



**HAL**  
open science

# Rôle des hélicases DDX5 et DDX17 dans l'expression et la maturation des ARN pré-messagers dans le contexte de pathologies neuro-développementales

Valentine Clerc

► **To cite this version:**

Valentine Clerc. Rôle des hélicases DDX5 et DDX17 dans l'expression et la maturation des ARN pré-messagers dans le contexte de pathologies neuro-développementales. Biologie moléculaire. Ecole normale supérieure de lyon - ENS LYON, 2023. Français. NNT : 2023ENSL0116 . tel-04529878

**HAL Id: tel-04529878**

**<https://theses.hal.science/tel-04529878>**

Submitted on 2 Apr 2024

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

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



# THESE

en vue de l'obtention du grade de Docteur, délivré par  
l'ECOLE NORMALE SUPERIEURE DE LYON

**Ecole Doctorale N° 340  
Biologie Moléculaire, Intégrative et Cellulaire (BMIC)**

**Discipline : Sciences de la Vie et de la Santé**

Soutenue publiquement le 15/12/2023, par :  
**Valentine Clerc**

---

**Rôle des hélicases DDX5 et DDX17 dans l'expression  
et la maturation des ARN pré-messagers dans le  
contexte de pathologies neuro-développementales**

---

Devant le jury composé de :

DUTERTRE Martin, Chargé de Recherche (HDR) – Institut Curie, Orsay  
MAZOYER Sylvie, Directrice de Recherche – CRNL, Lyon  
MOLLEREAU Bertrand, Professeur des Universités – ENS de Lyon  
BOURGEOIS Cyril, Chargé de Recherche (HDR) – LBMC, Lyon

Rapporteur  
Rapporteuse  
Examineur  
Directeur de Thèse

# Remerciements

Je voudrais tout d'abord commencer en remerciant les membres de mon jury. Merci aux deux rapporteurs de ma thèse, le Dr Martin Dutertre et le Dr Sylvie Mazoyer, d'avoir accepté de donner de leur temps pour relire mon manuscrit et pour leurs judicieuses suggestions. Je remercie également les autres membres de mon jury, le Pr Bertrand Mollereau et le Dr Cécile Martinat d'avoir accepté de faire partie de mon jury et d'évaluer mes travaux de recherche.

Je souhaite remercier les deux membres de mon comité de suivi de thèse, le Dr Stefania Millevoi et le Dr Benjamin Gibert, pour les conseils et les commentaires constructifs qu'ils ont pu me donner chaque année afin de faire avancer mon projet.

Je remercie la Fondation pour la Recherche Médicale pour la bourse de thèse que j'ai pu obtenir, m'ayant permis de mener mes travaux de thèse durant ces trois ans.

Je remercie le Dr Didier Auboeuf de m'avoir accueillie au sein de son équipe dès mon stage de master.

Je remercie mon encadrant, le Dr Cyril Bourgeois, qui m'a permis de réaliser mon stage de master 2 dans cette équipe et qui m'a ensuite donné l'opportunité de poursuivre en doctorat. Merci pour sa présence et ses nombreux conseils, particulièrement les corrections de rédaction. Un directeur de thèse qui a le sens du dialogue et qui est toujours disponible à ouvrir la porte de son bureau pour répondre à mes questions. Même s'il faut savoir que derrière cette porte de bureau, le temps se distord et que l'on n'est jamais certain de pouvoir en ressortir le même jour que l'on y est entré. Merci d'ailleurs à Franck Mortreux, partageant ce bureau, d'avoir supporté ces interruptions imprévisibles.

Merci bien sûr à toute l'équipe *ReGArDs*. Un merci spécial aux nombreux anciens membres de l'équipe ayant déjà fait leurs bagages. Lamya que je n'aurais pas connu longtemps, mais qui aura toujours répondu à mes questions. Gros merci à Jessica qui m'a encadré durant mon stage et qui aura été une des rares touches de féminité au sein du laboratoire. Ta contribution à l'ambiance musicale de nos manip sera à toujours regrettée. Bon courage pour combattre le crime dans le Vieux-Port ma gâtée. Merci au petit Pôle, qui aura tenté de maintenir un semblant d'ordre dans ce laboratoire du chaos. Un talentueux créateur de même un peu tête en l'air, mais avec plein de bonnes idées. Je suis sûre que tu trouveras ton chemin une fois que tu seras lassé du chômeur. Prends bien soin d'Olivia et de Ribosome.

Merci à Mathieu/Fabien, le maître du chaos en ces lieux, pour avoir gardé ma chaise de bureau chaude quand j'étais en manips. Ça fait bizarre de ne plus entendre tes cris à chaque coin de couloirs. Avec ta tête pleine d'idées aussi farfelues et ingénieuses les unes que les autres, je ne me fais pas de soucis pour ton futur, j'espère juste que Caro saura contrôler le génie fou qui sommeille en toi. Merci à Audrey Gibert pour les discussions engagées et surtout les parties de Scopa !

Merci à tous ceux étant encore parmi nous. Nicolas Fontrodona, Xavier Grand, Julien Ladet, Audrey Lapendry, Hélène Polvèche, Mateo Bazire, William D et Mia Croiset. Merci à la team bio-info : Hélène pour sa patience, ses conseils et sa bonne humeur. Xavier pour avoir eu le courage d'arriver à faire fonctionner le pipeline d'analyse des chimériques. Nicolas mon frère rat, qui est toujours disponible pour donner un coup de patte et me prévenir de l'approvisionnement en cappuccino-caramel. Mia qui est pour moi devenue un membre à part entière de cette équipe et indispensable à nos parties de pédantix. Bon courage au petit nouveau William, sache que tu es entre de bonnes mains. Merci spécial pour Audrey L, qui aura quasiment commencé et fini sa thèse en même temps que moi. Le plus mignon et gentil des hippo. Toujours là pour aider, conseiller et écouter. J'espère qu'il ne t'arrivera que des bonnes choses par la suite.

Merci à mes colocataires de bureau Julien et Mateo, parmi les personnes les plus drôlement débiles que je connaisse. Vos conneries m'auront vraiment manquées pendant la rédaction. Mateo qui s'est intégré bien trop vite à l'équipe, si bien qu'il a été assez fou pour nous rejoindre en thèse. Force et courage pour ton aventure au pays de la 3D. Julien, merci d'avoir été mon voisin de paillasse. Chaque équipe devrait compter un Julien parmi ses membres, tu es vraiment devenu l'âme de ce labo. Dans le cercle fermé des spécialistes des circRNAs, il ne te manque pas grand-chose pour boucler la boucle. Dédicace spéciale à Apo/Maité qui est une des rares personnes ayant assez de pouvoir pour arriver à instaurer un semblant de calme dans le bureau.

Merci à ceux ayant furtivement peuplé ces lieux durant leur stage mais qui nous auront durablement marqué. Lola qui a été un vrai petit soleil (un soleil qui n'hésitait pas à se coucher tard pour terminer ses extractions d'ARN) et Lavinia qui m'a permis, j'espère, d'améliorer mon anglais et qu'avoir encadré m'aura permis d'apprendre bien plus de choses que je ne lui en ai appris.

Petite dédicace à feu notre mascotte Sharky. Tu n'auras malheureusement pas supporté le départ de ton père de cœur. J'espère qu'au paradis des poissons tu m'es reconnaissant d'avoir creusé ta tombe quasiment à main nu au mépris de ma dignité. Ton manque sera à jamais irremplaçable mais Hutchy s'est déjà merveilleusement bien intégré à l'équipe. Puisse-tu nager encore longtemps brave combattant.

Je remercie aussi tous les autres membres du LBMC (mais pas que) qui auront pu partager leurs idées, conseils ou réactifs avec moi. Merci spécial à Laura Guiguettaz, la meilleure ingé qui soit ; Anissa Guillemin, qui passe probablement plus de temps sur ma chaise de bureau que moi ces derniers temps ; Philippe Mangeot, pour ses nombreux plasmides et idées, ainsi que les soirées jeux de société. Tous mes encouragements aux jeunes thésards : Amandine, Loane, Makram, Nicolas, Vinciane, Emmanuel, Sahra, Dana, Megha et tous ceux que j'ai pu oublier.

Merci aux personnes m'ayant accompagnées ces dernières années. Lorraine, il y a intérêt que tu m'emmènes dans tes bagages la prochaine fois que tu pars au Japon. Je suis fière de toi pour ce tour du monde en quasi solo. Lucie je n'ai qu'une hâte, c'est d'enfin partir à Londres avec toi ! Toulouse à de la chance de t'avoir. Merci à mes anciens camarades lyonnais m'ayant lâchement abandonné pour la capitale et la Normandie. Lois et Marine, je vous souhaite plein de bonheur et de réussite pour le futur. Marie bravo d'avance pour ton TOIEC, je sais que tu vas l'avoir. Je t'enverrai des grosses chaussettes pour que l'hiver soit moins dur à Barfleur. Hugo, le vrai Batman, celui que l'on ne mérite pas mais dont nous avons urgemment besoin. Merci d'être mon confident depuis toutes ces années. Merci à tous les autres, Ivan le terrible, Margaux la coiffeuse et ceux qui se reconnaîtront.

Je remercie les membres de ma famille qui ont eu la patience et le courage de me supporter depuis toutes ces années. Votre aide et votre soutien sont plus que je ne le mérite. Pour finir, last but not least, merci spécial à mon gros Naruto, la boule de poils qui m'aura adorablement tenue compagnie durant la rédaction de cette thèse.

# Résumé

## Rôle des hélicases DDX5 et DDX17 dans l'expression et la maturation des ARN pré-messagers dans le contexte de pathologies neuro-développementales

L'étape de maturation de l'extrémité 3' des ARNm est cruciale à la fois pour assurer leur libération de la chromatine et leur exportation au cytoplasme, mais permet aussi la production d'isoformes distinctes à travers la polyadénylation alternative. La régulation de cette étape est particulièrement importante dans les cellules neuronales. Son altération peut entraîner un readthrough transcriptionnel, résultant en la production de transcrits à l'extrémité 3' anormalement allongée, voire de transcrits chimériques. Certains de ces transcrits sont associés à la tumorigenèse, ainsi qu'à d'autres pathologies.

Les hélicases DDX5 et DDX17 sont impliquées dans quasiment toutes les étapes du métabolisme des ARNs. Elles participent notamment à la régulation de l'épissage et de la maturation en 3' des ARNm, leur déplétion induisant le readthrough transcriptionnel d'environ 800 gènes dans une lignée cellulaire de neuroblastome. L'analyse RNA-seq de 498 tumeurs de neuroblastome montre que l'expression de *DDX17* est fortement liée au pronostic des patients. Le neuroblastome est un cancer pédiatrique du système nerveux périphérique, dont les groupes à haut risque sont souvent associés à l'amplification de l'oncogène *MYCN*. Par ailleurs, des mutations hétérozygotes de *DDX17* ont été identifiées chez plusieurs individus présentant des troubles neurodéveloppementaux, suggérant son rôle dans le développement du système nerveux.

Mon travail de thèse a porté sur la caractérisation de l'effet d'une baisse d'expression des hélicases DDX17 et DDX5 dans des lignées cellulaires neuroblastiques. Plus spécifiquement, mon travail de thèse montre que la déplétion de DDX5 et DDX17 entraîne la production de transcrits chimériques dans les lignées cellulaires de neuroblastomes, dont une partie est aussi identifiée dans des tumeurs de patients. De plus, l'expression de *MYCN* apparaît favoriser le readthrough transcriptionnel de plusieurs gènes régulés par les hélicases, possiblement via sa fixation au niveau du site de terminaison des gènes impactés. D'autre part le travail collaboratif auquel j'ai participé montre que la baisse d'expression de DDX17 entraîne des troubles neurodéveloppementaux chez plusieurs modèles animaux, et j'ai pu confirmer dans les cellules humaines que la déplétion de DDX17 perturbe l'expression de gènes aux fonctions associées au neurodéveloppement. La caractérisation encore préliminaire de certaines mutations *DDX17* faux-sens suggère par ailleurs que celles-ci affectent la capacité de DDX17 à se lier à l'ARN, ainsi que sa solubilité.

Ces résultats suggèrent que DDX17 joue un rôle dans le développement neuronal, et que l'altération de son expression et de sa fonction pourrait favoriser certaines maladies liées au développement neuronal comme le neuroblastome et des troubles neurodéveloppementaux.

Mots clés : DDX5/17, polyadénylation, maturation des transcrits, neuroblastome, *MYCN*

# Abstract

## **Role of DDX5 and DDX17 helicases in pre-mRNA expression and processing in the context of neurodevelopmental diseases**

The processing of the 3' end of mRNAs is crucial both for their release from chromatin and their export to the cytoplasm, but also for the production of distinct isoforms through alternative polyadenylation. The regulation of this step is particularly important in neuronal cells. Its alteration can lead to transcriptional readthrough, resulting in the production of transcripts with abnormally elongated 3' ends, or even chimeric transcripts. Expression of some of these transcripts is associated with tumorigenesis as well as other diseases.

DDX5 and DDX17 helicases are involved in virtually all aspects of RNA metabolism. In particular, they are involved in the regulation of mRNA splicing and 3' processing, their silencing induces transcriptional readthrough of around 800 genes in a neuroblastoma cell line. RNA-seq analysis of 498 neuroblastoma tumors shows that *DDX17* expression is strongly linked to patient prognosis. Neuroblastoma is a pediatric cancer of the peripheral nervous system, in which high-risk groups are often associated with amplification of the *MYCN* oncogene. In addition, heterozygous mutations of *DDX17* have been identified in several individuals with neurodevelopmental disorders, suggesting its role in nervous system development. My thesis work focused on characterizing the effect of a decreased expression of the DDX17 and DDX5 helicases in neuroblastoma cell lines. More specifically, my thesis work shows that the silencing of *DDX5* and *DDX17* results in the production of chimeric transcripts in neuroblastoma cell lines, some of which are also identified in patient tumors. In addition, *MYCN* expression appears to promote the transcriptional readthrough of several helicase-regulated genes, possibly through its binding to the termination site of the affected genes. On the other hand, the collaborative work in which I took part shows that reduced DDX17 expression leads to neurodevelopmental disorders in several animal models, and I was able to confirm in human cells that DDX17 depletion disrupts the expression of genes with functions associated with neurodevelopment. The characterization of certain *DDX17* missense mutations also suggests that these affect DDX17's ability to bind RNA, as well as its solubility.

These results suggest that DDX17 is involved in neuronal development, and that alteration of its expression and function could favor certain diseases linked to neuronal development, such as neuroblastoma and neurodevelopmental disorders.

Keywords: DDX5/17, polyadenylation, mRNA processing, neuroblastoma, MYCN

## Table des matières

Liste des abréviations .....	1
Liste des figures et tableaux .....	3
Introduction .....	5
1 La transcription .....	6
2 L'épissage des ARN messenger .....	7
3 Maturation en 3' des ARN pre-messagers .....	10
3.1 Mécanisme de maturation de l'extrémité 3' des ARN pré-messenger .....	10
3.2 La polyadénylation alternative .....	12
3.3 Régulation de la polyadénylation alternative.....	14
3.3.1 Expression des facteurs de clivage et de polyadénylation.....	14
3.3.2 Régulation de la polyadénylation par les protéines de liaison à l'ARN .	15
3.3.3 Régulation par la cinétique de la Pol II.....	16
4 Terminaison de la transcription .....	17
4.1 Mécanisme de la terminaison de la transcription.....	17
4.1.1 Modèle allostérique .....	17
4.1.2 Modèle torpédo .....	18
4.1.3 Modèle unifié .....	18
4.2 Le readthrough transcriptionnel .....	19
4.2.1 Le readthrough transcriptionnel induit par des stress .....	19
4.2.2 Le readthrough transcriptionnel en contexte pathologique .....	22
4.2.3 Readthrough transcriptionnel et épissage des ARNm .....	26
4.2.4 Readthrough transcriptionnel et organisation 3D du génome .....	27
4.2.5 Conséquences du readthrough transcriptionnel .....	29
4.3 Transcrits chimériques .....	30
4.3.1 Les transcrits chimériques issus de <i>trans</i> -épissage .....	31



4.3.2	Les transcrits chimériques issus de l'épissage entre gènes adjacents ( <i>cis-SAGE</i> )	33
4.3.3	Identification des transcrits chimériques.....	35
4.3.4	Transcrits chimériques et cancer .....	36
5	Hélicases à ARN DDX5 et DDX17 .....	38
5.1	Généralité sur les hélicases à ARN .....	38
5.2	Structure des hélicases DDX5 et DDX17.....	39
5.3	Fonctions moléculaires des hélicases DDX5 et DDX17.....	42
5.3.1	Régulation de l'initiation de la transcription .....	42
5.3.2	Rôle de DDX5 et DDX17 dans l'épissage .....	43
5.3.3	Rôle de DDX5 et DDX17 dans la maturation en 3' des ARN et la terminaison transcriptionnelle .....	45
5.3.4	Autres fonctions moléculaires .....	54
5.4	Rôle de DDX5 et DDX17 dans la différenciation cellulaire .....	54
5.5	DDX5 et DDX17 en contexte pathologique .....	55
5.5.1	DDX5 et DDX17 dans les infections virales .....	55
5.5.2	Hélicases à ARN dans les maladies génétiques .....	58
5.5.3	DDX5 et DDX17 dans les cancers .....	60
6	Le neuroblastome.....	62
6.1	Epidémiologie.....	62
6.2	La pathogénèse et le développement du neuroblastome .....	63
6.3	Diagnostic et stratification du neuroblastome .....	64
6.4	Altération génétique et génomique du neuroblastome .....	65
6.5	MYCN.....	65
6.5.1	MYCN dans le neuroblastome.....	65
6.5.2	MYCN comme facteur de transcription .....	67
6.5.3	Rôles indépendants de l'initiation de la transcription.....	68

Problématique .....	72
Partie 1.....	74
Partie 2.....	120
Discussion .....	169
Matériels et méthodes .....	187
Annexes.....	188
Bibliographie.....	223

# Liste des abréviations

<b>3'ss</b>	3' Splice Site
<b>3'UTR</b>	3'UnTranslated Region
<b>3D</b>	Tridimensionnel
<b>5'ss</b>	5' Splice Site
<b>ADN</b>	Acide DésoxyriboNucléique
<b>APA</b>	Alternative PolyAdenylation
<b>AR</b>	Androgen Receptor
<b>ARN</b>	Acide RiboNucléique
<b>ARNdb</b>	ARN double brin
<b>ARNInc</b>	ARN long non codant
<b>ARNm</b>	Acide RiboNucléique messenger
<b>ARNnc</b>	ARN non codant
<b>ARNr</b>	Acide RiboNucléique ribosomique
<b>ARNt</b>	Acide RiboNucléique de transfert
<b>ATP</b>	Adénosine TriPhosphate
<b>CF</b>	Cleavage Factor
<b>ChIP</b>	Chromatin immunoprecipitation
<b>cis-SAGe</b>	cis-Splicing between adjacent genes
<b>CPA</b>	Cleavage and PolyAdenylation
<b>CPF</b>	Cleavage and Polyadenylation Factor
<b>CPSF</b>	Cleavage and Polyadenylation Specific factor
<b>CRISPR</b>	Clustered Regularly Interspaced Short Palindromic Repeats
<b>CstF</b>	Cleavage stimulation Factor
<b>CTCF</b>	CCCTC-binding Factor
<b>CTD</b>	Carboxy Terminal Domain
<b>DDX17</b>	DEAD-Box helicase 17
<b>DDX5</b>	DEAD-Box helicase 5
<b>DoG</b>	Downstream of Gene
<b>DSE</b>	Downstream Sequence Element
<b>ER</b>	Estrogen Receptor
<b>hnRNP</b>	Heterogeneous nuclear RiboNucleoProtein
<b>KO</b>	Knock-Out
<b>NB</b>	Neuroblastome
<b>NTP</b>	Nucléoside Tri-Phosphate
<b>pA</b>	polyAdenylation
<b>PABP</b>	PolyA Binding Protein
<b>PAP</b>	PolyA Polymerase
<b>PAS</b>	PolyAdenylation Signal
<b>PCR</b>	Polymerase Chain Reaction
<b>PIC</b>	Pre-Initiation Complex
<b>P-TEFB</b>	Positive Transcription Elongation Factor B
<b>RBP</b>	RNA Binding Protein

<b>RNA-seq</b>	RiboNucleic Acid sequencing
<b>RT</b>	Reverse Transcription
<b>shRNA</b>	Small hairpin RiboNucleic Acid
<b>siRNA</b>	Small interferent RiboNucleic Acid
<b>snRNP</b>	Small nuclear RiboNucleoProtein
<b>TAD</b>	Topologically Associated Domains
<b>TBP</b>	TATA-box-Binding Protein
<b>TSS</b>	Transcription Start Site
<b>TTS</b>	Transcription Termination Site
<b>USE</b>	Upstream Sequence Element

# Liste des figures et tableaux

Figure 1 : Différents modes d'épissage alternatif.....	8
Figure 2 : Séquences et facteurs de régulation de l'épissage.....	8
Figure 3 : Principaux facteurs de maturation de l'extrémité 3' et leurs séquences d'interaction dans l'ARN.....	11
Figure 4 : Types d'isoformes générées par la polyadénylation alternative.....	13
Figure 5 : Impact des protéines de liaison à l'ARN sur la maturation en 3' des pré-ARNm.....	15
Figure 6 : Modèle unifié de la terminaison transcriptionnelle.....	19
Figure 7 : Représentation du readthrough transcriptionnel.....	20
Figure 8 : Mécanismes de modulation de la maturation de l'extrémité 3' cellulaire par le virus de la grippe A (IAV) et le virus herpès simplex de type 1 (HSV-1).....	24
Figure 9 : Conséquence du readthrough transcriptionnel sur l'organisation 3D du génome.....	28
Figure 10 : Mécanisme de production des transcrits chimériques Cis-sage.....	34
Figure 11 : Superfamilles SF1 et SF2 des hélicases.....	38
Figure 12 : Structure des hélicases DDX5 et DDX17.....	39
Figure 13 : Mécanismes moléculaires des hélicases DEAD-box.....	40
Figure 14 : Régulation de l'épissage alternatif par DDX5 et DDX17.....	44
Figure 15 : THOC5/THOC6 favorise l'usage de PAS distaux en recrutant CDK12 et DDX5/17 pour réguler les R-loops et la vitesse d'élongation.....	48
Figure 16 : Modèle de régulation de l'extrémité 3' des ARNm par DDX5 et DDX17.....	51
Figure 17 : Modèle illustrant l'implication des G4 dans les interactions à longue distance de l'ADN liées à la Pol II.....	53
Figure 18 : Développement du neuroblastome à partir des cellules de la crête neurale.....	63
Figure 19 : Mécanismes de résolution des stress associés à la transcription par le biais de MYC/MYCN.....	68
Figure 20 : Localisation des substitutions d'acides aminés dans DDX17.....	165
Figure 21 : La mutation DDX17P203L entraîne une baisse de solubilité de la protéine.....	168

Figure 22 : Modèles de régulation de l'épissage par la modification de la cinétique de la Pol II induite par la liaison de facteur de transcription ..... 177

Figure 23 : Représentation des différents mécanismes par lesquels MYCN pourrait favoriser le readthrough transcriptionnel ..... 179

Tableau 1 : Interactions de DDX5 et DDX17 avec les virus ..... 57

Tableau 2 : Hélicases DEAD-box impliquées dans des maladies génétiques..... 59

Tableau 3 : Expression et mécanisme lié à DDX5 et/ou DDX17 dans différents cancers 61

Tableau 4 : Alignement clustal omega des orthologues de DDX17 ..... 189

# Introduction

## 1 La transcription

Pour qu'un gène soit exprimé, l'information contenue sur une séquence d'acide désoxyribonucléique (ADN) doit être transcrite en molécule d'acide ribonucléique (ARN) par l'ARN polymérase. La transcription peut grossièrement être divisée en trois phases : initiation, élongation et terminaison. L'étape d'initiation a été le focus de nombreux travaux, notamment du fait de son caractère hautement régulé. Néanmoins, l'étape de terminaison joue elle aussi un rôle important dans le contrôle de l'expression des gènes. Pour les gènes codant pour des protéines, cette étape de terminaison transcriptionnelle est intimement liée au processus de clivage et polyadénylation des transcrits, qui seront ensuite exportés vers le cytoplasme. Mes travaux de thèse ont essentiellement porté sur cette étape de terminaison transcriptionnelle et de maturation de l'extrémité 3' des transcrits.

Chez les eucaryotes, la transcription de différents types d'ARN est réalisée par différentes ARN polymérases. Les ARN ribosomique (ARNr) sont transcrits par l'ARN polymérase I (Grummt, 1998), les ARN de transfert (ARNt), ainsi que plusieurs petits ARN nucléaires (snARN) et l'ARNr 5S sont transcrits par l'ARN polymérase III (Willis, 1993), et enfin l'ARN polymérase II (Pol II) catalyse la formation des ARN messager (ARNm), de la plupart des snARN, ainsi que celle des ARN nucléolaires (snoARN), des ARN longs non-codant (ARNlnc), des microARN (miARN) et des petits ARN interférents (siARN).

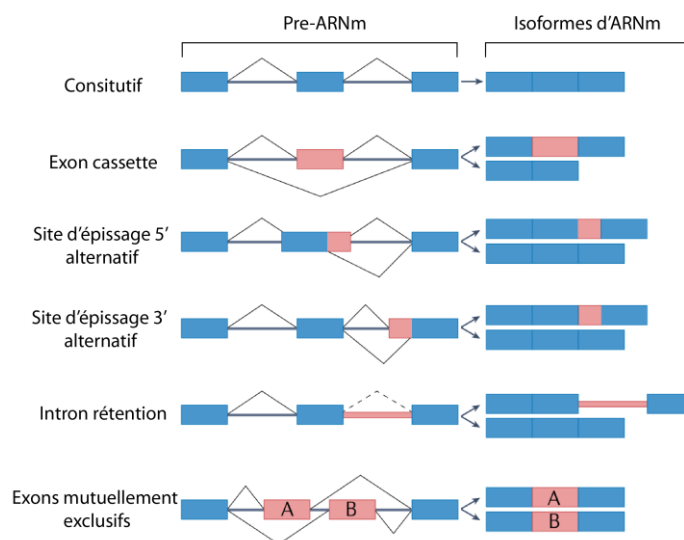
La Pol II est composée de 12 sous-unités RPB (RNA Binding Protein) dont la grande sous-unité RPB1 qui possède un domaine carboxy-terminal (CTD) composé d'un motif d'acides aminés (Tyr1-Ser2-Pro3-Thr4-Ser5-Pro6-Ser7), répété 52 fois chez l'homme, qui subit des modifications, principalement des phosphorylations, au cours de la transcription (Hsin and Manley, 2012). Le CTD joue un rôle important afin de coordonner la transcription et les étapes co-transcriptionnelles de maturation de l'ARN telles que l'ajout d'une coiffe en 5', l'épissage, le clivage et la polyadénylation de l'extrémité 3'. Ces étapes de maturation sont entre autres cruciales pour permettre la stabilité des ARNm, ainsi que leur libération de la chromatine et leur bonne exportation dans le cytoplasme. Une altération de l'épissage ou de la maturation de l'extrémité 3' de l'ARN pré-messager peuvent aussi impacter la séquence de l'ARNm et donc potentiellement celle de la protéine qu'il code, ainsi que sa localisation.



## 2 L'épissage des ARN messager

L'épissage des ARN pré-messagers est une étape indispensable de leur maturation en ARNm, qui permet l'élimination des introns des transcrits et la ligation des introns. L'épissage est assuré par le spliceosome, un complexe ribonucléoprotéique comprenant les petites ribonucléoprotéines (snRNP), résultant de l'association entre les snRNA et des protéines spécifiques. La majorité des ARNm est épissé par le spliceosome majeur, mais certains introns (0.5%) sont pris en charge par le spliceosome mineur (Turunen et al., 2013). Le spliceosome majeur est composé de plus de 200 protéines, comprenant celles formant les snRNP U1, U2, U4/U6, U5. Les composants du spliceosome reconnaissent et se lient à des séquences consensus situées aux extrémités 5' et 3' de chaque intron, correspondant au site d'épissage 5' (5'ss) et au site d'épissage 3' (3'ss), et catalysent deux réactions de transestérification consécutives qui aboutissent à l'excision de l'intron et à la ligature covalente des exons adjacents (Wilkinson et al., 2020).

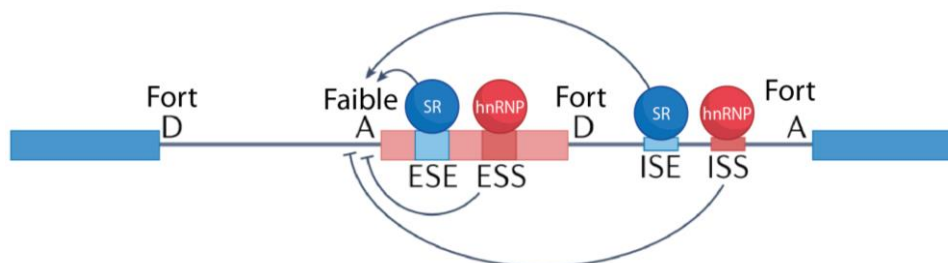
Tous les exons d'un transcrit ne sont pas forcément conservés dans l'ARNm mature final. Si les exons possédant des sites d'épissages forts, c'est-à-dire proche de la séquence consensus, sont généralement épissés de manière constitutive, les exons possédant des sites d'épissage faibles peuvent être épissés de façon différentielle selon les conditions cellulaires. On parle ainsi d'épissage alternatif, phénomène permettant à un gène de coder pour plusieurs transcrits matures différents et donc plusieurs protéines différentes. Un exemple extrême de ce phénomène est le gène *DSCAM* (Down Syndrome Cell Adhesion Molecule) chez la drosophile, qui code pour un récepteur de guidage axonal et dont l'épissage alternatif permet potentiellement de générer plus de 38 000 ARNm différents (Schmucker et al., 2000). Il existe plusieurs modes d'épissage alternatifs (Figure 1).



**Figure 1 : Différents modes d'épissage alternatif**

Épissage constitutif (en haut) et cinq différents modes d'épissage alternatif. L'épissage constitutif a généralement lieu lorsque tous les sites d'épissage sont optimaux. Cependant, il peut également se produire dans des exons dont les sites d'épissage ne sont pas optimaux, mais qui comportent de fortes séquences régulatrices favorisant l'épissage. (Adapté de [Marasco and Kornblihtt, 2023](#))

L'utilisation des sites d'épissage est régulée par des séquences dites *cis*-régulatrices conventionnellement classées en quatre catégories, selon leur position dans l'ARN et leur impact sur l'épissage alternatif : ESE (Exonic Splicing Enhancer), ESS (Exonic Splicing Silencer), ISE (Intronic Splicing Enhancer) et ISS (Intronic Splicing Silencer). Ces séquences recrutent des facteurs d'épissage, dits *trans*-régulateurs (**Figure 2**). Les plus connus sont les facteurs des familles des protéines SR (protéines Ser/Arg-rich) et hnRNP (heterogenous nuclear ribonucleoproteins). Si les protéines SR et hnRNP sont les plus connues, il existe une multitude d'autres facteurs régulant l'épissage. En particulier, des facteurs de régulation tissu-spécifiques tel que NOVA, ESRP ou FOX assurent la mise en place d'évènements d'épissage tissu-spécifiques ([Lee and Rio, 2015](#)).



**Figure 2 : Séquences et facteurs de régulation de l'épissage**

A : accepteur ; D : donneur ; ESE : exonic splicing enhancer ; ESS : exonic splicing silencer ; ISE : intronic splicing enhancer ; ISS : intronic splicing silencer. Adapté de ([Marasco and Kornblihtt, 2023](#))

L'épissage alternatif est aussi affecté par la présence de structures secondaires de l'ARN tel que les structures en tige-boucle. Ces structures peuvent par exemple masquer directement un site d'épissage ou bien masquer ou exposer un site de régulation en *cis* (Buratti and Baralle, 2004; Saldi et al., 2021).

L'environnement transcriptionnel peut aussi réguler l'épissage alternatif. La structure chromatinienne ainsi que la cinétique d'élongation de la Pol II peuvent ainsi influencer l'épissage alternatif (Lee and Rio, 2015). Deux modèles ont été proposés pour expliquer comment l'état transcriptionnel et épigénétique d'un gène donné régule l'épissage : Le modèle cinétique et le modèle de recrutement (Dvinge, 2018). Le "modèle cinétique" stipule que les altérations de la vitesse d'élongation de la Pol II favorisent ou entravent la reconnaissance de sites d'épissage concurrents ou de motifs *cis*-régulateurs. Ce modèle propose qu'une élongation transcriptionnelle lente favorise l'inclusion d'exons alternatifs possédant des séquences régulatrices faibles en permettant une "fenêtre d'opportunité" plus large pour leur reconnaissance par le spliceosome (Mata et al., 2003).

Le "modèle de recrutement" stipule que l'épissage alternatif est modulé par le recrutement direct ou indirect par la machinerie de transcription de facteurs d'épissage agissant sur le pré-ARNm naissant. La Pol II interagit effectivement directement ou indirectement avec des facteurs d'épissage (David et al., 2011; Misteli and Spector, 1999). Ce modèle est étayé par le fait que les protéines SR sont recrutées de manière cotranscriptionnelle sur le pré-ARNm naissant par la Pol II (Das et al., 2007). Le recrutement de nombreux facteurs d'épissage, tels que les snRNP U1 et U2, se fait principalement par l'intermédiaire du CTD de la Pol II (Hsin and Manley, 2012).

Les modèles cinétiques et de recrutement ne sont cependant pas mutuellement exclusifs. Le recrutement de facteurs d'épissage sur le pré-ARNm naissant peut affecter le taux d'élongation de la Pol II. Par exemple, des niveaux réduits de hnRNP A1 et A2 diminuent l'activité du complexe P-TEFb, facteur d'élongation de la transcription (Lemieux et al., 2015). Par ailleurs, la liaison de la protéine SRSF2 au pré-ARNm nouvellement transcrit peut entraîner la libération de la Pol II bloquée au niveau du promoteur (Xiong Ji et al., 2013).

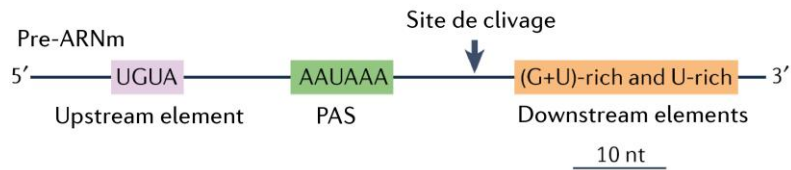
## 3 Maturation en 3' des ARN pre-messagers

### 3.1 Mécanisme de maturation de l'extrémité 3' des ARN pré-messager

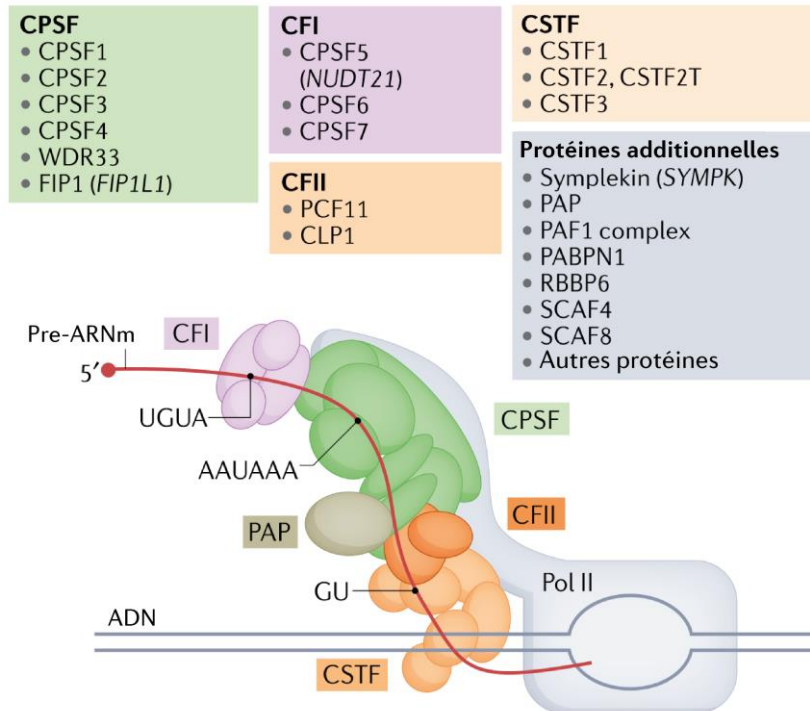
Chez les eucaryotes, la maturation de l'extrémité 3' des transcrits synthétisés par la Pol II consiste en le clivage de ce transcrit et en l'ajout d'une queue poly(A) composée d'environ 200 adénosines à l'extrémité 3' clivée. Le clivage est médié par le complexe multi-protéique CPSF (Cleavage and Polyadenylation Specificity factor), qui reconnaît la séquence du signal de polyadénylation (PAS), généralement AAUAAA, via les sous-unités CPSF30 et WDR33 (Chan et al., 2014) (Figure 3). La reconnaissance de la séquence ARN du PAS activerait l'activité endonucléase du facteur CPSF73, qui clive le transcrit environ 10-30 nucléotides en aval du PAS (Tian and Graber, 2012), généralement juste après le dinucléotide CA. Après clivage du transcrit, la poly(A) polymérase (PAP) ajoute une queue polyA à son extrémité 3' libre. Cette queue polyA est stabilisée via la liaison avec la protéine PABPN1. Le PAS est entouré de plusieurs motifs *cis*-régulateurs qui sont reconnus par des facteurs de clivage, qui régulent l'activité de CPSF et sont nécessaires à la maturation en 3' des transcrits. Ces facteurs incluent CStF (Cleavage stimulation factor), qui se lie aux éléments riches en G/U situés en aval du site de clivage (Pérez Cañadillas and Varani, 2003), ainsi que CFIm (mammalian cleavage factor II), qui se fixe à un élément riche en G qui se trouve sur certains ARNm, en aval du motif G/U (Schäfer et al., 2018). En plus de ces facteurs nécessaires au processus de polyadénylation existent aussi des facteurs non-essentiels, mais pouvant réguler la maturation en 3' des ARNm. Le complexe CFIm favorise ainsi l'usage de PAS possédant des motifs UGUA en amont, en se liant à ceux-ci et en recrutant CPSF (Zhu et al., 2018).

A noter qu'une exception existe à ce mode de maturation de l'extrémité 3' des ARNm : les gènes d'histones réplication-dépendants des métazoaires. Les transcrits de ces gènes ne possèdent pas d'introns et ne sont pas polyadénylés. La maturation de leur extrémité 3' implique un clivage en aval d'une structure tige-boucle conservée liée à la protéine SLBP (stem loop-binding protein) (Dominski and Marzluff, 2007). Ce clivage est réalisé par une machinerie de maturation en 3' spécialisée, contenant notamment la petite ribonucléoprotéine nucléaire U7 (U7 snRNP).

**a Principales séquences régulatrices**



**b Composants du complexe CPA**



**Figure 3 : Principaux facteurs de maturation de l'extrémité 3' et leurs séquences d'interaction dans l'ARN**

(a) Contexte séquentiel des PAS fonctionnels. Le clivage endonucléolytique de l'ARN se produit à environ 20 nucléotides en aval de l'hexamère PAS situé dans un contexte séquentiel approprié contenant le motif UGUA et le motif riche en (G+U) ou en U en amont et en aval, respectivement. Les couleurs des éléments ARN correspondent aux couleurs des complexes qui interagissent avec eux.

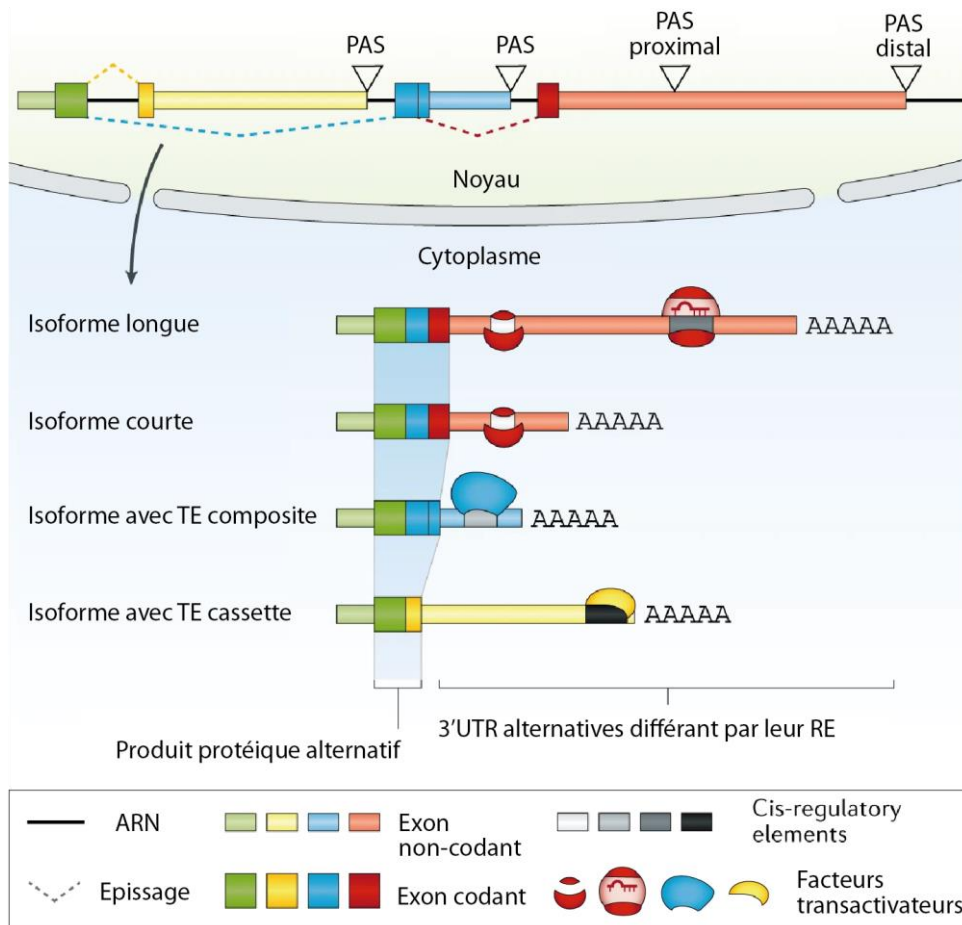
(b) Au cours de la transcription par l'ARN polymérase II (Pol II), la machinerie de clivage et de polyadénylation (CPA) lie le motif hexamère PAS AAUAAA et les éléments *cis*-régulateurs environnants dans l'ARN naissant. Les facteurs CPA représentés sont liés au pré-ARNm (ou ARN naissant) et au domaine C-ter de la Pol II pendant la transcription. Adapté de (Mitschka and Mayr, 2022)

### 3.2 La polyadénylation alternative

Environ 70% des gènes humains codant pour des ARNm possèdent plus d'un site PAS fonctionnel. Par conséquent, le clivage de l'extrémité 3' et la polyadénylation peuvent se produire à plusieurs sites par le biais d'un processus appelé polyadénylation alternative (APA) (Hoque et al., 2013). L'APA permet ainsi de générer des extrémités 3' alternatives à un même ARNm (Figure 4). Si le clivage a lieu avant le codon stop, l'APA peut changer l'identité de la protéine produite. Cela peut en effet induire la production d'une protéine à l'extrémité C-ter tronquée, pouvant affecter sa fonction, voir la rendre non-fonctionnelle. Si le PAS utilisé se trouve après le codon stop, la séquence peptidique de la protéine n'est pas affectée. Cependant, cela modifie la longueur de la région 3' non traduite (3' UTR) de l'ARNm, affectant potentiellement la stabilité et la localisation du transcrit. La séquence 3' UTR des transcrits possède généralement des sites pouvant être ciblés par complémentarité par des miARN, ainsi que des motifs reconnus par des protéines de liaison à l'ARN (RBP), pouvant entraîner la déadénylation ainsi que la dégradation de l'ARNm (Garneau et al., 2007). Les isoformes à extrémité 3' courte de certains oncogènes sont ainsi plus stables que leur équivalent possédant une extrémité 3' UTR longue (Mayr and Bartel, 2009). Toutefois, plusieurs travaux ont pu montrer que le raccourcissement de l'extrémité 3' UTR n'a qu'un effet modeste sur la stabilité générale des transcrits (Gruber et al., 2014; Spies et al., 2013).

Les régions 3'UTR peuvent aussi posséder des éléments permettant la localisation du transcrit, et donc de la protéine produite, à une localisation subcellulaire particulière. C'est notamment le cas de certains transcrits neuronaux, *BDNF* (*Brain-Derived Neurotrophic Factor*) et *Importin  $\beta$ 1* par exemple, dont l'isoforme possédant une extrémité 3' UTR longue est préférentiellement localisée aux neurites, alors que l'isoforme courte réside plutôt dans le corps cellulaire (An et al., 2008; Perry et al., 2012).

Il a de plus été démontré que la séquence 3' UTR peut affecter la localisation de la protéine codée. Dans le cas du gène *CD47*, il a pu être démontré que la région 3' UTR de l'isoforme longue de l'ARNm sert de plate-forme de liaison pour un complexe protéique qui s'assemble de manière co-translationnelle avec la protéine CD47 nouvellement synthétisée et permet sa localisation à la membrane plasmique (Berkovits and Mayr, 2015).



**Figure 4 : Types d'isoformes générées par la polyadénylation alternative**

Le clivage de l'extrémité 3' à un site poly(A) (PAS) distal génère une région 3' non traduite (UTR) plus longue que le clivage à un PAS proximal ; la 3' UTR longue contient des éléments *cis*-régulateurs (REs) supplémentaires qui régulent le devenir du transcrit et/ou de la protéine. Le clivage de l'extrémité 3' peut également avoir lieu au niveau de sites introniques. L'utilisation de PAS introniques situés immédiatement en aval des exons annotés conduit à la formation d'exons terminaux (TE) composites, tandis que l'épissage alternatif et la polyadénylation alternative au niveau des TE alternatifs donnent lieu à des TE cassettes. Les TE composites et les TE cassettes peuvent tous deux modifier la protéine exprimée, en tronquant la séquence codante et/ou en fournissant des extensions carboxy-terminales alternatives de la protéine. (Gruber and Zavalan, 2019)

### 3.3 Régulation de la polyadénylation alternative

#### 3.3.1 Expression des facteurs de clivage et de polyadénylation

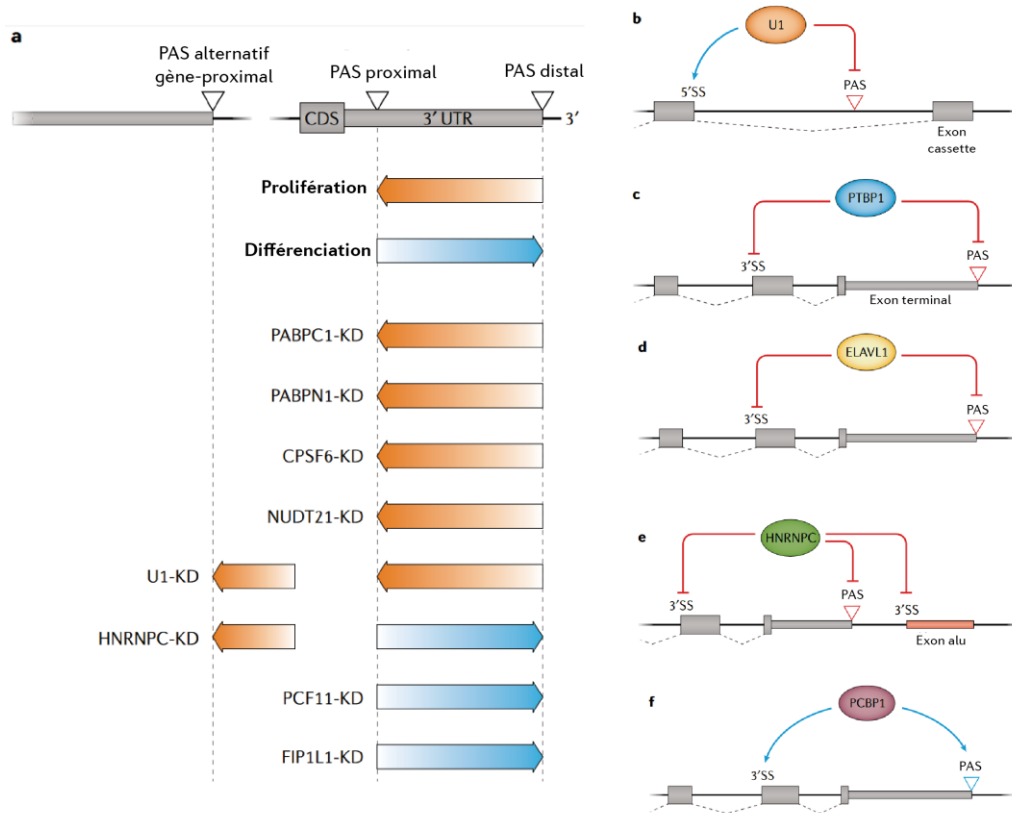
Un important mécanisme de régulation de l'APA est la modulation de l'expression des facteurs faisant partie de la machinerie de polyadénylation. L'augmentation de l'expression des facteurs appartenant au complexe CPA favorise généralement l'usage de sites PAS proximaux, qui apparaissent cinétiquement en premier mais sont souvent sous-optimaux. Cette corrélation inverse entre expression des facteurs de polyadénylation et expression globale des isoformes à 3' UTR longue a notamment été mise en évidence lors de la différenciation et dédifférenciation cellulaire (Sandberg et al., 2008). Cependant, les facteurs de clivage et de polyadénylation peuvent affecter spécifiquement la sélection de certains PAS, malgré leur implication générale dans la terminaison de la transcription de la plupart des gènes. Par exemple, au cours de la différenciation des lymphocytes B, des niveaux élevés de CStF64 favorisent le basculement de l'utilisation du PAS distal vers le PAS proximal du gène codant la chaîne lourde de l'IgM (Takagaki et al., 1996), ce qui conduit à la production d'anticorps IgM dont le domaine trans-membranaire est absent, ce qui permet à l'anticorps d'être sécrété. Les PAS proximaux régulés par CStF64 présentent un environnement nucléotidique inhabituel, enrichi en adénosines en amont du PAS et contenant le motif atypique de liaison à CStF64 GUKKU (où K = G ou U) en aval du PAS (Hwang et al., 2016). FIP1L1 et Pcf11 sont d'autres facteurs de la machinerie de polyadénylation favorisant le clivage au niveau des sites proximaux (Lackford et al., 2014; Li et al., 2015).

On peut noter que si le processus de CPA est contrôlé de manière co-transcriptionnelle, il peut encore être modulé de façon post-transcriptionnelle. Certains gènes apparaissent ainsi subir une CPA séquentielle. Le transcrit est en premier lieu clivé au niveau de son PAS distal, généralement plus fort, mais en raison d'une rétention d'introns non épissés, l'ARNm est retenu dans le noyau. Le clivage de l'ARNm au niveau de son PAS proximal ainsi qu'un épissage post-transcriptionnel supplémentaire permettraient ensuite l'export de l'ARNm dans le cytoplasme (Tang et al., 2022).



### 3.3.2 Régulation de la polyadénylation par les protéines de liaison à l'ARN

En plus de l'expression des facteurs de clivage et de polyadénylation, le choix du site de polyadénylation est aussi affecté par l'expression de protéines de liaison à l'ARN (RBP pour RNA binding protein) (Figure 5).



**Figure 5 : Impact des protéines de liaison à l'ARN sur la maturation en 3' des pré-ARNm**

**(a)** Tendances moyennes de l'augmentation de l'utilisation des sites de polyadénylation (PAS) distaux (flèches bleues) ou proximaux (flèches orange) lors de changements physiologiques de l'état cellulaire, ainsi que lors du knockdown (KD) de certaines protéines de liaison à l'ARN (RBP) et de U1snRNP. **(b-f)** Les effets répressifs (lignes rouges) et/ou activateurs (flèches bleues) des RBP sur l'utilisation du site d'épissage 5' (5'SS), le site d'épissage (3'SS) 3' et le PAS sont indiqués. Adapté de (Gruber and Zavolan, 2019).

Parmi les RBP pouvant réguler l'APA se trouve notamment des facteurs d'épissage. Par exemple, la liaison du composants du spliceosome U1 snRNP inhibe fortement l'utilisation des PAS introniques, favorisant l'usage de PAS plus distaux (Kaida et al., 2010). En plus des composants du spliceosome, de nombreux facteurs régulant l'épissage peuvent moduler l'APA. On peut citer les facteurs SRSF7, SRSF3, hnRNPC (Fischl et al., 2019), la RBP Nova (Licatalosi et al., 2008), PTBP1 (Gruber et al., 2018; Lou et al., 1996), PCBP1 (Xinjun Ji et al., 2013), ainsi que plusieurs RBP de la famille ELAV/Hu (Carrasco et al., 2020; Oktaba et al., 2015; Wei et al., 2020). Plusieurs facteurs d'épissage interagissent directement avec la machinerie de clivage et polyadénylation, ce qui peut suggérer que ces facteurs pourraient coopérer. Par ailleurs, considérant que les facteurs d'épissage et facteurs de clivage et de polyadénylation interagissent directement avec la Pol II, ils pourraient rentrer en compétition pour la Pol II en élongation. L'épissage et la polyadénylation sont tous les deux des étapes de maturation co-transcriptionnelle de l'ARN et de nombreux éléments viennent démontrer l'interconnexion existant entre les différentes étapes de la transcription. D'autres exemples de l'interconnexion entre la maturation en 3' des transcrits et l'épissage sont ainsi présentés plus loin dans la partie **4.2.3.**

### 3.3.3 Régulation par la cinétique de la Pol II

Tout comme l'épissage alternatif, l'APA, du fait de son aspect co-transcriptionnel, est affectée par la cinétique d'élongation de la Pol II. Plusieurs études montrent ainsi que l'augmentation de la pause de la Pol II à proximité d'un PAS fonctionnel favorise le clivage du transcrit au niveau de celui-ci et contribue de ce fait à l'APA (Gromak et al., 2006).

La cinétique de la Pol II est notamment affectée par la présence de structures secondaires de l'ARN et de l'ADN, les modifications de la chromatine, ainsi que l'état de phosphorylation de son CTD et des facteurs d'élongation associés (Muniz et al., 2021). L'effet de ces différents facteurs sur la cinétique de la Pol II permet aussi de passer d'une activité d'élongation transcriptionnelle à une activité de terminaison transcriptionnelle, étape fonctionnellement couplée à la maturation de l'extrémité 3' des ARNm.

## 4 Terminaison de la transcription

### 4.1 Mécanisme de la terminaison de la transcription

La maturation de l'extrémité 3' des transcrits est intimement liée à la terminaison de la transcription. En effet une perturbation de l'étape de clivage de l'extrémité 3', induite par exemple par une mutation du PAS, conduit à une terminaison défectueuse, résultant en un *readthrough* transcriptionnel, où le transcrit n'est pas clivé au niveau du site de clivage annoté et est allongé par la Pol II, de parfois plusieurs kilobases (kb) en aval (Connelly and Manley, 1988). Deux modèles, non mutuellement exclusifs, sont proposés pour expliquer comment la reconnaissance du PAS entraîne la terminaison transcriptionnelle ; le modèle allostérique et le modèle torpédo.

#### 4.1.1 Modèle allostérique

Le modèle allostérique postule que la reconnaissance d'un PAS fonctionnel par la Pol II entraînerait un changement conformationnel de celle-ci. Cela résulterait en une pause de la Pol II suivie par sa dissociation de la matrice d'ADN. Ce changement conformationnel serait médié par le complexe PNUITS-PP1, dont le recrutement entraînerait la déphosphorylation de la région C-terminale du facteur d'élongation transcriptionnelle SPT5, réduisant la vitesse de la Pol II (Cortazar et al., 2019). Ce modèle est appuyé par la démonstration que la transcription du PAS suffirait pour induire un changement conformationnel du complexe indépendamment du clivage de l'extrémité 3' du pre-ARNm (Zhang et al., 2015), même si comme indiqué plus bas, ces résultats ont été obtenus *in vitro* et peuvent ne représenter que partiellement le mécanisme à l'œuvre dans les cellules. On peut aussi ajouter que la phosphatase PP2A antagonise l'élongation en déphosphorylant le domaine C-ter de la Pol II et SPT5 (Huang et al., 2020). Cependant, PP2A est recrutée par le complexe Integrator près des promoteurs de gènes et agit sur des résidus distincts de ceux de PNUITS-PP1. La déphosphorylation par PP2A semble donc plutôt associée à la terminaison transcriptionnelle prématurée. On peut noter que la pause ou le ralentissement de la transcription est une caractéristique de la terminaison transcriptionnelle conservée au cours de l'évolution. Par exemple, chez les procaryotes, la terminaison intrinsèque repose sur des structures primaires et secondaires au sein de l'ARN pour entraîner la pause de l'ARN polymérase. Ainsi la présence d'une boucle GC-riche au niveau de l'ARN suivie d'une séquence polyU augmente l'instabilité des hybrides ARN:ADN qui sont formés par complémentarité du

brin d'ADN et du transcrit naissant (Ray-Soni et al., 2016). Cela pourrait à la fois induire une pause élémentaire et provoquer le changement de conformation initial de l'ARN polymérase bactérienne nécessaire à la terminaison en relâchant les contacts protéine-acide nucléique dans le complexe d'élongation et en rendant les complexes susceptibles d'être inactivés et dissociés par la suite.

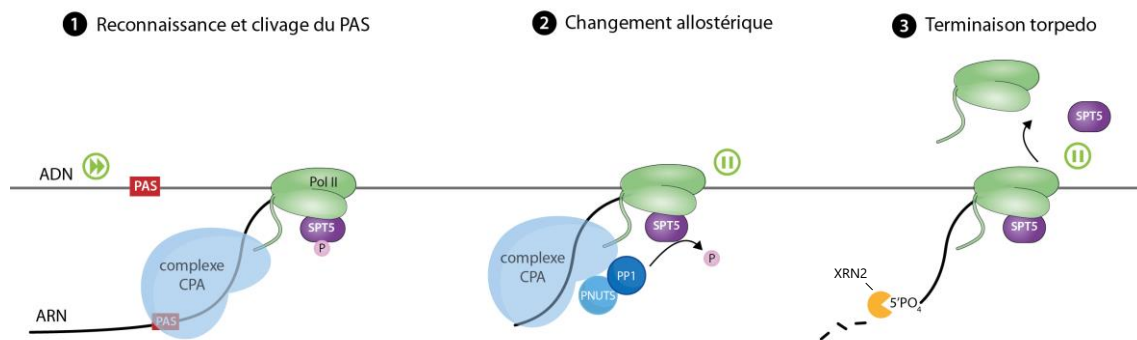
#### 4.1.2 Modèle torpédo

Selon le modèle de terminaison torpédo, le clivage du transcrit naissant fournit une extrémité 5'-PO<sub>4</sub> à l'exonucléase Xrn2 qui dégrade le fragment clivé de son extrémité 5' à 3', repoussant la Pol II et la délogeant finalement de la matrice d'ADN (West et al., 2004). Coupler le clivage avec le processus de terminaison transcriptionnelle permettrait notamment d'éviter de libérer la Pol II de transcrits non matures potentiellement dysfonctionnels. De manière cohérente avec ce modèle, la déplétion de Xrn2 par un système de dégron induit une perturbation globale de la terminaison transcriptionnelle des gènes codants pour des transcrits polyadénylés (Eaton et al., 2018). La déplétion du facteur de clivage CPSF73 induit elle aussi une perturbation de la terminaison se traduisant par un readthrough transcriptionnel et cela de manière plus importante que la déplétion de Xrn2. De plus l'expression de CPSF73 mutée sur son site d'activité entraîne une accumulation de transcrits présentant un readthrough transcriptionnel, même en présence de CPSF73 sauvage. Cela semble confirmer que le clivage du transcrit par CPSF73 est nécessaire à la terminaison transcriptionnelle. Pourtant, certaines études *in vitro* semblent indiquer que le clivage du transcrit n'est pas absolument nécessaire pour la terminaison transcriptionnelle (Zhang et al., 2015). Cette différence pourrait s'expliquer par la présence de facteurs cellulaires supplémentaires manquant dans les études *in vitro*.

#### 4.1.3 Modèle unifié

Les modèles allostérique et torpédo ne sont pas mutuellement exclusifs et sont en réalité compatibles. Dans un modèle unifié, la reconnaissance du PAS va à la fois entraîner le clivage du transcrit via l'activation de la machinerie de clivage/polyadénylation et le ralentissement de la Pol II via la déphosphorylation de SPT5 par le complexe PNUTS-PP1. Le clivage du transcrit permet l'action de Xrn2 en lui

donnant une extrémité 3' libre et le ralentissement de la Pol II permet à Xrn2 de « rattraper » celle-ci pour la déloger de la matrice d'ADN (Cortazar et al., 2019) (Figure 6).



**Figure 6 : Modèle unifié de la terminaison transcriptionnelle**

(1) Le complexe reconnaît le PAS fonctionnel et clive le transcrit (2) La reconnaissance du PAS induit par ailleurs la déphosphorylation de SPT5 par le complexe PNUTS-PP1, induisant le ralentissement de la Pol II (3) L'extrémité 5'PO<sub>4</sub> créée par le clivage du transcrit permet la dégradation du fragment clivé par exonucléase Xrn2, tandis que le ralentissement de la Pol II permettrait à Xrn2 de « rattraper » celle-ci pour la déloger de la matrice d'ADN

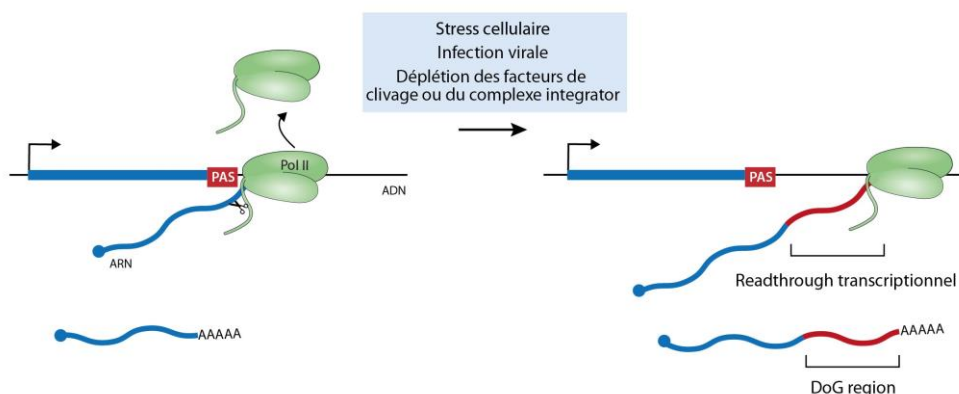
## 4.2 Le readthrough transcriptionnel

Le processus de terminaison transcriptionnelle étant directement lié à la maturation de l'extrémité 3' des transcrits, une perturbation de l'efficacité de celle-ci dû à divers stimuli peut entraîner des défauts de terminaison. L'induction d'un stress osmotique ou thermique ainsi que l'hypoxie induisent ainsi une perturbation de la terminaison transcriptionnelle, tout comme des infections virales et certains cancers.

### 4.2.1 Le readthrough transcriptionnel induit par des stress

Il a été rapporté en 2015 que l'induction d'un stress osmotique par traitement au KCl entraîne le readthrough transcriptionnel d'environ 10% des gènes codants pour des protéines, conduisant à la production de transcrits allongés (Vilborg et al., 2015). Ces transcrits ont été baptisé DoGs pour « Downstream-of-gene-containing-transcripts ». Les DoGs sont issus de la transcription du gène en amont et partagent le même site d'initiation de la transcription (TSS) que celui-ci, comme montré par l'analyse de Cap-seq,

une méthode permettant d'identifier les TSS des transcrits. De plus l'inhibition du gène en amont entraîne une diminution de l'expression du DoG, et la capture des DoGs par des sondes biotinylées permet la détection de la région DoG et du gène en amont. Cela démontre que les DoGs résultent bien de la transcription du gène en amont et donc d'un readthrough transcriptionnel. Un DoG est ainsi composé d'un segment 5' correspondant au gène en amont, appelé « *DoG-associated* », et d'un segment 3' désigné comme « *DoG-region* » (**Figure 7**). Par définition, les DoGs s'étendent sur au moins 5 kb au-delà du site de terminaison normal de leur gène d'origine. Ils peuvent faire plus de 200 kb mais ont une longueur moyenne de 14 kb. L'augmentation de leur expression est indépendante d'un effet sur l'expression du gène en amont (*DoG-associated*), suggérant que l'expression des DoGs est due à une augmentation du readthrough transcriptionnel et non pas à une augmentation de l'expression du gène *DoG-associated*. L'absence d'effet du traitement au KCl sur la demi-vie de deux DoGs ainsi que l'augmentation de la présence de la Pol II dans la région DoG suite au traitement viennent confirmer que le traitement au KCl n'affecte pas la dégradation des DoGs mais perturbe la terminaison de la Pol II, augmentant le readthrough transcriptionnel.



**Figure 7 : Représentation du readthrough transcriptionnel**

L'induction d'un stress ou la déplétion de facteurs du complexe CPA entraîne un défaut de clivage du transcrit et de la terminaison de la transcription induisant le readthrough transcriptionnel de la Pol II. Ce readthrough transcriptionnel conduit à la production d'un transcrit DoG comportant un segment 3' en aval du TTS annoté du gène, correspondant à région DoG.

Concernant le mécanisme par lequel ces stress induisent la production de DoGs par readthrough transcriptionnel, il a dans un premier temps été suggéré que l'induction de DoGs était médiée par la signalisation calcique via le récepteur inositol 1,4,5-trisphosphate (IP3) lors d'un stress hyperosmotique (Vilborg et al., 2015), mais pas lors d'un choc thermique. Néanmoins, il a été découvert plus tard que la diminution des

niveaux de DoGs observée après inhibition de la signalisation calcique était causée par une diminution générale de la transcription (Hennig et al., 2018). L'induction d'un stress cellulaire semble cependant bien altérer la présence de différents facteurs de terminaison transcriptionnelle au niveau de la chromatine. Par exemple, les analyses d'immunoprécipitation de la chromatine (ChIP) de plusieurs composants de la machinerie de clivage et de polyadénylation, notamment CPSF73, CPSF6 et CPSF1, montrent que leur fixation au niveau des TTS (transcription termination site) de certains gènes diminue en cas de stress hyperosmotique ou de choc thermique (Cardiello et al., 2018; Jalihal et al., 2020). De façon plutôt intrigante, il a été montré que suite à l'induction d'un stress hyperosmotique, CPSF6 ségrège dans des condensats formés par séparation de phase (Jalihal et al., 2020). Cela peut suggérer que cette séquestration de CPSF6 serait responsable de la diminution de sa liaison au TTS d'une partie des gènes, entraînant une altération fonctionnelle du clivage et de la polyadénylation des transcrits.

En plus des facteurs canoniques de clivage et de polyadénylation, le complexe Integrator serait lui aussi impliqué dans la régulation de l'expression des DoGs lors d'un stress cellulaire. Le complexe Integrator joue un rôle important dans de nombreux processus liés aux ARNs, telle la maturation de l'extrémité 3' de différents types d'ARNs non codants comme les snRNAs et certains miARNs viraux, mais aussi la terminaison transcriptionnelle des gènes d'histones dépendants de la réplication, des ARNs enhancer et ARNnc, ainsi que la levée de la pause transcriptionnelle de la Pol II (Kirstein et al., 2021; Mendoza-Figueroa et al., 2020). L'analyse de l'interactome de la Pol II à partir de fraction chromatiniennne de cellules contrôles ou soumises à un stress hyperosmotique a pu révéler une diminution de l'interaction entre la Pol II et plusieurs sous-unités du complexe Integrator dans les conditions de stress (Rosa-Mercado et al., 2021). De plus, la déplétion de la sous unité Int3 du complexe Integrator suffit à induire la production de DoGs (Dasilva et al., 2021; Rosa-Mercado et al., 2021). Les gènes produisant des DoGs suite à la déplétion de Int3 sont communs avec 25% des gènes produisant des DoGs lors d'un stress hyperosmotique. Cela suggère que le complexe Integrator est impliqué au moins partiellement dans la biogénèse des DoGs.

En plus du stress osmotique, la production de DoGs est aussi induite par d'autres stimuli stressants comme le choc thermique (Cardiello et al., 2018; Cugusi et al., 2022), le stress

oxydatif (Vilborg et al., 2017), l'infection par le virus de la grippe A (IAV) (Heinz et al., 2018; Nemeroff et al., 1998) et par le virus herpès simplex (HSV) (Rutkowski et al., 2015).

Une question importante est d'arriver à définir quelles sont les caractéristiques faisant que les gènes produisant des DoGs sont plus enclins à présenter un readthrough transcriptionnel que les autres lors de l'induction de stress. Les gènes produisant des DoGs partagent certaines caractéristiques communes, comme une séquence PAS généralement faible (qui s'éloigne du motif AAUAAA), communes aux DoGs induits par stress (Vilborg et al., 2017) et aussi ceux induits suite à l'infection par HSV-1 (Rutkowski et al., 2015). La région en aval des gènes DoGs associés (régions DoG) présentent une déplétion des motifs répétés AC et GU par rapport aux autres gènes (Vilborg et al., 2017). Cela peut suggérer que les gènes DoGs associés possèdent des sites de clivage et polyadénylation plus faibles, qui sont plus sensibles à l'induction de conditions stringentes comme la séquestration ou la diminution de l'expression de facteurs de clivage par exemple, en cas de stress cellulaire. Les régions DoG sont légèrement enrichies pour les marques d'histones associées à une élongation active de la transcription comme H3K36me3 (histone H3 lysine 36 triméthylé) et H3K79me2 au niveau des 5 premiers kb après leur TTS annoté (Vilborg et al., 2017). Ces marques sont enrichies même en l'absence de stress, montrant que les gènes présentant un readthrough induit par un stress présentent même avant l'induction de ce stress une signature chromatinienne indiquant un état de chromatine ouverte.

#### 4.2.2 Le readthrough transcriptionnel en contexte pathologique

##### 4.2.2.1 *Le readthrough transcriptionnel dans les cancers*

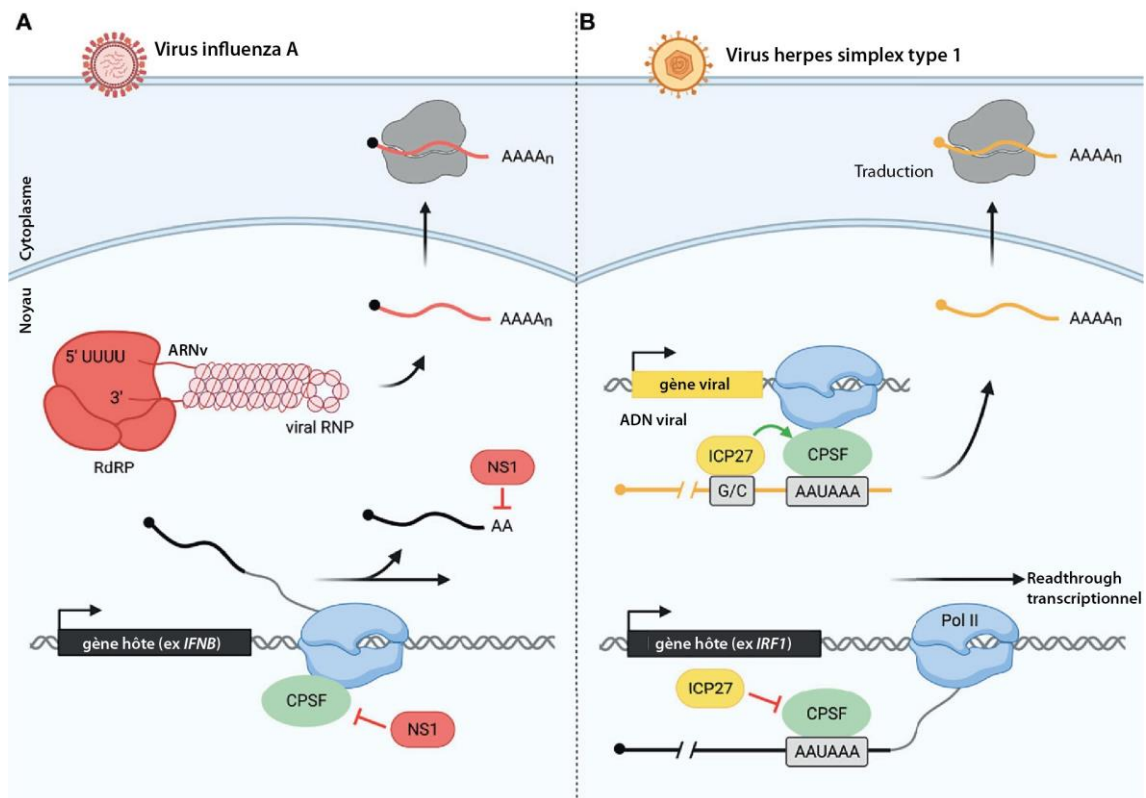
En plus de l'activation des DoGs en réaction au stress cellulaire, les DoGs ont également été associées à des maladies humaines. Des transcrits allongés résultant de readthrough transcriptionnel ont ainsi été observés dans des carcinomes rénaux à cellules claires (ccRCC) (Grosso et al., 2015). Des niveaux élevés de readthrough transcriptionnel sont inversement corrélés à la survie des patients souffrant de ccRCC. L'un des gènes les plus fréquemment mutés dans le ccRCC est SETD2, et sa perte est associée à une diminution spectaculaire de la marque d'histone H3K36me3. La marque H3K36me3 est généralement associée au corps des gènes activement transcrits, et plusieurs études ont



démonstré son implication dans la maturation des ARN via le recrutement de facteurs de maturation comme des facteurs d'épissage (Fahey and Davis, 2017). De manière intéressante, dans le ccRCC, la perte de SETD2 est corrélée à la production de DoGs. Des études récentes chez la drosophile ont également établi un lien entre la marque H3K36me3 et la polyadénylation des ARNs (Meers et al., 2017). Or comme indiqué précédemment, les régions DoG sont légèrement enrichies en H3K36me3, ce qui suggère un mécanisme potentiel pour l'implication de SETD2 dans l'expression des DoG. Toutefois cette hypothèse n'est basée principalement que sur des corrélations et l'on manque de preuves expérimentales, comme l'analyse de l'effet d'une perte de fonction de SETD2 sur la production de DoGs et le dépôt de la marque H3K36me3 afin d'établir un lien direct entre ces trois paramètres.

#### *4.2.2.2 Le readthrough transcriptionnel lors des infections virales*

Le readthrough transcriptionnel est aussi constaté dans plusieurs infections virales (Figure 8). Ainsi, l'infection par HSV-1 entraîne une perturbation de la terminaison transcriptionnelle d'environ 64% des gènes humains (Rutkowski et al., 2015; Wang et al., 2020). Le readthrough transcriptionnel induit par HSV-1 conduit à une importante transcription intergénique allant jusqu'à des dizaines de kb en aval de l'extrémité 3' annotée, jusque dans les gènes situés en aval. La déplétion de la protéine virale ICP27 réduit considérablement le readthrough transcriptionnel des gènes cellulaires, et l'expression ectopique de la protéine est suffisante pour induire le readthrough transcriptionnel de la Pol II (Wang et al., 2020). ICP27 s'associe à plusieurs sous-unités du complexe CPSF, notamment CPSF30, CPSF73, CPSF100, CPSF160 et hFip1. De plus, la protéine symplekin, qui est un composant central de la machinerie de maturation de l'extrémité 3' des transcrits s'associant normalement au complexe CPSF et CstF, semble ne pas interagir avec CPSF lorsque celui-ci est associé à ICP27 (Wang et al., 2020). Ces résultats indiquent que ICP27 peut interférer avec l'assemblage du complexe CPSF, entraînant une perturbation du clivage et de la polyadénylation. La déplétion de ICP27 ne suffit par contre pas à restaurer complètement une terminaison transcriptionnelle normale, suggérant qu'elle n'est pas la seule responsable de l'induction de readthrough transcriptionnel par l'infection de HSV-1 (Wang et al., 2020).



**Figure 8 : Mécanismes de modulation de la maturation de l'extrémité 3' cellulaire par le virus de la grippe A (IAV) et le virus herpès simplex de type 1 (HSV-1)**

(A) IAV utilise la protéine NS1 pour perturber la maturation de l'extrémité 3' cellulaire. L'inhibition du clivage et de la polyadénylation par NS1 entraîne le readthrough de nombreux gènes de l'hôte, dont IFNB. Parallèlement, IAV produit une polymérase ARN dépendante (RdRP) qui lui permet de transcrire et de processer l'ARN viral (ARNv, faisant partie de la ribonucléoprotéine virale vRNP). L'ARNm viral est également polyadénylé par RdRP lorsque la polymérase atteint une séquence 5'-U sur l'ARNv. Ces ARNm viraux sont ensuite exportés vers le cytoplasme pour être traduits. L'ARN et les protéines du virus de la grippe sont colorés en rouge, et l'ARN de l'hôte est coloré en noir. (B) HSV-1 utilise la protéine ICP27 pour perturber la maturation de l'extrémité 3' des ARNm cellulaires et favoriser la maturation de l'extrémité 3' des ARNm viraux. Le CPSF est incapable de s'assembler correctement en présence de l'ICP27. Cela conduit au readthrough de milliers de gènes hôtes, y compris IRF1. Le HSV-1 utilise la machinerie de l'hôte, Pol II et CPSF, pour la transcription virale et la maturation de l'extrémité 3', respectivement. Une séquence riche en GC (G/C) précédant le PAS viral est reconnue par ICP27, ce qui permet un clivage efficace du pré-ARNm viral par CPSF. L'ARN et les protéines du virus de l'herpès sont colorés en jaune, et l'ARN de l'hôte est coloré en noir. (Adapté de Vijayakumar et al., 2022)

Similairement, le virus de la grippe A (IAV) est capable d'inhiber la terminaison transcriptionnelle en interférant avec le clivage de l'extrémité 3' des ARNm. La protéine virale NS1 interagit directement avec CPSF30 pour empêcher le complexe CPSF de se lier au pré-ARNm (Nemeroff et al., 1998). La surexpression de la protéine NS1 conduit en effet à des niveaux plus élevés d'ARNm non clivés, ce qui confirme que NS1 inhibe le clivage de l'extrémité 3' par le complexe CPSF. Cette perturbation du clivage de l'extrémité 3' suite à l'infection par IAV entraîne une terminaison défectueuse de la transcription par la Pol II, ce qui se traduit par la production de longs transcrits readthrough (ou DoGs). Un article récemment publié confirme que l'infection par IAV perturbe globalement la polyadénylation, entraînant un allongement significatif de l'extrémité 3'UTR de 119 gènes cellulaires (Bergant et al., 2023). Les auteurs observent que les transcrits affectés sont bien exportés dans le cytoplasme et donc potentiellement traduits, ce qui peut sembler en contradiction avec les précédents résultats. Il faut noter que dans ce papier les analyses sont menées uniquement sur des ARNs polyadénylés, signifiant que les ARNs n'ayant pas été correctement clivés et polyadénylés sont écartés de l'étude.

Un effet direct de la perturbation de la terminaison transcriptionnelle induite par l'infection virale sur la virulence de ces virus n'a pas encore été clairement établi. Certains éléments indiquent tout de même que l'induction de readthrough affecte globalement l'expression des gènes cellulaires et pourrait perturber la réponse de l'hôte à l'infection. Lors de l'infection par IAV, la majorité des gènes présentant un readthrough voit leur expression diminuée. De plus, dans la souche grippale A/Udorn/72, la liaison du NS1 à CPSF30 inhibe la maturation en 3' de *IFN- $\beta$* , un régulateur clé de la réponse interféron de type I (Noah et al., 2003; Twu et al., 2006). Dans l'infection par HSV-1, le gène présentant le plus haut niveau de readthrough est le facteur de régulation de l'interféron 1 (*IRF1*). *IRF1* est un facteur de transcription de la signalisation interféron qui régule directement l'expression de nombreux gènes antiviraux (Hennig et al., 2018). Le fait que des gènes cruciaux de la régulation de la réponse immunitaire comme *IFN- $\beta$*  et *IRF1* subissent un readthrough transcriptionnel dans les cellules infectées par ces virus suggère que les virus pourraient atténuer la réponse antivirale de l'hôte en altérant la maturation de l'extrémité 3' des transcrits.

#### 4.2.3 Readthrough transcriptionnel et épissage des ARNm

L'épissage et la maturation en 3' des transcrits sont tous les deux des processus co-transcriptionnels et sont moléculairement couplés. Il existe en effet depuis longtemps des preuves d'interactions et d'une influence croisée entre ces deux processus (Baurén et al., 1998). Comme indiqué dans la partie 3.3.2, de nombreux facteurs d'épissages sont aussi impliqués dans la régulation de l'APA.

Le profil d'épissage exact des DoGs n'est pas forcément connu, mais des études menées dans des cellules de levures et de mammifères ont révélé que les pré-ARNm non épissés ont tendance à ne pas être clivés à leur extrémité 3' et qu'ils présentent un readthrough transcriptionnel. Chez *Saccharomyces cerevisiae*, la déplétion de Nab2 entraîne la production de transcrits présentant des défauts d'épissage (Alpert et al., 2020). Or l'analyse de séquençage « *long-read* » indique que ces défauts d'épissage semblent être dûs au readthrough transcriptionnel des gènes en amont. Par ailleurs, dans les cellules de mammifères, le séquençage dit « *long-read* » des ARNs associés à la chromatine a permis de déterminer la séquence/structure complète des pré-ARNm en cours de maturation, ce qui permet de déterminer avec précision leur statut d'épissage et la position de la Pol II le long du gène au moment de la préparation de l'échantillon (Reimer et al., 2021). Cette étude a mis en évidence la nature "tout ou rien" de la maturation des pré-ARNm : Les pré-ARNm complètement épissés sont correctement clivés à leur extrémité 3' et ne présentent pas de readthrough transcriptionnel, tandis que les transcrits qui conservent des introns ont tendance à être clivés de manière inefficace, comme indiqué par leur profil transcriptionnel se poursuivant au-delà du TTS annoté (Reimer et al., 2021). Ce couplage "tout ou rien" résulterait de l'important cross-talk existant entre les mécanismes d'épissage et de traitement de l'extrémité 3'. Comme l'épissage co-transcriptionnel précède généralement le traitement de l'extrémité 3', il a été suggéré que l'élimination des introns pourrait améliorer l'efficacité de la maturation de l'extrémité 3'. L'influence de l'épissage sur le clivage de l'extrémité 3' a été démontrée de manière plus directe dans le même article en étudiant les effets d'une mutation trouvée chez certains patients atteints de  $\beta$ -thalassémie. Une mutation ponctuelle située dans l'intron du gène de la  $\beta$ -globine crée un site d'épissage 3' cryptique qui est utilisé plus efficacement que le site d'épissage 3' canonique. L'efficacité accrue de l'épissage de

l'intron conduit également à une meilleure utilisation du PAS annoté et à une réduction du readthrough transcriptionnel du gène de la  $\beta$ -globine (Reimer et al., 2021).

En outre, un article rapporte qu'une augmentation du readthrough transcriptionnel des gènes sensibles au stress a également été observée suite à l'inhibition du facteur d'épissage SF3B1 par le Pladienolide B (PladB) (Castillo-Guzman et al., 2020). Les gènes produisant des transcrits présentant une rétention de leur intron terminal après le traitement au PladB sont ainsi plus susceptibles de présenter du readthrough transcriptionnel que les autres gènes, ce qui suggère que les perturbations de l'épissage interfèrent avec le clivage de l'extrémité 3' et la terminaison de la transcription dans au moins un sous-ensemble de gènes.

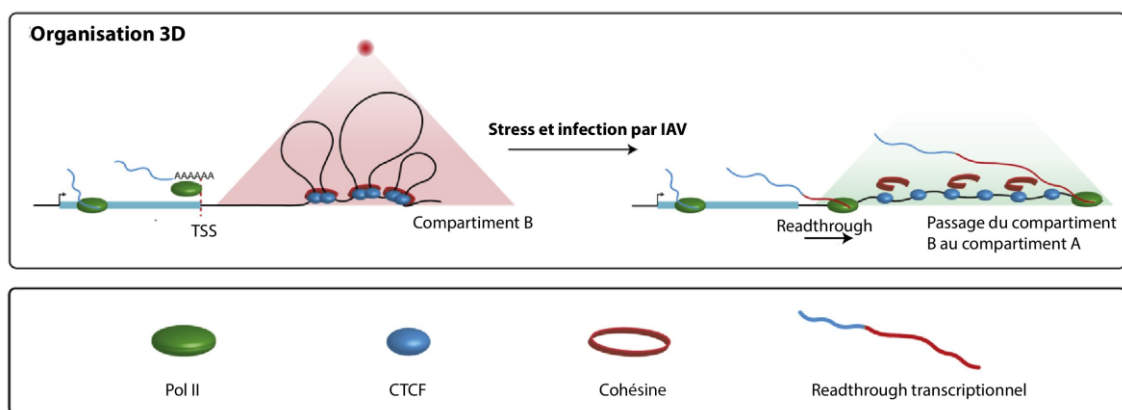
La concomitance d'une perturbation de l'épissage avec la production de DoGs a aussi pu être observée dans plusieurs conditions de stress, y compris en cas de choc thermique et l'infection par le HSV-1. Lors de l'infection par le HSV-1, la perturbation de l'épissage a été principalement observée dans les gènes présentant un read-in transcriptionnel, c'est-à-dire les gènes situés en aval du gène DoG-associé et dont la séquence est transcrite suite au readthrough transcriptionnel (Rutkowski et al., 2015).

#### 4.2.4 Readthrough transcriptionnel et organisation 3D du génome

Le développement de techniques d'étude de l'architecture du génome a permis de démontrer que le génome possède différents niveaux d'organisation (Misteli, 2020). A l'échelle de la mégabase, le génome est ainsi divisé en deux types de compartiments : le compartiment A, caractérisé par une transcription active, et le compartiment B, correspondant à une chromatine transcriptionnellement inactive. A une échelle allant de 0.2 à 1 Mb, le génome est découpé en TADs (Topologically Associated domains) correspondant à des séquences d'interactions préférentielles, dont les bornes sont délimitées par la fixation de CTCF (CCCTC binding Factor). Sur une échelle inférieure, deux régions du génome séparées par plusieurs kb peuvent interagir entre elles et former des boucles chromatiniennes, permettant notamment la mise en contact d'éléments régulateurs.

Plusieurs études semblent montrer que le readthrough transcriptionnel affecte l'organisation 3D du génome, et qu'inversement l'organisation 3D de la chromatine peut

moduler la terminaison transcriptionnelle. Ainsi, l'étude des modifications de la chromatine à la suite d'une infection par l'IAV a révélé que la protéine NS1 de l'IAV induit le readthrough transcriptionnel de gènes très actifs, entraînant le délogement de la cohésine de la chromatine, l'élimination des boucles chromatinienne et la décompaction de la chromatine dans les régions de readthrough (Heinz et al., 2018) (Figure 9). Les auteurs démontrent que la réorganisation de la chromatine est bien une conséquence du readthrough transcriptionnel en utilisant le flavopiridol, un inhibiteur de l'étape d'élongation (Chao and Price, 2001). Le traitement par flavopiridol induit bien une réduction du readthrough ainsi qu'une réduction des modifications chromatinienne normalement associées au readthrough (Heinz et al., 2018).



**Figure 9 : Conséquence du readthrough transcriptionnel sur l'organisation 3D du génome**

L'infection des cellules par le virus influenza A (IAV) conduit à l'inhibition de la terminaison de la transcription entraînant le readthrough transcriptionnel de la Pol II au-delà des sites de terminaison, une diminution des boucles chromatinienne et une décompaction locale de la chromatine. Adapté de (Morgan et al., 2022)

Ces résultats sont cohérents avec plusieurs études montrant que la transcription par la Pol II pourrait affecter l'organisation 3D de la chromatine. Chez la drosophile par exemple, la majorité des limites de TADs coïncide avec les promoteurs actifs plutôt que des sites CTCF (Ramírez et al., 2018) et l'activité transcriptionnelle permet de prédire les limites des TADs avec beaucoup plus de précision que la localisation CTCF et d'autres protéines insulatrices (Ulianov et al., 2016). Chez les mammifères, un sous-ensemble de bornes de

TADs se superpose à des gènes activement transcrits, mais pas à des sites CTCF (Dixon et al., 2012). Par ailleurs, l'inhibition de la transcription n'apparaît pas affecter la formation de compartiments chromatinien, mais entraîne par contre une diminution des contacts entre domaines géniques, correspondant à des régions de contacts enrichis le long du corps des gènes, indiquant que l'activité transcriptionnelle est partiellement responsable de la formation des domaines géniques (Zhang et al., 2021). De plus, des expériences de Hi-C et de CHIP confirmées par simulation suggèrent que la Pol II est une barrière d'extrusion non stationnaire (contrairement à CTCF qui est une barrière d'extrusion stationnaire de la cohésine), qui transloque et relocalise la cohésine (Banigan et al., 2023). L'extrusion de boucles et la transcription pourraient ainsi interagir pour générer et maintenir de manière dynamique les interactions entre les gènes et les éléments régulateurs et façonner l'organisation fonctionnelle du génome.

D'autres travaux montrent que l'organisation 3D de la chromatine affecte la terminaison transcriptionnelle et la maturation en 3' des transcrits. Il a été démontré que la formation de boucles chromatinien médiées par CTCF et la cohésine au niveau des PAS proximaux favorise l'usage de ceux-ci et donc l'expression d'isoformes courtes (Nanavaty et al., 2020). De plus, de récents travaux menés dans mon équipe d'accueil ont montré qu'altérer l'organisation 3D d'un gène peut moduler son readthrough transcriptionnel (Terrone et al., 2022). Ces travaux sont abordés plus en détail dans la section 5.3.3.

Ces résultats montrent la possible existence d'un cross-talk entre l'organisation 3D de la chromatine et le processus de maturation en 3' des transcrits.

#### 4.2.5 Conséquences du readthrough transcriptionnel

Les transcrits issus d'un readthrough transcriptionnel (ou DoGs) constituent des marques indéniables d'une perturbation de la terminaison transcriptionnelle et de la maturation des transcrits. L'étendue des conséquences de ces transcrits allongés sur les processus cellulaires n'est en revanche pas encore bien comprise, notamment à cause de la difficulté à cibler sélectivement ces transcrits. Les DoGs sont généralement retenus dans le noyau, restant associés à la chromatine, ils ne sont donc pas traduits mais ils possèdent une demi-vie similaire aux transcrits normaux (Vilborg et al., 2015). Lorsque

le readthrough d'un gène codant pour une protéine impacte une proportion importante de ses transcrits, on peut supposer que cela affecte la quantité de protéines traduites à partir de ceux-ci. Par exemple, dans l'infection par HSV-1, des expériences de ribosome-profiling ont révélé que la traduction des transcrits du gène *IRF1*, dont >75% présentent un readthrough, diminue de plus de 4 fois 8 heures après l'infection malgré une augmentation du niveau total des transcrits *IRF1*, coïncidant avec une augmentation de 4 à 6 fois de l'enrichissement nucléaire en transcrits *IRF1* après l'infection (Hennig et al., 2018). Ce résultat renforce l'idée que les transcrits allongés ne sont pas disponibles pour la traduction.

La production de DoGs et le readthrough transcriptionnel en général peuvent aussi affecter l'expression des gènes adjacents, plus précisément le gène directement en aval du gène DoG associé. En effet, lors d'un readthrough transcriptionnel, l'élongation transcriptionnelle de la Pol II peut se poursuivre jusqu'au gène situé en aval. Si le gène en aval se trouve en orientation inverse, cela peut potentiellement conduire à de l'interférence transcriptionnelle. Dans le cas où celui-ci se trouve dans la même orientation, il peut être transcrit sur le même brin d'ARN que le DoG. Ce phénomène est parfois appelé « read-in » transcriptionnel. Un même transcrit porte alors l'information de deux gènes différents. Ces transcrits peuvent être épissés en utilisant le site 5' d'épissage d'un exon du gène amont (généralement l'avant dernier exon) et le site 3' d'épissage du gène aval (généralement le deuxième exon). Ce phénomène d'épissage en *cis* d'exons de deux gènes différents permet ainsi la production de transcrits dits chimériques, parfois aussi appelé ARN chimériques *cis*-SAGE pour « *cis*-Splicing between adjacent genes ».

