expansions associated with myotonic dystrophy. *EMBO J*, 19(17) :4439–4448.
- Miriami, E., Margalit, H., and Sperling, R. (2003). Conserved sequence elements associated with exon skipping. *Nucleic Acids Res*, 31(7) :1974–1983.
- Miyaso, H., Okumura, M., Kondo, S., Higashide, S., Miyajima, H., and Imaizumi, K. (2003). An intronic splicing enhancer element in survival motor neuron (smn) pre-mrna. *J Biol Chem*, 278(18) :15825–15831.
- Modrek, B. and Lee, C. (2002). A genomic view of alternative splicing. *Nat Genet*, 30(1) :13–19.
- Mooney, S. M., Goel, A., D’Assoro, A. B., Salisbury, J. L., and Janknecht, R. (2010a). Pleiotropic effects of p300-mediated acetylation on p68 and p72 rna helicase. *J Biol Chem*, 285(40) :30443–30452.
- Mooney, S. M., Grande, J. P., Salisbury, J. L., and Janknecht, R. (2010b). Sumoylation of p68 and p72 rna helicases affects protein stability and transactivation potential. *Biochemistry*, 49(1) :1–10.
- Moore, H. C., Johnston, M., Nicol, S. M., Bourdon, J.-C., Thompson, A. M., Hutvagner, G., and Fuller-Pace, F. V. (2011). An evolutionarily conserved, alternatively spliced, intron in the p68/ddx5 dead-box rna helicase gene encodes a novel mirna. *RNA*, 17(4) :555–562.
- Muñoz, M. J., de la Mata, M., and Kornblihtt, A. R. (2010). The carboxy terminal domain of rna polymerase ii and alternative splicing. *Trends Biochem Sci*, 35(9) :497–504.
- Muh, S. J., Hovhannisyan, R. H., and Carstens, R. P. (2002). A non-sequence-specific double-stranded rna structural element regulates splicing of two mutually exclusive exons of fibroblast growth factor receptor 2 (fgfr2). *J Biol Chem*, 277(51) :50143–50154.

- Mulders, S. A. M., van den Broek, W. J. A. A., Wheeler, T. M., Croes, H. J. E., van Kuik-Romeijn, P., de Kimpe, S. J., Furling, D., Platenburg, G. J., Gourdon, G., Thornton, C. A., Wieringa, B., and Wansink, D. G. (2009). Triplet-repeat oligonucleotide-mediated reversal of rna toxicity in myotonic dystrophy. *Proc Natl Acad Sci U S A*, 106(33) :13915–13920.
- Nakamori, M., Gourdon, G., and Thornton, C. A. (2011). Stabilization of expanded (ctg) (cag) repeats by antisense oligonucleotides. *Mol Ther*, 19(12) :2222–2227.
- Nakamori, M., Kimura, T., Kubota, T., Matsumura, T., Sumi, H., Fujimura, H., Takahashi, M. P., and Sakoda, S. (2008). Aberrantly spliced alpha-dystrobrevin alters alpha-syntrophin binding in myotonic dystrophy type 1. *Neurology*, 70(9) :677–685.
- Narla, G., Difeo, A., Reeves, H. L., Schaid, D. J., Hirshfeld, J., Hod, E., Katz, A., Isaacs, W. B., Hebbring, S., Komiya, A., McDonnell, S. K., Wiley, K. E., Jacobsen, S. J., Isaacs, S. D., Walsh, P. C., Zheng, S. L., Chang, B.-L., Friedrichsen, D. M., Stanford, J. L., Ostrander, E. A., Chinnaiyan, A. M., Rubin, M. A., Xu, J., Thibodeau, S. N., Friedman, S. L., and Martignetti, J. A. (2005). A germline dna polymorphism enhances alternative splicing of the klf6 tumor suppressor gene and is associated with increased prostate cancer risk. *Cancer Res*, 65(4) :1213–1222.
- Naya, F. J., Wu, C., Richardson, J. A., Overbeek, P., and Olson, E. N. (1999). Transcriptional activity of mef2 during mouse embryogenesis monitored with a mef2-dependent transgene. *Development*, 126(10) :2045–2052.
- Neubauer, G., King, A., Rappsilber, J., Calvio, C., Watson, M., Ajuh, P., Sleeman, J., Lamond, A., and Mann, M. (1998). Mass spectrometry and est-database searching allows characterization of the multi-protein spliceosome complex. *Nat Genet*, 20(1) :46–50.
- Nogues, G., Kadener, S., Cramer, P., Bentley, D., and Kornblihtt, A. R. (2002). Transcriptional activators differ in their abilities to control alternative splicing. *J Biol Chem*, 277(45) :43110–43114.