### 4.3 Transcrits chimériques

Par définition, les transcrits chimériques sont des transcrits hybrides qui portent l'information de deux gènes originellement distincts (possédant chacun un TSS et un TTS propre). Les transcrits chimériques étaient traditionnellement considérés comme étant uniquement le résultat de la transcription de gènes fusions issus de réarrangements chromosomiques. Plusieurs de ces gènes fusions se retrouvent dans de nombreux cancers, cela de manière spécifique, permettant d'en faire d'importants biomarqueurs (Sun and Li, 2022). Par ailleurs certains de ces gènes fusions sont des oncogènes jouant



un rôle important dans la tumorigenèse et qui servent de cibles thérapeutiques dans le traitement de ces cancers. Un des exemples les plus connus est la fusion *BCR-ABL*, ou chromosome de Philadelphie, qui est due à la translocation entre les bras longs des chromosomes 9 et 22 (Kurzrock et al., 2003). Cette anomalie chromosomique se retrouve chez environ 95% des patients souffrant de leucémie myéloïde chronique et les inhibiteurs de tyrosine kinase ciblant spécifiquement *BCR-ABL* constituent le traitement de référence de ce cancer.

Toutefois, la multiplication de l'identification de transcrits chimériques, notamment grâce aux technologies de séquençage nouvelle génération, a permis de révéler l'existence de nouvelles catégories de transcrits chimériques n'étant pas dus à la transcription de gènes fusions. Ces transcrits chimériques sont parfois qualifiés de transcrits chimériques non-canoniques. Ainsi, deux autres catégories de transcrits chimériques dont la jonction a lieu au niveau du transcrit et non pas du génome ont été définis, les transcrits chimériques issus de *trans*-épissage ou de *cis*-épissage entre gènes adjacents (*cis*-SAGE).

#### 4.3.1 Les transcrits chimériques issus de *trans*-épissage

La formation de transcrits chimériques issus de *trans*-épissage implique que deux molécules d'ARN différentes soient épissées ensemble. Le *trans*-épissage peut être subdivisé en *trans*-épissage intergénique et intragénique. Les ARNs chimériques issus de *trans*-épissage intergénique peuvent provenir de deux gènes situés sur des chromosomes différents, ou de deux gènes situés sur le même brin ou sur des brins opposés du même chromosome. Le *trans*-épissage intragénique peut être généré entre les transcrits sens et antisens du même gène parental, ou même à partir de différentes copies des mêmes transcrits, et entraîner un remaniement des exons.

Le *trans*-épissage est un processus courant chez plusieurs espèces dites "inférieures", telles que les organismes unicellulaires, les nématodes ou les trypanosomes (Krause and Hirsh, 1987; Sutton and Boothroyd, 1986). Le mécanisme de *trans*-épissage dans ces organismes, spécifiquement connu sous le nom de *trans*-épissage SL (spliced-leader), implique le transfert d'une courte séquence leader de 15 à 50 nucléotides d'une molécule d'ARN non codant spécialisée à l'extrémité 5' des ARNm.

Le *trans*-épissage SL n'a pas été observé chez les humains ainsi que les autres vertébrés, néanmoins, plusieurs exemples de *trans*-épissage intergénique ont pu être identifiés (Lei et al., 2016). Chez l'homme, des exemples de *trans*-épissage intergénique incluent les transcrits chimériques *JAZF1-SUZ12* (Li et al., 2008) et *PAX3-FOXO1* (Yuan et al., 2013) qui ont été trouvés dans des cellules normales ainsi que dans des tumeurs. La différence est que ces transcrits de fusion sont le produit d'un *trans*-épissage dans les cellules normales, alors qu'ils sont le produit de translocations chromosomiques dans les tumeurs.

Le mécanisme de *trans*-épissage entre deux transcrits reste mal identifié. Il existe plusieurs modèles qui tentent d'expliquer ce phénomène (Lei et al., 2016) :

#### 4.3.1.1.1 Modèle de glissement transcriptionnel

Dans ce modèle, certains de ces transcrits chimériques dit « *trans*-épissés » ne seraient pas dûs à un mécanisme d'épissage en *trans* proprement dit mais à un phénomène de « glissement » transcriptionnel. Dans ce modèle, la machinerie transcriptionnelle pourrait se dissocier du gène associé au premier transcrit et venir se fixer sur le deuxième locus pour y reprendre la transcription (Li et al., 2009). Ce mécanisme reposerait sur la présence de courtes séquences homologues (SHS pour short homologous sequence).

#### 4.3.1.1.2 Modèle médié par les facteurs d'action en *trans*

Une étude menée chez le cochon a permis de constater l'existence de quatre séquences consensus présentes dans la région génomique des gènes 5' et 3' (Ma et al., 2012). Ces séquences sont similaires aux sites de fixation de CTCF. Il a ainsi été postulé que ces motifs pourraient être reconnus et recrutés par CTCF à la même machinerie transcriptionnelle. CTCF permettant la formation de boucle chromatinienne, sa fixation pourrait permettre de rapprocher des régions génomiques éloignées, facilitant les événements d'épissage en *trans*. En effet, après la déplétion de CTCF dans les cellules stromales de l'endomètre, l'expression du transcrit chimérique *JJAF1-JJAZ1* diminue (Li et al., 2008). Ce modèle pourrait être compatible avec celui de glissement transcriptionnel. CTCF en rapprochant les régions génomiques pourrait faciliter le glissement de la Pol II du gène en 5' à celui en 3'.

#### 4.3.1.1.3 Modèle de *trans*-épissage médié par l'ARNt

Il a aussi été proposé que deux transcrits distincts soient réunis par l'intermédiaire de séquences d'ARNt fréquentes, qui sont adjacentes aux gènes partenaires, via l'ARNt-intron lyase. Si des expériences *in vitro* indiquent que ce mode de génération de transcrits chimériques *trans*-épissés est possible sur des ARNm, il n'y a pas d'élément indiquant qu'il a bien lieu physiologiquement dans les cellules eucaryotes (Deidda et al., 2003).

#### 4.3.1.1.4 Modèle de *trans*-épissage médié par le spliceosome

Comme pour l'épissage en *cis*, le *trans*-épissage utilise généralement des sites d'épissage canoniques pour relier deux transcrits primaires distincts. Il est donc possible que le *trans*-épissage utilise la même machinerie d'épissage de base que l'épissage en *cis*. Plusieurs études *in vitro* ont démontré que les spliceosomes sont capables de générer des transcrits chimériques *trans*-épissés (Eul et al., 1995; Konarska et al., 1985).

Une question intéressante est de comprendre comment des ARNs transcrits à partir de chromosomes différents peuvent être épissés ensemble. Il est donc supposé que les gènes parents de transcrits *trans*-épissés doivent présenter une certaine proximité spatiale. Néanmoins, le manque d'évènements d'épissage limite la possibilité de vérifier cette hypothèse.

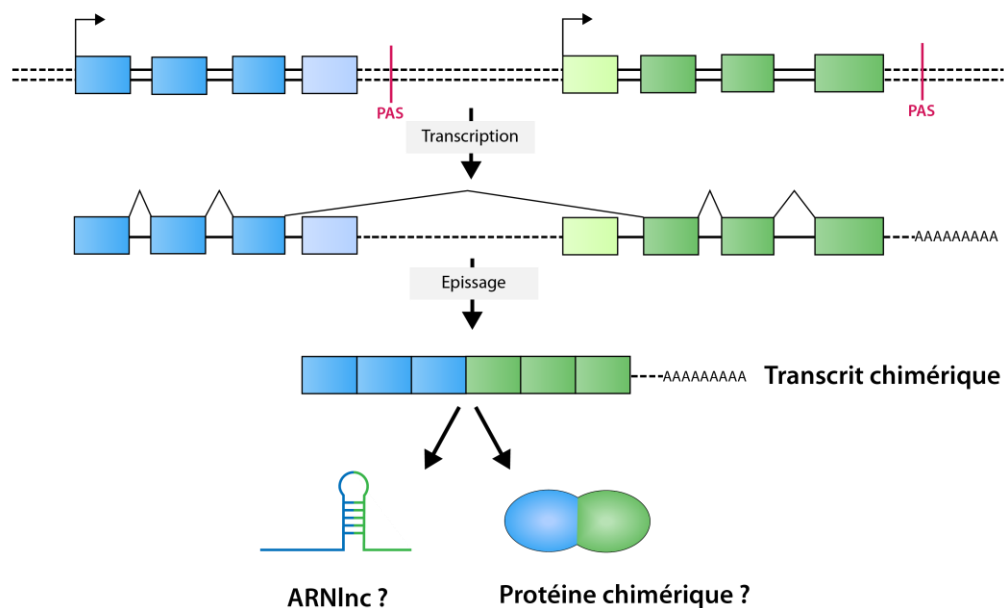
#### 4.3.2 Les transcrits chimériques issus de l'épissage entre gènes adjacents (*cis*-SAGe)

Comme indiqué précédemment, les transcrits chimériques issus d'épissage en *cis* sont produits suite à un évènement de readthrough transcriptionnel entraînant la transcription continue de deux gènes adjacents sur le même brin d'ARN (Figure 10). Ces transcrits sont parfois aussi qualifiés de chimères induites par transcription, ARNs chimériques en tandem, transcrits fusions de gènes adjacents (FuTAGs) et fusions readthrough. Le terme *cis*-SAGe n'explicite pas la nécessité d'un readthrough transcriptionnel pour permettre la production de ces transcrits chimériques, mais en raison de sa prévalence dans la littérature, ce sera celui utilisé dans ce document.

Les transcrits chimériques *cis*-SAGe partagent certaines caractéristiques communes. Le gène parental situé en 5' est activement transcrit et la distance intergénique entre les gènes parentaux est généralement inférieure à 30 kb, à comparer à une distance

intergénique moyenne de 50 kb pour le génome humain (Qin et al., 2015). Le site de jonction de la chimère correspond généralement à la liaison entre le site 5' du dernier intron du gène parental amont et le site 3' du premier intron du gène parental aval, même s'il existe des jonctions chimériques utilisant d'autres sites d'épissage voire parfois des sites cryptiques. Les transcrits *cis*-SAGE peuvent potentiellement être traduits en protéines chimériques, si la jonction préserve le cadre de lecture. Ils peuvent aussi aboutir à la production de protéines tronquées, être candidats à la dégradation des ARNs non-sens ou bien fonctionner comme des ARNs longs non-codants (ARNInc).

Il a été estimé qu'environ 4 % des gènes en tandem, et même 14 % des séquences codant pour des protéines humaines, peuvent contribuer à la génération d'ARNs chimériques *cis*-SAGE, et qu'ils pourraient représenter environ 5 % du transcriptome humain (Akiva et al., 2006; Greger et al., 2014).



**Figure 10 : Mécanisme de production des transcrits chimériques *Cis-sage***

Suite au readthrough transcriptionnel du gène 1, les gènes adjacents 1 et 2 sont transcrits sur le même brin d'ARN. Les exons du gène 1 et du gène 2 seraient épissés entre eux via la liaison d'un site 5' d'épissage du gène 1 et un site 3' d'épissage du gène 2.

### 4.3.3 Identification des transcrits chimériques

L'avènement des techniques de séquençage à haut débit ainsi que la mise à disposition de nombreux jeux de données, notamment provenant de tumeurs, a entraîné la mise au point de nombreux algorithmes et logiciels afin de permettre l'identification de transcrits chimériques. On comptait ainsi au moins 23 différentes méthodes d'identification de transcrits chimériques en 2019 (Haas et al., 2019). La plupart de ces méthodes se basent sur l'identification de paires de *reads* dont l'alignement au génome de référence présente une distance plus importante qu'attendue. Ces *reads* sont ainsi qualifiés de discordants. Pour prédire la position exacte du point de jonction entre les transcrits chimériques, ces outils s'appuient également sur l'identification de "*split-reads*" qui sont des séquences s'alignant à plus d'une localisation sur le génome. Après l'étape de cartographie, ces outils appliquent différents filtres afin d'écarter les possibles biais et de sélectionner les transcrits chimériques les plus fiables. Ces filtres comprennent (a) un filtre d'homologie : le transcrit chimérique est écarté si ses gènes parentaux sont des paralogues l'un de l'autre ou présentent une forte identité de séquence, (b) un filtre basé sur un seuil : un nombre minimum de lectures discordantes et fractionnées est utilisé pour sélectionner les candidats ARNs chimériques à forte probabilité.

Malgré une approche plutôt similaire, les résultats de ces différents outils sont relativement inconsistants. L'analyse de six outils informatiques déterminés comme étant les plus fiables et efficaces par Kumar *et al.* n'a permis de trouver que quatre transcrits chimériques communs en comparant l'analyse de ces outils sur un jeu de données test (Kumar et al., 2016). Cette divergence peut s'expliquer par la non-exhaustivité des différents outils, ainsi qu'à l'identification de faux positifs. En effet malgré les différents filtres appliqués pour la sélection des transcrits chimériques, plusieurs biais peuvent venir artificiellement enrichir la liste de jonctions identifiées. Par exemple, la présence de courtes séquences homologues (SHS) peut favoriser un changement de matrice (*template switching*) durant la préparation de la librairie, si une étape de rétrotranscription est utilisée (Houseley and Tollervey, 2010). Ce changement de matrice ARN au cours de la synthèse du cDNA peut ainsi générer artificiellement des produits « chimériques ».

Il est ainsi important de valider expérimentalement la présence de ces transcrits chimériques afin d'éliminer les artefacts des logiciels d'analyse. La validation peut se faire

par RT-PCR ou d'autres méthodes d'étude des ARNs. Idéalement, le produit PCR devrait être analysé par séquençage Sanger afin de confirmer la jonction.

#### 4.3.4 Transcrits chimériques et cancer

Si les transcrits chimériques issus de réarrangements chromosomiques sont fréquemment retrouvés dans des contextes cancéreux, les transcrits chimériques *cis-SAGe* sont plus rarement associés aux cancers. Néanmoins plusieurs ont pu être observés dans des tissus tumoraux et certains ont été rapportés avec des rôles oncogéniques clairs.

*SLC45A3-ELK4* a été découvert par deux groupes indépendants et constitue un biomarqueur potentiel du cancer de la prostate (Maher et al., 2009; Rickman et al., 2009). Plusieurs fusions *SLC45A3-ELK4* ont été rapportées ; la forme principale est une fusion de l'exon 1 de *SLC45A3* avec les quatre derniers exons d'*ELK4*. Il est intéressant de noter que, bien que *SLC45A3-ELK4* n'est pas traduit et fonctionne comme un lncRNA, sa déplétion dans les cellules cancéreuses entraîne une réduction de la prolifération cellulaire, malgré ses rôles oncogéniques peu clairs (Qin et al., 2017). Il faut toutefois noter que *SLC45A3-ELK4* n'est pas exclusivement retrouvé dans les tissus cancéreux de la prostate et a aussi été identifié dans le tissu artériel (Singh et al., 2020). L'expression de ce transcrit chimérique semble être régulée par la protéine CTCF qui se lie aux insulateurs situés à proximité du promoteur d'*ELK4* (Qin et al., 2016). Plus CTCF se lie aux insulateurs, moins *SLC45A3-ELK4* est exprimé.

Le transcrit chimérique *RBM6-RBM5* est présent dans plusieurs types de cancer. Son expression est par exemple associée à la taille des tumeurs du sein, suggérant que l'expression de *RBM6-RBM5* pourrait être un potentiel biomarqueur (Wang et al., 2007). Un lien direct entre l'expression de ce transcrit chimérique et la tumorigenèse n'a par contre pas été établie.

*DUS4L-BCAP29* est présent dans les cancers de l'estomac et de la prostate et semble jouer un rôle pro-tumoral dans le cancer de l'estomac (Kim et al., 2014; Nacu et al., 2011). Cependant, *DUS4L-BCAP29* est également présent dans les tissus normaux et semble promouvoir à la fois la croissance des tissus sains et cancéreux, et ne peut donc pas constituer un bon biomarqueur de cancer (Tang et al., 2017).

*CTSD-IFITM10* est un transcrit chimérique identifié dans des tumeurs primaires mammaires ainsi que dans des lignées cellulaires de cancer du sein, sans être présent dans les tissus sains (Varley et al., 2014). Il produit une protéine de 60kDa et sa déplétion par siRNA ciblant spécifiquement la jonction chimérique entraîne une diminution de la prolifération des cellules MCF7.

Le transcrit chimérique *ZNF451-BAG2* a été identifié dans plusieurs tumeurs de neuroblastome, et conduit à la production d'une protéine BAG2 tronquée (Shi et al., 2021). L'expression de cette protéine BAG2 tronquée empêche la différenciation des cellules de neuroblastome en réponse à l'acide rétinoïque, un traitement couramment utilisé pour traiter les patients atteints de neuroblastome à haut risque.

Les transcrits chimériques *cis-SAGE* dont l'expression a directement été reliée à la tumorigenèse sont assez rares, mais certaines études montrent une relative abondance de leur expression dans certains cancers. Dans une étude impliquant l'analyse d'échantillons de cancers de la prostate et de tissus non-cancéreux, plus de 300 transcrits chimériques ont été observés, dont 30 % ont été caractérisés comme des *cis-SAGE* (Qin et al., 2015). Une autre étude, portant toujours sur l'analyse RNA-seq d'échantillons de cancer de la prostate, classe 76% des transcrits chimériques identifiés comme des transcrits readthrough dérivant de gènes adjacents, c'est-à-dire *cis-SAGE* (Lai et al., 2015).

La question de savoir si les transcrits chimériques sont simplement une conséquence de la tumorigenèse ou peuvent en être la cause est controversée. Même si de nombreux transcrits chimériques ont été découverts dans plusieurs cancers, il est important de comprendre s'il s'agit réellement d'un facteur déterminant du développement cancéreux. Ainsi, la capacité oncogénique des transcrits chimériques doit être testée *in vitro* à l'aide de méthodes telles que le test de formation de colonies en agar mou, test d'invasion. *In vivo*, des tests de formation de tumeurs peuvent être conduits dans des souris nues (dépourvues de système immunitaire).

## 5 Hélicases à ARN DDX5 et DDX17

### 5.1 Généralité sur les hélicases à ARN

Les hélicases à ARN sont des protéines hautement conservées qui utilisent des nucléosides triphosphates (NTP) pour lier ou remodeler l'ARN, les complexes ARN-protéines ou les deux. Les hélicases à ARN sont présentes chez les bactéries, les archées et les eucaryotes, ainsi que dans de nombreux virus à ADN et à ARN. Chez les eucaryotes, les hélicases à ARN constituent la plus grande classe d'enzymes impliquées dans le métabolisme de l'ARN, avec plus de 70 membres chez l'humain et plus de 40 chez *Saccharomyces cerevisiae* (Bohnsack et al., 2023).

Les hélicases à ARN sont classées en deux superfamilles d'hélicases - la superfamille 1 (SF1) et la SF2 - qui englobent également la plupart des hélicases à ADN eucaryotes (Figure 11). Les deux superfamilles sont à leur tour divisées en familles d'hélicases. La superfamille SF1 inclut les hélicases à ARN de la famille Upf1-like, alors que les familles DEAD-box, DEAH-box/RNA helicase A-like (DEAH/RHA), Ski2-like et RIG-I-like appartiennent à la superfamille SF2. La sous-famille d'hélicases ARN à boîte DEAD constitue la plus grande famille des hélicases ARN, comportant 40 membres chez l'humain (Bourgeois et al., 2016).

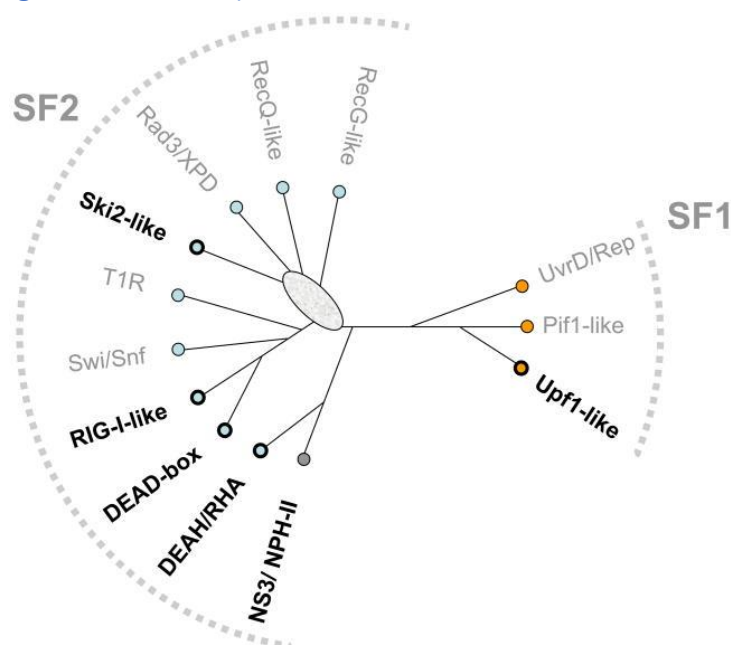


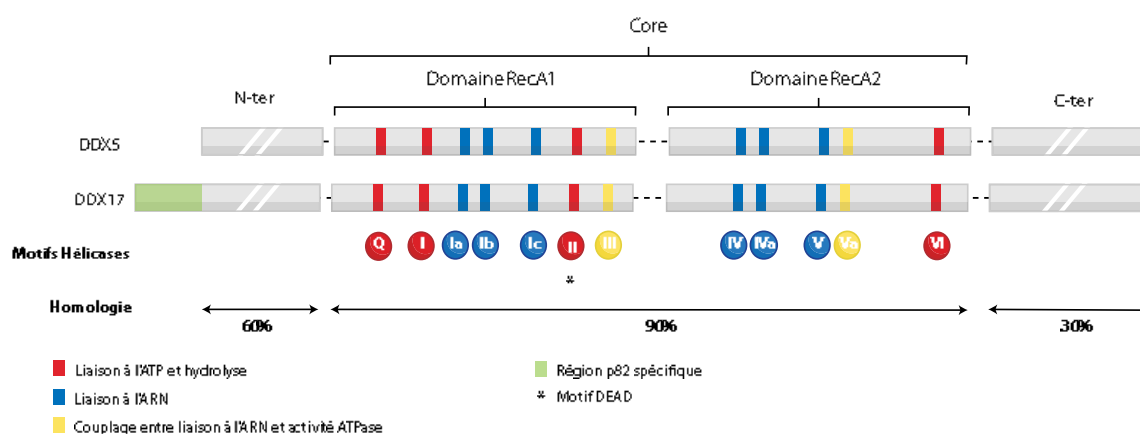
Figure 11 : Superfamilles SF1 et SF2 des hélicases

En gras : les familles contenant des hélicases ARN (Jankowsky, 2011).



## 5.2 Structure des hélicases DDX5 et DDX17

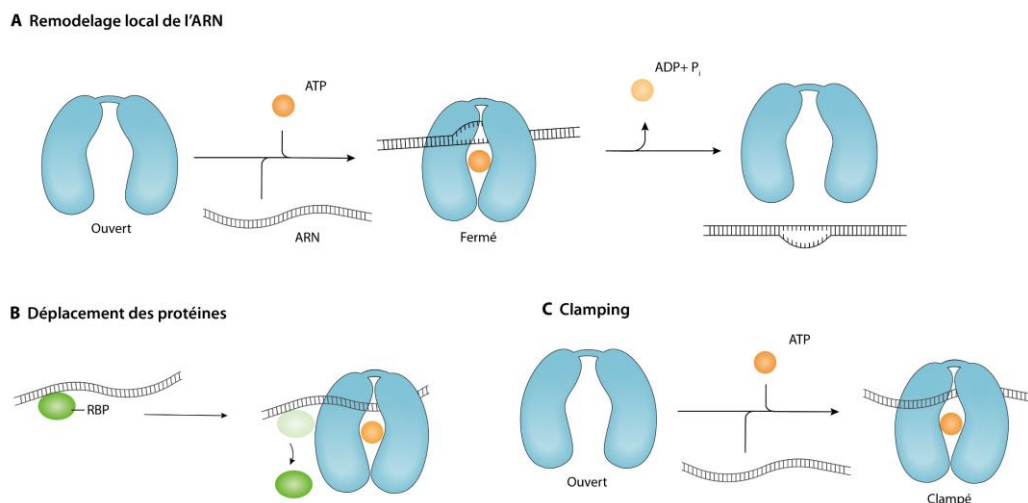
Les hélicases à ARN DDX5 et DDX17 (anciennement nommées p68 et p72) appartiennent à la famille des DEAD-box, comprise dans la superfamille 2 des hélicases à ARN. Comme toutes les hélicases à ARN eucaryotes, elles possèdent un domaine « core hélicase » contenant les sites de liaison à l'ARN et à l'adénosine triphosphate (ATP), mais possèdent en plus le motif DEAD (Asp-Glu-Ala-Asp), caractéristique de la famille DEAD-box. Le « core » est composé de deux domaines RecA-like, qui lient l'ARN et l'ATP de manière coopérative. Similairement aux autres hélicases DEAD-box, leur core contient au moins 12 motifs conservés : Q, I, Ia, Ib, Ic, II, III, IV, IVa, V, Va et VI (Linder and Jankowsky, 2011) (Figure 12 : Structure des hélicases DDX5 et DDX17). Le motif Q est nommé d'après le résidu glutamine conservé qui est présent dans plus de 99 % des hélicases DEAD. Le motif Q est nécessaire pour la liaison efficace de l'ARN simple brin (ARNsb) ainsi que pour les changements de conformation induits par la liaison des nucléotides et l'hydrolyse de l'ATP. Le motif I est crucial pour les activités ATPase et hélicase via son interaction avec le motif II et le motif III. Les motifs Ia et Ib participent à la liaison à l'ARN et aux réarrangements structuraux qui se produisent lors de la liaison et de l'hydrolyse de l'ATP. Les motifs II (contenant le motif DEAD) et III sont nécessaires à l'activité ATPase et interagissent avec les motifs I et VI pour former collectivement la poche de liaison de l'ATP. Le motif IV est supposé lier l'ARNsb et sert de connexion fonctionnelle entre les motifs IV et V. Les motifs V et VI sont tous deux importants pour la liaison à l'ARN en association avec les motifs Ia, Ib et IV.



**Figure 12 : Structure des hélicases DDX5 et DDX17**

Les différents domaines ne sont pas à l'échelle.

L'interaction avec l'ARN se fait via le squelette ribose-phosphate et non pas avec les bases nucléiques. La reconnaissance de l'ARN par les hélicases à DEAD-box est donc supposée plutôt être basée sur la structure de l'ARN que sur sa séquence. Comme les autres hélicases DEAD-box, DDX5 et DDX17 permettent le déroulement de petits duplex d'ARN, de manière non-processive et ATP-dépendante, sans translocation, par un mécanisme de séparation local des brins d'ARN (Yang et al., 2007) (Figure 13). La séparation des brins ne requiert que la fixation de l'ATP à l'hélicase, l'hydrolyse déclenchant le relargage de l'hélicase à l'ARN (Liu et al., 2008). En plus de la séparation de duplex d'ARN, les hélicases à DEAD-box peuvent aussi favoriser le déplacement et la séparation de protéines liées à l'ARN, leur conférant un rôle de remodelage des complexes ribonucléiques (Fairman et al., 2004).



**Figure 13 : Mécanismes moléculaires des hélicases DEAD-box**

(A) Les hélicases DEAD-box peuvent effectuer un déroulement local des brins, (B) déplacer les protéines de liaison à l'ARN (RBP) de leur substrat ou, (C) lorsque leur activité ATPase est bloquée, peuvent fonctionner comme des clamps au sein de RNP.

DDX5 et DDX17 présentent la plus grande homologie parmi les membres de la famille des DEAD-box, possédant 90% d'homologie au niveau du domaine hélicase, mais leurs extrémités N-Terminale et C-Terminale ne partagent respectivement que 60% et 30% d'homologie (Lamm et al., 1996). Ces extrémités sont supposées donner leur spécificité aux hélicases, régulant leurs interactions avec leurs partenaires protéiques. Néanmoins, ces extrémités C-ter et N-ter sont encore très mal caractérisées. On peut noter que le gène *DDX17* code deux protéines différentes, p72 et p82. L'isoforme p82 est produite via un site alternatif d'initiation de la traduction grâce à un codon non-AUG (CUG), se trouvant en amont du codon AUG qui définit le début de la traduction de l'isoforme p72 (Uhlmann-Schiffler et al., 2002). Du fait de leur importante homologie, les hélicases DDX5 et DDX17 partagent une redondance fonctionnelle importante. Néanmoins, elles possèdent aussi des rôles distincts et ne peuvent pas complètement se compenser mutuellement.

### 5.3 Fonctions moléculaires des hélicases DDX5 et DDX17

Les hélicases DDX5 et DDX17 sont impliquées dans quasiment toutes les étapes du métabolisme des ARNs, de leur synthèse à leur dégradation en passant par leur maturation.

#### 5.3.1 Régulation de l'initiation de la transcription

DDX5 et DDX17 ont été impliquées dans la corégulation transcriptionnelle de plusieurs gènes, en agissant comme corégulateurs de certains facteurs de transcription. Il a ainsi été démontré que DDX5 et DDX17 interagissent avec et activent le récepteur aux œstrogènes alpha ( $E\alpha$ ). Ce rôle de coactivateur transcriptionnel apparaît être indépendant de leur activité hélicase et ATPase (Wortham et al., 2009). À noter tout de même que le papier fondateur concernant le rôle de DDX5 comme co-activateur transcriptionnel de  $E\alpha$  a été rétracté en 2014 pour manipulation d'image (Endoh et al., 2014). Plusieurs autres groupes ont toutefois obtenu des résultats similaires, semblant donc confirmer le rôle de DDX5 et DDX17 dans la coactivation de  $E\alpha$  (Dutertre et al., 2010; Wortham et al., 2009).

Par ailleurs, il a été montré que DDX5 et DDX17 interagissent avec l'ARNInc SRA et il a été suggéré que DDX5 et DDX17 ont pour rôle de recruter SRA au sein de complexes de cofacteurs transcriptionnels. DDX5 et SRA coactivent les gènes cibles de la voie de signalisation Notch (Jung et al., 2013), et les deux hélicases ainsi que SRA coactivent MYOD (Carette et al., 2006).

DDX5 est aussi un cofacteur du récepteur des androgènes (AR). DDX5 est ainsi nécessaire au recrutement de AR et de la  $\beta$ -Caténine aux régions promotrices des gènes répondant aux androgènes (Clark et al., 2013). Des expériences utilisant un rapporteur Luciférase-PSA (Prostate-Specific Antigen) ont montré que la cotransfection des constructions DDX5 et  $\beta$ -Caténine avait un effet additif sur l'activité transcriptionnelle de AR en présence d'androgènes. Là encore, l'activité hélicase de DDX5 ne semble pas nécessaire. Dans cette étude, DDX17 ne coactive pas AR, que ce soit seul ou en conjonction avec DDX5. Cependant, il a été rapporté que DDX5 était recruté sur le promoteur du PSA bien qu'il n'ait pas été spécifiquement démontré qu'il coactivait AR (Wong et al., 2009).

DDX5 est également un coactivateur 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 (Nicol et al., 2013). D'autre part, DDX5 et DDX17 activent l'activité transcriptionnelle de MyoD, régulateur majeur de la myogenèse (Carette et al., 2006; Dardenne et al., 2014). DDX5 est également coactivateur de Runx2, facteur clé de la différenciation ostéoblastique (Jensen et al., 2008). Comme pour la coactivation des récepteurs hormonaux, leur activité hélicase apparaît de manière générale comme dispensable dans leur rôle de coactivateur de la transcription.

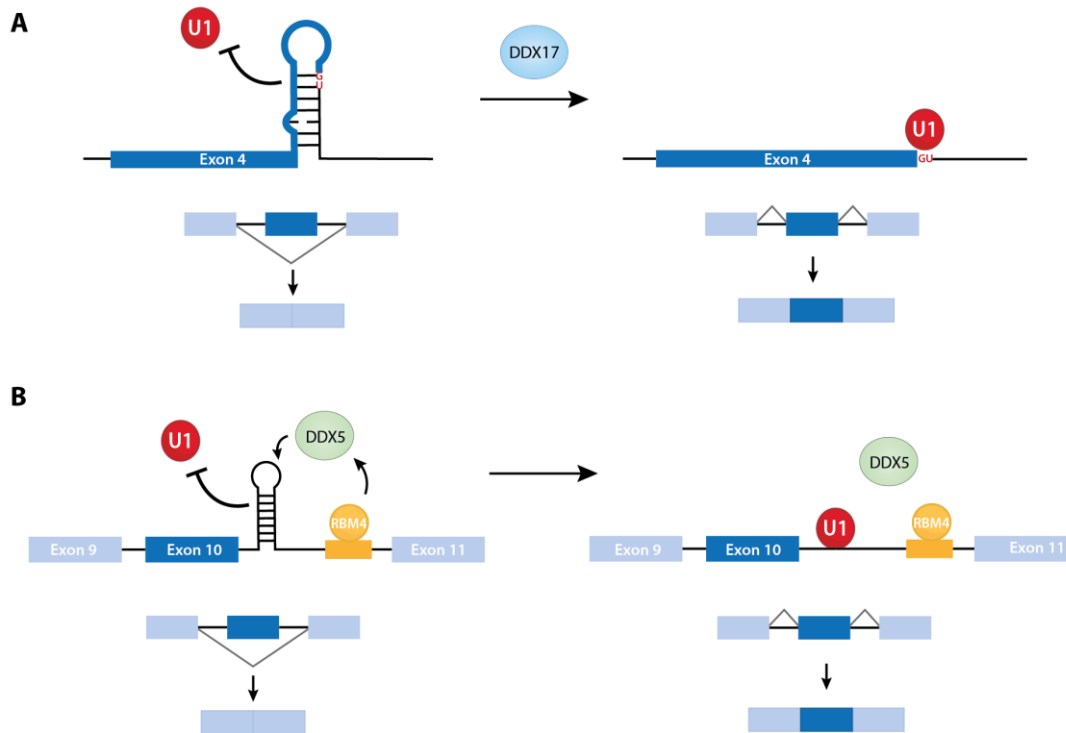
DDX5 et DDX17 sont aussi des cofacteurs de REST (Repressor Element 1-silencing transcription factor) (Lambert et al., 2018). Leur codéplétion couplée à celle de REST induit une augmentation de l'expression des gènes réprimés par REST plus forte qu'en cas d'une simple déplétion de REST. De plus, DDX17 interagit avec REST, et la codéplétion de DDX5 et DDX17 induit une diminution de la fixation de REST aux promoteurs de ses gènes cibles, suggérant que les deux hélicases sont nécessaires au recrutement de REST aux promoteurs des gènes qu'il réprime.

### 5.3.2 Rôle de DDX5 et DDX17 dans l'épissage

Les premiers indices d'une implication de DDX5 et DDX17 dans l'épissage ont été décelés en observant par spectrométrie de masse que DDX5 était un composant du spliceosome (Neubauer et al., 1998). Cela a ensuite été confirmé en montrant que DDX5 et DDX17 interagissent avec la snRNP U1 qui reconnaît le site 5' d'épissage (Lee, 2002; Liu, 2002). La déplétion de DDX5 inhibe la dissociation du duplex formé par la snRNP U1 et le site 5', entraînant l'accumulation d'ARN non épissés (Lin et al., 2005). Cette fonction est hélicase dépendante, car abolie par des mutations inactivant l'activité hélicase ou ATPase de DDX5 (Lin et al., 2005).

DDX5 et DDX17 ont par la suite été démontrées comme impliquées dans le contrôle de l'épissage alternatif d'une multitude de transcrits dans divers contextes cellulaires. L'étude plus poussée du mécanisme moléculaire sous-jacent sur plusieurs transcrits a permis de montrer que cette activité se fait par un mécanisme similaire par les deux hélicases, même si elles présentent une spécificité de substrat (Figure 14). DDX17 favorise ainsi l'inclusion de l'exon variable v4 du gène *CD44*, en déstabilisant la structure secondaire d'une tige boucle présente dans l'intron en aval l'exon v4 (Hönig et al., 2002). Le site 5' normalement masqué par la tige boucle serait ainsi rendu accessible pour l'hybridation avec la sous-unité snRNP U1 du spliceosome. Un mécanisme semblable a

été décrit permettant à DDX5 de promouvoir l'inclusion de l'exon 10 du gène *Tau* (Kar et al., 2011). DDX5 se fixe sur l'ARN pré-messager de *Tau* au niveau du 5'ss de l'intron 10 et permet de dérouler une structure de type tige-boucle qui empêche l'hybridation avec la snRNP U1.



**Figure 14 : Régulation de l'épissage alternatif par DDX5 et DDX17**

(A) Régulation de l'épissage alternatif de l'exon 4 de *CD44* par DDX17. DDX17 favorise l'inclusion de l'exon variable v4 du gène *CD44*, en déstabilisant la structure secondaire d'une tige boucle présente dans l'intron en aval l'exon v4. Cela permet au site 5' normalement masqué par la tige boucle d'être accessible pour l'hybridation avec la sous-unité snRNP U1 du spliceosome (B) Régulation de l'épissage alternatif de l'exon 10 de *Tau* par DDX5. DDX5 se fixe sur l'ARN pré-messager de *Tau* au niveau du 5'ss de l'intron 10 et permet de dérouler une la tige-boucle qui empêche normalement l'hybridation avec la snRNP U1. Le site 5' peut ainsi être reconnu par la snRNP U1, favorisant l'inclusion de l'exon 10 de *Tau*.

Par ailleurs, une analyse à large échelle réalisée dans une lignée de cellules épithéliales du sein, les MCF10A, a permis de montrer que DDX5 et DDX17 régulent l'inclusion d'environ 2000 exons (Dardenne et al., 2014). DDX5 et DDX17 sont ainsi capables de

réguler l'épissage à la fois directement en tant que composants du spliceosome mais aussi en tant que facteurs auxiliaires pouvant faciliter ou inhiber la reconnaissance des séquences régulatrices de l'épissage.

### 5.3.3 Rôle de DDX5 et DDX17 dans la maturation en 3' des ARN et la terminaison transcriptionnelle

Les premières indications d'un possible rôle de DDX5 et DDX17 dans le processus de maturation en 3' des transcrits et de la terminaison transcriptionnelle remontent à leur identification comme protéines appartenant au complexe de maturation de l'extrémité 3' des ARN (Shi et al., 2009). En effet, après purification du complexe de maturation en 3' et analyse de sa composition via la méthode MUDPIT (multidimensional protein identification technology), 85 protéines ont pu être identifiées, parmi lesquelles DDX5 et DDX17.

Plus tard, DDX5 fut indirectement reliée à la régulation de la terminaison transcriptionnelle par le complexe régulateur du rythme circadien PER-CRY (Padmanabhan et al., 2012). Il fut ainsi démontré que le complexe PER-CRY contient DDX5, ainsi que la grande sous-unité de l'ARN Pol II, les pré-ARNm de *Per* et *Cry* et la sénataxine. La sénataxine participe à la terminaison transcriptionnelle en résolvant les duplex ARN/ADN, ce qui facilite la dégradation de l'extrémité 3' de l'ARN et permet la libération de la Pol II. Il a été démontré que la Pol II s'accumule près de l'extrémité 3' des gènes *Per* et *Cry* lors du contrôle rétro-négatif du rythme circadien. Il fut constaté que le complexe PER-CRY s'associait à la Pol II lors de l'élongation et qu'il inhibait l'activité de la sénataxine, entravant ainsi la terminaison transcriptionnelle. Le rôle de DDX5 ne fut pas directement étudié dans cette étude.

#### 5.3.3.1 Régulation de la terminaison transcriptionnelle par la résolution de R-loops

Plus récemment, DDX5 et DDX17 ont été impliquées dans la terminaison transcriptionnelle du gène *HSPA1A* en coopération avec le complexe protéique THO (Katahira et al., 2019). Le complexe THO comprend six unités, THOC1, -2, -3, -5, -6, et -7, et permet entre autres de coupler la transcription et la maturation des ARNm à leur exportation au cytoplasme (Katahira, 2012; Katahira et al., 2013). L'analyse de l'interactome de hTHO (human THO) a permis d'identifier comme partenaires les protéines Ctsf64 et XRN2, respectivement impliquées dans la maturation de l'extrémité

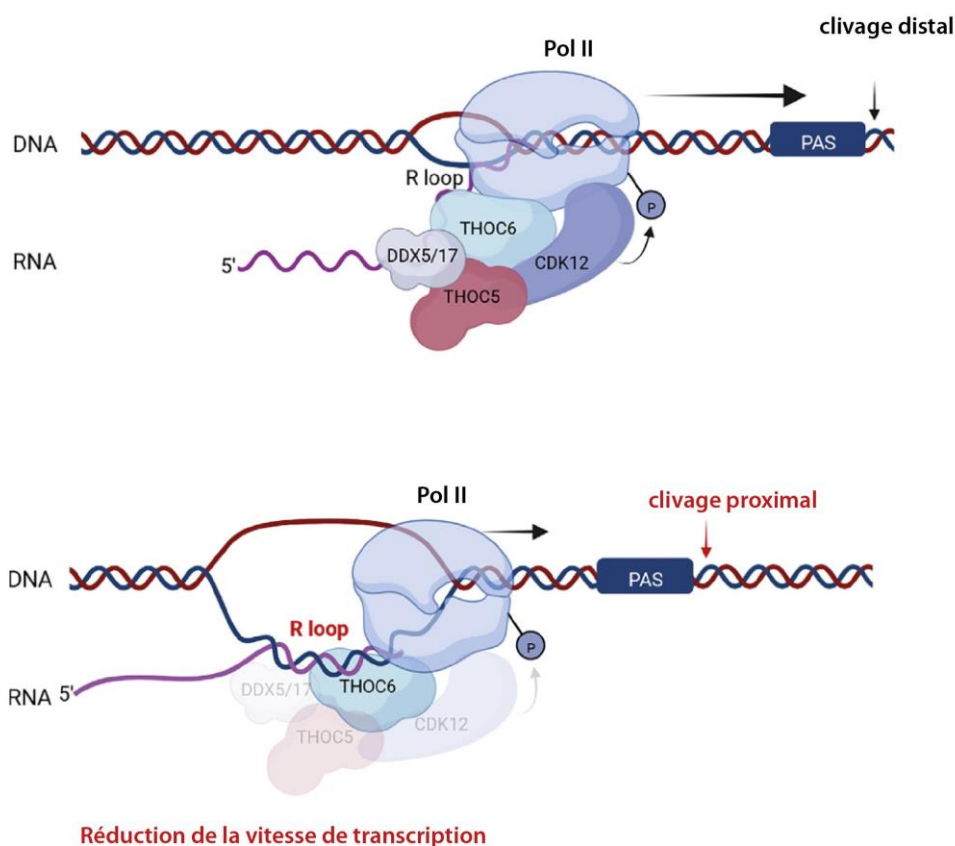
3' des transcrits et dans la terminaison transcriptionnelle, mais aussi les hélicases DDX5 et DDX17. De façon notable, la codéplétion de DDX5 et DDX17 par des siARNs entraîne une accumulation des transcrits *HSPA1A* dans des foci nucléaires, accompagnée d'une augmentation de l'expression de transcrits en aval du PAS de *HSPA1A*, ainsi qu'une accumulation de Ctsf64, XRN2 et THO au niveau du site de pause transcriptionnelle du gène *HSPA1A*, juste en aval du PAS. Il n'a toutefois pas été vérifié dans cette étude si le transcrit en aval du PAS est continu ou non avec le transcrit *HSPA1A* et donc que son expression résulte bien d'un readthrough transcriptionnel. Les auteurs n'observent cependant pas d'effet du siDDX5/17 ou du siTHOC5 sur le taux de clivage du transcrit *HSPA1A*. Cela suggère que DDX5 et DDX17 n'auraient pas d'effet sur le clivage du transcrit, mais seraient en revanche nécessaires à la bonne terminaison de la transcription et à la dégradation subséquente de l'extrémité clivée. La surexpression de la RNase H, une enzyme hydrolysant les hybrides ADN:ARN, permet de diminuer l'accumulation de transcrits *HSPA1A* dans les foci nucléaires normalement observée suite à la déplétion de DDX5 et DDX17. Les hybrides ADN:ARN sont notamment présents au niveau des R-loops, une structure à trois brins consistant en un hybride ADN:ARN et un brin d'ADN non apparié, qui se forment notamment durant la transcription. Ces résultats suggèrent donc que les hélicases DDX5 et DDX17 pourraient permettre de limiter la formation de R-loops au niveau du site de pause de la Pol II, facilitant ainsi la terminaison transcriptionnelle médiée par XRN2 et le relargage du transcrit mature *HSPA1A* (Katahira et al., 2019). À noter tout de même que l'étude ne présente pas de données permettant de contrôler l'effet de la codéplétion de DDX5 et DDX17 sur la formation de R-loops dans leur modèle.

Néanmoins, une étude menée par une autre équipe, portant aussi sur la régulation de la terminaison transcriptionnelle par THO5, a ensuite pu confirmer que DDX5 et DDX17 étaient des partenaires de THO5 mais surtout que leur déplétion entraîne bien l'accumulation de R-loops et qu'elles sont capables de résoudre des duplexes ARN:ADN *in vitro* (Polenkowski et al., 2023). Les auteurs montrent par ailleurs que DDX5 et DDX17 sont recrutées au niveau du TTS de plusieurs gènes cibles du complexe THO, et que la déplétion de THO5 entraîne une diminution de ce recrutement. La déplétion de DDX5 et DDX17 entraîne de plus une accumulation de R-loops au niveau des gènes cibles de THO.

On peut remarquer que cette étude à plus large échelle présente des résultats quelque peu en opposition avec la précédente. De manière assez notable, alors que Katahira et



*al.* n'observaient pas d'effet significatif de la déplétion de THO5 sur le clivage de HSP70 (Katahira et al., 2019), ici les auteurs constatent que la déplétion de THO5 entraîne une augmentation de 1,5 fois de l'utilisation du PAS proximal par rapport au distal pour 52-62% des gènes possédant au moins deux PAS (Polenkowski et al., 2023). Ils n'observent pas non plus d'effet sur l'exportation des ARNm. Bien sûr, il n'est pas choquant que le résultat d'une analyse large du transcriptome soit en contradiction avec une observation portant sur un seul gène. Par ailleurs, l'utilisation de lignées cellulaires exprimant des Pol II porteuses de mutations affectant leur vitesse d'élongation a permis de montrer que les cellules exprimant les Pol II lentes présentaient une augmentation de l'usage de PAS proximaux d'environ 52% des gènes possédant au moins deux PAS, suggérant qu'une vitesse d'élongation plus faible favorise l'usage des PAS proximaux. Les auteurs montrent que THO5 interagit avec CDK12, une protéine favorisant l'élongation de la Pol II en phosphorylant la CTD, et que la déplétion de CDK12 entraîne aussi une augmentation de l'usage du PAS proximal, suggérant que THO favorise l'élongation transcriptionnelle via CDK12 (Polenkowski et al., 2023). En prenant en compte les observations concernant DDX5 et DDX17, les auteurs suggèrent donc que THO5 recrute DDX5 et DDX17 et CDK12 au niveau des R-loops formées le long des gènes présentant une transcription ralentie, permettant ainsi une résolution des R-loops et une accélération de l'élongation qui favorise en conséquent l'usage de PAS distaux (Figure 15). L'effet de DDX5 et DDX17 sur l'utilisation de PAS proximaux ou distaux n'est cependant pas examiné. Les auteurs ne démontrent pas non plus d'effet direct de la présence de R-loops sur la vitesse d'élongation et la sélection du PAS.



**Figure 15 : THOC5/THOC6 favorise l'usage de PAS distaux en recrutant CDK12 et DDX5/17 pour réguler les R-loops et la vitesse d'élongation**

Lorsque le taux d'élongation de la transcription est faible et que des R-boucles se forment, le complexe THOC5/THOC6 recrute la CTD kinase CDK12, qui phosphoryle le domaine CTD de la Pol II sur la sérine 2, stimulant l'élongation. Parallèlement, le complexe THOC5/THOC6 recrute également les hélicases à ARN DDX5 et DDX17 pour résoudre les R-loops, empêchant ainsi l'altération de l'élongation de la transcription et sa terminaison prématurée. (Polenkowski et al., 2023)

Le rôle de DDX5 dans la résolution de R-loops a été confirmé dans plusieurs autres études. Un papier montre ainsi que la déplétion de DDX5 dans des cellules U2OS entraîne l'accumulation de R-loops au niveau du TTS des gènes à l'échelle du génome (Villarreal et al., 2020). Les mêmes auteurs ont aussi montré que DDX5 s'associe à XRN2 et résout les R-loops dans les régions de terminaison transcriptionnelle en aval du PAS

de plusieurs gènes. De plus la protéine arginine méthyltransférase 5 (PRMT5) se lie à DDX5 et la méthyle au niveau de son motif RGG/RG. Ce motif est nécessaire à l'interaction de DDX5 avec XRN2 et à la répression des R-loops, mais n'est pas essentiel à l'activité enzymatique de l'hélicase DDX5.

#### *5.3.3.2 Régulation de la maturation en 3' des transcrits par formation de boucles chromatinienne*

Des travaux récemment menés dans mon équipe d'accueil ont par ailleurs permis de montrer que la déplétion des hélicases DDX5 et DDX17 affecte bien la terminaison transcriptionnelle (Terrone et al., 2022). En effet, l'analyse RNA-seq d'une lignée de cellules de neuroblastome SH-SY5Y traitées par siRNA contre *DDX5* et *DDX17* montre qu'environ 800 gènes présentent un profil de readthrough transcriptionnel en aval de leur site de clivage annoté lorsque les hélicases sont déplétées. L'analyse des transcrits *NCKAP5L* et *SH3TC1* montrent qu'ils sont bien continus avec la région en aval leur site de clivage, démontrant que l'effet de la déplétion de DDX5 et DDX17 est une augmentation du readthrough transcriptionnel. De plus, la déplétion de DDX5 et DDX17 entraîne une augmentation de la présence de Pol II en aval du PAS des gènes présentant un readthrough transcriptionnel. Ces résultats suggèrent donc que DDX5 et DDX17 peuvent réguler le processus de maturation en 3' des transcrits et la terminaison transcriptionnelle. On peut noter que 34% des gènes présentant un readthrough transcriptionnel suite à la déplétion de DDX5 et DDX17 présentent également une altération de leur épissage, significativement plus qu'attendu, suggérant un possible lien entre la régulation de ces deux processus de maturation par DDX5 et DDX17. Cela apparait cohérent avec les études précédemment citées montrant un lien entre perturbation de l'épissage et readthrough transcriptionnel (Alpert et al., 2020; Castillo-Guzman et al., 2020; Reimer et al., 2021; Rutkowski et al., 2015).

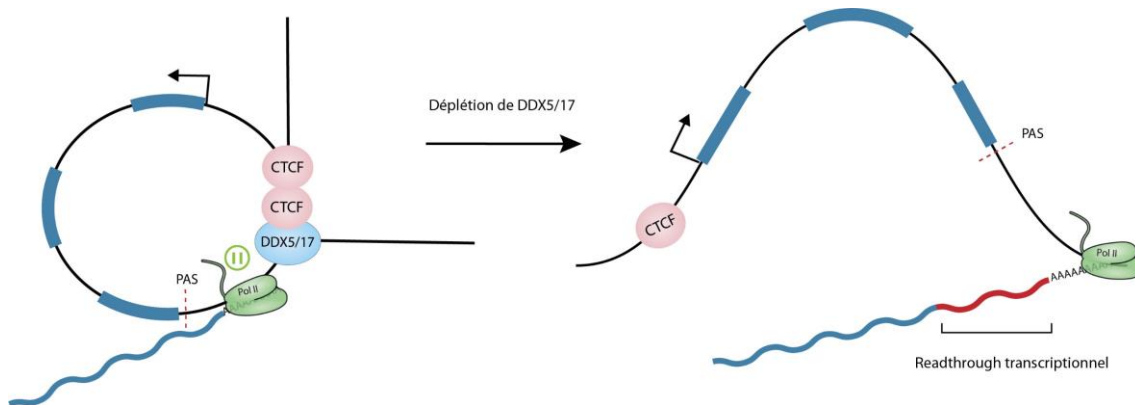
Par ailleurs, l'analyse de données publiques de CHIP-seq (Chromatin Immunoprecipitation and sequencing) permet de constater que les régions promotrices et terminatrices des gènes régulés par les hélicases DDX5 et DDX17 sont plus fréquemment associés à des sites de fixation de CTCF que les gènes non régulés. Combinée à celle des hélicases, la déplétion de CTCF amplifie le readthrough transcriptionnel induit par la seule déplétion de DDX5 et DDX17, suggérant que CTCF, DDX5 et DDX17 co-régulent la maturation en 3' et la terminaison transcriptionnelle d'un

ensemble de gènes. De plus, la délétion du site de fixation de CTCF présent en aval du gène *SH3TC1* entraîne une augmentation de son readthrough transcriptionnel, démontrant le rôle de CTCF dans la régulation de la maturation de l'extrémité 3' des ARNm.

L'équipe a aussi pu montrer que DDX5 et DDX17 régulent la formation de boucles chromatiennes entre le promoteur et la région terminatrice par le biais de CTCF. Des expériences de 3C (Chromosome Conformation Capture) réalisées sur deux gènes cibles, *NCS1* et *PRMT2*, ont confirmé notamment la proximité spatiale entre leur promoteur et leur région terminatrice, et montré que ces boucles sont déstabilisées par la déplétion de DDX5 et DDX17. De plus, la déplétion de DDX5 et DDX17 entraîne une diminution de la fixation de CTCF au niveau du promoteur et de la région 3' de ces deux gènes.

Pour finir, grâce à l'utilisation de la technique CLOuD9 (chromatin loop reorganization using CRISPR-dCas9) il a été possible d'artificiellement forcer la formation d'une boucle entre le promoteur et la région 3' du gène *FBLN1*, qui présente un readthrough transcriptionnel après déplétion de DDX5 et DDX17. Ce gène possède à la fois un PAS proximal et distal. L'utilisation du PAS proximal augmente lorsque la formation de la boucle est induite, montrant que la formation de celle-ci affecte bien la maturation en 3' de *FBLN1*. Cela peut suggérer que l'augmentation de readthrough induit par la déplétion de DDX5 et DDX17 passe par la déstabilisation de la boucle intragénique, même si cela n'est pas directement démontré.

Pris ensemble, ces résultats suggèrent donc les hélicases DDX5 et DDX17 sont nécessaires à la bonne terminaison transcriptionnelle d'un ensemble de gènes, en favorisant la fixation de CTCF au niveau du site 3' de ces gènes, ce qui permet la formation de boucles chromatiennes intragéniques entre le promoteur et le site de terminaison de ces gènes (**Figure 16**). Il a déjà été démontré par une précédente étude que la formation de boucles chromatiennes médiées par CTCF et cohésine au niveau des PAS proximaux favorise l'usage de ceux-ci et donc l'expression d'isoformes courtes ([Nanavaty et al., 2020](#)). Cela n'avait été démontré que sur deux locus et l'effet global de CTCF sur la polyadénylation n'a pas été étudié mais ces travaux constituent une preuve que CTCF est capable de moduler la maturation en 3' des transcrits.



**Figure 16 : Modèle de régulation de l'extrémité 3' des ARNm par DDX5 et DDX17**

En présence de DDX5/17, certains gènes forment des boucles intragéniques permises par la fixation de CTCF, favorisant l'usage du PAS annoté et la terminaison transcriptionnelle de la Pol II. Suite à la déplétion de DDX5/17, la fixation de CTCF et la présence de boucles diminuent. La Pol II envahit alors la région 3' en aval du gène, produisant un transcrit présentant un readthrough transcriptionnel.

Le mécanisme par lequel DDX5 et DDX17 régulent la liaison CTCF au niveau de l'extrémité 3' des gènes n'est pas examiné dans cette étude. Par ailleurs, le readthrough induit par la destruction du site CTCF présent en aval du PAS de *SH3TC1* est significativement augmenté lorsque les hélicases DDX5 et DDX17 sont déplétés. Ceci suggère que le rôle des deux hélicases dans la régulation de la maturation en 3' et/ou de la terminaison n'implique pas seulement CTCF et que d'autres facteurs pourraient être impliqués.

### 5.3.3.3 Régulation de la maturation en 3' des transcrits par la résolution de G-quadruplexes

De façon intéressante, les sites de terminaison régulés par DDX5 et DDX17 présentent un pourcentage GC plus élevé que les non-régulés. Les régions riches en GC sont plus susceptibles de former des structures secondaires telles que les G-quadruplexes (G4) et les R-loops. Comme précédemment indiqué, plusieurs papiers ont déjà rapporté le rôle de DDX5 et DDX17 dans la résolution de R-loops et certains ont pu relier ce rôle à la régulation de la terminaison transcriptionnelle. DDX5 et DDX17 sont aussi capables de résoudre les structures en G-quartets (G4) (Gao et al., 2019; Wu et al., 2019). Les G4 sont des structures à quatre brins, formées par l'auto-association de quatre guanines en quartets qui s'empilent en piégeant des ions potassium. Ils se forment au niveau de

séquences riches en guanine, et il existe des G4 d'ADN, d'ARN mais aussi des G4 hybrides ADN-ARN (Varshney et al., 2020). Les G4 ARN formés au niveau de la région 3' UTR des transcrits ont été identifiés comme augmentant l'efficacité de l'utilisation des sites de polyadénylation alternatifs, ce qui conduit à l'expression des isoformes possédant une extrémité 3' UTR plus courte (Beaudoin and Perreault, 2013). Il a aussi été proposé que le G4 ARN présent en aval du PAS du gène *TP53* serait nécessaire pour permettre la maturation en 3' correcte lors de l'induction de dommages à l'ADN (Decorsière et al., 2011). Le mécanisme reposerait sur la liaison de hnRNP H/F avec le G4 ARN.

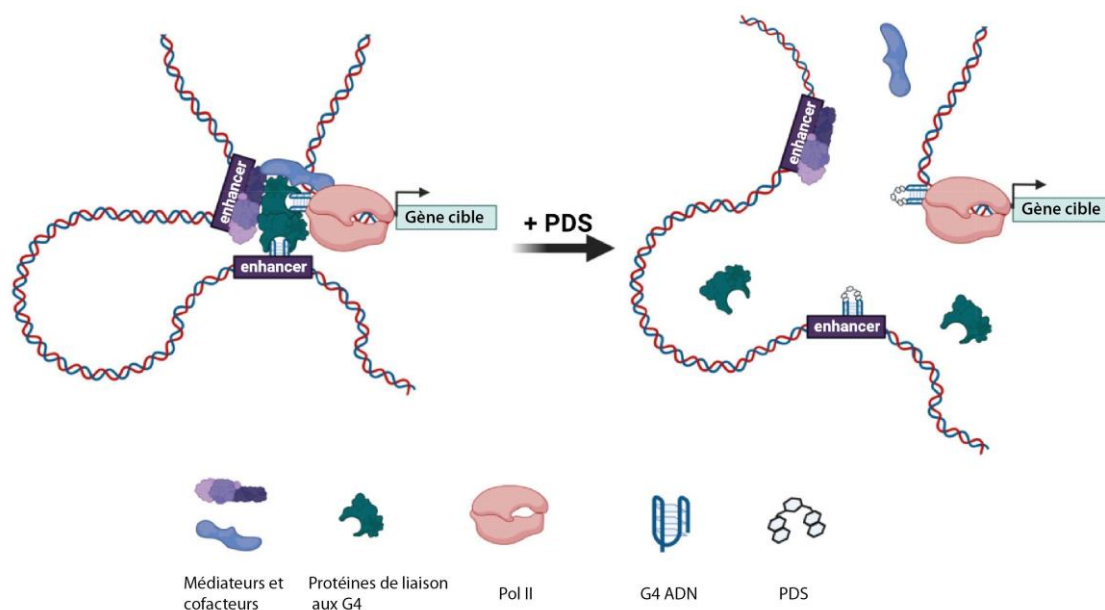
#### 5.3.3.4 *Lien entre résolution des structures secondaires et la formation de boucles chromatinienne ?*

Plusieurs études récentes indiquent que les structures secondaires telles que les G4 et les R-loops seraient impliquées dans la formation de boucles chromatinienne. Un ensemble de papiers récemment publiés cette année (2023) met en évidence la connexion entre G4, R-loops, CTCF et l'architecture 3D de la chromatine.

Les R-loops ont déjà précédemment été démontrées comme participant à la fixation de CTCF à la chromatine (Luo et al., 2022). Un récent article montre qu'indépendamment de CTCF, les R-loops peuvent constituer une barrière à la cohésine et pourraient participer à l'insulation génomique médiée par la cohésine (H. Zhang et al., 2023). De façon intéressante, un autre papier a précédemment montré que la présence de duplex ADN-ARN stabilise la formation de boucles chromatinienne induites par la transcription reliant l'extrémité 5' et 3' de plusieurs gènes (Pezone et al., 2019).

Il a de plus été montré que les régions liées par CTCF sont enrichies en R-loops et en G4. De plus, la présence de G4 près des sites consensus CTCF renforce la liaison de CTCF à ceux-ci (Wulfridge et al., 2023). Ainsi la délétion du motif permettant la formation d'un G4 (conservant le même contenu en GC) présent dans le premier intron du gène *Praf2*, 26 pb en aval du site CTCF, entraîne une diminution significative de la liaison de CTCF. Similairement, la stabilisation des G4 grâce à la Pyridostatin, un ligand des G4, résulte en l'augmentation globale de la fixation de CTCF à la chromatine. La stabilisation des G4 entraîne aussi une augmentation du nombre de boucles chromatinienne détectées par Hi-C.

Un second article met en évidence une forte corrélation entre les G4 d'ADN et les interactions chromatinienne à longue distance médiés par la Pol II (Yuan et al., 2023), via l'analyse de données publiques de ChIA-PET (chromatin interaction analysis with paired-end tag). Cela vient confirmer un lien possible entre G4 et formation de boucles chromatinienne. Toutefois, dans ce papier la stabilisation des G4 grâce à la Pyridostatin induit une diminution des contacts à longue distance médiés par la Pol II. Une hypothèse pour expliquer la différence avec le résultat obtenu dans l'étude précédente pourrait être que dans le cas des contacts ADN-Pol II, la formation de boucles dépend de protéines se liant aux G4. La liaison de la Pyridostatin avec les G4 empêcherait leur reconnaissance par les protéines de liaison (Figure 17). Le renforcement de la liaison de CTCF induite par la formation de G4 ne dépend donc vraisemblablement pas de protéines de liaison aux G4.



**Figure 17 : Modèle illustrant l'implication des G4 dans les interactions à longue distance de l'ADN liées à la Pol II.**

Un ligand de G4 de petite taille, la Pyridostatin (PDS), peut perturber la capacité de certaines protéines (par exemple YY1) à lier les G4 et altérer l'organisation 3D du génome. Adapté de (Yuan et al., 2023)

On peut donc supposer que DDX5 et DDX17 pourraient réguler la maturation en 3' des transcrits notamment via leur activité de régulation des structures secondaires.

Néanmoins on peut remarquer que les études précédentes montrent que la présence de G4 et de R-loops semble plutôt favoriser la formation de boucles chromatiniennes et la sélection de PAS proximal. Or la déplétion de DDX5 et DDX17 déstabilise les boucles et induit un readthrough transcriptionnel (Terrone et al., 2022), alors qu'elles ont une activité décrite de résolution des G4 et des R-loops (Gemmill et al., 2023; Polenkowski et al., 2023; Villarreal et al., 2020; Wu et al., 2019), et donc logiquement leur déplétion devrait induire une diminution de la présence des G4 et R-loops. Des travaux supplémentaires doivent donc être menés afin d'éclaircir ce point.

#### 5.3.4 Autres fonctions moléculaires

En plus des fonctions précédemment décrites, les hélicases DDX5 et DDX17 sont aussi impliquées dans la biogénèse des microARN et des ribosomes, la dégradation des ARNm non-sens, l'export des ARNm et la réparation de l'ADN (Xing et al., 2019), mécanismes que je ne décrirai pas ici, car ils n'ont pas été l'objet de mes travaux.

#### 5.4 Rôle de DDX5 et DDX17 dans la différenciation cellulaire

Leurs fonctions moléculaires multiples font de DDX5 et DDX17 des régulateurs clés de fonctions cellulaires majeures telles que la prolifération ou la différenciation. DDX5 et DDX17 sont ainsi impliquées dans la différenciation de plusieurs types cellulaires.

Les deux hélicases régulent la différenciation musculaire via la coactivation du facteur de transcription MyoD. La déplétion de DDX5 et DDX17 entraîne une absence de recrutement de TBP (TATA binding protein), de la Pol II et de certaines protéines de remodelage de la chromatine aux promoteurs des gènes musculaires, ainsi qu'une inhibition de la différenciation en cellules musculaires squelettiques (Caretto et al., 2006). La régulation de la différenciation musculaire par les hélicases passerait aussi par leur rôle de contrôle de l'épissage alternatif en coopération avec hnRNP H/F (Dardenne et al., 2014).

DDX5 et DDX17 participent aussi à la différenciation ostéoblastique en co-activant le facteur de transcription RunX2 (Jensen et al., 2008). DDX5 interagit avec RunX2 et est recrutée au niveau de ses promoteurs cibles, et la codéplétion des deux hélicases



entraîne une diminution de l'activité transcriptionnelle de RunX2. Cependant, de façon surprenante, leur déplétion entraîne une accélération de la maturation des ostéoblastes.

Plusieurs éléments indiquent que DDX5 et DDX17 régulent la différenciation neuronale. Comme indiqué précédemment dans la partie [5.3.1](#), DDX5 et DDX17 sont des cofacteurs de REST, un facteur de transcription réprimant un grand nombre de gènes neuronaux dans les cellules non neuronales ou dans les progéniteurs neuronaux. DDX5 et DDX17 contrôlent de plus la biogenèse de miARN pro-neuronaux ([Lambert et al., 2018](#)). Par ailleurs, une autre étude démontre que DDX17 et DDX5 sont mutuellement nécessaires pour contrôler l'expression transcriptionnelle de gènes associés à la neurogenèse, et sont essentielles à la différenciation neuronale précoce des cellules souches pluripotentes humaines ([Suthapot et al., 2022](#)).

## 5.5 DDX5 et DDX17 en contexte pathologique

### 5.5.1 DDX5 et DDX17 dans les infections virales

#### 5.5.1.1 Rôles de DDX5 et DDX17 dans l'infection virale

Les hélicases DDX5 et DDX17 régulent une multitude de voies moléculaires et cellulaires, et joue un rôle particulièrement crucial dans le métabolisme des ARN. Les virus dépendant de leur capacité à détourner la machinerie cellulaire pour assurer leur réplication, il n'est donc pas étonnant que de nombreux éléments viennent démontrer l'implication de DDX5 et/ou DDX17 dans plusieurs infections virales. DDX5 et DDX17 peuvent jouer un rôle proviral ou antiviral selon le type de virus. Le [Tableau 1](#) répertorie les interactions identifiées entre DDX5 et DDX17 et les virus. DDX5 favoriserait ainsi la réplication de plusieurs virus à ARN et interagit avec plusieurs protéines virales. DDX5 interagit par exemple avec Nsp13, une hélicase du coronavirus responsable du syndrome respiratoire aigu sévère (SARS-CoV-1) essentielle pour son cycle viral ([Chen et al., 2009](#)). La déplétion de DDX5 entraîne une diminution de la réplication virale, suggérant que son interaction avec Nsp13 est nécessaire pour la bonne réplication du virus. Les autres virus dont la réplication semblent être favorisée par DDX5 via son interaction avec les protéines virales incluent le virus de l'immunodéficience humaine (VIH-1) ([Zhou et al., 2013](#)), le virus de l'hépatite C (HCV) ([Goh et al., 2004](#)), le virus de l'encéphalite japonaise

(JEV) (Li et al., 2013), le virus du syndrome dysgénésique et respiratoire porcin (SDRP) (Zhao et al., 2015), et le virus de la grippe A (IAV) (Bortz et al., 2011).

DDX5 semble par contre plutôt jouer un rôle antiviral dans l'infection de virus à ADN comme le virus de l'hépatite B (HBV) (H. Zhang et al., 2016) et le virus de la myxomatose (MYXV) (Rahman et al., 2017). La déplétion de DDX5 augmente ainsi la transcription du mini-chromosome de HBV, ce qui suggère que DDX5 pourrait fonctionner comme un inhibiteur de l'infection par HBV (H. Zhang et al., 2016). Il est intéressant de remarquer que l'infection chronique par HBV est une cause des carcinomes hépatocellulaires (HCC), un des rares cancers où la surexpression de DDX5 peut être un facteur de pronostic favorable (même si différentes études rapportent des résultats plutôt contradictoires, cf **Tableau 3**).

Comme DDX5, DDX17 favorise la réplication du virus IAV en interagissant avec sa polymérase (Bortz et al., 2011). DDX17 est aussi impliquée dans l'infection par le VIH, mais par des mécanismes pouvant différer de DDX5. Comme son homologue, elle pourrait être impliquée dans l'export nucléaire de l'ARN du VIH en interagissant avec la protéine virale Rev (Yasuda-Inoue et al., 2013). Il a aussi été démontré que DDX17 contrôle l'épissage du VIH-1, indépendamment de DDX5 (Sithole et al., 2018) et que DDX17 favorise la production de particules infectieuses du VIH-1 (Lorgeoux et al., 2013). Certains articles rapportent aussi un rôle antiviral pour DDX17 dans l'infection par le VIH-1, elle servirait ainsi au recrutement de l'enzyme de clivage de coiffe, entraînant la dégradation des ARNm viraux (Zhu et al., 2011).

#### *5.5.1.2 Rôles de DDX5 et DDX17 dans la réponse immunitaire antivirale*

En plus de leur rôle dans le cycle viral, DDX5 et DDX17 pourraient aussi influencer la réponse immunitaire antivirale. Il a ainsi été montré que la surexpression de DDX5 inhibe la production d'IFN- $\beta$ , IL-6, DHX58, et p65 normalement induite par l'infection par IAV de cellules humaines A549 (Zhao et al., 2022). Sa déplétion entraîne une augmentation de l'expression de IFN- $\beta$ , IL-6, DHX58, mais pas de p65. La surexpression de METTL3, METTL14 ou YTHDF2 diminue significativement les niveaux d'ARNm de IFN- $\beta$ , IL-6 et DHX58, ce qui suggère que ces transcrits sont régulés par l'axe DDX5/METTL3/METTL14/YTHDF2.

Par ailleurs, les hélicases DDX5 et DDX17 régulent de nombreuses voies de signalisation (voir 5.5.3) et ces voies peuvent moduler la réponse cellulaire durant l'infection virale. On peut ainsi citer la voie Notch, qui a été identifiée dans les réponses immunitaires modulatrices induites par de nombreux virus, tels que le SDRP (Lu et al., 2020), le virus respiratoire syncytial (RSV) (Schaller et al., 2007, p. 4) et l'entérovirus 71 (Hao et al., 2018, p. 3), mais aussi la voie NF- $\kappa$ B, qui serait manipulée par plusieurs virus pour échapper à la réponse antivirale (Deng et al., 2018) et la voie Wnt/ $\beta$ -caténine, qui pourrait réguler la réplication de IAV (Xu et al., 2020). Néanmoins, l'implication de DDX5 et DDX17 dans le contrôle de la réponse antivirale via la régulation de ces voies n'a pas encore été démontrée, mis à part l'effet de DDX5 sur les transcrits p65 et IKK- $\gamma$  dans les cellules de souris.

DEAD-box helicase	Virus	proviral(+) antiviral(-)	Rôle/site d'interaction	References
DDX5	VIH-1	+	Interagit avec la région N-terminale de la protéine Rev.	(Zhou et al., 2013)
	HCV	+	Interagit avec l'ARN polymérase ARN-dépendante (NS5B).	(Dutta et al., 2012; Goh et al., 2004)
	HBV	-	Inhibe la transcription du minichromosome de HBV	(H. Zhang et al., 2016)
	IAV	+	Affecte la polymérase virale.	(Bortz et al., 2011)
	SARS-CoV	+	Interagit probablement avec l'hélicase virale (NSP13).	(Chen et al., 2009)
	JEV	+	Interagit avec les protéines NS3 et NS5 ainsi qu'avec le 3'-UTR.	(Li et al., 2013)
	SDRP	+	Interagit avec Nsp9	(Zhao et al., 2015)
	MYXV	-	Inconnu.	(Rahman et al., 2017)
DDX17	VIH-1	+	Favorise l'exportation de l'ARN du VIH-1 dépendant de Rev.	(Yasuda-Inoue et al., 2013)
		-	Favorise la production de particules infectieuses du VIH-1.	(Lorgeoux et al., 2013)
		-	Dégrade l'ARNm du VIH-1.	(Zhu et al., 2011)
	IAV	+	Affecte la polymérase virale.	(Bortz et al., 2011)

**Tableau 1 : Interactions de DDX5 et DDX17 avec les virus**

(Adapté de Ali, 2021)

### 5.5.2 Hélicases à ARN dans les maladies génétiques

Étant données leurs nombreuses fonctions moléculaires et leur implication dans de multiples voies et mécanismes cellulaires, une altération de l'expression ou de la séquence de *DDX5* et *DDX17* semble susceptible d'être à l'origine ou de moduler certaines maladies génétiques.

Il n'existe à ce jour que trois exemples répertoriés de maladies génétiques dans lesquels *DDX5* est impliquée : la polykystose rénale (ADPKD), la dystrophie myotonique de type 1 et la dystrophie myotonique de type 2. L'ADPKD est causée par la mutation des gènes *PDK1* et *PDK2*. *DDX5* a été identifiée comme étant surexprimée dans l'ADPKD et elle agit comme corépresseur avec p53 de l'expression de *PDK1*, et en promouvant l'expression et la maturation du miR-182 ciblant l'ARNm de *PDK1* (Zhang et al., 2020). Dans les dystrophies myotoniques de type 1 et 2, *DDX5* aurait par contre un effet protecteur. Ces deux maladies neuromusculaires sont causées par l'accumulation d'ARNs porteurs de triplets CUG et CCUG dans des agrégats toxiques. L'expression de *DDX5* est réduite dans les tissus malades et un rétablissement de son expression entraîne une dégradation des ARNs mutants et réduit l'atrophie des muscles squelettiques dans les modèles murins (Jones et al., 2015; Laurent et al., 2012).

On remarque qu'il s'agit de maladies où l'altération de l'expression de *DDX5* n'est pas la cause première, mais plutôt un facteur d'aggravation. Il n'existe à ce jour pas de maladie répertoriée comme étant associée à une mutation germinale des gènes *DDX5* ou *DDX17*.

D'autres hélicases DEAD-box sont en revanche impliquées dans plusieurs maladies génétiques comme indiqué dans le **Tableau 2**. On peut remarquer que la plupart de ces maladies correspondent à des troubles neuro-développementaux. La majorité ont pour origine l'apparition de mutations *de novo*. De plus, une grande partie de ces mutations touchent les domaines hélicase ou de liaison à l'ATP. Cela semble indiquer que la fonction de nombreuses hélicases DEAD-box est cruciale au bon développement neuronal.

De manière intéressante, il fut notamment montré que les mutations de *DDX3X* pourraient expliquer 1-3% des cas de déficience intellectuelle inexplicée chez les sujets féminins (Snijders Blok et al., 2015). La déplétion de *DDX3X* dans des modèles murins altère la différenciation neuronale et réduit la génération de neurones corticaux *in vivo*.

L'analyse de l'activité biochimique de plusieurs mutants montre que les mutations faux-sens de DDX3X cliniquement sévères présentent une activité hélicase réduite, qui est associée à une formation aberrante de granules RNP et à une traduction perturbée.

Comme indiqué précédemment, DDX5 et DDX17 sont impliquées à différents niveaux dans le contrôle de la différenciation neuronale, suggérant qu'une altération de leur expression ou fonction pourrait elle aussi être à l'origine de troubles neuro-développementaux.

Helicase	Maladie ou trouble	Origine	Refs.
eIF4A3 (DDX48)	Syndrome de Richieri-Costa-Pereira	Répétition non-codante	(Favaro et al., 2014)
DDX6	Trouble du développement intellectuel avec troubles du langage et faciès dysmorphique	Mutations faux-sens de novo	(Balak et al., 2019)
DDX54	Déficience intellectuelle et anomalies du système nerveux central	Mutations faux-sens de novo	(Paine et al., 2019)
DDX3X	Trouble du développement neurologique lié à DDX3X	Mutations faux-sens de novo	(Snijders Blok et al., 2015)
	Syndrome de Toriello-Carey	Mutations faux-sens de novo hétérozygotes	(Dikow et al., 2017)
DDX5	Polykystose rénale	Surexpression	(Zhang et al., 2020)
	Dystrophie myotonique de type 1, Dystrophie myotonique de type 2	Baisse d'expression	(Jones et al., 2015; Laurent et al., 2012)
DDX23	Trouble neuro-développemental syndromique	Mutations faux-sens de novo	(Burns et al., 2021)
DDX59	Syndrome oro-faciodigital de type 5	Mutations faux-sens de novo	(Shamseldin et al., 2013)

**Tableau 2 : Hélicases DEAD-box impliquées dans des maladies génétiques**

(Adapté de Bohnsack et al., 2023)

### 5.5.3 DDX5 et DDX17 dans les cancers

Notamment du fait du rôle de DDX5 et DDX17 dans le contrôle de la prolifération et la différenciation cellulaire, une altération de leur expression est susceptible d'entraîner un déséquilibre favorisant la tumorigenèse. L'expression des hélicases *DDX5* et *DDX17* est dérégulée dans de nombreux types de tumeurs, et les deux hélicases ont été impliquées dans le processus tumoral de plusieurs cancers. DDX5 et DDX17 interagissent ainsi avec de nombreuses molécules clés de la signalisation tumorale et participent à divers processus régulant la tumorigenèse, telles que la réparation des dommages à l'ADN, l'autophagie, le stress oxydatif et le métabolisme énergétique (Xu et al., 2022). Certaines études montrent aussi que DDX5 pourrait être impliquée dans l'inhibition de la réponse immunitaire. Le **Tableau 3** récapitule les différents cancers dans lesquels DDX5 et/ou DDX17 ont été identifiées comme étant dérégulées ainsi que les mécanismes ou voies de signalisation associées.

Les gènes *DDX5* et *DDX17* sont surexprimés dans la quasi-totalité des cancers où ils ont été analysés (Xu et al., 2022), à l'exception du carcinome hépatocellulaire (HCC) et de l'adénocarcinome canalaire pancréatique (PDAC), dans lesquels la surexpression de *DDX5* est associée à un meilleur pronostic (Morimachi et al., 2021; Zhang et al., 2019), même si les résultats concernant l'HCC sont contradictoires (Xue et al., 2018). On peut remarquer que *DDX17* est surexprimé et favorise la tumorigenèse dans ces cancers (Pan et al., 2021; Xue et al., 2019), mettant en évidence que malgré leur forte similarité, ces deux hélicases ont aussi leurs caractéristiques propres et peuvent avoir un impact différent sur la cellule selon le contexte.

Type de cancer	Hélicase	Expression	Mécanismes/signalisation	Référence
Cancer du sein	DDX5/17	↑	DDX5/DDX17 sont des activateurs auxiliaires de ER $\alpha$ DDX17 est un coactivateur transcriptionnel de Sox2 et MDM2 DDX5 participe à la voie Wnt/ $\beta$ -caténine	(Fuller-Pace et al., 2008) (Alqahtani et al., 2016) (Guturi et al., 2014)
Cancer du poumon non à petites cellules	DDX5/17	↑	Participent à la voie Wnt/ $\beta$ -caténine	(Mooney et al., 2010)
Leucémie/lymphome à cellules T de l'adulte	DDX5/17	↑	DDX17 participe à la voie NF- $\kappa$ B et l'épissage alternatif de CD44 DDX5 participe à la voie Notch	(Ameur et al., 2020) (Lin et al., 2013)
Tumeurs gliales	DDX5/17	↑	DDX5 participe à la voie NF- $\kappa$ B	(Wang et al., 2012)

			DDX17 inhibe l'autophagie des cellules gliales	(Z. Zhang et al., 2016)
Cancer de la prostate	DDX5	↑	DDX5 est un coactivateur transcriptionnel de l'AR	(You et al., 2019)
Cancer du côlon	DDX5	↑	Participe à la voie Wnt/ $\beta$ -caténine et NF- $\kappa$ B	(Khare et al., 2019; Sarkar et al., 2015)
Cancer de l'endomètre	DDX5	↑	Participe à la voie Wnt/ $\beta$ -caténine	(Liu et al., 2019)
Cancer de la thyroïde	DDX5	↑	Stimule la cascade de signalisation de l'ADN DSB	(Yuan et al., 2020)
Leucémie promyélocytaire aiguë	DDX5	↑	Inhibition de la production de ROS	(Wu et al., 2020)
Carcinome épidermoïde de l'œsophage	DDX5	↑	Régulation de l'autophagie cellulaire	(Ma et al., 2020)
Cancer du poumon à petites cellules	DDX5	↑	Favorise le métabolisme respiratoire	(Xing et al., 2020)
Rhabdomyosarcome alvéolaire	DDX5	↑	Stimule l'axe pro-survie EHMT2/PAX3-FOXO1 en stabilisant l'ARN de EHMT2	(Gualtieri et al., 2022)
Carcinome hépatocellulaire	DDX5	↑	Stimule la voie AKT	(Xue et al., 2018)
	DDX5	↓	Régulation de l'autophagie	(Zhang et al., 2019)
Adénocarcinome canalaire pancréatique	DDX5	↓	Une faible expression de DDX5 est associée à un mauvais pronostic	(Morimachi et al., 2021)
Adénocarcinome canalaire pancréatique	DDX17	↑	Participe à l'épissage alternatif de la Caspase 9, mH2A1	(Pan et al., 2021)
Carcinome hépatocellulaire	DDX17	↑	DDX17 est un cosuppresseur de Klf4 et favorise l'expression de PXN-AS1-IR3	(Xue et al., 2019)

**Tableau 3 : Expression et mécanisme lié à DDX5 et/ou DDX17 dans différents cancers**

adapté de (Xu et al., 2022)

Le rôle de DDX5 et DDX17 dans la tumorigenèse passe notamment par leur régulation de nombreuses voies de signalisation. Comme indiqué précédemment, DDX5 et DDX17 sont des coactivateurs des récepteurs nucléaires aux hormones stéroïdiennes comme le récepteur des œstrogènes  $Er\alpha$  et le récepteur des androgènes (AR), expliquant au moins en partie leur rôle dans les cancers du sein et de la prostate (Clark et al., 2008; Li et al., 2022). On peut cependant noter qu'une étude rapporte que la surexpression de *DDX17* est associée à une période plus longue de survie globale et sans rechute dans les cancers du sein ER positifs, ce qui n'est pas le cas de *DDX5* (Wortham et al., 2009). Le même article montre de plus que la déplétion de *DDX17* par siRNA, mais pas de *DDX5*, entraîne

une inhibition significative de la transcription des gènes répondant à l'Era dans des cellules mammaires cancéreuses. DDX5 et DDX17 stimulent aussi la voie Wnt/  $\beta$ -caténine en promouvant la translocation de la  $\beta$ -caténine au noyau et en agissant comme coactivateur de ses gènes cibles (Khare et al., 2019; Wang et al., 2015; Yang et al., 2006). Similairement, DDX5 induit la translocation nucléaire de la sous-unité p50 de NF- $\kappa$ B, promouvant la transcription de ses gènes cibles (Wang et al., 2012).

En plus des facteurs précédemment cités, DDX5 et DDX17 interagissent aussi fonctionnellement avec Notch, mTOR, Smad3, STAT3, Akt, c-Myc, TGF $\beta$ 1 (Xu et al., 2022). Le nombre important de voies de signalisations régulées par les deux hélicases explique qu'une altération de leur expression entraîne des perturbations du fonctionnement et de l'équilibre cellulaire favorisant la tumorigenèse dans de nombreux tissus.

Comme indiqué précédemment, les HCC et PDAC sont les seuls cancers dans lesquels de faibles niveaux d'expression de DDX5 seraient associés à une plus grande malignité du cancer et à un mauvais pronostic. On peut donc se demander ce qui peut expliquer cette inconsistance avec l'identification de DDX5 comme marqueur de la tumorigenèse dans les autres cancers. Le rôle de DDX5 dans la réparation de l'ADN et dans la stimulation de l'autophagie pourrait expliquer qu'une baisse de son expression favorise l'induction tumorale, mais l'on peut se demander pourquoi seulement dans ces deux cancers.

## 6 Le neuroblastome

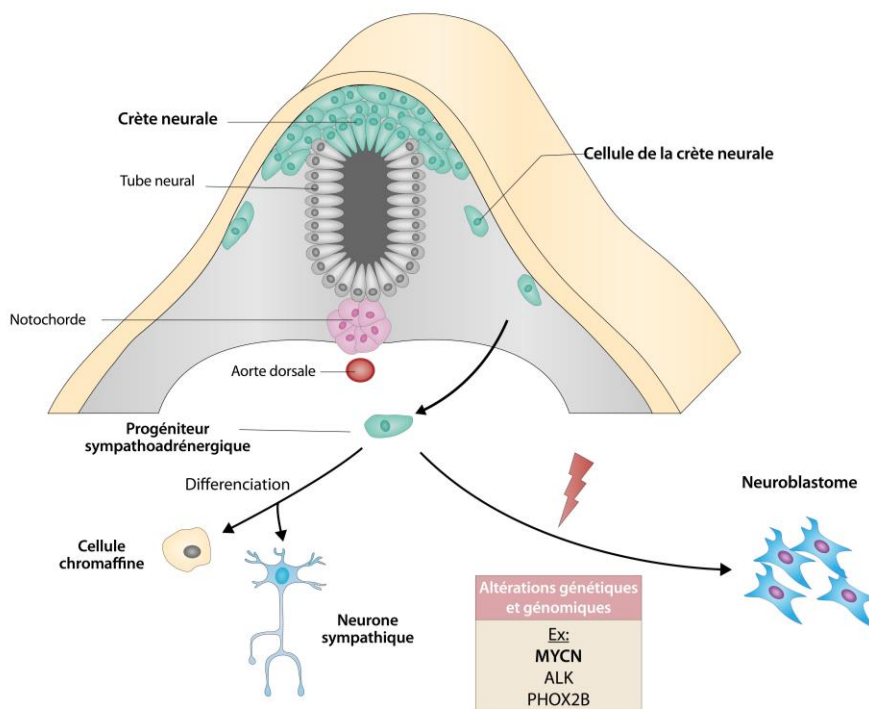
### 6.1 Epidémiologie

Le neuroblastome est un cancer relativement rare, avec moins de 1000 nouveaux cas par an en Amérique du Nord et environ 150 nouveaux cas par an en France. Il s'agit malgré cela de la tumeur solide extra-crânienne la plus fréquente chez les jeunes enfants. Ce cancer touche en effet essentiellement les nourrissons et les jeunes enfants, l'âge médian des patients à leur diagnostic est de 18 mois et 90% des patients sont diagnostiqués avant leurs 10 ans (London et al., 2005). Il est rare que le neuroblastome se développe à l'adolescence ou l'âge adulte, mais il est alors généralement associé à une survie à 5 ans plus faible que dans les groupes d'âge inférieur (Mossé et al., 2014). Le taux de mortalité global de ce cancer a significativement diminué de 1975 à 2005, malgré d'importantes différences entre les différentes classes de risques.



## 6.2 La pathogénèse et le développement du neuroblastome

Le neuroblastome est une tumeur neuro-endocrine du système nerveux sympathique. La majorité des tumeurs sont identifiées au niveau des glandes surrénales, mais on en trouve aussi le long des ganglions de la chaîne sympathique (Matthay et al., 2016). Le neuroblastome se développe à partir des cellules progénitrices du lignage sympathique, dérivant de la crête neurale (Figure 18). La crête neurale correspond à une structure transitoire embryonnaire propre aux vertébrés, se formant suite à l'invagination du tube neural, dont les cellules migrent et se différencient dans divers lignages cellulaires incluant les mélanocytes, les cellules des cartilages et os crânio-faciaux, les neurones sensitifs, entériques et sympathiques. Le neuroblastome est supposé prendre son origine dans un défaut de différenciation du lignage sympatho-adrénergique, qui se différencie normalement en neurones sympathiques et cellules à chromaffine (Zeineldin et al., 2022). Les causes exactes de ce défaut de différenciation sont encore floues, mais il est probable qu'elles soient en lien avec une altération des facteurs de transcription régulant la différenciation des progéniteurs sympatho-adrénergiques.



**Figure 18 : Développement du neuroblastome à partir des cellules de la crête neurale**

Lors du déroulement normal de l'embryogénèse, les cellules de la crête neurale délaminent et migrent pour se différencier en différents lignages cellulaires. Le neuroblastome survient suite à un défaut de différenciation du lignage sympatho-adrénergique, à l'origine des cellules à chromaffine et des neurones sympathiques.

### 6.3 Diagnostic et stratification du neuroblastome

Avant traitement, les patients sont classés en différents groupes de risques. Ce classement permet notamment de guider la prise en charge des patients et leur thérapie. Selon l'INRG (International Neuroblastoma Risk Group), les patients atteints de neuroblastome sont stratifiés en groupes à très faible risque, à faible risque, à risque intermédiaire et à risque élevé, avec une survie sans événement (SSE) à 5 ans de >85%, >75-≤85%, ≥50-≤75%, ou <50% pour chaque groupe, respectivement (Cohn et al., 2009). Parmi les facteurs associés à un pronostic défavorable et un risque élevé se trouvent entre autres un âge au moment du diagnostic supérieur à 18 mois, la perte du chromosome 11q et l'amplification de l'oncogène *MYCN*.

Pour les groupes à très faible et faible risque, le traitement se résume généralement à la résection chirurgicale, voire à la simple observation pour les patients de moins de six mois. Le neuroblastome présente en effet la caractéristique de pouvoir spontanément régresser chez certains patients (Brodeur, 2018). Les mécanismes moléculaires à l'origine de cette régression spontanée sont mal compris, mais pourraient être liés à la réponse immunitaire anti-tumorale de l'hôte, la perte de l'activité télomérase et la dépendance des cellules tumorales aux facteurs de croissance neurotrophiques.

Pour les groupes à risque intermédiaire, le traitement peut uniquement consister en une observation ou en la résection chirurgicale pour les nourrissons de moins de six mois. La chimiothérapie est utilisée en complément de la résection chirurgicale chez les autres patients, et peut être complétée par de la radiothérapie pour les tumeurs ne répondant pas suffisamment aux traitements.

Pour les groupes à haut-risque, le traitement est habituellement divisé en trois phases : induction, consolidation et maintenance (Smith and Foster, 2018). Le traitement d'induction consiste en une chimiothérapie intensive et en une résection chirurgicale. La phase de consolidation peut inclure de la radiothérapie ou l'administration de chimiothérapie à haute dose suivie de la transplantation de cellules souches hématopoïétiques myéloablatives. La phase de maintenance peut inclure de l'immunothérapie combinée, constituant en l'administration d'anticorps monoclonaux en combinaison avec de l'isotétrinoïne et/ou du GM-CSF (Lode et al., 2023)

## 6.4 Altération génétique et génomique du neuroblastome

Les mutations ponctuelles sont rarement identifiées au moment du diagnostic du neuroblastome. Les plus communément rapportées sont des mutations ponctuelles du récepteur de tyrosine kinase *ALK* et du facteur de transcription *PHOX2B*, dont les mutations germinales sont à l'origine de la majorité des cas familiaux de neuroblastomes (Janoueix-Lerosey et al., 2008; Mossé et al., 2008; Tolbert et al., 2017) (qui représentent environ 1-2% des cas). Des mutations de *ALK* sont aussi retrouvées dans 7-12% des neuroblastomes sporadiques (Chen et al., 2008).