- Nogués, G., Muñoz, M. J., and Kornblihtt, A. R. (2003). Influence of polymerase ii processivity on alternative splicing depends on splice site strength. *J Biol Chem*, 278(52) :52166–52171.
- Ogilvie, V. C., Wilson, B. J., Nicol, S. M., Morrice, N. A., Saunders, L. R., Barber, G. N., and Fuller-Pace, F. V. (2003). The highly related dead box rna helicases p68 and p72 exist as heterodimers in cells. *Nucleic Acids Res*, 31(5) :1470–1480.
- Olson, E. N. and Klein, W. H. (1994). bhlh factors in muscle development : dead lines and commitments, what to leave in and what to leave out. *Genes Dev*, 8(1) :1–8.
- Orngreen, M. C., Arlien-S, P., Duno, M., Hertz, J. M., and Vissing, J. (2012). Endocrine function in 97 patients with myotonic dystrophy type 1. *J Neurol*, 259(5) :912–920.
- O’Rourke, J. R., Georges, S. A., Seay, H. R., Tapscott, S. J., McManus, M. T., Goldhamer, D. J., Swanson, M. S., and Harfe, B. D. (2007). Essential role for *dicer* during skeletal muscle development. *Dev Biol*, 311(2) :359–368.
- Osborne, R. J., Lin, X., Welle, S., Sobczak, K., O’Rourke, J. R., Swanson, M. S., and Thornton, C. A. (2009). Transcriptional and post-transcriptional impact of toxic rna in myotonic dystrophy. *Hum Mol Genet*, 18(8) :1471–1481.
- Pan, Q., Shai, O., Lee, L. J., Frey, B. J., and Blencowe, B. J. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet*, 40(12) :1413–1415.
- Pandya-Jones, A. and Black, D. L. (2009). Co-transcriptional splicing of constitutive and alternative exons. *RNA*, 15(10) :1896–1908.
- Parker, M. H., Seale, P., and Rudnicki, M. A. (2003). Looking back to the embryo : defining transcriptional networks in adult myogenesis. *Nat Rev Genet*, 4(7) :497–507.



- Pascual, M., Vicente, M., Monferrer, L., and Artero, R. (2006). The muscleblind family of proteins : an emerging class of regulators of developmentally programmed alternative splicing. *Differentiation*, 74(2-3) :65–80.
- Paul, S., Dansithong, W., Jog, S. P., Holt, I., Mittal, S., Brook, J. D., Morris, G. E., Comai, L., and Reddy, S. (2011). Expanded cug repeats dysregulate rna splicing by altering the stoichiometry of the muscleblind 1 complex. *J Biol Chem*, 286(44) :38427–38438.
- Paul, S., Dansithong, W., Kim, D., Rossi, J., Webster, N. J. G., Comai, L., and Reddy, S. (2006). Interaction of muscleblind, cug-bp1 and hnrbp h proteins in dm1-associated aberrant ir splicing. *EMBO J*, 25(18) :4271–4283.
- Pelletier, R., Krasilnikova, M. M., Samadashwily, G. M., Lahue, R., and Mirkin, S. M. (2003). Replication and expansion of trinucleotide repeats in yeast. *Mol Cell Biol*, 23(4) :1349–1357.
- Penn, B. H., Bergstrom, D. A., Dilworth, F. J., Bengal, E., and Tapscott, S. J. (2004). A myod-generated feed-forward circuit temporally patterns gene expression during skeletal muscle differentiation. *Genes Dev*, 18(19) :2348–2353.
- Perbellini, R., Greco, S., Sarra-Ferraris, G., Cardani, R., Capogrossi, M. C., Meola, G., and Martelli, F. (2011). Dysregulation and cellular mislocalization of specific mirnas in myotonic dystrophy type 1. *Neuromuscul Disord*, 21(2) :81–88.
- Perdiguero, E., Sousa-Victor, P., Ballestar, E., and noz Cánoves, P. M. (2009). Epigenetic regulation of myogenesis. *Epigenetics*, 4(8) :541–550.
- Petri, H., Vissing, J., Witting, N., Bundgaard, H., and K, L. (2012). Cardiac manifestations of myotonic dystrophy type 1. *Int J Cardiol*, 160(2) :82–88.
- Philips, A. V., Timchenko, L. T., and Cooper, T. A. (1998). Disruption of splicing regulated by a cug-binding protein in myotonic dystrophy. *Science*, 280(5364) :737–741.