Les altérations résultantes de réarrangements génomiques sont par contre communes et souvent associées aux tumeurs dites à haut-risque. Plusieurs anomalies segmentaires sont ainsi associées au neuroblastome, correspondant à la perte ou le gain de fragments chromosomiques. Parmi les plus fréquentes, le gain du bras long du chromosome 17 (17q) a été observé dans plus de la moitié des cas de neuroblastome (Bown et al., 2001) et la perte du bras court du chromosome 1 (1p) et du bras long du chromosome 11 (11q) ont été observées dans respectivement 20% et 30% des cas (Attiyeh et al., 2005).

L'amplification de *MYCN* est aussi une altération fréquente du neuroblastome, retrouvée dans 20% des tumeurs. Les fonctions de ce facteur de transcription en contexte physiologique et dans le neuroblastome sont décrites plus en détails dans la partie suivante.

## 6.5 MYCN

### 6.5.1 MYCN dans le neuroblastome

Environ 20% des neuroblastomes présentent une amplification de l'oncogène *MYCN* au niveau du chromosome 2p24 au diagnostic. Cette amplification, conduisant à une augmentation du nombre de copies du locus *MYCN*, est fortement associée avec un stade cancéreux avancé et un pronostic défavorable (Brodeur et al., 1984).

*MYCN* code pour le facteur de transcription MYCN, aussi nommé N-Myc, faisant partie de la famille MYC, qui comprend aussi *c-MYC* et *MYCL*. Contrairement à *c-MYC*, qui est exprimé de façon ubiquitaire et persiste à l'âge adulte, l'expression de *MYCN* est essentiellement restreinte au système nerveux ainsi qu'aux tissus mésenchymateux

durant le développement embryonnaire (Zimmerman et al., 1986). *MYCN* est ainsi fortement exprimé dans les cellules de la crête neurale, son expression étant importante pour la migration et la prolifération de celles-ci. L'expression de *MYCN* diminue ensuite progressivement lors de la différenciation des neurones sympathiques (Zimmerman et al., 1986). Des études chez le poisson zèbre ont montré que l'expression forcée de *MYCN* dans les précurseurs sympatho-adrénergiques bloque le développement des cellules à chromaffine, entraînant le développement de neuroblastomes (Zhu et al., 2012). De plus, des souris transgéniques où l'ADNc *MYCN* humain est placé sous contrôle du promoteur de la tyrosine hydroxylase (Th-*MYCN*) développent spontanément des neuroblastomes qui sont morphologiquement et phénotypiquement similaires aux neuroblastomes humains à haut risque (Weiss et al., 1997). Cela suggère que la surexpression de *MYCN* dans les cellules de la crête neurale de façon prolongée est suffisante pour le développement de neuroblastomes.

De par son rôle de facteur de transcription favorisant la prolifération et la migration des précurseurs sympatho-adrénergiques, la surexpression de *MYCN* favorise la tumorigenèse en modulant de multiples voies (Huang and Weiss, 2013). *MYCN* peut activer la transcription de gènes impliqués dans les métastases, la survie cellulaire, la prolifération, la pluripotence, l'auto-renouvellement et l'angiogenèse. De plus, *MYCN* peut réprimer l'expression de gènes qui favorisent la différenciation, l'arrêt du cycle cellulaire, la surveillance immunitaire et les gènes réprimant la métastase et l'angiogenèse.

*MYCN* semble donc constituer une cible thérapeutique idéale pour le traitement des neuroblastomes à haut-risque. Toutefois, les différentes tentatives faites pour développer des inhibiteurs de *MYCN* n'ont pas permis d'aboutir à un traitement suffisamment fiable et efficace (Fletcher et al., 2018). Les difficultés rencontrées sont à la fois inhérentes à celles rencontrées pour l'inhibition des facteurs de transcription en général, auxquelles s'ajoute aussi le manque de sites d'interactions spécifiques de la protéine, ainsi que la forte homologie entre les protéines de la famille *MYC* (Bartolucci et al., 2022). De nouvelles stratégies sont actuellement développées, la majorité ciblant indirectement *MYCN* en inhibant ses cofacteurs, des protéines favorisant son expression ou la stabilité de sa protéine. De récentes avancées techniques ont relancé les efforts dans le ciblage direct de *MYCN*, notamment en ciblant son ARNm ou son gène, par

exemple via l'utilisation d'oligonucléotides appelés antigène peptide nucleic acid (agPNA) spécifique à *MYCN* qui permettent de bloquer sa transcription (Tonelli et al., 2005).

#### 6.5.2 *MYCN* comme facteur de transcription

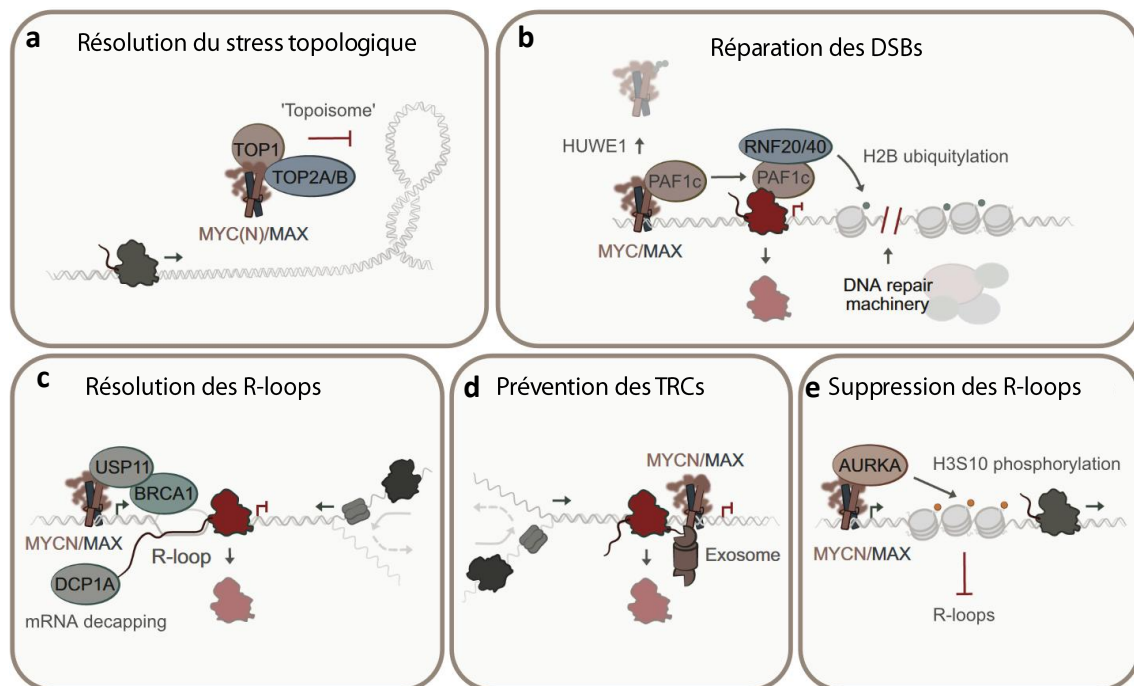
Comme les autres membres de la famille MYC, *MYCN* est un facteur de transcription possédant un domaine hélice-boucle-hélice (bHLH), lui permettant de se fixer à l'ADN au niveau de séquences consensus appelées E-box, dont le motif canonique est « CACGTG ». *MYCN* se fixe à ces E-box sous forme d'hétérodimère avec la protéine MAX. Plus précisément, le domaine de régulation transcriptionnelle N-terminal de *MYCN* contient deux régions connues sous le nom de Myc Homology Box (MB), MBI et MBII. La région centrale contient trois autres MB ainsi qu'un signal de localisation nucléaire. La région basique (BR) C-terminale est impliquée dans la liaison à l'ADN, tandis que le HLH permet l'hétérodimérisation avec Max.

*MYCN* active la transcription en recrutant de multiples co-activateurs et complexes protéiques au niveau de la E-box. Le complexe acétyltransférase TIP60 et l'histone acétyltransférase GCN5 se lient aux protéines MYC indirectement par l'intermédiaire de la protéine adaptatrice TRRAP (Cheng et al., 2023; Liu et al., 2003). Deux autres protéines, TIP48 et TIP49, présentes dans le complexe TIP60 et impliquées dans le remodelage de la chromatine, se lient à l'extrémité N-terminale de *MYCN*. *MYCN* interagit par ailleurs avec la déméthylase H3K9me3/me2 KDM4B et, lorsqu'il est surexprimé, *MYCN* recrute KDM4B aux E-box, diminuant ainsi les niveaux de H3K9me3. Par le recrutement de ces complexes, *MYCN* permet le maintien de la chromatine dans un état actif. De plus, *MYCN* interagit également avec TFIIC et RAD21 pour permettre à la Pol II de sortir de son état de pause et de passer à une activité d'élongation transcriptionnelle (Büchel et al., 2017). Par ailleurs, BRCA1 interagit avec *MYCN* et permet à ce dernier de réprimer la formation de R-loops dans les régions proximales du promoteur, ce qui renforce l'activité transcriptionnelle (Herold et al., 2019).

*MYCN* peut aussi fonctionner comme un répresseur transcriptionnel en se fixant indirectement à l'ADN en interagissant avec SP1 et MIZ1, induisant le recrutement de HDAC1 (Gherardi et al., 2013). Par ailleurs, *MYCN* s'associe à EZH2, une méthyltransférase membre du complexe PRC2 (Polycomb Repressive Complex 2), pour réprimer la transcription des gènes (Corvetta et al., 2013).

### 6.5.3 Rôles indépendants de l'initiation de la transcription

En plus de son rôle classique de facteur de transcription, plusieurs travaux ont plus récemment pu mettre en évidence le rôle de MYCN dans d'autres fonctions liées à la transcription, comme de permettre de soulager le stress transcriptionnel (Papadopoulos et al., 2023) (Figure 19).



**Figure 19 : Mécanismes de résolution des stress associés à la transcription par le biais de MYC/MYCN**

MYC et MYCN interviennent contre les stress associés à la transcription par plusieurs mécanismes. (a) MYC et MYCN mobilisent les topoisomérases 1 (TOP1) et 2 (TOP2A/B) dans un "topoisome" et stimulent leurs activités pour résoudre les super-enroulements d'ADN et le stress de torsion. (b) L'ubiquitylation de MYC par HUWE1 permet le transfert de PAF1C de MYC à la Pol II. PAF1C interagit ensuite avec RNF20/40, qui mono-ubiquitine l'histone H2B, permettant ainsi la réparation des cassures doubles brins. (c) MYCN s'associe via USP11 avec BRCA1, permettant la résolution des R-loops proches du promoteur. Le complexe MYCN-USP11-BRCA1 recrute ensuite des enzymes de décoiffage de l'ARNm, dont DCP1A, qui enlèvent la coiffe 5' de la transcription naissante, ce qui entraîne l'éviction de la Pol II de la chromatine et la terminaison de la transcription. (d) MYCN s'associe à l'exosome, qui dégage la Pol II de la chromatine en dégradant le transcrit associé à partir de l'extrémité 3'. La transcription est ainsi interrompue, ce qui réduit l'occurrence des conflits transcription-réplication (TRCs). (e) MYCN forme un complexe avec AURKA qui phosphoryle H3S10, augmentant ainsi l'incorporation de H3.3 dans la chromatine et empêchant la formation de R-loops sur les sites promoteurs liés par MYCN. (Papadopoulos et al., 2023)

La transcription peut induire plusieurs types de stress, comme la torsion de l'ADN, la formation de R-loops, ainsi que des conflits avec la machinerie de réplication (Hamperl et al., 2017; Teves and Henikoff, 2014). Ces stress peuvent entraîner une diminution de la transcription ainsi que l'apparition de cassures doubles brins (DSB), pouvant induire la mort cellulaire dans des cellules en forte prolifération (Hamperl et al., 2017).

Les protéines MYC pouvant virtuellement se fixer à tous les promoteurs et leur surexpression entraînant l'entrée précoce en phase S de réplication (Macheret and Halazonetis, 2018), il apparaît donc nécessaire d'avoir des mécanismes permettant de prendre en charge le stress induit par leur activité transcriptionnelle.

L'activité transcriptionnelle importante induite par MYC et MYCN induit ainsi un stress topologique se traduisant notamment par des supertours positifs ou négatifs de l'ADN. Ces supertours peuvent être résolus grâce à l'activité de topoisomérases. MYCN interagit avec les topoisomérases TOP1 et TOP2B pour former un complexe appelé « toposome », stimulant l'activité de TOP1 et TOP2B (Das et al., 2022). Cela permet d'augmenter l'engagement de ces topoisomérases au niveau des promoteurs et enhancers liés par MYCN et leur résolution du stress topologique, permettant d'augmenter la transcription liée à MYCN.

MYCN permet aussi de prévenir les conflits transcription-réplication en libérant la Pol II en état de pause ou de blocage de la chromatine grâce au recrutement de BRCA1 par l'intermédiaire de l'enzyme de déubiquitination USP11 (Herold et al., 2019). La formation d'un complexe MYCN-BRCA1-USP11 permet l'élimination des R-loops au niveau des promoteurs et engage des complexes de décoiffage des ARNm comprenant DCP1A, DCP2 et EDC4. Ces complexes retirent la coiffe 5' du transcrit naissant et entraînent l'éviction de la Pol II de l'ADN, mettant fin à la transcription (Herold et al., 2019). Cela permet d'éliminer les sources locales de stress et de libérer l'espace pour une nouvelle transcription.

Par ailleurs, MYCN interagit avec l'exosome et le recrute au niveau des promoteurs, permettant d'éviter l'accumulation de fourches de réplication immobilisées et de DSBs (Papadopoulos et al., 2022). Ceci permettrait d'assurer une élongation transcriptionnelle productive (non interrompue et permettant la production efficace d'ARNm). La déplétion

de la sous unité catalytique EXOSC10 de l'exosome entraîne en effet la rétention de la Pol II dans le corps des gènes, associée à une diminution significative des niveaux de transcrits naissants indiquant une transcription improductive (Papadopoulos et al., 2022).

MYCN favoriserait ainsi la terminaison transcriptionnelle promoteur-proximale pour prévenir les conflits transcription-réplication lors d'évènements de transcription non optimaux. La terminaison transcriptionnelle médiée par l'exosome n'est cependant pas constatée pour tous les gènes liés par MYCN. Durant la phase S, MYCN interagit directement avec la kinase AURKA (Richards et al., 2016). MYCN et AURKA incorporent de manière coopérative l'histone H3.3 dans les sites promoteurs par le biais de la phosphorylation S10 de l'histone H3.3 induite par AURKA. L'inhibition de l'AURKA freine l'incorporation de H3.3, et entraîne une accumulation de Pol II ainsi qu'une augmentation des R-loops au niveau des promoteurs liés par MYCN (Roeschert et al., 2021). Cette situation entraîne à son tour des dommages généralisés à l'ADN, l'activation de la voie ATR (qui indique un stress de réplication), une réduction de la progression de la fourche de réplication et une augmentation des conflits entre la transcription et la réplication. En limitant la formation de R-loops, le complexe MYCN-AURKA permet de prévenir le stress lié à la réplication. Les auteurs font l'hypothèse que l'incorporation de H3.3 prévient la formation de R-loops. Ils observent néanmoins que l'élimination des R-loops augmente l'incorporation de l'histone H3.3 en aval des sites de liaison de MYCN, et le mécanisme par lequel l'incorporation de H3.3 pourrait prévenir la formation de R-loops n'est pas étudié.

On peut par ailleurs noter qu'un papier de la même équipe indique que AURKA permettrait d'inhiber la sortie de pause de la Pol II dépendante de MYCN (Büchel et al., 2017), ce qui pourrait aussi expliquer en partie la réduction de R-loops et des conflits transcription-réplication, puisque la transcription serait réduite. De plus, AURKA stabilise la protéine MYCN (Otto et al., 2009), ce qui pourrait favoriser son interaction avec BRCA1 et les facteurs de terminaison (Herold et al., 2019). AURKA pourrait donc potentiellement agir contre le stress transcriptionnel à la fois en limitant la formation de R-loops via l'incorporation de H3.3, limitant ainsi la terminaison de la transcription, mais aussi en stabilisant MYCN, favorisant le recrutement de BRCA1 et de facteurs de terminaison, permettant l'arrêt de la transcription en cas de condition défavorable. Par



ces deux mécanismes ayant un effet opposé sur la terminaison de la transcription promoteur-proximale, AURKA permettrait de diminuer les conflits transcription-réplication.

# Problématique

Bien qu'étant moins étudiée que les étapes d'initiation de la transcription ou d'épissage, l'étape de maturation de l'extrémité 3' des ARNm, étroitement couplée à celle de la terminaison transcriptionnelle, est elle aussi hautement régulée. Son bon déroulement est nécessaire afin de produire des transcrits matures et stables et elle participe à la diversité du transcriptome à travers la polyadénylation alternative. Une perturbation de cette étape est susceptible d'entraîner un readthrough transcriptionnel, phénomène qui est observé lors de divers stress cellulaires, ainsi que lors d'infections virales et dans cancers. Le readthrough transcriptionnel peut moduler le transcriptome en affectant l'expression et la traduction du transcrit canonique, mais aussi des transcrits adjacents, lorsque le readthrough se prolonge jusqu'au gène en aval de celui-ci. On observe alors la production de transcrits dits chimériques, portant l'information de deux gènes adjacents sur un même brin d'ARN. Certains de ces transcrits chimériques sont enrichis dans certains tissus cancéreux et possèdent une activité protumorale.

Les hélicases DDX5 et DDX17 sont des régulateurs de la terminaison transcriptionnelle dont la déplétion entraîne le readthrough d'environ 800 gènes dans des cellules de neuroblastomes (Terrone et al., 2022). Le neuroblastome est un cancer pédiatrique du système nerveux périphérique, dont les patients classés dans les groupes à haut risque, dont la tumeur présente une amplification de l'oncogène *MYCN*, présentent une survie à 5 ans inférieure à 50%. L'analyse de l'expression de DDX17 dans les tumeurs d'une cohorte de 498 patients révèle qu'une faible expression de DDX17 est fortement associée à une survie à 5 ans plus faible. De plus, de façon surprenante, une analyse de CHIP-seq de *MYCN* montre un enrichissement de *MYCN* au niveau des sites de terminaison de la transcription régulés par DDX5 et DDX17.

**L'objectif premier et principale de mes travaux de thèse était donc de caractériser l'impact du readthrough induit par la déplétion de DDX5 et DDX17 dans un contexte cancéreux, le neuroblastome. Celui-ci s'est notamment concentré sur l'étude de la production de transcrits chimériques ainsi que de l'effet de *MYCN* sur le readthrough transcriptionnel.**

Par ailleurs, des mutations du gène *DDX17* ont été identifiées chez plusieurs patients souffrant de troubles neuro-développementaux, suggérant un possible rôle de cette hélicase dans le développement et le fonctionnement du système nerveux. Il s'agit des premiers cas où une mutation de *DDX17* pourrait être liée à une maladie génétique.

**L'objectif secondaire de mes travaux de thèse était ainsi de caractériser l'effet d'une perte d'expression de *DDX17* dans des cellules neuronales ainsi que de caractériser certains des mutants identifiés chez des patients souffrant de troubles neuro-développementaux.**

# Partie 1

Le neuroblastome est le cancer solide extracrânien pédiatrique le plus fréquent. Il s'agit d'une tumeur neuroendocrine embryonnaire, issue des cellules progénitrices de la crête neurale. Le neuroblastome présente une hétérogénéité clinique importante, allant de la régression spontanée à une métastase généralisée. Bien que des augmentations globales de la survie sans séquelles à 5 ans aient été rapportées, les patients classés à haut risque présentent une survie à 5 ans encore inférieure à 50% (Johnsen et al., 2019). Les groupes à haut risque sont associés à l'amplification de l'oncogène *MYCN*. L'analyse de l'expression de *DDX5* et *DDX17* dans les tumeurs d'une cohorte de 498 patients révèle qu'une faible expression de *DDX17*, ainsi que de *DDX5* dans une moindre mesure, est fortement associée à une survie à 5 ans plus faible. L'expression de *DDX17* apparaît de plus être significativement plus faible dans les tumeurs à haut risque, et elle est inversement corrélée à l'amplification de *MYCN*. Nous avons cherché à savoir si la corrélation d'une faible expression de *DDX17* avec un pronostic défavorable dans les neuroblastomes pourrait être due à la perturbation de la maturation de l'extrémité 3' des transcrits régulés par *DDX5/17*. L'analyse RNA-seq de cellules de neuroblastomes révèle en effet que la déplétion de *DDX5* et *DDX17* par siARN entraîne en effet un readthrough transcriptionnel résultant en la production de transcrits anormalement allongés à leur extrémité 3', ainsi que de transcrits chimériques résultant de la transcription continue de deux gènes adjacents. De plus, mes résultats montrent que la surexpression de *MYCN* augmente l'expression de plusieurs de ces transcrits, suggérant que *MYCN* favorise le readthrough transcriptionnel de ces gènes. *MYCN* interagit avec *DDX17* et se fixe fréquemment à proximité des séquences régulées par *DDX5* et *DDX17*, suggérant que la présence de *MYCN* perturbe la régulation de la maturation en 3' des ARNm par les hélicases. Enfin, le recrutement forcé de *MYCN* au niveau de l'extrémité 3' de *MAPK7*, un gène produisant le transcrit chimérique *MAPK7-RNF112* lorsque *DDX5* et *DDX17* sont déplétées ou *MYCN* surexprimée, induit une augmentation de l'expression de la forme chimérique. Cela suggère que la présence de *MYCN* à proximité du PAS d'un gène suffit à induire un readthrough transcriptionnel. Ces travaux suggèrent donc que *MYCN* perturbe directement la maturation de l'extrémité 3' des gènes, ce qui constitue un nouveau rôle pour ce facteur de transcription

Ma contribution à ces travaux a été la suivante :

Figure 1 :

-RT-qPCR de validation des observations faites à partir du RNA-seq (présence de transcrits chimériques)

Figure 4 :

-Western blot de l'expression de MYCN et de DDX5 et DDX17

-RT-qPCR de l'expression de *MYCN* et de *DDX5* et *DDX17*

Figure 5 :

-Western blot de validation de la déplétion par siRNA de MYCN ou de sa surexpression par transfection transitoire

-RT-qPCR validant l'effet de la surexpression et de la déplétion de MYCN sur l'expression de plusieurs transcrits readthrough et transcrits chimériques

Figure 6 :

-ChIP-qPCR validant la fixation de MYCN à proximité de l'extrémité 3' des gènes régulés par DDX5/17

-Western blot de l'immunoprécipitation de MYCN avec DDX17 et DDX5

Figure 7 :

-ChIP-qPCR validant le recrutement de dCas9MYCN et dCas9CT au niveau du PAS de *MAPK7*

-RT-qPCR de l'expression de *MAPK7* et de *MAPK7-RNF112* suite au recrutement de dCas9MYCN ou dCas9CT

Figure S2 :

-quantification RT-qPCR de l'effet du siDDX5/17 sur l'expression des transcrits des gènes parentaux amonts (via à vis des transcrits chimériques)

Figure S4 :

-quantification RT-qPCR de l'effet de la surexpression ou la déplétion par siRNA de MYCN l'expression des transcrits des gènes parentaux amonts (vis-à-vis des transcrits chimériques et readthrough)

# The oncoprotein MYCN antagonizes the effect of helicases DDX17 and DDX5 and stimulates the production of chimeric transcripts

Valentine Clerc<sup>1</sup>, Jessica Valat<sup>1</sup>, Xavier Grand<sup>1</sup>, Nicolas Fontrodona<sup>1</sup>, Lavinia Attili<sup>1</sup>, Elisa Polewiak<sup>1</sup>, Benjamin Gibert<sup>2</sup>, Nicolas Rama<sup>2</sup>, Cyril F. Bourgeois<sup>1</sup>

<sup>1</sup> Equipe labellisée "La Ligue", Laboratoire de Biologie et Modélisation de la Cellule, Ecole Normale Supérieure de Lyon, CNRS, UMR 5239, INSERM U1293, Université Claude Bernard Lyon 1, Lyon, France.

<sup>2</sup> Apoptosis, Cancer and Development Laboratory- Equipe labellisée "La Ligue", LabEx DEVweCAN, Institut Convergence PLAsCAN, Centre de Recherche en Cancérologie de Lyon, INSERM U1052-CNRS 5286, Université Claude Bernard Lyon 1, Lyon, France.

Corresponding author: Cyril F. Bourgeois ([cyril.bourgeois@inserm.fr](mailto:cyril.bourgeois@inserm.fr)).

## ABSTRACT

DEAD box helicases DDX17 and DDX5 control the correct termination of transcription and the associated cleavage of the 3' end of transcripts. Here we show that the transcriptional readthrough induced by their depletion in neuroblastoma cells also results in an enhanced production of chimeric transcripts from tandem genes. An analysis of neuroblastoma tumours, in which chimeric transcripts are abundant, revealed that a low expression of DDX17 and DDX5 genes is associated with a poor 5-year survival of patients. A low DDX17 expression is also significantly associated with high risk tumours and it is inversely correlated with MYCN oncogene amplification, suggesting a link between those two factors. Surprisingly, changes in MYCN expression do not affect the expression of both helicases, but they impact transcription termination and chimeric transcript production. We provide evidence that MYCN directly promotes transcriptional readthrough by directly binding to the 3' regions of genes, and that it interacts with DDX17, suggesting that it may disrupt alter the activity of the helicase. Altogether our work reveals a new function of MYCN on transcription termination and suggests that the combined deregulation of MYCN and DDX17 expression could lead to readthrough-associated transcriptome alterations in neuroblastoma.

## INTRODUCTION

During transcription by RNA polymerase II (RNAPII) in eukaryotes, the nascent RNA undergoes different processing steps, including 5' capping, splicing, 3' end cleavage and addition of the poly-A tail. These steps are mostly co-transcriptional and are impacted not only by RNA binding proteins, but also by transcription and chromatin-associated factors [1,2]. The tight connection that exists between transcription termination and 3' end RNA processing is archetypal of this coordination of events. The slow down and release of RNAPII after the transcription of the polyadenylation site (PAS) and the cleavage of the nascent transcript are mutually dependent on each other, in a complex interplay [3–6]. The regulation of this process involves a number of factors that act on chromatin, at the level of the RNAPII complex or on the RNA molecule, or which can have an effect at several levels.

An alteration in the expression or activity of factors involved in transcription termination and/or RNA cleavage classically results in a transcriptional readthrough and in the extension of the 3' extremity of the transcript. Readthrough transcription has been observed in cells exposed to various stresses or viral infection, as well as in cancer [7–15], as recently reviewed [16,17]. The fate of these extended transcripts is uncertain, but it has been shown that stress-induced DoGs (downstream of gene containing transcripts) lack a defined endpoint and are non-coding RNAs which remain associated to the chromatin in the nucleus [12,13].

In some cases, RNAPII that fails to terminate past the PAS can invade the downstream gene on the same DNA strand, generating *cis*-spliced transcripts containing exons from both genes. These RNA molecules have been called transcription-induced chimeras [18,19], tandem RNA chimeras [20], products of conjoined genes [21,22], readthrough fusions [23] or *cis*-splicing products of adjacent genes [24,25]. Below we will simply define them as chimeric transcripts, keeping in mind that in this study we are not considering RNA molecules resulting from chromosomal rearrangements. The formation of chimeric transcripts can be associated to splicing defects in the adjacent invaded gene, highlighting the necessary coordination of events that ensures the correct and autonomous expression of neighbouring genes [26,27]. Thanks to the increasing number and deeper analysis of RNA-seq experiments, chimeric transcripts have been identified in many types of cancers and they were associated with oncogenesis [28,29]. However, they are also found in normal tissues [30] and their production is considered as a possible mechanism to increase protein diversity [19]. Indeed, chimeric transcripts can retain a functional open reading frame (ORF) allowing the production of chimeric or fusion proteins [23,31–33].

Several factors have been shown to affect the formation of chimeric transcripts in human cells, including splicing and/or polyadenylation factors [34]. Chemical splicing inhibition can repress the formation of chimeric transcripts in neuroblastoma, a cancer in which most fusions are

intrachromosomal, likely resulting from transcriptional readthrough and *cis*-splicing [35]. Chromatin-associated factors are also involved in chimeric RNA production, such as the histone methyltransferase SETD2 [8] or the CCCTC-binding factor CTCF [24,25], which plays important roles in the spatial genome organization and gene expression regulation via chromatin looping or insulation [36]. Recent reports further supported a function of CTCF in controlling PAS choice and transcription termination [37,38]. DEAD box ATP-dependent RNA helicase DDX17 and its paralog DDX5 have a variety of functions in gene expression, especially in transcription and RNA processing [39,40]. Thanks to their helicase activity, they act by modulating local RNA secondary structures or DNA/RNA structures such as R-loops and G-quartets, impacting promoter and exon choice [41–45]. Recently, we and others have shown that DDX17 and DDX5 control the 3' end processing of RNAPII transcripts, and that their knockdown leads to a decreased termination downstream of the expected PAS [38,42,43,46]. Although an impact on R-loops or 3'UTR RNA structure has been described [42,43,46], we showed that this function is linked to the binding of CTCF near the PAS and to the specific 3D organization of their target genes, and we hypothesized that the helicases could control the dynamics of transcription and RNA processing across DNA and/or RNA structured regions [38].

Here, we report that the 3' end processing defect induced by *DDX17* and *DDX5* silencing in neuroblastoma cells results in the production of several hundreds of chimeric transcripts, some of which are also found in aggressive neuroblastoma tumours. Interestingly, we found that *DDX17* expression correlates with a good prognostic factor in neuroblastomas and that it is inversely correlated to the expression of *MYCN*, the main oncogene in this cancer. We present evidence supporting a direct role of *MYCN* in controlling transcription termination through a direct binding near the PAS of helicase-targeted genes. Altogether our data reveal a new function of *MYCN* in transcription and suggest that the combined overexpression of *MYCN* and suboptimal expression of *DDX17* could be determinant factors to explain the increased expression of chimeric transcripts in aggressive neuroblastomas.

## RESULTS

### **DDX17 and DDX5 depletion enhances the expression of chimeric transcripts**

We recently described a genome-wide effect of *DDX17/DDX5* depletion on transcription termination in neuroblastoma cells SH-SY5Y cells [38]. While analysing our RNA-seq results, we observed several examples of si*DDX17/DDX5*-induced transcriptional readthrough that also displayed splicing junction reads across adjacent genes. Typically, splicing skipped the terminal exon of the altered gene and linked its penultimate exon to the second exon of the following gene, generating a chimeric mRNA molecule (Figure 1A, Supplementary Figure S1A). However, we also frequently observed complex



alternative splicing patterns across exons from both genes (Supplementary Figure S1B). Interestingly, in some cases chimeric junctions linked more than 2 adjacent genes, suggesting that DDX17/DDX5 depletion could induce the formation of multimeric mRNA molecules (*NTRK1-PEAR1-LRRC71* and *NDUFA13-YJEFN2-CILP2*, Supplementary Figure S1B).

We next used the previously described Arriba algorithm [47] to determine more precisely the number of chimeric transcripts whose expression is modified upon DDX17/DDX5 depletion. We identified 282 chimeric mRNA, 77% of which (218) were upregulated upon DDX17/DDX5 silencing (Figure 1B, top panel, Table S1). Applying a more stringent threshold of selection confirmed that the depletion of both helicases indeed mostly increased the expression of chimeric transcripts (Figure 1B, bottom panel), indicating that this phenomenon was likely a side effect of the transcriptional readthrough phenotype we described earlier [38]. In fact, a large proportion of the genes displaying an induced chimeric transcript had been identified in our previous study (Figure 1C). From those two analyses, we defined a list of 953 genes whose termination was inhibited upon depletion of DDX17 and DDX5 (Table S1).

We next sought to experimentally validate the expression of a selection of chimeric transcripts. We first carried out RT-PCR experiments using a primer overlapping the RNA fusion between both genes, to specifically detect the chimeric transcript, that we systematically compared to the canonical form of the gene. We observed a clear increase of all tested chimeric transcripts upon DDX17/DDX5 depletion, which did not match with the expression of the canonical mRNA (Supplementary Figure S2A). We confirmed this result by RT-qPCR on a panel of chimeric transcripts, which exhibited a normalized expression increased from 2 to 40 fold in absence of DDX17/DDX5, compared to the canonical transcript (Figure 1D and Supplementary Figure S2B). Note that the validated *CTSD-IFITM10* and *NDUFA13-YJEFN2* chimeric RNAs were identified while browsing through our RNA-seq data (Supplementary Figures S1A and S1B), but were not predicted by Arriba. Besides, the *TRIM3-HPX* transcript, which was strongly induced by DDX17/DDX5 depletion, did not reach the significance threshold we had set in Arriba. This means the number of chimeric transcripts is probably underestimated, but we preferred not to enlarge our prediction further, to avoid including false positive events.

### **Chimeric transcripts can produce chimeric proteins**

Some chimeric transcripts maintain a coding sequence that matches the reading frame of their respective canonical transcripts, allowing them to avoid a likely degradation by the nonsense-mediated mRNA decay pathway, which is triggered if a frameshift creates a premature stop codon. For example, the *CTSD-IFITM10* chimera lacks the last exon of *CTSD* and the first exon of *IFITM10*, but it is cleaved and polyadenylated at the canonical PAS of the *IFITM10* gene. This should result in the production of

a putative chimeric protein of 557 aminoacids (60 kDa), that lacks 55 and 28 aminoacids from the C-terminal part of CTSD and N-terminal part of IFITM10 proteins, respectively (Figure 2A).

We carried out a western-blot analysis to detect this new protein, using an antibody against CTSD. The gene normally encodes a unique precursor that is processed into two peptidic chains that compose Cathepsin D. As expected, we identified the two subunits between 25 and 45 kDa (Figure 2B), but we also detected a discrete 60 kDa band that appeared only upon depletion of DDX17 and DDX5. All bands disappeared when cells were treated with a siRNA targeting the body of the *CTSD* transcript. However, treating cells with another siRNA specifically targeted to the *CTSD-IFITM10* junction more specifically reduced the 60 kDa band, while other bands remained visible. This result demonstrated that the siDDX17/DDX5-induced chimeric RNA can produce chimeric proteins, that may potentially change dramatically the functions of both canonical proteins.

### **The expression of the *DDX17* gene is reduced in high-risk neuroblastoma tumours**

Chimeric transcripts were recently identified in large amounts in the most aggressive neuroblastomas [35]. Among the 282 DDX17/DDX5-regulated chimeric transcripts from our analysis, 51 were also found in tumours, mostly coming from the pool of upregulated transcripts (Figure 3A, top panel). This result was confirmed when focusing on the most significant chimers, as 27% of upregulated transcripts but none of the downregulated transcripts were also found in tumours (Figure 3A, bottom panel).

A few years ago, we showed that DDX17 plays an important role in the early phases of differentiation of neuroblastoma cells, upon induction by retinoic acid [48]. More recently a high expression DDX5 was associated with an unfavourable outcome in neuroblastoma patients [49]. To explore further the link between the two helicases and this cancer, we analysed the expression of *DDX17* and *DDX5* genes in the clinically annotated SEQC neuroblastoma cohort [50]. Interestingly, a better survival probability of patients significantly correlated with a high expression of *DDX17* (Figure 3B). A low *DDX17* level was also significantly associated with the most aggressive stage 4 tumours (classified according to the International Neuroblastoma Staging System INSS) and with high-risk tumours (Figures 3C and 3D). Finally, in agreement with the driving role of *MYCN* in neuroblastoma, *DDX17* expression was inversely correlated with the amplification of this oncogene (Figure 3E). The same analysis showed a more contrasted result for *DDX5*, but a low *DDX5* expression was also found to be correlated with a lower survival probability and with *MYCN* amplification (Supplementary Figure S3).

In conclusion, those results established a link between the expression of *DDX17*, and to a lower extent of *DDX5*, and the severity of neuroblastoma. This raised the interesting possibility that a suboptimal

expression of DDX17 and DDX5, which alters transcription termination of hundreds of genes, could contribute to the overexpression of chimeric transcripts in aggressive neuroblastomas.

### **MYCN does not impact the expression of DDX17 and DDX5**

The inverse correlation between *MYCN* gene amplification and a lower expression of both *DDX17* and *DDX5* genes (Figure 3E and Supplementary Figure S3) suggested that MYCN could control the expression of both helicases. To test this hypothesis, we modulated the expression of MYCN in two different neuroblastoma cell lines, that display or not an amplification of *MYCN*.

First, we transfected SH-SY5Y cells with increasing amounts of a HA-tagged MYCN encoding plasmid to mimic MYCN overexpression. As expected, MYCN protein was barely detectable in control cells and strongly induced upon transfection, but this did not result in any significant change in DDX17 and DDX5 protein levels (Figure 4A). At the mRNA level, DDX17 expression was not impacted either, while *DDX5* expression was moderately increased at the two lowest MYCN concentrations, but not at maximal MYCN expression (Figure 4B).

In parallel, we used the *MYCN*-amplified SK-N-BE(2) cell line to downregulate MYCN expression with a mixture of specific siRNAs. This treatment resulted in a near 60% decrease in MYCN protein level but in these conditions, we did not detect any change in DDX17 and DDX5 expression, both at protein and RNA level (Figures 4C and 4D).

Therefore, we concluded that variations in the expression of MYCN do not have any significant impact on the expression of DDX17 and DDX5 helicases.

### **MYCN overexpression inhibits transcription termination and induces the formation of chimeric transcripts**

As chimeric transcripts are more detected in aggressive neuroblastomas [35] where *MYCN* is frequently amplified, we used the two cellular models described above to test whether this oncoprotein could by itself have an effect on the production of these abnormal mRNAs. Interestingly, we observed an increased production of chimeric transcripts upon overexpression of MYCN in SH-SY5Y cells (Figures 5A and 5B, Supplementary Figure S4). MYCN also induced a transcriptional readthrough at the 3' end of genes that we previously identified as being regulated by DDX17 and DDX5 [38] (Figure 5B right panel, Supplementary Figure S4).

We then carried out the opposite experiment and monitored the expression of the same transcripts in conditions of MYCN knock-down in SK-N-BE(2) cells (Figure 5C). In these cells, we observed a significant reduction of the expression of chimeric transcripts, as well as a reduced transcriptional readthrough

(Figure 5D). These results suggest that MYCN could have an impact on gene expression which differs from its promoter-associated activity.

### **MYCN binds to the 3' end of DDX17/DDX5-regulated genes and interacts with DDX17**

To test whether the effect of MYCN on transcription termination and 3' end processing is direct, we re-analysed previously published ChIP-seq datasets performed in several neuroblastoma cell lines. We found that the 3' region of genes displaying transcriptional readthrough or chimeric RNA production upon DDX17/DDX5 knock-down or MYCN overexpression generally exhibited at least one MYCN peak, often detected across several cell lines (Figure 6A and Supplementary Figure S5). A global analysis of all DDX17/DDX5-regulated genes confirmed this observation, as the chromatin region overlapping their terminal exon was significantly enriched in MYCN binding sites, compared to control genes (Figure 6B). We also performed ChIP-qPCR in our SK-N-BE(2) cells to validate the binding of MYCN to 3' end proximal sites of our model genes (Figure 6C). Together with our results from Figure 4 showing that MYCN does not affect DDX17 and DDX5 expression, these experiments strongly suggested that the effect of MYCN on transcription termination is direct.

We next hypothesized that through its binding to chromatin at the 3' end of DDX17/DDX5 target genes, MYCN could interfere with the function of helicases. Therefore, we tested whether MYCN could interact with DDX17 and DDX5 by carrying out co-immunoprecipitation (co-IP) assays in SK-N-BE(2) cells. As shown in Figure 6D, endogenous DDX17 and MYCN efficiently co-immunoprecipitated, whatever protein was targeted. The co-IP was even enhanced when extracts were treated with benzonase, excluding the involvement of nucleic acids and suggesting a possible direct interaction between the two proteins. In contrast, no co-IP was detected between MYCN and DDX5 (Figure 6E), which underlines the specificity of the MYCN-DDX17 interaction.

Altogether these results indicate that MYCN impacts transcription termination in a direct manner, possibly by inhibiting the normal function of DDX17.

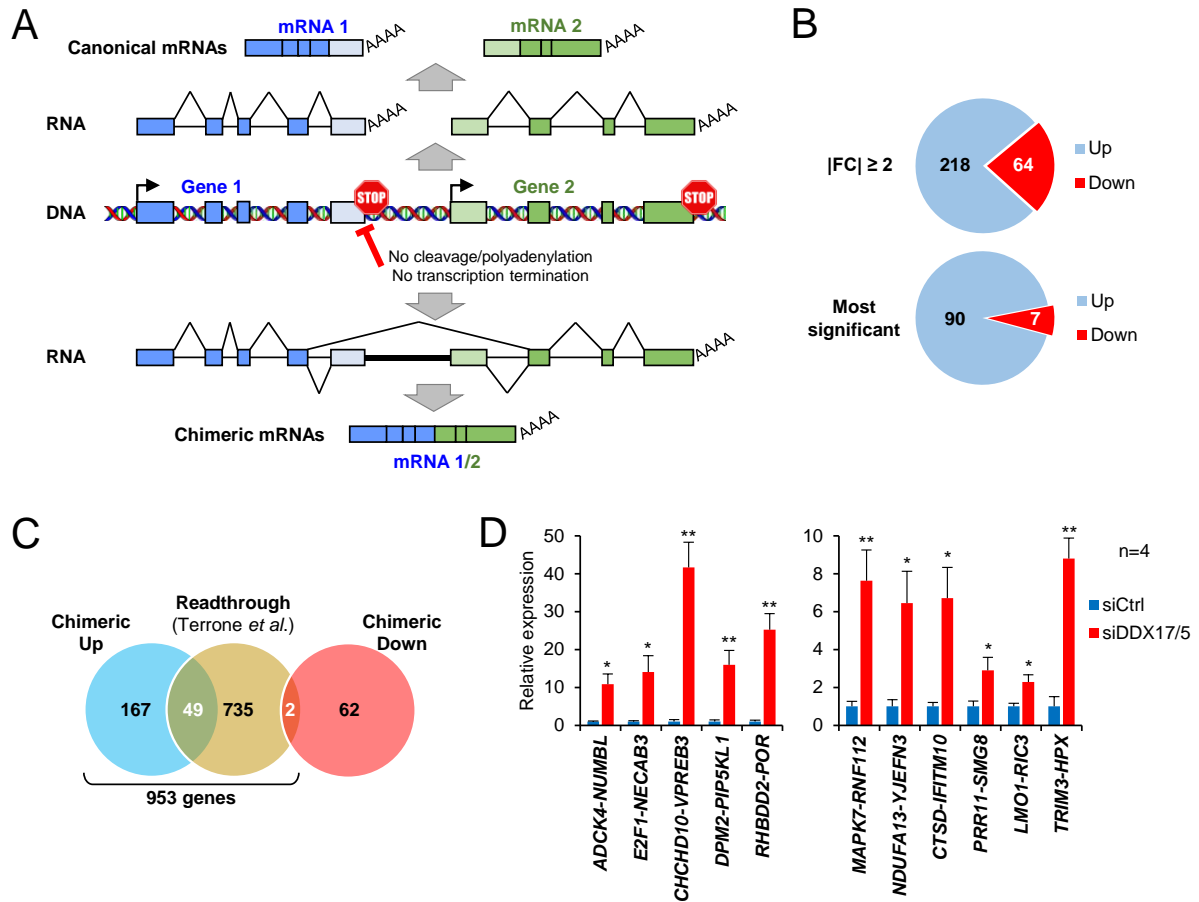
### **The forced recruitment of MYCN near the termination region is sufficient to induce the formation of chimeric transcripts.**

To provide a direct proof that MYCN binding near the termination region of a gene is sufficient to promote transcriptional readthrough and the subsequent production of a chimeric transcript, we designed an experiment in SH-SY5Y cells (in which the endogenous MYCN level is very low) in which we could force the recruitment of MYCN downstream of the PAS of the *MAPK7* gene. For this, we engineered a plasmid expressing the MYCN protein in a fused for with a dead Cas9 protein (dCas9-MYCN), that we could then target to a specific locus thanks to the co-transfection of a plasmid

expressing specific RNA guides (Figure 7A). We verified by ChIP-qPCR that the dCas9-MYCN protein was indeed bound at its target site upon transfection of RNA guides (Figure 7B). A control dCas9-GFP fusion was used as a control, to rule out a possible non-specific action of the dCas9 fusion protein on transcription through steric hindrance. We then quantified the expression of the MAPK7 gene, which remained unaffected by the recruitment of either dCas9-MYCN or dCas9-GFP, as did the control ADCK4 gene (Figure 7B, left panel). In contrast, the chimeric MAPK7-RNF112 chimera was induced only upon the binding of the dCas9-MYCN between the two genes while the ADCK4-NUMBL chimera remained unchanged (Figure 7B, right panel).

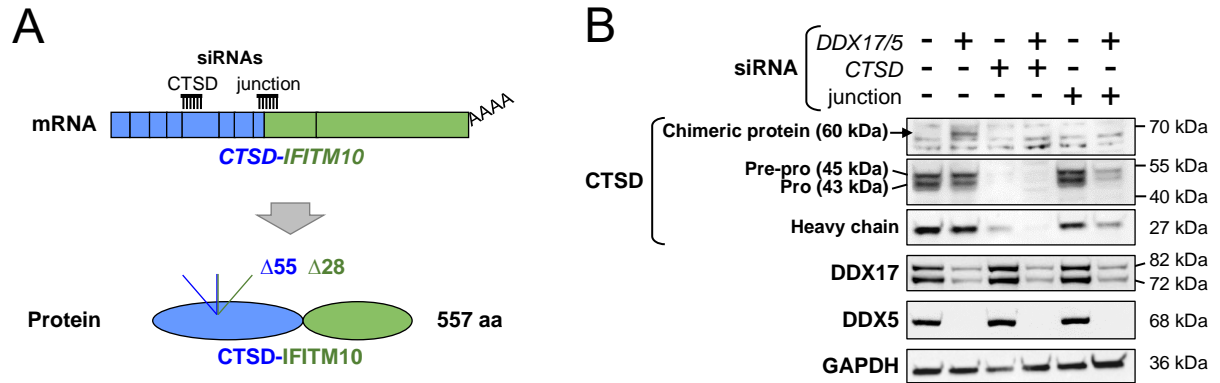
This experiment demonstrated that the binding of MYCN near the termination site can inhibit the normal termination process and promote the formation of a chimeric transcript.

# Figure 1



**Figure 1 : DDX17 and DDX5 depletion enhances the expression of chimeric transcripts.**  
**(A)** Schematic representation of the production of readthrough induced chimeric transcripts. **(B)** Diagram showing the number of up and downregulated chimeric transcripts displaying at least a fold change of 2 and the most significant transcripts (p-val < 0.1). **(C)** Venn diagram showing the overlap between genes producing chimeric transcripts and the genes previously identified as displaying transcriptional readthrough events. The list of 953 genes corresponds to the genes on which we carried out global analyses. **(D)** Quantification of chimeric transcripts. Expression of chimeric transcripts was monitored by RT-qPCR using primers spanning the chimeric junction and then normalized on the expression of the 5' parental gene (see also Supplementary Figure S2B). Two-tailed paired *t*-test (\* P-val < 0.05 ; \*\* P-val < 0.01).

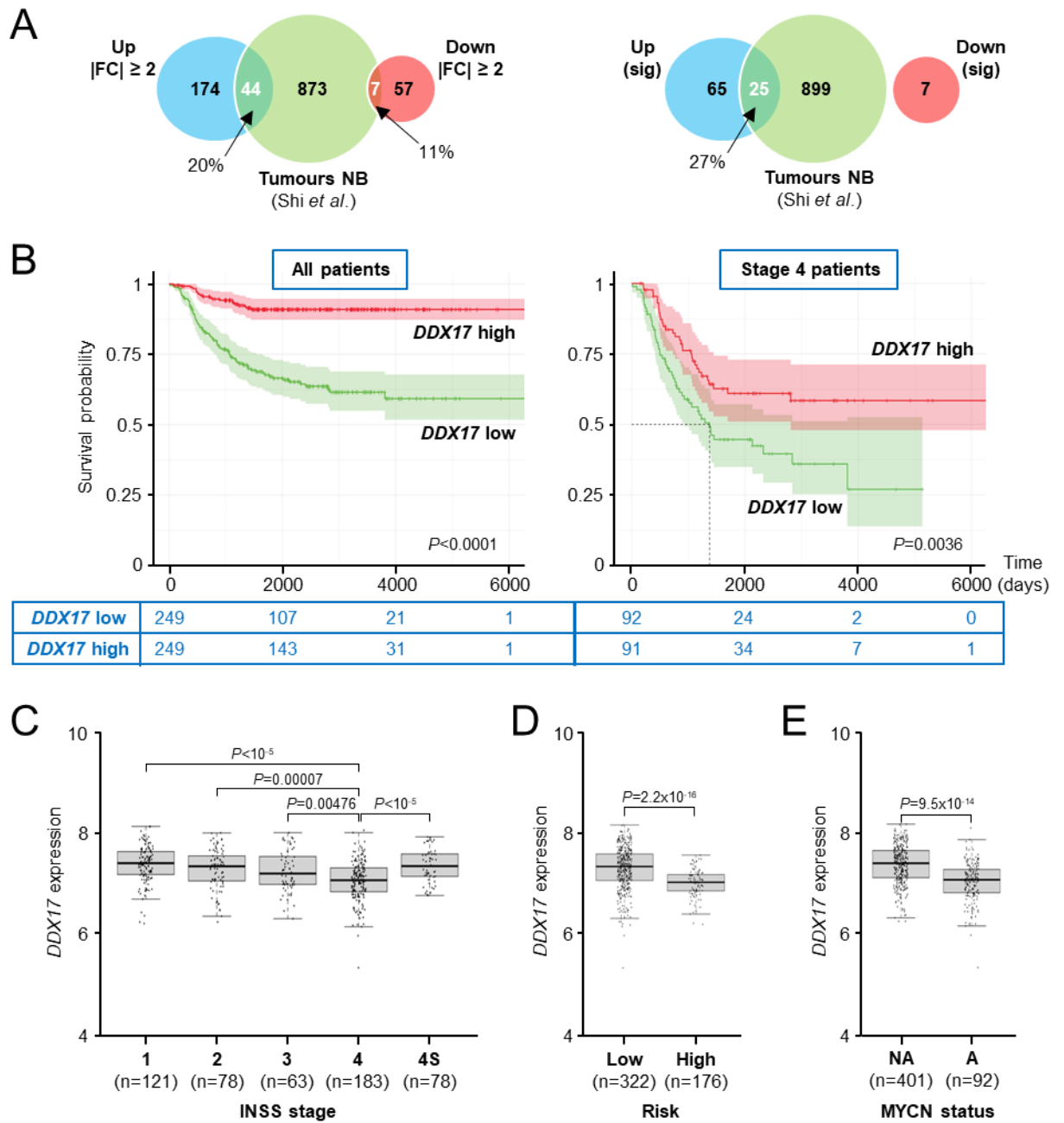
## Figure 2



**Figure 2 : Chimeric transcripts can produce chimeric proteins.**

(A) Schematic representation of CTSD-IFITM10 chimeric transcripts and putative chimeric protein. The sites targeted by the siRNAs used to validate the specificity of the band corresponding to CTSD-IFITM10 protein are indicated. (B) Western-blot showing the expression of the chimeric CTSD-IFITM10 protein and the canonical protein CTSD, as well as the validation of DDX5 and DDX17 silencing.

**Figure 3**



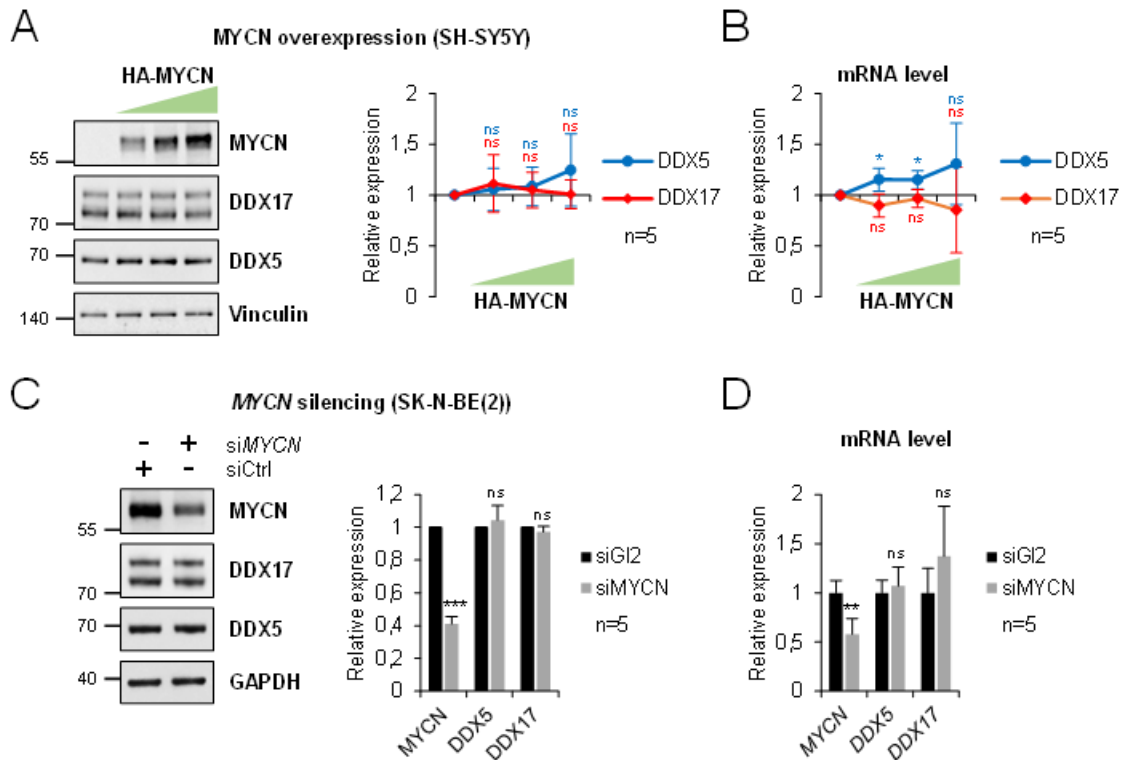
**Figure 3 : The expression of the *DDX17* gene is reduced in high-risk neuroblastomas.**

(A) Venn diagrams showing the numbers of chimeric transcripts that are both up or down-regulated after *DDX17/5* depletion and identified in neuroblastoma tumors. (B) Kaplan-Meier curves indicating the overall survival of 498 well-defined NB patients (GEO: GSE62564), and for stage 4 patients only, with high or low *DDX17* expression. (C) Box plots of *DDX17* expression relative to tumor stage. ANOVA corrected for multiple comparisons with Tukey's tests. (D-E) Box plots of *DDX17* expression relative to low/high risk classification and amplified (A) or non amplified (NA) *MYCN* status. Two-tailed *t*-test.





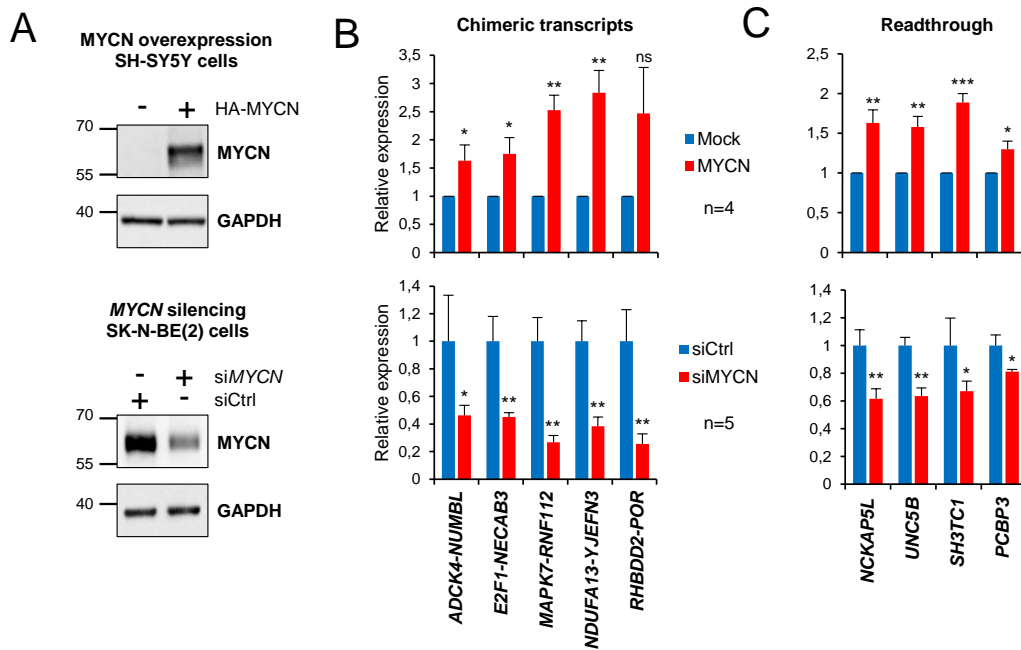
## Figure 4



**Figure 4 : MYCN does not impact the expression of DDX17 and DDX5.**

(A-B) Western-blot (A) and RT-qPCR experiments (B) showing that MYCN overexpression of MYCN in SH-SY5Y cells does not significantly affect DDX17 and DDX5 expression. Paired *t*-test (\**P*-val<0.05). (C-D) Western-blot (C) and RT-qPCR experiments (D) showing that treatment by siMYCN of SK-N-BE(2) cells does not significantly affect DDX17 and DDX5 expression. mRNA levels are normalized to *GAPDH* expression. Paired *t*-test (\*\**P*-val<0.01 ; \*\*\**P*-val<0.001).

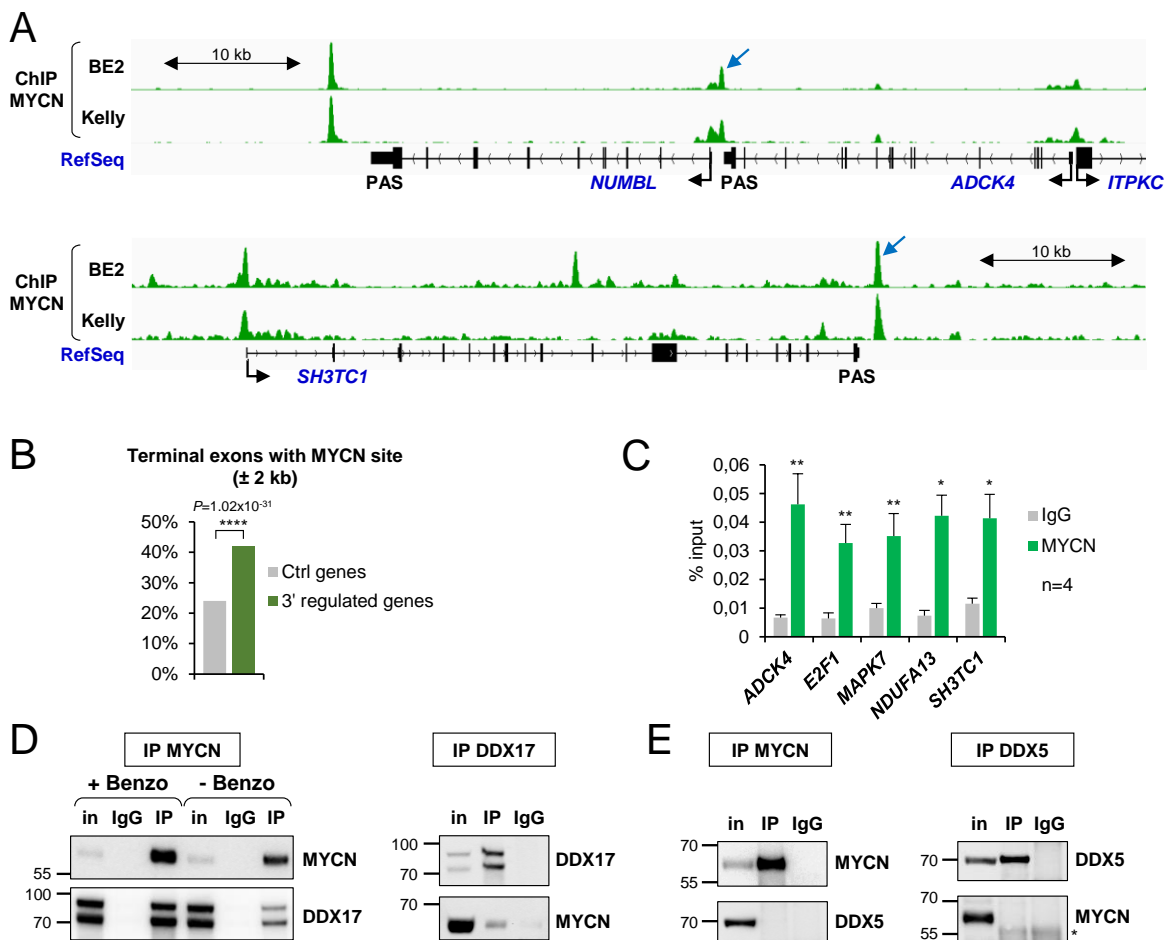
**Figure 5**



**Figure 5 : MYCN overexpression inhibits transcription termination and induces the formation of chimeric transcripts.**

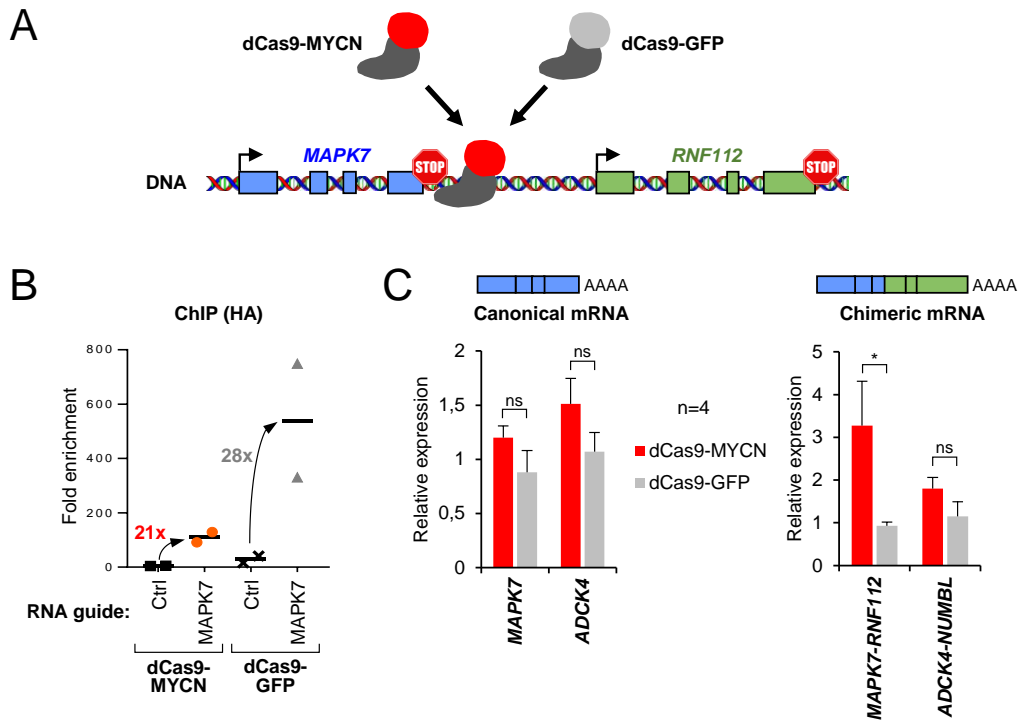
(A) Western-blot validation of MYCN overexpression (top) and silencing (bottom). (B) Quantification of chimeric transcripts expression after MYCN overexpression (top) or silencing (bottom). Expression was measured by RT-qPCR using junction primers and normalized on the expression of the parental gene (see also Supplementary Figure S4). Statistics for overexpression and MYCN silencing, respectively: ratio paired *t*-test and unpaired Mann-Whitney test (\**P*-val<0.05 ; \*\**P*-val<0.01). (C) Quantification of readthrough transcripts expression after MYCN overexpression (top) or silencing (bottom). Expression of was measured by RT-qPCR using primers localized downstream of the 3' end of the gene, and was normalized to the total amount of the corresponding canonical mRNA, measured with primers localized near the 3' end of the gene (see also Supplementary Figure S4). Statistics for overexpression and MYCN silencing, respectively: ratio paired *t*-test and unpaired Mann-Whitney test (\**P*-val<0.05 ; \*\**P*-val<0.01).

## Figure 6



**Figure 6 : MYCN binds to the 3' end of DDX17/DDX5-regulated genes and interacts with DDX17.** (A) Typical examples of DDX17/DDX5-regulated genes displaying MYCN binding near their 3' end. (B) Genome wide analysis of MYCN binding near terminal exons of DDX17/DDX5-regulated genes and control genes in neuroblastoma cell lines. Tukey's test. (C) ChIP-qPCR analysis of MYCN binding near the 3' end of DDX17/DDX5-regulated genes. Paired *t*-test (\**P*-val<0.05 ; \*\**P*-val<0.01). (D-E) Co-immunoprecipitation of endogenous MYCN with DDX17 (D) and DDX5 (E) in SK-N-BE(2) cells.

## Figure 7



**Figure 7 : The forced recruitment of MYCN near the termination region is sufficient to induce the formation of chimeric transcripts.**

(A) Strategy for the recruitment of MYCN (fused to a catalytically inactive dCas9) at the 3' end of the *MAPK7* gene. (B) ChIP-qPCR analysis of dCas9-MYCN and dCas9-GFP binding near the 3' end of *MAPK7*. (C) Quantification of *MAPK7* canonical transcript and *MAPK7-RNF112* chimeric transcript in cells transfected with dCas9-MYCN or dCas9-GFP, relative to transfection without specific gRNA. Paired *t*-test (\**P*-val<0.05 ; \*\**P*-val<0.01).

## DISCUSSION

In this work we show that the readthrough induced by depletion of the DDX17 and DDX5 helicases induces the production of chimeric transcripts in neuroblastoma cells, and that the reduced expression of both helicases, especially of DDX17, is associated with a poorer prognosis. Interestingly, 25 of the most significant chimeric transcripts induced by DDX17/DDX5 depletion were also identified in neuroblastoma tumors. Only 1 of these 25 transcripts was also detected in 161 samples of healthy adrenal gland tissue [35], while 4 were found in at least 10% of the 172 analysed neuroblastoma tumours. Being highly expressed in neuroblastomas but not in normal tissues could indicate that these transcripts could potentially serve in the future as biomarkers or help to guide patients for a better treatment and management.

Two of four chimeric transcripts found in at least 10% of patient's tumors are predicted to potentially produce chimeric proteins. These chimeric proteins may generate highly immunogenic neo-antigens, which would be unique to the tumor [51] and that could be targeted by the immune system [52]. In a recent case study, the authors identified a novel chimeric transcript produced by a fusion gene and demonstrated that it produces a neoantigen that can specifically elicit a host cytotoxic T-cell response in metastatic head and neck cancer [53]. In the case of readthrough induced chimeric RNA, it nevertheless should be noted that the expression of chimeric proteins might be far less important. In case of sufficient production, the expression of chimeric proteins could alter the protein interaction network and activate cell growth pathways by functioning as oncoproteins or alter the cell response to antitumoral treatments.

Even if not translated, chimeric transcripts can act as functional lncRNAs. For example, in prostate cancer, SLC45A3-ELK4 chimeric transcripts can act as a lncRNA and regulate cell growth, and the higher expression level of this chimeric transcript was found to be associated with disease progression and prostate cancer metastasis[54,55]. Further analysis would be required to determine whether the chimeric transcripts we identified indeed possess protumoral properties or can influence tumor response to treatment.

DDX17 and DDX5 are usually associated with tumor progression in other cancers, with the exception of pancreatic and liver cancer [56]. The fact that their expression appears to be strongly associated with better survival in neuroblastoma is therefore of interest. The reduction of their expression in high-risk neuroblastoma might just be a consequence of the different genomic and transcriptomic landscape between risk groups, such as MYCN amplification. However, the fact that a better survival of stage 4 patients is still associated with a high DDX17 expression reduces this tight link with MYCN, thus it could be interesting to see if this variation of expression of the two helicases may be associated

with the mechanism of oncogenesis of neuroblastoma or if it can affect patient's outcome. In addition to their role in repressing readthrough transcript, the role of DDX5 and DDX17 in genomic stability and DNA repair [57,58] might also explain how their low expression can be associated with high-risk tumor, especially in neuroblastoma, a cancer that is associated with complex combinations of acquired genetic aberrations [59,60].

We show that MYCN, a transcription factor whose gene amplification is a high-risk factor in neuroblastoma, promotes the production of transcriptional readthrough and chimeric transcripts. The recruitment of a dCas9-MYCN near the 3' end of the MAPK7 gene is sufficient to increase the expression of the chimeric MAPK7-RNF112 transcript produced from this gene. It was recently demonstrated that recruitment of dCas9 downstream of a gene induces Pol II pausing and promotes transcriptional termination, reducing stress-induced readthrough. Nevertheless, in our work, recruitment of dCas9-CT does not induce a significant increase in chimeric transcript production, compared with cells transfected with dCas9-CT without guides, suggesting that MYCN recruitment increases chimeric transcript production and that this effect is not dependent on dCas9 recruitment. In addition to its effect on transcription, MYCN could therefore also modulate the maturation of the 3' end of transcripts. Several transcription factors have already been demonstrated to affect mRNA processing events such as splicing, by binding to the gene body [61,62]. Transcription factors have been shown to modulate splicing through chromatin remodeling, altering Pol II elongation rate, as well as direct recruitment of splicing factors. RNA Pol II elongation rate and chromatin modification are also parameters that can affect mRNA 3' end processing and transcription termination [63–66]. MYCN itself interacts and recruits multiple chromatin remodeling complexes, whose ectopic recruitment to the 3' end of gene may alter the local chromatin state and affect Pol II kinetics and interactome [67–69]. Furthermore, MYCN and DDX17 interaction can suggest that MYCN might directly disturb DDX17 function.

In summary, our results provide evidence that both DDX17 and DDX5 loss of expression as well as MYCN overexpression induce the readthrough transcription of several common genes, inducing the production of chimeric transcripts. Coincidentally, DDX17 and DDX5 reduced expression and MYCN amplification are both linked to worse survival in neuroblastoma patient. Further work will be necessary to unveil if the induction of readthrough and chimeric transcripts can indeed play a functional role in neuroblastoma.

## MATERIALS AND METHODS

### Cell Culture and Transfections

Human SH-SY5Y neuroblastoma cells (ECACC) were grown and transfected essentially as described previously [48]. For standard silencing experiments, 20 nM of siRNA (Table 1) were mixed with Lipofectamine RNAiMax (Thermo Fisher Scientific) following the manufacturer's instructions and cells were harvested 48h after transfection. For MYCN overexpression, cells were plated in 6-well-plates to reach 70% confluency, were then transfected with 0.5 to 1 µg (pCDNA3-HA-human MYCN, Addgene, pcDNA™3.1/myc-His) using JetPrime® (Polyplus transfection) following the manufacturer's instructions. Cells were harvested 48h after transfection. For dCas9 experiments, we cloned the sequences corresponding to the two RNA guides into the *BsmBI* site of the CRIZI plasmid (provided by Philippe Mangeot, CIRI, Lyon). SH-SY5Y cells 6-well-plates to reach 70% confluency were then transfected with 1 µg of dCas9-HA-MYCN plasmid and 1 µg of gRNA-containing plasmid (500 ng of each guide, sequences in Table 2) using jetPRIME (PolyPlus Transfection).

Human SK-N-BE(2) neuroblastoma cells (ECACC) were grown and transfected essentially as described previously[48]. In standard transfections experiments on SK-N-BE(2) cells, 20 nM of siRNA was used and cells were harvested 48h later.

### Western Blot

Co-immunoprecipitation and western blotting following total protein extraction was carried out as previously described[70]. Primary antibodies used for western-blotting were: DDX5 (ab10261, Abcam), DDX17 (ab24601, Abcam), GAPDH (sc-32233, SantaCruz), MYCN (sc-53993, SantaCruz), CTSD (21327-1-AP, Proteintech). For co-immunoprecipitation, SH-SY5Y cells were harvested and gently lysed for 5 min on ice in a buffer containing 10 mM Tris-HCl pH 8.0, 140 mM NaCl, 1.5 mM MgCl<sub>2</sub>, 10 mM EDTA, 0.5% NP40, completed with protease and phosphatase inhibitors (Roche #11697498001 and #5892970001), to isolate the nuclei from the cytoplasm. After centrifugation, the nuclei were lysed in the IP buffer (20 mM Tris-HCl pH 7.5, 150 mM NaCl, 2 mM EDTA, 1% NP40, 10% glycerol and protease/phosphatase inhibitors) for 30 min at 4°C under constant mixing. The nuclear lysate was centrifuged for 15 min to remove debris and soluble proteins were quantified by BCA (ThermoFisher). The lysate was pre-cleared with 30µg of Dynabeads Protein G/A (ThermoFisher) for 30 min under rotatory mixing, and then split in 1.5 mg aliquots of proteins for each assay. Each fraction received 5 µg of antibody and the incubation was left overnight at 4°C under rotation. The following antibodies were used for IP: rabbit antiDDX17 (ProteinTech 19910-1-AP) or a control rabbit IgG (ThermoFisher), goat anti-DDX5 (Abcam ab10261) or control goat IgG (Santa Cruz), and mouse anti-MYCN (Santa Cruz sc-53993) or control mice IgG. The next day, the different lysate/antibody mixtures were divided, treated or not with benzonase (Merck-Millipore) for 30 min at 37°C, and then incubated with 50 µg Dynabeads Protein G/A (ThermoFisher) blocked with bovine serum albumin, for 4 h at 4°C under rotation. Bead were then washed 5 times with IP buffer. Elution was performed by boiling for 5 min in SDS-PAGE loading buffer prior to analysis by western-blotting.

### RNA extraction and Real-time quantitative PCR

Total RNA were isolated using TriPure Isolation Reagent (Roche). For reverse transcription, 2 µg of purified RNAs were treated with Dnase I (ThermoFisher) and retrotranscribed using Maxima reverse transcriptase (ThermoFisher), as recommended by the supplier. Potential genomic DNA contamination was systematically verified by performing negative RT controls in absence of enzyme, and by including



controls with water instead of cDNA in qPCR assays. For qPCR analyses, the specificity and linear efficiency of all primers (sequences are in Table 3) was first verified by establishing a standard expression curve with various amounts of human genomic DNA or cDNA. qPCR reactions were carried out on 0.625 ng cDNA using a LightCycler 480 System (Roche), with the SYBR® Premix Ex Taq (Tli RNaseH Plus) (Takara), under thermocycling conditions that were recommended by the manufacturers. Melting curves were controlled to rule out the existence of non-specific products. Relative DNA levels were calculated using the  $2^{-\Delta\Delta Ct}$  method (using the average Ct obtained from technical duplicates or triplicates) and were normalized to the expression of GAPDH RNA

### **Chromatin immunoprecipitation**

A total of  $10^7$  cells were crosslinked with 1% formaldehyde for 10 min at room temperature. Crosslinking was quenched by addition of 0.125 M glycine. Nuclei were isolated by sonication using a Covaris S220 (2 min, Peak Power: 75; Duty Factor: 2; Cycles/burst: 200), pelleted by centrifugation at  $1000\times g$  for 5 min at 4 °C, washed once with FL buffer (5 mM HEPES pH 8.0, 85 mM KCl, 0.5% NP-40) and resuspended in 1 ml shearing buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA, 0.1% SDS). Chromatin was sheared in order to obtain fragments ranging from 200 to 800 bp using Covaris S220 (20 min, Peak Power: 140; Duty Factor: 5; Cycles/burst: 200). Chromatin was next immunoprecipitated overnight at 4 °C with 4 µg of mouse anti-MYCN II antibody (sc53993, Santa Cruz) or equivalent amounts of their corresponding IgG Isotype control (ThermoFisher) and 30 µl of Dynabeads® Protein A/G (ThermoFisher) were added. Complexes were washed with 4 different buffers: Low salt buffer (1% Triton X-100, 0.1% SDS, 150 mM NaCl, 20 mM Tris-HCl pH 8, 2 mM EDTA), High salt buffer (1% Triton X-100, 0.1% SDS, 500 mM NaCl, 20 mM Tris-HCl pH 8, 2 mM EDTA), Low LiCl buffer (1% Na-DOC, 1% NP-40, 10 mM Tris-HCl pH 8, 0.5M LiCl), Tris/EDTA (10 mM Tris-HCl pH 8, 1 mM EDTA), and were eluted in Elution buffer (1% SDS, 200 mM NaCl, 0.1M NaHCO<sub>3</sub>, 20 µg Proteinase K) overnight at 65°C. The immunoprecipitated chromatin was purified by phenol-chloroform extraction, and quantitative PCR was performed using Rotor-Gene 3000 cycler (Corbett) or LightCycler 480 II (Roche, Mannheim, Germany). Values were expressed relative to the signal obtained for the immunoprecipitation with control IgG. Chromatin was precipitated with 4 µg of mouse anti-MYCN II antibody (sc53993, Santa Cruz) or equivalent amounts of their corresponding IgG Isotype control (ThermoFisher).

### **Bioinformatic prediction of chimeric RNAs**

Raw reads were pre-processed using fastp (v0.23.2) [71] and then mapped using STAR (v2.7.8a) [72] to the human reference genome (hg19, GRCh37.87), additional parameters are set according to recommendation of Arriba software to detect gene fusions. Mapped reads were filtered using samtools (v1.11) [73]. Gene fusions were detected using Arriba (v2.3.0) [47] then results were parsed using R script. Differentially fused genes (DFG) were analyzed using DESeq2 package (v1.40.1) [74]. Split and discordant reads identified thanks to Arriba are counted for each fusions and compared to control condition. Quality check of raw and pre-processed reads was performed using fastqc (v0.11.5) and compiled using multiqc (v1.11) [75,76]. Complete pipeline in Nextflow[77] is available at ([https://gitbio.ens-lyon.fr/LBMC/regards/nextflow/-/blob/master/src/arriba\\_fusion.nf](https://gitbio.ens-lyon.fr/LBMC/regards/nextflow/-/blob/master/src/arriba_fusion.nf))

### **Analysis of MYCN ChIP-seq data**

Raw reads were pre-processed using fastp (v0.23.2) [71] and then mapped using bowtie2 v2.5.2 [78] to the human reference genome (hg19, GRCh37.87). Mapped reads were filtered using samtools (v1.11) [73] then formatted using deeptools (v3.5.1) [79]. Peak calling analysis was carried out with MACS (v3.0.0a6) [80]. Nearest exons from peak summits were identified thanks to BEDtools (v2.25.0) [6]. Quality check of raw and pre-processed reads was performed using fastqc (v0.11.5) and compiled using multiqc (v1.11) [76,77]. Complete pipeline in Nextflow [77] is available at (<https://gitbio.ens-lyon.fr/xgrand/ChIPster>).

Table 1 : siRNA

siRNA	sequence
siMYCN_1	AGAGGAAGAUGAAGAGGAA[dT][dT]
siMYCN_2	CCACAGAAGAAGAUAAAGA[dT][dT]
siDDX5/17	GGCUAGAUGUGGAAGAUGU[dT][dT]
siCTL(GI2/Luc)	CGUACGCGAAUACUUCGA[dT][dT]
siCTSD.E6	GUAUUACAAGGGUUCUCUG[dT][dT]
siCTSD.E8- IFITM10.E2	ACUACACGCUCAAGGCCCA[dT][dT]

Table 2 : guideRNA

sgRNA	sequence
sgRNA_dCas9MYCN_MAPK7_1_F	CACCGACTCCCTGTCCTACCACAG
sgRNA_dCas9MYCN_MAPK7_1_R	aaacCTGTGGTAGGACAGGGAGTC
sgRNA_dCas9MYCN_MAPK7_2_F	CACCGCGACTGTGAGAGATATCACG
sgRNA_dCas9MYCN_MAPK7_2_R	aaacCGTGATATCTCTCACAGTCGC

Table 3 : primers

Target	Name of primer	Sequence	size (pb)	Experiment
GAPDH	GAPDH_F	ACCACAGTCCATGCCATCACT	71	RT-qPCR
	GAPDH_R	CCATCACGCCACAGTTTCC		
DDX5	DDX5_F	CAGGCACAGCATACACTTTC	158	
	DDX5_R	CTTCATGCCTCCTCTACC		
DDX17	DDX17_F	GGATGTCTGCATGGAAAGTG	122	
	DDX17_R	TCAGATCATCACAGCGTCTC		
MYCN	MYCN_F	TCAGTACCTCCGGAGAGGAC	96	
	MYCN_R	TCTCCACAGTGACCACGTCG		
CTSD	CTSD_F	ACACAGGCACTTCCCTCATG	146	
	CTSD_R	TCCCAGCTTCAGTGTGATCG		
CTSD-IFITM10	CTSD-IFITM10_F	TGCTGCTCTTGACTCCATG	266	
	CTSD-IFITM10_R	GGACTACAGCTCAAGGCC		
NDUFA13	NDUFA13_F	ATGCTGGCCATAGGGATTGG	139	
	NDUFA13_R	GGTCGGTTTCTGCCTGTAAC		
NDUFA13- YJEFN3	NDUFA13-YJEFN3_F	CGTGCCCCGACTGGAAGGGCC	128	
	NDUFA13-YJEFN3_R	CCTGCGTTCCAAATCTGCTC		
DPM2	DPM2_F	TCGCCGTTAGCCTGATCATC	154	
	DPM2_R	ACAAACAGGAGCAGCAGGAG		
DPM2-PIP5KL1	DPM2-PIP5KL1_F	TCCTGCTGCTCCTGTTTGTG	136	
	DPM2-PIP5KL1_R	CCGGGCTGATCTCAAACAGG		
ADCK4	ADCK4_F	TGCTGGACTTTGGTGCAAGC	115	
	ADCK4_R	TCCCTGGACTTCTGCAGGAC		

ADCK4-NUMBL	ADCK4-NUMBL_F	AGGCTTTGAAACCAAGGGCG	128
	ADCK4-NUMBL_R	CGTAACTTGTTTCATGGTGCC	
CHCHD10	CHCHD10_F	CCTACGAGATCAGGCAGTTCC	202
	CHCHD10_R	ACAATCTGGTGTGTGGTCTG	
CHCHD10-VPREB3	CHCHD10-VPREB3_F	GCCGTCTCTAAGGTCGCC	187
	CHCHD10-VPREB3_R	CGTAGTCCCTGATGGTGACG	
E2F1	E2F1_F	TATGGTGATCAAAGCCCTCCT	112
	E2F1_R	AGGGCACAGAAAACATCGATC	
E2F1-NECAB3	E2F1-NECAB3_F	GTGTCACCACCACCATCATCTC	101
	E2F1-NECAB3_R	ATCATTCTGTCTGCTCTGCGG	
MAPK7	MAPK7_F	ACGGTGTTGGCTTTGACCTG	130
	MAPK7_R	CCTTCGAGCCAGTCAGCAAG	
MAPK7-RNF112	MAPK7-RNF112_F	GCCTTGCTGCCACAGTAGAC	141
	MAPK7-RNF112_R	AGGCTGCGGTAGGTAGAAGG	
LM01	LM01_F	GACTACCTGAGGCTCTTTGGCA	108
	LM01_R	GTCGAGGTGATACACGTTGTCC	
LM01-RIC3	LM01-RIC3_F	CTTCGAGATGGTGATGCGGG	138
	LM01-RIC3_R	GAGTCTGGCCATCTGAGGGT	
TRIM3	TRIM3_F	TGGCAGACTATGACAACCGTTG	163
	TRIM3_R	TGGAAGGTAAAGACGCAGCAAG	
TRIM3-HPX	TRIM3-HPX_F	AACAGCCGCATCCAGGACTA	125
	TRIM3-HPX_R	CATTGTCATCCAGGGTGGTAGC	
RHBDD2	RHBDD2_F	GGCTCATTCCCCAGACCTCT	146
	RHBDD2_R	CCTCATCAGGCTGAAGGGGA	
RHBDD2-POR	RHBDD2-POR_F	TCAGGGTCTTCAGCCGAGAG	162
	RHBDD2-POR_R	ACCCACGATGAGCGAAAACA	
PRR11	PRR11_F	CAGCTGGTGCCAGAACTGCA	148
	PRR11_R	CAGTGCTTCTGGAGCTTGCA	
PRR11-SMG8	PRR11-SMG8_F	AACAGGAACAGGACTGACTCCA	154
	PRR11-SMG8_R	GAGGTGCTTGTTTCCTTCCACA	
NCKAP5L total	NCKAP5L_F	TGGAGACCTCGGAGTCTCTC	133
	NCKAP5L_R	CATGAAGAGAGCCGGTCCAG	
NCKAP5L do_1.3kb	NCKAP5L.do_F	ATCTTCTGCAGAATTCAGG	211
	NCKAP5L.do_R	CAAAGTGCCAGTGAAGTTG	
SH3TC1 total	SH3TC1_F	TGCTCGGTGACATCATCTTC	175
	SH3TC1_R	AAGAGCGACTTCTCACGGTC	
SH3TC1 do_0.7kb	SH3TC1.do_F	CATCTCAGTCCCAGGAATC	143
	SH3TC1.do_R	GTTTGTGAGAATCACGTGGG	
PCBP3 total	PCBP3_F	CAGATCAAATCGCCAACGC	145
	PCBP3_R	AGGATTACAGCGTGCCCATC	
PCBP3 do_13kb	PCBP3.do_F	TTGTGTCCTGCATCCATTGG	175
	PCBP3.do_R	ACATGCTCACAGACACGTTT	
UNC5B total	UNC5B_F	AATGACTGGCGGATGTTAGC	179
	UNC5B_R	ACCAGCATCTCACTTTGCC	
UNC5B do_10.3kb	UNC5B.do_F	TTGGCACCACACTACATTTGGG	124

	UNC5B.do_R	AATGGTGACGTCTGTTCCAC		
MAPK7_do	MAPK7_D_F	CTGGCAGAGGAGAGAAGGGT	108	ChIP-qPCR
	MAPK7_D_R	CCAGGCTGCCACGTGATATC		
NDUFA13_do	NDUFA13_D_F	GGGAGAGGGAAGCTGGAGAC	139	
	NDUFA13_D_R	CTGCCGTTGAACCTCCATCC		
SH3TC1_do	SH3TC1_D_F	TCTCTTGCCCCAGGAAACA	129	
	SH3TC1_D_R	AAAAGAGGGGCCTGGGAGAG		
E2F1_do	E2F1_D_F	ACCCAATTGTGCACCCTGA	88	
	E2F1_D_R	TGTCGGGAGCAGAACCTAGG		
ADCK4_do	ADCK4_D_F	ACACACACAGTGCCTCTTGG	83	
	ADCK4_D_R	CAGGGACATTAGAGCCCATTAG		

## REFERENCES

1. Neugebauer KM. Nascent RNA and the Coordination of Splicing with Transcription. *Cold Spring Harb. Perspect. Biol.* 2019 ; 11 : a032227.
2. Tellier M, Maudlin I, Murphy S. Transcription and splicing: A two-way street. *Wiley Interdiscip. Rev. RNA* 2020 ; 11 : e1593.
3. Boreikaitė V, Passmore LA. 3'-End Processing of Eukaryotic mRNA: Machinery, Regulation, and Impact on Gene Expression. *Annu. Rev. Biochem.* 2023 ;
4. Porrua O, Libri D. Transcription termination and the control of the transcriptome: why, where and how to stop. *Nat. Rev. Mol. Cell Biol.* 2015 ; 16 : 190–202.
5. Proudfoot NJ. Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut. *Science* 2016 ; 352 : aad9926.
6. Rodríguez-Molina JB, West S, Passmore LA. Knowing when to stop: Transcription termination on protein-coding genes by eukaryotic RNAPII. *Mol. Cell* 2023 ; 83 : 404–415.
7. Cardiello JF, Goodrich JA, Kugel JF. Heat Shock Causes a Reversible Increase in RNA Polymerase II Occupancy Downstream of mRNA Genes, Consistent with a Global Loss in Transcriptional Termination. *Mol. Cell. Biol.* 2018 ; 38 : e00181-18.
8. Grosso AR, Leite AP, Carvalho S, *et al.* Pervasive transcription read-through promotes aberrant expression of oncogenes and RNA chimeras in renal carcinoma. *eLife* 2015 ; 4 .
9. Rutkowski AJ, Erhard F, L'Hernault A, *et al.* Widespread disruption of host transcription termination in HSV-1 infection. *Nat. Commun.* 2015 ; 6 : 7126.
10. Heinz S, Texari L, Hayes MGB, *et al.* Transcription Elongation Can Affect Genome 3D Structure. *Cell* 2018 ; 174 : 1522-1536.e22.
11. Rosa-Mercado NA, Zimmer JT, Apostolidi M, *et al.* Hyperosmotic stress alters the RNA polymerase II interactome and induces readthrough transcription despite widespread transcriptional repression. *Mol. Cell* 2021 ; 81 : 502-513.e4.
12. Vilborg A, Passarelli MC, Yario TA, *et al.* Widespread Inducible Transcription Downstream of Human Genes. *Mol. Cell* 2015 ; 59 : 449–461.
13. Vilborg A, Sabath N, Wiesel Y, *et al.* Comparative analysis reveals genomic features of stress-induced transcriptional readthrough. *Proc. Natl. Acad. Sci. U. S. A.* 2017 ; 114 : E8362–E8371.
14. Hennig T, Michalski M, Rutkowski AJ, *et al.* HSV-1-induced disruption of transcription termination resembles a cellular stress response but selectively increases chromatin accessibility downstream of genes. *PLOS Pathog.* 2018 ; 14 : e1006954.
15. Bauer DLV, Tellier M, Martínez-Alonso M, *et al.* Influenza Virus Mounts a Two-Pronged Attack on Host RNA Polymerase II Transcription. *Cell Rep.* 2018 ; 23 : 2119-2129.e3.
16. Rosa-Mercado NA, Steitz JA. Who let the DoGs out? – biogenesis of stress-induced readthrough transcripts. *Trends Biochem. Sci.* 2021 ;
17. Morgan M, Shiekhattar R, Shilatifard A, *et al.* It's a DoG-eat-DoG world—altered transcriptional mechanisms drive downstream-of-gene (DoG) transcript production. *Mol. Cell* 2022 ; 82 : 1981–1991.
18. Akiva P, Toporik A, Edelheit S, *et al.* Transcription-mediated gene fusion in the human genome. *Genome Res.* 2006 ; 16 : 30–36.

19. Parra G, Reymond A, Dabbouseh N, *et al.* Tandem chimerism as a means to increase protein complexity in the human genome. *Genome Res.* 2006 ; 16 : 37–44.
20. Greger L, Su J, Rung J, *et al.* Tandem RNA Chimeras Contribute to Transcriptome Diversity in Human Population and Are Associated with Intronic Genetic Variants. *PLOS ONE* 2014 ; 9 : e104567.
21. Prakash T, Sharma VK, Adati N, *et al.* Expression of Conjoined Genes: Another Mechanism for Gene Regulation in Eukaryotes. *PLoS ONE* 2010 ; 5 : e13284.
22. Kim RN, Kim A, Choi S-H, *et al.* Novel mechanism of conjoined gene formation in the human genome. *Funct. Integr. Genomics* 2012 ; 12 : 45–61.
23. Varley KE, Gertz J, Roberts BS, *et al.* Recurrent read-through fusion transcripts in breast cancer. *Breast Cancer Res. Treat.* 2014 ; 146 : 287–297.
24. Qin F, Song Z, Babiceanu M, *et al.* Discovery of CTCF-Sensitive Cis-Spliced Fusion RNAs between Adjacent Genes in Human Prostate Cells. *PLOS Genet.* 2015 ; 11 : e1005001.
25. Qin F, Song Y, Zhang Y, *et al.* Role of CTCF in Regulating SLC45A3-ELK4 Chimeric RNA. *PLOS ONE* 2016 ; 11 : e0150382.
26. Alpert T, Straube K, Carrillo Oesterreich F, *et al.* Widespread Transcriptional Readthrough Caused by Nab2 Depletion Leads to Chimeric Transcripts with Retained Introns. *Cell Rep.* 2020 ; 33 : 108324.
27. Hadar S, Meller A, Saida N, *et al.* Stress-induced transcriptional readthrough into neighboring genes is linked to intron retention. *iScience* 2022 ; 25 : 105543.
28. Dorney R, Dhungel BP, Rasko JEJ, *et al.* Recent advances in cancer fusion transcript detection. *Brief. Bioinform.* 2023 ; 24 : bbac519.
29. Sun Y, Li H. Chimeric RNAs Discovered by RNA Sequencing and Their Roles in Cancer and Rare Genetic Diseases. *Genes* 2022 ; 13 : 741.
30. Babiceanu M, Qin F, Xie Z, *et al.* Recurrent chimeric fusion RNAs in non-cancer tissues and cells. *Nucleic Acids Res.* 2016 ; 44 : 2859–2872.
31. Yun SM, Yoon K, Lee S, *et al.* PPP1R1B-STAR3D3 chimeric fusion transcript in human gastric cancer promotes tumorigenesis through activation of PI3K/AKT signaling. *Oncogene* 2014 ; 33 : 5341–5347.
32. Han P, Chen R-H, Wang F, *et al.* Novel chimeric transcript RRM2-c2orf48 promotes metastasis in nasopharyngeal carcinoma. *Cell Death Dis.* 2017 ; 8 : e3047.
33. Egashira S, Jinnin M, Makino K, *et al.* Recurrent Fusion Gene ADCK4-NUMBL in Cutaneous Squamous Cell Carcinoma Mediates Cell Proliferation. *J. Invest. Dermatol.* 2019 ; 139 : 954–957.
34. Chwalenia K, Qin F, Singh S, *et al.* A cell-based splicing reporter system to identify regulators of cis-splicing between adjacent genes. *Nucleic Acids Res.* 2019 ; 47 : e24.
35. Shi Y, Yuan J, Rraklli V, *et al.* Aberrant splicing in neuroblastoma generates RNA-fusion transcripts and provides vulnerability to spliceosome inhibitors. *Nucleic Acids Res.* 2021 ; 49 : 2509–2521.
36. Braccioli L, Wit E de. CTCF: a Swiss-army knife for genome organization and transcription regulation. *Essays Biochem.* 2019 ; 63 : 157–165.
37. Nanavaty V, Abrash EW, Hong C, *et al.* DNA Methylation Regulates Alternative Polyadenylation via CTCF and the Cohesin Complex. *Mol. Cell* 2020 ; 78 : 752-764.e6.

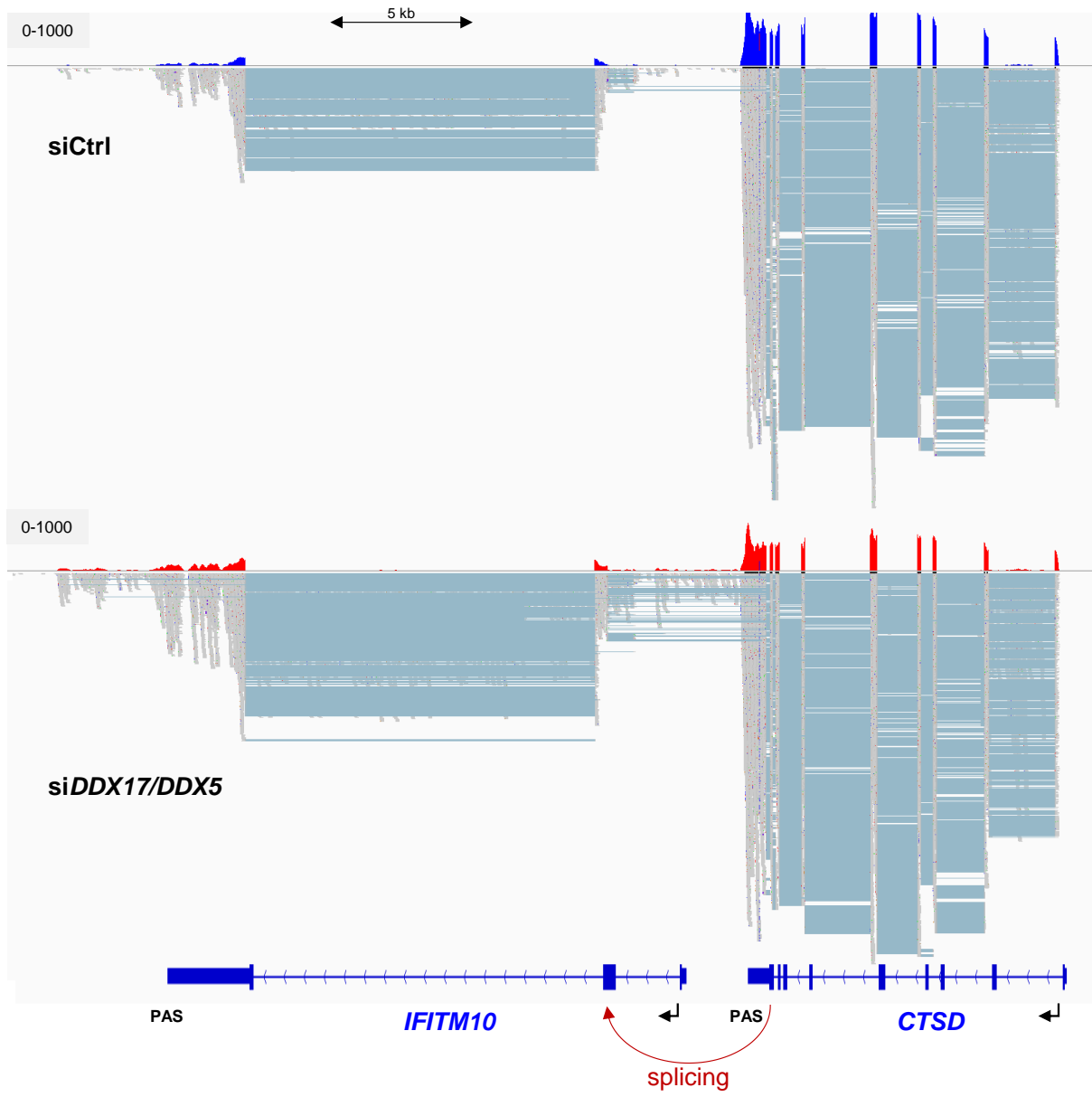
38. Terrone S, Valat J, Fontrodona N, *et al.* RNA helicase-dependent gene looping impacts messenger RNA processing. *Nucleic Acids Res.* 2022 ; 50 : 9226–9246.
39. Giraud G, Terrone S, Bourgeois and CF. Functions of DEAD box RNA helicases DDX5 and DDX17 in chromatin organization and transcriptional regulation. *BMB Rep.* 2018 ; 51 : 613–622.
40. Xing Z, Ma WK, Tran EJ. The DDX5/Dbp2 subfamily of DEAD-box RNA helicases. *Wiley Interdiscip. Rev. RNA* 2019 ; 10 : e1519.
41. Camats M, Guil S, Kokolo M, *et al.* P68 RNA Helicase (DDX5) Alters Activity of Cis- and Trans-Acting Factors of the Alternative Splicing of H-Ras. *PLOS ONE* 2008 ; 3 : e2926.
42. Lai Y-H, Choudhary K, Cloutier SC, *et al.* Genome-Wide Discovery of DEAD-Box RNA Helicase Targets Reveals RNA Structural Remodeling in Transcription Termination. *Genetics* 2019 ; 212 : 153–174.
43. Mersaoui SY, Yu Z, Coulombe Y, *et al.* Arginine methylation of the DDX5 helicase RGG/RG motif by PRMT5 regulates resolution of RNA:DNA hybrids. *EMBO J.* 2019 ; 38 : e100986.
44. Wu G, Xing Z, Tran EJ, *et al.* DDX5 helicase resolves G-quadruplex and is involved in MYC gene transcriptional activation. *Proc. Natl. Acad. Sci.* 2019 ; 116 : 20453–20461.
45. Kar A, Fushimi K, Zhou X, *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.* 2011 ; 31 : 1812–1821.
46. Katahira J, Ishikawa H, Tsujimura K, *et al.* Human THO coordinates transcription termination and subsequent transcript release from the HSP70 locus. *Genes Cells* 2019 ; 24 : 272–283.
47. Uhrig S, Ellermann J, Walther T, *et al.* Accurate and efficient detection of gene fusions from RNA sequencing data. *Genome Res.* 2021 ; 31 : 448–460.
48. Lambert M-P, Terrone S, Giraud G, *et al.* The RNA helicase DDX17 controls the transcriptional activity of REST and the expression of proneural microRNAs in neuronal differentiation. *Nucleic Acids Res.* 2018 ; 46 : 7686–7700.
49. Zhao X, Li D, Yang F, *et al.* Long Noncoding RNA NHEG1 Drives  $\beta$ -Catenin Transactivation and Neuroblastoma Progression through Interacting with DDX5. *Mol. Ther.* n.d. ;
50. Zhang W, Yu Y, Hertwig F, *et al.* Comparison of RNA-seq and microarray-based models for clinical endpoint prediction. *Genome Biol.* 2015 ; 16 : 1–12.
51. Weber D, Ibn-Salem J, Sorn P, *et al.* Accurate detection of tumor-specific gene fusions reveals strongly immunogenic personal neo-antigens. *Nat. Biotechnol.* 2022 ; 40 : 1276–1284.
52. Brennick CA, George MM, Corwin WL, *et al.* Neoepitopes as cancer immunotherapy targets: key challenges and opportunities. *Immunotherapy* 2017 ; 9 : 361–371.
53. Yang W, Lee K-W, Srivastava RM, *et al.* Immunogenic neoantigens derived from gene fusions stimulate T cell responses. *Nat. Med.* 2019 ; 25 : 767–775.
54. Qin F, Zhang Y, Liu J, *et al.* SLC45A3-ELK4 functions as a long non-coding chimeric RNA. *Cancer Lett.* 2017 ; 404 : 53–61.
55. Zhang Y, Gong M, Yuan H, *et al.* Chimeric Transcript Generated by cis-Splicing of Adjacent Genes Regulates Prostate Cancer Cell Proliferation. *Cancer Discov.* 2012 ; 2 : 598–607.
56. Xu K, Sun S, Yan M, *et al.* DDX5 and DDX17—multifaceted proteins in the regulation of tumorigenesis and tumor progression. *Front. Oncol.* 2022 ; 12 : 943032.
57. Bader AS, Luessing J, Hawley BR, *et al.* DDX17 is required for efficient DSB repair at DNA:RNA hybrid deficient loci. *Cell Biology*, 2021 : p.



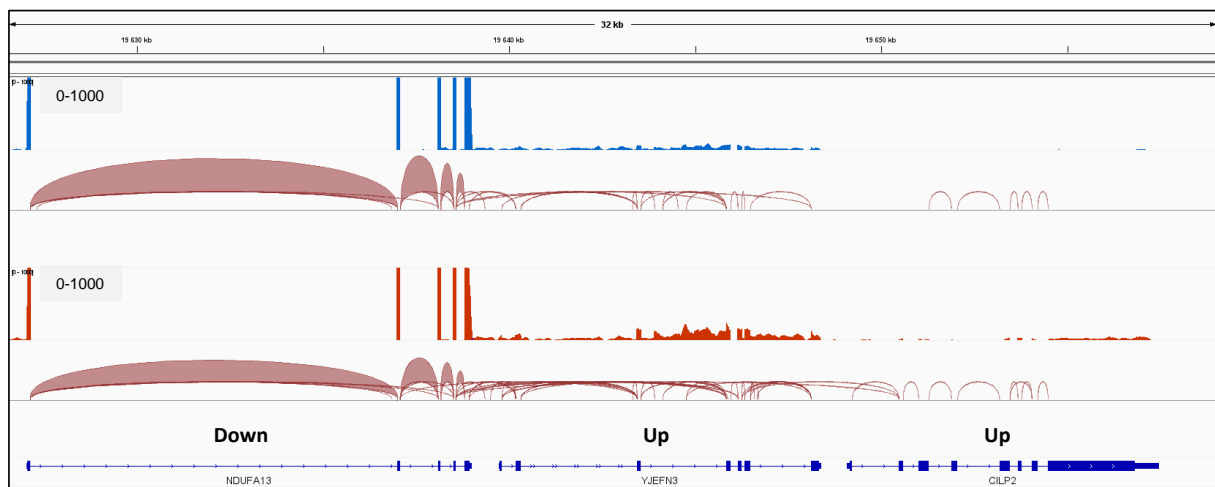
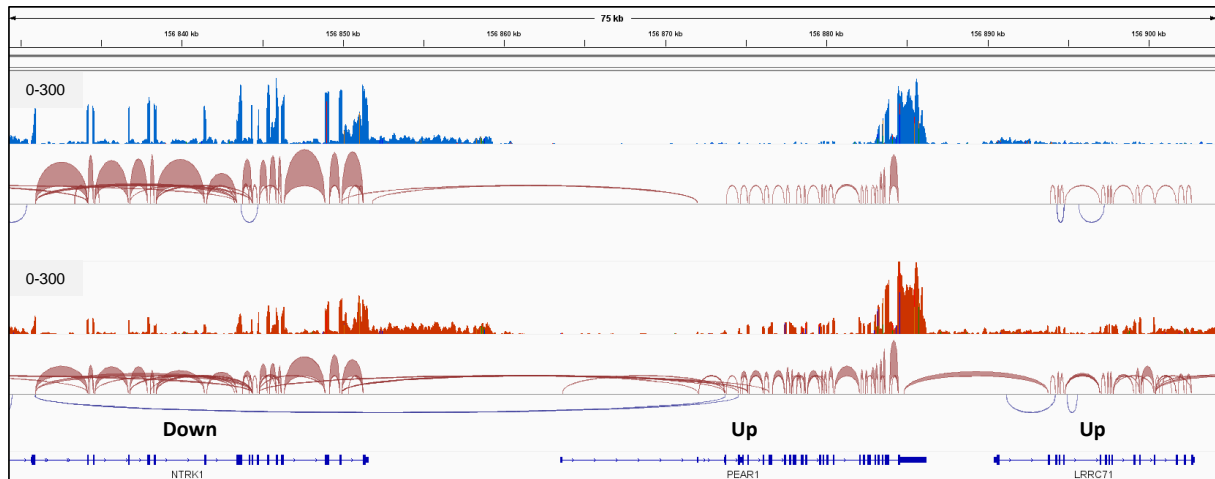
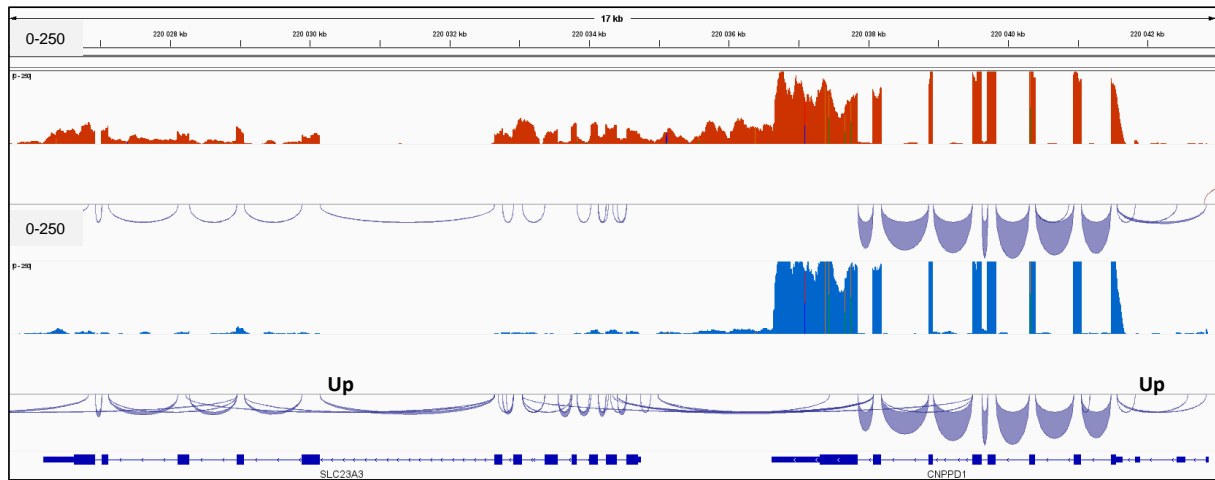
58. Sessa G, Gómez-González B, Silva S, *et al.* BRCA2 promotes DNA-RNA hybrid resolution by DDX5 helicase at DNA breaks to facilitate their repair. *EMBO J.* 2021 ; 40 : e106018.
59. Bown N. Neuroblastoma tumour genetics: clinical and biological aspects. *J. Clin. Pathol.* 2001 ; 54 : 897–910.
60. Paolini L, Hussain S, Galardy PJ. Chromosome instability in neuroblastoma: A pathway to aggressive disease. *Front. Oncol.* 2022 ; 12 .
61. Boumpas P, Merabet S, Carnesecchi J. Integrating transcription and splicing into cell fate: Transcription factors on the block. *WIREs RNA* 2023 ; 14 : e1752.
62. Rambout X, Dequiedt F, Maquat LE. Beyond Transcription: Roles of Transcription Factors in Pre-mRNA Splicing. *Chem. Rev.* 2018 ; 118 : 4339–4364.
63. Fusby B, Kim S, Erickson B, *et al.* Coordination of RNA Polymerase II Pausing and 3' End Processing Factor Recruitment with Alternative Polyadenylation. *Mol. Cell. Biol.* 2016 ; 36 : 295–303.
64. Lin J, Hung F-Y, Ye C, *et al.* HDA6 dependent chromatin deacetylation orchestrates mRNA polyadenylation. *Genome Res.* 2020 ; gr.255232.119.
65. Zhang J, Zhang Y-Z, Jiang J, *et al.* The Crosstalk Between Epigenetic Mechanisms and Alternative RNA Processing Regulation. *Front. Genet.* 2020 ; 11 .
66. Soles LV, Shi Y. Crosstalk Between mRNA 3'-End Processing and Epigenetics. *Front. Genet.* 2021 ; 12 .
67. McMahon SB, Buskirk HAV, Dugan KA, *et al.* The Novel ATM-Related Protein TRRAP Is an Essential Cofactor for the c-Myc and E2F Oncoproteins. *Cell* 1998 ; 94 : 363–374.
68. Tu WB, Helander S, Pilstål R, *et al.* Myc and its interactors take shape. *Biochim. Biophys. Acta BBA - Gene Regul. Mech.* 2015 ; 1849 : 469–483.
69. Sun Y, Bell JL, Carter D, *et al.* WDR5 Supports an N-Myc Transcriptional Complex That Drives a Protumorigenic Gene Expression Signature in Neuroblastoma. *Cancer Res.* 2015 ; 75 : 5143–5154.
70. Dardenne E, Polay Espinoza M, Fattet L, *et al.* RNA Helicases DDX5 and DDX17 Dynamically Orchestrate Transcription, miRNA, and Splicing Programs in Cell Differentiation. *Cell Rep.* 2014 ; 7 : 1900–1913.
71. Chen S, Zhou Y, Chen Y, *et al.* fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 2018 ; 34 : i884–i890.
72. Dobin A, Davis CA, Schlesinger F, *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 2013 ; 29 : 15–21.
73. Danecek P, Bonfield JK, Liddle J, *et al.* Twelve years of SAMtools and BCFtools. *GigaScience* 2021 ; 10 : giab008.
74. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014 ; 15 : 1–21.
75. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data n.d. ;.
76. Ewels P, Magnusson M, Lundin S, *et al.* MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 2016 ; 32 : 3047–3048.
77. Di Tommaso P, Chatzou M, Floden EW, *et al.* Nextflow enables reproducible computational workflows. *Nat. Biotechnol.* 2017 ; 35 : 316–319.

78. Langmead B, Wilks C, Antonescu V, *et al.* Scaling read aligners to hundreds of threads on general-purpose processors. *Bioinformatics* 2019 ; 35 : 421–432.
79. Ramírez F, Ryan DP, Grüning B, *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* 2016 ; 44 : W160-165.
80. Zhang Y, Liu T, Meyer CA, *et al.* Model-based Analysis of CHIP-Seq (MACS). *Genome Biol.* 2008 ; 9 : 1–9.

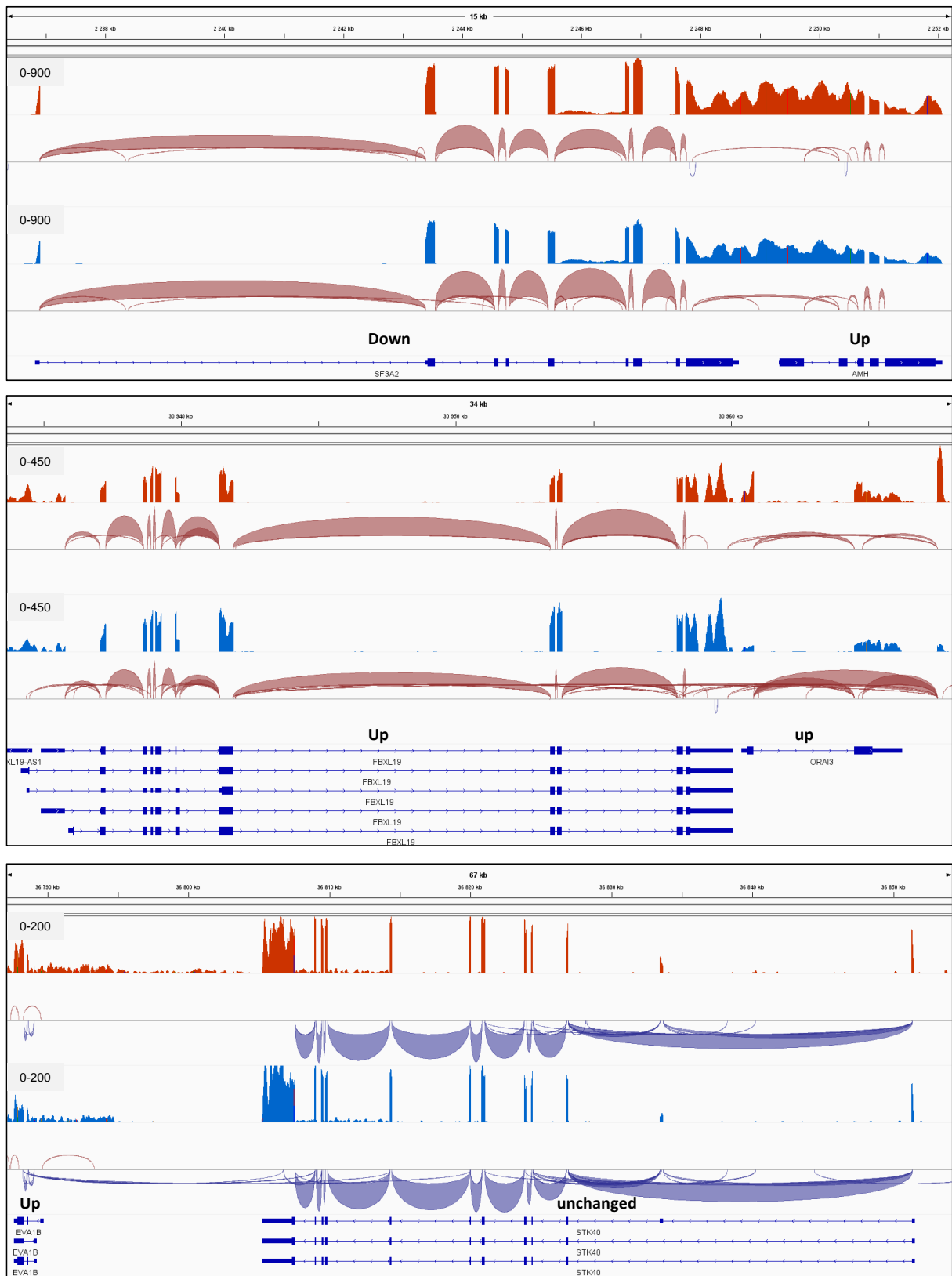
Figure S1A (related to Figure 1)



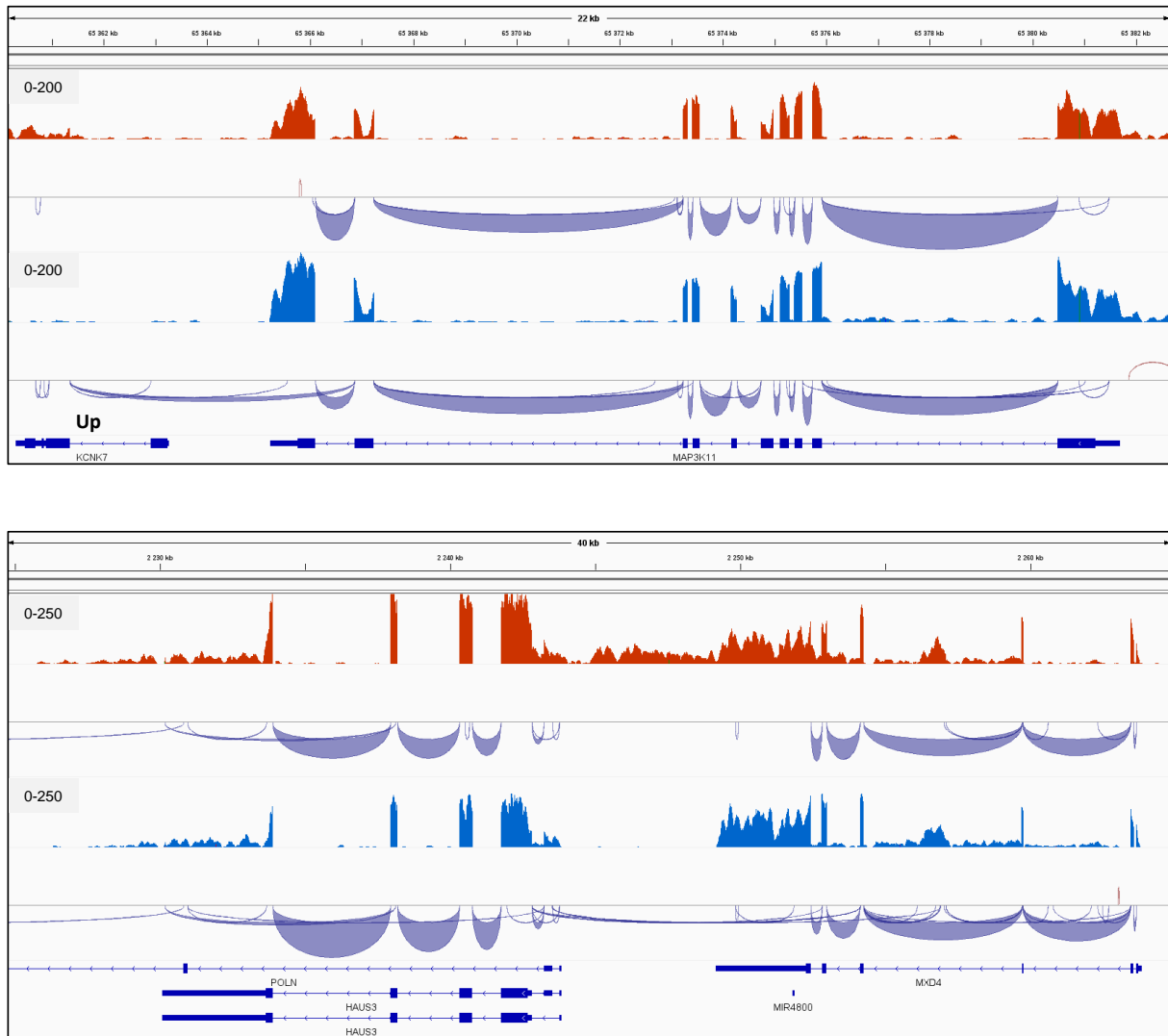
**Figure S1B (related to Figure 1)**



**Figure S1B (continuation)**



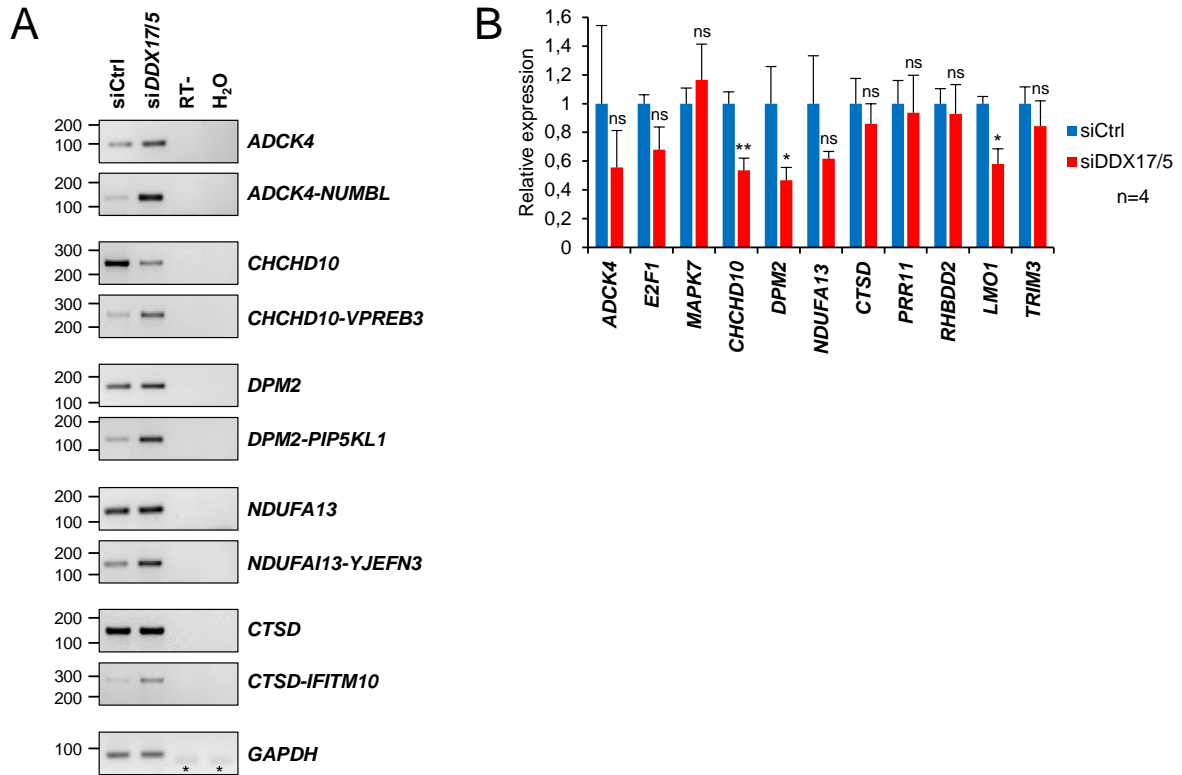
## Figure S1B (continuation)



### Figure S1 : Visualization of chimeric transcripts expression in the absence of DDX5/17.

- (A) IGV (Integrative Genomics Viewer) screenshots of a representative example of genes (CTDS-IFITM10) presenting an increase of reads mapping both to exons of an upstream gene and exon of a downstream gene in condition silencing of both DDX5 and DDX17 (red), compared to a control siRNA (blue). The RefSeq genomic annotation is shown for each gene (in blue), as well as the gene orientation (black arrow).
- (B) IGV screenshots of the splicing pattern of chimeric transcripts induced by silencing of both DDX5 and DDX17 (in blue) compared to a control siRNA (red). The RefSeq genomic annotation is shown for each gene (in blue).

## Figure S2 (related to figure 1)

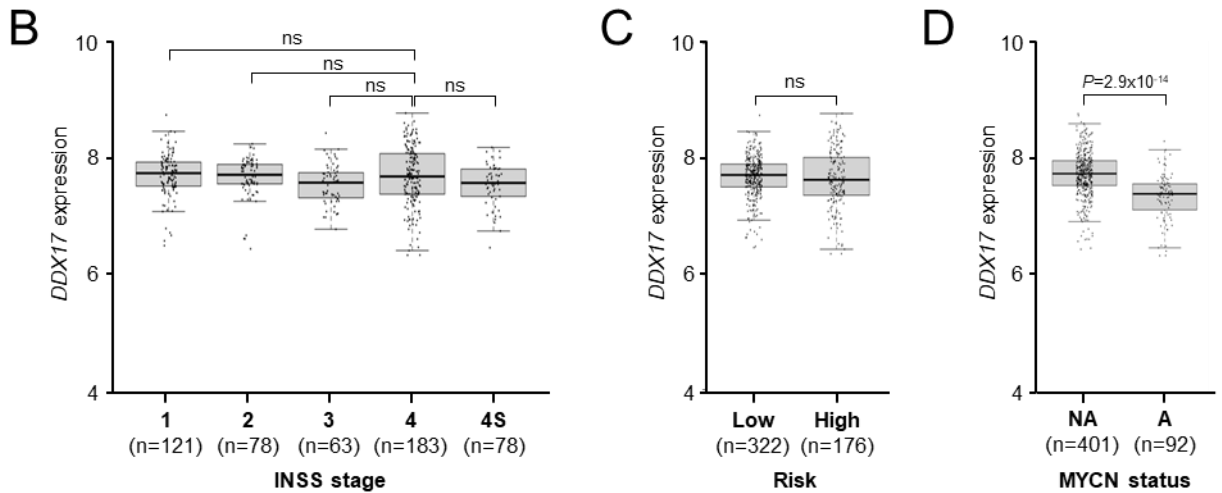
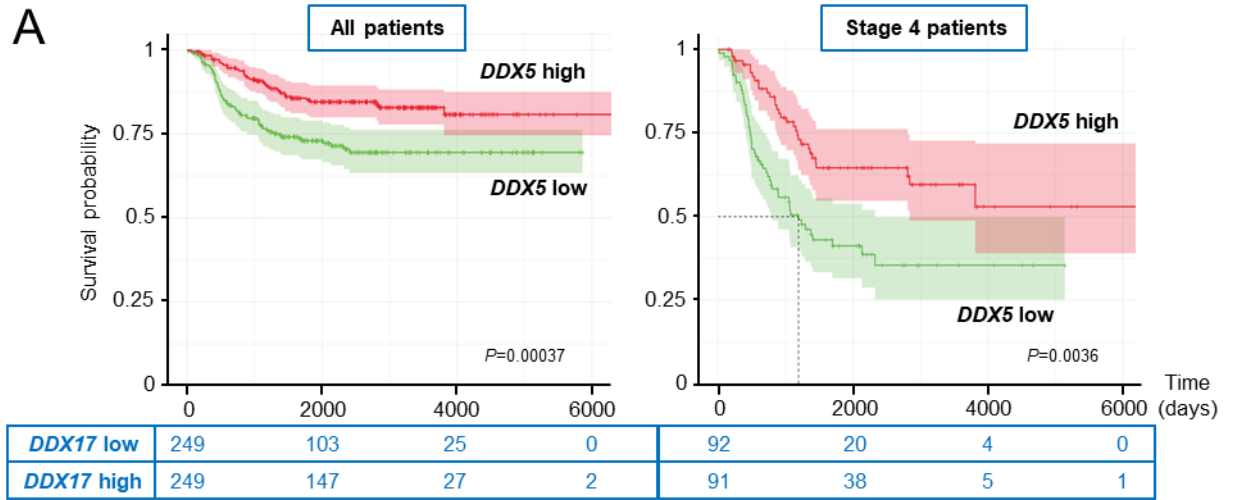


**Figure S2 : Effect of DDX5 and DDX17 silencing on parental genes expression**

(A) RT-PCR gel showing the effect of DDX5 and DDX17 silencing on upstream parental gene transcript expression and chimeric expression

(B) Quantification of upstream parental gene transcript expression by RT-qPCR

**Figure S3 (related to figure 3)**



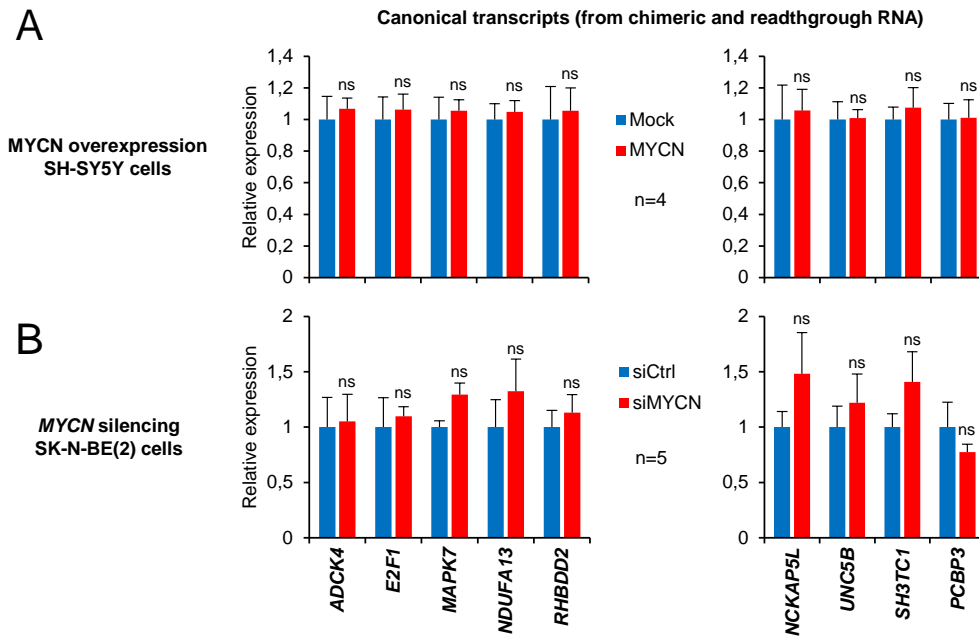
**Figure S3 : Expression of DDX5 in neuroblastoma tumors.**

(A) Kaplan-Meier curves indicating the overall survival of 498 well-defined NB patients (GEO: GSE62564) and for stage 4 patients only, with high or low DDX5 expression.

(B-D) Box plots of DDX5 expression relative to tumor stage (B), low/high risk classification (C) and amplified (A) or non amplified (NA) MYCN status (D).



## Figure S4 (related to figure 5)

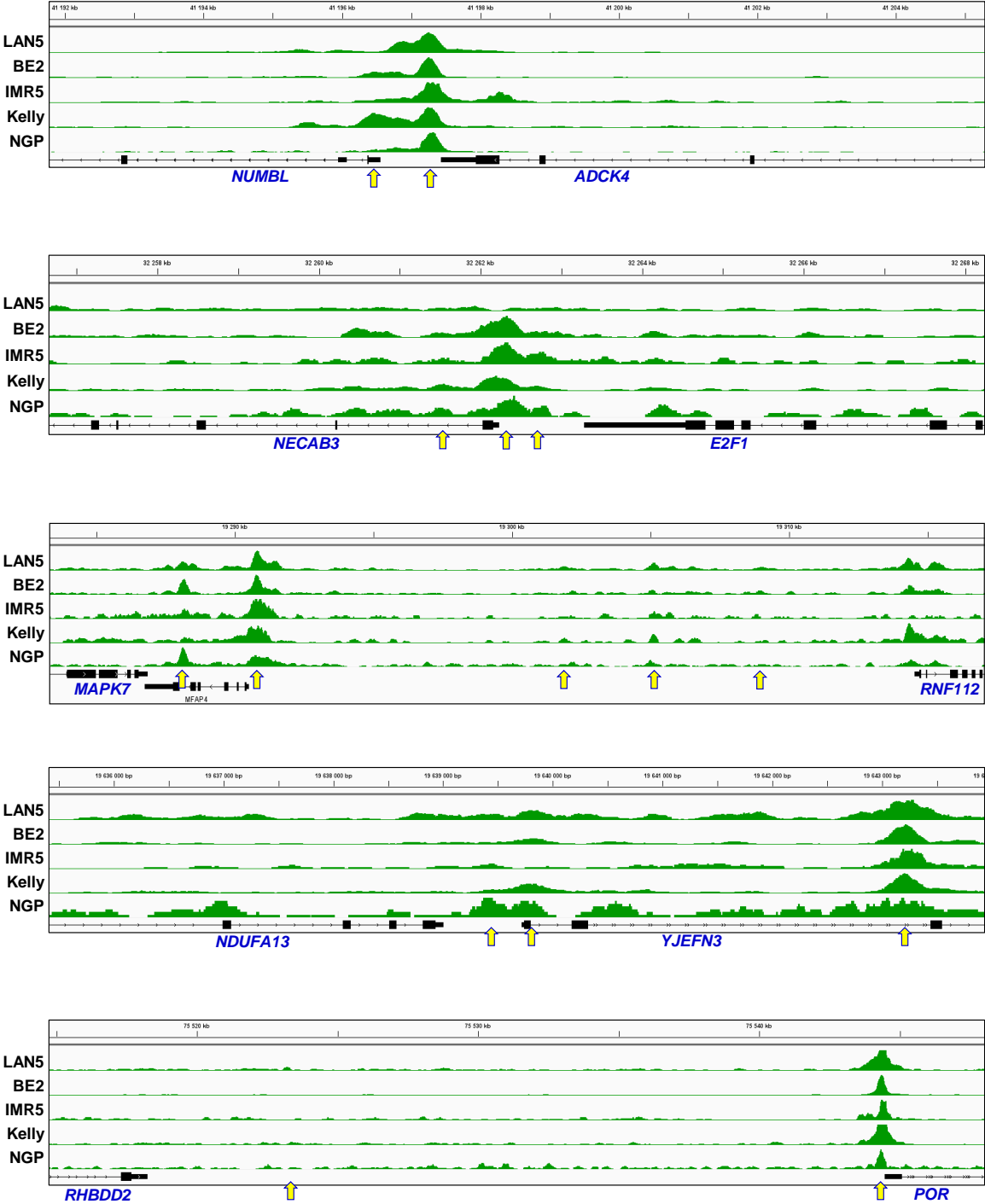


**Figure S4 : Effect of MYCN silencing and overexpression on parental genes expression**

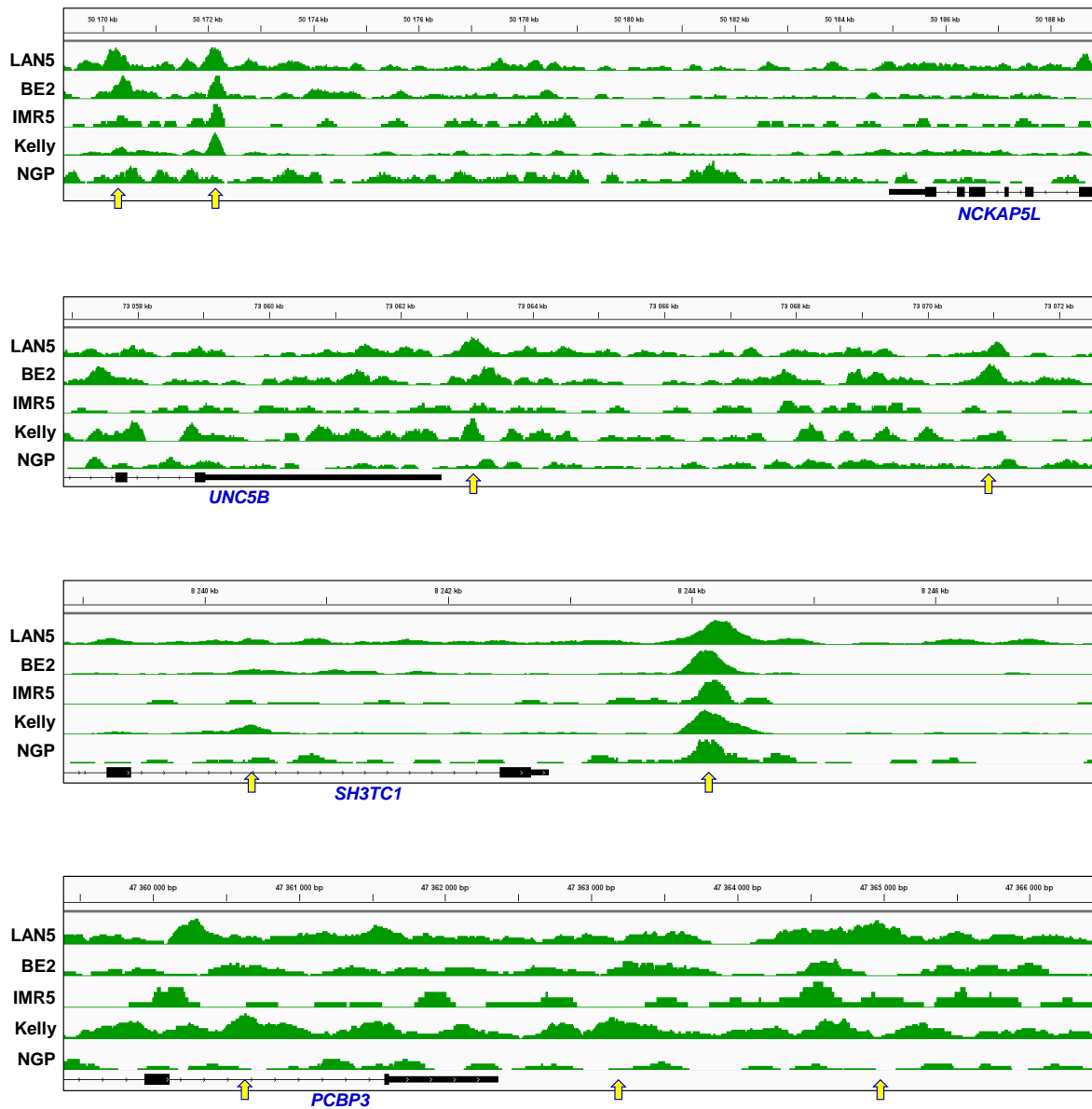
(A) Quantification of the effect of MYCN overexpression on upstream parental gene transcript expression by RT-qPCR

(B) Quantification of the effect of MYCN silencing on upstream parental gene transcript expression by RT-qPCR

Figure S5 (related to Figure 6)



## Figure S5 (continuation)



### Figure S5 : MYCN binds to the 3' of DDX5/17 regulated genes

IGV screenshots of gene regulated by DDX5/17 and displaying MYCN binding to their 3' end in neuroblastoma MYCN amplified cell lines. Yellow arrows indicate MYCN binding sites identified as statistically significant.

**Table S1 : chimeric rna siDDX5/17 SHSY5Y**

ID	baseMean	log2FoldChange	FC	pvalue
PHOX2A_RP11-807H22.7	27,13	8,01	257,78	5,40E-10
<b>RASA4B_POLR2J</b>	44,88	2,57	5,94	6,90E-08
CRIP2_CRIP1	20,37	5,01	32,22	5,16E-07
<b>CACNG4_CACNG1</b>	16,81	4,49	22,47	3,41E-06
DPM2_PIP5KL1	16,56	4,08	16,91	1,24E-05
PDIA6_RP11-791G15.2	20,2	2,57	5,94	0,000122188
C19orf48_CTD-2568A17.1	10,91	4,81	28,05	0,000140729
LRRC61_RP4-584D14.7	10,11	4,75	26,91	0,000206149
AEN_ISG20	10,9	4,26	19,16	0,000207083
ARF3_FKBP11	11,37	3,74	13,36	0,000278486
CNPPD1_SLC23A3	5,39	5,67	50,91	0,000361406
SF3A2_AMH	17,59	2,56	5,90	0,000380093
FBXL19_ORAI3	4,97	5,55	46,85	0,00059978
<b>CHCHD10_VPREB3</b>	9,74	3,57	11,88	0,000685812
PSEN2_ADCK3	6,67	4,91	30,06	0,000837031
STK40_EVA1B	5,19	5,31	39,67	0,001016476
SNHG15_MYO1G	4,45	5,41	42,52	0,001107631
TONSL_VPS28	6,28	4,93	30,48	0,001199475
<b>MAP3K11_KCNK7</b>	5,86	4,83	28,44	0,001536009
EEF1D_NAPRT1	13,48	2,49	5,62	0,00154833
<b>NTRK1_PEAR1</b>	8,09	4,14	17,63	0,001557685
UBXN1_C11orf48	7,42	4,27	19,29	0,001624956
PBXIP1_PMVK	6,05	4,59	24,08	0,001665892
MXD4_HAUS3	4,39	5,14	35,26	0,002111682
CNPY3_RP3-475N16.1	17,84	2,06	4,17	0,002340504
STARD3_TCAP	3,9	5,17	36,00	0,00257504
DCHS1_TPP1	3,89	5,17	36,00	0,002589754
<b>ZYX_EPHA1-AS1</b>	3,94	5,14	35,26	0,00269328
SLC35E2_SLC35E2B	20,01	-2,04	<b>-4,11</b>	0,003435513
<b>PEAR1_LRRC71</b>	4,88	4,55	23,43	0,004432672
ARID3C_DCTN3	9,48	2,72	6,59	0,004496015
MBTPS1_SLC38A8	3,78	4,89	29,65	0,004949202
UBE2L3_CCDC116	3,36	5	32,00	0,005008772
RGP1_RP11-112J3.16	19,62	1,89	3,71	0,005102278
BSDC1_FAM229A	5,96	3,88	14,72	0,005570787
<b>PFKP_PITRM1-AS1</b>	8,92	2,78	6,87	0,006244018
<b>E2F1_NECAB3</b>	6,19	3,52	11,47	0,006283783
<b>FAM173A_HAGHL</b>	3,37	4,88	29,45	0,006344402
IL17RC_CRELD1	3,08	4,8	27,86	0,008710359
<b>TUBGCP2_ADAM8</b>	3,28	4,7	25,99	0,009302271
TMEM258_RP11-467L20.10	2,98	4,79	27,67	0,009357547
ALKBH6_SYNE4	4,11	4,32	19,97	0,009383918
ARHGEF25_SLC26A10	11,23	2,17	4,50	0,009710483
CLPP_ALKBH7	2,94	4,73	26,54	0,010726252
<b>C19orf60_TMEM59L</b>	2,87	4,7	25,99	0,011858426
CXXC5_CTB-35F21.1	4,96	3,66	12,64	0,013096983
SNAPC4_CARD9	3,98	4,15	17,75	0,013424112
GFOD2_ENKD1	2,71	4,68	25,63	0,013532667
TMEM234_RP4-622L5.7	2,87	4,58	23,92	0,014408725
<b>DDX56_NPC1L1</b>	2,74	4,58	23,92	0,015766955
FUOM_RP11-122K13.7	4,3	3,77	13,64	0,015846008

FAM200B_BST1	3,68	3,98	15,78	0,016211946
IL18BP_RP11-849H4.4	3,57	4,1	17,15	0,016739445
AC002116.8_SYNE4	3,59	4,1	17,15	0,016858258
TMEM222_SYTL1	2,7	4,53	23,10	0,017498204
DCUN1D2_ADRHL1	3,44	4	16,00	0,017637715
ADCK4_NUMBL	11,07	2,03	4,08	0,018037179
LDLRAP1_RP11-70P17.1	5,15	3,18	9,06	0,018929917
HOXD3_HOXD4	3,29	3,94	15,35	0,020988499
MAPKAPK3_RP11-804H8.6	3,84	3,77	13,64	0,024662565
RANGAP1_CHADL	2,3	4,41	21,26	0,027271029
AP5S1_MAVS	3,17	3,88	14,72	0,028458677
CYB561_RP11-269G24.4	3,33	3,73	13,27	0,029502459
<b>PHC2_A3GALT2</b>	2,27	4,36	20,53	0,029651184
<b>ZDHHC1_TPPP3</b>	5,73	2,72	6,59	0,030483175
SIRT7_PCYT2	3,03	3,86	14,52	0,030543805
<b>SPATA5L1_C15orf48</b>	2,17	4,37	20,68	0,031451416
<b>TNK2_AC124944.5</b>	2,16	4,37	20,68	0,031587826
<b>SLC29A1_HSP90AB1</b>	3,46	3,67	12,73	0,031901816
TFB1M_RP11-477D19.2	4,76	-3,06	-8,34	0,035095235
NES_RP11-284F21.7	2,05	4,22	18,64	0,041806729
C8orf82_LRRC24	2,04	4,22	18,64	0,042133332
AC069368.3_ANKDD1A	3,9	3,16	8,94	0,04222042
BOP1_KM-PA-2	7,82	-1,88	-3,68	0,043468013
TMEFF1_MURC	11,7	-1,61	-3,05	0,044937821
YJEFN3_CILP2	2,03	4,17	18,00	0,044993272
<b>SORBS3_PDLIM2</b>	1,89	4,14	17,63	0,051721521
PLEKHO2_ANKDD1A	4,23	2,78	6,87	0,053005447
<b>RUNDC1_RPL27</b>	2,61	3,6	12,13	0,053699384
UBE2Q2_FBXO22	3,88	-2,82	-7,06	0,061959724
PAQR4_LINC00514	1,77	3,97	15,67	0,068671237
<b>PELI3_DPP3</b>	3,22	2,94	7,67	0,073100007
PRKAA1_TTC33	3,68	-2,64	-6,23	0,073102045
PLEKHM1_ARHGAP27	1,75	3,92	15,14	0,073104518
XPO7_NPM2	2,76	3,2	9,19	0,073310504
C6orf47_BAG6	10,05	1,48	2,79	0,074772528
<b>PRDX5_CCDC88B</b>	1,77	3,87	14,62	0,07664202
GPR153_ICMT	1,62	3,94	15,35	0,078784832
<b>PLEKHA6_GOLT1A</b>	1,61	3,93	15,24	0,080200273
HTT_MSANTD1	1,64	3,88	14,72	0,083265142
GS1-124K5.12_GS1-124K5.11	2,52	-3,1	-8,57	0,084351683
<b>MAPK7_RNF112</b>	1,64	3,86	14,52	0,084448243
RNF121_IL18BP	1,64	3,86	14,52	0,085143595
ST6GALNAC4_ST6GALNAC6	1,64	3,86	14,52	0,085143595
SIRT2_RINL	1,63	3,86	14,52	0,085790108
TMEM223_RP11-727F15.11	3,68	2,43	5,39	0,088022055
<b>POLD2_POLM</b>	2,1	3,3	9,85	0,095966405
SLC5A3_MRPS6	2,39	-3,09	-8,51	0,100629971
ZNF500_RP11-127I20.7	1,47	3,79	13,83	0,101757417
TBC1D16_RP11-353N14.5	1,5	3,76	13,55	0,102662522
AC003665.1_RP11-6N17.4	2,68	-2,71	-6,54	0,104379174
PSME3_AOC2	1,48	3,75	13,45	0,106083266
<b>KRIT1_LRRD1</b>	1,58	3,68	12,82	0,106289065
OTUD6B_LRRC69	1,63	-3,45	-10,93	0,106589053
CNPY3_GNMT	3,92	2,32	4,99	0,106600308

PSMB2_RP4-728D4.2	3,23	2,49	5,62	0,109968457
TMEM203_TPRN	1,96	3,19	9,13	0,111751028
DSE_RP1-93H18.1	3,56	-2,17	-4,50	0,115737146
CDR2L ICT1	3,01	2,61	6,11	0,116419531
PLEKHA6_KISS1	1,51	3,62	12,30	0,117117479
CD276_RP11-8P11.3	1,51	3,62	12,30	0,117117479
MLLT6_CISD3	10,55	1,24	2,36	0,123617055
MAPK11_MAPK12	1,34	3,68	12,82	0,124115754
TMEM141_CCDC183	1,34	3,64	12,47	0,129964538
CHRNB4_CHRNA3	1,36	3,62	12,30	0,130056801
SLC2A8_ZNF79	1,36	3,62	12,30	0,130056801
FASTK_RP11-148K1.12	5,36	1,72	3,29	0,130495659
HCG18_HCG17	12,59	1,11	2,16	0,133723485
NUDT14_JAG2	1,33	3,61	12,21	0,133982786
PSD2_CTB-35F21.4	1,33	3,61	12,21	0,133982786
TMEM98_SPACA3	1,33	3,61	12,21	0,133982786
NAA40_COX8A	1,82	3,06	8,34	0,136323066
METTL21B_TSFM	1,48	-3,3	-9,85	0,14021565
ATP7A_RP5-1000K24.2	1,97	-2,78	-6,87	0,14218239
VPS45_PLEKHO1	4,89	-1,72	-3,29	0,14243429
SAFB2_TINCR	1,89	3	8,00	0,144394064
MAPKAPK5_ADAM1A	2,32	-2,54	-5,82	0,148247835
PAM16_AC005356.1	1,77	2,98	7,89	0,150657846
PLEKHG5_TNFRSF25	1,82	2,97	7,84	0,151101483
TAS2R6_TAS2R5	1,92	-2,84	-7,16	0,153620805
RIMS4_RP11-445H22.4	1,22	3,55	11,71	0,153813525
CNRIP1_PPP3R1	1,99	2,89	7,41	0,155528426
MZT2B_AC018804.6	2,41	2,47	5,54	0,157415567
IVD_BAHD1	1,23	3,51	11,39	0,157476971
FAM96B_RRAD	2,16	2,74	6,68	0,157711067
IGSF3_CD58	1,22	3,51	11,39	0,158752418
RBM12B_LINC00535	1,21	3,5	11,31	0,160734763
CHCHD5_AC012442.5	10,68	1,13	2,19	0,162980674
SCRT1_DGAT1	1,69	2,96	7,78	0,163699609
PDXP_LGALS1	3,47	2,06	4,17	0,164516632
ARMC6_SLC25A42	5,32	1,58	2,99	0,166210317
LGALS1_NOL12	2,14	2,65	6,28	0,166539248
CCDC154_C16orf91	1,83	2,78	6,87	0,171071145
TARDBP_RP4-635E18.8	3,9	1,95	3,86	0,171365694
KMT2D_PRKAG1	1,68	2,89	7,41	0,171811241
TPCN2_RP11-554A11.9	2,02	2,73	6,63	0,172511881
VAC14_MTSS1L	1,2	3,42	10,70	0,17262051
COQ4_SLC27A4	2,03	2,65	6,28	0,172655357
ORAI3_SETD1A	2,73	2,24	4,72	0,172717437
ACYP1_MLH3	1,58	-2,96	-7,78	0,174760845
HACL1_COLQ	2,31	-2,38	-5,21	0,1755789
NSUN4_FAAH	3,56	1,86	3,63	0,178999362
STAG3L3_TRIM74	3,59	1,86	3,63	0,17920914
WDR90_RHOT2	1,69	2,86	7,26	0,180050812
OBSL1_CHPF	3,44	1,99	3,97	0,181920616
SLC35A2_TIMM17B	1,56	2,85	7,21	0,190861056
STYXL1_TMEM120A	8,41	-1,19	-2,28	0,19158136
DSCC1_TAF2	2,08	-2,35	-5,10	0,19233742
GNAI2_SEMA3B	1,09	3,36	10,27	0,195397254

SLC39A7_HSD17B8	1,08	3,36	10,27	0,196957062
JPH4_AP1G2	1,56	2,78	6,87	0,198167358
ZNF707_MAPK15	1,08	3,35	10,20	0,198177139
ANKRD34A_LIX1L	2,07	-2,35	-5,10	0,19847529
MAGEA12_CSAG4	1,51	2,82	7,06	0,199687052
XKR6_SOX7	1,09	3,31	9,92	0,203355202
NSFP1_NSF	3,15	-1,92	-3,78	0,205066668
ARF5_FSCN3	1,5	2,79	6,92	0,205436161
IL17RD_ARHGEF3	4,18	1,75	3,36	0,206906955
IRX5_CTD-3032H12.2	1,53	2,73	6,63	0,207740607
TEAD2_CTC-30107.4	2,11	2,25	4,76	0,214442758
WIZ_AKAP8L	1,56	2,71	6,54	0,217812042
RHBDD2_POR	3,26	1,86	3,63	0,220387044
ARMC9_B3GNT7	3,89	1,73	3,32	0,220863468
MSANTD3_TMEFF1	6,55	-1,31	-2,48	0,223763566
KIF18A_RP11-406D1.2	2,1	-2,24	-4,72	0,224627196
SLC25A29_AL157871.2	2,12	2,24	4,72	0,22508793
ZNF212_ZNF783	1,42	2,7	6,50	0,227063928
GLYCKT_DNAH1	3,07	1,85	3,61	0,227378742
DISP1_RP11-452F19.3	3,25	-1,73	-3,32	0,2279522
NELFB_TOR4A	1,75	2,51	5,70	0,22980362
KLF16_REXO1	1,75	2,51	5,70	0,22980362
RASL10B_C17orf50	1,42	2,7	6,50	0,230372424
DHRS13_FLOT2	1,38	2,68	6,41	0,23624706
FAM175A_HELQ	1,55	-2,6	-6,06	0,236780983
FKBP11_RP11-302B13.5	1,38	2,68	6,41	0,237269056
ZNF738_CTD-2561J22.2	9,01	-1,03	-2,04	0,237790301
USP4_C3orf62	6,26	-1,26	-2,39	0,237899091
WEE2-AS1_KIAA1147	2,8	-1,89	-3,71	0,240384327
ARMC5_TGFB1I1	0,95	3,17	9,00	0,246668821
PCDHB14_PCDHB15	0,95	3,17	9,00	0,246668821
TXLNA_CCDC28B	0,95	3,17	9,00	0,246668821
POLR2J2_POLR2J3	4,7	-1,33	-2,51	0,247329438
ASPDH_JOSD2	2,51	1,96	3,89	0,247554713
PTGFRN_CD101	0,95	3,16	8,94	0,2480766
HPS1_PYROXD2	2,48	1,95	3,86	0,251128581
FAM89B_EHBP1L1	1,4	2,54	5,82	0,25596459
TRAI_PUBA7	1,38	-2,78	-6,87	0,259232685
MON1B_SYCE1L	1,85	2,13	4,38	0,26386542
SLC2A11_MIF	1,69	2,39	5,24	0,265634327
CTBS_GNG5	3,09	-1,59	-3,01	0,270581559
U73166.2_SEMA3B	6,33	1,12	2,17	0,278399759
ERI1_RP11-10A14.3	3,48	-1,46	-2,75	0,280179709
PTGES2_RP11-395P17.3	3,09	1,57	2,97	0,281844391
KCNIP2_MGEA5	5,49	1,31	2,48	0,282276242
Z83851.1_TCF20	1,61	2,28	4,86	0,282655009
PMS2P3_POM121C	1,87	2,07	4,20	0,285269825
SRSF1_RP11-159D12.2	1,41	-2,25	-4,76	0,289132585
FAM167A_AF131216.6	1,29	2,44	5,43	0,29006802
ZNF606_C19orf18	1,29	2,44	5,43	0,29006802
MAP2K5_SKOR1	1,24	2,48	5,58	0,290280535
SSSCA1_EHBP1L1	1,27	2,44	5,43	0,293074915
ADM2_MIOX	2,47	1,77	3,41	0,293513835
EIF4G3_HP1BP3	2,46	-1,81	-3,51	0,29373597

UBL7-AS1_RP11-10017.3	3,81	1,34	2,53	0,297691113
SLC6A2_CES1P2	2,69	1,68	3,20	0,297988909
SLC35G2_NCK1	1,75	-2,01	-4,03	0,2997893
PRIM1_NACA	4,4	-1,22	-2,33	0,301053534
PAGR1_MVP	3,16	1,58	2,99	0,301861689
SFT2D2_TBX19	6,25	1,17	2,25	0,304072068
C21orf33_AP001057.1	0,81	2,96	7,78	0,308569806
HERC3_FAM13A-AS1	1,21	-2,4	-5,28	0,309894812
ZNF772_VN1R1	1,63	2,2	4,59	0,31151292
NFYC-AS1_RP4-739H11.3	5,62	1,07	2,10	0,313385368
SLC35A3_HIAT1	4,8	-1,13	-2,19	0,31889545
TMED4_DDX56	3,99	-1,21	-2,31	0,324177945
MFGE8_HAPLN3	3,37	1,34	2,53	0,327517986
CIRBP_C19orf24	1,57	2,08	4,23	0,333095668
NTN3_TBC1D24	1,14	2,31	4,96	0,338537704
AP005482.1_PSMG2	1,86	-1,8	-3,48	0,340097112
APEH_RNF123	2,81	1,42	2,68	0,341972355
TBCEL_TECTA	4,93	-1,07	-2,10	0,34251974
THBS4_RP11-168A11.4	5,18	1,14	2,20	0,346578787
AC021188.4_SNRNP200	2,64	-1,4	-2,64	0,359985686
ADORA2A_UPB1	1,58	1,82	3,53	0,367492544
ACCS_EXT2	3,17	1,35	2,55	0,368938255
UNC79_COX8C	2,43	1,52	2,87	0,369768288
GPBP1L1_CCDC17	1,34	2,01	4,03	0,374561373
DCAF7_TACO1	1,66	1,81	3,51	0,382385582
EGLN2_CTC-490E21.12	1,89	1,61	3,05	0,386222449
SPECC1L-ADORA2A_UPB1	1,59	1,73	3,32	0,387581805
NSL1_BATF3	1,35	1,93	3,81	0,388964596
NSUN6_RP11-499P20.2	1,52	-1,71	-3,27	0,39986038
DLX6-AS2_DLX6-AS1	2,3	-1,44	-2,71	0,399988162
L3MBTL1_SGK2	1,55	1,82	3,53	0,400242454
RAD51AP1_DYRK4	1,9	-1,56	-2,95	0,404117605
CLUHP3_ZNF720	3,96	-1,07	-2,10	0,405540804
CEP152_RP11-227D13.4	1,9	-1,52	-2,87	0,411150902
NBPF15_NBPF16	2,98	-1,33	-2,51	0,416385585
FEN1_FADS2	2,22	-1,38	-2,60	0,438582307
ZNF551_AC003006.7	2,96	1,17	2,25	0,438735533
MTAP_RP11-149I2.5	2,85	1,13	2,19	0,439870764
HARS2_ZMAT2	2,66	-1,14	-2,20	0,440580361
CTNNBIP1_CLSTN1	3,09	-1,09	-2,13	0,445430457
SIDT2_TAGLN	3,77	1,02	2,03	0,455241377
MPND_AC007292.6	3,9	1,02	2,03	0,458526301
TRADD_B3GNT9	3,16	1,05	2,07	0,461346584
AC009961.3_BAZ2B	1,37	-1,54	-2,91	0,4646653
PRR11_SMG8	2,1	1,32	2,50	0,467746707
MRPS10_GUCA1B	2,66	-1,09	-2,13	0,470481543
DIXDC1_DLAT	1,34	-1,53	-2,89	0,472281994
UBE2D2_CXXC5	2,26	-1,24	-2,36	0,483066911
SUMO1_AC079354.5	2,83	1,02	2,03	0,494944726
HCG15_HCG16	1,15	1,74	3,34	0,508169887
STRADA_LIMD2	1,66	1,28	2,43	0,508304027
NRBP1_KRTCAP3	1,79	1,26	2,39	0,510576633
RNF216P1_RBAK	1,74	1,27	2,41	0,511219085
METTL10_FAM53B	1,76	-1,31	-2,48	0,524955518



EARS2_GGA2	1,63	1,24	2,36	0,525990131
VN1R108P_RP13-401N8.1	2	1,15	2,22	0,526957408
GALNT13_PHBP4	1,49	-1,29	-2,45	0,528065576
ANGEL1_RP11-488C13.5	1,78	-1,1	-2,14	0,557916228
EFEMP2_CFL1	1,18	1,29	2,45	0,561604059
RNF152_RP11-879F14.1	1,01	1,44	2,71	0,564840435
SH3BP1_LGALS1	1,64	1,12	2,17	0,565308933
PNKP_AC018766.4	1,49	-1,26	-2,39	0,566130674
RCCD1_PRC1-AS1	2,19	1	2,00	0,570800085
PPRC1_NOLC1	1,67	-1,04	-2,06	0,572158335
AC092574.2_ZNF721	1,47	-1,21	-2,31	0,573010727
BCKDHB_RP11-250B2.3	1,65	-1,02	-2,03	0,580455203
TBC1D24_ATP6VOC	1,66	1,04	2,06	0,595402356
ZNF619_ZNF620	1,73	-1,01	-2,01	0,620763944
ZNF747_ZNF768	1,37	1,17	2,25	0,621744534

**Table s1 : chimeric rna identified in siDDX5/17 SHSY5Y**

Blue and orange boxes define up and down transcripts respectively. Those in green are the most significant ( $p\text{-val} < 0.1$ ). IDs in blue are those found in tumors, and those in bold are the most significant.

## Partie 2

Le papier en pré-publication présenté est le résultat d'une collaboration avec trois autres équipes.

L'analyse du génome d'un ensemble de 13.494 trios de parents/enfant souffrants de maladies rares et l'utilisation de l'outil Genematcher a permis d'identifier 11 patients souffrant de troubles neuro-développementaux similaires et porteurs de mutations hétérozygotes *de novo* dans le gène *DDX17*. *DDX17* est déjà connu être un coactivateur transcriptionnel de REST, un facteur de transcription critique dans les premières phases de la différenciation neuronale (Lambert et al., 2018). Par ailleurs, *DDX17* et *DDX5* sont nécessaires pour contrôler l'expression transcriptionnelle de gènes associés à la neurogenèse, et sont essentielles à la différenciation neuronale précoce des CSPh (Suthapot et al., 2022). Cependant, ces études ont utilisé une approche de co-déplétion de *DDX17* et *DDX5* pour étudier la fonction de ces facteurs dans la neurogenèse, et des informations concernant la contribution spécifique de chaque hélicase à ce processus manquent. Si plusieurs hélicases DEAD-box ont déjà été impliquées dans des maladies génétiques neuro-développementales, les hélicases *DDX5* et *DDX17* n'avaient jusqu'à maintenant jamais été associées à des problèmes développementaux chez l'humain. Nous avons donc cherché dans un premier temps à caractériser le rôle de *DDX17* dans le développement neuronal à l'aide de plusieurs modèles : cellules humaines, modèles murin et xénope. Notre équipe s'est concentrée sur le modèle humain, en étudiant l'effet de la déplétion de *DDX17* sur le transcriptome de cellules neuronales. Nos résultats révèlent que la déplétion de *DDX17* entraîne la dérégulation de 350 gènes, qui présentent un enrichissement dans plusieurs fonctions associées au développement et au fonctionnement neuronal. L'expression de plusieurs facteurs de transcription associés au développement ainsi que celle de gènes associés à la guidance axonale est ainsi dérégulée en l'absence de *DDX17*. Parallèlement, les travaux menés par nos collaborateurs montrent que la déplétion de *DDX17* entraîne une diminution de la complexité et de la croissance axonale chez la souris. Enfin, chez le xénope, les animaux KO hétérozygotes pour *DDX17* présentent une diminution de la croissance axonale ainsi que de la mémoire de travail

# 1 Monoallelic *de novo* variants in *DDX17* cause a novel neurodevelopmental disorder

2  
3 Eleanor G. Seaby<sup>1,2&</sup>, Annie Godwin<sup>3</sup>, Valentine Clerc<sup>4</sup>, Géraldine Meyer-Dilhet<sup>5</sup>, Xavier Grand<sup>6,7</sup>, Tia  
4 Fletcher<sup>3</sup>, Laloe Monteiro<sup>5</sup>, Valerio Carelli<sup>8,9</sup>, Flavia Palombo<sup>8</sup>, Marco Seri<sup>10</sup>, Giulia Olivucci<sup>10</sup>, Mina  
5 Grippa<sup>10</sup>, Claudia Ciaccio<sup>11</sup>, Stefano D'Arrigo<sup>11</sup>, Maria Iascone<sup>12</sup>, Marion Bermudez<sup>13</sup>, Jan Fischer<sup>13</sup>,  
6 Nataliya Di Donato<sup>13</sup>, Sophie Goesswein<sup>13</sup>, Marco L. Leung<sup>14</sup>, Daniel C. Koboldt<sup>15</sup>, Cortlandt Myers<sup>16</sup>,  
7 Dennis Bartholomew<sup>16</sup>, Gudny Anna Arnadottir<sup>17</sup>, Kari Stefansson<sup>17</sup>, Patrick Sulem<sup>17</sup>, Ethan M.  
8 Goldberg<sup>18,19</sup>, Ange-Line Bruel<sup>20,21</sup>, Frederic Tran Mau Them<sup>20,21</sup>, Marjolaine Willems<sup>22</sup>, Hans Tomas  
9 Bjornsson<sup>23</sup>, Hakon Bjorn Hognason<sup>23</sup>, Eirny Tholl Thorolfsdottir<sup>23</sup>, Emanuele Agolini<sup>24</sup>, Antonio Novelli<sup>24</sup>,  
10 Giuseppe Zampino<sup>25,26</sup>, Roberta Onesimo<sup>27</sup>, Katherine Lachlan<sup>1,28</sup>, Diana Baralle<sup>1</sup>, Heidi L. Rehm<sup>2,29</sup>,  
11 Anne O'Donnell-Luria<sup>2,29,30</sup>, Julien Courchet<sup>5</sup>, Matt Guille<sup>3</sup>, Cyril F. Bourgeois<sup>4\*</sup>, Sarah Ennis<sup>1\*</sup>

12  
13 \*Joint senior authorship

14 &Corresponding author

## 15 Author affiliations:

- 16 1. Human Genetics and Genomic Medicine, Faculty of Medicine, University of Southampton, Southampton, SO16 6YD, UK
- 17 2. Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA
- 18 3. European Xenopus Laboratory, University of Portsmouth, UK
- 19 4. Laboratoire de Biologie et Modélisation de la Cellule, Ecole Normale Supérieure de Lyon, CNRS, UMR 5239, Inserm, U1293,  
20 Université Claude Bernard Lyon 1, 46 allée d'Italie, F-69364 Lyon, France
- 21 5. Univ Lyon, Univ Lyon 1, CNRS, INSERM, Physiopathologie et Génétique du Neurone et du Muscle, UMR5261, U1315, Institut  
22 NeuroMyoGène, 69008 Lyon, France
- 23 6. INSERM U1052 - Cancer Research Center of Lyon (CRCL), 69008 Lyon France
- 24 7. University of Lyon, UMR\_S1052, CRCL 69008 Lyon France
- 25 8. IRCCS Istituto delle Scienze Neurologiche di Bologna, Programma di Neurogenetica, Bologna, Italy
- 26 9. Department of Biomedical and Neuromotor Sciences (DIBINEM), University of Bologna, Bologna, Italy
- 27 10. IRCCS Azienda Ospedaliero-Universitaria di Bologna, U.O. Genetica Medica, Bologna, Italy
- 28 11. Department of Pediatric Neurosciences, Fondazione IRCCS Istituto Neurologico Carlo Besta, Milan, Italy
- 29 12. Laboratorio di genetica Medica, Ospedale papà Giovanni XXIII, Bergamo, Italy
- 30 13. Institute for Clinical Genetics, University Hospital Carl Gustav Carus at the Technische Universität Dresden, Fetscherstrasse  
31 74, 01307 Dresden, Germany
- 32 14. Departments of Pathology and Pediatrics, The Ohio State University, Columbus, OH
- 33 15. Department of Pediatrics, The Ohio State University, Columbus, OH
- 34 16. Department of Pediatrics and Clinical Genetics, Nationwide Children's Hospital & Ohio State University, Columbus, Ohio,  
35 United States
- 36 17. deCODE Genetics/Amgen, Reykjavik, Iceland
- 37 18. Division of Neurology, Department of Pediatrics, The Children's Hospital of Philadelphia, Philadelphia, PA, USA;
- 38 19. Department of Neurology and Neuroscience, The University of Pennsylvania Perelman School of Medicine, USA
- 39 20. Unité Fonctionnelle Innovation en Diagnostic génomique des maladies rares, FHU-TRANSLAD, CHU Dijon Bourgogne, Dijon,  
40 France.
- 41 21. INSERM UMR1231 GAD, F-21000, Dijon, France
- 42 22. University Hospital of Montpellier, 191 avenue du Doyen Gaston Giraud, 34295 Montpellier, France
- 43 23. Department of Genetic and Molecular Medicine, Landspítali Hospital, Reykjavik, Iceland
- 44 24. Laboratory of Medical Genetics, Translational Cytogenomics Research Unit, Bambino Gesù Children's Hospital, IRCCS,  
45 Rome, Italy.
- 46 25. Paediatric Unit - Fondazione Policlinico Universitario A. Gemelli IRCCS, Rome, Italy
- 47 26. Medicine and Surgery School - Università Cattolica del S. Cuore, Rome, Italy
- 48 27. Rare Diseases Unit - Fondazione Policlinico Universitario A. Gemelli IRCCS – Rome
- 49 28. Wessex Clinical Genetics Service, University Hospital Southampton NHS Foundation Trust, Southampton, SO16 5YA, UK
- 50 29. Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA 02114, USA
- 51 30. Division of Genetics and Genomics, Boston Children's Hospital, Boston, MA, 02115, USA

55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

Corresponding author:

Dr Eleanor G. Seaby  
MP 808, Duthie Building  
University Hospital Southampton  
Southampton  
SO16 6YD  
UK  
E.Seaby@soton.ac.uk

Funding

EGS is supported by The Kerkut Charitable Trust, The Foulkes Foundation, and a University of Southampton Presidential Scholarship. Mouse work was performed within the framework of the LABEX CORTEX (ANR-11-LABX-0042 / ANR-11-IDEX-0007) and with support from the ERC Starting Grant (678302-NEUROMET). AODL and HLR were funded by U01HG011755. The *Xenopus* work was supported by MRC award MR/V012177/1 to MG and SE; the EXRC is supported by the Wellcome Trust (212942/Z/18/Z) and BBSRC (BB/R014841/1). DB is supported by National Institute for Health Research (NIHR) (RP-2016-07-011) research professorship.

Conflicts of Interest

Valerio Carelli acts as consultant and investigator in clinical trials for Chiesi Pharmaceuticals (Leber hereditary optic neuropathy), GenSight Biologics (Leber hereditary optic neuropathy), and Stealth BioTherapeutics (mitochondrial myopathies). Heidi Rehm received funding for rare disease research from Illumina and Microsoft.

Acknowledgements

Claudia Ciaccio and Stefano D'Arrigo are members of the ITHACA-ERN. Mouse experiments were performed with support from the Service Commun des Animaleries de Rockefeller (SCAR) from the University of Lyon.

Ethics

Genomics England (Ethics approval by the Health Research Authority [NRES Committee East of England] REC: 14/EE/1112; IRAS: 166046) approved the project RR357 (Translational genomics:

90 Optimising novel gene discovery for 100,000 rare disease patients), facilitating the identification and  
91 dissemination of a novel gene and phenotype in the index case. For all other cases included, written  
92 consent (as approved by each institute's institutional review board/ethics committee) was provided to  
93 publish any phenotype and genotype data. All *Xenopus* work was completed in accordance with the  
94 Home Office Code of Practice under PP4353452 following ethical approval from the University of  
95 Portsmouth's Animal Welfare and Ethical Review Body. Mouse breeding and handling was performed  
96 according to experimental protocols approved by the CECCAPP Ethics committee (C2EA15) of the  
97 University of Lyon, and in accordance with the French and European legislation.

98  
99 Data code and availability

100  
101 Transcriptomic data were deposited in the Gene Expression Omnibus (GEO) database under the  
102 record GSE223072. The following secure token has been created to allow review of record while it  
103 remains in private status: wrwfgmgibzedhwx. The embargo will be released upon acceptance of the  
104 manuscript. The published article includes all remaining data generated or analyzed during this study.

105  
106 **Abstract**

107  
108 Introduction

109 DDX17 is an RNA helicase shown to be involved in critical processes during the early phases of neuronal  
110 differentiation. Globally, we identified 11 patients with neurodevelopmental phenotypes with *de novo*  
111 monoallelic variants in *DDX17*. All 11 patients had a neurodevelopmental phenotype, whereby intellectual  
112 disability, delayed speech and language, and motor delay predominated.

113  
114 Materials and methods

115 We performed *in utero* cortical electroporation in the brain of developing mice, assessing axon complexity  
116 and outgrowth of electroporated neurons, comparing wild-type and *Ddx17* knockdown. We then  
117 undertook *ex vivo* cortical electroporation on neuronal progenitors to quantitatively assess axonal  
118 development at a single cell resolution. Homozygous and heterozygous *ddx17* crispant knockouts in  
119 *Xenopus tropicalis* were generated for assessment of morphology, performed behavioural assays, and

120 neuronal outgrowth measurements. We further undertook transcriptomic analysis of neuroblastoma SH-  
121 SY5Y cells, to identify differentially expressed genes in DDX17-KD cells compared to controls.

122

### 123 Results

124 Knockdown of Ddx17 in electroporated mouse neurons *in vivo* showed delayed neuronal migration as  
125 well as decreased cortical axon complexity. Mouse primary cortical neurons revealed reduced axon  
126 outgrowth upon knockdown of *Ddx17 in vitro*. The axon outgrowth phenotype was replicated in crisprant  
127 *ddx17* tadpoles, including in a heterozygous model. Crisprant tadpoles had clear functional neural defects  
128 and showed an impaired neurobehavioral phenotype. Transcriptomic analysis identified a statistically  
129 significant number of differentially expressed genes involved in neurodevelopmental processes in  
130 DDX17-KD cells compared to control cells.

131

### 132 Discussion

133 We have identified a new gene, *DDX17*, representing a rare cause of neurodevelopmental delay. We  
134 provide evidence for the role of the gene and mechanistic basis of dysfunctional neurodevelopment in  
135 both mammalian and non-mammalian species.

136

### 137 **Introduction**

138

139 RNA helicases have essential biochemical roles in all aspects of RNA metabolism, including unwinding  
140 and annealing RNA molecules and remodelling ribonucleoprotein complexes. The DEAD-box proteins are  
141 highly conserved across species, comprising the largest family of RNA helicases.<sup>1,2</sup> They share twelve  
142 conserved motifs including the signature DEAD motif comprising the amino acid sequence Asp-Glu-  
143 Ala-Asp. DDX17, also known as DEAD box protein 17, (and its close homolog DDX5), are highly energy  
144 dependent DEAD-box RNA helicases involved in diverse cellular processes, notably gene expression,  
145 biogenesis of miRNAs via their interaction with the Drosha/DGCR8 complex, and the regulation of cell  
146 fate switches and biological transitions.<sup>3,4</sup> They are coregulators of several transcription factors including  
147 MYOD, a master regulator of muscle differentiation and SMAD proteins, which mediate transforming

148 growth factor beta induced epithelial-to-mesenchymal transition.<sup>4,5</sup> Additionally, they are components of  
149 the spliceosome and regulate alternative splicing.

150  
151 *DDX17*, located on chromosome 22q13.2, has been shown to be involved in the control of Repressor  
152 Element 1-silencing transcription factor (REST) related processes that are critical during the early phases  
153 of neuronal differentiation.<sup>6</sup> Through its association with REST, *DDX17* promotes its binding to the  
154 promoter of certain REST-targeted genes and coregulates the transcriptional repression activity of REST.  
155 *DDX17* and the REST complex are downregulated during neuroblastoma cell differentiation, affecting  
156 activation of neuronal genes. Furthermore, *DDX17* and *DDX5* regulate the expression of multiple  
157 proneural microRNAs which target the REST complex during neurogenesis, implicating *DDX17* in  
158 neuronal gene repression.<sup>6</sup> In 2022, Suthapot *et al.*<sup>7</sup> focused on characterising chromatin occupancy of  
159 *DDX17* and *DDX5* in hPSCs NTERA2 and their neuronal derivatives. They showed that the expression of  
160 both helicases is abundant throughout neural differentiation of the hPSCs NTERA2, preferentially  
161 localised within the nucleus and that they occupy chromatin genome-wide at regions associated with  
162 genes related to neurogenesis. Both *DDX17* and *DDX5* are mutually required for controlling  
163 transcriptional expression of these neurogenesis-associated genes but are not important for maintenance  
164 of the stem cell state of hPSCs. In contrast, they are critical for early neural differentiation of hPSCs,  
165 possibly due to their role in the upregulation of key neurogenic transcription factors such as *SOX1*,  
166 *SOX21*, *SOX2*, *ASCL1*, *NEUROG2* and *PAX6*. Critically, *DDX17* and *DDX5* are important for  
167 differentiation of hPSCs towards *NESTIN* and *TUBB3* positive cells, which represent neural progenitors  
168 and mature neurons, respectively. However, those studies used a *DDX17* and *DDX5* co-depletion  
169 approach to address the function of these factors in neurogenesis, and information regarding the specific  
170 contribution of each helicase to this process is lacking.

171  
172 To date, *DDX17* has no disease-gene relationship. The gene is highly constrained for loss-of-function  
173 (LoF), i.e. fewer LoF variants in *DDX17* are observed in population datasets than would be expected  
174 under a null mutational hypothesis (37.7 expected, 1 observed, pLI=1.0 in gnomAD). Genes can be  
175 quantified by constraint to LoF using the Loss-of-function Observed/Expected Upper-bound Fraction  
176 (LOEUF) score, which places genes along a continuous spectrum of intolerance to haploinsufficiency.<sup>8</sup>

177 Genes highly constrained for LoF, represented by low LOEUF scores, are highly associated with known  
178 haploinsufficient disease genes.<sup>8,9</sup> However, the majority of genes in the lowest LOEUF decile are not yet  
179 associated with a disease phenotype but may be expected to cause disease if mutated through LoF.<sup>10</sup>  
180 *DDX17* has a LOEUF score of 0.13, suggesting that haploinsufficiency of the gene is not tolerated.  
181 Considering its role in neuronal differentiation, muscle differentiation and alternative splicing, one might  
182 expect that *DDX17* is an essential gene in neurodevelopment and may present such a phenotype in  
183 humans. Therefore, identification of patients with LoF variants in *DDX17* may help characterize a new  
184 gene-disease relationship.

185

## 186 **Methods**

187

### 188 Accessing patient data

189

190 We obtained access to the 100,000 Genomes Project data through membership of a Genomics England  
191 Clinical Interpretation Partnership, with approved project RR359: Translational genomics: Optimising  
192 novel gene discovery for 100,000 rare disease patients. Deidentified whole genome sequencing and  
193 phenotype data (stored as human phenotype ontology terms) were accessible in the Genomics England  
194 Research Environment.

195

196 Additional study subjects were identified through Matchmaker Exchange after deposition in  
197 GeneMatcher.<sup>11,12</sup> In total, 11 patients consented to participate in our study. Parents and legal guardians  
198 of all affected individuals provided written consent for the publication of their results alongside genetic and  
199 clinical information. Guardians of patients 6, 7 and 10 explicitly consented to have photographs published.

200

### 201 Sequencing and data analysis

202 All patients, except for patient 3, had trio exome sequencing performed. Data processing and variant  
203 filtering and prioritization were carried out by in house pipelines at respective host centres. Patient 3 had  
204 trio whole genome sequencing undertaken as part of the 100,000 Genomes Project<sup>13</sup> and their data was  
205 filtered using the DeNovoLOEUF filtering strategy.<sup>9</sup> DeNovoLOEUF is a tool that can be applied at scale



206 to genomics datasets, extracting rare *de novo* predicted loss-of-function variants in LOEUF-constrained  
207 genes.<sup>9</sup>

208

209 *Xenopus tropicalis*

210

211 Adult Nigerian strain *Xenopus tropicalis* were housed within the European *Xenopus* Resource Centre  
212 (EXRC; <https://xenopusresource.org>), University of Portsmouth, in recirculating MBK Ltd systems  
213 maintained at 24°C - 27°C (13-11-hour light-dark cycle) with 10% daily water changes. All *Xenopus* work  
214 was completed in accordance with the Animals (Scientific Procedures) Act 1986 under licence  
215 PP4353452 following ethical approval from the University of Portsmouth's Animal Welfare and Ethical  
216 Review Body. Detailed methods on: the generation of *X. tropicalis* founder animals and *ddx17* crispants;  
217 wholemount *in situ* hybridisation; phenotypic analysis; experimental design and statistical analysis are  
218 available in **Supplementary methods**.

219

220 RNA-seq

221 Detailed methods are available in **Supplementary methods**. In summary, human SH-SY5Y  
222 neuroblastoma cells (ECACC #94030304) were grown and transfected with 60 nM of a mixture of 2  
223 different siRNA against *DDX17* (Merck-Millipore, see sequences in **Table S1**). Protein extraction was  
224 carried out as previously described<sup>14</sup> and total RNA were isolated. Directional RNA libraries were  
225 prepared from total RNA after removal of ribosomal RNA (lncRNA library, Novogene). High throughput  
226 sequencing of 150 bp paired-end reads was carried out on an Illumina Novaseq 6000 platform  
227 (Novogene), generating an average number of 75 million matched pairs of reads per sample. Raw reads  
228 were pre-processed and mapped reads were filtered using SAMtools<sup>15</sup> and the number of reads per gene  
229 was counted using HTSeq.<sup>16</sup> Differential gene analysis was carried out with the DESeq2<sup>17</sup> package.  
230 Parameters for differential expression: P-value < 0.05, [log<sub>2</sub>(FC)] ≥ 0.50 and base mean ≥ 10. Gene  
231 ontology and gene-set enrichment analyses were carried out using the ShinyGO 0.76.3 web interface.<sup>18</sup>

232

233

234 Mouse experiments  
235

236 Mouse breeding and handling was performed according to experimental protocols approved by the  
237 CECCAPP Ethics committee (C2EA15) of the University of Lyon, and in accordance with French and  
238 European legislation. Detailed methods on *ex vivo* cortical electroporation and primary neuronal cultures;  
239 immunostaining; *in utero* cortical electroporation; immunohistochemistry; image acquisition; and  
240 quantifications and statistical analyses are available in **Supplementary methods**.

241  
242 **Results**

243 We applied the DeNovoLOEUF filtering strategy, as previously described<sup>9</sup>, to 13,494 parent/offspring trios  
244 in the 100,00 Genomes Project, focusing on genes with a LOEUF score <0.2 with no prior disease gene  
245 association. We identified one individual harboring a heterozygous pLoF variant in *DDX17*. Using the  
246 GeneMatcher platform, we identified a further 10 patients all with *de novo* variants in *DDX17* presenting  
247 with neurodevelopmental phenotypes. Representatives for these participants were then invited to join our  
248 research study and referring clinicians were asked to complete a standardized phenotype table  
249 (**Supplementary File B**). The summary of the phenotypic features of the 11 patients (from 11  
250 independent families) harbouring *de novo* heterozygous variants in *DDX17* are provided in **Table 2**. All  
251 variants were absent from gnomAD<sup>8</sup> v2.1.1 and v.3.1.2.  
252

TABLE 2 | Core phenotypic features of cohort with heterozygous *de novo* variants in *DDX17*

Patient	Predicted Loss of Function					Missense					
	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11
<b>Age (at last visit)</b>	0-5y	11-15y	11-15y	16-20y	16-20y	0-5y	6-10y	11-15y	16-20y	16-20y	21-25y
<b>Sex</b>	F	M	M	M	M	M	M	F	M	F	M
<b>Variant coordinate (GRCh38)</b>	22:38499457_38499458del	22:38494928C>CAT	22:38495939T>C	22:38494973TC>T	22:38493717del	22:38493736T>C	22:38495016C>T	22:38498504G>A	22:38495900T>C	22:38498463T>C	22:38506032G>T
<b>Variant consequence</b>	Arg161Glyfs*7	c.997_998dupAT	c.739-2A>G	Arg318Hifs*36	Asn462Metfs*16	Gln454Arg	Arg304His	Pro203Leu	Gln259Arg	Thr217Ala	Ala69Asp
<b>Macrocephaly</b>	N	N	N	Y	At birth	N	N	N	Y	N	N
<b>Dysmorphic facial features</b>	Y	Y	N	Y	Y	Y	N	Y	N	Y	N
<b>Delayed walking</b>	Y	Y	Y	N	N	Y	Y	Y	N	Y	unknown

<b>independently</b>											
<b>Intellectual disability</b>	mild-moderate	N	mild-moderate	mild	mild	unknown	N	moderate	mild	N	moderate
<b>ADHD</b>	N	N	Y	Y	Y	unknown	N	N	Y	unknown	Y
<b>Autism spectrum disorder</b>	N	Y	Y	N	N	N	N	N	Y	Y	N
<b>Delayed speech and language development</b>	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y
<b>Global developmental delay</b>	Y	N	N	Y	Y	Y	Y	Y	Y	Y	Y
<b>Neurodevelopmental delay</b>	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y
<b>Gross</b>	Y	Y	N	N	N	Y	Y	Y	Y	Y	N

<b>motor delay</b>	[Blue shaded]					[Blue shaded]					
<b>Fine motor delay</b>	Y	N	Y	N	Y	Y	Y	Y	Y	Y	Y
<b>Stereotypy</b>	N	Y	N	N	N	Y	Y	N	N	Y	N
<b>Generalised hypotonia</b>	Y	Y	N	N	N	Y	N	Y	N	Y	N

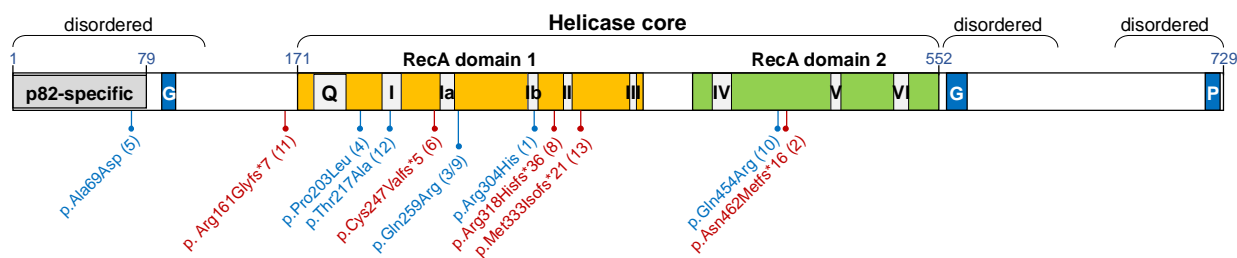
ADHD - attention deficit hyperactivity disorder; F – female; M – male; m – months; N – no; Y – yes; y – years. Blue cells denote presence of a feature. Grey cells represent unavailable data.