- Pinotsis, N., Chatziefthimiou, S. D., Berkemeier, F., Beuron, F., Mavridis, I. M., Konarev, P. V., Svergun, D. I., Morris, E., Rief, M., and Wilmanns, M. (2012). Superhelical architecture of the myosin filament-linking protein myomesin with unusual elastic properties. *PLoS Biol*, 10(2) :e1001261.
- Pradeepa, M. M., Sutherland, H. G., Ule, J., Grimes, G. R., and Bickmore, W. A. (2012). Psp1/ledgf p52 binds methylated histone h3k36 and splicing factors and contributes to the regulation of alternative splicing. *PLoS Genet*, 8(5) :e1002717.
- Puri, P. L. and Sartorelli, V. (2000). Regulation of muscle regulatory factors by dna-binding, interacting proteins, and post-transcriptional modifications. *J Cell Physiol*, 185(2) :155–173.
- Putnam, A. A. and Jankowsky, E. (2013). Dead-box helicases as integrators of rna, nucleotide and protein binding. *Biochim Biophys Acta*, 1829(8) :884–893.
- Pyle, A. M. (2008). Translocation and unwinding mechanisms of rna and dna helicases. *Annu Rev Biophys*, 37 :317–336.
- Pyle, A. M., Fedorova, O., and Waldsich, C. (2007). Folding of group ii introns : a model system for large, multidomain rnas ? *Trends Biochem Sci*, 32(3) :138–145.
- Raghavachari, N., Barb, J., Yang, Y., Liu, P., Woodhouse, K., Levy, D., O'Donnell, C. J., Munson, P. J., and Kato, G. J. (2012). A systematic comparison and evaluation of high density exon arrays and rna-seq technology used to unravel the peripheral blood transcriptome of sickle cell disease. *BMC Med Genomics*, 5 :28.
- Raj, B., O'Hanlon, D., Vessey, J. P., Pan, Q., Ray, D., Buckley, N. J., Miller, F. D., and Blencowe, B. J. (2011). Cross-regulation between an alternative splicing activator and a transcription repressor controls neurogenesis. *Mol Cell*, 43(5) :843–850.
- Rao, P. K., Kumar, R. M., Farkhondeh, M., Baskerville, S., and Lodish, H. F. (2006). Myogenic factors that regulate expression of muscle-specific micrnas. *Proc Natl Acad Sci U S A*, 103(23) :8721–8726.

- Rau, F. (2011). *Etude de la dérégulation de miR-1 dans le cœur de patients atteints de Dystrophies Myotoniques*. PhD thesis, Université de Strasbourg.
- Rau, F., Freyermuth, F., Fugier, C., Villemin, J.-P., Fischer, M.-C., Jost, B., Dembele, D., Gourdon, G., Nicole, A., Duboc, D., Wahbi, K., Day, J. W., Fujimura, H., Takahashi, M. P., Auboeuf, D., Dreumont, N., Furling, D., and Charlet-Berguerand, N. (2011). Misregulation of mir-1 processing is associated with heart defects in myotonic dystrophy. *Nat Struct Mol Biol*, 18(7) :840–845.
- Rebbapragada, I. and Lykke-Andersen, J. (2009). Execution of nonsense-mediated mrna decay : what defines a substrate? *Curr Opin Cell Biol*, 21(3) :394–402.
- Reddy, S., Smith, D. B., Rich, M. M., Leferovich, J. M., Reilly, P., Davis, B. M., Tran, K., Rayburn, H., Bronson, R., Cros, D., Balice-Gordon, R. J., and Housman, D. (1996). Mice lacking the myotonic dystrophy protein kinase develop a late onset progressive myopathy. *Nat Genet*, 13(3) :325–335.
- Rochlin, K., Yu, S., Roy, S., and Baylies, M. K. (2010). Myoblast fusion : when it takes more to make one. *Dev Biol*, 341(1) :66–83.
- Roof, D. J., Hayes, A., Adamian, M., Chishti, A. H., and Li, T. (1997). Molecular characterization of ablim, a novel actin-binding and double zinc finger protein. *J Cell Biol*, 138(3) :575–588.
- Rössler, O. G., Hloch, P., Schütz, N., Weitzenegger, T., and Stahl, H. (2000). Structure and expression of the human p68 rna helicase gene. *Nucleic Acids Res*, 28(4) :932–939.
- Rossow, K. L. and Janknecht, R. (2003). Synergism between p68 rna helicase and the transcriptional coactivators cbp and p300. *Oncogene*, 22(1) :151–156.
- Salzman, D. W., Shubert-Coleman, J., and Furneaux, H. (2007). P68 rna helicase unwinds the human let-7 microrna precursor duplex and is required for let-7-directed silencing of gene expression. *J Biol Chem*, 282(45) :32773–32779.
- Samaan, S. (2013). *Implication of alternative splicing in signaling pathways regulation*. PhD thesis, Université Paris Diderot (Paris 7).

- Samaan, S., Tranchevent, L.-C., Dardenne, E., Espinoza, M. P., Zonta, E., Germann, S., Gratadou, L., Dutertre, M., and Auboeuf, D. (2013). The *ddx5* and *ddx17* rna helicases are cornerstones in the complex regulatory array of steroid hormone-signaling pathways. *Nucleic Acids Res.*
- Sammeth, M., Foissac, S., and Guigó, R. (2008). A general definition and nomenclature for alternative splicing events. *PLoS Comput Biol*, 4(8) :e1000147.
- Sandre-Giovannoli, A. D. and Lévy, N. (2006). Altered splicing in prelamin a-associated premature aging phenotypes. *Prog Mol Subcell Biol*, 44 :199–232.
- Santerre, R. F., Bales, K. R., Janney, M. J., Hannon, K., Fisher, L. F., Bailey, C. S., Morris, J., Ivarie, R., and Smith, 2nd, C. (1993). Expression of bovine *myf5* induces ectopic skeletal muscle formation in transgenic mice. *Mol Cell Biol*, 13(10) :6044–6051.
- Saporita, A. J., Chang, H.-C., Winkeler, C. L., Apicelli, A. J., Kladney, R. D., Wang, J., Townsend, R. R., Michel, L. S., and Weber, J. D. (2011). Rna helicase *ddx5* is a p53-independent target of arf that participates in ribosome biogenesis. *Cancer Res*, 71(21) :6708–6717.
- Saulière, J., Sureau, A., Expert-Bezançon, A., and Marie, J. (2006). The polypyrimidine tract binding protein (*ptb*) represses splicing of exon 6b from the beta-tropomyosin pre-mrna by directly interfering with the binding of the *u2af65* subunit. *Mol Cell Biol*, 26(23) :8755–8769.
- Savkur, R. S., Philips, A. V., and Cooper, T. A. (2001). Aberrant regulation of insulin receptor alternative splicing is associated with insulin resistance in myotonic dystrophy. *Nat Genet*, 29(1) :40–47.
- Schaal, T. D. and Maniatis, T. (1999). Multiple distinct splicing enhancers in the protein-coding sequences of a constitutively spliced pre-mrna. *Mol Cell Biol*, 19(1) :261–273.
- Schor, I. E., Rascovan, N., Pelisch, F., Alló, M., and Kornblihtt, A. R. (2009).