The cohort comprises 8 males and 3 females, all of whom are alive and have a median age of 15y at the latest available follow up. The median age of walking was 18 months. Intellectual disability, ranging from moderate to mild (IQ 56-83), is prevalent in 7/10 (70%) of patients. Seven of eleven (64%) of the cohort have dysmorphic facial features. Overlapping facial dysmorphology between patients includes: telecanthus; synophrys; upslanting palpebral fissures; depressed nasal bridge; posteriorly rotated ears; high arched eyebrows; epicanthus; telecanthus; frontal bossing; micrognathia and strabismus. Three patients have 5<sup>th</sup> finger clinodactyly, two have clubfeet, and three have 2,3 toe syndactyly (some of these features are shown in **Figure 1A**). Fifty-six percent (5/9) have attention deficit hyperactivity disorder (ADHD) and 4/11 (36%) have features of autism. Ninety-one percent (10/11) have delayed speech and language development, 9/11 (82%) have global developmental delay, and 11/11 (100%) have neurodevelopmental delay. Gross motor delay is prevalent in 8/11 (73%), and 9/11 (82%) have fine motor delay. Thirty-six percent (4/11) have stereotypy and 5/11 (45%) have generalized hypotonia. Patient 9 (height 178cm (Z=0.40); weight, 55.7kg (Z= -0.89)) and patient 4 (height 165cm (7th percentile); weight, 51.5kg (4<sup>th</sup> percentile)) have signs of macrocephaly with Z-scores of 3.97 and 3.23, respectively. Patient 5 had signs of macrocephaly at birth, which normalized through infancy. Eight participants had brain MRI scans of which 4 patients showed abnormalities including: left lateral compartment greater than right; monolateral temporal cortical dysplasia; asymmetry of the cerebral cortex and right sided nonspecific demyelination; generalized brain demyelination, and periventricular white matter hyperintensities. No other obvious asymmetry was observed.

### Molecular genetic findings

Following variant filtering and prioritization of exome and genome data, no likely pathogenic or pathogenic variants as curated using American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG-AMP) guidelines<sup>19</sup> were identified that fully explained the patients' phenotypes. Eleven participating research programs identified *de novo* variants of uncertain significance in *DDX17* that were of sufficient interest to submit to Matchmaker Exchange. Five variants were pLoF and 6 variants were missense (**Figure 1B**).

FIGURE 1 | **DDX17 patient variants**



Gene ideogram, whereby variants in blue are missense, and variants in red are loss of function.

Four patients in the cohort had additional variants of uncertain significance reported. Patient 7 had compound heterozygous pathogenic variants in *ACADM* associated with medium chain fatty acid dehydrogenase deficiency; NM\_000016.5(*ACADM*):c.799G>A,p.(Gly267Arg) and NM\_000016.5(*ACADM*):c.985A>G,p.(Lys329Glu). Patient 3 had a *de novo* Yq11.21-qter deletion and *de novo* Ypter-p11.3 duplication for which the significance is unknown. Patient 6 had a VUS in *HCFC1* associated with methylmalonic aciduria and homocysteinemia; NM\_005334.3(*HCFC1*):c.4418C>T.p.(Thr1407Met) and patient 10 harboured a maternally inherited 16q23.3 (81,477,800 – 81,552,781) VUS.

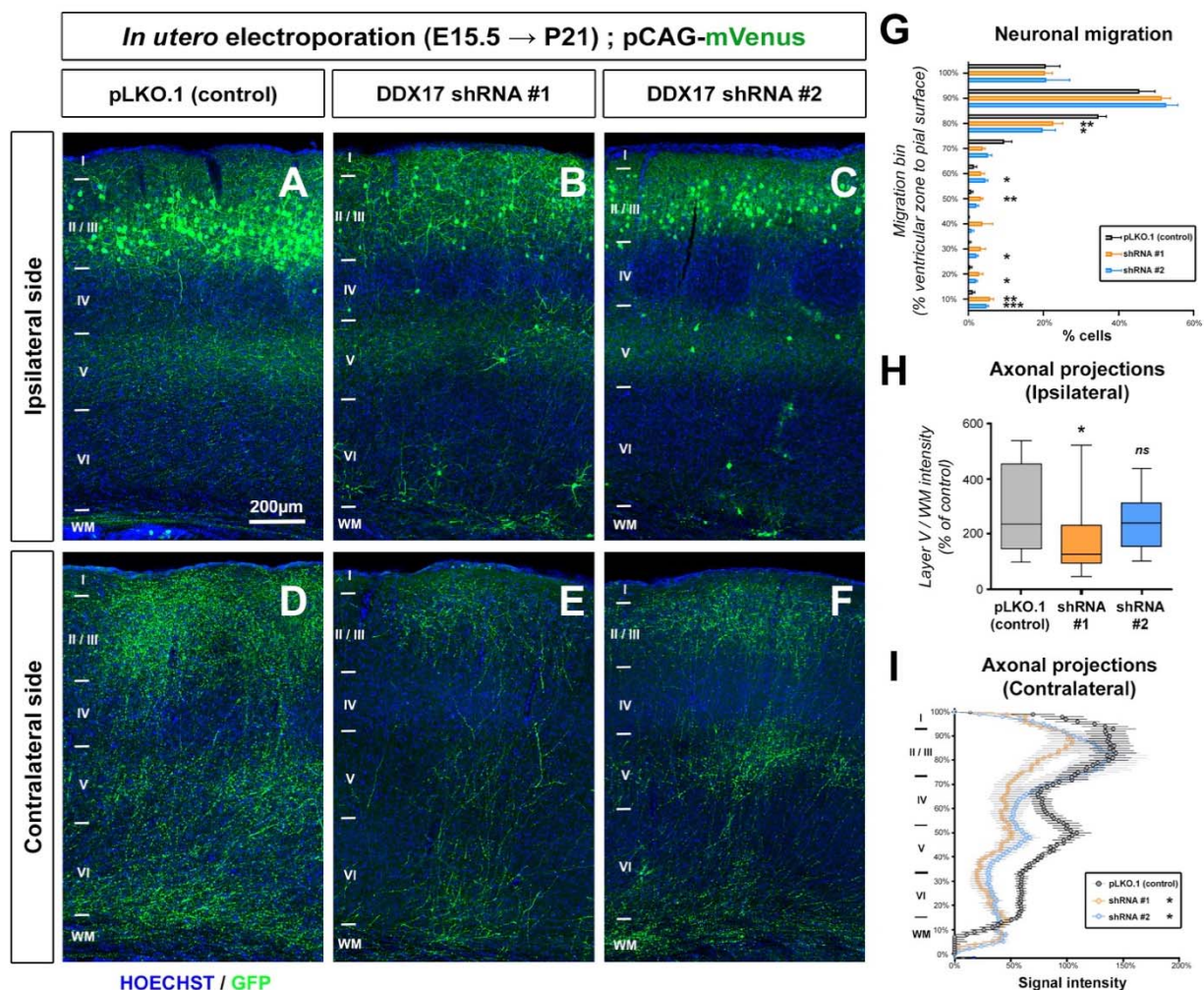
### DDX17 supports cortical neuron development in the mouse

The identification of several variants in *DDX17* associated with neurodevelopmental features prompted assessment of *DDX17* reduction on cortical development in animal models. We first turned to *in utero* cortical electroporation (IUCE) in the mouse using two distinct shRNA plasmids targeting *Ddx17* (shDDX17 #1 or shDDX17 #2). Electroporations were performed at embryonic day (E)15.5, the developmental stage at which progenitors give rise to callosal-projecting pyramidal neurons. By P21, in control conditions, all electroporated neurons (visualized by mVenus fluorescence) reached the superficial layers of the cortex (layers II/III) (**Fig 2A**). In contrast, we could observe defects in neuronal migration upon knockdown of *Ddx17* (**Fig 2B-C**) and after quantification, a statistically significant fraction

of neurons did not reach the most superficial cortical layers in conditions electroporated with shRNA plasmids (**Fig 2G**). Despite this, neuronal polarization and axon formation was not impaired. In control conditions, axons of layer II/III neurons progress through the corpus callosum to reach the contralateral hemisphere, and branch extensively on ipsilateral layer V (**Fig 2A**), as well as contralateral layers II/III (**Fig 2D**). In shRNA-electroporated animals, we observed a trend toward a decrease of axon density in the ipsilateral side (**Fig 2B-C**, quantified in **Fig 2H**), and especially a strong reduction of contralateral axon density (**Fig 2E-F**, quantified in **Fig 2I**). There was no difference in axon density in the white matter (WM), indicating that the same proportion of axons reached the contralateral hemisphere regardless of *Ddx17* expression. Our results demonstrate that DDX17 is required for cortical development in the developing mouse brain.



FIGURE 2 | Knockdown of Ddx17 decreases cortical axon complexity in the mouse *in vivo*

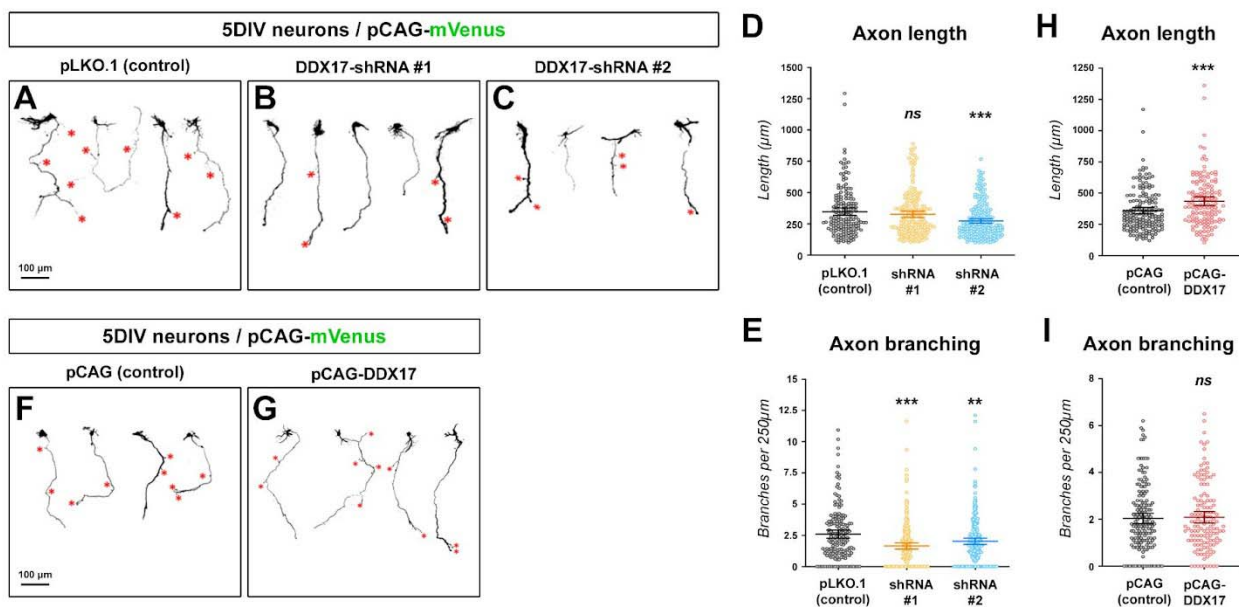


(A-F) Histochemistry of the ipsilateral or contralateral side of mice at P21 following *in utero* electroporation with pLKO (A,D) or after loss of function of DDX17 (B-C and E-F) and the fluorescent protein mVenus. (G) Quantification of neuronal migration defects upon knockdown of DDX17. Some position was quantified on a ventricular zone to pial surface axis. Each bin represents 10% migration. Data: average + SEM, N = 6 sections out of 3 animals (2 sections per animal). Analysis: Two-way ANOVA with multiple comparisons. \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ . (H) Quantification of normalized mVenus fluorescence in layer V of the ipsilateral cortex (min, max, median, 25<sup>th</sup>, and 75<sup>th</sup> percentile). NpLKO=13, NshDDX17-1=18, NshDDX17-2=25. Analysis: One-way ANOVA with Dunn's multiple comparisons. ns: $p > 0.05$ , \* $p < 0.05$ . (I) Quantification of normalized mVenus fluorescence along a radial axis in the contralateral cortex (Average  $\pm$  SEM) (H) in control condition pLKO) or after knockdown of DDX17. NpLKO=13, NshDDX17-1=18, NshDDX17-2=25. Analysis: Two-way ANOVA. \* $p < 0.05$ .

Because the wiring of the brain results from sequential biological processes, defects in the early steps (e.g. neurogenesis or neuronal migration) can lead to alterations in the later biological processes such as

axon development. To make sure that the axonal phenotypes observed *in vivo* do not result from abnormal neuronal migration, we subsequently turned to *in vitro* neuronal cultures, which allow quantitative assessment of axonal development at a single cell resolution. We performed *ex vivo* cortical electroporation (EVCE) at E15.5 to target neuronal progenitors in the dorsal telencephalon and cultured neurons for 5 days *in vitro* (5 DIV). In both shRNA conditions, we observed an effect on axonal development compared to the control condition (pLKO.1) (**Fig 3A-C**). The inhibition of DDX17 expression this decreased axon length and reduced collateral branch formation (**Fig 3D-E**). Two independent shRNA plasmids produced markedly similar phenotypes, indicating that this phenotype is unlikely to be an off-target effect of the shRNAs. To confirm this result, we tested the consequence of overexpressing human DDX17 by EVCE. Following electroporation, we observed an increased axon length compared to the control condition, mirroring the effect of DDX17 knockdown (**Fig 3F-G**, quantified in **Fig 3H**). Interestingly, although we counted more collaterals per neuron, this increase was due to the increase in axon length and axon branching did not differ from the control condition when normalized per axon length. Overall, our results demonstrate that DDX17 is important for axon development in mouse cortical neurons.

FIGURE 3 | **DDX17 is necessary and sufficient for axon development**



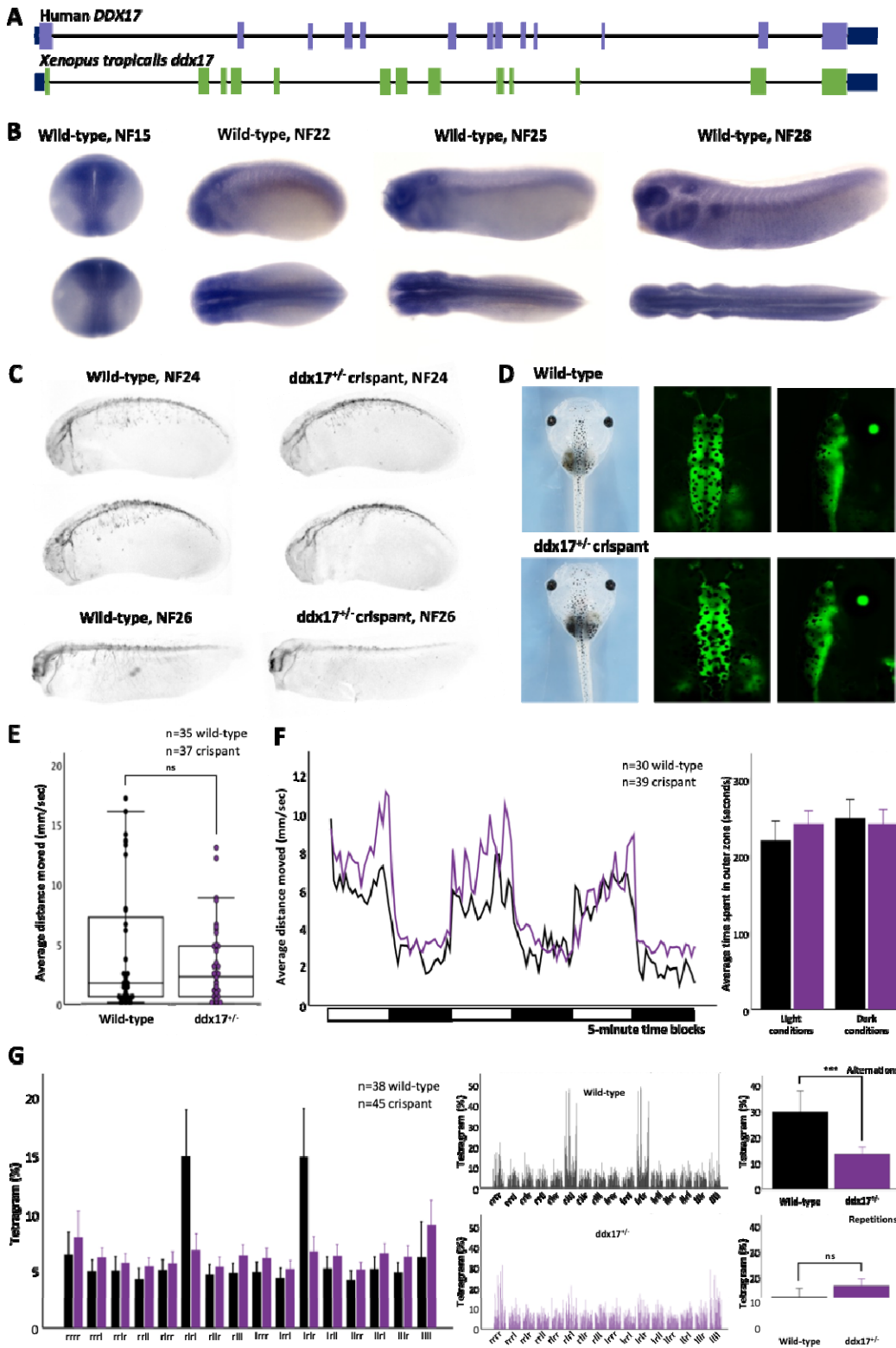
(A-C) Representative images of mVenus expressing cortical neurons (5 DIV) in control condition

(pLKO.1) or after loss of function (DDX17-shRNA #1 and DDX17-shRNA #2). Red star (\*) point to collateral branches of the axon. **(D-E)** Quantification of axon length and number of collateral branches of 5 DIV neurons in the indicated conditions. Bars represent the average and 95% CI. Statistical tests: Kruskal-Wallis test with Dunn's post-test (each condition compared to control condition). **(F-G)** Representative images of mVenus expressing cortical neurons (5 DIV) in control conditions, or upon overexpression of DDX17. Red star (\*) points to branch/collateral position. **(H-I)** Quantification of axon length and number of collateral branches of 5 DIV neurons in the indicated conditions. Bars represent the average and 95% CI. Statistical tests: Kruskal-Wallis test with Dunn's post-test. (D-E)  $N_{(pLKO.1)}=169$  ,  $N_{(shDDX17 \#1)}=205$  ,  $N_{(shDDX17 \#2)}=228$ . (H-I)  $N_{(pCAG)}=168$  ,  $N_{(pCAG-DDX17)}=134$ . ns:  $p>0.05$ , \*\*:  $p<0.01$ , \*\*\*:  $p<0.001$

#### Xenopus ddx17 crispants have reduced axon outgrowth and working memory

To test the effect of loss-of-function *ddx17* variants in an intact animal, crispant *X. tropicalis* models were used; the exon structure of the human and *Xenopus* genes are similar (**Fig 4A**) and the proteins produced have 68% amino acid identity (**Fig S2A**). The expression pattern of *ddx17* mRNA has not previously been reported in *Xenopus* and *in situ* hybridisation shows it to be expressed most highly in neural tissues including the migratory neural crest, brain, eye, and otic vesicle (see the purple staining in **Fig 4B**).

**FIGURE 4 | Heterozygous *ddx17* *X. tropicalis* crispants appear morphologically normal but show reduced axon outgrowth and have a working memory deficit**



**A.** *Xenopus tropicalis* and *Homo sapiens* have the same exon-intron structure. **B.** A developmental series of wild-type *X. tropicalis* were fixed and underwent in situ hybridisation with a probe specific for *ddx17*, the

blue stain shows where this gene is expressed. The highest levels of *ddx17* mRNA are in neural tissues although it is detectable more widely. **C.** Control and heterozygous crispant embryos were fixed at the stages shown and stained for neuron bodies and axons using HNK1 monoclonal antibody. The extension of axons ventrally from the neural tube is reduced in crispants at stage 24 (4/4 embryos) although growth does continue (see stage 26), scoring was blind and prior to genotyping. **D.** Brightfield microscopy showed no clear distinctions between control and crispant tadpoles across a range of stages and, when the neural tissue was labeled transgenically this too failed to reveal any gross-morphological distinctions. **E.** Tadpoles at stage NF42, similar to those shown in D, underwent automated movement analysis in a Zantiks MWP unit. In all cases the analysis was performed blind (with genotyping subsequent to measurements) the black data points represent wild-type animals with purple showing crispant data. **F.** The control and crispant animals move similar distances over time but crispants show a greater initial startle response than controls. Interestingly all tadpoles spend the majority of each 300 second block in the most proximal arena, closest to the edge of the dish (F). **G.** The main change caused by heterozygous *ddx17* LOF becomes clear when working memory is tested in the free movement pattern Y-maze; the crispants have lost the alternating search pattern shown by all vertebrates.

First, a mosaic homozygous crispant knockout of *ddx17* was generated using two non-overlapping CRISPR/Cas9 gene-editing complexes designed to disrupt exon 7 (**Fig S2B-C**) and these were mated to produce F1 heterozygous and compound heterozygous tadpoles (Fig S1A). The introduction of indels into the *ddx17* locus was tested by Sanger sequencing of the target region and demonstrated strong penetrance of indels in the founder animals (**Fig S1B**), producing a *ddx17* mosaic knockout (*Xtr.ddx17<sup>em1EXRC</sup>*). The phenotypes produced by each sgRNA were indistinguishable, showing they were not due to off target effects. The heterozygous offspring produced by outcrossing mosaic founders were first identified using the T7E1 mismatch detection assay (**Fig S1C**), with the genotype confirmed by Sanger sequencing (**Fig S1D**), revealing frequently occurring deletions of 5bp, 22bp, 131bp and 136bp (**Fig S1E**), resulting in a frameshift leading to protein truncation within exon 7 (*Xtr.ddx17<sup>em2EXRC</sup>*). Inbreeding of the mosaic founders produced homozygous or compound heterozygous tadpoles as shown by Sanger sequencing (Fig S1F) (*Xtr.ddx17<sup>em3EXRC</sup>*). The homozygous tadpoles bearing larger deletions (131bp or 136bp) in one allele (**Fig S1G**), were found at later stages to have a series of further, progressive deletions as demonstrated by polyacrylamide gel electrophoresis (**Fig S1H**).

Phenotypically, more than half of the founder crispant embryos showed evidence of gastrulation defects (**Fig S3A**). The remaining animals showed no gross morphological defects (**Fig S3B,C**). The gross craniofacial morphology of *ddx17* crispants was tested by injecting CRISPR/Cas9 complexes into one-cell of a dividing two-cell embryo. This results in the effects of the protein truncation being concentrated on one side of the embryo along the left-right axis, with the other side acting as an internal control; no altered morphology was observed in the crispants (**Fig S3C**). A decrease in head size was observed prior to free-feeding stages (average 5.93 mm<sup>2</sup> for control tadpoles and 4.89 mm<sup>2</sup> for crispant tadpoles (NF41, n = 16),  $t(30)=3.85$ ,  $p<0.001$ ). This significant decrease in head size was no longer apparent at later stages of development (average 15.9 mm<sup>2</sup> for control tadpoles and 16.5 mm<sup>2</sup> for crispant tadpoles (NF48, n = 16),  $t(30)=-5.05$ ,  $p=0.617$ , **Fig S3D**).

The phenotype of patients with *ddx17* variants include neurodevelopmental deficits and these can now be modeled in *X. tropicalis*. Here, in addition to the Free-movement Pattern (FMP) Y-maze that is already validated in *X. tropicalis*,<sup>20</sup> another assay has been adapted to assess working memory in *Xenopus*. In this assay, animals are exposed to 5-minute alternating periods of light and dark; their initial response is to move (startle) as the lights go off, but they quickly learn that the dark period does not signal danger and reduce their response to it. This attenuated response to the changing environmental conditions demonstrates a second measure of short-term working memory in the tadpoles. Importantly this behavior can be probed by pharmacological agents which disrupt working memory and attenuate anxiolytic responses. Following treatment with the NMDA receptor antagonist MK-801, tadpoles can be observed to startle repeatedly in response to the light-dark transition period. The startle response is also abolished following administration of 5  $\mu$ M diazepam (**Fig S4A**). Once established using pharmacological agents, these assays were applied to test the neurodevelopmental phenotypes of the *ddx17* crispants.

Comparing the movement of control tadpoles and F<sub>0</sub> *ddx17* crispant animals at NF48 showed that crispant animals move less over the 10-minute trial period (average 4.0 mm/sec for control tadpoles and 2.7 mm/sec for crispant tadpoles (n = 48,  $t(82)=-2.6$ ,  $p=0.012$ ), **Fig S3F**). A significant reduction in locomotion was additionally observed in a second assay investigating the tadpoles' response to 5-minute light-dark transition periods (average 3.9 mm/sec for control tadpoles and 2.3 mm/sec for crispant

tadpoles ( $n = 24$ ,  $t(46)=3.4$ ,  $p=0.001$ ), **Fig S3G**). Unusually, these mosaic crispant tadpoles were not observed to startle at these light-dark transitions. In the FMP Y-maze assay that tests working memory, founder crispant tadpoles performed fewer alternations in their search patterns than controls (average alternations 22.8% for control tadpoles and 19.5% for crispant tadpoles ( $n = 20$ ,  $F(1, 33) = 0.366$ ,  $p = 0.550$ )) and more repetitions (average repetitions 9% for control tadpoles and 12.8% for crispant tadpoles ( $n = 20$ ,  $F(1, 33) = 3.272$ ,  $p = 0.080$ )) (**Fig S3H**).

Since there was clear evidence of movement and neural defects in the  $F_0$  crispants, together with reduced axon length observed in electroporated mouse brains deficient in DDX17, neuron outgrowth was examined in *Xenopus* using an anti-HNK-1 antibody. Embryos injected with control (*tyr*)<sup>21</sup> and *ddx17* gene-editing complexes restricted to one side of the embryo reveal reduced axon outgrowth in *ddx17* crispants on the ipsilateral side (compare the injected and uninjected lateral views in **Fig S4E**). Similarly, in the heterozygous *ddx17* model axon outgrowth is visibly reduced compared with controls at NF stage 24 with evidence of continued axon outgrowth at later stages (NF26 onwards (**Fig 4C & Fig S4B**)).

In comparison to the founder animals,  $F_1$  crispant tadpoles bearing non-mosaic heterozygous or compound heterozygous/homozygous indels in *ddx17* show no gross morphological or developmental abnormalities including in gastrulation (**Figs 4D, S4C & S5A**), hatching rate and head size (**Fig S5B**). Gross structural differences between control and crispants were not seen in the forebrain, midbrain or hindbrain regions of heterozygous *ddx17* crispant animals, even when bred in a [Xtr.Tg(tubb2b:GFP)Amaya] RRID: EXRC\_3001 background (**Fig S4C**) to make differences more obvious. Post-hatching, all animals moved away from tactile stimuli (to the head) and at later stages adopted a normal, head down filter-feeding posture with the ability to navigate freely within their respective environments. Further, no abnormalities in locomotive behavior consistent with descriptors of seizure activity in tadpoles were observed.<sup>22,23</sup> The locomotive activity of wild-type tadpoles, heterozygous or homozygous, non-mosaic crispants in *ddx17* was indistinguishable at NF48 when tracked across a 10-minute trial period (**Fig 4E & S5C**). Similarly, comparative average locomotive activity in the light-dark assay was not significantly different from controls in either non-mosaic crispant group, with all tadpoles noted to spend the majority of each of the 5-minute blocks at the edge of the dish.



Unlike wild-type tadpoles however, both non-mosaic crispant tadpole models were noted to respond with increased locomotion to multiple dark-light transition periods, although startling was more pronounced in the homozygous/compound heterozygous model (**Fig 4F & S5D**). Since this indicated reduced memory, we used a well characterized working memory assay, the 1 hour free-movement pattern Y-maze to compare wild-type and homozygous or heterozygous *ddx17* crispant tadpoles. Although not significant, homozygous *ddx17* tadpoles performed fewer alternations (average alternations 22.6% for control tadpoles and 15.2% for homozygous *ddx17* tadpoles ( $n = 49$ ,  $F(1, 95) = 3.761$ ,  $p = 0.055$ )) but significantly more repetitions ( $n = 49$ ,  $F(1, 95) = 10.983$ ,  $p = 0.001$ ) than controls (**Fig S5E**). This increase in repetitions taken together with the unattenuated startle response suggests increased levels of anxiety in the homozygous *ddx17* tadpole group.<sup>24</sup> In comparison, the heterozygous *ddx17* model demonstrated significantly fewer alternations ( $F(1, 80) = 14.25$ ,  $p < 0.001$ ) without a significant increase in repetitions ( $F(1, 80) = 3.29$ ,  $p = 0.074$ ) when compared to wild-type tadpoles (**Fig 4G**). When considered alongside the response observed in the light-dark assay, this shows a clear deficit in the short-term working memory.

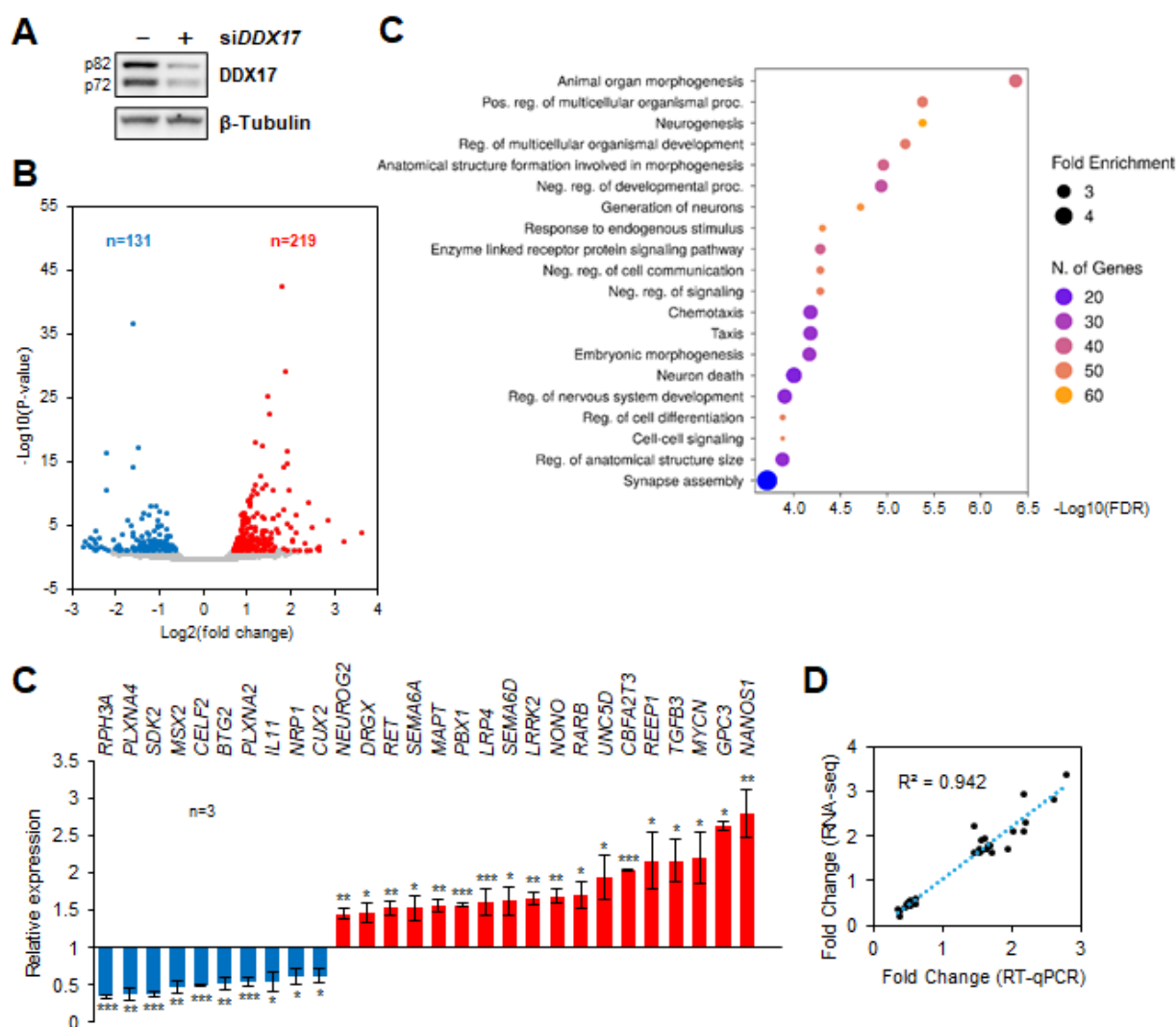
Overall, the crispant *ddx17* models show very similar phenotypes, with evidence of reduced axon outgrowth and working memory consistent between them. The mosaic founder animals however move less than the wild-type tadpoles unlike the other crispant models and have a transiently reduced head size.

### RNA-seq analysis

Finally, to gain insight into the possible functions and target genes of DDX17 in a human cellular context, we carried out a transcriptomic analysis of neuroblastoma SH-SY5Y cells in which we knocked-down the expression of the *DDX17* gene, using a mixture of 2 different siRNAs (**Fig 5A**). We identified 350 genes that were differentially expressed in *DDX17*-KD cells compared to control cells (**Fig 5B** and **Supplementary File C**). The functions of this set of genes were significantly associated to developmental processes, in particular the development and functions of the nervous system (**Fig 5C, S6**, and **Supplementary File A**). For instance, the expression of several development-associated transcription

factors (*MSX2*, *TBX3*, *GATA3*, *FOSL1*, *NEUROG2*, *SMAD6* and *SMAD9*, *SOX13*, *DRGX*, *RARB*, *MYCN*...) was deregulated upon *DDX17* KD. Of note, looking at the molecular functions associated with those genes also revealed the presence of a significant number of *trans*-membrane receptors (31 genes) and receptor ligands (15 genes) (**Supplementary File C**), including several receptors/ligands associated with axon guidance (*DCC*, *EFNB2*, *PLXNA2* and *PLXNA4*, *SEMA6A* and *SEMA6D*, *RET*, *ROBO2*, *UNC5D*...).

FIGURE 5 | *DDX17* controls the expression of genes involved in nervous system development



**A.** Western-blot showing the siRNA-mediated depletion of *DDX17* protein in SH-SY5Y cells. **B.** Volcano plot showing the genes that are impacted by *DDX17* KD in SH-SY5Y cells, as predicted from the RNA-

*seq analysis. Significantly altered genes (downregulated in blue and upregulated in red) were identified as described in the Methods section.*

**C.** *Gene ontology analysis using ShinyGO for the genes impacted by DDX17 KD. Only the top 20 of the GO enriched biological processes are shown (see Supplementary File A for the full list of enriched terms).*

**D.** *Validation of the effect of DDX17 knockdown on the steady-state expression of a selection of genes. RT-qPCR data were first normalized to GAPDH mRNA level in each condition, and the normalized mRNA level of each gene in the DDX17 knockdown condition was then normalized to the control condition, set to 1. Data are expressed as the mean value  $\pm$  S.E.M. of independent experiments ( $n = 3$ ). Unpaired Student's *t*-test (\**P*-val < 0.05; \*\**P*-val < 0.01; \*\*\**P*-val < 0.001).*

**E.** *Correlation between the measured fold change of expression (x-axis) and the corresponding predicted fold change value (y-axis) for the 28 genes shown in panel E.*

We then analyzed separately the subsets of genes of which steady-state expression was negatively (131 genes) or positively (219 genes) altered by *DDX17* KD (**Supplementary files A & C**). This analysis showed again a significant occurrence of GO terms associated with neurogenesis in both groups of genes, but it also underlined a link between downregulated genes and body morphogenesis, while the group of upregulated genes was associated more specifically to cell-signaling pathways (**Fig S7**). To validate our computational analysis, we selected 28 genes from the two subgroups of mis-regulated genes and measured their mRNA level by RT-qPCR assays in mock-depleted and *DDX17*-depleted SH-SY5Y cells. Consistently, the expression of each tested gene was altered as predicted from the RNA-seq, with a strong combined correlation score ( $R^2=0.942$ ) (**Figs 5D and 5E**).

Collectively, our results indicate that the *DDX17* gene is involved in several processes during the development of vertebrates, in particular in the development of the nervous system, most likely because they control the expression of subsets of genes with key functions in the neural system.

## **Discussion**

To the best of our knowledge, this is the first study to describe *de novo* heterozygous *DDX17* variants associated with features describing a novel neurodevelopmental disorder. Using mouse and frog animal models, we provide strong evidence that *DDX17* plays important roles in the developing nervous system,

in mouse and frog models. More specifically, *Ddx17* knockdown impaired neuronal migration and axon development in the brain of newborn mice and reduced axon outgrowth and branching in primary cortical neurons *in vitro*. In agreement with these results, crispant tadpole *ddx17* models, including a heterozygous F<sub>1</sub> model, also presented a reduced axon outgrowth phenotype. Crispant tadpoles also had clear functional neural defects. Since the region and developmental state of the central nervous system in the mice and tadpoles in which this effect has been noted are distinct, it suggests that *DDX17* has an important role in widely distributed neurodevelopmental processes. The conservation of function across evolutionarily distant species such as mice and frogs strongly support that the role of *DDX17* is conserved in humans too. These *in vivo* results are therefore consistent with the hypothesis that heterozygous loss-of-function *DDX17* variants identified in patients induce a significant alteration of the function of the protein during neuronal development, resulting in the observed phenotype in our cohort.

The phenotype associated with *de novo* variants in *DDX17* is consistent with a neurodevelopmental disorder, typified by mild-moderate intellectual disability, delayed speech and language development and global developmental delay. Sixty-four percent (7/11) of the cohort have dysmorphic facial features, although for many this is subtle. Overlapping dysmorphology between patients includes: synophrys; upslanting palpebral fissures; depressed nasal bridge; posteriorly rotated ears; high arched eyebrows; epicanthus; telecanthus; and strabismus. Some patients have gross and fine motor delay, generalized hypotonia, stereotypy, and evidence of autism spectrum disorder.

There were no substantial differences in phenotype severity between patients harboring missense variants versus loss-of-function variants, suggesting *DDX17* haploinsufficiency causes the observed phenotype. Since all missense variants but one fall within the helicase domain (**Fig 1**), it suggests that they impact the structure and/or activity of this domain in a way that deeply alters the overall function of *DDX17*, similarly to loss-of-function variants. Our preliminary molecular modelling analyses did not reveal any significant modification of the *DDX17* structure in which patient missense mutations were introduced (data not shown). However, our analysis was based on the only known 3D structure of human *DDX17*, which is limited to the helicase core domain<sup>25</sup>. Recently it has been shown that the two disordered and

flexible flanking domains also strongly affect the helicase activity of Dbp2, the yeast DDX17 ortholog.<sup>26</sup> It is thus currently impossible to accurately predict the impact of variants on DDX17 function, without taking into account the interactions between the structured and unstructured regions of the protein.

We, and others<sup>6,7</sup> have shown previously a role of DDX17 in the retinoic acid induced differentiation of neuroblastoma SH-SY5Y and pluripotent embryonal NTERA2 cells, respectively. However, this effect was most evident when DDX17 knockdown was combined with the concomitant depletion of its paralog DDX5. We now demonstrate that the downregulation of DDX17 alone is sufficient to alter neuronal development, both *in vivo* and *in vitro*. Both DDX17 and DDX5 have largely redundant functions, which probably explains why their joint depletion has such a strong effect compared to single protein depletion. Interestingly, two distinct shRNAs targeting DDX17 alter neuronal migration in the mouse cortex. Although shRNA-based strategies are prone to off-target disruption of neuronal migration in the mouse cortex<sup>27</sup>, the migration phenotype is compatible with our previous observation that DDX17 controls the activity of the Repressor Element 1-silencing transcription factor (REST) complex during neurogenesis<sup>6</sup> and that the REST/CoREST complex regulates neuronal migration.<sup>28</sup> Future studies using genetic knockout models will demonstrate the specificity of the migration phenotype. Furthermore, we report that DDX17 plays a role in axon morphogenesis that is independent of its function in neuronal migration.

*Xenopus* frogs have been used as pioneer model organisms since the mid-twentieth century, mainly in discovery research<sup>29</sup>. Gene editing was found to be exceptionally effective in them and their application as tools for studying disease has increased.<sup>30</sup> *X. tropicalis* are diploid tetrapods with very few gene duplications. Their genome structure has high levels of synteny with humans<sup>31</sup> and the initial determination that 80% of human disease genes have orthologues in this species is now thought to be an underestimate.<sup>32</sup> We and others have shown them to be highly suited to testing the links between a variant of uncertain significance and human disease phenotypes.<sup>20,33</sup> This can often be achieved without breeding the animals due to the efficiency of CRISPR/Cas resulting in very low levels of mosaicism in founders. Hence, they represent a rapid and cost-effective assay for gene-disease associations, filling an important gap between the mouse and zebrafish models. Here we have used mosaic founders, and both

heterozygous and compound heterozygous/homozygous  $F_1$  non-mosaic models to test the effect of a truncation in *ddx17*. Here, we have expanded the use of *Xenopus* in disease modelling; we applied a behavioral assay to tadpoles (the light dark-transition assay) that has been used previously in other models. We then confirmed the measurements of both working memory and anxiety with small molecule inhibitors. Additionally, we directly compared mosaic homozygous  $F_0$  crispants with non-mosaic  $F_1$  homozygous and  $F_1$  heterozygous animals. There were stronger phenotypic effects in mosaic founder animals. This is a known phenomenon, which may be associated with a failure to activate compensatory mechanisms in mosaic animals, including crispants in another aquatic model, zebrafish (reviewed by Rouf *et al.*<sup>34</sup>). This suggests that *Xenopus* behave like zebrafish in this respect.

Our data offer some limited insights into the mechanism whereby DDX17 variants affecting its function relate to the disease phenotype. Since DDX17 is known to regulate gene expression at multiple levels, the different pathological features associated with *DDX17* mutations likely result from the altered expression of some of its target genes and transcripts. Indeed, our transcriptomic analysis showed that 350 genes may be impacted, a large proportion of which are important for development and morphogenesis, and most particularly for neurogenesis. This includes several key transcription factors (*NEUROG2*, *RARB*, *MYCN*...), the deregulation of which could have direct and indirect effects on many other genes in the course of embryonic development. Furthermore, the DDX17-dependent regulation of several genes coding for trans-membrane receptors and ligands associated with axon guidance is also of particular significance, considering the altered axonal development observed upon DDX17 knockdown in mice and tadpoles, and the neurological phenotype observed in patients. Whilst further work is needed, the goal of this study is to establish DDX17 as a novel neurodevelopmental disease gene and enable identification of more patients to further elucidate the genotype-phenotype relationship.

## Conclusion

We have identified 11 patients with neurodevelopmental phenotypes harbouring monoallelic *de novo* variants in *DDX17*. Functional experiments (*in vitro* and *in vivo*) show that *DDX17* is important in neurodevelopmental processes, in keeping with the observed human phenotype. *Ddx17* knockdown of

newborn mice showed impaired axon outgrowth, and reduced axon outgrowth and branching was observed in primary cortical neurons *in vitro*. The axon outgrowth phenotype was replicated in crispant *ddx17* tadpoles, including in a heterozygous (F<sub>1</sub>) model. Crispant tadpoles had clear functional neural defects and showed an impaired neurobehavioral phenotype. Transcriptomic analysis further supports the role of *DDX17* in neurodevelopmental processes, particularly neurogenesis. These results strongly support that monoallelic loss-of-function variants in *DDX17* cause a neurodevelopmental phenotype.

## References

1. Bourgeois, C.F., Mortreux, F., and Auboeuf, D. (2016). The multiple functions of RNA helicases as drivers and regulators of gene expression. *Nat Rev Mol Cell Biol* 17, 426-438. 10.1038/nrm.2016.50.
2. Bohnsack, K.E., Yi, S., Venus, S., Jankowsky, E., and Bohnsack, M.T. (2023). Cellular functions of eukaryotic RNA helicases and their links to human diseases. *Nat Rev Mol Cell Biol*. 10.1038/s41580-023-00628-5.
3. Xing, Z., Ma, W.K., and Tran, E.J. (2019). The DDX5/Dbp2 subfamily of DEAD-box RNA helicases. *Wiley Interdiscip Rev RNA* 10, e1519. 10.1002/wrna.1519.
4. Giraud, G., Terrone, S., and Bourgeois, C.F. (2018). Functions of DEAD box RNA helicases DDX5 and DDX17 in chromatin organization and transcriptional regulation. *BMB Rep* 51, 613-622. 10.5483/BMBRep.2018.51.12.234.
5. Fuller-Pace, F.V. (2013). The DEAD box proteins DDX5 (p68) and DDX17 (p72): multi-tasking transcriptional regulators. *Biochim Biophys Acta* 1829, 756-763. 10.1016/j.bbagr.2013.03.004.
6. Lambert, M.-P., Terrone, S., Giraud, G., Benoit-Pilven, C., Cluet, D., Combaret, V., Mortreux, F., Auboeuf, D., and Bourgeois, C.F. (2018). The RNA helicase DDX17 controls the transcriptional activity of REST and the expression of proneural microRNAs in neuronal differentiation. *Nucleic acids research* 46, 7686-7700.
7. Suthapot, P., Xiao, T., Felsenfeld, G., Hongeng, S., and Wongtrakoongate, P. (2022). The RNA helicases DDX5 and DDX17 facilitate neural differentiation of human pluripotent stem cells NTERA2. *Life Sciences* 291, 120298.

8. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., et al. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 581, 434-443. 10.1038/s41586-020-2308-7.
9. Seaby, E.G., Thomas, N.S., Webb, A., Brittain, H., Taylor Tavares, A.L., Baralle, D., Rehm, H.L., O'Donnell-Luria, A., and Ennis, S. (2023). Targeting de novo loss-of-function variants in constrained disease genes improves diagnostic rates in the 100,000 Genomes Project. *Human Genetics* 142, 351-362.
10. Seaby, E.G., Smedley, D., Tavares, A.L.T., Brittain, H., van Jaarsveld, R.H., Baralle, D., Rehm, H.L., O'Donnell-Luria, A., Ennis, S., and Consortium, G.E.R. (2022). A gene-to-patient approach uplifts novel disease gene discovery and identifies 18 putative novel disease genes. *Genetics in Medicine* 24, 1697-1707.
11. Sobreira, N., Schiettecatte, F., Valle, D., and Hamosh, A. (2015). GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. *Human mutation* 36, 928-930.
12. Philippakis, A.A., Azzariti, D.R., Beltran, S., Brookes, A.J., Brownstein, C.A., Brudno, M., Brunner, H.G., Buske, O.J., Carey, K., Doll, C., et al. (2015). The Matchmaker Exchange: a platform for rare disease gene discovery. *Hum Mutat* 36, 915-921. 10.1002/humu.22858.
13. 100, G.P.P.I. (2021). 100,000 genomes pilot on rare-disease diagnosis in health care—preliminary report. *New England Journal of Medicine* 385, 1868-1880.
14. Dardenne, E., Polay Espinoza, M., Fattet, L., Germann, S., Lambert, M.P., Neil, H., Zonta, E., Mortada, H., Gratadou, L., Deygas, M., et al. (2014). RNA helicases DDX5 and DDX17 dynamically orchestrate transcription, miRNA, and splicing programs in cell differentiation. *Cell Rep* 7, 1900-1913. 10.1016/j.celrep.2014.05.010.
15. Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., and Li, H. (2021). Twelve years of SAMtools and BCFtools. *Gigascience* 10. 10.1093/gigascience/giab008.
16. Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166-169. 10.1093/bioinformatics/btu638.

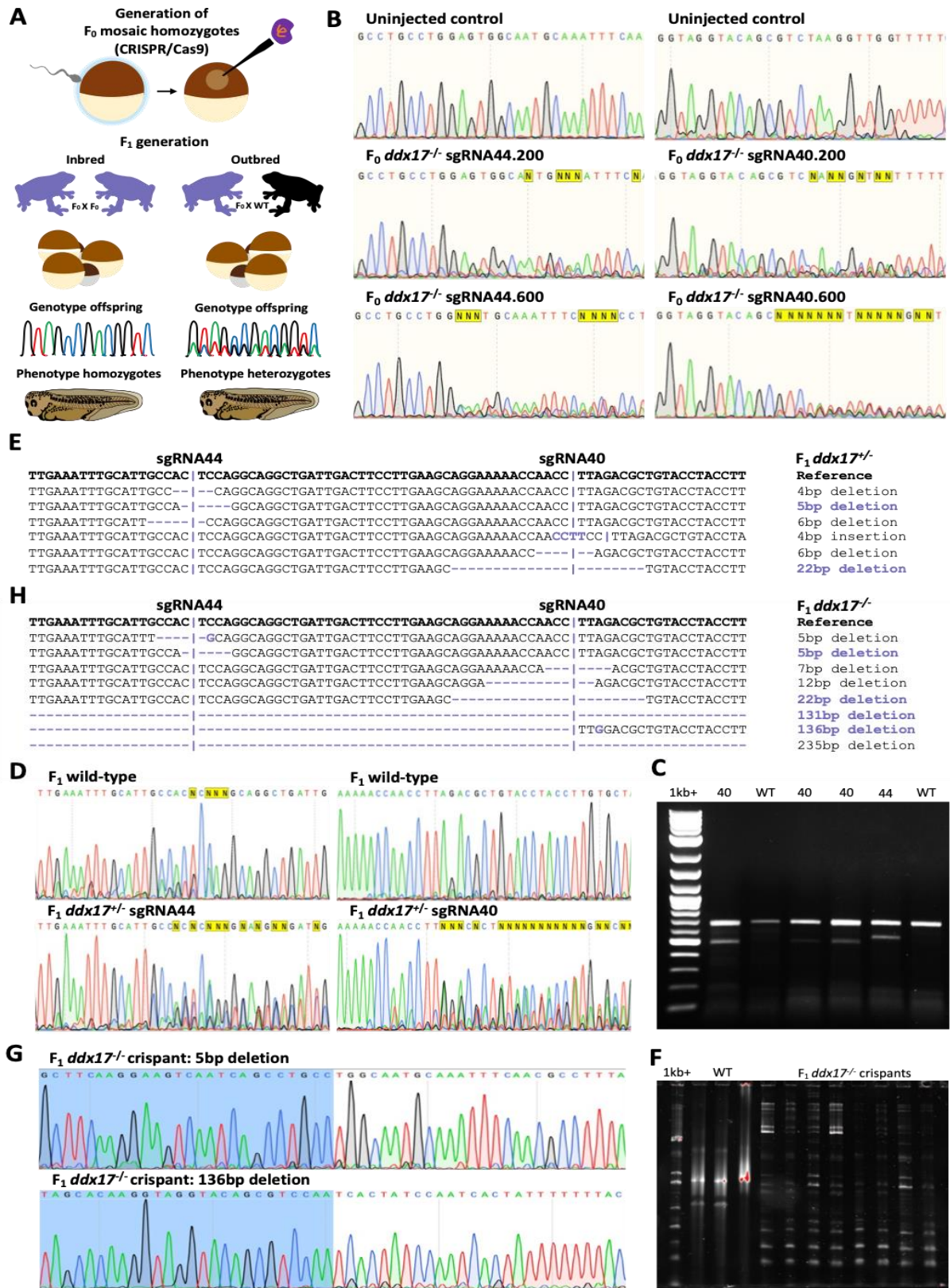


17. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15, 550. 10.1186/s13059-014-0550-8.
18. Ge, S.X., Jung, D., and Yao, R. (2020). ShinyGO: a graphical gene-set enrichment tool for animals and plants. *Bioinformatics* 36, 2628-2629. 10.1093/bioinformatics/btz931.
19. Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., Voelkerding, K., and Rehm, H.L. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in Medicine* 17, 405-424. 10.1038/gim.2015.30.
20. Ismail, V., Zachariassen, L.G., Godwin, A., Sahakian, M., Ellard, S., Stals, K.L., Baple, E., Brown, K.T., Foulds, N., Wheway, G., et al. (2022). Identification and functional evaluation of GRIA1 missense and truncation variants in individuals with ID: An emerging neurodevelopmental syndrome. *Am J Hum Genet* 109, 1217-1241. 10.1016/j.ajhg.2022.05.009.
21. Nakayama, T., Fish, M.B., Fisher, M., Oomen-Hajagos, J., Thomsen, G.H., and Grainger, R.M. (2013). Simple and efficient CRISPR/Cas9-mediated targeted mutagenesis in *Xenopus tropicalis*. *Genesis* 51, 835-843. 10.1002/dvg.22720.
22. Panthi, S., Chapman, P.A., Szyszka, P., and Beck, C.W. (2023). Characterisation and automated quantification of induced seizure-related behaviours in *Xenopus laevis* tadpoles. *J Neurochem*. 10.1111/jnc.15836.
23. Hewapathirane, D.S., and Haas, K. (2009). The Albino *Xenopus laevis* Tadpole as a Novel Model of Developmental Seizures. *Animal Models of Epilepsy: Methods and Innovations*, 45-57.
24. Fontana, B.D., Alnassar, N., and Parker, M.O. (2022). The zebrafish (*Danio rerio*) anxiety test battery: comparison of behavioral responses in the novel tank diving and light-dark tasks following exposure to anxiogenic and anxiolytic compounds. *Psychopharmacology (Berl)* 239, 287-296. 10.1007/s00213-021-05990-w.
25. Ngo, T.D., Partin, A.C., and Nam, Y. (2019). RNA Specificity and Autoregulation of DDX17, a Modulator of MicroRNA Biogenesis. *Cell Rep* 29, 4024-4035.e4025. 10.1016/j.celrep.2019.11.059.

26. Song, Q.X., Liu, N.N., Liu, Z.X., Zhang, Y.Z., Rety, S., Hou, X.M., and Xi, X.G. (2023). Nonstructural N- and C-tails of Dbp2 confer the protein full helicase activities. *J Biol Chem* 299, 104592. [10.1016/j.jbc.2023.104592](https://doi.org/10.1016/j.jbc.2023.104592).
27. Baek, S.T., Kerjan, G., Bielas, S.L., Lee, J.E., Fenstermaker, A.G., Novarino, G., and Gleeson, J.G. (2014). Off-target effect of doublecortin family shRNA on neuronal migration associated with endogenous microRNA dysregulation. *Neuron* 82, 1255-1262.
28. Volvert, M.-L., Prevot, P.-P., Close, P., Laguesse, S., Pirotte, S., Hemphill, J., Rogister, F., Kruzy, N., Sacheli, R., and Moonen, G. (2014). MicroRNA targeting of CoREST controls polarization of migrating cortical neurons. *Cell reports* 7, 1168-1183.
29. De Robertis, E.M., and Gurdon, J.B. (2021). A Brief History of *Xenopus* in Biology. *Cold Spring Harb Protoc* 2021. [10.1101/pdb.top107615](https://doi.org/10.1101/pdb.top107615).
30. Kostiuk, V., and Khokha, M.K. (2021). *Xenopus* as a platform for discovery of genes relevant to human disease. *Curr Top Dev Biol* 145, 277-312. [10.1016/bs.ctdb.2021.03.005](https://doi.org/10.1016/bs.ctdb.2021.03.005).
31. Hellsten, U., Harland, R.M., Gilchrist, M.J., Hendrix, D., Jurka, J., Kapitonov, V., Ovcharenko, I., Putnam, N.H., Shu, S., Taher, L., et al. (2010). The genome of the Western clawed frog *Xenopus tropicalis*. *Science* 328, 633-636. [10.1126/science.1183670](https://doi.org/10.1126/science.1183670).
32. Blum, M., and Ott, T. (2018). *Xenopus*: An Undervalued Model Organism to Study and Model Human Genetic Disease. *Cells Tissues Organs* 205, 303-313. [10.1159/000490898](https://doi.org/10.1159/000490898).
33. Macken, W.L., Godwin, A., Wheway, G., Stals, K., Nazlamova, L., Ellard, S., Alfares, A., Aloraini, T., AlSubaie, L., Alfadhel, M., et al. (2021). Biallelic variants in *COPB1* cause a novel, severe intellectual disability syndrome with cataracts and variable microcephaly. *Genome Med* 13, 34. [10.1186/s13073-021-00850-w](https://doi.org/10.1186/s13073-021-00850-w).
34. Rouf, M.A., Wen, L., Mahendra, Y., Wang, J., Zhang, K., Liang, S., Wang, Y., Li, Z., Wang, Y., and Wang, G. (2023). The recent advances and future perspectives of genetic compensation studies in the zebrafish model. *Genes & Diseases* 10, 468-479. <https://doi.org/10.1016/j.gendis.2021.12.003>.

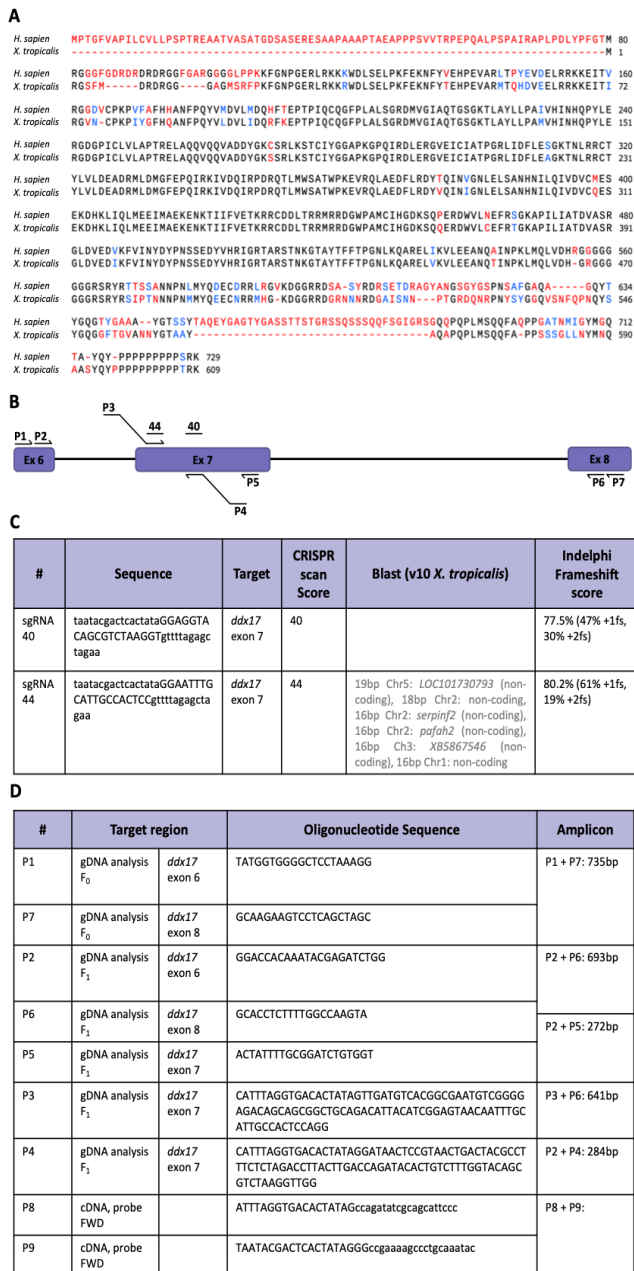
## Supplementary Figures

**Figure S1** | Crispr/Cas9 microinjection into *X. tropicalis* eggs produces mosaic homozygous crispant tadpoles encoding truncated Ddx17 which is inherited in the F1 generation.



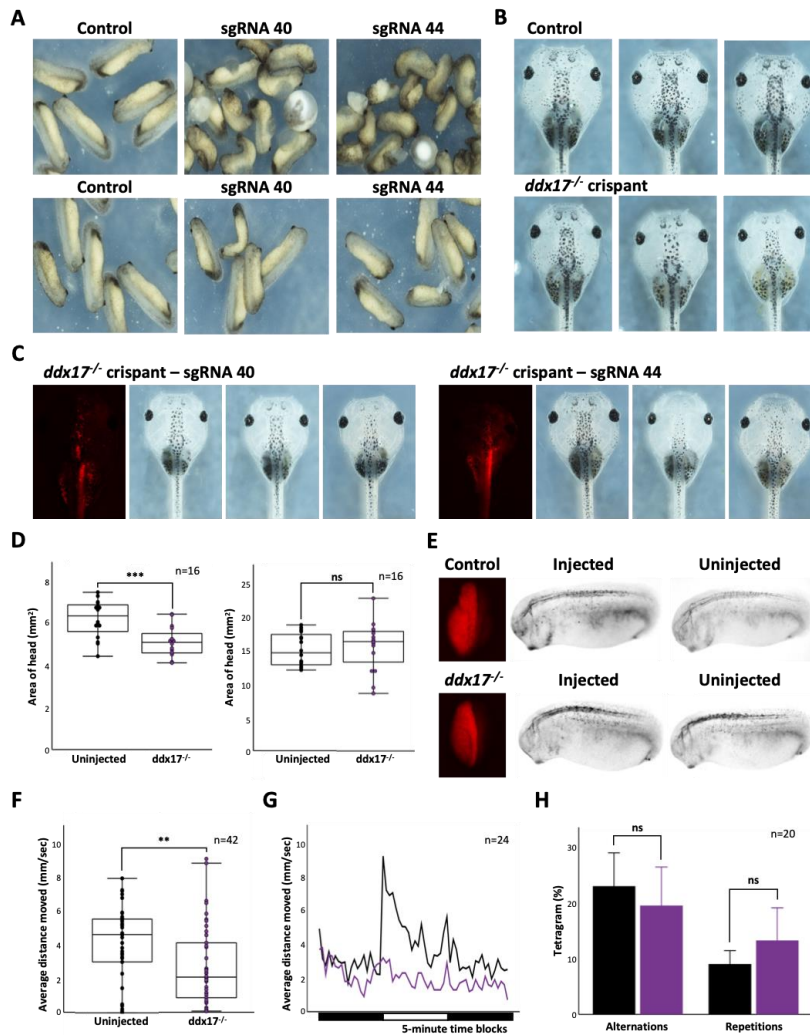
*F0* tadpoles were microinjected with increasing amounts of CRISPR/Cas9 complex containing either sgRNA44 or sgRNA40 (see fig S2C). Sequencing of PCR products containing the target region (A) confirms that indels have been produced in the embryo genome in a dose dependent manner (B). These tadpoles were grown to adulthood and either outbred to produce F1 heterozygotes (E, D, C) similar to patients or inbred to produce compound heterozygotes that produce only truncated *Ddx17* (H, G, F). A T7E1 assay was used to identify heterozygotes and the genotype was confirmed by Sanger sequencing (C, D). Sequence analysis of PCR products spanning the CRISPR target site confirms the predominant genotype as a 5 bp deletion (sgRNA44) or 22 bp deletion (sgRNA40) for the heterozygotes (E). Cloning and sequence analysis of the compound heterozygous F1 model shows a variety of deletions from 131 to 235 bp (H, G). Further, this model produces an unusual genotype in which complex mixtures of larger deletions are also seen; they can be identified by polyacrylamide electrophoresis (F) of PCR products spanning the target region.

**Figure S2**



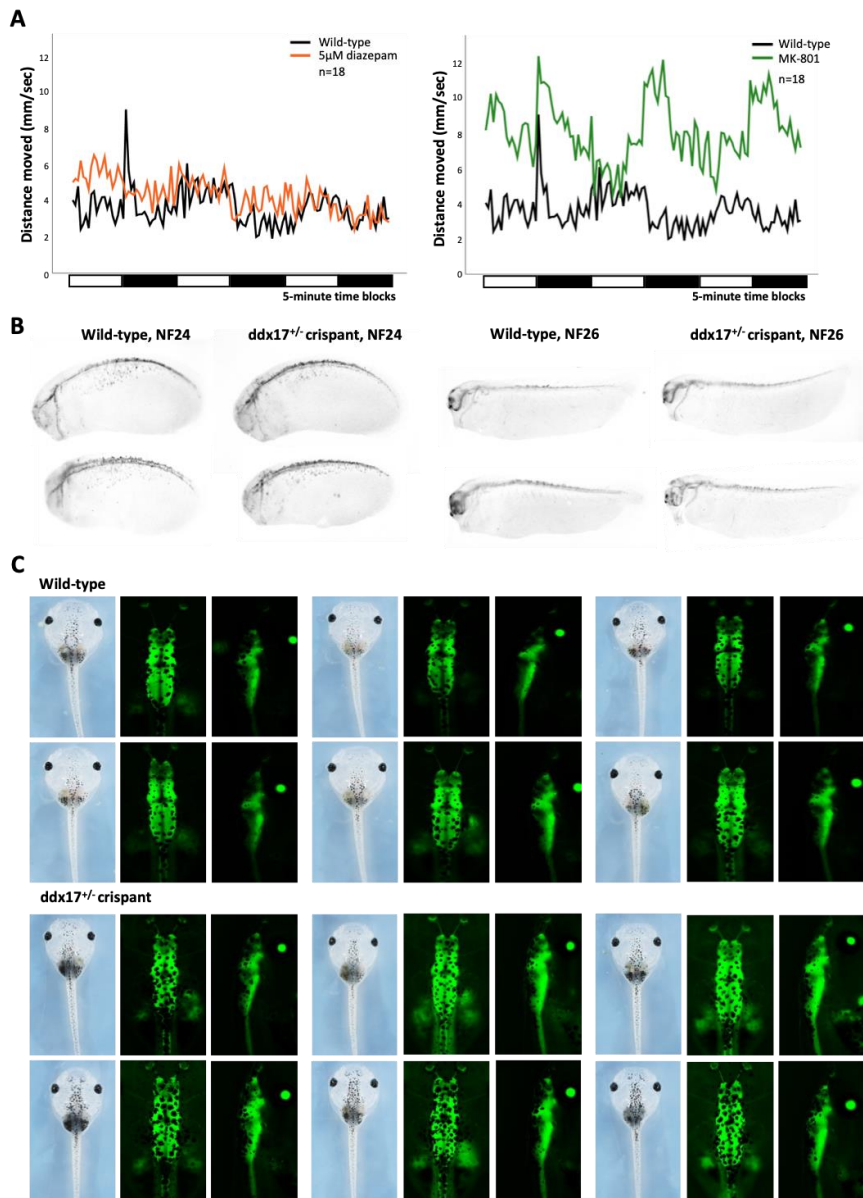
A. The amino acid alignment between the *H. sapiens* and *X. tropicalis* Ddx17 proteins. B. The positions of sgRNAs and genotyping PCR primers in the *X. tropicalis* ddx17 gene. The sequences of the different oligonucleotides used for the *Xenopus* experiments in this study: sgRNAs (C) and PCR primers (D).

**Figure S3** | F<sub>0</sub> mosaic homozygous *X. tropicalis* display reduced axon outgrowth, and working memory like F<sub>1</sub> models, but also gastrulation defects and short term microcephaly



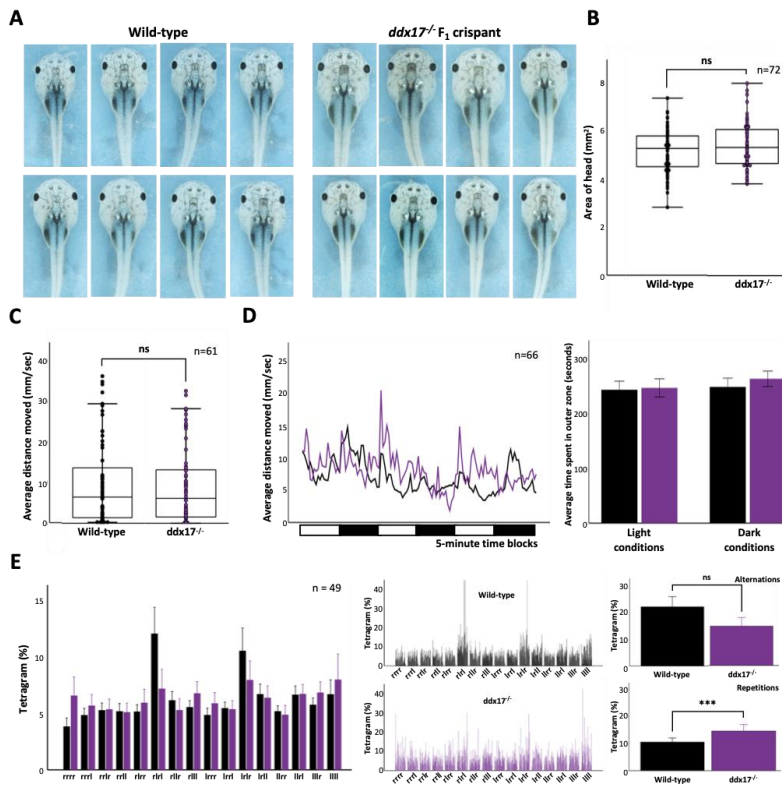
A. Gastrulation defects are observed in approximately 40% of embryos microinjected with either *ddx17* sgRNA; these embryos were removed from subsequent analysis. B. *ddx17* F<sub>0</sub> crispants show normal gross morphology. C. One of 2-cell microinjection of *ddx17* CRISPR/cas does not reveal any injection dependent morphological changes but does show reduced axon outgrowth on the injected side when embryos are stained with HNK1 antibody (E). D. At stage 41 the head area of crispants is significantly reduced compared with controls but by stage 48 this difference is no longer present, suggesting a delay to growth rather than an absence. F. Crispants show reduced locomotion when compared to control tadpoles. G. This is also seen during the light-dark behaviour challenge. H. In the FMP Y-maze test F<sub>0</sub> crispants show fewer alternations and more repetitions than controls, although these are not significant.

**Figure S4**



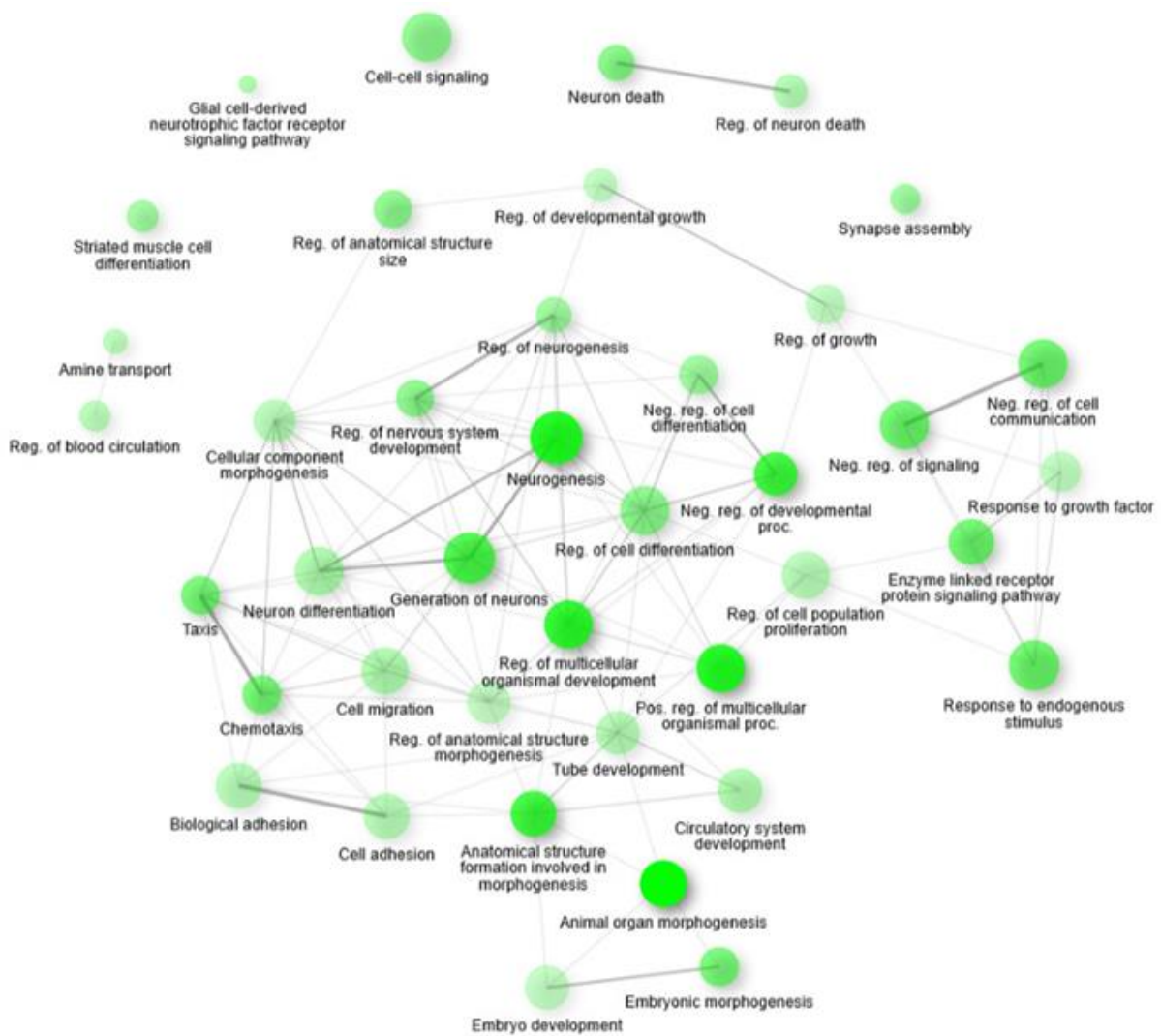
*A.* Dark-light transitions cause initial movement responses. Evidence shows that the movement is inhibited by diazepam treatment at concentrations that are anxiolytic; treatment with the NMDA inhibitor MK801 causes the transition response to be repeated. Together these show that this assay measures both anxiousness and working memory in tadpoles. *B.* Heterozygous *ddx17* crispant *X. tropicalis* have reduced neuronal outgrowth as seen by HNK1 staining and *C.* normal morphology in bright field microscopy or when crispants are made in a *tubb2b*-GFP transgenic line.

**Figure S5** | Compound heterozygous *ddx17*<sup>-/-</sup> tadpoles are morphologically normal but show working memory deficits



*A. Brightfield images of control and *ddx17*<sup>-/-</sup> show no clear morphological differences and measuring the area of their heads similarly reveals no difference (B). C. Wild-type and *ddx17*<sup>-/-</sup> tadpoles have the same locomotion activity but *ddx17*<sup>-/-</sup> tadpoles retain their startle response during repeated light-dark transitions indicating a memory deficit (D). E. In the FMP Y-maze assay the reduction in working memory indicated by reduced alternations in the maze is not-significant although a clear trend but the number of repetitions increases significantly. Together there is strong evidence for a memory deficit in these animals.*

Figure S6



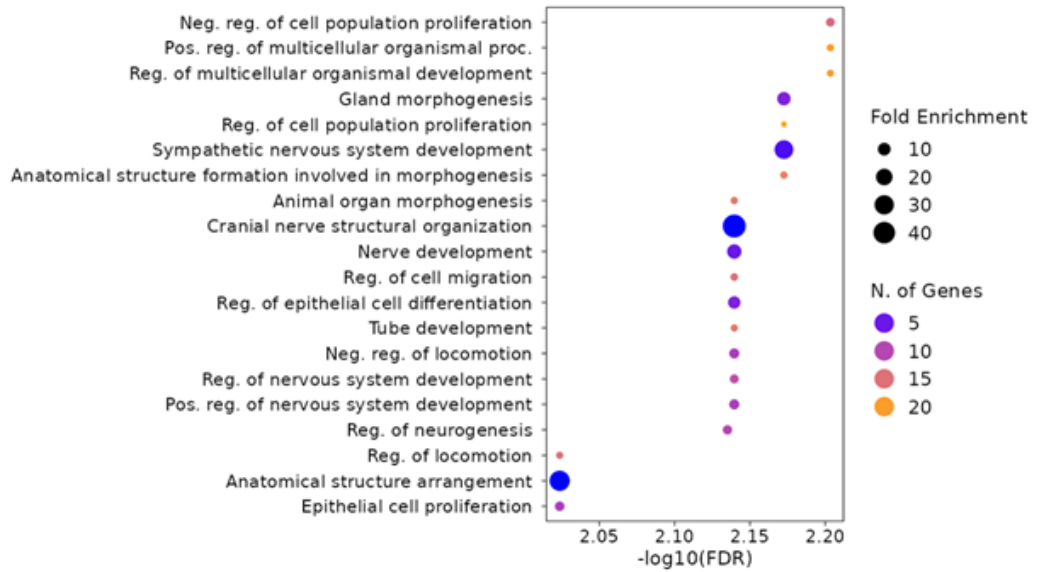
Network representation of the top 40 enriched biological processes. The thickness of the lines connecting related GO terms reflects the percent of overlapping genes. The size and darkness of the nodes correspond respectively to the number of genes and to the significance of enriched gene sets.



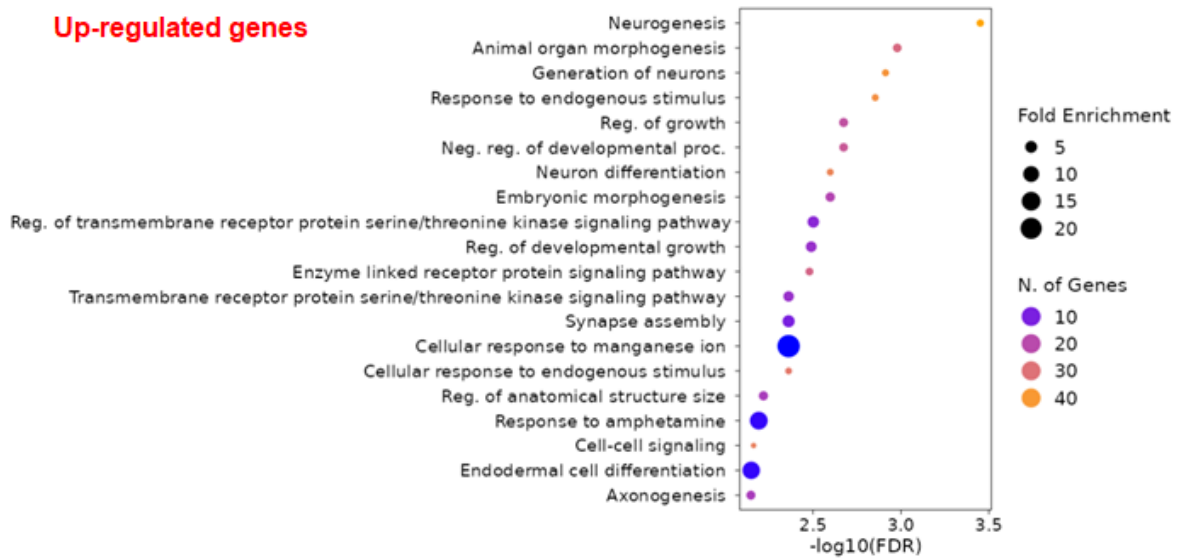
Figure S7

**A**

**Down-regulated genes**

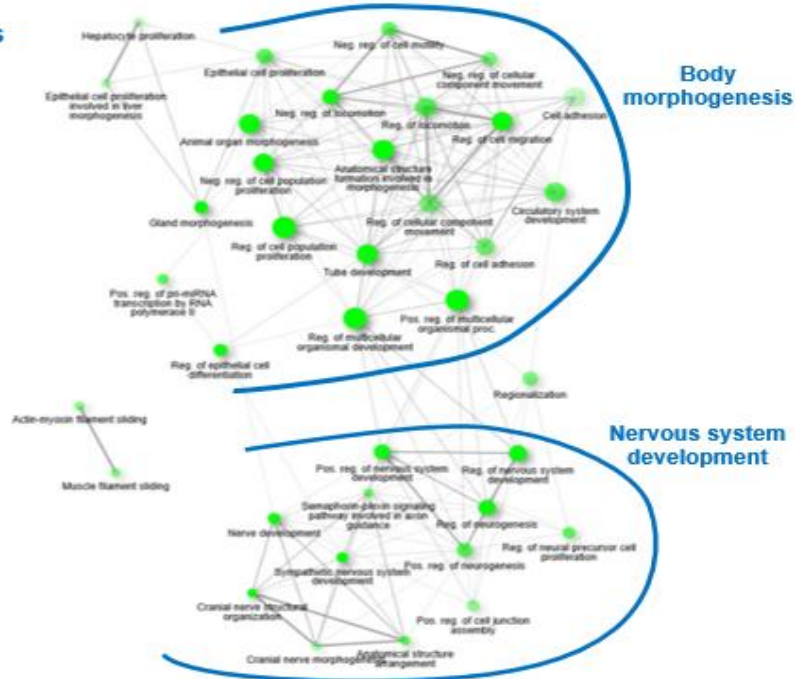


**Up-regulated genes**

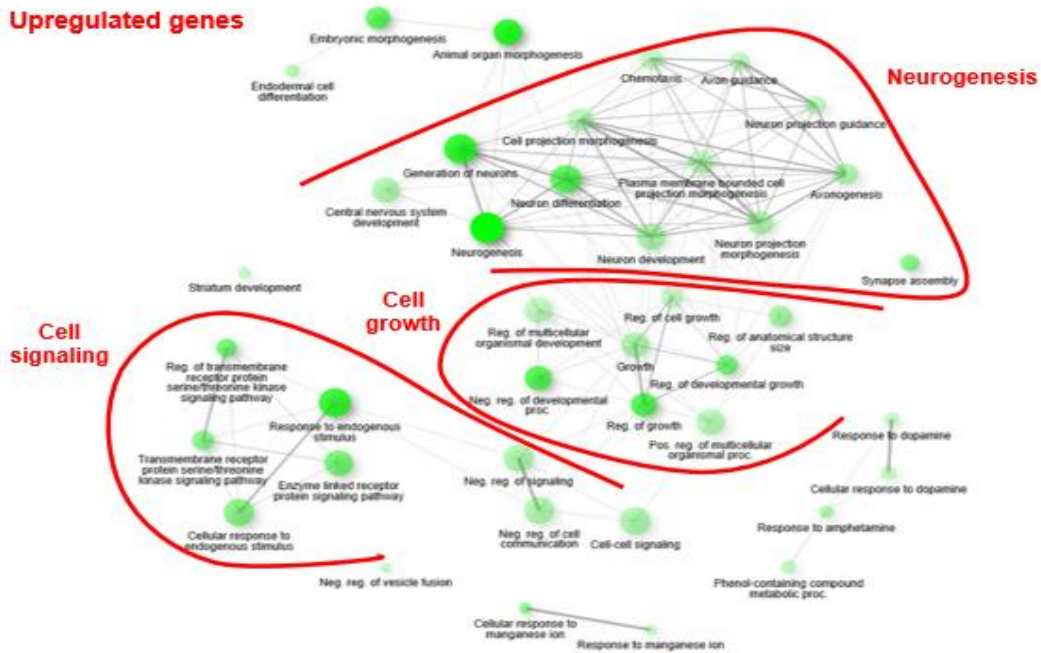


**B**

**Downregulated genes**



**Upregulated genes**



A. ShinyGO analysis showing the top 20 enriched biological processes impacted by DDX17 knockdown for down-regulated and up-regulated genes.

B. Network representation of the top 40 enriched biological processes for down-regulated and up-regulated genes. Details are as in Supplementary File A.

## Supplementary Tables

**Table S1**

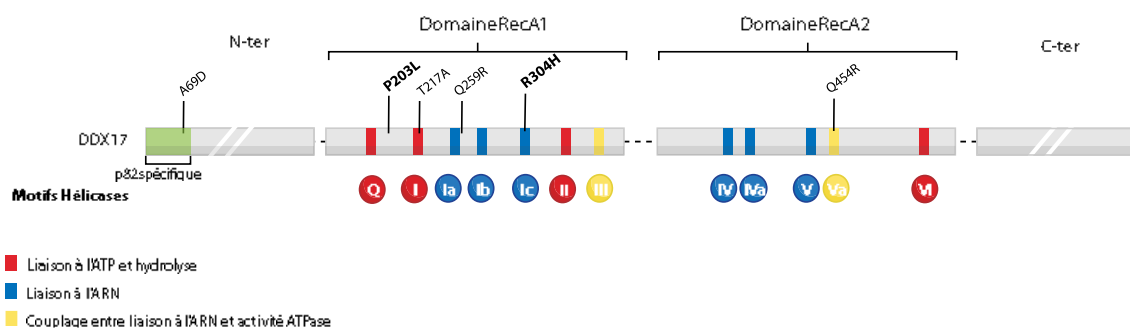
REAGENTS or RESOURCE	SOURCE	IDENTIFIER
Experimental Models : Organisms/Strains		
Mice : E15.5 pregnant females - Swiss (outbred) - 129Sv/B6 (F1 hybrid of inbred mice)	Janvier Labs In house	N/A
Critical Commercial Assays		
Plasmid DNA, endotoxin free Midi prep kit from Nucleobond Xtra Midi EF	Macherey-Nagel	Cat#740420.50
Chemicals, Peptides, and Recombinant Proteins		
Isoflurane (Isoflu-Vet 1000mg/g)	Dechra	Cat#ISO008 (Centravet)
Buprenorphin (Buprecare) 73422 BUPRECARE 0.3mg/ml INJ 10 ML	Ecuphar	Cat#73422
Povidone-iodine (Betadine) Scrub 4%	Vetoquinol	Cat#VET003 (Centravet)
Povidone-iodine (Vétédine) Solution	Vetoquinol	Cat#VET002 (Centravet)
Ocry-gel 10g	TVM Laboratory	Cat#OCR002 (Centravet)
Paraformaldehyde 32%	Electron Microscopy Sciences, Fisher	Cat#15714-S
HEPES buffer ; Ph7.4 ; 2.5mM	Fisher	Cat#BP299
Calcium chloride 1mM	Sigma, Merck	Cat#C1016

Magnesium sulfate 1mM	Sigma, Merck	Cat#M2643
Sodium bicarbonate 7,5% solution	Gibco, Fisher	Cat#25080
D-glucose	Fisher	Cat#47829
HBSS (10X)	Gibco, Fisher	Cat#14185-045
Papain	Worthington	Cat#LK0033176
Deoxyribonuclease I from bovine pancreas	Sigma, Merck	Cat#D5025-150KU
Neurobasal medium™ (1X)	Gibco, Fisher	Cat# 21103-049
B27™ serum free-supplement (50X)	Gibco, Fisher	Cat# 17504-044
N2 serum-free supplement (100X)	Gibco, Fisher	Cat# A13707.01
Glutamax™ supplement (100X)	Gibco, Fisher	Cat# 35050-038
Penicillin-Streptomycin solution (100X)	Gibco, Fisher	Cat# 15140-122
Sterile PBS pH7,4 (1X)	Gibco	Cat#10010-015
Phosphate buffered saline (PBS)	Sigma	Cat#P4417-50TAB
Triton™ X-100	MP Biomedical, Fisher	Cat#9002-93-1
Corning BioCoat 12mm#1 German Glass Coverslips, Round	Corning, Fisher	Cat#P35G-1.5
Fluoromount G	Invitrogen	00-4958-02
Fast Green (1:20) ratio	Sigma, Merck	Cat#F7252
Nuclease free water	Ambion	Cat#AM9937

Recombinant DNA		
pLKO.1		NA
pCAG	Courchet, J et al., 2013	NA
pCAG-mVENUS	Hand and Polleux, 2011	NA
shDDX17 #1		
shDDX17 #2		
pCAG-DDX17-WT	Vector Builder	NA
Antibodies		
Chicken anti-GFP	Rockland	mAb# 600-901-B12
Hoechst 33258 solution	Merck, Sigma	Cat#94403
Rabbit anti-MAP2	Invitrogen	Cat#PA5-17646
Mouse anti-neurofilament, clone SMI-312	Biologend	Cat#837904
Goat anti-chicken Alexa Fluor Plus 488	Invitrogen	mAb#A32931
Donkey anti-mouse Alexa Fluor Plus 647	Invitrogen	mAb#A31571
Donkey anti-rabbit Alexa Fluor 546	Invitrogen	mAb#A10040
Software		
NIS-Element Advanced Research	Nikon Corporation, Melville, NY	NA

GraphPad PRISM	GraphPad Software Inc.	Version 9
Excel®	Microsoft Corporation	NA
ImageLab	Bio-Rad™	NA
Adobe Photoshop®	Adobe	NA
Microscopy		
Nikon Ti-E microscope	Nikon	NA
Objective lenses (Nikon) : x10 PlanApo ; NA 0.45, x20 PlanApo VC ; NA 0.75	Nikon	NA
C2 laser scanning confocal microscope	Nikon	NA
Other		
Vibratome VT1000S	Leica Biosystem	NA
Micro-injector Picospritzer III	Harvard Apparatus	
Electroporator ECM 830 BTX and Generator Footswitch (model 1250FS)	Harvard Apparatus	
Gold-coated electrodes (GenePads 5x7mm BTX)	Harvard Apparatus	NA
Tweezertrodes kit, 3mm platinum BTX	Harvard Apparatus	Cat#45-0487
Borosilicate Glass capillaries O.D. :1mm, I.D. :0,50m, 10cm legth, FIMT	World Precision Instrument	Cat# GBF100-50-10

En parallèle avec l'étude de l'effet d'une déplétion de DDX17 sur le transcriptome, nous avons aussi cherché à caractériser moléculairement les mutants DDX17. Ces travaux ne font pas encore l'objet d'un article. Nous nous sommes concentrés sur deux mutations identifiées chez deux patients différents, correspondant aux variants Arg304His et Pro203Leu (Figure 20). Ces deux mutations ont été sélectionnées car il s'agit de mutations faux-sens, produisant des protéines mutées, permettant ainsi d'analyser l'effet de ces mutations sur le comportement de l'hélicase. Les deux mutations se trouvent dans le domaine RecA1 de DDX17. Plus précisément, la mutation R304H touche le domaine Ic (motif «TPGRLID » parfois désigné aussi comme le domaine Ib), qui est impliqué dans la liaison à l'ARN, alors que la mutation P203L touche la région se trouvant entre le motif Q et le domaine ATPase I. Une simulation de la capacité de DDX17 à fixer l'ARN, basée sur la seule structure 3D connue de DDX17 (Ngo et al., 2019) et réalisée par Stéphane Réty, chercheur au LBMC, prédit d'ailleurs que le mutant DDX17-R304H fixe moins efficacement l'ARN que DDX17-WT (Figure 21A).



**Figure 20 : Localisation des substitutions d'acides aminés dans DDX17**

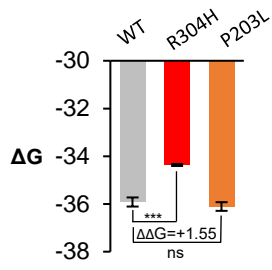
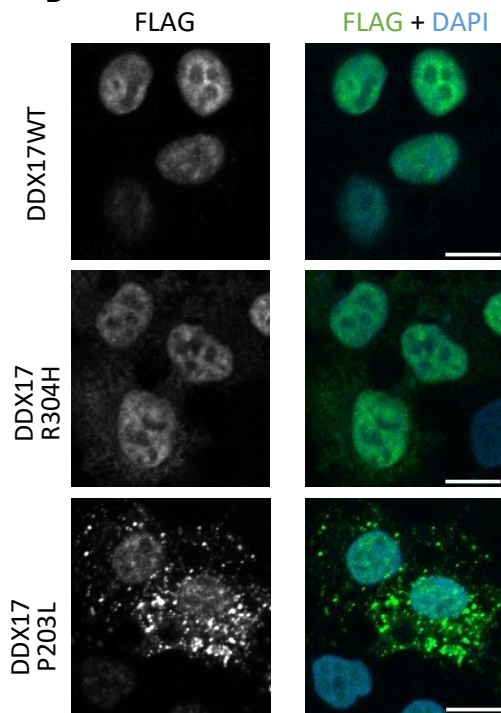
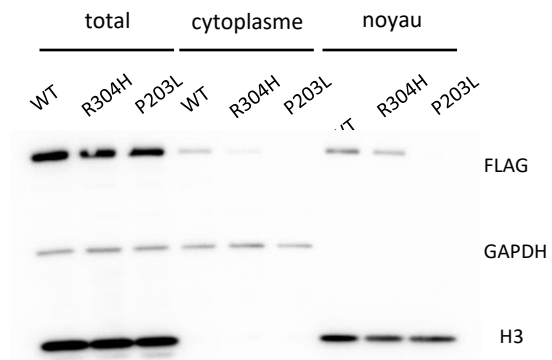
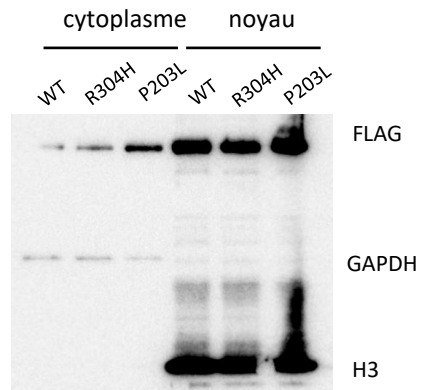
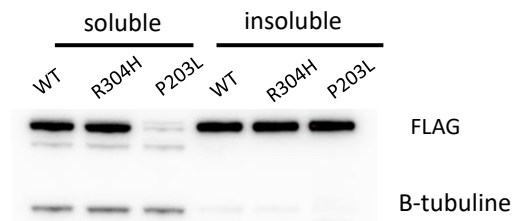
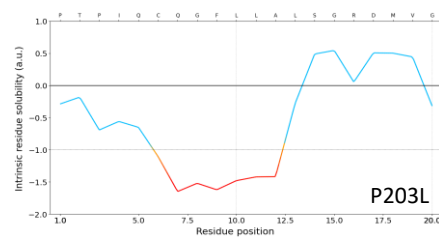
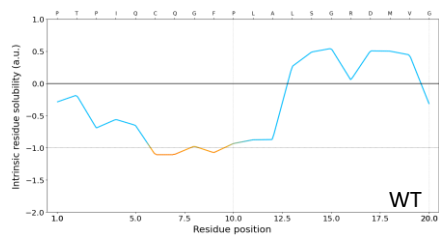
Représentation schématique de la localisation des 6 substitutions résultant des mutations faux-sens identifiées chez 6 des 11 patients identifiés. Les 5 autres variants conduisent à l'apparition d'un codon stop prématurés entraînant une perte de fonction. Les variants en gras correspondent à ceux étudiés lors de mes travaux

L'analyse de la localisation des protéines DDX17 mutées par immunofluorescence ne montre pas de différence significative entre DDX17R304H et DDX17WT, qui sont localisées essentiellement dans les noyaux des cellules transfectées, même si on détecte

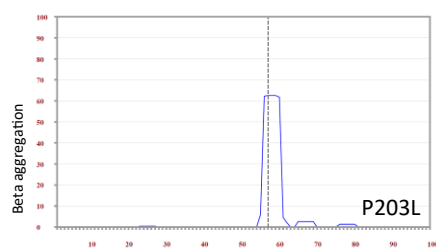
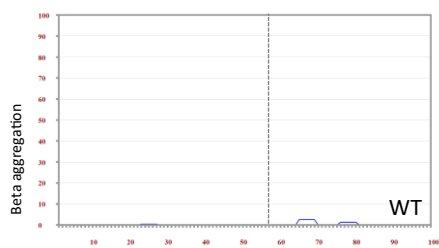
un peu plus de protéine R304H dans le cytoplasme. Dans le cas de la protéine P203L par contre, on observe un comportement radicalement différent puisque ces protéines mutées semblent en effet s'agréger en foci, majoritairement cytoplasmiques mais aussi nucléaires (Figure 21B). Une analyse par western blot des fractions cytoplasmiques et nucléaires de cellules transfectées extraites avec un tampon peu stringent permet de confirmer la localisation majoritaire de DDX17 WT et R304H dans le noyau (Figure 21C). On observe par contre une quantité très faible de P203L dans les deux fractions. La protéine P203L étant bien détectée dans la fraction totale, préparée dans des conditions plus stringentes, nous avons fait l'hypothèse que la perte de P203L lors du fractionnement pouvait être une conséquence de l'agrégation observée en immunofluorescence et d'une solubilité plus faible de cette protéine. Pour valider cette hypothèse, j'ai récupéré le culot des deux fractions après l'étape de centrifugation et je les ai ré-extrait avec un tampon de lyse stringent. Dans ces conditions, on observe à nouveau la protéine P203L dans les fractions cytoplasmiques et nucléaires (Figure 21D). Par ailleurs, l'analyse par western blot de fractions insoluble et soluble montre que la protéine P203L est quasiment absente de la fraction soluble mais se trouve bien dans la fraction insoluble (Figure 21E). On peut noter que, de façon surprenante, si les protéines DDX17 WT et R304H se trouvent bien dans la fraction soluble, elles sont aussi détectées dans la fraction insoluble, à un niveau quasiment identique à P203L.

Ces résultats suggèrent que le variant P203L entraîne une baisse de la solubilité de DDX17 et une forte augmentation de la propension de DDX17 à former des agrégats. Ces résultats expérimentaux sont par ailleurs confirmés par des prédictions informatiques de la capacité des séquences peptidiques à s'agréger ainsi que de leur solubilité (Figure 21F-G).



**A****B****C****D****E****F**

cohsoftware

**G**

Tango.crg

**Figure 21 : La mutation DDX17P203L entraîne une baisse de solubilité de la protéine DDX17 ainsi que son agrégation sous forme de foci cytoplasmiques et nucléaires**

(A) Prédiction de l'affinité de DDX17 WT et mutée pour l'ARNdb. (B) Les cellules HEK293T transfectées avec un plasmide exprimant une DDX17-FLAG sauvage ou mutée ont été immuno-marquées avec un anticorps anti-FLAG. Les noyaux ont été marqués au DAPI. Les barres d'échelle correspondent à 10  $\mu$ m. (C) Western blot du fractionnement des fractions cytoplasmiques et nucléaires de cellules HEK293T exprimant DDX17-FLAG sauvage ou mutée. La présence de histone H3 et la GAPDH permet de contrôler la qualité du fractionnement. (D) Western blot correspondant au culot (fraction insoluble) des fractions nucléaires et cytoplasmiques. (E) Western blot des fractions solubles et insolubles de cellules HEK293T exprimant DDX17-FLAG sauvage ou mutée. (F-G) Prédiction de la capacité d'agrégation et d'insolubilité de DDX17WT et DDX17P203L avec (F) CamSol ([Sormanni et al., 2015](#)) (acides aminés 115 à 134) (G) Tango ([Fernandez-Escamilla et al., 2004](#)) (acides aminés 68 à 167). La barre verticale grise indique l'acide aminé 203.

# Discussion

## **Régulation de la production de transcrits chimériques par les hélicases DDX5 et DDX17 et par MYCN dans le neuroblastome**

### **Impact du readthrough transcriptionnel induit par la déplétion de DDX5 et DDX17 dans les cellules de neuroblastome :**

La perturbation de la maturation de l'extrémité 3' des transcrits peut conduire à un readthrough transcriptionnel. Le readthrough transcriptionnel est notamment observé suite à l'induction d'un stress cellulaire, d'une infection virale et dans certains cancers. Les conséquences du readthrough et les possibles fonctions des transcrits alternatifs dont il entraîne la production restent mal caractérisées. Le readthrough transcriptionnel peut ainsi entraîner la production de transcrits chimériques, résultant de la transcription de deux gènes (voire plus) sur un même brin d'ARN.

L'analyse du RNA-seq de cellules SH-SY5Y traitées par siDDX5/17 nous a permis d'identifier 218 transcrits chimériques dont le taux est augmenté. 11 de ces transcrits ont pu être validés par RT-qPCR, confirmant qu'il ne s'agit pas du résultat d'un biais de séquençage. Parmi les 91 transcrits chimériques les plus significativement induits, 25 (27%) sont communs à une liste de transcrits chimériques identifiés dans des tumeurs de patients souffrant de neuroblastome.

Certains de ces transcrits chimériques sont prédits pour respecter le cadre de lecture et pouvoir produire des protéines chimériques. Parmi les transcrits validés par RT-qPCR, *ADCK4-NUMBL*, *CHCHD10-VPREB3*, *CTSD-IFITM10* et *RHBDD2-POR*, soit quatre transcrits sur onze, ont une ORF chimérique. Il a été possible de confirmer l'expression de la protéine chimérique CTSD-IFITM10 dans les cellules traitées par siDDX5/17, confirmant que le transcrit chimérique peut bien être traduit en protéine chimérique.

Un des paramètres limitant notre caractérisation de transcrits chimériques produit par la déplétion des deux hélicases est que l'on ne connaît pas leur profil exact d'épissage. Plusieurs articles ont rapporté que les gènes présentant un readthrough transcriptionnel présentaient aussi des défauts d'épissage (voir 4.2.3). La technique de séquençage short-read ne permet par définition que l'assemblage de fragments courts et ne permet donc

pas d'avoir la séquence exacte complète d'un transcrit donné. Un séquençage dit « long-read » pourrait permettre d'accéder à cette information. Il faut toutefois noter qu'une des limites du séquençage « long-read » est son taux d'erreur plus élevé, ce qui peut poser problème lorsque l'on examine des événements rares comme les transcrits chimériques. Par exemple, les plateformes actuelles de séquençage « long-read », notamment Pacific Biosciences (PacBio) et Oxford Nanopore Technologies (ONT), présentaient précédemment un taux d'erreur moyen compris entre 5 et 20 % (Rang et al., 2018; Weirather et al., 2017). De récents progrès ainsi que l'application de méthodes de correction semblent néanmoins permettre d'augmenter significativement la fiabilité du séquençage « long-read » (Wenger et al., 2019). Les algorithmes d'identification de transcrits chimériques présentant déjà une faible reproductibilité de leurs résultats entre eux, il est important de faire usage d'une technique de séquençage hautement fiable, afin de ne pas encore augmenter les faux positifs ou événements non détectés.

Il faut aussi ajouter que les RNA-seq analysés furent établis à partir de bibliothèques enrichies en ARN polyA+, écartant les ARNs non-polyadénylés. Or certains articles rapportent que les transcrits readthrough, produits suite à l'induction d'un stress cellulaire par exemple, ne sont pas forcément polyadénylés (Nemeroff et al., 1998; Vilborg et al., 2015). Cela pourrait expliquer pourquoi dans nos expériences de fractionnement cellulaire suivi de RT-qPCR, les transcrits chimériques testés sont bien présents dans le cytoplasme (résultats non montrés), alors que de précédents papiers indiquent que les transcrits issus de readthrough tel les DoGs sont retenus à la chromatine (Vilborg et al., 2015). Il pourrait donc être judicieux de procéder à un séquençage non spécifique des ARNs, afin de vérifier si l'on n'identifie pas davantage de transcrits présentant des défauts d'épissage et de terminaison transcriptionnelle.

#### Rôle de DDX5/17 et du readthrough transcriptionnel dans le neuroblastome :

Étant donné l'association entre une faible expression des gènes *DDX17* et *DDX5* et un mauvais pronostic des patients, il pourrait être intéressant d'explorer plus en détail le lien entre *DDX5/17* et l'oncogenèse ainsi que la progression du neuroblastome. Cette association est d'autant plus intrigante que *DDX5* et *DDX17* sont généralement surexprimés dans les autres cancers analysés (voir 5.5.3). Même si cette association pourrait n'être qu'une corrélation, il resterait intéressant de voir si une baisse d'expression des hélicases peut affecter la réponse des patients aux traitements ou

constituer un biomarqueur permettant de déterminer avec plus de précision le profil des patients et le traitement thérapeutique adapté.

Le readthrough induit par une baisse d'expression de DDX5 et DDX17 pourrait moduler l'agressivité des tumeurs de neuroblastomes. L'expression de transcrits chimériques pourrait affecter certaines fonctions cellulaires, via la production de protéines chimériques ou de protéines tronquées. On peut rappeler l'exemple du transcrit chimérique *ZNF451-BAG2*, exprimé dans les tumeurs et les lignées cellulaires de neuroblastome, et dont l'expression affecte la différenciation induite par l'acide rétinoïque (Shi et al., 2021). Comme indiqué précédemment, 25 des transcrits chimériques induits suite à la déplétion de DDX5 et DDX17 ont aussi été identifiés dans des tumeurs de neuroblastome. De plus, parmi ces 25 transcrits, seulement 1 est aussi détecté par l'analyse de 161 échantillons de tissus sains de glandes surrénales. Parmi les 24 restants, 4 sont identifiés dans au moins 10% des 498 tumeurs de neuroblastome analysées. Des analyses supplémentaires seraient nécessaires afin de déterminer si ces transcrits chimériques possèdent des propriétés pro-tumorales ou peuvent influencer la réponse des tumeurs aux traitements, mais ils pourraient potentiellement servir de biomarqueurs ou permettre de mieux orienter la prise en charge des patients.

Le transcrit *CTSD-IFITM10*, dont nous avons pu valider la traduction en protéine chimérique dans nos cellules, a aussi été identifié dans des échantillons de tumeurs mammaires, et sa déplétion spécifique par traitement avec des siRNA ciblant la jonction chimérique induit une baisse de prolifération des cellules MCF7 (Varley et al., 2014). Il serait donc intéressant de tester l'effet de ce transcrit sur la prolifération dans les cellules de neuroblastome, en analysant l'effet de sa surexpression ou de sa déplétion ciblée par des siARN dirigés contre la jonction chimérique.

Nous souhaitons pouvoir accéder aux données RNA-seq de tumeurs de patients afin de pouvoir vérifier s'il est possible de retrouver un lien entre la production de transcrits chimériques et les niveaux d'expression des hélicases DDX5 et DDX17, ainsi que MYCN dans les tumeurs. Cela nous permettrait aussi de rechercher si certains transcrits chimériques sont associés à certaines caractéristiques des tumeurs et des patients, ainsi que le pronostic de ceux-ci. Nous n'avons pas encore réussi à obtenir l'accès à ces données mais espérons pouvoir mener ces analyses par la suite.

L'induction de readthrough pourrait aussi plus largement affecter les fonctions cellulaires en induisant la production d'isoformes à longue 3' UTR, pouvant lier des miARN et des RBP différentes pouvant moduler leur demi-vie, leur localisation cellulaire et leur traduction.

#### Possibles autres effets pro-tumoraux d'une déplétion de DDX5 et DDX17 dans le neuroblastome

En addition de leur effet sur la maturation en 3' des transcrits, DDX5 et DDX17 sont aussi impliqués dans la réparation de l'ADN (Bader et al., 2022; Fortuna et al., 2021; Sessa et al., 2021; Yu et al., 2020), ainsi que dans la régulation des structures secondaires des acides nucléiques comme les R-loops et les G-quadruplexes (G4) (Polenkowski et al., 2023; Villarreal et al., 2020; Wu et al., 2019). Ces structures secondaires peuvent à la fois posséder des rôles pro-cancéreux mais aussi constituer des cibles thérapeutiques (Wells et al., 2019; R. Zhang et al., 2023). Les R-loops et G4 participent eux-mêmes à l'instabilité génomique (Castillo-Guzman and Chédin, 2021; Crossley et al., 2019; Linke et al., 2021). DDX5 participe ainsi à la réparation des cassures double brins (DSB) en déroulant les R-loops (Yu et al., 2020). À noter tout de même que la formation de R-loops semblerait pouvoir participer au processus de réparation des cassures d'ADN double brin. En effet, la présence d'hybrides ADN:ARN a été détectée autour des DSBs et il a été démontré qu'ils sont nécessaires à une réparation efficace (Bader and Bushell, 2020; Chakraborty et al., 2016; Lu et al., 2018). De récents travaux démontrent ainsi que DDX17 permet la formation de ces hybrides ADN:ARN au niveau des DSBs présents dans des loci déficients en hybrides (Bader et al., 2022). Cette activité pourrait expliquer que DDX17 favorise la réparation des DSBs. Cela n'a pas été démontré, mais il est possible que la présence de R-loops soit dans un premier temps nécessaire lors des premières étapes de la réparation des DSBs, puis qu'il faille dans un deuxième temps les résoudre pour permettre le bon déroulement des dernières étapes de la réparation des DSBs. On peut ajouter que les R-loops jouant aussi un rôle dans la maturation en 3' et la terminaison transcriptionnelle (Sanz et al., 2016; Skourti-Stathaki et al., 2011; Yang et al., 2019; Yanling Zhao et al., 2016), le rôle de DDX5 et DD17 dans la régulation de la maturation de l'extrémité 3' des ARNm et la stabilité du génome ne sont peut-être pas à considérer de façon complètement indépendante. Plusieurs travaux indiquent de plus l'existence de liens

entre la maturation de l'extrémité 3' des ARNm et la stabilité génomique ([Dutertre et al., 2021](#)).

Il faut noter que le neuroblastome est un cancer qui se caractérise notamment par d'importants réarrangements génomiques. Les altérations les plus fréquentes ne sont pas des mutations ponctuelles de gènes mais l'amplification de *MYCN* (qui peut atteindre une centaine de copies) et des anomalies chromosomiques segmentaires comme la délétion des bras chromosomiques 1p et 11q, ainsi que le gain de 17q ([Bown, 2001](#)). Un contexte favorisant les cassures doubles brin ou une altération des mécanismes de réparation pourrait favoriser la survenue de ces altérations génomiques lors des évènements de recombinaison ([Alves et al., 2017](#)).

Il faut préciser ici qu'un papier publié en 2020 rapporte un rôle plutôt pro-tumoral à DDX5 dans le neuroblastome. Les auteurs de cette étude associent notamment une forte expression de DDX5 à une diminution de la survie à 5 ans des patients ([Zhao et al., 2020](#)). Cependant, la cohorte analysée ne comporte que 42 patients, à comparer à l'analyse du RNA-seq d'une cohorte de 498 patients utilisée dans notre étude présentant une association entre une meilleure survie des patients et une forte expression de DDX5.

#### Mécanisme de régulation de la production de transcrits chimériques par DDX5 et DDX17

Le mécanisme par lequel DDX5 et DDX17 régulent la maturation de l'extrémité 3' des transcrits n'a pas été l'objet de mes travaux de thèse. Le mécanisme par lequel DDX5 et DDX17 permettent de limiter le readthrough transcriptionnel entraînant la production de transcrits allongés a été exposé dans un papier précédent de l'équipe ([Terrone et al., 2022](#)). Dans ce modèle, DDX5 et DDX17 régulent la terminaison transcriptionnelle en favorisant la formation de boucles intragéniques dépendantes de CTCF. Il est probable qu'un mécanisme similaire soit à l'œuvre dans le contrôle de la production de transcrits chimériques. Néanmoins, il pourrait être intéressant pour compléter ces travaux de tester *in vitro* l'effet d'une déplétion de DDX5 et DDX17 sur l'activité de clivage du complexe CPA.

Pour comparer l'effet du readthrough induit par la déplétion des hélicases avec celui induit par l'induction de stress ou d'infection viral, on pourrait aussi examiner les modifications de la chromatine au niveau des gènes présentant un readthrough DDX5/17

dépendant avant et après déplétion. Il pourrait aussi être pertinent de réaliser ou d'analyser des données de CLIP-seq ou de CHIP-seq afin d'observer si DDX5 et DDX17 se fixent directement au niveau de la région 3' des gènes et des transcrits présentant un readthrough suite à leur déplétion.

Par ailleurs, les hélicases DDX5 et DDX17 étant des régulateurs des structures secondaires des acides nucléiques comme les G4 et les R-loop (Gemmill et al., 2023; Polenkowski et al., 2023; Wu et al., 2019), il serait intéressant de tester l'effet de leur déplétion sur la formation de ces structures à proximité du PAS des gènes présentant un readthrough transcriptionnel.

#### Effet d'une variation d'expression de MYCN sur la terminaison transcriptionnelle

L'analyse de l'effet d'une déplétion ou surexpression de MYCN permet d'observer, respectivement, une diminution ou une augmentation de plusieurs transcrits chimériques, ainsi que de plusieurs transcrits allongés. L'expression de MYCN n'apparaît pas affecter significativement l'expression de DDX5 et DDX17, suggérant que l'effet de MYCN n'est pas dû à un contrôle transcriptionnel de l'expression des hélicases. L'analyse de données de CHIP-seq révèle par contre que MYCN est significativement enrichie à proximité des exons terminaux des gènes subissant un readthrough transcriptionnel en l'absence de DDX5/17. Ces résultats sont confirmés au niveau de plusieurs gènes par la réalisation de CHIP-qPCR dans nos cellules. Ces données peuvent surprendre du fait du rôle de facteur de transcription de MYCN, qui se fixe normalement au niveau du promoteur des gènes, en reconnaissant les séquences E-box auxquelles il se lie en hétérodimère avec MAX (myc-associated factor X) (Huang and Weiss, 2013). Ces résultats peuvent suggérer que la fixation ectopique de MYCN, lorsque son gène est surexprimé ou amplifié (comme dans les tumeurs de neuroblastome à haut risque) au niveau de site de terminaison transcriptionnelle des gènes, pourrait interférer avec la régulation de la maturation de l'extrémité 3' de leurs transcrits.

De plus, le recrutement d'une dCas9-MYCN au niveau du site de terminaison transcriptionnelle du gène *MAPK7*, produisant le transcrit chimérique *MAPK7-RNF112* lorsque DDX5 et DDX17 sont déplétées ou MYCN surexprimée, induit une augmentation de l'expression de la forme chimérique. Cette expérience ciblée sur un gène constitue une preuve de concept que le recrutement de MYCN en 3' du gène suffit à induire un



readthrough et la production de transcrits chimériques. Le recrutement de dCas9-CT n'induit pas d'augmentation significative de la production de transcrits chimériques, comparé aux cellules transfectées avec la dCas9-CT sans guides, suggérant que le recrutement de MYCN permet d'augmenter la production de transcrits chimériques et que cet effet ne dépend pas seulement du recrutement de la dCas9. En effet il a été récemment montré que le recrutement de dCas9 en aval d'un gène induit la pause de la Pol II et favorise la terminaison transcriptionnelle, permettant de réduire le readthrough induit par le stress ([Zukher et al., 2023](#)).

Le mécanisme par lequel le recrutement de MYCN pourrait favoriser le readthrough transcriptionnel reste assez éluusif. MYCN co-immunoprécipite avec DDX17, cela de manière acide nucléique indépendante. Cela indique que MYCN et DDX17 pourraient se trouver dans les mêmes complexes protéiques dans les cellules, et que le maintien de ces complexes ne dépend pas de la présence d'ARN ou d'ADN. Le fait que MYCN interagisse spécifiquement avec DDX17 et non pas avec DDX5 semble indiquer que MYCN interagit soit directement avec DDX17, soit avec une protéine se fixant préférentiellement à DDX17 plutôt que DDX5. La présence de DDX17 et MYCN dans les mêmes complexes suggère que MYCN pourrait directement perturber la régulation de la maturation en 3' des transcrits par cette hélicase. Il pourrait être intéressant d'identifier le domaine permettant l'interaction de MYCN avec DDX17 et voir si en empêchant cette interaction via une mutation de ce domaine par exemple, il est possible d'entraver le readthrough induit par l'expression de MYCN.

L'analyse de données de CHIP-seq semble aussi indiquer que MYCN se fixe fréquemment à proximité des sites de liaison de CTCF. CTCF joue un rôle dans la sélection du PAS et la terminaison transcriptionnelle en favorisant la formation de boucles intergéniques ([Nanavaty et al., 2020](#); [Terrone et al., 2022](#)). Il est donc possible que la fixation de MYCN puisse interférer avec la fixation ou la fonction de CTCF. Pour l'instant, mes données préliminaires ne permettent pas d'observer une variation significative de la fixation de CTCF à la chromatine suite à la déplétion de MYCN (résultat non montré). Néanmoins, je n'ai examiné qu'un nombre restreint de gènes. On peut ajouter que si la fixation de MYCN à proximité des sites CTCF n'affecte pas forcément sa fixation à la chromatine, cela pourrait tout de même affecter ses fonctions, par exemple en limitant sa capacité à délimiter des boucles chromatinienne en affectant son interaction avec la cohésine. Il

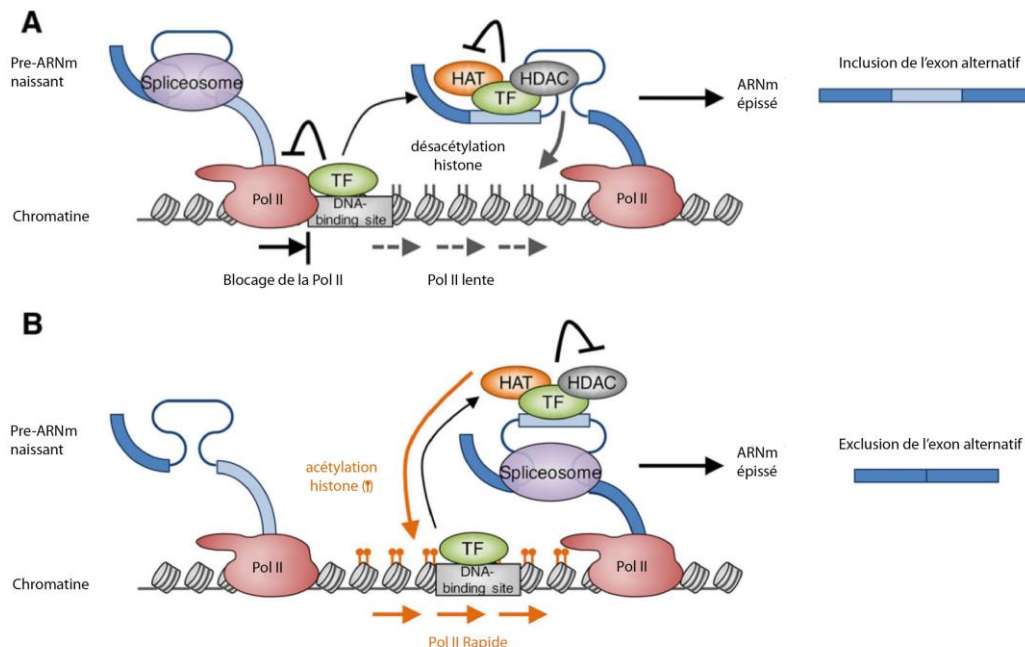
faut tout de même noter que des résultats récemment publiés suggèrent que MYC, un homologue de MYCN, favorise au contraire la fixation de CTCF (Wei et al., 2023). Les auteurs rapportent en effet que la surexpression de MYC entraîne l'émergence de 4 045 pics CTCF et la diminution de 6 676 pics CTCF. La surexpression de MYC semble ainsi augmenter légèrement plus la fixation de CTCF au niveau des sites communs de fixation de MYC et CTCF. On pourrait tout de même regretter qu'il n'y ait pas vraiment de vérification que la surexpression de MYC n'ait pas d'effet sur l'expression de CTCF. MYCN et MYC, malgré leur forte similarité, peuvent néanmoins posséder des fonctions distinctes (Westermann et al., 2008).

#### Régulation de la maturation des transcrits par un facteur de transcription

Comme indiqué précédemment, la fixation d'un facteur de transcription comme MYCN au niveau de l'extrémité 3' des gènes peut paraître atypique. Néanmoins, il ne s'agit pas du premier exemple de facteur de transcription se fixant au niveau du corps des gènes impactant la maturation des ARNm. Il y a bien sûr l'exemple de CTCF, dont la fixation à l'ADN est associée à la pause de la Pol II (Chernukhin et al., 2007; Kang and Lieberman, 2011; Shukla et al., 2011). Il a ainsi été montré que CTCF favorise l'inclusion de l'exon 5 du gène *CD45* en augmentant la pause de la Pol II. La fixation de CTCF module aussi la maturation en 3' des transcrits. Ainsi, la mutation du site CTCF en aval du gène *SH3TC1* entraîne une augmentation de son readthrough transcriptionnel, et la mutation du site CTCF en aval du locus *NFYA* entraîne une augmentation de la production de l'isoforme distale (Nanavaty et al., 2020; Terrone et al., 2022). Il n'est pas toujours clair si l'effet de CTCF passe par la formation d'une boucle chromatinienne, comme c'est le cas dans la régulation de l'APA du locus *NFYA* (Nanavaty et al., 2020), ou par d'autres mécanismes. MAZ et VEZF1 sont deux autres facteurs de transcription qui favorisent l'inclusion d'exons alternatifs en augmentant la pause de la Pol II au niveau de leur site de fixation (Gowher et al., 2012; Robson-Dixon and Garcia-Blanco, 2004). D'autres facteurs de

transcription modulent l'épissage via leur activité de remodelage de la chromatine ([Figure 22](#)).

(A) Les facteurs de transcription (TF) liés au corps du gène peuvent lier et bloquer la Pol II en aval d'un



**Figure 22 : Modèles de régulation de l'épissage par la modification de la cinétique de la Pol II induite par la liaison de facteur de transcription**

site d'épissage 3' faible pour promouvoir sa reconnaissance par le spliceosome et l'inclusion de l'exon alternatif dans l'ARNm mature. Alternativement ou simultanément, les TFs liés aux corps des gènes recrutent les HDAC ou inhibent les HAT afin d'augmenter la désacétylation des histones, de ralentir la Pol II et de promouvoir l'inclusion de l'exon. (B) Les TF liés au corps du gène peuvent « imprimer » les pré-ARNm naissants et recruter les HAT ou inhiber les HDAC afin d'augmenter l'acétylation des histones, d'augmenter les taux d'élongation de la Pol II et de promouvoir l'exclusion des exons. (Adapté de [Rambout et al., 2018](#))

C'est notamment le cas de DBIRD, qui favorise l'élongation locale de la transcription et le saut des exons riches en A/U qu'il lie et inhibe le corépresseur transcriptionnel histone désacétylase 3 (HDAC3) par l'intermédiaire de sa sous-unité DBC1 ([Close et al., 2012](#)).

Certains facteurs de transcription pourraient aussi moduler l'épissage en interagissant avec les facteurs d'épissage, comme WT1 qui pourrait inhiber la liaison à l'ARN du facteur d'épissage RBM4 en interagissant avec lui ([Markus et al., 2006](#)). On peut aussi citer le facteur de transcription NF- $\kappa$ B RELA, qui est capable de réguler l'épissage en recrutant DDX17 au niveau des exons sur le génome ([Ameur et al., 2020](#)).

Les exemples précédemment cités concernent majoritairement l'épissage, mais il existe aussi des éléments montrant que les facteurs de transcription peuvent réguler la maturation de l'extrémité 3' des ARN. L'activateur transcriptionnel GAL4-VP16 stimule la polyadénylation co-transcriptionnelle en recrutant PAF1c (RNA polymerase II-associated factor 1 homolog complex) (Nagaike et al., 2011), un complexe impliqué dans la régulation des différentes étapes transcriptionnelle de la Pol II qui stimulent le clivage et la polyadénylation des transcrits (Van Oss et al., 2017) et qui est par ailleurs nécessaire à la méthylation des lysines des histones, telles que H3-K4, -K36, -K79 (Krogan et al., 2003). GAL4-VP16 se fixe au niveau du promoteur du gène et non pas de son PAS, mais cela démontre qu'un facteur de transcription est susceptible de moduler le processus de clivage des transcrits.

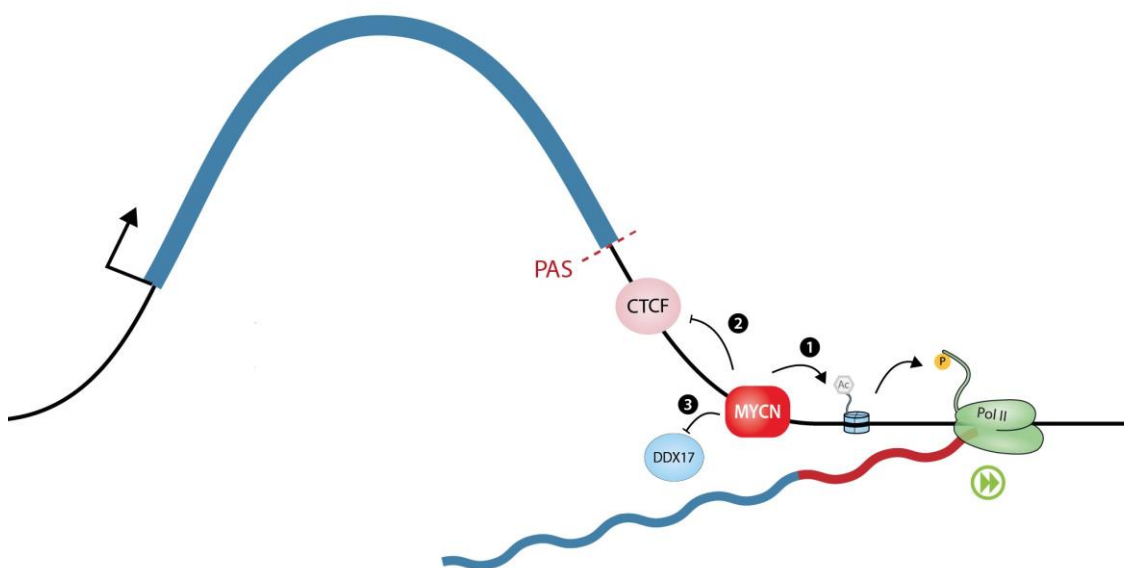
MYCN interagit lui-même avec plusieurs complexes de remodelage de la chromatine. Leur recrutement au niveau du PAS des gènes pourrait altérer le statut de la chromatine et affecter la cinétique de la Pol II. Les protéines de la famille MYC interagissent par exemple avec TRRAP (McMahon et al., 1998; Tu et al., 2015), un composant de plusieurs complexes multiprotéiques d'histones acétyltransférases (HAT). Le recrutement d'HAT pourrait affecter la maturation en 3' des transcrits par différents moyens. Certaines marques d'histones acétylées peuvent recruter le facteur d'élongation P-TEFb par le biais de BRD4. P-TEFb favorise l'élongation de la transcription en phosphorylant le CTD de la Pol II, principalement sur le résidu Ser2 (Fujinaga et al., 2023). La présence de P-TEFb au niveau du PAS pourrait donc interférer avec la pause de la Pol II normalement induite par la reconnaissance du PAS. La modification du statut de phosphorylation du CTD de la Pol II pourrait aussi perturber son interaction avec des facteurs du complexe CPA ainsi que d'autres facteurs protéiques modulant la maturation de l'extrémité 3'.

MYCN peut également interagir avec WDR5 (Sun et al., 2015), qui est une sous-unité du complexe histone H3K4 méthyltransférase et avec la H3K9me3/me2 déméthylase KDM4B (Yang et al., 2015), qui lui permettent d'activer la transcription. Sa déplétion entraîne la perte des marques euchromatiniennes H3K4me3 et H3K9ac (Cotterman et al., 2008).

Par ailleurs, comme indiqué en 6.5.3, MYCN peut indirectement moduler la terminaison transcriptionnel promoteur-proximale au niveau de ses promoteurs cibles, il serait donc

intéressant de voir si les mêmes mécanismes peuvent être mis en jeu lors de la fixation de MYCN au niveau de l'extrémité 3' des gènes.

Il serait possible d'analyser l'effet d'une variation de l'expression de MYCN ou de son recrutement direct sur les modifications de la chromatine à proximité de la région 3' de gènes présentant un readthrough afin de vérifier si l'effet de MYCN peut effectivement passer par l'altération du statut de la chromatine. Il pourrait aussi être intéressant d'utiliser des protéines MYCN mutées dans leurs domaines d'interaction avec certains composants des complexes de remodelage de la chromatine.



**Figure 23 : Représentation des différents mécanismes par lesquels MYCN pourrait favoriser le readthrough transcriptionnel**

(1) MYCN pourrait favoriser le recrutement de complexes de remodelage de la chromatine entraînant le dépôt de marques favorisant l'élongation transcriptionnelle. (2) MYCN pourrait affecter le recrutement ou la fonction de CTCF. (3) MYCN pourrait altérer la fonction de DDX17 en le séquestrant hors de ses sites d'action par exemple.

### Conclusion partie 1

Dans l'ensemble, ces travaux montrent que le readthrough induit par la déplétion des hélicases DDX5 et DDX17 participe à l'induction de transcrits chimériques dans les cellules de neuroblastome et que la baisse de leur expression est associée à un pronostic plus défavorable. De plus, MYCN, un facteur de transcription dont l'amplification est un

facteur de mauvais pronostic dans le neuroblastome, favorise l'induction de readthrough transcriptionnel et de transcrits chimériques. En plus de réguler la transcription, MYCN pourrait donc aussi moduler la maturation de l'extrémité 3' des transcrits. Cet effet de MYCN, associé à une expression réduite de DDX17 et DDX5, pourrait au moins en partie expliquer la présence accrue de transcrits chimériques dans les tumeurs de neuroblastomes par rapport aux tissus sains. Ces travaux ouvrent des perspectives sur un possible rôle protecteur de DDX5 et DDX17 dans le neuroblastome, et la compréhension du mécanisme par lequel MYCN favorise le readthrough transcriptionnel appelle davantage de travaux dans le futur.

## **Partie DDX17 et développement neuronal**

### **Effet d'une baisse d'expression de DDX17 sur le développement neuronal et caractérisation de deux mutations :**

Sur la base de l'identification de patients présentant des troubles neuro-développementaux et porteurs de mutations hétérozygotes du gène *DDX17*, nous avons cherché à identifier les effets d'une perte de fonction de *DDX17* pouvant expliquer les phénotypes observés chez les patients. Les résultats obtenus suite à l'analyse de l'impact moléculaire et physiologique d'une baisse d'expression de *DDX17* à l'aide de plusieurs modèles semblent montrer que *DDX17* joue en effet un rôle important dans le développement du système nerveux. Les études menées chez la souris et le xénope par nos collaborateurs montrent en effet qu'une perte d'expression de *DDX17* entraîne une diminution de la croissance des axones *in vivo* dans les deux modèles. De plus, l'étude *in vitro* de neurones corticaux de souris montre aussi une réduction de la complexité axonale. Enfin, des tests comportementaux menés chez le xénope montrent une réduction de la mémoire de travail.

L'analyse de l'effet d'une déplétion de *DDX17* sur le transcriptome de cellules humaines SH-SY5Y, réalisée dans notre équipe, montre que celle-ci entraîne une altération de l'expression de 350 gènes. Ces gènes présentent un enrichissement significatif en fonctions impliquées notamment dans la neurogenèse, la morphogenèse, la communication inter-cellules et l'assemblage synaptique. Ces fonctions semblent cohérentes avec les défauts de développement du système nerveux observés dans les modèles animaux, mais aussi chez les patients présentant des mutations hétérozygotes de *DDX17*. Parmi les gènes réprimés en condition si*DDX17*, on peut ainsi citer *PLXNA4* et *PLXNA2*, qui codent respectivement pour les co-récepteurs de SEMA3A et SEMA6A et jouent tous les deux un rôle dans la guidance axonale ([Andermatt et al., 2014](#); [Suto et al., 2005](#)). *PLXNA4* serait aussi impliqué dans la morphogenèse des dendrites ([Danelon et al., 2020](#)). Chez la souris, le KO de *PLXNA2* entraîne une baisse de la prolifération dans le gyrus denté et de la neurogenèse adulte. Cela s'accompagne de changements comportementaux comme une réduction de la mémoire de la peur et la sociabilité ([Zhao et al., 2018](#)). Chez l'humain, une étude d'association pan-génomique a permis de lier la présence de variant de *PLXNA2* à la schizophrénie ([Mah et al., 2006](#)). On peut aussi citer le gène *RPH3A*, qui régule entre autres l'organisation, l'activité et la plasticité synaptique

(Deák et al., 2006; Franchini et al., 2022; Yang et al., 2023). La présence d'un variant de *RPH3A* a été associée à l'apparition de troubles neurologiques multi-systémiques chez une jeune patiente (Maselli et al., 2018).

Le but n'est pas de faire une liste commentée exhaustive des fonctions de tous les gènes dérégulés par la déplétion de *DDX17*, mais ces exemples illustrent comment l'altération du transcriptome suite à la baisse d'expression de *DDX17* pourrait expliquer les phénotypes observés dans les modèles animaux ainsi que chez les patients. Cette hypothèse serait confortée si nous pouvions identifier des variations de gènes similaires dans ces autres modèles, ce qui constitue un objectif futur majeur.

Il est par ailleurs intéressant de noter que les orthologues de *DDX17* présents chez *D. melanogaster* (CG10777) et *C. elegans* (*ddx-17*) ont tous les deux été identifiés lors de cribles génétiques comme entraînant des défauts de la morphologie dendritique suite à leur knock-down par ARNi (Antonacci et al., 2015; Olesnický et al., 2013). Ces résultats semblent en accord avec les défauts de croissance axonale constatés chez la souris et le xénope, et suggèrent que le rôle de *DDX17* dans la morphogénèse des neurites est conservé à travers l'évolution.

#### Lien entre le rôle de *DDX17* dans le neuroblastome et le neuro-développement

Il faut ajouter que bien que nous n'ayons pour le moment étudié l'effet d'une déplétion de *DDX17* que sur l'expression globale des gènes, il est aussi possible que *DDX17* affecte le développement neuronal par d'autres moyens. Comme nos travaux précédents l'ont montré, *DDX5* et *DDX17* régulent de nombreuses autres étapes de la maturation et du métabolisme des ARN, notamment l'épissage, la terminaison transcriptionnelle, ainsi que l'expression et la maturation des miARN (Lambert et al., 2018; Terrone et al., 2022). Les effets spécifiques à *DDX17* sont peu connus et étudiés, mais il est possible qu'une baisse de son expression puisse affecter l'épissage et la maturation en 3' des transcrits. Le contrôle de l'APA et de la longueur de la 3' UTR des transcrits est particulièrement crucial dans le développement et le fonctionnement des neurones. Lors de la différenciation neuronale, on observe ainsi un allongement de l'extrémité 3' UTR de nombreux transcrits (Govenek and Tian, 2018; Kiltchewskij et al., 2023). Les neurones ont tendance à exprimer des isoformes avec de longues 3' UTR comparés à d'autres



cellules présentes dans le cerveau comme la microglie ou les cellules endothéliales (Guvenek and Tian, 2018).

La longueur de l'extrémité 3'UTR permet aussi de réguler la localisation des transcrits dans les neurones, les isoformes longues étant préférentiellement exportées dans les neurites, alors que les isoformes courtes restent dans le corps cellulaire. La capacité des ARNm localisés au niveau axonal ou dendritique à subir une traduction locale est essentielle pour les fonctions synaptiques. La concentration synaptique des canaux ioniques, tels que Kv1.1 (*Kcna1*), est ainsi régulée par la localisation et la traduction locale de son ARNm, qui dépendent elles-mêmes de l'extrémité 3'UTR (Raab-Graham et al., 2006). De façon intéressante, un papier en pré-publication indique que l'analyse de la polyadénylation alternative utilisant le séquençage des ARNm sur cellules uniques (ScRNA-Seq) à partir de tissu cérébral *post-mortem* a révélé des 3' UTR majoritairement plus longues chez les personnes autistes par rapport aux individus sains (Göpferich et al., 2020).

Il est par ailleurs intéressant d'observer que parmi les gènes significativement surexprimés suite à la déplétion de DDX17 dans les SH-SY5Y se trouve *MYCN*. Ce résultat est d'autant plus intrigant que la co-déplétion de DDX5 et DDX17 dans les SK-N-BE(2) induit une diminution de l'expression de *MYCN* (non montré). DDX17 seulement et non DDX5 aurait ainsi un rôle négatif sur l'expression de *MYCN*, bien que l'on ignore si le mécanisme est direct ou non. On peut aussi remarquer qu'un variant gain de fonction de *MYCN* a été relié avec un phénotype de déficience intellectuelle chez une jeune patiente (Yu et al., 2018). Le rôle de *MYCN* dans le développement du système nerveux périphérique est bien connu, mais il possède aussi des fonctions dans le système nerveux central, même si elles ont moins été étudiées. *MYCN* est en effet exprimé dans le cerveau humain fœtal (Grady et al., 1987) et la mutation hétérozygote de *MYCN* provoque le syndrome de Feingold, caractérisé par une réduction de la taille du cerveau et des difficultés d'apprentissage (van Bokhoven et al., 2005). Il est donc possible que la variation d'expression de *MYCN* induite par la déplétion de DDX17 puisse participer aux phénotypes observés chez les modèles animaux et les patients.

## Caractérisation de deux mutations faux sens de DDX17

La caractérisation des mutants avait pour but de comprendre l'effet des mutations faux-sens sur le comportement et le fonctionnement de DDX17 et de voir à terme si leur effet peut expliquer le phénotype observé chez les patients. Le variant R304H est prédit pour présenter une diminution de sa capacité à lier l'ARN. Cette prédiction pourrait être vérifiée en testant la capacité de DDX17-R304H à se lier à l'ARN *in vitro* par EMSA (RNA electrophoretic mobility shift assay). Ni R304H ni P203L ne sont prédites par simulation pour affecter l'activité ATPase de DDX17 (résultat non-montré). Toutefois, les simulations utilisées ne prennent pas en compte la structure entière de l'hélicase, dont les extrémités N-ter et C-ter intrinsèquement désordonnées n'ont pas encore été résolues par cristallographie (Ngo et al., 2019). Cela pourrait donc tout de même être pertinent de vérifier expérimentalement l'effet de ces mutations sur l'activité ATPase de DDX17, mais aussi sur son activité hélicase.

Il pourrait être intéressant de tester plus globalement l'impact des mutations identifiées sur l'interactome ARN et protéique de DDX17, respectivement par des expériences de CLIP-seq et de spectrométrie de masse. Il faut toutefois noter que le caractère insoluble du variant P203L peut rendre certaines méthodes moins efficaces.

On peut par ailleurs remarquer que les résidus touchés par cinq des six mutations faux-sens identifiées chez les patients sont conservés chez les orthologues de *DDX17* jusqu'à *C. elegans* et *S. cerevisiae* (voir annexe) Cela suggère l'importance de ces cinq résidus dans la fonction de l'hélicase. Le seul résidu touché par une mutation faux-sens non conservé est celui présent dans la région spécifique à l'isoforme p82.

Les expériences d'immunofluorescence et de fractionnement nous ont permis d'observer que la mutation P203L entraîne une agrégation et une baisse de solubilité de DDX17. Il est probable que la formation de tels agrégats séquestrent DDX17 et l'empêchent d'exercer ses fonctions cellulaires. Certains partenaires de DDX17 pourraient peut-être aussi être séquestrés dans ces agrégats, bien que DDX5 par exemple, qui est un partenaire de DDX17, ne semble pas voir sa localisation modifiée dans nos expériences (résultat non montré). Mes données préliminaires (résultat non montré) semblent indiquer une colocalisation entre DDX17-P203L et HDCA6, une désacétylase jouant un rôle essentiel dans la réponse cellulaire aux protéines mal

repliées et agrégées, qui participe au recrutement des agrégats de protéines ubiquitinyliées dans l'agrésome (Boyault et al., 2007). L'aspect des agrégats de DDX17-P203L, qui est plutôt dissemblable à celui des agrésomes « classiques », pourrait en effet être similaire aux agrésomes « ribbon-like » décrits par Garcia-Mata et al. (Garcia-Mata et al., 2002). La formation d'agrésomes est normalement un mécanisme protecteur permettant l'élimination des protéines mal repliées formant des agrégats. Néanmoins, la persistance de ces agrésomes est susceptible d'être néfaste à la cellule. Un article rapporte ainsi que la persistance d'agrésomes est associée à une morphologie nucléaire anormale, à l'induction de cassures d'ADN double brin et à l'arrêt du cycle cellulaire dépendant de p53. Les agrésomes pourraient avoir un effet néfaste sur la mitose en interférant de manière stérique avec l'alignement des chromosomes, le positionnement des centrosomes et la formation du fuseau (Lu et al., 2015). Si les agrésomes formés par DDX17-P203L ne sont pas éliminés efficacement par la cellule, on peut donc imaginer un effet néfaste sur les cellules exprimant *DDX17* à des niveaux importants. Toutefois, nos expériences ayant été réalisées via une transfection de plasmides dans les cellules, il nous faut vérifier si DDX17-P203L forme aussi des agrésomes en condition physiologique, à des niveaux d'expression similaires à la protéine DDX17 endogène.

Un objectif important est d'identifier l'impact de ces mutations sur le transcriptome des cellules. Nous avons tenté de faire des expériences de « rescue » (sauvetage) en exprimant DDX17 WT ou mutées dans les cellules traitées par siDDX17, mais nous n'avons pas réussi à obtenir un effet de rétablissement suffisant avec la DDX17 WT pour nous permettre d'analyser l'effet des mutations. Il s'agit d'une des limites d'un système basé sur la transfection transitoire de plasmides codants pour les protéines. Nous espérons que l'utilisation de lignées cellulaires mutées sur le gène *DDX17* nous permettra de comparer plus efficacement l'impact de ces mutations sur les fonctions de *DDX17*. Afin de caractériser l'effet des mutations de *DDX17* sur la différenciation et le fonctionnement neuronal, l'objectif de l'équipe est de développer en collaboration des modèles de cellules souches pluripotentes humaines porteuses de mutations de *DDX17*, dont on pourrait par la suite induire la différenciation neuronale.

En parallèle, l'effet des mutations de *DDX17* sur le développement neuronal pourra être examiné chez la souris par nos collaborateurs, sur des neurones en culture et, à plus long

terme, il serait aussi envisageable de développer des modèles knock-out et knock-in de souris.

### Conclusion partie 2

Ces travaux présentent des résultats d'expériences fonctionnelles (réalisées par nos partenaires) et transcriptomiques permettant d'identifier *DDX17* comme un nouveau régulateur du développement neuronal. Ces résultats supportent l'hypothèse que les mutations monoalléliques de *DDX17* identifiées chez 11 patients soient une cause de leurs troubles neuro-développementaux. Il s'agit de plus de la première étude présentant un lien entre *DDX17* et une maladie génétique. La caractérisation des mutations faux-sens identifiées pourrait nous permettre de mettre à jour de nouvelles fonctions cellulaires et moléculaires de cette hélicase à ARN.

**Dans l'ensemble, mes travaux de thèse permettent de montrer que l'altération de l'expression des hélicases DDX5 et surtout DDX17 entraîne des modifications du transcriptome à différents niveaux dans les cellules neuronales. L'altération de leur expression semble aussi pouvoir être liée au développement de pathologies du développement du système nerveux, à la fois cancéreuses comme le neuroblastome mais aussi de troubles neuro-développementaux. Des travaux supplémentaires seront toutefois nécessaires pour déterminer comment les fonctions de ces hélicases sont impliquées dans le développement de ces pathologies.**

# Matériels et méthodes

## Immunofluorescence :

Les cellules HEKT293 sont transfectées avec pCDNA.4-p72-FLAG WT ou muté sur des lamelles de verre puis sont fixées pendant 20 min dans de la formaline. Elles sont ensuite perméabilisées avec dans du PBS 1X – 0,5 % Triton pendant 10 min puis bloquées dans du PBS-Tween 0,2% - BSA 5% 1h à RT. Par la suite elles sont incubées 1h avec l'anticorps primaire à RT, rincées au PBS, incubées 1 h avec un anticorps secondaire conjugué à un fluorochrome, et rincées avec du PBS. Les lames furent montées dans du ProLong™ Diamond avec DAPI (P36966). Les acquisitions furent réalisées avec le Confocal Zeiss LSM800 avec un objectif x63 dans de l'huile d'immersion.

## Fractionnement cellulaire :

Les cellules HEKT293 sont lavées et récoltées dans du PBS 1X puis centrifugées 3 min à 3000 rpm à 4°C. Après un deuxième lavage au PBS 1X, le culot de cellules est resuspendu dans 200 µL de tampon "Dautry" complété avec de l'EDTA (10 mM Tris-HCl pH 8, 140 mM NaCl, 1.5 mM MgCl<sub>2</sub>, 10 mM EDTA, 0.5%NP40) et incubé 10 min sur glace. Le lysat est centrifugé 5 min à 3000 rpm à 4°C, le surnageant correspondant à la fraction cytoplasmique est récupéré et le culot constitué de noyaux est lavé avec 200 µL de tampon "Dautry" + EDTA. Le culot est ensuite resuspendu dans 200 µL de tampon Dautry. Les fraction cytoplasmique et nucléaire sont ensuite soniquées pendant 5 minutes puis centrifugées. Le surnageant correspondant au fraction « soluble » est récupéré, et le culot correspondant à la fraction « insoluble » est resuspendu dans 10µl de tampon à haute force ionique (10 mM Tris-HCl pH 8, 420 mM NaCl, 3 mM MgCl<sub>2</sub>, 0.5% NP40).

## Fractionnement des fractions solubles et insolubles :

Les cellules HEKT293 sont lavées et récoltées dans du PBS 1X puis centrifugées 3 min à 3000 rpm à 4°C. Après un deuxième lavage au PBS 1X, le culot de cellules est resuspendu dans 200 µL de tampon NP40 (0.5% NP40, 10 mM Tris HCl pH 7.8, 10 mM EDTA, 150 mM NaCl, 1 × protease inhibitor cocktail) et incubé 10 min sur glace. Le lysat est ensuite soniqué au bioruptor 5 min, puis centrifugé 25 min à 13000 rpm à 4°C. le surnageant correspondant à la fraction soluble est récupéré et le culot est lavé avec 200 µL de tampon de lavage (50 mM Tris HCl pH 7.4, 150 mM NaCl), pour être ensuite resuspendu dans 200 µl de tampon de resolubilisation (50 mM Tris HCl pH 6.8, 5% SDS, 10% glycerol) et soniqué pendant 5 min puis centrifugé 10 min à 13000 rpm à RT. Le surnageant correspondant à la fraction insoluble est ensuite collectée.

# Annexes



Q9XUW5_CAEEL	GLKMTCLFGGASKGFPQARDLERGVDIVVATPGRLLDFLDNGTTNMKKCSYLVLDEADRML	289
DBP2_YEAST	RIRNTCVYGGVPKSQQIRDLRSRGSEIVIATPGRLIDMLEIGKTNLKRVTYLVLDEADRML	273
Q8MZI3_DROME	HVRNTCIFGGAPKQQARDLERGVEIVIATPGRLIDFLERGTSLKRCTYLVLDEADRML	317
F7B0D7_XENTR	RLKSTCIYGGAPKGFQIRDLERGVEICIAATPGRLIDFLEAGKTNLRRCTYLVLDEADRML	242
A0A5H1ZRQ2_HUMAN	RLKSTCIYGGAPKGFQIRDLERGVEICIAATPGRLIDFLESGKTNLRRCTYLVLDEADRML	331
Q3U741_MOUSE	RLKSTCIYGGAPKGFQIRDLERGVEICIAATPGRLIDFLESGKTNLRRCTYLVLDEADRML	252

:: \*\*:\*\*:\*\* . \* . \* \*\* . \*\* : \* :\*\*\*\*\*.\*\*: \* .\*.::: :\*\*\*\*\*  
Ib
Ic R304H
II

Q9XUW5_CAEEL	DMGFEPQIKKIIGQIRPDRQTLMFSATWPKEVRALASDFQKDA AFLNVGSLELAANHNT	349
DBP2_YEAST	DMGFEPQIRKIVDQIRPDRQTLMWSATWPKEVKQLAADYLNDPIQVQVGSLELSASHNIT	333
Q8MZI3_DROME	DMGFEPQIRKIMQQIRPDRQVLMWSATWPKEVRQLAEFLNNYIQVNIIGSLSLSANHNIL	377
F7B0D7_XENTR	DMGFEPQIRKIVDQIRPDRQTLMWSATWPKEVRQLAEDFLRDYVQINIGNLELSANHNIL	302
A0A5H1ZRQ2_HUMAN	DMGFEPQIRKIVDQIRPDRQTLMWSATWPKEVRQLAEDFLRDYTQINVGNLELSANHNIL	391
Q3U741_MOUSE	DMGFEPQIRKIVDQIRPDRQTLMWSATWPKEVRQLAEDFLRDYTQINVGNLELSANHNIL	312

\*\*\*\*\*:\*\*: \*\*\*\*\*.\*\*:\*\*\*\*\*: \*\* :: .: :\*:\*.\*\*:\*.\*\*\*  
III

Q9XUW5_CAEEL	QVVDILEEHAKQAKMELLNHIMNQKECKTII FVETKRRKADELTRAMRRDGPWPTLCIHGD	409
DBP2_YEAST	QIVEVVSDFEKDRDLNKYLETASQDNEYKTLIFASTKRMCCDDITKYLREDGWPALAIHGD	393
Q8MZI3_DROME	QIVDVCENEKMLKLIKLLTDISAENETKTIIFVETKRRVDEITRNI SRQWRACAIHGD	437
F7B0D7_XENTR	QIVDVCQESEKDHKLIQLMEEIMAEKENKTIIFVETKRRCCDLTRMRRDGPWAMCIHGD	362
A0A5H1ZRQ2_HUMAN	QIVDVCMESEKDHKLIQLMEEIMAEKENKTIIFVETKRRCCDLTRMRRDGPWAMCIHGD	451
Q3U741_MOUSE	QIVDVCMESEKDHKLIQLMEEIMAEKENKTIIFVETKRRCCDLTRMRRDGPWAMCIHGD	372

\*:\*: : \* : \* : : :\*: \*\*.\*\*\*.\*\*\*: \*:\*: : .: \*\* : .\*\*\*\*  
IV
IVa

Q9XUW5_CAEEL	KNQGERDWVLQEFKAGKTPIMLATDVAARGLDVDDIKFVINYDYPNNSSEDYVHRIGRTGR	469
DBP2_YEAST	KDQERERDWVLQEFRNGRSPIMVATDVAARGIDVKGINVINYDMPGNIEDYVHRIGRTGR	453
Q8MZI3_DROME	KSQQERDFVLSFRNGRSHILVATDVAARGLDVDDVKFVINYDYPNNSSEDYVHRIGRTGR	497
F7B0D7_XENTR	KSQQERDWVLCEFRGTGKAPILIATDVASRGLDVEDIKFVINYDYPNNSSEDYVHRIGRTAR	422
A0A5H1ZRQ2_HUMAN	KSQPERDWVLNEFRSGKAPILIATDVASRGLDVEDVKFVINYDYPNNSSEDYVHRIGRTAR	511
Q3U741_MOUSE	KSQPERDWVLNEFRSGKAPILIATDVASRGLDVEDVKFVINYDYPNNSSEDYVHRIGRTAR	432

\*. \* \*\*\*:\*\* .\*: \* : \*:\*\*\*\*\*:\*\*\*.\*\*\*.:::\*\*\*\*\* \*.. \*\*\*\*\*.\*  
Q454R

Q9XUW5_CAEEL	SDKKGTAYTFFTHTNASKAKDLLKVLDEAKQTVPQALRDMANRS-YGGNSRGRYGGGGF	528
DBP2_YEAST	AGATGTAISFFTEQNKGLGAKLISIMREANQNIPELLKYDRRS-YGGGH--PRYGGGRG	510
Q8MZI3_DROME	SNNTGTAYTLFTHSNANKANDLIQVLEANQTINPKLMNAMNGGYNKRG--GGMGGGYR	555
F7B0D7_XENTR	STNKGTAITFFTPGNLQARELVKVEEANQTINPKLMQLVDHGR-----GGGG---G	472
A0A5H1ZRQ2_HUMAN	STNKGTAITFFTPGNLQARELIKVLEEANQAINPKLMQLVDHRG-----GGGGGGKG	564
Q3U741_MOUSE	STNKGTAITFFTPGNLQARELIKVLEEANQAINPKLMQLVDHRG-----GGGGGGKG	485

: .\*\*\* ::\*\* \* . .\*:::: \*\*:\* : \* . . \*





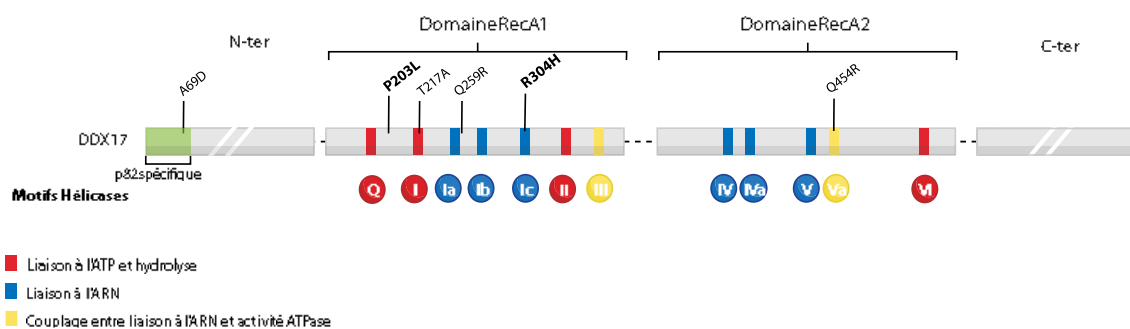
## Projet splicing lore

Splicing Lore est une interface web, liée à une base de données qui regroupe 160 jeux de données publiques (GEO Omnibus ou ENCODE) issues de séquençages d'ARN réalisés dans une dizaine de lignées cellulaires différentes, dans lesquelles l'expression de 75 facteurs d'épissage différents a été inhibée par traitement siARN ou shARN.

Le but de cet outil est de prédire, à partir d'une liste d'exons d'intérêt, si tout ou partie de ces exons est régulée par l'un des facteurs d'épissage analysés. Un score de fiabilité a été mis en place pour faciliter aux utilisateurs la recherche et l'étude des facteurs d'épissage candidats et de leurs exons d'intérêts. Ce score compris entre 0 et 1 est lié à un test statistique de permutation, qui donne aussi une p-value. Plus le score est proche de 1 plus, le facteur d'épissage est corrélé ou anti-corrélé à notre liste d'exons.

Ma contribution à ce projet a été de valider par RT-PCR la co-régulation de plusieurs exons par DDX5/17 et HNRNPC dans la lignée cellulaire SHSY5Y.

En parallèle avec l'étude de l'effet d'une déplétion de DDX17 sur le transcriptome, nous avons aussi cherché à caractériser moléculairement les mutants DDX17. Ces travaux ne font pas encore l'objet d'un article. Nous nous sommes concentrés sur deux mutations identifiées chez deux patients différents, correspondant aux variants Arg304His et Pro203Leu (Figure 20). Ces deux mutations ont été sélectionnées car il s'agit de mutations faux-sens, produisant des protéines mutées, permettant ainsi d'analyser l'effet de ces mutations sur le comportement de l'hélicase. Les deux mutations se trouvent dans le domaine RecA1 de DDX17. Plus précisément, la mutation R304H touche le domaine Ic (motif «TPGRLID » parfois désigné aussi comme le domaine Ib), qui est impliqué dans la liaison à l'ARN, alors que la mutation P203L touche la région se trouvant entre le motif Q et le domaine ATPase I. Une simulation de la capacité de DDX17 à fixer l'ARN, basée sur la seule structure 3D connue de DDX17 (Ngo et al., 2019) et réalisée par Stéphane Réty, chercheur au LBMC, prédit d'ailleurs que le mutant DDX17-R304H fixe moins efficacement l'ARN que DDX17-WT (Figure 21A).



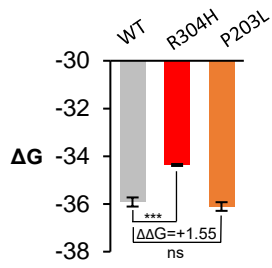
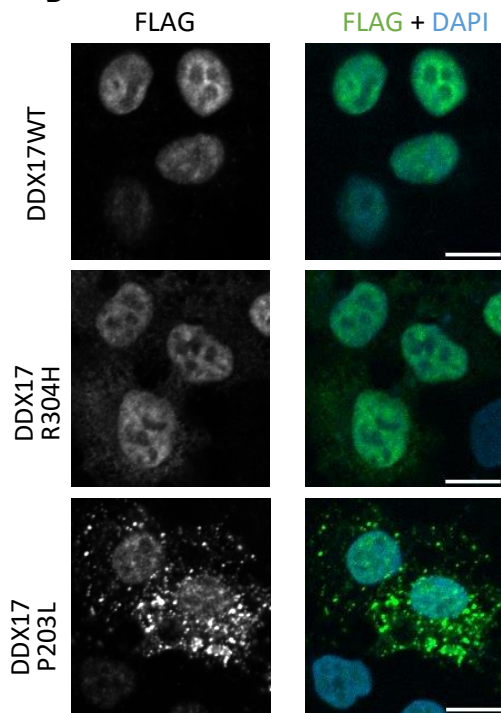
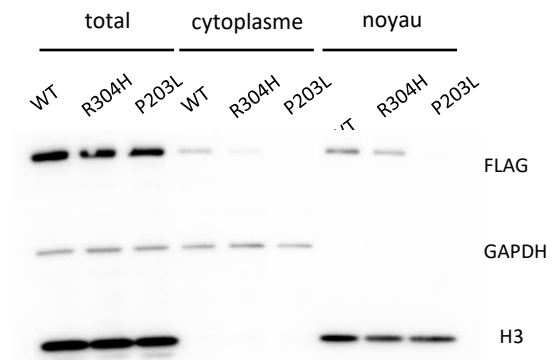
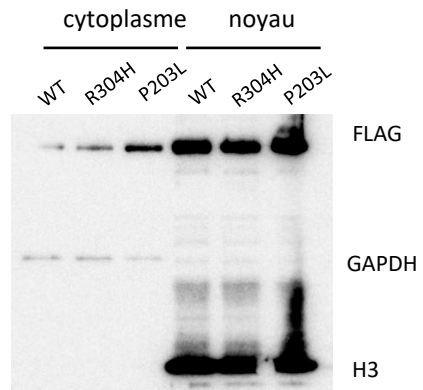
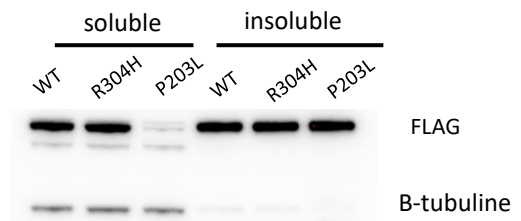
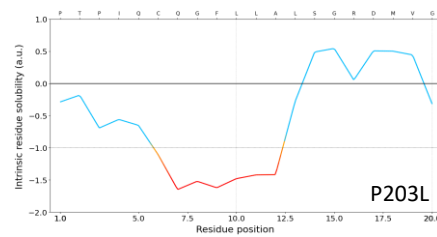
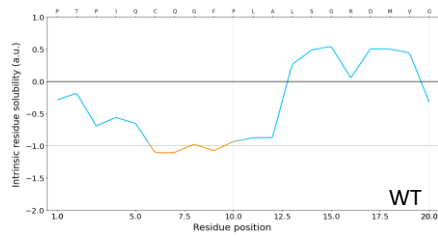
**Figure 20 : Localisation des substitutions d'acides aminés dans DDX17**

Représentation schématique de la localisation des 6 substitutions résultant des mutations faux-sens identifiées chez 6 des 11 patients identifiés. Les 5 autres variants conduisent à l'apparition d'un codon stop prématurés entraînant une perte de fonction. Les variants en gras correspondent à ceux étudiés lors de mes travaux

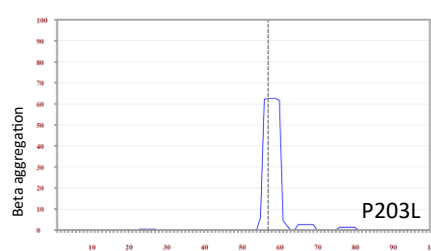
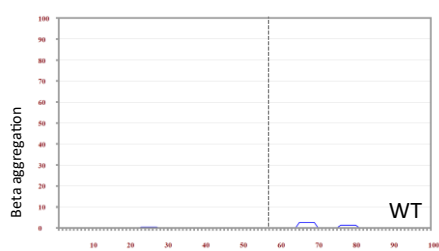
L'analyse de la localisation des protéines DDX17 mutées par immunofluorescence ne montre pas de différence significative entre DDX17R304H et DDX17WT, qui sont localisées essentiellement dans les noyaux des cellules transfectées, même si on détecte

un peu plus de protéine R304H dans le cytoplasme. Dans le cas de la protéine P203L par contre, on observe un comportement radicalement différent puisque ces protéines mutées semblent en effet s'agréger en foci, majoritairement cytoplasmiques mais aussi nucléaires (Figure 21B). Une analyse par western blot des fractions cytoplasmiques et nucléaires de cellules transfectées extraites avec un tampon peu stringent permet de confirmer la localisation majoritaire de DDX17 WT et R304H dans le noyau (Figure 21C). On observe par contre une quantité très faible de P203L dans les deux fractions. La protéine P203L étant bien détectée dans la fraction totale, préparée dans des conditions plus stringentes, nous avons fait l'hypothèse que la perte de P203L lors du fractionnement pouvait être une conséquence de l'agrégation observée en immunofluorescence et d'une solubilité plus faible de cette protéine. Pour valider cette hypothèse, j'ai récupéré le culot des deux fractions après l'étape de centrifugation et je les ai ré-extrait avec un tampon de lyse stringent. Dans ces conditions, on observe à nouveau la protéine P203L dans les fractions cytoplasmiques et nucléaires (Figure 21D). Par ailleurs, l'analyse par western blot de fractions insoluble et soluble montre que la protéine P203L est quasiment absente de la fraction soluble mais se trouve bien dans la fraction insoluble (Figure 21E). On peut noter que, de façon surprenante, si les protéines DDX17 WT et R304H se trouvent bien dans la fraction soluble, elles sont aussi détectées dans la fraction insoluble, à un niveau quasiment identique à P203L.

Ces résultats suggèrent que le variant P203L entraîne une baisse de la solubilité de DDX17 et une forte augmentation de la propension de DDX17 à former des agrégats. Ces résultats expérimentaux sont par ailleurs confirmés par des prédictions informatiques de la capacité des séquences peptidiques à s'agréger ainsi que de leur solubilité (Figure 21F-G).

**A****B****C****D****E****F**

cohsoftware

**G**

Tango.crg

**Figure 21 : La mutation DDX17P203L entraîne une baisse de solubilité de la protéine DDX17 ainsi que son agrégation sous forme de foci cytoplasmiques et nucléaires**

(A) Prédiction de l'affinité de DDX17 WT et mutée pour l'ARNdb. (B) Les cellules HEK293T transfectées avec un plasmide exprimant une DDX17-FLAG sauvage ou mutée ont été immuno-marquées avec un anticorps anti-FLAG. Les noyaux ont été marqués au DAPI. Les barres d'échelle correspondent à 10  $\mu$ m. (C) Western blot du fractionnement des fractions cytoplasmiques et nucléaires de cellules HEK293T exprimant DDX17-FLAG sauvage ou mutée. La présence de histone H3 et la GAPDH permet de contrôler la qualité du fractionnement. (D) Western blot correspondant au culot (fraction insoluble) des fractions nucléaires et cytoplasmiques. (E) Western blot des fractions solubles et insolubles de cellules HEK293T exprimant DDX17-FLAG sauvage ou mutée. (F-G) Prédiction de la capacité d'agrégation et d'insolubilité de DDX17WT et DDX17P203L avec (F) CamSol ([Sormanni et al., 2015](#)) (acides aminés 115 à 134) (G) Tango ([Fernandez-Escamilla et al., 2004](#)) (acides aminés 68 à 167). La barre verticale grise indique l'acide aminé 203.

# Discussion

## **Régulation de la production de transcrits chimériques par les hélicases DDX5 et DDX17 et par MYCN dans le neuroblastome**

### **Impact du readthrough transcriptionnel induit par la déplétion de DDX5 et DDX17 dans les cellules de neuroblastome :**

La perturbation de la maturation de l'extrémité 3' des transcrits peut conduire à un readthrough transcriptionnel. Le readthrough transcriptionnel est notamment observé suite à l'induction d'un stress cellulaire, d'une infection virale et dans certains cancers. Les conséquences du readthrough et les possibles fonctions des transcrits alternatifs dont il entraîne la production restent mal caractérisées. Le readthrough transcriptionnel peut ainsi entraîner la production de transcrits chimériques, résultant de la transcription de deux gènes (voire plus) sur un même brin d'ARN.

L'analyse du RNA-seq de cellules SH-SY5Y traitées par siDDX5/17 nous a permis d'identifier 218 transcrits chimériques dont le taux est augmenté. 11 de ces transcrits ont pu être validés par RT-qPCR, confirmant qu'il ne s'agit pas du résultat d'un biais de séquençage. Parmi les 91 transcrits chimériques les plus significativement induits, 25 (27%) sont communs à une liste de transcrits chimériques identifiés dans des tumeurs de patients souffrant de neuroblastome.

Certains de ces transcrits chimériques sont prédits pour respecter le cadre de lecture et pouvoir produire des protéines chimériques. Parmi les transcrits validés par RT-qPCR, *ADCK4-NUMBL*, *CHCHD10-VPREB3*, *CTSD-IFITM10* et *RHBDD2-POR*, soit quatre transcrits sur onze, ont une ORF chimérique. Il a été possible de confirmer l'expression de la protéine chimérique CTSD-IFITM10 dans les cellules traitées par siDDX5/17, confirmant que le transcrit chimérique peut bien être traduit en protéine chimérique.

Un des paramètres limitant notre caractérisation de transcrits chimériques produit par la déplétion des deux hélicases est que l'on ne connaît pas leur profil exact d'épissage. Plusieurs articles ont rapporté que les gènes présentant un readthrough transcriptionnel présentaient aussi des défauts d'épissage (voir 4.2.3). La technique de séquençage short-read ne permet par définition que l'assemblage de fragments courts et ne permet donc

pas d'avoir la séquence exacte complète d'un transcrit donné. Un séquençage dit « long-read » pourrait permettre d'accéder à cette information. Il faut toutefois noter qu'une des limites du séquençage « long-read » est son taux d'erreur plus élevé, ce qui peut poser problème lorsque l'on examine des événements rares comme les transcrits chimériques. Par exemple, les plateformes actuelles de séquençage « long-read », notamment Pacific Biosciences (PacBio) et Oxford Nanopore Technologies (ONT), présentaient précédemment un taux d'erreur moyen compris entre 5 et 20 % (Rang et al., 2018; Weirather et al., 2017). De récents progrès ainsi que l'application de méthodes de correction semblent néanmoins permettre d'augmenter significativement la fiabilité du séquençage « long-read » (Wenger et al., 2019). Les algorithmes d'identification de transcrits chimériques présentant déjà une faible reproductibilité de leurs résultats entre eux, il est important de faire usage d'une technique de séquençage hautement fiable, afin de ne pas encore augmenter les faux positifs ou événements non détectés.

Il faut aussi ajouter que les RNA-seq analysés furent établis à partir de bibliothèques enrichies en ARN polyA+, écartant les ARNs non-polyadénylés. Or certains articles rapportent que les transcrits readthrough, produits suite à l'induction d'un stress cellulaire par exemple, ne sont pas forcément polyadénylés (Nemeroff et al., 1998; Vilborg et al., 2015). Cela pourrait expliquer pourquoi dans nos expériences de fractionnement cellulaire suivi de RT-qPCR, les transcrits chimériques testés sont bien présents dans le cytoplasme (résultats non montrés), alors que de précédents papiers indiquent que les transcrits issus de readthrough tel les DoGs sont retenus à la chromatine (Vilborg et al., 2015). Il pourrait donc être judicieux de procéder à un séquençage non spécifique des ARNs, afin de vérifier si l'on n'identifie pas davantage de transcrits présentant des défauts d'épissage et de terminaison transcriptionnelle.

#### Rôle de DDX5/17 et du readthrough transcriptionnel dans le neuroblastome :

Étant donné l'association entre une faible expression des gènes *DDX17* et *DDX5* et un mauvais pronostic des patients, il pourrait être intéressant d'explorer plus en détail le lien entre *DDX5/17* et l'oncogenèse ainsi que la progression du neuroblastome. Cette association est d'autant plus intrigante que *DDX5* et *DDX17* sont généralement surexprimés dans les autres cancers analysés (voir 5.5.3). Même si cette association pourrait n'être qu'une corrélation, il resterait intéressant de voir si une baisse d'expression des hélicases peut affecter la réponse des patients aux traitements ou



constituer un biomarqueur permettant de déterminer avec plus de précision le profil des patients et le traitement thérapeutique adapté.

Le readthrough induit par une baisse d'expression de DDX5 et DDX17 pourrait moduler l'agressivité des tumeurs de neuroblastomes. L'expression de transcrits chimériques pourrait affecter certaines fonctions cellulaires, via la production de protéines chimériques ou de protéines tronquées. On peut rappeler l'exemple du transcrit chimérique *ZNF451-BAG2*, exprimé dans les tumeurs et les lignées cellulaires de neuroblastome, et dont l'expression affecte la différenciation induite par l'acide rétinoïque (Shi et al., 2021). Comme indiqué précédemment, 25 des transcrits chimériques induits suite à la déplétion de DDX5 et DDX17 ont aussi été identifiés dans des tumeurs de neuroblastome. De plus, parmi ces 25 transcrits, seulement 1 est aussi détecté par l'analyse de 161 échantillons de tissus sains de glandes surrénales. Parmi les 24 restants, 4 sont identifiés dans au moins 10% des 498 tumeurs de neuroblastome analysées. Des analyses supplémentaires seraient nécessaires afin de déterminer si ces transcrits chimériques possèdent des propriétés pro-tumorales ou peuvent influencer la réponse des tumeurs aux traitements, mais ils pourraient potentiellement servir de biomarqueurs ou permettre de mieux orienter la prise en charge des patients.

Le transcrit *CTSD-IFITM10*, dont nous avons pu valider la traduction en protéine chimérique dans nos cellules, a aussi été identifié dans des échantillons de tumeurs mammaires, et sa déplétion spécifique par traitement avec des siRNA ciblant la jonction chimérique induit une baisse de prolifération des cellules MCF7 (Varley et al., 2014). Il serait donc intéressant de tester l'effet de ce transcrit sur la prolifération dans les cellules de neuroblastome, en analysant l'effet de sa surexpression ou de sa déplétion ciblée par des siARN dirigés contre la jonction chimérique.

Nous souhaitons pouvoir accéder aux données RNA-seq de tumeurs de patients afin de pouvoir vérifier s'il est possible de retrouver un lien entre la production de transcrits chimériques et les niveaux d'expression des hélicases DDX5 et DDX17, ainsi que MYCN dans les tumeurs. Cela nous permettrait aussi de rechercher si certains transcrits chimériques sont associés à certaines caractéristiques des tumeurs et des patients, ainsi que le pronostic de ceux-ci. Nous n'avons pas encore réussi à obtenir l'accès à ces données mais espérons pouvoir mener ces analyses par la suite.

L'induction de readthrough pourrait aussi plus largement affecter les fonctions cellulaires en induisant la production d'isoformes à longue 3' UTR, pouvant lier des miARN et des RBP différentes pouvant moduler leur demi-vie, leur localisation cellulaire et leur traduction.

#### Possibles autres effets pro-tumoraux d'une déplétion de DDX5 et DDX17 dans le neuroblastome

En addition de leur effet sur la maturation en 3' des transcrits, DDX5 et DDX17 sont aussi impliqués dans la réparation de l'ADN (Bader et al., 2022; Fortuna et al., 2021; Sessa et al., 2021; Yu et al., 2020), ainsi que dans la régulation des structures secondaires des acides nucléiques comme les R-loops et les G-quadruplexes (G4) (Polenkowski et al., 2023; Villarreal et al., 2020; Wu et al., 2019). Ces structures secondaires peuvent à la fois posséder des rôles pro-cancéreux mais aussi constituer des cibles thérapeutiques (Wells et al., 2019; R. Zhang et al., 2023). Les R-loops et G4 participent eux-mêmes à l'instabilité génomique (Castillo-Guzman and Chédin, 2021; Crossley et al., 2019; Linke et al., 2021). DDX5 participe ainsi à la réparation des cassures double brins (DSB) en déroulant les R-loops (Yu et al., 2020). À noter tout de même que la formation de R-loops semblerait pouvoir participer au processus de réparation des cassures d'ADN double brin. En effet, la présence d'hybrides ADN:ARN a été détectée autour des DSBs et il a été démontré qu'ils sont nécessaires à une réparation efficace (Bader and Bushell, 2020; Chakraborty et al., 2016; Lu et al., 2018). De récents travaux démontrent ainsi que DDX17 permet la formation de ces hybrides ADN:ARN au niveau des DSBs présents dans des loci déficients en hybrides (Bader et al., 2022). Cette activité pourrait expliquer que DDX17 favorise la réparation des DSBs. Cela n'a pas été démontré, mais il est possible que la présence de R-loops soit dans un premier temps nécessaire lors des premières étapes de la réparation des DSBs, puis qu'il faille dans un deuxième temps les résoudre pour permettre le bon déroulement des dernières étapes de la réparation des DSBs. On peut ajouter que les R-loops jouant aussi un rôle dans la maturation en 3' et la terminaison transcriptionnelle (Sanz et al., 2016; Skourti-Stathaki et al., 2011; Yang et al., 2019; Yanling Zhao et al., 2016), le rôle de DDX5 et DDX17 dans la régulation de la maturation de l'extrémité 3' des ARNm et la stabilité du génome ne sont peut-être pas à considérer de façon complètement indépendante. Plusieurs travaux indiquent de plus l'existence de liens

entre la maturation de l'extrémité 3' des ARNm et la stabilité génomique ([Dutertre et al., 2021](#)).

Il faut noter que le neuroblastome est un cancer qui se caractérise notamment par d'importants réarrangements génomiques. Les altérations les plus fréquentes ne sont pas des mutations ponctuelles de gènes mais l'amplification de *MYCN* (qui peut atteindre une centaine de copies) et des anomalies chromosomiques segmentaires comme la délétion des bras chromosomiques 1p et 11q, ainsi que le gain de 17q ([Bown, 2001](#)). Un contexte favorisant les cassures doubles brin ou une altération des mécanismes de réparation pourrait favoriser la survenue de ces altérations génomiques lors des évènements de recombinaison ([Alves et al., 2017](#)).

Il faut préciser ici qu'un papier publié en 2020 rapporte un rôle plutôt pro-tumoral à DDX5 dans le neuroblastome. Les auteurs de cette étude associent notamment une forte expression de DDX5 à une diminution de la survie à 5 ans des patients ([Zhao et al., 2020](#)). Cependant, la cohorte analysée ne comporte que 42 patients, à comparer à l'analyse du RNA-seq d'une cohorte de 498 patients utilisée dans notre étude présentant une association entre une meilleure survie des patients et une forte expression de DDX5.

#### Mécanisme de régulation de la production de transcrits chimériques par DDX5 et DDX17

Le mécanisme par lequel DDX5 et DDX17 régulent la maturation de l'extrémité 3' des transcrits n'a pas été l'objet de mes travaux de thèse. Le mécanisme par lequel DDX5 et DDX17 permettent de limiter le readthrough transcriptionnel entraînant la production de transcrits allongés a été exposé dans un papier précédent de l'équipe ([Terrone et al., 2022](#)). Dans ce modèle, DDX5 et DDX17 régulent la terminaison transcriptionnelle en favorisant la formation de boucles intragéniques dépendantes de CTCF. Il est probable qu'un mécanisme similaire soit à l'œuvre dans le contrôle de la production de transcrits chimériques. Néanmoins, il pourrait être intéressant pour compléter ces travaux de tester *in vitro* l'effet d'une déplétion de DDX5 et DDX17 sur l'activité de clivage du complexe CPA.

Pour comparer l'effet du readthrough induit par la déplétion des hélicases avec celui induit par l'induction de stress ou d'infection viral, on pourrait aussi examiner les modifications de la chromatine au niveau des gènes présentant un readthrough DDX5/17

dépendant avant et après déplétion. Il pourrait aussi être pertinent de réaliser ou d'analyser des données de CLIP-seq ou de CHIP-seq afin d'observer si DDX5 et DDX17 se fixent directement au niveau de la région 3' des gènes et des transcrits présentant un readthrough suite à leur déplétion.

Par ailleurs, les hélicases DDX5 et DDX17 étant des régulateurs des structures secondaires des acides nucléiques comme les G4 et les R-loop (Gemmill et al., 2023; Polenkowski et al., 2023; Wu et al., 2019), il serait intéressant de tester l'effet de leur déplétion sur la formation de ces structures à proximité du PAS des gènes présentant un readthrough transcriptionnel.

#### Effet d'une variation d'expression de MYCN sur la terminaison transcriptionnelle

L'analyse de l'effet d'une déplétion ou surexpression de MYCN permet d'observer, respectivement, une diminution ou une augmentation de plusieurs transcrits chimériques, ainsi que de plusieurs transcrits allongés. L'expression de MYCN n'apparaît pas affecter significativement l'expression de DDX5 et DDX17, suggérant que l'effet de MYCN n'est pas dû à un contrôle transcriptionnel de l'expression des hélicases. L'analyse de données de CHIP-seq révèle par contre que MYCN est significativement enrichie à proximité des exons terminaux des gènes subissant un readthrough transcriptionnel en l'absence de DDX5/17. Ces résultats sont confirmés au niveau de plusieurs gènes par la réalisation de CHIP-qPCR dans nos cellules. Ces données peuvent surprendre du fait du rôle de facteur de transcription de MYCN, qui se fixe normalement au niveau du promoteur des gènes, en reconnaissant les séquences E-box auxquelles il se lie en hétérodimère avec MAX (myc-associated factor X) (Huang and Weiss, 2013). Ces résultats peuvent suggérer que la fixation ectopique de MYCN, lorsque son gène est surexprimé ou amplifié (comme dans les tumeurs de neuroblastome à haut risque) au niveau de site de terminaison transcriptionnelle des gènes, pourrait interférer avec la régulation de la maturation de l'extrémité 3' de leurs transcrits.

De plus, le recrutement d'une dCas9-MYCN au niveau du site de terminaison transcriptionnelle du gène *MAPK7*, produisant le transcrit chimérique *MAPK7-RNF112* lorsque DDX5 et DDX17 sont déplétées ou MYCN surexprimée, induit une augmentation de l'expression de la forme chimérique. Cette expérience ciblée sur un gène constitue une preuve de concept que le recrutement de MYCN en 3' du gène suffit à induire un

readthrough et la production de transcrits chimériques. Le recrutement de dCas9-CT n'induit pas d'augmentation significative de la production de transcrits chimériques, comparé aux cellules transfectées avec la dCas9-CT sans guides, suggérant que le recrutement de MYCN permet d'augmenter la production de transcrits chimériques et que cet effet ne dépend pas seulement du recrutement de la dCas9. En effet il a été récemment montré que le recrutement de dCas9 en aval d'un gène induit la pause de la Pol II et favorise la terminaison transcriptionnelle, permettant de réduire le readthrough induit par le stress ([Zukher et al., 2023](#)).

Le mécanisme par lequel le recrutement de MYCN pourrait favoriser le readthrough transcriptionnel reste assez éluusif. MYCN co-immunoprécipite avec DDX17, cela de manière acide nucléique indépendante. Cela indique que MYCN et DDX17 pourraient se trouver dans les mêmes complexes protéiques dans les cellules, et que le maintien de ces complexes ne dépend pas de la présence d'ARN ou d'ADN. Le fait que MYCN interagisse spécifiquement avec DDX17 et non pas avec DDX5 semble indiquer que MYCN interagit soit directement avec DDX17, soit avec une protéine se fixant préférentiellement à DDX17 plutôt que DDX5. La présence de DDX17 et MYCN dans les mêmes complexes suggère que MYCN pourrait directement perturber la régulation de la maturation en 3' des transcrits par cette hélicase. Il pourrait être intéressant d'identifier le domaine permettant l'interaction de MYCN avec DDX17 et voir si en empêchant cette interaction via une mutation de ce domaine par exemple, il est possible d'entraver le readthrough induit par l'expression de MYCN.

L'analyse de données de CHIP-seq semble aussi indiquer que MYCN se fixe fréquemment à proximité des sites de liaison de CTCF. CTCF joue un rôle dans la sélection du PAS et la terminaison transcriptionnelle en favorisant la formation de boucles intergéniques ([Nanavaty et al., 2020](#); [Terrone et al., 2022](#)). Il est donc possible que la fixation de MYCN puisse interférer avec la fixation ou la fonction de CTCF. Pour l'instant, mes données préliminaires ne permettent pas d'observer une variation significative de la fixation de CTCF à la chromatine suite à la déplétion de MYCN (résultat non montré). Néanmoins, je n'ai examiné qu'un nombre restreint de gènes. On peut ajouter que si la fixation de MYCN à proximité des sites CTCF n'affecte pas forcément sa fixation à la chromatine, cela pourrait tout de même affecter ses fonctions, par exemple en limitant sa capacité à délimiter des boucles chromatinienne en affectant son interaction avec la cohésine. Il

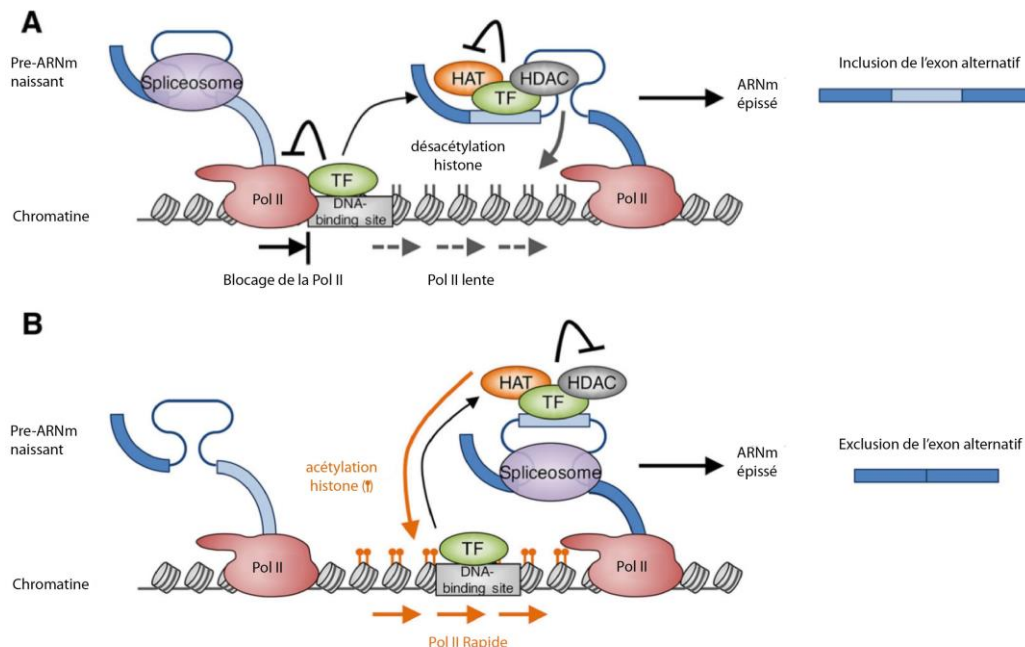
faut tout de même noter que des résultats récemment publiés suggèrent que MYC, un homologue de MYCN, favorise au contraire la fixation de CTCF (Wei et al., 2023). Les auteurs rapportent en effet que la surexpression de MYC entraîne l'émergence de 4 045 pics CTCF et la diminution de 6 676 pics CTCF. La surexpression de MYC semble ainsi augmenter légèrement plus la fixation de CTCF au niveau des sites communs de fixation de MYC et CTCF. On pourrait tout de même regretter qu'il n'y ait pas vraiment de vérification que la surexpression de MYC n'ait pas d'effet sur l'expression de CTCF. MYCN et MYC, malgré leur forte similarité, peuvent néanmoins posséder des fonctions distinctes (Westermann et al., 2008).