- Neuronal cell depolarization induces intragenic chromatin modifications affecting ncam alternative splicing. *Proc Natl Acad Sci U S A*, 106(11) :4325–4330.
- Schwander, M., Shirasaki, R., Pfaff, S. L., and Müller, U. (2004). Beta1 integrins in muscle, but not in motor neurons, are required for skeletal muscle innervation. *J Neurosci*, 24(37) :8181–8191.
- Sen, S., Talukdar, I., Liu, Y., Tam, J., Reddy, S., and Webster, N. J. G. (2010). Muscleblind-like 1 (mbnl1) promotes insulin receptor exon 11 inclusion via binding to a downstream evolutionarily conserved intronic enhancer. *J Biol Chem*, 285(33) :25426–25437.
- Sen, S., Talukdar, I., and Webster, N. J. G. (2009). Srp20 and cug-bp1 modulate insulin receptor exon 11 alternative splicing. *Mol Cell Biol*, 29(3) :871–880.
- Sergeant, N., Sablonnière, B., Schraen-Maschke, S., Ghestem, A., Maurage, C. A., Wattez, A., Vermersch, P., and Delacourte, A. (2001). Dysregulation of human brain microtubule-associated tau mrna maturation in myotonic dystrophy type 1. *Hum Mol Genet*, 10(19) :2143–2155.
- Seznec, H., Agbulut, O., Sergeant, N., Savouret, C., Ghestem, A., Tabti, N., Willer, J. C., Ourth, L., Duros, C., Brisson, E., Fouquet, C., Butler-Browne, G., Delacourte, A., Junien, C., and Gourdon, G. (2001). Mice transgenic for the human myotonic dystrophy region with expanded ctg repeats display muscular and brain abnormalities. *Hum Mol Genet*, 10(23) :2717–2726.
- Shin, S. and Janknecht, R. (2007). Concerted activation of the mdm2 promoter by p72 rna helicase and the coactivators p300 and p/caf. *J Cell Biochem*, 101(5) :1252–1265.
- Shin, S., Rossow, K. L., Grande, J. P., and Janknecht, R. (2007). Involvement of rna helicases p68 and p72 in colon cancer. *Cancer Res*, 67(16) :7572–7578.
- Sicot, G. (2013). *Conséquences pathologiques des expansions CTG sur le système nerveux central d'un modèle murin de la dystrophie myotonique de Steinert :*

- Approches moléculaires, protéomiques et cellulaires*. PhD thesis, Université Paris Descartes / Paris V.
- Sicot, G., Gourdon, G., and Gomes-Pereira, M. (2011). Myotonic dystrophy, when simple repeats reveal complex pathogenic entities : new findings and future challenges. *Hum Mol Genet*, 20(R2) :R116–R123.
- Singh, N. N., Singh, R. N., and Androphy, E. J. (2007). Modulating role of rna structure in alternative splicing of a critical exon in the spinal muscular atrophy genes. *Nucleic Acids Res*, 35(2) :371–389.
- Singh, R. K. and Cooper, T. A. (2012). Pre-mrna splicing in disease and therapeutics. *Trends Mol Med*, 18(8) :472–482.
- Singleton, M. R., Dillingham, M. S., and Wigley, D. B. (2007). Structure and mechanism of helicases and nucleic acid translocases. *Annu Rev Biochem*, 76 :23–50.
- Skinner, M. K., Rawls, A., Wilson-Rawls, J., and Roalson, E. H. (2010). Basic helix-loop-helix transcription factor gene family phylogenetics and nomenclature. *Differentiation*, 80(1) :1–8.
- Soleimani, V. D., Yin, H., Jahani-Asl, A., Ming, H., Kockx, C. E. M., van Ijcken, W. F. J., Grosveld, F., and Rudnicki, M. A. (2012). Snail regulates myod binding-site occupancy to direct enhancer switching and differentiation-specific transcription in myogenesis. *Mol Cell*, 47(3) :457–468.
- Speel, E. J., van der Ven, P. F., Albrechts, J. C., Ramaekers, F. C., Fürst, D. O., and Hopman, A. H. (1998). Assignment of the human gene for the sarcomeric m-band protein myomesin (myom1) to 18p11.31-p11.32. *Genomics*, 54(1) :184–186.
- Spellman, R. and Smith, C. W. J. (2006). Novel modes of splicing repression by ptb. *Trends Biochem Sci*, 31(2) :73–76.
- Sperling, J., Azubel, M., and Sperling, R. (2008). Structure and function of the pre-mrna splicing machine. *Structure*, 16(11) :1605–1615.