#### Régulation de la maturation des transcrits par un facteur de transcription

Comme indiqué précédemment, la fixation d'un facteur de transcription comme MYCN au niveau de l'extrémité 3' des gènes peut paraître atypique. Néanmoins, il ne s'agit pas du premier exemple de facteur de transcription se fixant au niveau du corps des gènes impactant la maturation des ARNm. Il y a bien sûr l'exemple de CTCF, dont la fixation à l'ADN est associée à la pause de la Pol II (Chernukhin et al., 2007; Kang and Lieberman, 2011; Shukla et al., 2011). Il a ainsi été montré que CTCF favorise l'inclusion de l'exon 5 du gène *CD45* en augmentant la pause de la Pol II. La fixation de CTCF module aussi la maturation en 3' des transcrits. Ainsi, la mutation du site CTCF en aval du gène *SH3TC1* entraîne une augmentation de son readthrough transcriptionnel, et la mutation du site CTCF en aval du locus *NFYA* entraîne une augmentation de la production de l'isoforme distale (Nanavaty et al., 2020; Terrone et al., 2022). Il n'est pas toujours clair si l'effet de CTCF passe par la formation d'une boucle chromatinienne, comme c'est le cas dans la régulation de l'APA du locus *NFYA* (Nanavaty et al., 2020), ou par d'autres mécanismes. MAZ et VEZF1 sont deux autres facteurs de transcription qui favorisent l'inclusion d'exons alternatifs en augmentant la pause de la Pol II au niveau de leur site de fixation (Gowher et al., 2012; Robson-Dixon and Garcia-Blanco, 2004). D'autres facteurs de

transcription modulent l'épissage via leur activité de remodelage de la chromatine (Figure 22).

(A) Les facteurs de transcription (TF) liés au corps du gène peuvent lier et bloquer la Pol II en aval d'un



**Figure 22 : Modèles de régulation de l'épissage par la modification de la cinétique de la Pol II induite par la liaison de facteur de transcription**

site d'épissage 3' faible pour promouvoir sa reconnaissance par le spliceosome et l'inclusion de l'exon alternatif dans l'ARNm mature. Alternativement ou simultanément, les TFs liés aux corps des gènes recrutent les HDAC ou inhibent les HAT afin d'augmenter la désacétylation des histones, de ralentir la Pol II et de promouvoir l'inclusion de l'exon. (B) Les TF liés au corps du gène peuvent « imprimer » les pré-ARNm naissants et recruter les HAT ou inhiber les HDAC afin d'augmenter l'acétylation des histones, d'augmenter les taux d'élongation de la Pol II et de promouvoir l'exclusion des exons. (Adapté de [Rambout et al., 2018](#))

C'est notamment le cas de DBIRD, qui favorise l'élongation locale de la transcription et le saut des exons riches en A/U qu'il lie et inhibe le corépresseur transcriptionnel histone désacétylase 3 (HDAC3) par l'intermédiaire de sa sous-unité DBC1 ([Close et al., 2012](#)).

Certains facteurs de transcription pourraient aussi moduler l'épissage en interagissant avec les facteurs d'épissage, comme WT1 qui pourrait inhiber la liaison à l'ARN du facteur d'épissage RBM4 en interagissant avec lui ([Markus et al., 2006](#)). On peut aussi citer le facteur de transcription NF- $\kappa$ B RELA, qui est capable de réguler l'épissage en recrutant DDX17 au niveau des exons sur le génome ([Ameur et al., 2020](#)).

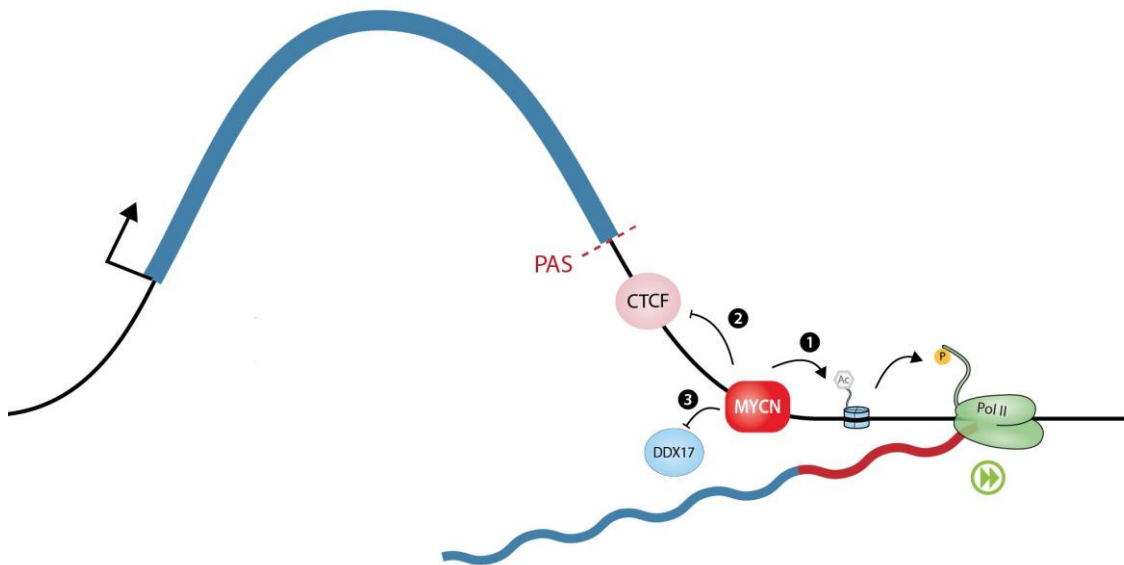
Les exemples précédemment cités concernent majoritairement l'épissage, mais il existe aussi des éléments montrant que les facteurs de transcription peuvent réguler la maturation de l'extrémité 3' des ARN. L'activateur transcriptionnel GAL4-VP16 stimule la polyadénylation co-transcriptionnelle en recrutant PAF1c (RNA polymerase II-associated factor 1 homolog complex) (Nagaike et al., 2011), un complexe impliqué dans la régulation des différentes étapes transcriptionnelle de la Pol II qui stimulent le clivage et la polyadénylation des transcrits (Van Oss et al., 2017) et qui est par ailleurs nécessaire à la méthylation des lysines des histones, telles que H3-K4, -K36, -K79 (Krogan et al., 2003). GAL4-VP16 se fixe au niveau du promoteur du gène et non pas de son PAS, mais cela démontre qu'un facteur de transcription est susceptible de moduler le processus de clivage des transcrits.

MYCN interagit lui-même avec plusieurs complexes de remodelage de la chromatine. Leur recrutement au niveau du PAS des gènes pourrait altérer le statut de la chromatine et affecter la cinétique de la Pol II. Les protéines de la famille MYC interagissent par exemple avec TRRAP (McMahon et al., 1998; Tu et al., 2015), un composant de plusieurs complexes multiprotéiques d'histones acétyltransférases (HAT). Le recrutement d'HAT pourrait affecter la maturation en 3' des transcrits par différents moyens. Certaines marques d'histones acétylées peuvent recruter le facteur d'élongation P-TEFb par le biais de BRD4. P-TEFb favorise l'élongation de la transcription en phosphorylant le CTD de la Pol II, principalement sur le résidu Ser2 (Fujinaga et al., 2023). La présence de P-TEFb au niveau du PAS pourrait donc interférer avec la pause de la Pol II normalement induite par la reconnaissance du PAS. La modification du statut de phosphorylation du CTD de la Pol II pourrait aussi perturber son interaction avec des facteurs du complexe CPA ainsi que d'autres facteurs protéiques modulant la maturation de l'extrémité 3'.

MYCN peut également interagir avec WDR5 (Sun et al., 2015), qui est une sous-unité du complexe histone H3K4 méthyltransférase et avec la H3K9me3/me2 déméthylase KDM4B (Yang et al., 2015), qui lui permettent d'activer la transcription. Sa déplétion entraîne la perte des marques euchromatiniennes H3K4me3 et H3K9ac (Cotterman et al., 2008).



Il serait possible d'analyser l'effet d'une variation de l'expression de MYCN ou de son recrutement direct sur les modifications de la chromatine à proximité de la région 3' de gènes présentant un readthrough afin de vérifier si l'effet de MYCN peut effectivement passer par l'altération du statut de la chromatine. Il pourrait aussi être intéressant d'utiliser des protéines MYCN mutées dans leurs domaines d'interaction avec certains composants des complexes de remodelage de la chromatine.



**Figure 23 : Représentation des différents mécanismes par lesquels MYCN pourrait favoriser le readthrough transcriptionnel**

(1) MYCN pourrait favoriser le recrutement de complexes de remodelage de la chromatine entraînant le dépôt de marques favorisant l'élongation transcriptionnelle. (2) MYCN pourrait affecter le recrutement ou la fonction de CTCF. (3) MYCN pourrait altérer la fonction de DDX17 en le séquestrant hors de ses sites d'action par exemple.

### Conclusion partie 1

Dans l'ensemble, ces travaux montrent que le readthrough induit par la déplétion des hélicases DDX5 et DDX17 participe à l'induction de transcrits chimériques dans les cellules de neuroblastome et que la baisse de leur expression est associée à un pronostic plus défavorable. De plus, MYCN, un facteur de transcription dont l'amplification est un facteur de mauvais pronostic dans le neuroblastome, favorise l'induction de readthrough transcriptionnel et de transcrits chimériques. En plus de réguler la transcription, MYCN

pourrait donc aussi moduler la maturation de l'extrémité 3' des transcrits. Cet effet de MYCN, associé à une expression réduite de DDX17 et DDX5, pourrait au moins en partie expliquer la présence accrue de transcrits chimériques dans les tumeurs de neuroblastomes par rapport aux tissus sains. Ces travaux ouvrent des perspectives sur un possible rôle protecteur de DDX5 et DDX17 dans le neuroblastome, et la compréhension du mécanisme par lequel MYCN favorise le readthrough transcriptionnel appelle davantage de travaux dans le futur.

## **Partie DDX17 et développement neuronal**

### **Effet d'une baisse d'expression de DDX17 sur le développement neuronal et caractérisation de deux mutations :**

Sur la base de l'identification de patients présentant des troubles neuro-développementaux et porteurs de mutations hétérozygotes du gène *DDX17*, nous avons cherché à identifier les effets d'une perte de fonction de *DDX17* pouvant expliquer les phénotypes observés chez les patients. Les résultats obtenus suite à l'analyse de l'impact moléculaire et physiologique d'une baisse d'expression de *DDX17* à l'aide de plusieurs modèles semblent montrer que *DDX17* joue en effet un rôle important dans le développement du système nerveux. Les études menées chez la souris et le xénope par nos collaborateurs montrent en effet qu'une perte d'expression de *DDX17* entraîne une diminution de la croissance des axones *in vivo* dans les deux modèles. De plus, l'étude *in vitro* de neurones corticaux de souris montre aussi une réduction de la complexité axonale. Enfin, des tests comportementaux menés chez le xénope montrent une réduction de la mémoire de travail.

L'analyse de l'effet d'une déplétion de *DDX17* sur le transcriptome de cellules humaines SH-SY5Y, réalisée dans notre équipe, montre que celle-ci entraîne une altération de l'expression de 350 gènes. Ces gènes présentent un enrichissement significatif en fonctions impliquées notamment dans la neurogenèse, la morphogenèse, la communication inter-cellules et l'assemblage synaptique. Ces fonctions semblent cohérentes avec les défauts de développement du système nerveux observés dans les modèles animaux, mais aussi chez les patients présentant des mutations hétérozygotes de *DDX17*. Parmi les gènes réprimés en condition si*DDX17*, on peut ainsi citer *PLXNA4* et *PLXNA2*, qui codent respectivement pour les co-récepteurs de SEMA3A et SEMA6A et jouent tous les deux un rôle dans la guidance axonale ([Andermatt et al., 2014](#); [Suto et al., 2005](#)). *PLXNA4* serait aussi impliqué dans la morphogenèse des dendrites ([Danelon et al., 2020](#)). Chez la souris, le KO de *PLXNA2* entraîne une baisse de la prolifération dans le gyrus denté et de la neurogenèse adulte. Cela s'accompagne de changements comportementaux comme une réduction de la mémoire de la peur et la sociabilité ([Zhao et al., 2018](#)). Chez l'humain, une étude d'association pan-génomique a permis de lier la présence de variant de *PLXNA2* à la schizophrénie ([Mah et al., 2006](#)). On peut aussi citer le gène *RPH3A*, qui régule entre autres l'organisation, l'activité et la plasticité synaptique

(Deák et al., 2006; Franchini et al., 2022; Yang et al., 2023). La présence d'un variant de *RPH3A* a été associée à l'apparition de troubles neurologiques multi-systémiques chez une jeune patiente (Maselli et al., 2018).

Le but n'est pas de faire une liste commentée exhaustive des fonctions de tous les gènes dérégulés par la déplétion de *DDX17*, mais ces exemples illustrent comment l'altération du transcriptome suite à la baisse d'expression de *DDX17* pourrait expliquer les phénotypes observés dans les modèles animaux ainsi que chez les patients. Cette hypothèse serait confortée si nous pouvions identifier des variations de gènes similaires dans ces autres modèles, ce qui constitue un objectif futur majeur.

Il est par ailleurs intéressant de noter que les orthologues de *DDX17* présents chez *D. melanogaster* (CG10777) et *C. elegans* (*ddx-17*) ont tous les deux été identifiés lors de cribles génétiques comme entraînant des défauts de la morphologie dendritique suite à leur knock-down par ARNi (Antonacci et al., 2015; Olesnický et al., 2013). Ces résultats semblent en accord avec les défauts de croissance axonale constatés chez la souris et le xénope, et suggèrent que le rôle de *DDX17* dans la morphogénèse des neurites est conservé à travers l'évolution.

#### Lien entre le rôle de *DDX17* dans le neuroblastome et le neuro-développement

Il faut ajouter que bien que nous n'ayons pour le moment étudié l'effet d'une déplétion de *DDX17* que sur l'expression globale des gènes, il est aussi possible que *DDX17* affecte le développement neuronal par d'autres moyens. Comme nos travaux précédents l'ont montré, *DDX5* et *DDX17* régulent de nombreuses autres étapes de la maturation et du métabolisme des ARN, notamment l'épissage, la terminaison transcriptionnelle, ainsi que l'expression et la maturation des miARN (Lambert et al., 2018; Terrone et al., 2022). Les effets spécifiques à *DDX17* sont peu connus et étudiés, mais il est possible qu'une baisse de son expression puisse affecter l'épissage et la maturation en 3' des transcrits. Le contrôle de l'APA et de la longueur de la 3' UTR des transcrits est particulièrement crucial dans le développement et le fonctionnement des neurones. Lors de la différenciation neuronale, on observe ainsi un allongement de l'extrémité 3' UTR de nombreux transcrits (Govenek and Tian, 2018; Kiltchewskij et al., 2023). Les neurones ont tendance à exprimer des isoformes avec de longues 3' UTR comparés à d'autres

cellules présentes dans le cerveau comme la microglie ou les cellules endothéliales (Guvenek and Tian, 2018).

La longueur de l'extrémité 3'UTR permet aussi de réguler la localisation des transcrits dans les neurones, les isoformes longues étant préférentiellement exportées dans les neurites, alors que les isoformes courtes restent dans le corps cellulaire. La capacité des ARNm localisés au niveau axonal ou dendritique à subir une traduction locale est essentielle pour les fonctions synaptiques. La concentration synaptique des canaux ioniques, tels que Kv1.1 (*Kcna1*), est ainsi régulée par la localisation et la traduction locale de son ARNm, qui dépendent elles-mêmes de l'extrémité 3'UTR (Raab-Graham et al., 2006). De façon intéressante, un papier en pré-publication indique que l'analyse de la polyadénylation alternative utilisant le séquençage des ARNm sur cellules uniques (ScRNA-Seq) à partir de tissu cérébral *post-mortem* a révélé des 3' UTR majoritairement plus longues chez les personnes autistes par rapport aux individus sains (Göpferich et al., 2020).

Il est par ailleurs intéressant d'observer que parmi les gènes significativement surexprimés suite à la déplétion de DDX17 dans les SH-SY5Y se trouve *MYCN*. Ce résultat est d'autant plus intrigant que la co-déplétion de DDX5 et DDX17 dans les SK-N-BE(2) induit une diminution de l'expression de *MYCN* (non montré). DDX17 seulement et non DDX5 aurait ainsi un rôle négatif sur l'expression de *MYCN*, bien que l'on ignore si le mécanisme est direct ou non. On peut aussi remarquer qu'un variant gain de fonction de *MYCN* a été relié avec un phénotype de déficience intellectuelle chez une jeune patiente (Yu et al., 2018). Le rôle de *MYCN* dans le développement du système nerveux périphérique est bien connu, mais il possède aussi des fonctions dans le système nerveux central, même si elles ont moins été étudiées. *MYCN* est en effet exprimé dans le cerveau humain fœtal (Grady et al., 1987) et la mutation hétérozygote de *MYCN* provoque le syndrome de Feingold, caractérisé par une réduction de la taille du cerveau et des difficultés d'apprentissage (van Bokhoven et al., 2005). Il est donc possible que la variation d'expression de *MYCN* induite par la déplétion de DDX17 puisse participer aux phénotypes observés chez les modèles animaux et les patients.

## Caractérisation de deux mutations faux sens de DDX17

La caractérisation des mutants avait pour but de comprendre l'effet des mutations faux-sens sur le comportement et le fonctionnement de DDX17 et de voir à terme si leur effet peut expliquer le phénotype observé chez les patients. Le variant R304H est prédit pour présenter une diminution de sa capacité à lier l'ARN. Cette prédiction pourrait être vérifiée en testant la capacité de DDX17-R304H à se lier à l'ARN *in vitro* par EMSA (RNA electrophoretic mobility shift assay). Ni R304H ni P203L ne sont prédites par simulation pour affecter l'activité ATPase de DDX17 (résultat non-montré). Toutefois, les simulations utilisées ne prennent pas en compte la structure entière de l'hélicase, dont les extrémités N-ter et C-ter intrinsèquement désordonnées n'ont pas encore été résolues par cristallographie (Ngo et al., 2019). Cela pourrait donc tout de même être pertinent de vérifier expérimentalement l'effet de ces mutations sur l'activité ATPase de DDX17, mais aussi sur son activité hélicase.

Il pourrait être intéressant de tester plus globalement l'impact des mutations identifiées sur l'interactome ARN et protéique de DDX17, respectivement par des expériences de CLIP-seq et de spectrométrie de masse. Il faut toutefois noter que le caractère insoluble du variant P203L peut rendre certaines méthodes moins efficaces.

On peut par ailleurs remarquer que les résidus touchés par cinq des six mutations faux-sens identifiées chez les patients sont conservés chez les orthologues de *DDX17* jusqu'à *C. elegans* et *S. cerevisiae* (voir annexe) Cela suggère l'importance de ces cinq résidus dans la fonction de l'hélicase. Le seul résidu touché par une mutation faux-sens non conservé est celui présent dans la région spécifique à l'isoforme p82.

Les expériences d'immunofluorescence et de fractionnement nous ont permis d'observer que la mutation P203L entraîne une agrégation et une baisse de solubilité de DDX17. Il est probable que la formation de tels agrégats séquestrent DDX17 et l'empêchent d'exercer ses fonctions cellulaires. Certains partenaires de DDX17 pourraient peut-être aussi être séquestrés dans ces agrégats, bien que DDX5 par exemple, qui est un partenaire de DDX17, ne semble pas voir sa localisation modifiée dans nos expériences (résultat non montré). Mes données préliminaires (résultat non montré) semblent indiquer une colocalisation entre DDX17-P203L et HDCA6, une désacétylase jouant un rôle essentiel dans la réponse cellulaire aux protéines mal

repliées et agrégées, qui participe au recrutement des agrégats de protéines ubiquitinyliées dans l'agrésome (Boyault et al., 2007). L'aspect des agrégats de DDX17-P203L, qui est plutôt dissemblable à celui des agrésomes « classiques », pourrait en effet être similaire aux agrésomes « ribbon-like » décrits par Garcia-Mata et al. (Garcia-Mata et al., 2002). La formation d'agrésomes est normalement un mécanisme protecteur permettant l'élimination des protéines mal repliées formant des agrégats. Néanmoins, la persistance de ces agrésomes est susceptible d'être néfaste à la cellule. Un article rapporte ainsi que la persistance d'agrésomes est associée à une morphologie nucléaire anormale, à l'induction de cassures d'ADN double brin et à l'arrêt du cycle cellulaire dépendant de p53. Les agrésomes pourraient avoir un effet néfaste sur la mitose en interférant de manière stérique avec l'alignement des chromosomes, le positionnement des centrosomes et la formation du fuseau (Lu et al., 2015). Si les agrésomes formés par DDX17-P203L ne sont pas éliminés efficacement par la cellule, on peut donc imaginer un effet néfaste sur les cellules exprimant *DDX17* à des niveaux importants. Toutefois, nos expériences ayant été réalisées via une transfection de plasmides dans les cellules, il nous faut vérifier si DDX17-P203L forme aussi des agrésomes en condition physiologique, à des niveaux d'expression similaires à la protéine DDX17 endogène.

Un objectif important est d'identifier l'impact de ces mutations sur le transcriptome des cellules. Nous avons tenté de faire des expériences de « rescue » (sauvetage) en exprimant DDX17 WT ou mutées dans les cellules traitées par siDDX17, mais nous n'avons pas réussi à obtenir un effet de rétablissement suffisant avec la DDX17 WT pour nous permettre d'analyser l'effet des mutations. Il s'agit d'une des limites d'un système basé sur la transfection transitoire de plasmides codants pour les protéines. Nous espérons que l'utilisation de lignées cellulaires mutées sur le gène *DDX17* nous permettra de comparer plus efficacement l'impact de ces mutations sur les fonctions de *DDX17*. Afin de caractériser l'effet des mutations de *DDX17* sur la différenciation et le fonctionnement neuronal, l'objectif de l'équipe est de développer en collaboration des modèles de cellules souches pluripotentes humaines porteuses de mutations de *DDX17*, dont on pourrait par la suite induire la différenciation neuronale.

En parallèle, l'effet des mutations de *DDX17* sur le développement neuronal pourra être examiné chez la souris par nos collaborateurs, sur des neurones en culture et, à plus long

terme, il serait aussi envisageable de développer des modèles knock-out et knock-in de souris.

### Conclusion partie 2

Ces travaux présentent des résultats d'expériences fonctionnelles (réalisées par nos partenaires) et transcriptomiques permettant d'identifier *DDX17* comme un nouveau régulateur du développement neuronal. Ces résultats supportent l'hypothèse que les mutations monoalléliques de *DDX17* identifiées chez 11 patients soient une cause de leurs troubles neuro-développementaux. Il s'agit de plus de la première étude présentant un lien entre *DDX17* et une maladie génétique. La caractérisation des mutations faux-sens identifiées pourrait nous permettre de mettre à jour de nouvelles fonctions cellulaires et moléculaires de cette hélicase à ARN.

**Dans l'ensemble, mes travaux de thèse permettent de montrer que l'altération de l'expression des hélicases DDX5 et surtout DDX17 entraîne des modifications du transcriptome à différents niveaux dans les cellules neuronales. L'altération de leur expression semble aussi pouvoir être liée au développement de pathologies du développement du système nerveux, à la fois cancéreuses comme le neuroblastome mais aussi de troubles neuro-développementaux. Des travaux supplémentaires seront toutefois nécessaires pour déterminer comment les fonctions de ces hélicases sont impliquées dans le développement de ces pathologies.**



# Matériels et méthodes

## Immunofluorescence :

Les cellules HEKT293 sont transfectées avec pCDNA.4-p72-FLAG WT ou muté sur des lamelles de verre puis sont fixées pendant 20 min dans de la formaline. Elles sont ensuite perméabilisées avec dans du PBS 1X – 0,5 % Triton pendant 10 min puis bloquées dans du PBS-Tween 0,2% - BSA 5% 1H à RT. Par la suite elles sont incubées 1h avec l'anticorps primaire à RT, rincées au PBS, incubées 1 h avec un anticorps secondaire conjugué à un fluorochrome, et rincées avec du PBS. Les lames furent montées dans du ProLong™ Diamond avec DAPI (P36966). Les acquisitions furent réalisées avec le Confocal Zeiss LSM800 avec un objectif x63 dans de l'huile d'immersion.

## Fractionnement cellulaire :

Les cellules HEKT293 sont lavées et récoltées dans du PBS 1X puis centrifugées 3 min à 3000 rpm à 4°C. Après un deuxième lavage au PBS 1X, le culot de cellules est resuspendu dans 200 µL de tampon "Dautry" complété avec de l'EDTA (10 mM Tris-HCl pH 8, 140 mM NaCl, 1.5 mM MgCl<sub>2</sub>, 10 mM EDTA, 0.5%NP40) et incubé 10 min sur glace. Le lysat est centrifugé 5 min à 3000 rpm à 4°C, le surnageant correspondant à la fraction cytoplasmique est récupéré et le culot constitué de noyaux est lavé avec 200 µL de tampon "Dautry" + EDTA. Le culot est ensuite resuspendu dans 200 µL de tampon Dautry. Les fraction cytoplasmique et nucléaire sont ensuite soniquées pendant 5 minutes puis centrifugées. Le surnageant correspondant au fraction « soluble » est récupéré, et le culot correspondant à la fraction « insoluble » est resuspendu dans 10µl de tampon à haute force ionique (10 mM Tris-HCl pH 8, 420 mM NaCl, 3 mM MgCl<sub>2</sub>, 0.5% NP40).

## Fractionnement des fractions solubles et insolubles :

Les cellules HEKT293 sont lavées et récoltées dans du PBS 1X puis centrifugées 3 min à 3000 rpm à 4°C. Après un deuxième lavage au PBS 1X, le culot de cellules est resuspendu dans 200 µL de tampon NP40 (0.5% NP40, 10 mM Tris HCl pH 7.8, 10 mM EDTA, 150 mM NaCl, 1 × protease inhibitor cocktail) et incubé 10 min sur glace. Le lysat est ensuite soniqué au bioruptor 5 min, puis centrifugé 25 min à 13000 rpm à 4°C. le surnageant correspondant à la fraction soluble est récupéré et le culot est lavé avec 200 µL de tampon de lavage (50 mM Tris HCl pH 7.4, 150 mM NaCl), pour être ensuite resuspendu dans 200 µl de tampon de resolubilisation (50 mM Tris HCl pH 6.8, 5% SDS, 10% glycerol) et soniqué pendant 5 min puis centrifugé 10 min à 13000 rpm à RT. Le surnageant correspondant à la fraction insoluble est ensuite collectée.

# Annexes

### Tableau 4 : Alignement clustal omega des orthologues de DDX17

CLUSTAL O(1.2.4) multiple sequence alignment

Q9XUW5_CAEEL	-----MGDRGYGGSRSYGGSSGG-----	18
DBP2_YEAST	-----	0
Q8MZI3_DROME	-----MN-MY---NGQMNAFGGGGGGAGGLAGGMVPMRMGGGAPGGGPG-----	41
F7B0D7_XENTR	-----	0
A0A5H1ZRQ2_HUMAN	MPTGFVAPILCVLLPSPPTREAATVASATGDSA-----SERESAAPAAAPTAEAPP	51
Q3U741_MOUSE	-----	0
Q9XUW5_CAEEL	-----GSRGGYGGGRGGGGGYSGRGGGYGGGGGGY	53
DBP2_YEAST	-----MTYGGRDQQYNKNTNYKSRGGDFRGGGRNSDRNSY-----NDRPQGGNY	42
Q8MZI3_DROME	-----GAGMFQRNRSAPYP--GFNGHGPANGGQRRMGGGP-----SMGPGGPRN	84
F7B0D7_XENTR	-----MRG-----S-----FMDRD----	9
A0A5H1ZRQ2_HUMAN	SVVTRPEPQALPSPAIRAFLP--DLYPFGTMRGG-----G-----FGDRDRDRD	93
Q3U741_MOUSE	-----MRGG-----G-----FGDRDRDRD	14
*		
Q9XUW5_CAEEL	GGGYGGGRGGGRGGSNGSAGGLRDLVDWS--AENLTPIEKDFYHENAASVSRREQYEIDQ	111
DBP2_YEAST	RGGFGGRSN-----YNQPQELIKPNWDEELPKLPTFEKNFYVEHESVDRSDSEIAQ	94
Q8MZI3_DROME	QDGGFQQNGG---QRSSNHGAHLPIKIVWS--EVNLTFFRKNFYKPCDSVLARTVGETET	138
F7B0D7_XENTR	RGG---GAGMSRFPKFGNPGERLRKKRWD--LSELPKFEKNFYTEHPEVARMTQHDVEE	63
A0A5H1ZRQ2_HUMAN	RGGFGARGGGGLPPKFKGNPGERLRKKKWD--LSELPKFEKNFYVEHPEVARLTPYEVDE	151
Q3U741_MOUSE	RGGFGARGGSGGLPPKFKGNPGERLRKKKWD--LSELPKFEKNFYVEHPEVARLTPYEVDE	72
.* . . . . * . * . : * : . : * * * * * * :		
Q9XUW5_CAEEL	WVSANQVTLEG-RGVPRPVFEFNEAPLPGQIH-ELLYGKFKQKPTVIQSIWFIAMSGRDI	169
DBP2_YEAST	FRKENEMTISG-HDIPKPIITTFDEAGFPDYVLNEVKAEGFDKPTGIQCQGWFIAMSGRDM	153
Q8MZI3_DROME	FLTSNEITIKG-DQVPTPSIEFEEGGFPDYVMNEIRKQGFAPKPTAIQAQGWFIAMSGRDL	197
F7B0D7_XENTR	LRRKKEITIRG-VNCPKPIYGFHQANFPQYVLDVLIDQRFKEPTPIQCQGFELALSGRDM	122
A0A5H1ZRQ2_HUMAN	LRRKKEITVRGGDVCPKPVFAFHANFPQYVMDVMDQHFTEPTPIQCQGFELALSGRDM	211
Q3U741_MOUSE	LRRKKEITVRGGDVCPKPVFAFHANFPQYVMDVMDQHFTEPTPIQCQGFELALSGRDM	132
:::* * * * * . . . : * : : * : * * * * . . : * : * * * * :		
Q <span style="color: red;">P203L</span>		
Q9XUW5_CAEEL	ISIAKTGSGKTLAFMLPALVHITKQHRQRGEGPAVLVLLPTRELAQQVQVEVSIDFCHSL	229
DBP2_YEAST	VGIAATGSGKTLSCYCLPGIVHINAQPLLAPGDGPIVLVLAPTRELAQQIQTECSKFGHSS	213
Q8MZI3_DROME	VGVAQTGSGKTLAYVLPVAVHINNQPRLERGDGPIALVLAPTRELAQQIQVAIEFGSNT	257
F7B0D7_XENTR	VGIAQTGSGKTLAYLLPAMVHINHQPYLERGDGPICLVLAPTRELAQQVQVADDYGKSS	182
A0A5H1ZRQ2_HUMAN	VGIAQTGSGKTLAYLLPAIVHINHQPYLERGDGPICLVLAPTRELAQQVQVADDYGKCS	271
Q3U741_MOUSE	VGIAQTGSGKTLAYLLPAIVHINHQPYLERGDGPICLVLAPTRELAQQVQVADDYGKCS	192
: . : * * * * * : : * . : * :		
<span style="color: red;">I T217A</span> <span style="color: green;">la</span> <span style="color: red;">Q259R</span>		

Q9XUW5_CAEEL	GLKMTCLFGGASKGFPQARDLERGVDIVVATPGRLLDFLDNGTTNMKKCSYLVLDEADRML	289
DBP2_YEAST	RIRNTCVYGGVPKSQQIRDLRSRGSEIVIATPGRLIDMLEIGKTNLKRVTYLVLDEADRML	273
Q8MZI3_DROME	HVRNTCIFGGAPKQGQARDLERGVEIVIATPGRLIDFLERGTSLKRCTYLVLDEADRML	317
F7B0D7_XENTR	RLKSTCIYGGAPKGFQIRDLERGVEICIAATPGRLIDFLEAGKTNLRRCTYLVLDEADRML	242
A0A5H1ZRQ2_HUMAN	RLKSTCIYGGAPKGFQIRDLERGVEICIAATPGRLIDFLESGKTNLRRCTYLVLDEADRML	331
Q3U741_MOUSE	RLKSTCIYGGAPKGFQIRDLERGVEICIAATPGRLIDFLESGKTNLRRCTYLVLDEADRML	252

:: \*\*: \*\*: \* . \* \* \* . \* \* : \* : \* \* \* \* \* . \* : \* : \* . \* . : : \* \* \* \* \* \* \* \* \* \*  
Ib
Ic R304H
II

Q9XUW5_CAEEL	DMGFEPQIKKIIGQIRPDRQTLMFSAITWPKEVRALASDFQKDA AFLNVGSLELAANHNT	349
DBP2_YEAST	DMGFEPQIRKIVDQIRPDRQTLMWSATWPKEVKQLAADYLNDP IQVQVGSLELSASHNIT	333
Q8MZI3_DROME	DMGFEPQIRKIMQQIRPDRQVLMWSATWPKEVRQLAEFLNNYIQVNI GLSLSANHNIL	377
F7B0D7_XENTR	DMGFEPQIRKIVDQIRPDRQTLMWSATWPKEVRQLAE DFLRDYVQINIGNLELSANHNIL	302
A0A5H1ZRQ2_HUMAN	DMGFEPQIRKIVDQIRPDRQTLMWSATWPKEVRQLAE DFLRDYTQINVGNELSANHNIL	391
Q3U741_MOUSE	DMGFEPQIRKIVDQIRPDRQTLMWSATWPKEVRQLAE DFLRDYTQINVGNELSANHNIL	312

\* \* \* \* \* : \* \* : \* \* \* \* \* . \* : \* \* \* \* \* : \* \* : : . : : : \* . \* : \* . \* \* \*  
III

Q9XUW5_CAEEL	QVVDILEEHAKQAKMELLNHIMNQKECKTIIFVETKRRKADELTRAMRRD GWPTLCIHGD	409
DBP2_YEAST	QIVEVVSDFEKDRDLNKYLETASQDNEYKTLIFASTKRCDDITKYLRE DGPALAIHGD	393
Q8MZI3_DROME	QIVDVCENEKMLKLIKLLTDISAENETKTIIFVETKRRVDEITRNI SRQWRACAIHGD	437
F7B0D7_XENTR	QIVDVCQESEKDHKLIQLMEEIMAEKENKTIIFVETKRRCCDLTRR MRDGPAMCIHGD	362
A0A5H1ZRQ2_HUMAN	QIVDVCMESEKDHKLIQLMEEIMAEKENKTIIFVETKRRCCDLTRR MRDGPAMCIHGD	451
Q3U741_MOUSE	QIVDVCMESEKDHKLIQLMEEIMAEKENKTIIFVETKRRCCDLTRR MRDGPAMCIHGD	372

\* . \* : : \* : \* : : : \* \* \* . \* . \* : \* : \* : : . : \* \* : . \* \* \* \*  
IV
IVa

Q9XUW5_CAEEL	KNQGERDWVLQEFKAGKTPIMLATDVAARGLDVDDIKFVINYDYPNS SEDYVHRIGRTGR	469
DBP2_YEAST	KDQERERDWVLQEFRNGRSPIMVATDVAARGIDVKGINYVINYDMPG NI EYVHRIGRTGR	453
Q8MZI3_DROME	KSQQERDFVLSFRNGRSHILVATDVAARGLDVDDVKFVINYDYPNS SEDYVHRIGRTGR	497
F7B0D7_XENTR	KSQQERDWVLCEFRTGKAPILIATDVASRGLDVEDIKFVINYDYPNS SEDYVHRIGRTAR	422
A0A5H1ZRQ2_HUMAN	KSQPERDWVLNEFRSGKAPILIATDVASRGLDVEDVKFVINYDYPNS SEDYVHRIGRTAR	511
Q3U741_MOUSE	KSQPERDWVLNEFRSGKAPILIATDVASRGLDVEDVKFVINYDYPNS SEDYVHRIGRTAR	432

\* . \* \* \* \* . \* : \* : \* : \* \* \* \* . \* : \* \* \* \* \* \* . \* . \* \* \* \* \* \* \* \* \* \* \*  
Q454R

Q9XUW5_CAEEL	SDKKGTAYTFFTHTNASKAKDLLKVLDEAKQTVPAALRDMANRS-YGG SNSRGRYGGGGF	528
DBP2_YEAST	AGATGTAISFFTEQNKGLGAKLISIMREANQNIPELLKYDRRS-YGGGH --PRYGGGGRG	510
Q8MZI3_DROME	SNNTGTAYTLFTHSNANKANDLIQVLEANQTINPKLMNAMNGGYNKRG --GGMGGGYR	555
F7B0D7_XENTR	STNKGTA YTFFTPGNLQARELVKVEEANQTINPKLMQLVDHGR-----GGGG---G	472
A0A5H1ZRQ2_HUMAN	STNKGTA YTFFTPGNLQARELIKVLEEANQAINPKLMQLVDHRG-----GGGGGGKG	564
Q3U741_MOUSE	STNKGTA YTFFTPGNLQARELIKVLEEANQAINPKLMQLVDHRG-----GGGGGGKG	485

: . \* \* \* : : \* \* \* \* . \* : \* : \* : \* \* \* : \* . . \*

Q9XUW5_CAEEL	QKRGYGGN-----DN--FAPKRPYDNGGYGGSGGGGRW-----	561
DBP2_YEAST	GRGGYGR-----GGYGGRRGGYGGNRQRDGGWGNR-----G	542
Q8MZI3_DROME	GGNGYQGR---NPQ-----MGGGYNGGNNYRNNN--GPGATMNRNSFNGGSAGG	599
F7B0D7_XENTR	GRSRYRSIPTNNPNMMYQEECNRRMHGKDGGRRDGRNNN--RDGAIS---NNPTGRDQN	527
A0A5H1ZRQ2_HUMAN	GRSRYRTTSSANNPNLMYQDECRRRLRGVKDGGRRDSASY--RDRSETDRAGYANGSGYG	622
Q3U741_MOUSE	GRSRYRTTSSANNPNLMYQDECRRRLRGVKDGGRRDSTSY--RDRSETDRASYANGSGYG	543

\* \* \*

Q9XUW5_CAEEL	-----	561
DBP2_YEAST	RSNY-----	546
Q8MZI3_DROME	PPRFDQKPRTSPPNQGGQGGYRGQGGGGYQGGQQQQQQNGGVFFSRFNPNAACFEPNKP	659
F7B0D7_XENTR	RPNYS-----YGGQVSNFQPNQY	545
A0A5H1ZRQ2_HUMAN	SPNSA-----FGAQ-----AGQY	635
Q3U741_MOUSE	SPNSA-----FGAQ-----AGQY	556

Q9XUW5_CAEEL	-----	561
DBP2_YEAST	-----	546
Q8MZI3_DROME	QNGPGSGPPQAQHQQQQQAQAQQHLLPDASVQAAMTAANYGVDQKRS-----	708
F7B0D7_XENTR	SYGQGGFTGVA-----NNY-----	559
A0A5H1ZRQ2_HUMAN	TYGQGTGAAA-----YGTSSYTAQEYAGAGTYGASSTTSTGRSSQ	675
Q3U741_MOUSE	TYAQTGTYGAAA-----YGTSGYTAQEYAAGTYGASSTASAGRSSQ	596

Q9XUW5_CAEEL	-----	561
DBP2_YEAST	-----	546
Q8MZI3_DROME	---RFSYNNMNFANMPPAMNQQAQAQAQQQQQLQTPQQQQQQMQPQAYGQYSS	764
F7B0D7_XENTR	-----G-TAAYAQAQPPLMSQQ-----FAPPS-----	580
A0A5H1ZRQ2_HUMAN	SSSQQFSG-IGRSGQQPPLMSQQ-----FAQPP-----	703
Q3U741_MOUSE	SSSQQFSG-IGRSGQQPPLMSQQ-----FAQPP-----	624

Q9XUW5_CAEEL	-----	561
DBP2_YEAST	-----	546
Q8MZI3_DROME	MSSSMATVSLNGGAAAVASSYRAPHYAVPTQAAPYMMANGGDI FAYPPPLPVQN----	818
F7B0D7_XENTR	-----SSGLLNYM--NQAASYQYPPPPPPPPPPTRK	609
A0A5H1ZRQ2_HUMAN	--G-----ATNMIGYM--GQ-TAYQYPPPPPPPPSRK-	731
Q3U741_MOUSE	--G-----ATNMIGYM--GQ-TAYQYPPPPPPPPSRK-	652

## Projet splicing lore

Splicing Lore est une interface web, liée à une base de données qui regroupe 160 jeux de données publiques (GEO Omnibus ou ENCODE) issues de séquençages d'ARN réalisés dans une dizaine de lignées cellulaires différentes, dans lesquelles l'expression de 75 facteurs d'épissage différents a été inhibée par traitement siARN ou shARN.

Le but de cet outil est de prédire, à partir d'une liste d'exons d'intérêt, si tout ou partie de ces exons est régulée par l'un des facteurs d'épissage analysés. Un score de fiabilité a été mis en place pour faciliter aux utilisateurs la recherche et l'étude des facteurs d'épissage candidats et de leurs exons d'intérêts. Ce score compris entre 0 et 1 est lié à un test statistique de permutation, qui donne aussi une p-value. Plus le score est proche de 1 plus, le facteur d'épissage est corrélé ou anti-corrélé à notre liste d'exons.

Ma contribution à ce projet a été de valider par RT-PCR la co-régulation de plusieurs exons par DDX5/17 et HNRNPC dans la lignée cellulaire SHSY5Y.



DATABASE

**SplicingLore : a web resource for studying the regulation of cassette exons by human splicing factors**

Journal:	<i>DATABASE</i>
Manuscript ID	DATABASE-2023-0073.R1
Manuscript Type:	Original Article
Date Submitted by the Author:	06-Nov-2023
Complete List of Authors:	Polvèche, Hélène; Université Claude Bernard Lyon 1; I-Stem Valat, Jessica; Université Claude Bernard Lyon 1 Fontrodona, Nicolas; Université Claude Bernard Lyon 1 Lapendry, Audrey; Université Claude Bernard Lyon 1 Clerc, Valentine; Université Claude Bernard Lyon 1 Janczarski, Stéphane; Université Claude Bernard Lyon 1 Mortreux, Franck; Université Claude Bernard Lyon 1 Auboeuf, Didier; Université Claude Bernard Lyon 1 Bourgeois, Cyril; Université Claude Bernard Lyon 1
Keywords:	Alternative splicing, Splicing regulation, Splicing factor, web resource, Database

SCHOLARONE™  
Manuscripts

# SplicingLore : a web resource for studying the regulation of cassette exons by human splicing factors

Hélène Polvèche<sup>1,3\*</sup>, Jessica Valat<sup>1,2\*</sup>, Nicolas Fontrodona<sup>1,2\*</sup>, Audrey Lapendry<sup>1,2</sup>, **Valentine Clerc<sup>1,2</sup>**, Stéphane Janczarski<sup>1</sup>, Franck Mortreux<sup>1,2</sup>, Didier Auboeuf<sup>1,2#</sup>, Cyril F. Bourgeois<sup>1,2#</sup>

<sup>1</sup> Laboratoire de Biologie et Modelisation de la Cellule, Ecole Normale Supérieure de Lyon, CNRS, UMR 5239, Inserm, U1293, Université Claude Bernard Lyon 1, 46 allée d'Italie F-69364 Lyon, France

<sup>2</sup> "Equipe Labellisée Ligue Contre le Cancer"

<sup>3</sup> CECS/AFM, I-STEM, 28 rue Henri Desbrières, F-91100, Corbeil-Essonnes, France

\* These authors contributed equally

# These authors contributed equally

Corresponding author : [cyril.bourgeois@inserm.fr](mailto:cyril.bourgeois@inserm.fr)

## Abstract

One challenge faced by scientists from the alternative RNA splicing field is to decode the cooperative or antagonistic effects of splicing factors to understand and eventually predict splicing outcomes on a genome-wide scale. In this manuscript, we introduce SplicingLore, an open access database and web resource that help to fill this gap in a straightforward manner. The database contains a collection of RNA-seq-derived lists of alternative exons regulated by a total of 75 different splicing factors. All datasets were processed in a standardized manner, ensuring valid comparisons and correlation analyses. The user can easily retrieve a factor-specific set of differentially included exons from the database, or provide a list of exons and search which splicing factor(s) control(s) their inclusion. Our simple workflow is fast and easy to run, and it ensures a reliable calculation of correlation scores between the tested datasets. As a proof of concept, we predicted and experimentally validated a novel functional cooperation between the RNA helicases DDX17 and DDX5 and the HNRNPC protein. SplicingLore is available at <https://splicinglore.ens-lyon.fr/>.



## Introduction

Eukaryotic genes are transcribed into pre-messenger RNAs (pre-mRNAs) that are most of the time composed of a succession of exons separated by introns, the latter being removed by the spliceosome to form mature RNA molecules or mRNAs. The spliceosome is a large ribonucleoprotein complex which recognizes the 5' and 3' splice sites delimitating exons from introns and catalyzes the splicing reaction, a process that is modulated by numerous auxiliary factors, including RNA binding proteins (RBP) <sup>1,2</sup>. The split nature of genes and the contribution of many different parameters enables alternative splicing, which is defined as the differential selection of exonic or intronic sequences to produce different mRNA isoforms from the same gene. Alternative splicing is a prevalent phenomenon that greatly expands proteome diversity and that also contributes to the quantitative regulation of gene expression <sup>3,4</sup>. Various splicing dedicated databases, such as VastDB <sup>5</sup>, ASpedia <sup>6</sup> or FasterDB <sup>7</sup>, serve as repositories of alternative splicing profiles and inclusion levels across cell types or organisms. ASpedia also provides useful functional information about the alternative exons, as Exon Ontology <sup>8</sup> which helps to predict the biological consequences of genome-wide splicing variations.

Mechanistically, the binding of RBPs, like serine/arginine-rich (SR) proteins and hnRNP proteins, to short degenerated RNA motifs located within exons or in their flanking introns, is determinant for the control of alternative exon inclusion <sup>9-12</sup>. The activity of a given RBP can also be favoured or antagonized by other factors, hereafter defined generally as splicing factors (SF), and their joint action or competition can promote or inhibit the assembly of the spliceosome onto nearby splice sites <sup>13,14</sup>. The specificity and intrinsic properties of each SF, even those of general constituents of the spliceosome, combined to the unique sequence and structural organization of each exon, explain why the knockdown of a specific factor in cells only results in the differential splicing of a limited number of transcripts, highlighting the relative plasticity of the spliceosome. The identification and quantification of splicing changes controlled by a given factor can be monitored by RNA sequencing (RNA-seq) technologies <sup>15</sup>. However, it is more difficult to understand or to predict the cooperative or antagonistic effects of SFs on alternative splicing, especially on a genome-wide scale.

To address this question, we present SplicingLore, a database and web resource that contains the lists of alternative exons that are differentially included upon knockdown of 75 different SFs in various human cell lines. This collection is integrated into a website from which the user can easily download a given dataset of SF-regulated exons or submit a single exon or a list of exons to predict their potential regulation by the SFs of the database. As a proof of concept of the utility of our resource, we predicted and experimentally validated a new functional cooperation between RNA helicases DDX17 and DDX5 and the HNRNPC protein. SplicingLore is freely available at <https://splicinglore.ens-lyon.fr/>.

## Results

### Data collection and analysis

To feed the SplicingLore database, we searched the GEO omnibus<sup>16</sup> and ENCODE<sup>17</sup> databases for datasets generated from Illumina RNA seq experiments that all consisted in knocking down a given splicing factor (SF) in a human cell line, mostly by means of a treatment with siRNA or shRNA (Figure 1A). We filtered out the datasets for which the sequencing quality or depth was too low to reach our objective, which was to obtain robust and reliable lists of alternative exons regulated by the tested SF (see Methods for details). We eventually obtained a curated list of 160 datasets, corresponding to 75 SFs tested in 21 different cell lines (Figure 1A, Supplementary Table S1). Three cell lines (HepG2, K562 and 293T) represented more than 75% of the datasets (Figure 1B), and some SFs were tested in up to 5 cell lines.

The selected datasets were all processed in the same way using FaRLine, a splicing-dedicated pipeline developed earlier in our lab<sup>18</sup>. We thus obtained uniformly formatted lists of alternative splicing events for each dataset. Although FaRLine detects and quantifies variations in alternative 5' and 3' splice sites, as well as multiply spliced exons, we decided to focus only on single cassette exons, which represent the vast majority of regulated splicing events and which are easier to handle when comparing multiple datasets. Note that the annotation of all exons is based on the FasterDB database (<https://fasterdb.ens-lyon.fr/faster/home.pl>)<sup>7</sup>.

Integrating data from various cell lines could introduce a bias as splicing regulation is inherently variable across cell lines. Indeed, some exons may exhibit a discordant regulation by the same SF in two different cell lines, most likely because cofactors may contribute to their cell-specific regulation. To address this question, we looked whether differential exon inclusion (DeltaPSI) changed across cell lines, for each SF analysed in at least 3 datasets. These analyses globally showed that each SF displayed a similar regulation of its target exons across different cell lines (Supplementary Figure S1). We then performed a broader clustering analysis for all exons regulated by all SFs for which at least 3 cell lines were available. This analysis showed that exons were clustered according to the SF rather than the cell line (Supplementary Figure S2), ruling out a possible cell line-based bias. Note that the analysis presented in Supplementary Figure 1 also allowed us to verify that the way the cells were treated to achieve SF knockdown (mostly siRNA or shRNA) did not significantly affect the result.

To underline further the relevance of our alternative splicing analyses, we also tested whether SF binding was enriched in the regions surrounding their corresponding alternative exons identified by FaRLine. First, we recovered available CLIP-seq data for 29 SF and calculated the enrichment of peaks at and around the exons that are up- and down-regulated by the corresponding factors, as determined by FaRLine, compared to control exons. As shown in Figure 2A, exons regulated by each tested SF

1  
2  
3 displayed an enriched binding of the corresponding SF at an/or around the exon, with sometimes a  
4 differential enrichment pattern between positively and negatively regulated exons. For example,  
5 HNRNPC binding is strongly enriched at exons upregulated upon HNRNPC KD, consistent with its  
6 general binding-associated repressive role on splicing<sup>12,19</sup>. In contrast, SR proteins such as SRSF3,  
7 SRSF1, SRSF7 or TRA2A bind more strongly to the exons that they promote the inclusion of (Figure 2A)  
8<sup>20</sup>. Second, we recovered the consensus binding sites for 21 SF and similarly looked for their enrichment  
9 at/around their down- and up-regulated exons (Figure 2B). Again, the differential enrichment of  
10 binding sites matched the known function of the corresponding SF. For example, we observed a  
11 positive enrichment of SR protein binding sites on exons that are positively regulated by these proteins,  
12 and vice versa.  
13  
14  
15  
16  
17  
18  
19

### 20 21 22 **Web interface and graphic visualisation**

23 SplicingLore can be queried in two different ways (Figures 1A and 3A). The first method allows the user  
24 to analyse a list of exons associated with parameters of splicing regulation. The default format  
25 requested for this analysis includes the gene symbol and exon number, chromosomal coordinates, as  
26 well as values of differential exon inclusion ( $\Delta$ PSI) and  $p$ -value (Method 1, Figure 3A, left window).  
27 Gene symbols and exon numbers correspond to gene annotations of the FasterDB database<sup>7</sup>, but the  
28 critical details to provide here are the chromosomal coordinates. Incorrect gene symbol and/or exon  
29 number will not impede the search but instead will issue a warning message and a proposition for  
30 correcting the input, based on coordinates (see below). To facilitate the navigation and the conversion  
31 of exon coordinates in both directions between the hg19 and hg38 version of the human genome, the  
32 SplicingLore interface includes a tab that allows the user to directly launch a conversion from the UCSC  
33 reference chain files with the R package liftOver (<https://bioconductor.org/packages/liftOver>). (Figure  
34 3A).  
35  
36  
37  
38  
39  
40  
41  
42

43 The query list of exons can be automatically retrieved from one of the projects referenced in the  
44 database, upon selection of the dataset of interest from the drop-down menu (Figure 3A, top right  
45 window). Alternatively, the user can upload a list of exons of interest to query SplicingLore for their  
46 potential regulation by the SFs of the database. As each exon is associated to a  $\Delta$ PSI value, this first  
47 method will allow to look for positive and negative correlations between the input regulation and the  
48 regulation by SFs from the database.  
49  
50  
51  
52

53 The second query method is less restrictive as it does not require any parameter of exon inclusion ( $\Delta$ PSI  
54 and  $p$ -value), if these values are not available (Method 2, Figure 3A). In this mode, the user can upload  
55 a list of exons that are only identified by their chromosomal coordinates (Figure 3A, bottom right  
56 window).  
57  
58  
59  
60

1  
2  
3 Once the query is launched, the first displayed page ("List of exons" tab) informs the user about exons  
4 which were not properly recognized, either because they were not detected in any of the 160 datasets,  
5 or because they were incorrectly formatted. In the latter case, corrected features are given to allow  
6 the user to restart the analysis (Supplementary Figure S3).  
7  
8

9 From the "Scores" tab, the user obtains a downloadable table (Figure 3B) which recapitulates the  
10 correlation scores between the query list and each of the SplicingLore datasets, as well as different  
11 comparison parameters that are explained below. Since this table is sorted based on our custom Score,  
12 the user can immediately visualize on top of the list which SF is more susceptible to be involved in the  
13 regulation of the tested exons. Scores indicating a significant positive or negative correlation are shown  
14 in different colours to facilitate their rapid visualization on the website (downloaded files are  
15 colourless). The "percent.common.query" and "percent.common.SF" columns indicate the fraction of  
16 common exons between the compared datasets, calculated from the query list or from the tested  
17 dataset, respectively (see Supplementary Figure S4 for a representation of these fractions).  
18  
19

20 The "Graphics" tab presents the same information in the form of a correlation graph, which is available  
21 in a fixed or interactive mode (Figure 3C). The values used to generate these graphs can also be  
22 downloaded as a table. Finally, a more comprehensive table is available in the "Details Up/Down" tab  
23 (Figure 3D). Alongside the information described above, it also provides the number (#) and fraction  
24 (%) of exons that are shared by both compared datasets, taking into account the direction of their  
25 regulation (up-regulation or down-regulation of their inclusion upon SF knock-down).  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35

### 36 **Example of application**

37 In order to check out the validity of our tool, we retrieved from SplicingLore the list of exons whose  
38 inclusion was altered upon the knock-down of DEAD box RNA helicases DDX5 and DDX17 in the 293T  
39 cell line (GSE123752, Supplementary Table S3). We then used this list as a query to search for possible  
40 correlative (or anti-correlative) effects between these splicing regulators and other SFs.  
41  
42  
43

44 The first lines of the resulting table show that beside the query list itself (score of 1), the best  
45 correlation scores correspond to the other two DDX5/DDX17 datasets included in SplicingLore, in SH-  
46 SY5Y and MCF7 cell lines, with respective scores of 0.59 and 0.53 (Figure 3B, Supplementary Table S4).  
47 Interestingly, we also observed a good correlation (with a score of 0.41-0.43) between the tested  
48 DDX5/DDX17 dataset and five different HNRNPH1 datasets (Figure 3B, Supplementary Table S4). This  
49 is in agreement with the fact that HNRNPH1 was previously described as a co-regulator of  
50 DDX5/DDX17-mediated splicing in myoblasts and epithelial cells<sup>21</sup>.  
51  
52  
53  
54  
55

56 Results from Supplementary Table S4 also revealed HNRNPC as one of the top predicted factors, with  
57 positive correlation scores in 4 datasets (score= 0.44 in 293T, K562 and MG63 cell lines, and 0.38 in  
58 HepG2 cells) (Figure 3B, Supplementary Table S4). When considering separately the subclasses of up-  
59  
60

1  
2  
3 and down-regulated exons, we noticed that the positive correlation between DDX5/DDX17 and  
4 HNRNPC exons was especially visible for down-regulated exons (Supplementary Table S5,  
5 Supplementary Figure S5). Indeed, up to 15.6% of exons down-regulated upon HNRNPC knockdown  
6 were regulated in the same manner (11.5% in average for the 4 correlated datasets), while only about  
7 1% were regulated in the opposite manner (Supplementary Tables S5 and S6). For up-regulated exons,  
8 results were more variable between the different datasets, although a trend for a positive correlation  
9 was also observed, especially in 293T and K562 cells. These 2 datasets displayed more than 11% of  
10 positive correlation with the query, these values being the highest among all datasets (apart from  
11 DDX5/DDX17 datasets, Supplementary Table S5).

12  
13 To explore further the possibility of a functional link between DDX5/DDX17 and HNRNPC, we first  
14 queried SplicingLore in a reverse manner, uploading on the website a list of 3030 exons corresponding  
15 to the union of all exons regulated upon HNRNPC knock-down in 4 datasets of our database (293T,  
16 K562, HepG2 and MG63). This analysis identified the 3 DDX5/DDX17 datasets among those showing a  
17 significant positive correlation (scores from 0.46 to 0.50) with the query list (Supplementary Table S7,  
18 Supplementary Figure S6).

19  
20 Finally, to experimentally validate our predictions, we silenced the expression of DDX5/DDX17 and  
21 HNRNPC in 293T cells, independently or together (Figure 4A and Supplementary Figure S7A), and we  
22 monitored the effect of these treatments on a selection of alternative exons whose inclusion was  
23 found to be impacted by the knock-down of these factors (10 more skipped exons and 7 more included  
24 exons). In conditions of single depletion, the inclusion of all exons was modified according to the  
25 predictions and interestingly, the combined depletion of DDX5/DDX17 and HNRNPC significantly  
26 enhanced the effect of single siRNA treatment on splicing (Figures 4B and 4C). **We observed the same  
27 cooperative effect of DDX5/DDX17 and HNRNPC on exon inclusion in the neuroblastoma SH-SY5Y cell  
28 line (Supplementary Figures S7B-D).** This cooperation prompted us to test whether these factors  
29 interact with each other. Indeed, endogenous HNRNPC and DDX17 co-immunoprecipitated in an RNA-  
30 independent manner in 293T cells, suggesting a possible direct interaction (Figure 4D). No clear sign of  
31 interaction was observed between DDX5 and HNRNPC (Supplementary Figure S7E).

32  
33 Altogether, these results disclosed a novel functional partnership between HNRNPC and the helicase  
34 DDX17, and illustrated the usefulness of SplicingLore to improve our knowledge on splicing regulation.

## 35 Discussion

36  
37 One challenge faced by scientists from the alternative splicing field is to decode the cooperative or  
38 antagonistic effects of SFs to understand and eventually predict splicing outcomes on a genome-wide  
39 scale<sup>22</sup>. Exploiting a large collection of RNA-seq-derived lists of exons regulated by 75 SFs, SplicingLore  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 helps to fill this knowledge gap in a straightforward manner. RNA-seq datasets from ENCODE have  
4 previously been analysed in a transverse manner, for example along with RBP-RNA interactions  
5 datasets, which helped to understand the functional roles of these proteins in alternative splicing<sup>11,12</sup>.  
6 SplicingLore includes not only 100 ENCODE datasets, but also 60 datasets from other sources. Our  
7 resource also enables cross-comparisons between datasets, and allows to directly search for SFs that  
8 control the inclusion of a list of exons provided by the user. Therefore, SplicingLore complements other  
9 splicing databases such as VastDB<sup>5</sup> or ASpedia<sup>6</sup>, that are repositories of alternative splicing profiles  
10 and inclusion levels across cell types or organisms, and which also provide useful information about  
11 the functional impact of splicing variations. SplicingLore only requires to format correctly the list of  
12 exons of interest, and it is therefore fast and easy to run. All datasets were selected based on the  
13 quality and coverage depth of the sequencing experiment, and they were processed in a standardized  
14 manner. This unique experimental workflow ensures a reliable calculation of correlation scores  
15 between the tested dataset and good validation efficiency when testing the inclusion of the predicted  
16 exons experimentally.

17  
18 Of note, what we describe here is the most straightforward way to use SplicingLore, but our data can  
19 be exploited in a different and deeper way, depending on the user's interest. For example, we  
20 previously used some of the SF-specific alternative exon lists stored in SplicingLore to explore the  
21 intrinsic nature of these exons and the links that exist between their biased nucleotidic composition  
22 and splicing regulatory sequences<sup>23,24</sup>.

23  
24 As a proof of concept, we searched for SFs that could stimulate or antagonize the inclusion of  
25 alternative exons regulated by RNA helicases DDX17 and DDX5. These closely related paralog proteins  
26 belong to the large family of evolutionarily conserved DEAD-box ATP-dependent RNA helicases<sup>25</sup>.  
27 Previous reports, from our lab and others, have shown that DDX5 and DDX17 control the inclusion of  
28 a large number of exons by modulating the folding of their target transcripts, thanks to their helicase  
29 activity, and by modulating the binding of splicing regulators to RNA<sup>21,26-31</sup>. This is the case of HNRNPH  
30 for example, whose binding to RNA is facilitated by DDX5/DDX17<sup>21</sup>, and which was predicted by  
31 SplicingLore among the top predicted SFs for positive correlation with the helicases.

32  
33 SplicingLore predicted a good correlation in the regulation of several hundreds of exons by  
34 DDX5/DDX17 and HNRNPC, including exons that are either activated or repressed by those factors. We  
35 experimentally validated the additive effect of these factors on a subset of both classes of exons, in  
36 two different cell lines, and found that HNRNPC and DDX17 associate with each other in cells in an  
37 RNA-independent manner. The molecular nature of this functional relationship is not clear, and  
38 understanding how HNRNPC stimulates exon inclusion while it is often described as a splicing repressor  
39<sup>12</sup> will require further investigation.

1  
2  
3 In conclusion, SplicingLore represents a straightforward and reliable tool to investigate alternative  
4 splicing regulation by a large panel of SFs. SplicingLore also complements other web resources  
5 developed by our group, FasterDB <sup>7</sup> and Exon Ontology <sup>8</sup>, which together form a suite of tools to  
6 facilitate the understanding of alternative splicing, from its regulation to its biological consequences.  
7 SplicingLore will be useful not only to splicing specialists, but also to scientists less familiar with the  
8 complex mechanisms that underlie splicing regulation.  
9  
10  
11  
12  
13  
14  
15  
16

## 17 **Methods**

### 18 **Database and web interface**

19 SplicingLore database was implemented on Linux Ubuntu 22.04, using a MySQL database  
20 (www.mysql.com) and organized in 9 tables. The web interface was developed using the programming  
21 languages PHP, CSS and JavaScript. Tools performing statistical tests were driven in R (v4.2.0) using R  
22 packages Tidyverse ([doi.org/10.21105/joss.01686](https://doi.org/10.21105/joss.01686)) or in Python (v3.10), using Plotly <sup>32</sup> for graphics.  
23  
24  
25  
26  
27  
28

### 29 **Data collection**

30 SplicingLore is a database of 160 public datasets retrieved from GEO omnibus <sup>16</sup> and ENCODE  
31 consortium <sup>17</sup>, linked to a user-friendly web interface (Supplementary Table S1). All datasets were  
32 derived from Illumina RNA-seq experiments performed in 21 different human cell lines (293T, 786-O,  
33 A498, A673, Endoc-BH1, GM19238, HeLa, HepG2, SH-SY5Y, HMLE, Huh-7, K562, MCF7, MDA-MB-231,  
34 MG63, RD, RH30, RH41, RKO, LNCaP, LM2) in which the expression of 75 splicing factors was inhibited  
35 by siRNA or shRNA treatment. The quality control of the sequencing data was evaluated using FastQC  
36 (v0.11.9). Reads were trimmed using Prinseq-lite (v0.20.4) <sup>33</sup> (--trim-right 20) and filtered by average  
37 quality score (--trim-qual 20) and cutadapt (v1.18) <sup>34</sup>. Sequencing projects for which the filtered read  
38 count was lower than 20 millions pairs of reads were eliminated.  
39  
40  
41  
42  
43  
44  
45  
46

### 47 **Differential exon inclusion analysis**

48 For each dataset, we performed a qualitative and quantitative analysis of splicing variations using the  
49 FaRLine tool <sup>7</sup> (hg19 genome), comparing the condition in which the expression of the splicing factor  
50 was inhibited to the control condition. The following parameters were used to define exons that were  
51 differentially included:  $|\Delta\text{PSI}| \geq 10\%$ , adjusted  $p\text{-Val} \leq 0.05$ . The  $\Delta\text{PSI}$  corresponds to the change in  
52 exon inclusion (PSI: percent spliced-in) between two conditions. We then retrieved and fed the  
53 SplicingLore database with the lists of differentially included cassette exons corresponding to each  
54 dataset.  
55  
56  
57  
58  
59  
60

## Clustering analyses

Heatmaps of Supplementary Figures S1 and S2 were created with R (v4.2.0) using R packages Tidyverse (doi.org/10.21105/joss.01686) and pheatmap (<https://github.com/raivokolde/pheatmap>).

## Randomization test and scoring

- Input list of exons with  $\Delta$ PSI and  $p$ -values

We defined a Pearson correlation between the  $\Delta$ PSI of two sets of exons. The computed empirical  $p$ -value is the probability of observing a Pearson correlation as high or higher as the observed correlation when considering a set of randomly associated exons. This probability was computed from the empirical cumulative distribution function generated by computing the Pearson correlation for  $10^4$  random sets of exon (which leads to a maximal  $p$ -value resolution of  $10^{-4}$ ).

The "percent.sig.input" is the fraction of significantly regulated exons from the input list of exons that are also found in the list of exons regulated by the indicated SF. The "percent.sig.SF" is the fraction of significantly regulated exons from the list of exons regulated by the indicated SF that are also found in the input list of exons.

A confidence score (Score) was set up to facilitate the identification of candidate SF. It is linked to a permutation statistic test, which also gives a  $p$ -value, the "percent.sig.input" and the "percent.sig.SF". The Score ranges between 0 and 1, which indicates the degree of correlation or anti-correlation between the effect of a given SF and the input list of exons provided by the user.

$$Score = mean\left(1 - pvalue, \frac{percent.sig.input}{100}, \frac{percent.sig.sf}{100}\right)$$

- Basic input list of exons (chromosome coordinates only)

For an input set of exons  $E$ , randomization tests were performed to test if the number of exons  $N$  regulated by a given SF  $S$  is enriched or impoverished. For this,  $10^4$  sets of control exons with the same size as  $E$  were sampled. Then the number of exons regulated by  $S$  was computed for each control set. Finally, an empirical  $p$ -value was computed for  $S$  as:

$$P_{emp} = \frac{\min(k,l) + 1}{10,000 + 1}$$

where  $k$  is the number of controls sets with a number of exons regulated by  $S$  higher or equal to  $N$ , and  $l$  is the number of controls sets with a number of exon regulated by  $S$  lower or equal to  $N$ . This  $p$ -value was computed for each SF and was then corrected using the Benjamini-Hochberg procedure.



### Analysis of CLIP-seq data

Processed bed peak files from various CLIP-seq experiment projects, targeting 29 different SFs in different cell lines, were recovered from various sources, mainly POSTAR2<sup>35</sup> and ENCODE<sup>36</sup>, see Supplementary Sable S8. Peaks obtained from experiments targeting the same SF were merged together using BEDtools<sup>37</sup>. We then defined the sets of exons in the Splicing Lore database that were significantly more included or skipped by the 29 splicing factors that were knocked down in at least one sample. Exons that were significantly regulated in opposite ways by the same SF in different samples were removed. Factors with less than 100 significantly more included and skipped exons were discarded. The ratio  $R_i$  was then calculated for a splicing factor  $i$ :

$$R_i = \log_2\left(\frac{P_{reg} + 0.01}{P_{ctrl} + 0.01}\right)$$

Where  $P_{reg}$  is the proportion of exons more included or skipped upon knockdown of a given SF that overlap a peak of that SF and  $P_{ctrl}$  is the proportion of control exons that overlap the same peaks. Control exons correspond to all other exons detected by FaRLine in at least one project, excluding those significantly skipped or included upon knockdown of factor  $i$ . Finally, logistic regression was performed to test whether the proportions  $P_{reg}$  and  $P_{ctrl}$  were significantly different. All calculated  $p$ -values were corrected using the Benjamini-Hochberg procedure<sup>38</sup>. Non-significant  $R_i$  are shown in grey. The same method was applied to six 100-nucleotide regions surrounding regulated exons.

### Analysis of RNA binding motifs

The binding motifs preferentially recognized by 21 human SF were recovered from mCrossBase (<https://zhanglab.c2b2.columbia.edu/mCrossBase/>)<sup>39</sup> and cisBP-RNA Database (<http://cisbp-rna.cabr.utoronto.ca/>)<sup>40</sup>. Supplementary Table S9 describes the binding motifs used for further analysis and converted to MEME format. The set of sequences corresponding to significantly more included or skipped exons upon knockdown of each of these 21 SF was recovered. In addition, for each of these sets a control set of exon sequences was defined, corresponding to all other exons detected by FaRLine in at least one project, excluding those that were significantly regulated upon knockdown of the SF. These sequences were then extended by 200 nucleotides in both directions. A motif enrichment analysis was performed with the SEA program<sup>41</sup> using the binding motif, the extended sequences regulated by a given SF and control sequences. As a negative enrichment is not calculated by the SEA tool, motif impoverishment analysis was performed using the same principle, but with the regulated sequences as control and control sequences as primary sequences. The  $V$  value was then calculated for each set of exons up or down-regulated upon knockdown of a SF, using the following formula:

$$V = \min(1 - P_e, 1 - P_i) \times s \text{With} \begin{cases} s = -1 & \text{if } P_i < P_e \\ s = 1 & \text{if } P_i > P_e \end{cases}$$

Where  $P_e$  is the SEA  $p$ -value for positive enrichment of a binding motif in regulated sequences by a given SF and  $P_i$  is the same type of  $p$ -value but for negative enrichment. Finally, a heatmap was generated showing these  $V$ -values for exons that were up or down-regulated in each SF knockdown.  $V$ -values above and below 0.95 are shown in red and blue, respectively. Non-significant  $V$ -values (between -0.95 and 0.95) are shown in grey and  $V$ -values for SF with less than 100 exons regulated upon knockdown are shown in white.

### Cell culture and transfections

Human embryonic kidney 293T and SH-SY5Y cells were grown as recommended by the manufacturer and transfected as described previously<sup>26,42</sup>. For knockdown experiments, we used a total of 40 nM siRNA as follows : 40 nM siCtrl for control experiments, 20 nM of siDDX5/DDX17 or siHNRNPC + 20 nM siCtrl for single factor depletion, or 20 nM siDDX5/DDX17 + 20 nM siHNRNPC for double knockdown. Cells were harvested 48h later. Sequences of siRNAs are given in Supplementary Table S2.

### Co-immunoprecipitation and Western blotting

Total protein extraction was carried out as previously described<sup>21</sup>. Primary antibodies used for Western-blotting: DDX5 (ab10261, Abcam), DDX17 (ab24601, Abcam), HNRNPC (D6S3N, Cell Signaling), GAPDH (sc-1616, SantaCruz).

For co-immunoprecipitation, cells were harvested and gently lysed for 5 minutes on ice in a buffer containing 10 mM Tris-HCl pH 8.0, 140 mM NaCl, 1.5 mM MgCl<sub>2</sub>, 10 mM EDTA, 0.5% NP40, completed with protease and phosphatase inhibitors (Roche #11697498001 and #5892970001), to isolate the nuclei from the cytoplasm. After centrifugation, the nuclei were lysed in the IP buffer (20 mM Tris-HCl pH 7.5, 150 mM NaCl, 2 mM EDTA, 1% NP40, 10% Glycerol and protease/phosphatase inhibitors) for 30 minutes at 4°C under constant mixing. The nuclear lysate was centrifuged for 15 minutes to remove debris and soluble proteins were quantified by BCA (Thermo Fisher Scientific). The lysate was pre-cleared with 30 μL of Dynabeads Protein G (Thermo Fisher Scientific) for 30 minutes under rotatory mixing, and then split in 1.5 mg aliquots of proteins for each assay. Each fraction received 5 μg of antibody and the incubation was left overnight at 4°C under rotation. The following antibodies were used for IP: rabbit anti-DDX17 (19910-1-AP, ProteinTech) and anti-HNRNPC (PA522280, Thermo Fisher Scientific) or a control rabbit IgG (Thermo Fisher Scientific), goat anti-DDX5 (ab10261, Abcam) or control goat IgG (Santa Cruz). The next day, the different lysate/antibody mixtures were incubated with 50 μL Dynabeads Protein G (Thermo Fisher Scientific) blocked with bovine serum albumin, for 4

1  
2  
3 hours at 4°C under rotation. Bead were then washed 5 times with IP buffer. Elution was performed by  
4 boiling for 5 minutes in SDS-PAGE loading buffer prior to analysis by Western-blotting.  
5  
6  
7

### 8 **RNA extraction and PCR analyses**

9  
10 Total RNA were isolated using TriPure Isolation Reagent (Roche). For reverse transcription, 2 µg of  
11 purified RNAs were treated with Dnase I (Thermo Fisher Scientific) and retrotranscribed using Maxima  
12 reverse transcriptase (Thermo Fisher Scientific), as recommended by the manufacturer. Potential  
13 genomic DNA contamination was systematically verified by performing negative RT controls in absence  
14 of enzyme, and by including controls with water instead of cDNA in PCR assays. All PCR analyses were  
15 performed on 0.5 ng cDNA using 0.5 U GoTaq® DNA polymerase (Promega). Quantification of PCR  
16 products was performed using the Image Lab software (BioRad) after agarose gel electrophoresis. The  
17 PSI (percent spliced-in) value of each alternative exon was calculated in each condition using the  
18 following formula: inclusion product / (inclusion product + skipping product) x100. The ΔPSI  
19 corresponds to the difference between the PSI for each silencing condition and the PSI of the siCtrl  
20 condition. Sequences of all primers are given in Supplementary Table S2.  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31

### 32 **Data availability**

33 All raw data supporting the findings of this study (Supplementary Table 1) were retrieved from GEO  
34 omnibus<sup>16</sup> and ENCODE consortium<sup>17</sup>. All processed data can be freely retrieved from the SplicingLore  
35 database.  
36  
37  
38  
39  
40  
41

### 42 **Code availability**

43 The code used to process the data presented in the manuscript is available here:

44 [https://github.com/helenepolveche/splicinglore\\_website](https://github.com/helenepolveche/splicinglore_website)  
45  
46  
47  
48  
49  
50

### 51 **References**

- 52  
53 1 Kastner, B., Will, C. L., Stark, H. & Luhrmann, R. Structural Insights into Nuclear pre-mRNA  
54 Splicing in Higher Eukaryotes. *Cold Spring Harb Perspect Biol* **11**,  
55 doi:10.1101/cshperspect.a032417 (2019).  
56 2 Wilkinson, M. E., Charenton, C. & Nagai, K. RNA Splicing by the Spliceosome. *Annu Rev Biochem*  
57 **89**, 359-388, doi:10.1146/annurev-biochem-091719-064225 (2020).  
58 3 Marasco, L. E. & Kornblihtt, A. R. The physiology of alternative splicing. *Nature reviews.*  
59 *Molecular cell biology* **24**, 242-254, doi:10.1038/s41580-022-00545-z (2023).  
60

- 1  
2  
3 4 Wright, C. J., Smith, C. W. J. & Jiggins, C. D. Alternative splicing as a source of phenotypic  
4 diversity. *Nature Reviews Genetics* **23**, 697-710, doi:10.1038/s41576-022-00514-4 (2022).
- 5 5 Tapial, J. *et al.* An atlas of alternative splicing profiles and functional associations reveals new  
6 regulatory programs and genes that simultaneously express multiple major isoforms. *Genome*  
7 *research* **27**, 1759-1768, doi:10.1101/gr.220962.117 (2017).
- 8 6 Hyung, D., Kim, J., Cho, S. Y. & Park, C. ASpedia: a comprehensive encyclopedia of human  
9 alternative splicing. *Nucleic acids research* **46**, D58-D63, doi:10.1093/nar/gkx1014 (2018).
- 10 7 Mallinjou, P. *et al.* Endothelial, epithelial, and fibroblast cells exhibit specific splicing programs  
11 independently of their tissue of origin. *Genome research* **24**, 511-521,  
12 doi:10.1101/gr.162933.113 (2014).
- 13 8 Tranchevent, L. C. *et al.* Identification of protein features encoded by alternative exons using  
14 Exon Ontology. *Genome research* **27**, 1087-1097, doi:10.1101/gr.212696.116 (2017).
- 15 9 Witten, J. T. & Ule, J. Understanding splicing regulation through RNA splicing maps. *Trends in*  
16 *genetics : TIG* **27**, 89-97, doi:10.1016/j.tig.2010.12.001 (2011).
- 17 10 Dominguez, D. *et al.* Sequence, Structure, and Context Preferences of Human RNA Binding  
18 Proteins. *Molecular cell* **70**, 854-867 e859, doi:10.1016/j.molcel.2018.05.001 (2018).
- 19 11 Yee, B. A., Pratt, G. A., Graveley, B. R., Van Nostrand, E. L. & Yeo, G. W. RBP-Maps enables robust  
20 generation of splicing regulatory maps. *RNA* **25**, 193-204, doi:10.1261/rna.069237.118 (2019).
- 21 12 Van Nostrand, E. L. *et al.* A large-scale binding and functional map of human RNA-binding  
22 proteins. *Nature* **583**, 711-719, doi:10.1038/s41586-020-2077-3 (2020).
- 23 13 Fu, X. D. & Ares, M., Jr. Context-dependent control of alternative splicing by RNA-binding  
24 proteins. *Nature reviews. Genetics* **15**, 689-701, doi:10.1038/nrg3778 (2014).
- 25 14 Ule, J. & Blencowe, B. J. Alternative Splicing Regulatory Networks: Functions, Mechanisms, and  
26 Evolution. *Molecular cell* **76**, 329-345, doi:10.1016/j.molcel.2019.09.017 (2019).
- 27 15 Jiang, W. & Chen, L. Alternative splicing: Human disease and quantitative analysis from high-  
28 throughput sequencing. *Comput Struct Biotechnol J* **19**, 183-195, doi:10.1016/j.csbj.2020.12.009  
29 (2021).
- 30 16 Edgar, R., Domrachev, M. & Lash, A. E. Gene Expression Omnibus: NCBI gene expression and  
31 hybridization array data repository. *Nucleic acids research* **30**, 207-210, doi:DOI  
32 10.1093/nar/30.1.207 (2002).
- 33 17 Dunham, I. *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature*  
34 **489**, 57-74, doi:10.1038/nature11247 (2012).
- 35 18 Benoit-Pilven, C. *et al.* Complementarity of assembly-first and mapping-first approaches for  
36 alternative splicing annotation and differential analysis from RNAseq data. *Scientific reports* **8**,  
37 4307, doi:10.1038/s41598-018-21770-7 (2018).
- 38 19 Zarnack, K. *et al.* Direct Competition between hnRNP C and U2AF65 Protects the Transcriptome  
39 from the Exonization of Alu Elements. *Cell* **152**, 453-466, doi:10.1016/j.cell.2012.12.023 (2013).
- 40 20 Howard, J. M. & Sanford, J. R. The RNAissance family: SR proteins as multifaceted regulators of  
41 gene expression. *Wiley interdisciplinary reviews. RNA* **6**, 93-110, doi:10.1002/wrna.1260 (2015).
- 42 21 Dardenne, E. *et al.* RNA helicases DDX5 and DDX17 dynamically orchestrate transcription,  
43 miRNA, and splicing programs in cell differentiation. *Cell reports* **7**, 1900-1913,  
44 doi:10.1016/j.celrep.2014.05.010 (2014).
- 45 22 Barash, Y. *et al.* Deciphering the splicing code. *Nature* **465**, 53-59, doi:10.1038/nature09000  
46 (2010).
- 47 23 Fontrodona, N. *et al.* Interplay between coding and exonic splicing regulatory sequences.  
48 *Genome research* **29**, 711-722, doi:10.1101/gr.241315.118 (2019).
- 49 24 Lemaire, S. *et al.* Characterizing the interplay between gene nucleotide composition bias and  
50 splicing. *Genome biology* **20**, 259, doi:10.1186/s13059-019-1869-y (2019).
- 51 25 Bourgeois, C. F., Mortreux, F. & Auboeuf, D. The multiple functions of RNA helicases as drivers  
52 and regulators of gene expression. *Nature reviews. Molecular cell biology* **17**, 426-438,  
53 doi:10.1038/nrm.2016.50 (2016).
- 54  
55  
56  
57  
58  
59  
60

- 1  
2  
3 26 Ameur, L. B. *et al.* Intragenic recruitment of NF-kappaB drives splicing modifications upon  
4 activation by the oncogene Tax of HTLV-1. *Nature communications* **11**, 3045,  
5 doi:10.1038/s41467-020-16853-x (2020).  
6  
7 27 Camats, M., Guil, S., Kokolo, M. & Bach-Elias, M. P68 RNA helicase (DDX5) alters activity of cis-  
8 and trans-acting factors of the alternative splicing of H-Ras. *PloS one* **3**, e2926,  
9 doi:10.1371/journal.pone.0002926 (2008).  
10  
11 28 Dardenne, E. *et al.* Splicing switch of an epigenetic regulator by RNA helicases promotes tumor-  
12 cell invasiveness. *Nature structural & molecular biology* **19**, 1139-1146, doi:10.1038/nsmb.2390  
13 (2012).  
14  
15 29 Kar, A. *et al.* RNA helicase p68 (DDX5) regulates tau exon 10 splicing by modulating a stem-loop  
16 structure at the 5' splice site. *Molecular and cellular biology* **31**, 1812-1821,  
17 doi:10.1128/MCB.01149-10 (2011).  
18  
19 30 Lambert, M. P. *et al.* The RNA helicase DDX17 controls the transcriptional activity of REST and  
20 the expression of proneural microRNAs in neuronal differentiation. *Nucleic acids research* **46**,  
21 7686-7700, doi:10.1093/nar/gky545 (2018).  
22  
23 31 Lee, Y. J., Wang, Q. & Rio, D. C. Coordinate regulation of alternative pre-mRNA splicing events  
24 by the human RNA chaperone proteins hnRNPA1 and DDX5. *Genes & development* **32**, 1060-  
25 1074, doi:10.1101/gad.316034.118 (2018).  
26  
27 32 Sievert, C. Interactive web-based data visualization with R, plotly, and shiny. online resource  
28 (CRC Press, Taylor and Francis Group, Boca Raton, FL) (2020).  
29  
30 33 Schmieder, R. & Edwards, R. Quality control and preprocessing of metagenomic datasets.  
31 *Bioinformatics* **27**, 863-864, doi:10.1093/bioinformatics/btr026 (2011).  
32  
33 34 Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads.  
34 *EMBnet.journal* **17**, 10-12, doi:<https://doi.org/10.14806/ej.17.1.200> (2011).  
35  
36 35 Zhu, Y. *et al.* POSTAR2: deciphering the post-transcriptional regulatory logics. *Nucleic acids*  
37 *research* **47**, D203-D211, doi:10.1093/nar/gky830 (2019).  
38  
39 36 Luo, Y. *et al.* New developments on the Encyclopedia of DNA Elements (ENCODE) data portal.  
40 *Nucleic acids research* **48**, D882-D889, doi:10.1093/nar/gkz1062 (2020).  
41  
42 37 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features.  
43 *Bioinformatics* **26**, 841-842, doi:10.1093/bioinformatics/btq033 (2010).  
44  
45 38 Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate - a Practical and Powerful  
46 Approach to Multiple Testing. *J R Stat Soc B* **57**, 289-300, doi:DOI 10.1111/j.2517-  
47 6161.1995.tb02031.x (1995).  
48  
49 39 Feng, H. *et al.* Modeling RNA-Binding Protein Specificity In Vivo by Precisely Registering Protein-  
50 RNA Crosslink Sites. *Molecular cell* **74**, 1189-1204 e1186, doi:10.1016/j.molcel.2019.02.002  
51 (2019).  
52  
53 40 Ray, D. *et al.* A compendium of RNA-binding motifs for decoding gene regulation. *Nature* **499**,  
54 172-177, doi:10.1038/nature12311 (2013).  
55  
56 41 Bailey, T. L. & Grant, C. E. SEA: Simple Enrichment Analysis of motifs. *bioRxiv*,  
57 2021.2008.2023.457422, doi:10.1101/2021.08.23.457422 (2021).  
58  
59 42 Terrone, S. *et al.* RNA helicase-dependent gene looping impacts messenger RNA processing.  
60 *Nucleic acids research* **50**, 9226-9246, doi:10.1093/nar/gkac717 (2022).

## Author contributions

Conceptualization: H.P., D.A. and C.F.B.; methodology and statistical analysis: H.P and N.F.; data collection and processing: H.P., A.L. and N.F.; experimental validation and interpretation: J.V., V.C. and

1  
2  
3 C.F.B.; software: H.P., N.F. and S.J.; writing of the original draft: H.P. and C.F.B.; review & editing: H.P.,  
4 F.M., D.A. and C.F.B.  
5  
6  
7  
8  
9

## 10 **Acknowledgements**

11 I-Stem is part of the Biotherapies Institute for Rare Diseases (BIRD) supported by the Association  
12 Française contre les Myopathies (AFM-Téléthon). We gratefully acknowledge support from the PSMN  
13 (Pôle Scientifique de Modélisation Numérique) of the ENS of Lyon for computing resources. We also  
14 thank Laurent Modolo and other members of the LBMC Biocomputing Hub for helpful discussions. This  
15 work was supported by grants from the Ligue contre le Cancer ("Equipe Labellisée"). J.V. received  
16 doctoral fellowships from the french Ministry of Research and Education and from the Ligue contre  
17 le Cancer. A.L. was supported by a doctoral fellowship from AFM-Téléthon.  
18  
19  
20  
21  
22  
23  
24  
25  
26

## 27 **Competing interests**

28 The authors declare no competing interests.  
29  
30  
31  
32  
33

## 34 **Figure legends**

35 **Figure 1.** General features of SplicingLore.  
36

37 **A.** Outline of the SplicingLore workflow. RNA-seq datasets from GEO and ENCODE databases were  
38 uniformly processed with FaRLine to generate lists of alternative exons regulated by 75 splicing factors  
39 in 21 human cell lines. These datasets can be used to compute correlation scores of exon regulation  
40 with a list of exons of interest, which is either retrieved from the database or provided by the user  
41 (query Method 1). The user can alternatively query SplicingLore with a list of exons associated only  
42 with their chromosomal coordinates (Method 2). In the latter case, the list of exons can include or not  
43 a  $\Delta$ PSI and *p*-value.  
44  
45  
46  
47  
48

49 **B.** Diagram showing the repartition of the SplicingLore datasets between cell lines.  
50  
51  
52

53 **Figure 2.** Comparison with splicing factor binding sites.