- Spicer, D. B., Rhee, J., Cheung, W. L., and Lassar, A. B. (1996). Inhibition of myogenic bhlh and mef2 transcription factors by the bhlh protein twist. *Science*, 272(5267) :1476–1480.
- Spies, N., Nielsen, C. B., Padgett, R. A., and Burge, C. B. (2009). Biased chromatin signatures around polyadenylation sites and exons. *Mol Cell*, 36(2) :245–254.
- Spranger, M., Spranger, S., Tischendorf, M., Meinck, H. M., and Cremer, M. (1997). Myotonic dystrophy. the role of large triplet repeat length in the development of mental retardation. *Arch Neurol*, 54(3) :251–254.
- Srebrow, A. and Kornblihtt, A. R. (2006). The connection between splicing and cancer. *J Cell Sci*, 119(Pt 13) :2635–2641.
- Stamm, S., Ben-Ari, S., Rafalska, I., Tang, Y., Zhang, Z., Toiber, D., Thanaraj, T. A., and Soreq, H. (2005). Function of alternative splicing. *Gene*, 344 :1–20.
- Stevenson, R. J., Hamilton, S. J., MacCallum, D. E., Hall, P. A., and Fuller-Pace, F. V. (1998). Expression of the 'dead box' rna helicase p68 is developmentally and growth regulated and correlates with organ differentiation/maturation in the fetus. *J Pathol*, 184(4) :351–359.
- Summerbell, D., Halai, C., and Rigby, P. W. J. (2002). Expression of the myogenic regulatory factor mrf4 precedes or is contemporaneous with that of myf5 in the somitic bud. *Mech Dev*, 117(1-2) :331–335.
- Sun, H. and Chasin, L. A. (2000). Multiple splicing defects in an intronic false exon. *Mol Cell Biol*, 20(17) :6414–6425.
- Sun, Q., Zhang, Y., Yang, G., Chen, X., Zhang, Y., Cao, G., Wang, J., Sun, Y., Zhang, P., Fan, M., Shao, N., and Yang, X. (2008). Transforming growth factor-beta-regulated mir-24 promotes skeletal muscle differentiation. *Nucleic Acids Res*, 36(8) :2690–2699.
- Sureau, A., Saulière, J., Expert-Bezançon, A., and Marie, J. (2011). Celf and ptb proteins modulate the inclusion of the  $\beta$ -tropomyosin exon 6b during myogenic differentiation. *Exp Cell Res*, 317(1) :94–106.



- Sweetman, D., Goljanek, K., Rathjen, T., Oustanina, S., Braun, T., Dalmay, T., and Münsterberg, A. (2008). Specific requirements of mrfs for the expression of muscle specific micrnas, mir-1, mir-206 and mir-133. *Dev Biol*, 321(2) :491–499.
- Tange, T. O., Damgaard, C. K., Guth, S., Valcárcel, J., and Kjems, J. (2001). The hnrnp a1 protein regulates hiv-1 tat splicing via a novel intron silencer element. *EMBO J*, 20(20) :5748–5758.
- Tanner, N. K. and Linder, P. (2001). Dexe/h box rna helicases : from generic motors to specific dissociation functions. *Mol Cell*, 8(2) :251–262.
- Theissen, B., Karow, A. R., Köhler, J., Gubaev, A., and Klostermeier, D. (2008). Cooperative binding of atp and rna induces a closed conformation in a dead box rna helicase. *Proc Natl Acad Sci U S A*, 105(2) :548–553.
- Theodorou, E., Dalembert, G., Heffelfinger, C., White, E., Weissman, S., Corcoran, L., and Snyder, M. (2009). A high throughput embryonic stem cell screen identifies oct-2 as a bifunctional regulator of neuronal differentiation. *Genes Dev*, 23(5) :575–588.
- Tijerina, P., Bhaskaran, H., and Russell, R. (2006). Nonspecific binding to structured rna and preferential unwinding of an exposed helix by the cyt-19 protein, a dead-box rna chaperone. *Proc Natl Acad Sci U S A*, 103(45) :16698–16703.
- Tilgner, H., Knowles, D. G., Johnson, R., Davis, C. A., Chakraborty, S., Djebali, S., Curado, J. a., Snyder, M., Gingeras, T. R., and Guigó, R. (2012). Deep sequencing of subcellular rna fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncnas. *Genome Res*, 22(9) :1616–1625.
- Tilgner, H., Nikolaou, C., Althammer, S., Sammeth, M., Beato, M., Valcárcel, J., and Guigó, R. (2009). Nucleosome positioning as a determinant of exon recognition. *Nat Struct Mol Biol*, 16(9) :996–1001.
- Timchenko, N. A., Cai, Z. J., Welm, A. L., Reddy, S., Ashizawa, T., and Timchenko, L. T. (2001a). Rna cug repeats sequester cugbp1 and alter protein levels and activity of cugbp1. *J Biol Chem*, 276(11) :7820–7826.

- Timchenko, N. A., Iakova, P., Cai, Z. J., Smith, J. R., and Timchenko, L. T. (2001b). Molecular basis for impaired muscle differentiation in myotonic dystrophy. *Mol Cell Biol*, 21(20) :6927–6938.
- Timchenko, N. A., Patel, R., Iakova, P., Cai, Z.-J., Quan, L., and Timchenko, L. T. (2004). Overexpression of cug triplet repeat-binding protein, cugbp1, in mice inhibits myogenesis. *J Biol Chem*, 279(13) :13129–13139.
- Timchenko, N. A., Wang, G.-L., and Timchenko, L. T. (2005). Rna cug-binding protein 1 increases translation of 20-kda isoform of ccaat/enhancer-binding protein beta by interacting with the alpha and beta subunits of eukaryotic initiation translation factor 2. *J Biol Chem*, 280(21) :20549–20557.
- Timchenko, N. A., Welm, A. L., Lu, X., and Timchenko, L. T. (1999). Cug repeat binding protein (cugbp1) interacts with the 5' region of c/ebpbeta mrna and regulates translation of c/ebpbeta isoforms. *Nucleic Acids Res*, 27(22) :4517–4525.
- Tran, H., Gourrier, N., Lemercier-Neuillet, C., Dhaenens, C.-M., Vautrin, A., Fernandez-Gomez, F. J., Arandel, L., Carpentier, C., Obriot, H., Eddarkaoui, S., Delattre, L., Van Brussels, E., Holt, I., Morris, G. E., Sablonnière, B., Buée, L., Charlet-Berguerand, N., Schraen-Maschke, S., Furling, D., Behm-Ansmant, I., Branlant, C., Caillet-Boudin, M.-L., and Sergeant, N. (2011). Analysis of exonic regions involved in nuclear localization, splicing activity, and dimerization of muscleblind-like-1 isoforms. *J Biol Chem*, 286(18) :16435–16446.
- Turner, C. and Hilton-Jones, D. (2010). The myotonic dystrophies : diagnosis and management. *J Neurol Neurosurg Psychiatry*, 81(4) :358–367.
- Uhlmann-Schiffler, H., Rössler, O. G., and Stahl, H. (2002). The mrna of dead box protein p72 is alternatively translated into an 82-kda rna helicase. *J Biol Chem*, 277(2) :1066–1075.
- Vargas, D. Y., Shah, K., Batish, M., Levandoski, M., Sinha, S., Marras, S. A. E.,

- Schedl, P., and Tyagi, S. (2011). Single-molecule imaging of transcriptionally coupled and uncoupled splicing. *Cell*, 147(5) :1054–1065.
- Venables, J. P., Lapasset, L., Gadea, G., Fort, P., Klinck, R., Irimia, M., Vignal, E., Thibault, P., Prinos, P., Chabot, B., Abou Elela, S., Roux, P., Lemaitre, J.-M., and Tazi, J. (2013). Mbnl1 and rbfox2 cooperate to establish a splicing programme involved in pluripotent stem cell differentiation. *Nat Commun*, 4 :2480.
- Veyckemans, F. and Scholtes, J.-L. (2013). Myotonic dystrophies type 1 and 2 : anesthetic care. *Paediatr Anaesth*, 23(9) :794–803.
- Vousden, K. H. and Prives, C. (2009). Blinded by the light : The growing complexity of p53. *Cell*, 137(3) :413–431.
- Wagner, M., Rid, R., Maier, C. J., Maier, R. H., Laimer, M., Hintner, H., Bauer, J. W., and Onder, K. (2012). Ddx5 is a multifunctional co-activator of steroid hormone receptors. *Mol Cell Endocrinol*, 361(1-2) :80–91.
- Wang, D., Huang, J., and Hu, Z. (2012a). Rna helicase ddx5 regulates microrna expression and contributes to cytoskeletal reorganization in basal breast cancer cells. *Mol Cell Proteomics*, 11(2) :M111.011932.
- Wang, E. T., Cody, N. A. L., Jog, S., Biancolella, M., Wang, T. T., Treacy, D. J., Luo, S., Schroth, G. P., Housman, D. E., Reddy, S., Lécuyer, E., and Burge, C. B. (2012b). Transcriptome-wide regulation of pre-mrna splicing and mrna localization by muscleblind proteins. *Cell*, 150(4) :710–724.
- Wang, G.-S., Kearney, D. L., De Biasi, M., Taffet, G., and Cooper, T. A. (2007). Elevation of rna-binding protein cugbp1 is an early event in an inducible heart-specific mouse model of myotonic dystrophy. *J Clin Invest*, 117(10) :2802–2811.
- Wang, H., Gao, X., Huang, Y., Yang, J., and Liu, Z.-R. (2009). P68 rna helicase is a nucleocytoplasmic shuttling protein. *Cell Res*, 19(12) :1388–1400.
- Wang, H., Gao, X., Yang, J. J., and Liu, Z.-R. (2013). Interaction between p68 rna helicase and ca<sup>2+</sup>-calmodulin promotes cell migration and metastasis. *Nat Commun*, 4 :1354.

- Wang, L. and Qian, L. (2014). mir-24 regulates intrinsic apoptosis pathway in mouse cardiomyocytes. *PLoS One*, 9(1) :e85389.
- Wang, L.-C., Hung, W.-T., Pan, H., Chen, K.-Y., Wu, Y.-C., Liu, Y.-F., and Hsiao, K.-M. (2008). Growth-dependent effect of muscleblind knockdown on *caenorhabditis elegans*. *Biochem Biophys Res Commun*, 366(3) :705–709.
- Wang, R., Jiao, Z., Li, R., Yue, H., and Chen, L. (2012c). p68 rna helicase promotes glioma cell proliferation in vitro and in vivo via direct regulation of nf- $\kappa$ b transcription factor p50. *Neuro Oncol*, 14(9) :1116–1124.
- Wang, S. J., Zhang, C., You, Y., and Shi, C. M. (2012d). Overexpression of rna helicase p68 protein in cutaneous squamous cell carcinoma. *Clin Exp Dermatol*, 37(8) :882–888.
- Ward, A. J., Rimer, M., Killian, J. M., Dowling, J. J., and Cooper, T. A. (2010). Cugbp1 overexpression in mouse skeletal muscle reproduces features of myotonic dystrophy type 1. *Hum Mol Genet*, 19(18) :3614–3622.
- Warf, M. B., Diegel, J. V., von Hippel, P. H., and 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(23) :9203–9208.
- Warzecha, C. C., Sato, T. K., Nabet, B., Hogenesch, J. B., and Carstens, R. P. (2009). Esrp1 and esrp2 are epithelial cell-type-specific regulators of fgfr2 splicing. *Mol Cell*, 33(5) :591–601.
- Weintraub, H., Davis, R., Tapscott, S., Thayer, M., Krause, M., Benezra, R., Blackwell, T. K., Turner, D., Rupp, R., and Hollenberg, S. (1991). The myod gene family : nodal point during specification of the muscle cell lineage. *Science*, 251(4995) :761–766.
- Wheeler, T. M., Leger, A. J., Pandey, S. K., MacLeod, A. R., Nakamori, M., Cheng, S. H., Wentworth, B. M., Bennett, C. F., and Thornton, C. A. (2012). Targeting nuclear rna for in vivo correction of myotonic dystrophy. *Nature*, 488(7409) :111–115.

- Will, C. L. and Lührmann, R. (2011). Spliceosome structure and function. *Cold Spring Harb Perspect Biol*, 3(7).
- Williams, A. H., Liu, N., van Rooij, E., and Olson, E. N. (2009). MicroRNA control of muscle development and disease. *Curr Opin Cell Biol*, 21(3) :461–469.
- Wilson, B. J., Bates, G. J., Nicol, S. M., Gregory, D. J., Perkins, N. D., and Fuller-Pace, F. V. (2004). The p68 and p72 dead box rna helicases interact with hdac1 and repress transcription in a promoter-specific manner. *BMC Mol Biol*, 5 :11.
- Win, A. K., Perattur, P. G., Pulido, J. S., Pulido, C. M., and Lindor, N. M. (2012). Increased cancer risks in myotonic dystrophy. *Mayo Clin Proc*, 87(2) :130–135.
- Winblad, S., Lindberg, C., and Hansen, S. (2005). Temperament and character in patients with classical myotonic dystrophy type 1 (dm-1). *Neuromuscul Disord*, 15(4) :287–292.
- Winter, J., Jung, S., Keller, S., Gregory, R. I., and Diederichs, S. (2009). Many roads to maturity : microRNA biogenesis pathways and their regulation. *Nat Cell Biol*, 11(3) :228–234.
- Witten, J. T. and Ule, J. (2011). Understanding splicing regulation through rna splicing maps. *Trends Genet*, 27(3) :89–97.
- Wong, C. F. and Tellam, R. L. (2008). MicroRNA-26a targets the histone methyltransferase enhancer of zeste homolog 2 during myogenesis. *J Biol Chem*, 283(15) :9836–9843.
- Wortham, N. C., Ahamed, E., Nicol, S. M., Thomas, R. S., Periyasamy, M., Jiang, J., Ochocka, A. M., Shousha, S., Huson, L., Bray, S. E., Coombes, R. C., Ali, S., and Fuller-Pace, F. V. (2009). The dead-box protein p72 regulates eralpha-/oestrogen-dependent transcription and cell growth, and is associated with improved survival in eralpha-positive breast cancer. *Oncogene*, 28(46) :4053–4064.
- Xu, W., Seok, J., Mindrinos, M. N., Schweitzer, A. C., Jiang, H., Wilhelmy, J., Clark, T. A., Kapur, K., Xing, Y., Faham, M., Storey, J. D., Moldawer, L. L.,

- Maier, R. V., Tompkins, R. G., Wong, W. H., Davis, R. W., Xiao, W., , I., and to Injury Large-Scale Collaborative Research Program, H. R. (2011). Human transcriptome array for high-throughput clinical studies. *Proc Natl Acad Sci U S A*, 108(9) :3707–3712.
- Yaffe, D. and Saxel, O. (1977). Serial passaging and differentiation of myogenic cells isolated from dystrophic mouse muscle. *Nature*, 270(5639) :725–727.
- Yahi, H., Philipot, O., Guasconi, V., Fritsch, L., and Ait-Si-Ali, S. (2006). Chromatin modification and muscle differentiation. *Expert Opin Ther Targets*, 10(6) :923–934.
- Yan, B. (2010). Numb—from flies to humans. *Brain Dev*, 32(4) :293–298.
- Yang, L., Lin, C., and Liu, Z.-R. (2006). P68 rna helicase mediates pdgf-induced epithelial mesenchymal transition by displacing axin from beta-catenin. *Cell*, 127(1) :139–155.
- Yang, L., Lin, C., Zhao, S., Wang, H., and Liu, Z.-R. (2007a). Phosphorylation of p68 rna helicase plays a role in platelet-derived growth factor-induced cell proliferation by up-regulating cyclin d1 and c-myc expression. *J Biol Chem*, 282(23) :16811–16819.
- Yang, Q., Del Campo, M., Lambowitz, A. M., and Jankowsky, E. (2007b). Dead-box proteins unwind duplexes by local strand separation. *Mol Cell*, 28(2) :253–263.
- Yang, Q. and Jankowsky, E. (2006). The dead-box protein ded1 unwinds rna duplexes by a mode distinct from translocating helicases. *Nat Struct Mol Biol*, 13(11) :981–986.
- Yeo, G. W., Van Nostrand, E., Holste, D., Poggio, T., and Burge, C. B. (2005). Identification and analysis of alternative splicing events conserved in human and mouse. *Proc Natl Acad Sci U S A*, 102(8) :2850–2855.
- Yotova, V., Labuda, D., Zietkiewicz, E., Gehl, D., Lovell, A., Lefebvre, J.-F., Bourgeois, S., Lemieux-Blanchard, E., Labuda, M., Vézina, H., Houde, L.,

- Tremblay, M., Toupance, B., Heyer, E., Hudson, T. J., and Laberge, C. (2005). Anatomy of a founder effect : myotonic dystrophy in northeastern quebec. *Hum Genet*, 117(2-3) :177–187.
- Yu, Z., Teng, X., and Bonini, N. M. (2011). Triplet repeat-derived sirnas enhance rna-mediated toxicity in a drosophila model for myotonic dystrophy. *PLoS Genet*, 7(3) :e1001340.
- Yuan, Y., Compton, S. A., Sobczak, K., Stenberg, M. G., Thornton, C. A., Griffith, J. D., and Swanson, M. S. (2007). Muscleblind-like 1 interacts with rna hairpins in splicing target and pathogenic rnas. *Nucleic Acids Res*, 35(16) :5474–5486.
- Yun, K. and Wold, B. (1996). Skeletal muscle determination and differentiation : story of a core regulatory network and its context. *Curr Opin Cell Biol*, 8(6) :877–889.
- Zerbe, L. K., Pino, I., Pio, R., Cospers, P. F., Dwyer-Nield, L. D., Meyer, A. M., Port, J. D., Montuenga, L. M., and Malkinson, A. M. (2004). Relative amounts of antagonistic splicing factors, hnrnp a1 and asf/sf2, change during neoplastic lung growth : implications for pre-mrna processing. *Mol Carcinog*, 41(4) :187–196.
- Zhang, X. H.-F., Kangsamaksin, T., Chao, M. S. P., Banerjee, J. K., and Chasin, L. A. (2005). Exon inclusion is dependent on predictable exonic splicing enhancers. *Mol Cell Biol*, 25(16) :7323–7332.
- Zhu, J., Mayeda, A., and Krainer, A. R. (2001). Exon identity established through differential antagonism between exonic splicing silencer-bound hnrnp a1 and enhancer-bound sr proteins. *Mol Cell*, 8(6) :1351–1361.
- Zonta, E., Bittencourt, D., Samaan, S., Germann, S., Dutertre, M., and Auboeuf, D. (2013). The rna helicase ddx5/p68 is a key factor promoting c-fos expression at different levels from transcription to mrna export. *Nucleic Acids Res*, 41(1) :554–564.
- Zu, T., Gibbens, B., Doty, N. S., Gomes-Pereira, M., Huguet, A., Stone, M. D., Margolis, J., Peterson, M., Markowski, T. W., Ingram, M. A. C., Nan, Z., Forster,



---

C., Low, W. C., Schoser, B., Somia, N. V., Clark, H. B., Schmechel, S., Bitterman, P. B., Gourdon, G., Swanson, M. S., Moseley, M., and Ranum, L. P. W. (2011). Non-atg-initiated translation directed by microsatellite expansions. *Proc Natl Acad Sci U S A*, 108(1) :260–265.