54 **A.** Enrichment of SF binding sites (based on CLIP-seq data) at exons that are up or down regulated by  
55 the corresponding factor, and in their flanking regions. Each square on both sides of the exon  
56 represents a 100 nucleotide window upstream and upstream of the exon.  
57  
58  
59  
60

1  
2  
3 **B. Enrichment of SF-specific binding sequences at exons (+/- 200 nucleotides) that are up or down**  
4 **regulated by the corresponding factor. Non significant enrichment values are represented by grey**  
5 **squares. Squares are white when less than 100 exons were identified by FaRLine as more included or**  
6 **excluded upon knockdown of the corresponding SF.**  
7  
8  
9

10  
11 **Figure 3.** Visualization of the SplicingLore web Interface.  
12

13 **A.** View of the home page, with the instructions to prepare the input list of exons and the query  
14 window. This includes the drop-down menu that allows to retrieve a given list of SF-regulated exons  
15 from the database.  
16

17  
18 **B.** View of the top of the main output file, showing for each dataset the different correlation  
19 parameters with the query list (here the exons regulated by DDX5/DDX17 in 293T cells). The Score and  
20 associated *p*-value for positive or negative correlation are provided for each dataset of the database,  
21 along with the total number of exons in the corresponding dataset and the fraction of common exons  
22 between the compared datasets, relative to the query list or to the tested dataset.  
23

24  
25 **C.** Correlation graph of the  $\Delta$ PSI values for common exons from the query list and one of the stored  
26 datasets. Orange dots represent exons that are significantly regulated by both SF. The file containing  
27 the corresponding  $\Delta$ PSI values for these exons can be downloaded here.  
28

29  
30 **D.** View of the top of the detailed output file, which provides the number (#) and fraction (%) of shared  
31 exons between the query and the dataset, separated in up-regulated and down-regulated sub-classes  
32 of exons.  
33  
34  
35  
36  
37

38 **Figure 4.** Validation of a functional relationship between DDX5/DDX17 and HNRNPC.  
39

40 **A.** Western-blot showing the expression of DDX5, DDX17 and HNRNPC following treatment with siRNA  
41 targeted against luciferase (negative control), DDX5/DDX17 and HNRNPC. Quantification of this  
42 experiment is shown in Supplementary Figure S7A.  
43

44  
45 **B.** RT-PCR analysis showing the inclusion of a selection of alternative exons after the depletion of  
46 DDX5/DDX17 and/or HNRNPC in 293T cells. The corresponding gene and exon number (according to  
47 FasterDB annotation) is indicated. Exons down-regulated and up-regulated upon SF knock-down are  
48 represented in red and blue, respectively.  
49

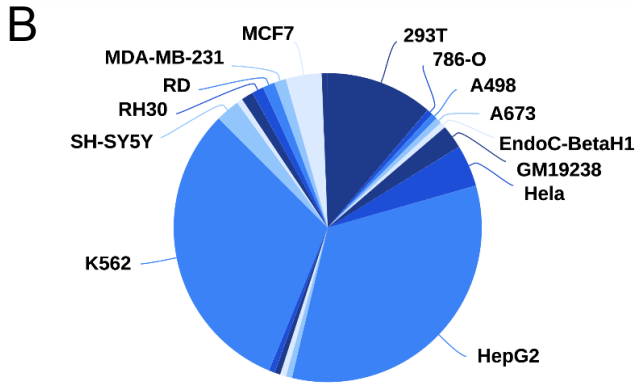
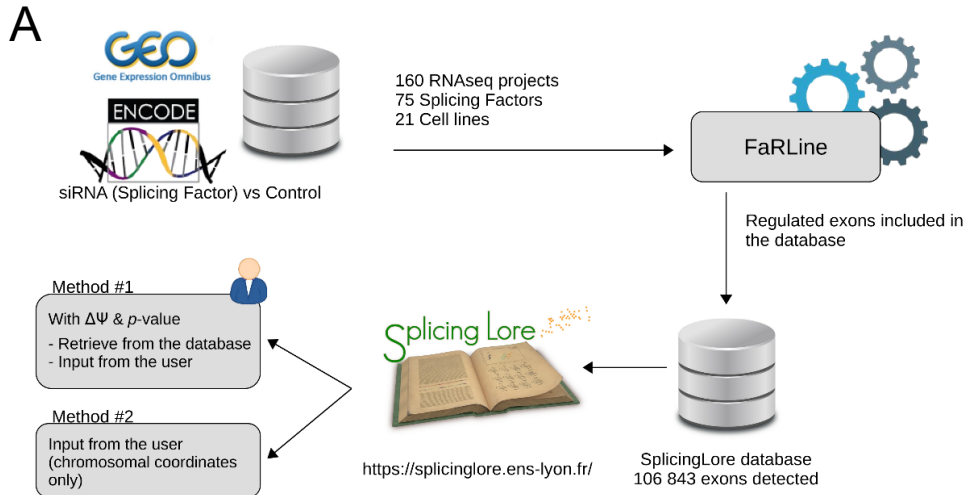
50  
51 **C.** Quantification of the RT-PCR experiment shown in C. The indicated  $\Delta$ PSI values correspond to the  
52 difference between the PSI (percent spliced-in) value of each depleted sample and the control sample.  
53 Statistical comparison between each condition (including the unshown control condition) was  
54 calculated using a one-way ANOVA (Holm–Sidak’s multiple comparison tests: \* *P*-val < 0.05; \*\* *P*-val <  
55 0.01; \*\*\* *P*-val < 0.001).  
56  
57  
58  
59  
60

1  
2  
3 **D.** Co-immunoprecipitation assays between endogenous HNRNPC and DDX17 in 293T cells, in the  
4 absence of presence of RNase A.  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

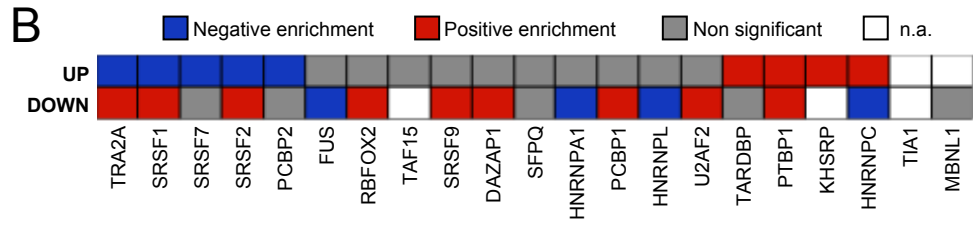
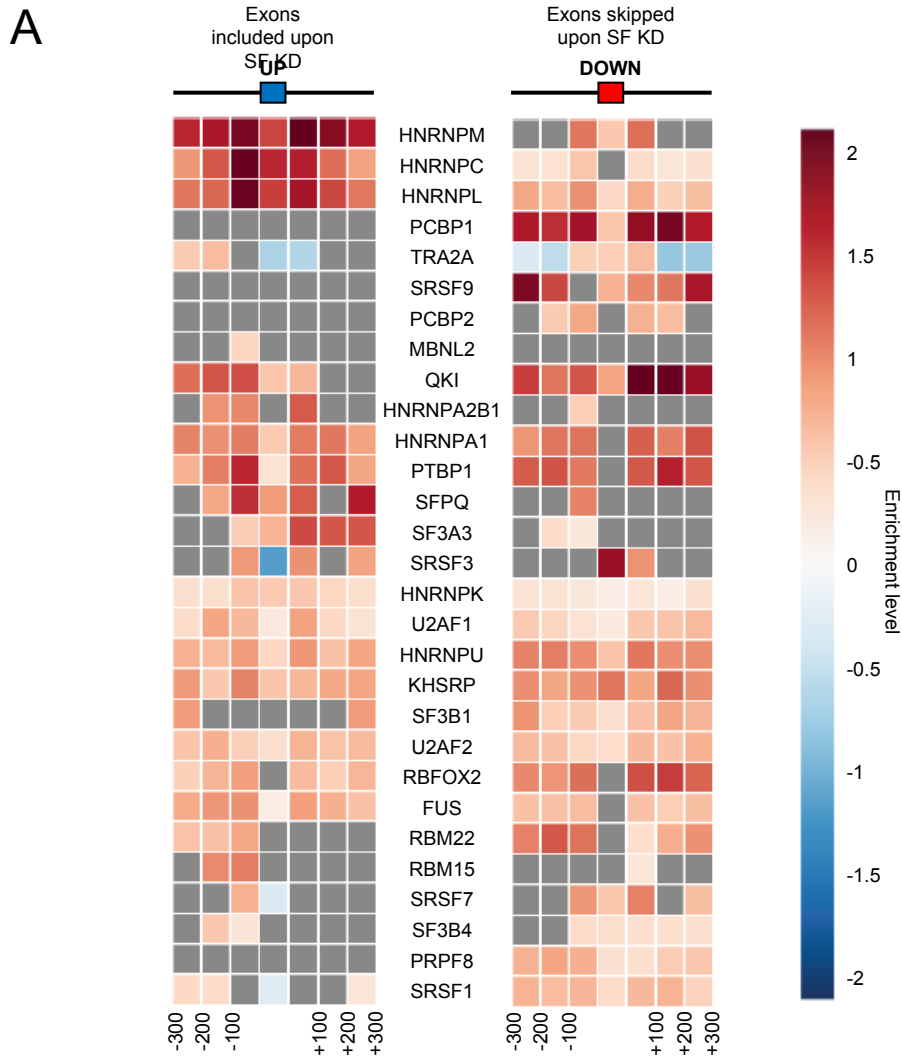
For Peer Review



# Figure 1



# Figure 2

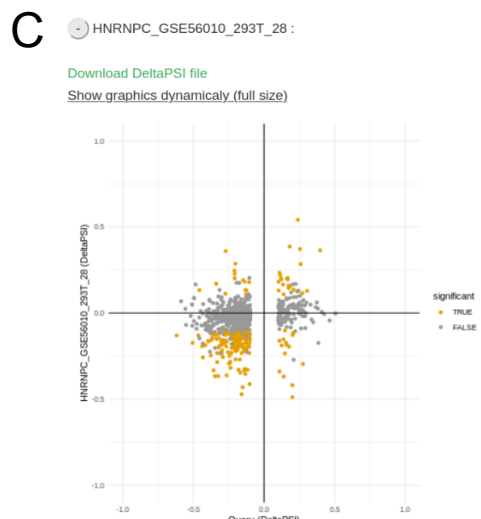


# Figure 3

**A**

**B**

Projects	Score	p-value	Direction	#exons	percent.sig.input	percent.sig.SF
DDX5_DDX17_GSE123752_293T_80	0.44	1e-04	p-value < X / correlate	1391	100	100
DDX5_DDX17_GSE98871_SH.SY5Y_79	0.50	1e-04	p-value < X / correlate	1126	53.49	24.47
DDX5_DDX17_LBMC_MCF7_172	0.53	1e-04	p-value < X / correlate	950	37.38	20.37
U2AF1_E_MTAB.2682_293T_66	0.47	0.0034	p-value = X / correlate	1069	28.04	12.49
HNRNPG.RBMX_KD1_GSE74085_293T_201	0.48	1e-04	p-value < X / correlate	1135	13.95	19.62
HNRNPC_GSE56010_293T_28	0.44	4e-04	p-value = X / correlate	1125	13.73	16.87
HNRNPC_ENC SR634KBO_K562_128	0.44	1e-04	p-value < X / correlate	705	14.52	17.34
HNRNPC_GSE63086_MG63_188	0.44	0.002	p-value = X / correlate	1012	16.53	15.02
HNRNPH1_GSE104559_RH30_214	0.43	5e-04	p-value = X / correlate	803	11.07	16.96
HNRNPH1_GSE104559_RH41_216	0.43	4e-04	p-value = X / correlate	1018	10.78	19.28
RBFOX2_GSE60392_293T_33	0.42	2e-04	p-value = X / correlate	1276	6.33	18.97
SRSF9_ENC SR597XHH_HepG2_110	0.42	0.004	p-value = X / correlate	688	4.03	22.4
HNRNPH1_GSE104559_RD_212	0.42	0.0142	p-value = X / correlate	927	11.86	16.72
HNRNPH1_GSE104559_RH41_215	0.42	0.0011	p-value = X / correlate	1025	9.56	17.14
U2AF1_ENC SR372UWV_HepG2_105	0.41	0.0034	p-value = X / correlate	717	14.16	10
HNRNPH1_GSE104559_RH30_213	0.41	0.0192	p-value = X / correlate	838	8.7	17.69

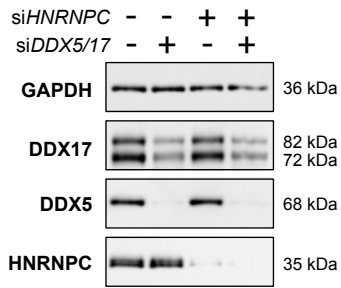


**D**

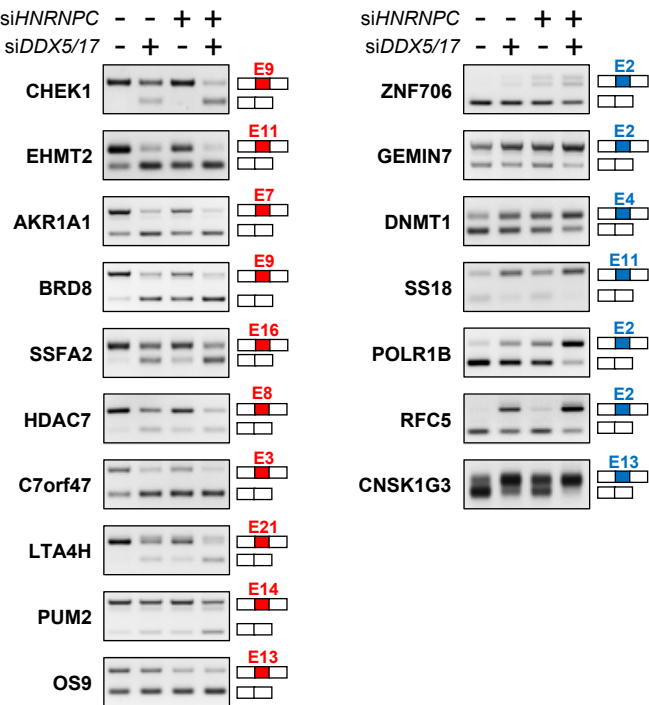
Projects	UP		UP		DOWN		DOWN		#exons	#exons common	% common	% significant	% significant SplicingFactor
	% negative corr	% positive corr	% negative corr	% positive corr	% negative corr	% positive corr	% negative corr	% positive corr					
DDX5_DDX17_GSE123752_293T_80	0	100	0	206	0	100	0	1185	1391	1391	100	100	
DDX5_DDX17_GSE98871_SH.SY5Y_79	0.49	32.04	1	66	0	57.13	0	677	1126	744	53.49	24.47	
DDX5_DDX17_LBMC_MCF7_172	1.46	20.39	3	42	0.25	39.83	3	472	950	520	37.38	20.37	
U2AF1_E_MTAB.2682_293T_66	11.65	9.71	24	20	2.36	26.84	28	318	1069	390	28.04	12.49	
HNRNPC_GSE63086_MG63_188	9.71	7.77	20	16	0.76	15.61	9	185	1012	230	16.53	15.02	
HNRNPC_ENC SR634KBO_K562_128	4.37	11.65	9	24	0.51	13.76	6	163	705	202	14.52	17.34	
U2AF1_ENC SR372UWV_HepG2_105	6.31	6.31	13	13	1.69	12.74	20	151	717	197	14.16	10	
HNRNPG.RBMX_KD1_GSE74085_293T_201	4.37	4.37	9	9	1.18	13.67	14	162	1135	194	13.95	19.62	
HNRNPC_GSE56010_293T_28	7.28	11.17	15	23	1.18	11.73	14	139	1125	191	13.73	16.87	
AQR_ENC SR624OUL_K562_225	4.85	7.77	10	16	0.68	12.15	8	144	817	178	12.8	8.64	
HNRNPH1_GSE104559_RD_212	4.37	6.31	9	13	1.69	10.38	20	123	927	165	11.86	16.72	

# Figure 4

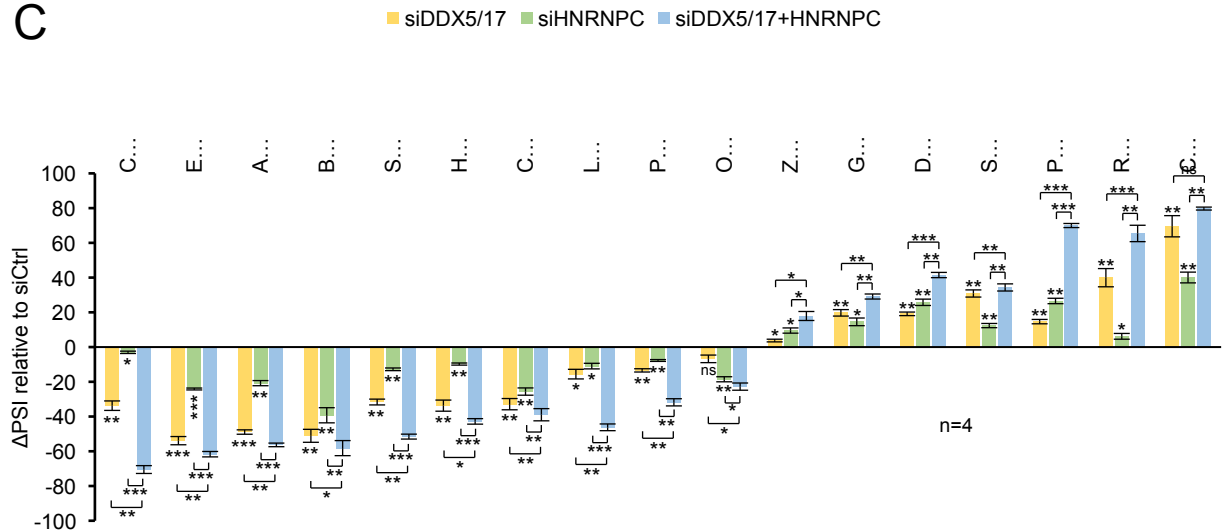
## A



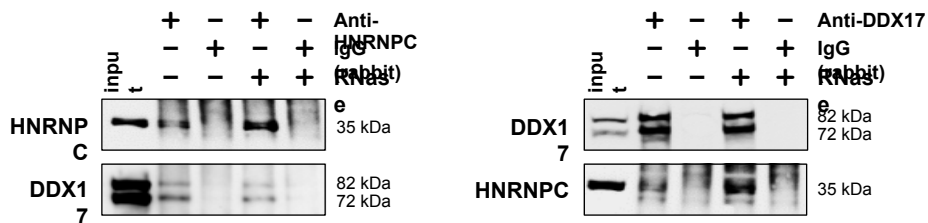
## B



## C



## D

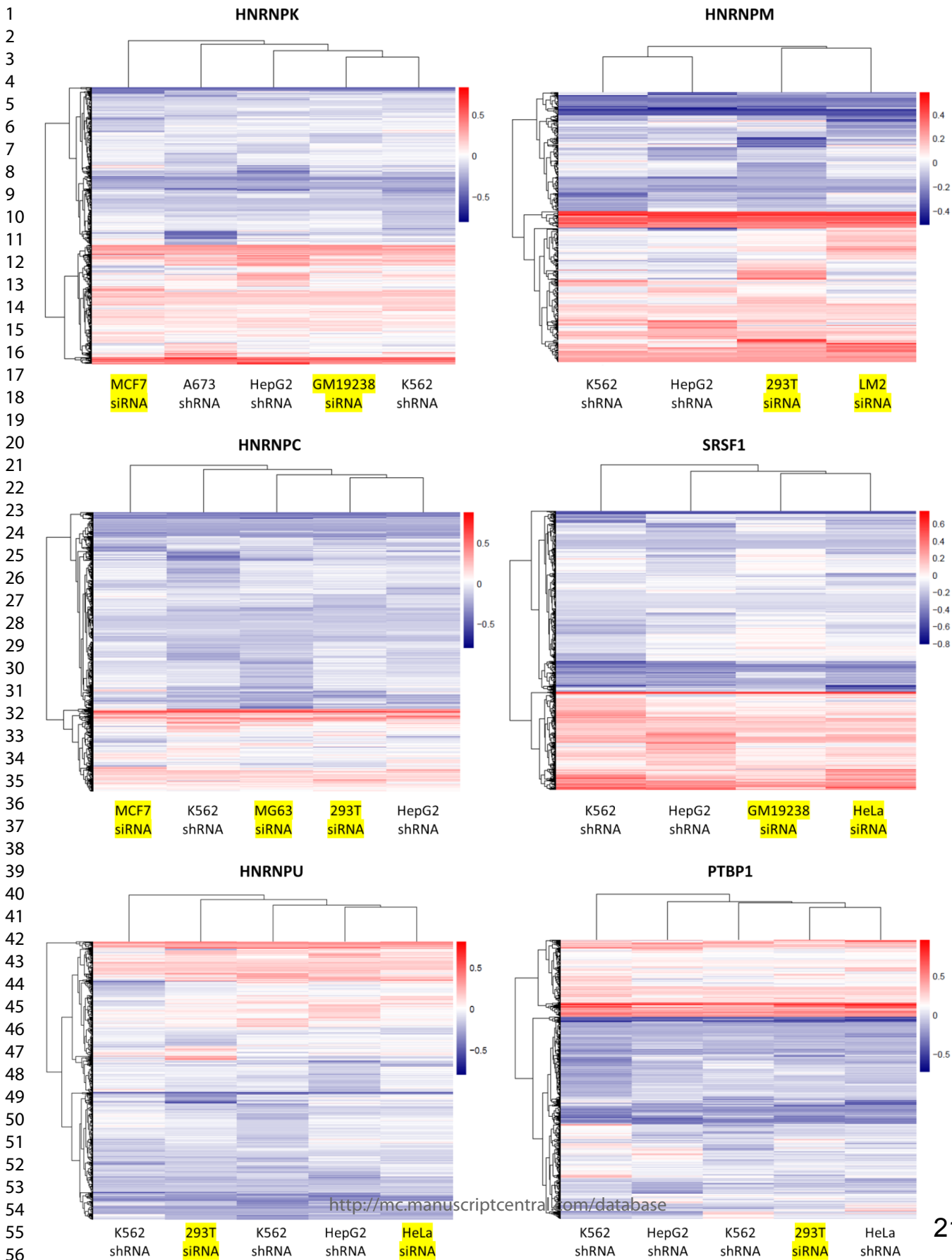


# Supplementary Figures

## Supplementary Figure S1 (next pages)

Heatmap analyses of the DeltaPSI of exons regulated by a given splicing factor in at least 3 different cell lines. The list of exons includes all exons that were significantly regulated in at least 1 cell lines (according to FaRLine thresholds). The cell line and mode of knockdown is indicated (datasets corresponding to a siRNA-mediated knockdown are highlighted in yellow).

# Supplementary Figure S1

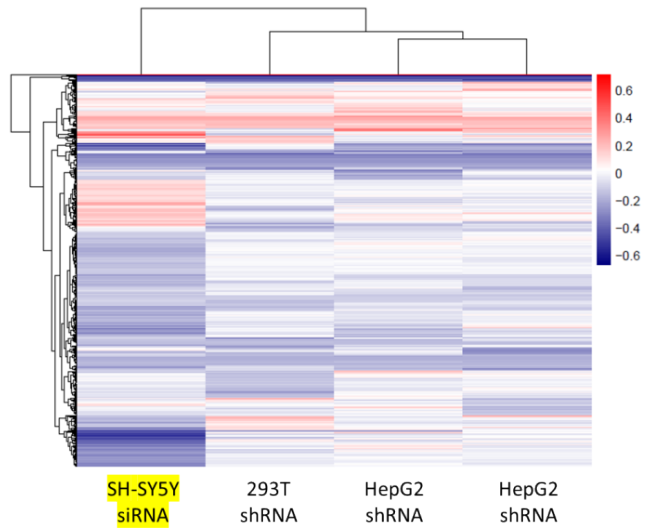
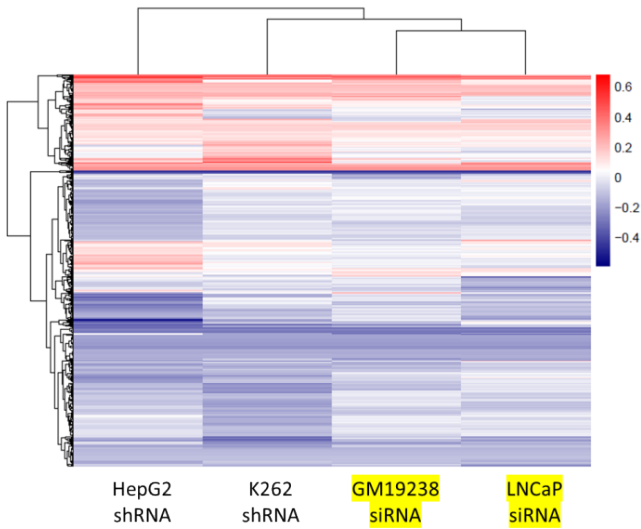


# Supplementary Figure S1

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56

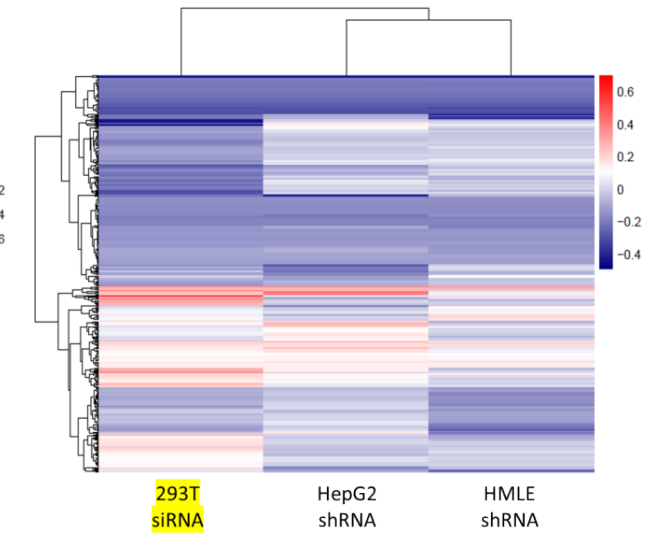
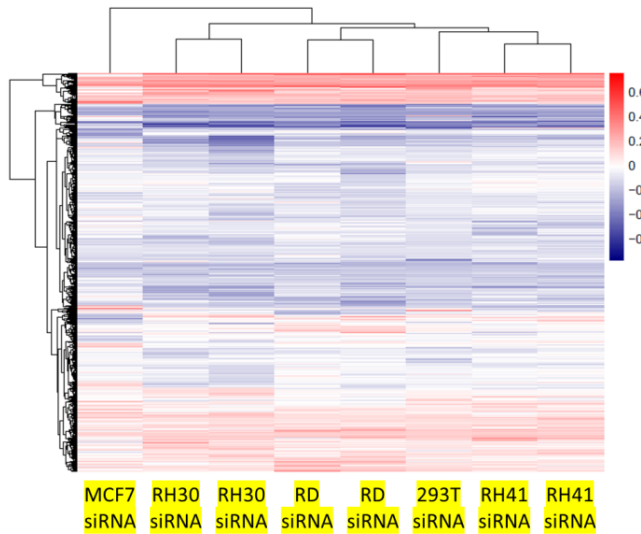
**HNRNPL**

**DAZAP1**



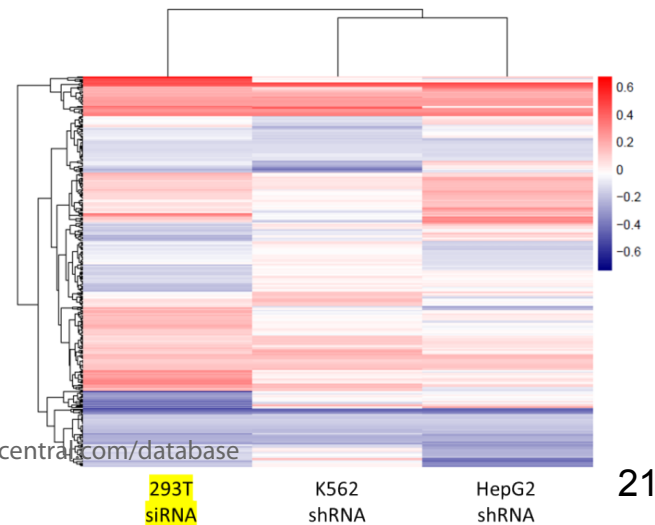
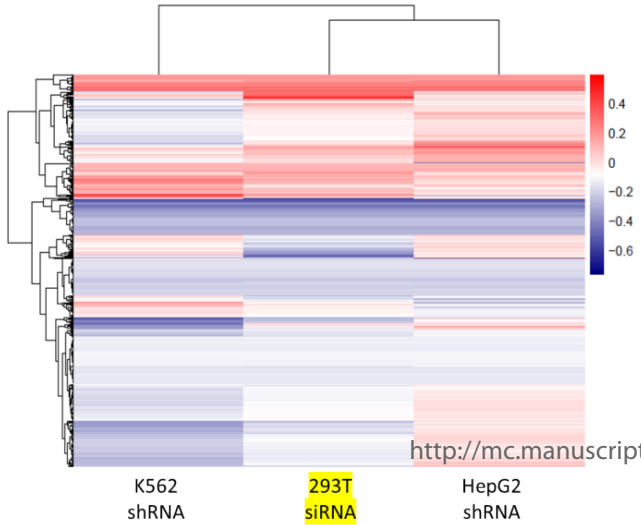
**HNRNPH1**

**HNRNPF**



**HNRNPA2B1**

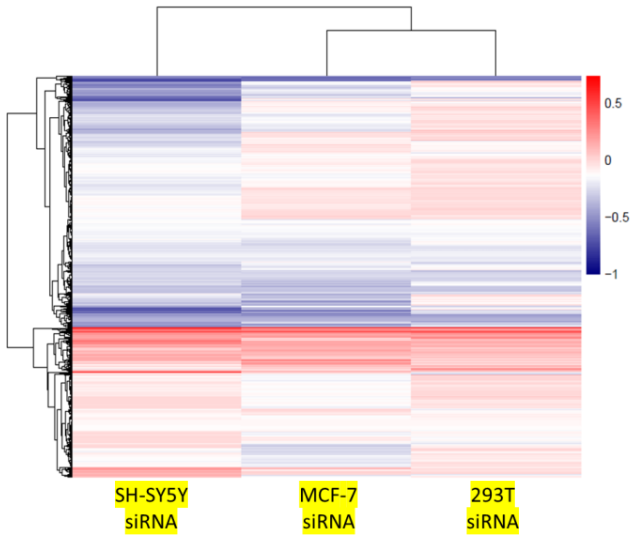
**HNRNPA1**



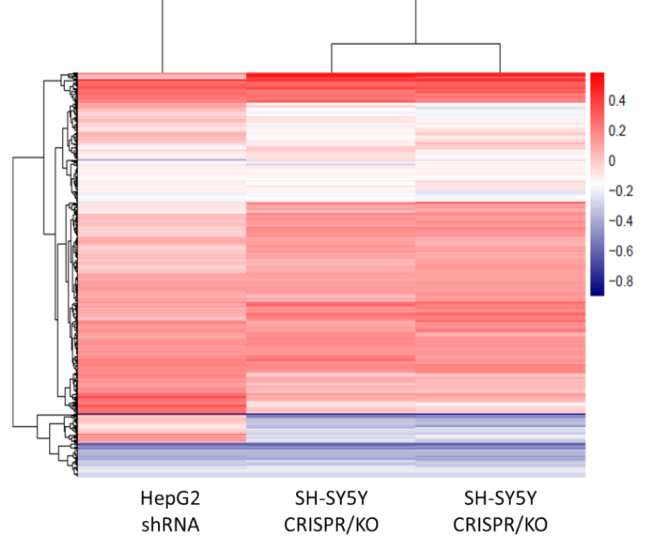
# Supplementary Figure S1

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56

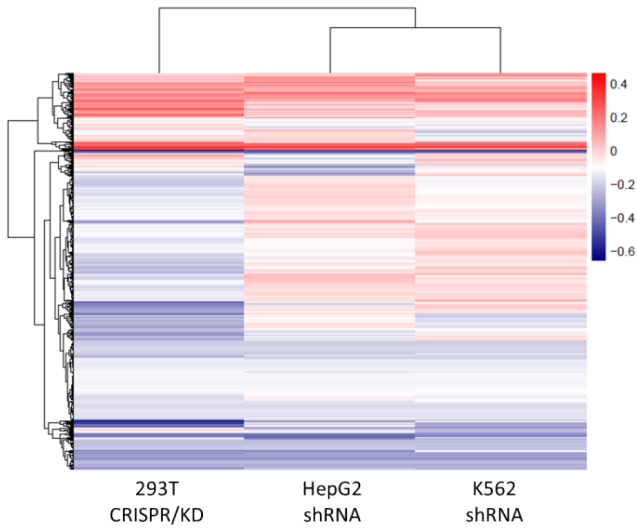
**DDX5/DDX17**



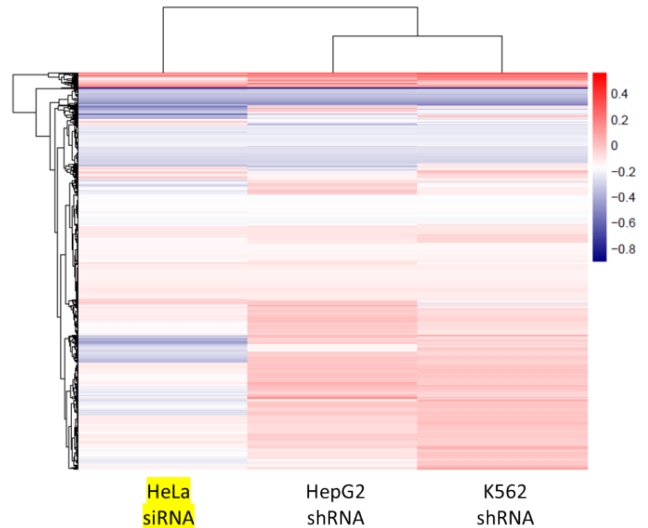
**FUS/TLS**



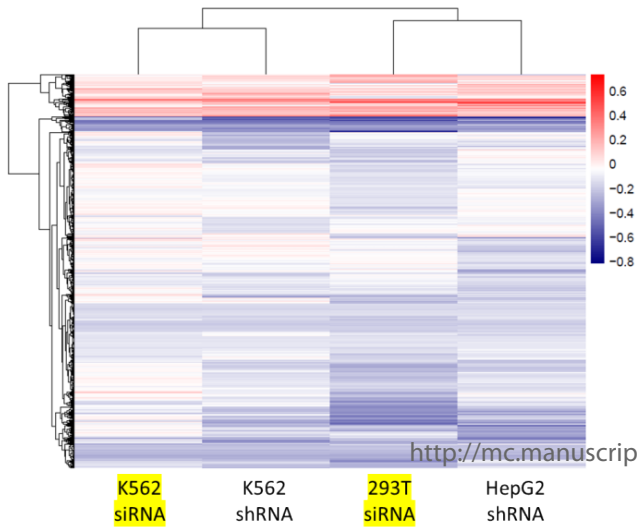
**RBM25**



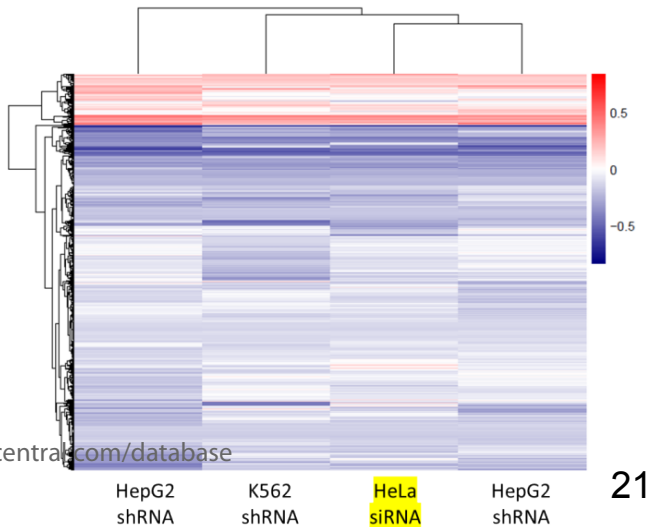
**SF3B1**



**U2AF1**

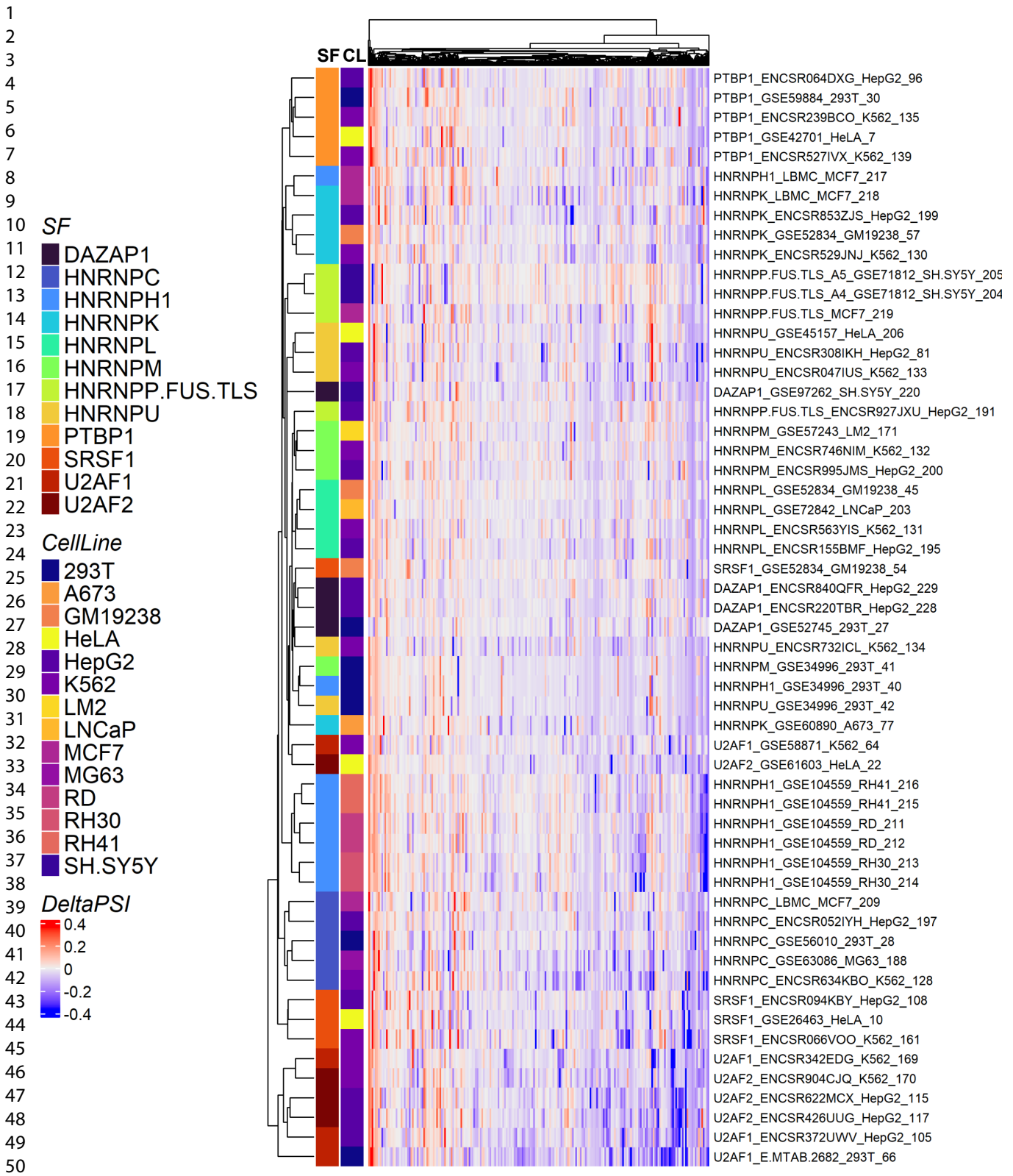


**U2AF2**





# Supplementary Figure S2



Heatmap analysis of the DeltaPSI of exons regulated by all SFs for which at least 3 different cell lines are available. The list of exons includes all exons that are significantly regulated in at least 2 cell lines for each SF.

## Supplementary Figure S3

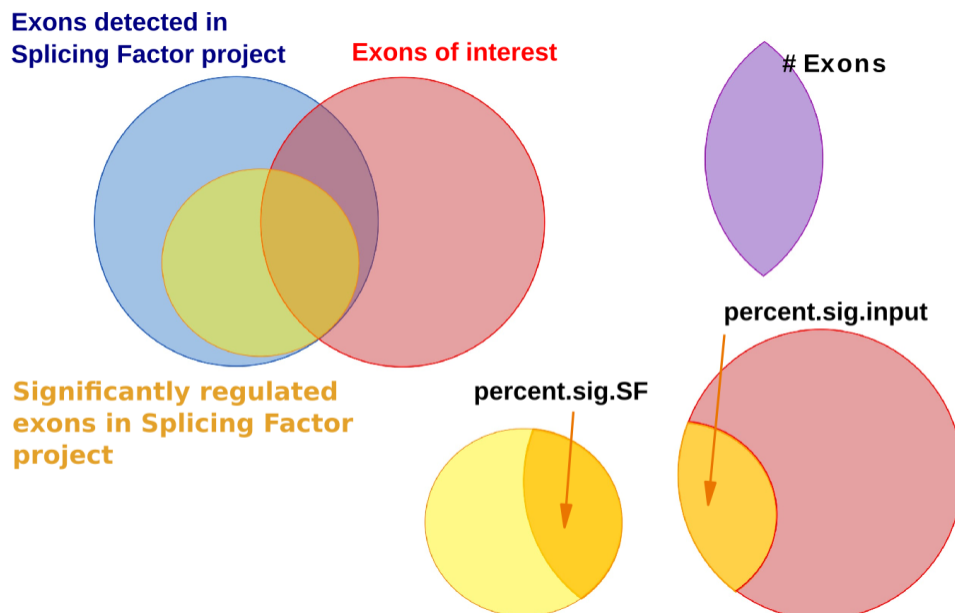
List of Exons Scores Graphics Details Up/Down

- The number of cells is **not equal to 6** for the row(s):  
PRMT7\_19;16;68390604;68390700;8.86575e-09

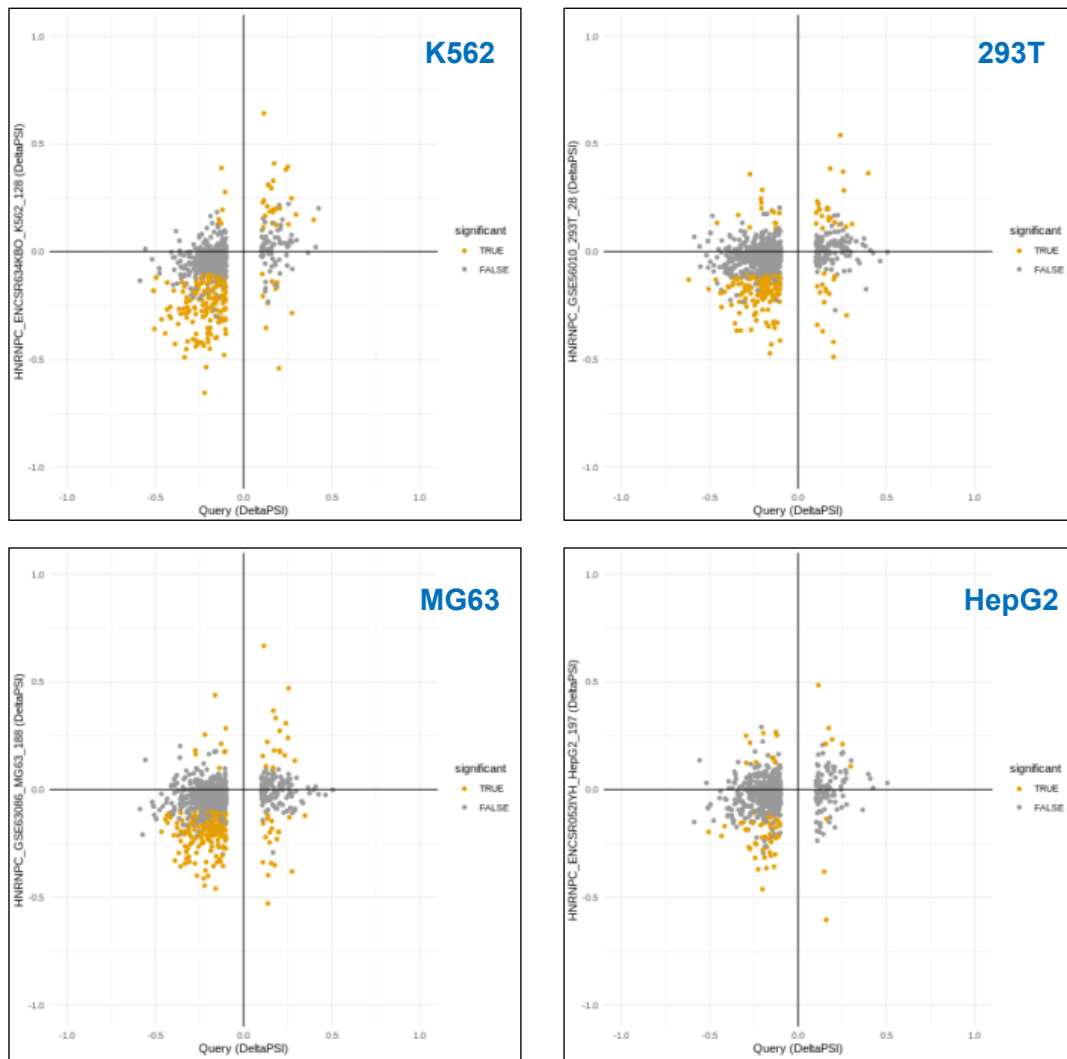
- The name **KIAA146\_25** does not exist. The chromosomal coordinates **18:59947593-59947675** indicate as exon name : **KIAA1468\_25**  
- There is no exon detected on these coordinates **17:7230480-7230549 ( NEURL4\_3 )** in our database.

Exon Name	chr	start	end	DeltaPSI	p-value
HMG20B_8	19	3576890	3577105	-0.200238	1.02413e-06
FLNA_30	X	153585619	153585642	0.208556	0.0182072
CCNE1_9	19	30312903	30313037	-0.506494	2.99897e-10
LRRC51_2	11	71799343	71799426	-0.143145	0.0397017
DOK1_5	2	74783021	74783205	-0.393464	0.000262855

## Supplementary Figure S4

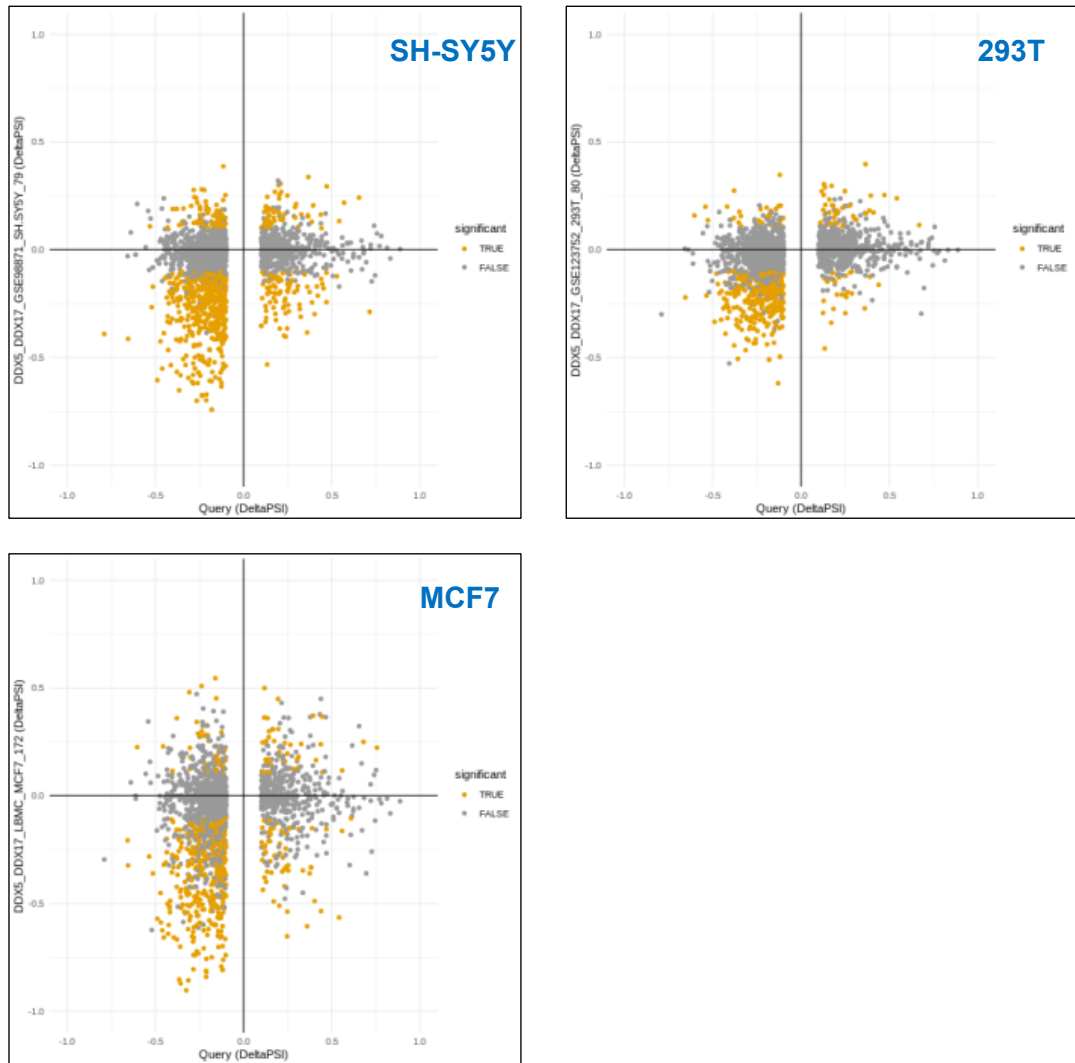


# Supplementary Figure S5



Correlation graphs between the query (DDX5/DDX17\_293T) and the different HNRNPC datasets.

# Supplementary Figure S6



Correlation graphs between the query (HNRNPC\_union) and the different DDX5/DDX17 datasets.

# Bibliographie

Akiva, P., Toporik, A., Edelheit, S., Peretz, Y., Diber, A., Shemesh, R., Novik, A., Sorek, R., 2006. Transcription-mediated gene fusion in the human genome. *Genome Res.* 16, 30–36. <https://doi.org/10.1101/gr.4137606>

Ali, M.A.M., 2021. DEAD-box RNA helicases: The driving forces behind RNA metabolism at the crossroad of viral replication and antiviral innate immunity. *Virus Res.* 296, 198352. <https://doi.org/10.1016/j.virusres.2021.198352>

Alpert, T., Straube, K., Carrillo Oesterreich, F., Herzel, L., Neugebauer, K.M., 2020. Widespread Transcriptional Readthrough Caused by Nab2 Depletion Leads to Chimeric Transcripts with Retained Introns. *Cell Rep.* 33, 108324. <https://doi.org/10.1016/j.celrep.2020.108324>

Alqahtani, H., Gopal, K., Gupta, N., Jung, K., Alshareef, A., Ye, X., Wu, F., Li, L., Lai, R., 2016. DDX17 (P72), a Sox2 binding partner, promotes stem-like features conferred by Sox2 in a small cell population in estrogen receptor-positive breast cancer. *Cell. Signal.* 28, 42–50. <https://doi.org/10.1016/j.cellsig.2015.11.004>

Alves, I., Houle, A.A., Hussin, J.G., Awadalla, P., 2017. The impact of recombination on human mutation load and disease. *Philos. Trans. R. Soc. B Biol. Sci.* 372, 20160465. <https://doi.org/10.1098/rstb.2016.0465>

Ameur, L.B., Marie, P., Thenoz, M., Giraud, G., Combe, E., Claude, J.-B., Lemaire, S., Fontrodona, N., Polveche, H., Bastien, M., Gessain, A., Wattel, E., Bourgeois, C.F., Auboeuf, D., Mortreux, F., 2020. Intragenic recruitment of NF- $\kappa$ B drives splicing modifications upon activation by the oncogene Tax of HTLV-1. *Nat. Commun.* 11, 3045. <https://doi.org/10.1038/s41467-020-16853-x>

An, J.J., Gharami, K., Liao, G.-Y., Woo, N.H., Lau, A.G., Vanevski, F., Torre, E.R., Jones, K.R., Feng, Y., Lu, B., Xu, B., 2008. Distinct role of long 3' UTR BDNF mRNA in spine morphology and synaptic plasticity in hippocampal neurons. *Cell* 134, 175–187. <https://doi.org/10.1016/j.cell.2008.05.045>

Andermatt, I., Wilson, N.H., Bergmann, T., Mauti, O., Gesemann, M., Sockanathan, S., Stoeckli, E.T., 2014. Semaphorin 6B acts as a receptor in post-crossing commissural axon guidance. *Dev. Camb. Engl.* 141, 3709–3720. <https://doi.org/10.1242/dev.112185>

Antonacci, S., Forand, D., Wolf, M., Tyus, C., Barney, J., Kellogg, L., Simon, M.A., Kerr, G., Wells, K.L., Younes, S., Mortimer, N.T., Olesnick, E.C., Killian, D.J., 2015. Conserved RNA-Binding Proteins Required for Dendrite Morphogenesis in *Caenorhabditis elegans* Sensory Neurons. *G3 GenesGenomesGenetics* 5, 639–653. <https://doi.org/10.1534/g3.115.017327>

Attiyeh, E.F., London, W.B., Mossé, Y.P., Wang, Q., Winter, C., Khazi, D., McGrady, P.W., Seeger, R.C., Look, A.T., Shimada, H., Brodeur, G.M., Cohn, S.L., Matthay, K.K., Maris, J.M., 2005. Chromosome 1p and 11q Deletions and Outcome in Neuroblastoma. *N. Engl. J. Med.* 353, 2243–2253. <https://doi.org/10.1056/NEJMoa052399>

- Bader, A.S., Bushell, M., 2020. DNA:RNA hybrids form at DNA double-strand breaks in transcriptionally active loci. *Cell Death Dis.* 11, 1–7. <https://doi.org/10.1038/s41419-020-2464-6>
- Bader, A.S., Luessing, J., Hawley, B.R., Skalka, G.L., Lu, W.-T., Lowndes, N.F., Bushell, M., 2022. DDX17 is required for efficient DSB repair at DNA:RNA hybrid deficient loci. *Nucleic Acids Res.* 50, 10487–10502. <https://doi.org/10.1093/nar/gkac843>
- Balak, C., Benard, M., Schaefer, E., Iqbal, S., Ramsey, K., Ernoult-Lange, M., Mattioli, F., Llaci, L., Geoffroy, V., Courel, M., Naymik, M., Bachman, K.K., Pfundt, R., Rump, P., Beest, J. ter, Wentzensen, I.M., Monaghan, K.G., McWalter, K., Richholt, R., Béché, A.L., Jepsen, W., Both, M.D., Belnap, N., Boland, A., Piras, I.S., Deleuze, J.-F., Szelinger, S., Dollfus, H., Chelly, J., Muller, J., Campbell, A., Lal, D., Rangasamy, S., Mandel, J.-L., Narayanan, V., Huentelman, M., Weil, D., Piton, A., 2019. Rare De Novo Missense Variants in RNA Helicase DDX6 Cause Intellectual Disability and Dysmorphic Features and Lead to P-Body Defects and RNA Dysregulation. *Am. J. Hum. Genet.* 105, 509–525. <https://doi.org/10.1016/j.ajhg.2019.07.010>
- Banigan, E.J., Tang, W., van den Berg, A.A., Stocsits, R.R., Wutz, G., Brandão, H.B., Busslinger, G.A., Peters, J.-M., Mirny, L.A., 2023. Transcription shapes 3D chromatin organization by interacting with loop extrusion. *Proc. Natl. Acad. Sci.* 120, e2210480120. <https://doi.org/10.1073/pnas.2210480120>
- Bartolucci, D., Montemurro, L., Raieli, S., Lampis, S., Pession, A., Hrelia, P., Tonelli, R., 2022. MYCN Impact on High-Risk Neuroblastoma: From Diagnosis and Prognosis to Targeted Treatment. *Cancers* 14, 4421. <https://doi.org/10.3390/cancers14184421>
- Baurén, G., Belikov, S., Wieslander, L., 1998. Transcriptional termination in the Balbiani ring 1 gene is closely coupled to 3'-end formation and excision of the 3'-terminal intron. *Genes Dev.* 12, 2759–2769. <https://doi.org/10.1101/gad.12.17.2759>
- Beaudoin, J.-D., Perreault, J.-P., 2013. Exploring mRNA 3'-UTR G-quadruplexes: evidence of roles in both alternative polyadenylation and mRNA shortening. *Nucleic Acids Res.* 41, 5898–5911. <https://doi.org/10.1093/nar/gkt265>
- Bergant, V., Schnepf, D., de Andrade Krätzig, N., Hubel, P., Urban, C., Engleitner, T., Dijkman, R., Ryffel, B., Steiger, K., Knolle, P.A., Kochs, G., Rad, R., Staeheli, P., Pichlmair, A., 2023. mRNA 3'UTR lengthening by alternative polyadenylation attenuates inflammatory responses and correlates with virulence of Influenza A virus. *Nat. Commun.* 14, 4906. <https://doi.org/10.1038/s41467-023-40469-6>
- Berkovits, B.D., Mayr, C., 2015. Alternative 3' UTRs act as scaffolds to regulate membrane protein localization. *Nature* 522, 363–367. <https://doi.org/10.1038/nature14321>
- Bohnsack, K.E., Yi, S., Venus, S., Jankowsky, E., Bohnsack, M.T., 2023. Cellular functions of eukaryotic RNA helicases and their links to human diseases. *Nat. Rev. Mol. Cell Biol.* <https://doi.org/10.1038/s41580-023-00628-5>
- Bortz, E., Westera, L., Maamary, J., Steel, J., Albrecht, R.A., Manicassamy, B., Chase, G., Martínez-Sobrido, L., Schwemmle, M., García-Sastre, A., 2011. Host- and Strain-Specific

Regulation of Influenza Virus Polymerase Activity by Interacting Cellular Proteins. *mBio* 2, 10.1128/mbio.00151-11. <https://doi.org/10.1128/mbio.00151-11>

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

Bown, N., 2001. Neuroblastoma tumour genetics: clinical and biological aspects. *J. Clin. Pathol.* 54, 897–910.

Bown, N., Łastowska, M., Cotterill, S., O'Neill, S., Ellershaw, C., Roberts, P., Lewis, I., 2001. 17q gain in neuroblastoma predicts adverse clinical outcome. *Med. Pediatr. Oncol.* 36, 14–19. [https://doi.org/10.1002/1096-911X\(20010101\)36:1<14::AID-MPO1005>3.0.CO;2-G](https://doi.org/10.1002/1096-911X(20010101)36:1<14::AID-MPO1005>3.0.CO;2-G)

Boyault, C., Zhang, Y., Fritah, S., Caron, C., Gilquin, B., Kwon, S.H., Garrido, C., Yao, T.-P., Vourc'h, C., Matthias, P., Khochbin, S., 2007. HDAC6 controls major cell response pathways to cytotoxic accumulation of protein aggregates. *Genes Dev.* 21, 2172–2181. <https://doi.org/10.1101/gad.436407>

Brodeur, G.M., 2018. Spontaneous regression of neuroblastoma. *Cell Tissue Res.* 372, 277–286. <https://doi.org/10.1007/s00441-017-2761-2>

Brodeur, G.M., Seeger, R.C., Schwab, M., Varmus, H.E., Bishop, J.M., 1984. Amplification of N-myc in untreated human neuroblastomas correlates with advanced disease stage. *Science* 224, 1121–1124. <https://doi.org/10.1126/science.6719137>

Büchel, G., Carstensen, A., Mak, K.-Y., Roeschert, I., Leen, E., Sumara, O., Hofstetter, J., Herold, S., Kalb, J., Baluapuri, A., Poon, E., Kwok, C., Chesler, L., Maric, H.M., Rickman, D.S., Wolf, E., Bayliss, R., Walz, S., Eilers, M., 2017. Association with Aurora-A Controls N-MYC-Dependent Promoter Escape and Pause Release of RNA Polymerase II during the Cell Cycle. *Cell Rep.* 21, 3483–3497. <https://doi.org/10.1016/j.celrep.2017.11.090>

Buratti, E., Baralle, F.E., 2004. Influence of RNA Secondary Structure on the Pre-mRNA Splicing Process. *Mol. Cell. Biol.* 24, 10505–10514. <https://doi.org/10.1128/MCB.24.24.10505-10514.2004>

Burns, W., Bird, L.M., Heron, D., Keren, B., Ramachandra, D., Thiffault, I., Del Viso, F., Amudhavalli, S., Engleman, K., Parenti, I., Kaiser, F.J., Wierzba, J., Riedhammer, K.M., Liptay, S., Zadeh, N., Pormann, J., Fischer, A., Gößwein, S., McLaughlin, H.M., Telegrafi, A., Langley, K.G., Steet, R., Louie, R.J., Lyons, M.J., 2021. Syndromic neurodevelopmental disorder associated with de novo variants in DDX23. *Am. J. Med. Genet. A.* 185, 2863–2872. <https://doi.org/10.1002/ajmg.a.62359>

Cardiello, J.F., Goodrich, J.A., Kugel, J.F., 2018. Heat Shock Causes a Reversible Increase in RNA Polymerase II Occupancy Downstream of mRNA Genes, Consistent with a Global Loss in Transcriptional Termination. *Mol. Cell. Biol.* 38, e00181-18. <https://doi.org/10.1128/MCB.00181-18>

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., Sartorelli, V., 2006. The RNA Helicases p68/p72 and

the Noncoding RNA SRA Are Coregulators of MyoD and Skeletal Muscle Differentiation. *Dev. Cell* 11, 547–560. <https://doi.org/10.1016/j.devcel.2006.08.003>

Carrasco, J., Rauer, M., Hummel, B., Grzejda, D., Alfonso-Gonzalez, C., Lee, Y., Wang, Q., Puchalska, M., Mittler, G., Hilgers, V., 2020. ELAV and FNE Determine Neuronal Transcript Signatures through EXon-Activated Rescue. *Mol. Cell* 80, 156-163.e6. <https://doi.org/10.1016/j.molcel.2020.09.011>

Castillo-Guzman, D., Chédin, F., 2021. Defining R-loop classes and their contributions to genome instability. *DNA Repair* 106, 103182. <https://doi.org/10.1016/j.dnarep.2021.103182>

Castillo-Guzman, D., Hartono, S.R., Sanz, L.A., Chédin, F., 2020. SF3B1-targeted Splicing Inhibition Triggers Global Alterations in Transcriptional Dynamics and R-Loop Metabolism. *bioRxiv* 2020.06.08.130583. <https://doi.org/10.1101/2020.06.08.130583>

Chakraborty, A., Tapryal, N., Venkova, T., Horikoshi, N., Pandita, R.K., Sarker, A.H., Sarker, P.S., Pandita, T.K., Hazra, T.K., 2016. Classical non-homologous end-joining pathway utilizes nascent RNA for error-free double-strand break repair of transcribed genes. *Nat. Commun.* 7, 13049. <https://doi.org/10.1038/ncomms13049>

Chan, S.L., Huppertz, I., Yao, C., Weng, L., Moresco, J.J., Yates, J.R., Ule, J., Manley, J.L., Shi, Y., 2014. CPSF30 and Wdr33 directly bind to AAUAAA in mammalian mRNA 3' processing. *Genes Dev.* 28, 2370–2380. <https://doi.org/10.1101/gad.250993.114>

Chao, S.-H., Price, D.H., 2001. Flavopiridol Inactivates P-TEFb and Blocks Most RNA Polymerase II Transcription in Vivo \*. *J. Biol. Chem.* 276, 31793–31799. <https://doi.org/10.1074/jbc.M102306200>

Chen, J.-Y., Chen, W.-N., Poon, K.-M.V., Zheng, B.-J., Lin, X., Wang, Y.-X., Wen, Y.-M., 2009. Interaction between SARS-CoV helicase and a multifunctional cellular protein (Ddx5) revealed by yeast and mammalian cell two-hybrid systems. *Arch. Virol.* 154, 507–512. <https://doi.org/10.1007/s00705-009-0323-y>

Chen, Y., Takita, J., Choi, Y.L., Kato, M., Ohira, M., Sanada, M., Wang, L., Soda, M., Kikuchi, A., Igarashi, T., Nakagawara, A., Hayashi, Y., Mano, H., Ogawa, S., 2008. Oncogenic mutations of ALK kinase in neuroblastoma. *Nature* 455, 971–974. <https://doi.org/10.1038/nature07399>

Cheng, C., He, T., Chen, K., Cai, Y., Gu, Y., Pan, L., Duan, P., Wu, Y., Wu, Z., 2023. P300 Interacted With N-Myc and Regulated Its Protein Stability via Altering Its Post-Translational Modifications in Neuroblastoma. *Mol. Cell. Proteomics* 22, 100504. <https://doi.org/10.1016/j.mcpro.2023.100504>

Chernukhin, I., Shamsuddin, S., Kang, S.Y., Bergström, R., Kwon, Y.-W., Yu, W., Whitehead, J., Mukhopadhyay, R., Docquier, F., Farrar, D., Morrison, I., Vigneron, M., Wu, S.-Y., Chiang, C.-M., Loukinov, D., Lobanenkova, V., Ohlsson, R., Klenova, E., 2007. CTCF Interacts with and Recruits the Largest Subunit of RNA Polymerase II to CTCF Target Sites Genome-Wide. *Mol. Cell. Biol.* 27, 1631–1648. <https://doi.org/10.1128/MCB.01993-06>



Clark, E.L., Coulson, A., Dalglish, C., Rajan, P., Nicol, S.M., Fleming, S., Heer, R., Gaughan, L., Leung, H.Y., Elliott, D.J., Fuller-Pace, F.V., Robson, C.N., 2008. The RNA helicase p68 is a novel Androgen Receptor Co-Activator involved in splicing and is over-expressed in Prostate Cancer. *Cancer Res.* 68, 7938–7946.

<https://doi.org/10.1158/0008-5472.CAN-08-0932>

Clark, E.L., Hadjimichael, C., Temperley, R., Barnard, A., Fuller-Pace, F.V., Robson, C.N., 2013. p68/Ddx5 Supports  $\beta$ -Catenin & RNAP II during Androgen Receptor Mediated Transcription in Prostate Cancer. *PLoS ONE* 8, e54150.

<https://doi.org/10.1371/journal.pone.0054150>

Close, P., East, P., Dirac-Svejstrup, A.B., Hartmann, H., Heron, M., Maslen, S., Chariot, A., Söding, J., Skehel, M., Svejstrup, J.Q., 2012. DBIRD complex integrates alternative mRNA splicing with RNA polymerase II transcript elongation. *Nature* 484, 386–389.

<https://doi.org/10.1038/nature10925>

Cohn, S.L., Pearson, A.D.J., London, W.B., Monclair, T., Ambros, P.F., Brodeur, G.M., Faldut, A., Hero, B., Iehara, T., Machin, D., Mosseri, V., Simon, T., Garaventa, A., Castel, V., Matthay, K.K., 2009. The International Neuroblastoma Risk Group (INRG) Classification System: An INRG Task Force Report. *J. Clin. Oncol.* 27, 289–297.

<https://doi.org/10.1200/JCO.2008.16.6785>

Connelly, S., Manley, J.L., 1988. A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. *Genes Dev.* 2, 440–452.

<https://doi.org/10.1101/gad.2.4.440>

Cortazar, M.A., Sheridan, R.M., Erickson, B., Fong, N., Glover-Cutter, K., Brannan, K., Bentley, D.L., 2019. Control of RNA Pol II Speed by PNUITS-PP1 and Spt5 Dephosphorylation Facilitates Termination by a “Sitting Duck Torpedo” Mechanism. *Mol. Cell* 76, 896–908.e4.

<https://doi.org/10.1016/j.molcel.2019.09.031>

Corvetta, D., Chayka, O., Gherardi, S., D’Acunzio, C.W., Cantilena, S., Valli, E., Piotrowska, I., Perini, G., Sala, A., 2013. Physical interaction between MYCN oncogene and polycomb repressive complex 2 (PRC2) in neuroblastoma: functional and therapeutic implications. *J. Biol. Chem.* 288, 8332–8341.

<https://doi.org/10.1074/jbc.M113.454280>  
Cotterman, R., Jin, V.X., Krig, S.R., Lemen, J.M., Wey, A., Farnham, P.J., Knoepfler, P.S., 2008. N-Myc regulates a widespread euchromatic program in the human genome partially independent of its role as a classical transcription factor. *Cancer Res.* 68, 9654–9662.

<https://doi.org/10.1158/0008-5472.CAN-08-1961>

Crossley, M.P., Bocek, M., Cimprich, K.A., 2019. R-Loops as Cellular Regulators and Genomic Threats. *Mol. Cell* 73, 398–411.

<https://doi.org/10.1016/j.molcel.2019.01.024>

Cugusi, S., Mitter, R., Kelly, G.P., Walker, J., Han, Z., Pisano, P., Wierer, M., Stewart, A., Svejstrup, J.Q., 2022. Heat shock induces premature transcript termination and reconfigures the human transcriptome. *Mol. Cell* 82, 1573–1588.e10.

<https://doi.org/10.1016/j.molcel.2022.01.007>

Danelon, V., Goldner, R., Martinez, E., Gokhman, I., Wang, K., Yaron, A., Tran, T.S., 2020. Modular and Distinct Plexin-A4/FARP2/Rac1 Signaling Controls Dendrite

Morphogenesis. *J. Neurosci.* 40, 5413–5430. <https://doi.org/10.1523/JNEUROSCI.2730-19.2020>

Dardenne, E., Polay Espinoza, M., Fattet, L., Germann, S., Lambert, M.-P., Neil, H., Zonta, E., Mortada, H., Gratadou, L., Deygas, M., Chakrama, F.Z., Samaan, S., Desmet, F.-O., Tranchevent, L.-C., Dutertre, M., Rimokh, R., Bourgeois, C.F., Auboeuf, D., 2014. RNA Helicases DDX5 and DDX17 Dynamically Orchestrate Transcription, miRNA, and Splicing Programs in Cell Differentiation. *Cell Rep.* 7, 1900–1913. <https://doi.org/10.1016/j.celrep.2014.05.010>

Das, R., Yu, J., Zhang, Z., Gygi, M.P., Krainer, A.R., Gygi, S.P., Reed, R., 2007. SR Proteins Function in Coupling RNAP II Transcription to Pre-mRNA Splicing. *Mol. Cell* 26, 867–881. <https://doi.org/10.1016/j.molcel.2007.05.036>

Das, S.K., Kuzin, V., Cameron, D.P., Sanford, S., Jha, R.K., Nie, Z., Rosello, M.T., Holewinski, R., Andresson, T., Wisniewski, J., Natsume, T., Price, D.H., Lewis, B.A., Kouzine, F., Levens, D., Baranello, L., 2022. MYC assembles and stimulates topoisomerases 1 and 2 in a “topoisome.” *Mol. Cell* 82, 140-158.e12. <https://doi.org/10.1016/j.molcel.2021.11.016>

Dasilva, L.F., Blumenthal, E., Beckedorff, F., Cingaram, P.R., Gomes Dos Santos, H., Edupuganti, R.R., Zhang, A., Dokaneheifard, S., Aoi, Y., Yue, J., Kirstein, N., Tayari, M.M., Shilatifard, A., Shiekhatter, R., 2021. Integrator enforces the fidelity of transcriptional termination at protein-coding genes. *Sci. Adv.* 7, eabe3393. <https://doi.org/10.1126/sciadv.abe3393>

David, C.J., Boyne, A.R., Millhouse, S.R., Manley, J.L., 2011. The RNA polymerase II C-terminal domain promotes splicing activation through recruitment of a U2AF65–Prp19 complex. *Genes Dev.* 25, 972–983. <https://doi.org/10.1101/gad.2038011>

Deák, F., Shin, O.-H., Tang, J., Hanson, P., Ubach, J., Jahn, R., Rizo, J., Kavalali, E.T., Südhof, T.C., 2006. Rabphilin regulates SNARE-dependent re-priming of synaptic vesicles for fusion. *EMBO J.* 25, 2856–2866. <https://doi.org/10.1038/sj.emboj.7601165>

Decorsière, A., Cayrel, A., Vagner, S., 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. <https://doi.org/10.1101/gad.607011>

Deidda, G., Rossi, N., Tocchini-Valentini, G.P., 2003. An archaeal endoribonuclease catalyzes cis- and trans- nonspliceosomal splicing in mouse cells. *Nat. Biotechnol.* 21, 1499–1504. <https://doi.org/10.1038/nbt908>

Deng, L., Zeng, Q., Wang, M., Cheng, A., Jia, R., Chen, S., Zhu, D., Liu, M., Yang, Q., Wu, Y., Zhao, X., Zhang, S., Liu, Y., Yu, Y., Zhang, L., Chen, X., 2018. Suppression of NF-κB Activity: A Viral Immune Evasion Mechanism. *Viruses* 10, 409. <https://doi.org/10.3390/v10080409>

Dikow, N., Granzow, M., Graul-Neumann, L.M., Karch, S., Hinderhofer, K., Paramasivam, N., Behl, L.-J., Kaufmann, L., Fischer, C., Evers, C., Schlesner, M., Eils, R., Borck, G., Zweier, C., Bartram, C.R., Carey, J.C., Moog, U., 2017. DDX3X mutations in two girls with

a phenotype overlapping Toriello–Carey syndrome. *Am. J. Med. Genet. A.* 173, 1369–1373. <https://doi.org/10.1002/ajmg.a.38164>

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., Ren, B., 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485, 376–380. <https://doi.org/10.1038/nature11082>

Dominski, Z., Marzluff, W.F., 2007. Formation of the 3' end of histone mRNA: Getting closer to the end. *Gene* 396, 373–390. <https://doi.org/10.1016/j.gene.2007.04.021>

Dutertre, M., Gratadou, L., Dardenne, E., Germann, S., Samaan, S., Lidereau, R., Driouch, K., de la Grange, P., Auboeuf, D., 2010. Estrogen regulation and physiopathologic significance of alternative promoters in breast cancer. *Cancer Res.* 70, 3760–3770. <https://doi.org/10.1158/0008-5472.CAN-09-3988>

Dutertre, M., Sfaxi, R., Vagner, S., 2021. Reciprocal Links between Pre-messenger RNA 3'-End Processing and Genome Stability. *Trends Biochem. Sci.* 46, 579–594. <https://doi.org/10.1016/j.tibs.2021.01.009>

Dutta, S., Gupta, G., Choi, Y.-W., Kotaka, M., Fielding, B.C., Song, J., Tan, Y.-J., 2012. The variable N-terminal region of DDX5 contains structural elements and auto-inhibits its interaction with NS5B of hepatitis C virus. *Biochem. J.* 446, 37–46. <https://doi.org/10.1042/BJ20120001>

Dvinge, H., 2018. Regulation of alternative mRNA splicing: old players and new perspectives. *FEBS Lett.* 592, 2987–3006. <https://doi.org/10.1002/1873-3468.13119>

Eaton, J.D., Davidson, L., Bauer, D.L.V., Natsume, T., Kanemaki, M.T., West, S., 2018. Xrn2 accelerates termination by RNA polymerase II, which is underpinned by CPSF73 activity. *Genes Dev.* 32, 127–139. <https://doi.org/10.1101/gad.308528.117>

Endoh, H., Maruyama, K., Masuhiro, Y., Kobayashi, Y., Goto, M., Tai, H., Yanagisawa, J., Metzger, D., Hashimoto, S., Kato, S., 2014. Retraction for Endoh 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.* 34, 915. <https://doi.org/10.1128/MCB.01458-13>

Eul, J., Graessmann, M., Graessmann, A., 1995. Experimental evidence for RNA trans-splicing in mammalian cells. *EMBO J.* 14, 3226–3235.

Fahey, C.C., Davis, I.J., 2017. SETting the Stage for Cancer Development: SETD2 and the Consequences of Lost Methylation. *Cold Spring Harb. Perspect. Med.* 7, a026468. <https://doi.org/10.1101/cshperspect.a026468>

Fairman, M.E., Maroney, P.A., Wang, W., Bowers, H.A., Gollnick, P., Nilsen, T.W., Jankowsky, E., 2004. Protein Displacement by DEXH/D “RNA Helicases” Without Duplex Unwinding. *Science* 304, 730–734. <https://doi.org/10.1126/science.1095596>

Favaro, F.P., Alvizi, L., Zechi-Ceide, R.M., Bertola, D., Felix, T.M., de Souza, J., Raskin, S., Twigg, S.R.F., Weiner, A.M.J., Armas, P., Margarit, E., Calcaterra, N.B., Andersen, G.R., McGowan, S.J., Wilkie, A.O.M., Richieri-Costa, A., de Almeida, M.L.G., Passos-Bueno, M.R., 2014. A Noncoding Expansion in EIF4A3 Causes Richieri-Costa-Pereira Syndrome,

a Craniofacial Disorder Associated with Limb Defects. *Am. J. Hum. Genet.* 94, 120–128. <https://doi.org/10.1016/j.ajhg.2013.11.020>

Fernandez-Escamilla, A.-M., Rousseau, F., Schymkowitz, J., Serrano, L., 2004. Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nat. Biotechnol.* 22, 1302–1306. <https://doi.org/10.1038/nbt1012>

Fischl, H., Neve, J., Wang, Z., Patel, R., Louey, A., Tian, B., Furger, A., 2019. hnRNP C regulates cancer-specific alternative cleavage and polyadenylation profiles. *Nucleic Acids Res.* 47, 7580–7591. <https://doi.org/10.1093/nar/gkz461>

Fletcher, J.I., Ziegler, D.S., Trahair, T.N., Marshall, G.M., Haber, M., Norris, M.D., 2018. Too many targets, not enough patients: rethinking neuroblastoma clinical trials. *Nat. Rev. Cancer* 18, 389–400. <https://doi.org/10.1038/s41568-018-0003-x>

Fortuna, T.R., Kour, S., Anderson, E.N., Ward, C., Rajasundaram, D., Donnelly, C.J., Hermann, A., Wyne, H., Shewmaker, F., Pandey, U.B., 2021. DDX17 is involved in DNA damage repair and modifies FUS toxicity in an RGG-domain dependent manner. *Acta Neuropathol. (Berl.)* 142, 515–536. <https://doi.org/10.1007/s00401-021-02333-z>

Franchini, L., Stanic, J., Barzasi, M., Zianni, E., Mauceri, D., Diluca, M., Gardoni, F., 2022. Rabphilin-3A Drives Structural Modifications of Dendritic Spines Induced by Long-Term Potentiation. *Cells* 11, 1616. <https://doi.org/10.3390/cells11101616>

Fujinaga, K., Huang, F., Peterlin, B.M., 2023. P-TEFb: The master regulator of transcription elongation. *Mol. Cell* 83, 393–403. <https://doi.org/10.1016/j.molcel.2022.12.006>

Fuller-Pace, F., Ahamed, E., Wortham, N., Ali, S., 2008. Coactivation of estrogen receptor alpha by the DEAD-box RNA helicases p68 and p72 and its role in breast cancer. *Breast Cancer Res. BCR* 10, P16. <https://doi.org/10.1186/bcr1900>

Gao, J., Byrd, A.K., Zybailov, B.L., Marecki, J.C., Guderyon, M.J., Edwards, A.D., Chib, S., West, K.L., Waldrip, Z.J., Mackintosh, S.G., Gao, Z., Putnam, A.A., Jankowsky, E., Raney, K.D., 2019. DEAD-box RNA helicases Dbp2, Ded1 and Mss116 bind to G-quadruplex nucleic acids and destabilize G-quadruplex RNA. *Chem. Commun.* 55, 4467–4470. <https://doi.org/10.1039/C8CC10091H>

Garcia-Mata, R., Gao, Y.-S., Sztul, E., 2002. Hassles with Taking Out the Garbage: Aggravating Aggresomes. *Traffic* 3, 388–396. <https://doi.org/10.1034/j.1600-0854.2002.30602.x>

Garneau, N.L., Wilusz, J., Wilusz, C.J., 2007. The highways and byways of mRNA decay. *Nat. Rev. Mol. Cell Biol.* 8, 113–126. <https://doi.org/10.1038/nrm2104>

Gemmill, D., Nelson, C., Badmalia, M., Pereira, H.S., Kerr, L., Wolfinger, M.T., Patel, T.R., 2023. The 3' terminal region of Zika virus RNA contains a conserved G-quadruplex and is unfolded by human DDX17. *Biochem. Cell Biol.* <https://doi.org/10.1139/bcb-2023-0036>

- Gherardi, S., Valli, E., Erriquez, D., Perini, G., 2013. MYCN-mediated transcriptional repression in neuroblastoma: the other side of the coin. *Front. Oncol.* 3. <https://doi.org/10.3389/fonc.2013.00042>
- Goh, P.-Y., Tan, Y.-J., Lim, S.P., Tan, Y.H., Lim, S.G., Fuller-Pace, F., Hong, W., 2004. Cellular RNA Helicase p68 Relocalization and Interaction with the Hepatitis C Virus (HCV) NS5B Protein and the Potential Role of p68 in HCV RNA Replication. *J. Virol.* 78, 5288–5298. <https://doi.org/10.1128/JVI.78.10.5288-5298.2004>
- Göpferich, M., George, N.O., Muelas, A.D., Bizyn, A., Pascual, R., Fijalkowska, D., Kalamakis, G., Müller, U., Krijgsveld, J., Mendez, R., Fariñas, I., Huber, W., Anders, S., Martin-Villalba, A., 2020. Single cell 3'UTR analysis identifies changes in alternative polyadenylation throughout neuronal differentiation and in autism. <https://doi.org/10.1101/2020.08.12.247627>
- Gowher, H., Brick, K., Camerini-Otero, R.D., Felsenfeld, G., 2012. Vezf1 protein binding sites genome-wide are associated with pausing of elongating RNA polymerase II. *Proc. Natl. Acad. Sci. U. S. A.* 109, 2370–2375. <https://doi.org/10.1073/pnas.1121538109>
- Grady, E.F., Schwab, M., Rosenau, W., 1987. Expression of N-myc and c-src during the development of fetal human brain. *Cancer Res.* 47, 2931–2936.
- Greger, L., Su, J., Rung, J., Ferreira, P.G., Consortium, G., Lappalainen, T., Dermitzakis, E.T., Brazma, A., 2014. Tandem RNA Chimeras Contribute to Transcriptome Diversity in Human Population and Are Associated with Intronic Genetic Variants. *PLOS ONE* 9, e104567. <https://doi.org/10.1371/journal.pone.0104567>
- Gromak, N., West, S., Proudfoot, N.J., 2006. Pause Sites Promote Transcriptional Termination of Mammalian RNA Polymerase II. *Mol. Cell. Biol.* 26, 3986–3996. <https://doi.org/10.1128/MCB.26.10.3986-3996.2006>
- Grosso, A.R., Leite, A.P., Carvalho, S., Matos, M.R., Martins, F.B., Vítor, A.C., Desterro, J.M., Carmo-Fonseca, M., de Almeida, S.F., 2015. Pervasive transcription read-through promotes aberrant expression of oncogenes and RNA chimeras in renal carcinoma. *eLife* 4, e09214. <https://doi.org/10.7554/eLife.09214>
- Gruber, A.J., Schmidt, R., Ghosh, S., Martin, G., Gruber, A.R., van Nimwegen, E., Zavolan, M., 2018. Discovery of physiological and cancer-related regulators of 3' UTR processing with KAPAC. *Genome Biol.* 19, 44. <https://doi.org/10.1186/s13059-018-1415-3>
- Gruber, A.J., Zavolan, M., 2019. Alternative cleavage and polyadenylation in health and disease. *Nat. Rev. Genet.* 20, 599–614. <https://doi.org/10.1038/s41576-019-0145-z>
- Gruber, A.R., Martin, G., Müller, P., Schmidt, A., Gruber, A.J., Gumienny, R., Mittal, N., Jayachandran, R., Pieters, J., Keller, W., van Nimwegen, E., Zavolan, M., 2014. Global 3' UTR shortening has a limited effect on protein abundance in proliferating T cells. *Nat. Commun.* 5, 5465. <https://doi.org/10.1038/ncomms6465>
- Grummt, I., 1998. Regulation of Mammalian Ribosomal Gene Transcription by RNA Polymerase I, in: Moldave, K. (Ed.), *Progress in Nucleic Acid Research and Molecular*

Biology. Academic Press, pp. 109–154. [https://doi.org/10.1016/S0079-6603\(08\)60506-1](https://doi.org/10.1016/S0079-6603(08)60506-1)

Gualtieri, A., Bianconi, V., Renzini, A., Pieroni, L., Licursi, V., Mozzetta, C., 2022. The RNA helicase DDX5 cooperates with EHMT2 to sustain alveolar rhabdomyosarcoma growth. *Cell Rep.* 40, 111267. <https://doi.org/10.1016/j.celrep.2022.111267>

Guturi, K.K.N., Sarkar, M., Bhowmik, A., Das, N., Ghosh, M.K., 2014. DEAD-box protein p68 is regulated by  $\beta$ -catenin/transcription factor 4 to maintain a positive feedback loop in control of breast cancer progression. *Breast Cancer Res.* 16, 1–15. <https://doi.org/10.1186/s13058-014-0496-5>

Guvenek, A., Tian, B., 2018. Analysis of alternative cleavage and polyadenylation in mature and differentiating neurons using RNA-seq data. *Quant. Biol.* 6, 253–266. <https://doi.org/10.1007/s40484-018-0148-3>

Haas, B.J., Dobin, A., Li, B., Stransky, N., Pochet, N., Regev, A., 2019. Accuracy assessment of fusion transcript detection via read-mapping and de novo fusion transcript assembly-based methods. *Genome Biol.* 20, 213. <https://doi.org/10.1186/s13059-019-1842-9>

Hamperl, S., Bocek, M.J., Saldivar, J.C., Swigut, T., Cimprich, K.A., 2017. Transcription-Replication Conflict Orientation Modulates R-Loop Levels and Activates Distinct DNA Damage Responses. *Cell* 170, 774-786.e19. <https://doi.org/10.1016/j.cell.2017.07.043>

Hao, J., Li, P., Tian, Y., Wang, Y., Li, Suqing, Wang, L., Li, Shujun, 2018. Crosstalk between Toll-like receptor 3 and Notch signaling contributes to CD14+ monocytes activity in enterovirus 71 infected hand, foot, and mouth disease. *Int. Immunopharmacol.* 60, 26–33. <https://doi.org/10.1016/j.intimp.2018.04.031>

Heinz, S., Texari, L., Hayes, M.G.B., Urbanowski, M., Chang, M.W., Givarkes, N., Rialdi, A., White, K.M., Albrecht, R.A., Pache, L., Marazzi, I., García-Sastre, A., Shaw, M.L., Benner, C., 2018. Transcription Elongation Can Affect Genome 3D Structure. *Cell* 174, 1522-1536.e22. <https://doi.org/10.1016/j.cell.2018.07.047>

Hennig, T., Michalski, M., Rutkowski, A.J., Djakovic, L., Whisnant, A.W., Friedl, M.-S., Jha, B.A., Baptista, M.A.P., L'Hernault, A., Erhard, F., Dölken, L., Friedel, C.C., 2018. HSV-1-induced disruption of transcription termination resembles a cellular stress response but selectively increases chromatin accessibility downstream of genes. *PLOS Pathog.* 14, e1006954. <https://doi.org/10.1371/journal.ppat.1006954>

Herold, S., Kalb, J., Büchel, G., Ade, C.P., Baluapuri, A., Xu, J., Koster, J., Solvie, D., Carstensen, A., Klotz, C., Rodewald, S., Schüle-Völk, C., Döbelstein, M., Wolf, E., Molenaar, J., Versteeg, R., Walz, S., Eilers, M., 2019. Recruitment of BRCA1 limits MYCN-driven accumulation of stalled RNA polymerase. *Nature* 567, 545–549. <https://doi.org/10.1038/s41586-019-1030-9>

Hönig, A., Auboeuf, D., Parker, M.M., O'Malley, B.W., Berget, S.M., 2002. Regulation of Alternative Splicing by the ATP-Dependent DEAD-Box RNA Helicase p72. *Mol. Cell. Biol.* 22, 5698–5707. <https://doi.org/10.1128/MCB.22.16.5698-5707.2002>

- Hoque, M., Ji, Z., Zheng, D., Luo, W., Li, W., You, B., Park, J.Y., Yehia, G., Tian, B., 2013. Analysis of alternative cleavage and polyadenylation by 3' region extraction and deep sequencing. *Nat. Methods* 10, 133–139. <https://doi.org/10.1038/nmeth.2288>
- Houseley, J., Tollervey, D., 2010. Apparent Non-Canonical Trans-Splicing Is Generated by Reverse Transcriptase In Vitro. *PLOS ONE* 5, e12271. <https://doi.org/10.1371/journal.pone.0012271>
- Hsin, J.-P., Manley, J.L., 2012. The RNA polymerase II CTD coordinates transcription and RNA processing. *Genes Dev.* 26, 2119–2137. <https://doi.org/10.1101/gad.200303.112>
- Huang, K.-L., Jee, D., Stein, C.B., Elrod, N.D., Henriques, T., Mascibroda, L.G., Baillat, D., Russell, W.K., Adelman, K., Wagner, E.J., 2020. Integrator Recruits Protein Phosphatase 2A to Prevent Pause Release and Facilitate Transcription Termination. *Mol. Cell* 80, 345–358.e9. <https://doi.org/10.1016/j.molcel.2020.08.016>
- Huang, M., Weiss, W.A., 2013. Neuroblastoma and MYCN. *Cold Spring Harb. Perspect. Med.* 3. <https://doi.org/10.1101/cshperspect.a014415>
- Hwang, H.-W., Park, C.Y., Goodarzi, H., Fak, J.J., Mele, A., Moore, M.J., Saito, Y., Darnell, R.B., 2016. PAPERCLIP Identifies MicroRNA Targets and a Role of CstF64/64tau in Promoting Non-canonical poly(A) Site Usage. *Cell Rep.* 15, 423–435. <https://doi.org/10.1016/j.celrep.2016.03.023>
- Jalihal, A.P., Pitchiaya, S., Xiao, L., Bawa, P., Jiang, X., Bedi, K., Parolia, A., Cieslik, M., Ljungman, M., Chinnaiyan, A.M., Walter, N.G., 2020. Multivalent Proteins Rapidly and Reversibly Phase-Separate upon Osmotic Cell Volume Change. *Mol. Cell* 79, 978–990.e5. <https://doi.org/10.1016/j.molcel.2020.08.004>
- Janoueix-Lerosey, I., Lequin, D., Brugières, L., Ribeiro, A., de Pontual, L., Combaret, V., Raynal, V., Puisieux, A., Schleiermacher, G., Pierron, G., Valteau-Couanet, D., Frebourg, T., Michon, J., Lyonnet, S., Amiel, J., Delattre, O., 2008. Somatic and germline activating mutations of the ALK kinase receptor in neuroblastoma. *Nature* 455, 967–970. <https://doi.org/10.1038/nature07398>
- 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., Westendorf, J.J., 2008. p68 (Ddx5) interacts with Runx2 and regulates osteoblast differentiation. *J. Cell. Biochem.* 103, 1438–1451. <https://doi.org/10.1002/jcb.21526>
- Ji, Xinjun, Wan, J., Vishnu, M., Xing, Y., Liebhaber, S.A., 2013.  $\alpha$ CP Poly(C) Binding Proteins Act as Global Regulators of Alternative Polyadenylation. *Mol. Cell. Biol.* 33, 2560–2573. <https://doi.org/10.1128/MCB.01380-12>
- Ji, Xiong, Zhou, Y., Pandit, S., Huang, J., Li, H., Lin, C.Y., Xiao, R., Burge, C.B., Fu, X.-D., 2013. SR Proteins Collaborate with 7SK and Promoter-Associated Nascent RNA to Release Paused Polymerase. *Cell* 153, 855–868. <https://doi.org/10.1016/j.cell.2013.04.028>
- Johnsen, J.I., Dyberg, C., Wickström, M., 2019. Neuroblastoma—A Neural Crest Derived Embryonal Malignancy. *Front. Mol. Neurosci.* 12. <https://doi.org/10.3389/fnmol.2019.00009>

- Jones, K., Wei, C., Schoser, B., Meola, G., Timchenko, N., Timchenko, L., 2015. Reduction of toxic RNAs in myotonic dystrophies type 1 and type 2 by the RNA helicase p68/DDX5. *Proc. Natl. Acad. Sci.* 112, 8041–8045. <https://doi.org/10.1073/pnas.1422273112>
- Jung, C., Mittler, G., Oswald, F., Borggrefe, T., 2013. RNA helicase Ddx5 and the noncoding RNA SRA act as coactivators in the Notch signaling pathway. *Biochim. Biophys. Acta BBA - Mol. Cell Res.* 1833, 1180–1189. <https://doi.org/10.1016/j.bbamcr.2013.01.032>
- Kaida, D., Berg, M.G., Younis, I., Kasim, M., Singh, L.N., Wan, L., Dreyfuss, G., 2010. U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature* 468, 664–668. <https://doi.org/10.1038/nature09479>
- Kang, H., Lieberman, P.M., 2011. Mechanism of Glycyrrhizic Acid Inhibition of Kaposi's Sarcoma-Associated Herpesvirus: Disruption of CTCF-Cohesin-Mediated RNA Polymerase II Pausing and Sister Chromatid Cohesion. *J. Virol.* 85, 11159–11169. <https://doi.org/10.1128/JVI.00720-11>
- Kar, A., Fushimi, K., Zhou, X., Ray, P., Shi, C., Chen, X., Liu, Z., Chen, S., 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. <https://doi.org/10.1128/MCB.01149-10>
- Katahira, J., 2012. mRNA export and the TREX complex. *Biochim. Biophys. Acta BBA - Gene Regul. Mech., Nuclear Transport and RNA Processing* 1819, 507–513. <https://doi.org/10.1016/j.bbagr.2011.12.001>
- Katahira, J., Ishikawa, H., Tsujimura, K., Kurono, S., Hieda, M., 2019. Human THO coordinates transcription termination and subsequent transcript release from the HSP70 locus. *Genes Cells* 24, 272–283. <https://doi.org/10.1111/gtc.12672>
- Katahira, J., Okuzaki, D., Inoue, H., Yoneda, Y., Maehara, K., Ohkawa, Y., 2013. Human TREX component Thoc5 affects alternative polyadenylation site choice by recruiting mammalian cleavage factor I. *Nucleic Acids Res.* 41, 7060–7072. <https://doi.org/10.1093/nar/gkt414>
- Khare, V., Tabassum, S., Chatterjee, U., Chatterjee, S., Ghosh, M.K., 2019. RNA helicase p68 deploys  $\beta$ -catenin in regulating RelA/p65 gene expression: implications in colon cancer. *J. Exp. Clin. Cancer Res.* 38, 330. <https://doi.org/10.1186/s13046-019-1304-y>
- Kiltschewskij, D.J., Harrison, P.F., Fitzsimmons, C., Beilharz, T.H., Cairns, M.J., 2023. Extension of mRNA poly(A) tails and 3'UTRs during neuronal differentiation exhibits variable association with post-transcriptional dynamics. *Nucleic Acids Res.* 51, 8181–8198. <https://doi.org/10.1093/nar/gkad499>
- Kim, H.-P., Cho, G.-A., Han, S.-W., Shin, J.-Y., Jeong, E.-G., Song, S.-H., Lee, W.-C., Lee, K.-H., Bang, D., Seo, J.-S., Kim, J.-I., Kim, T.-Y., 2014. Novel fusion transcripts in human gastric cancer revealed by transcriptome analysis. *Oncogene* 33, 5434–5441. <https://doi.org/10.1038/onc.2013.490>



- Kirstein, N., Gomes Dos Santos, H., Blumenthal, E., Shiekhatar, R., 2021. The Integrator complex at the crossroad of coding and noncoding RNA. *Curr. Opin. Cell Biol., Cell Nucleus* 70, 37–43. <https://doi.org/10.1016/j.ceb.2020.11.003>
- Konarska, M.M., Padgett, R.A., Sharp, P.A., 1985. Trans splicing of mRNA precursors in vitro. *Cell* 42, 165–171. [https://doi.org/10.1016/s0092-8674\(85\)80112-4](https://doi.org/10.1016/s0092-8674(85)80112-4)
- Krause, M., Hirsh, D., 1987. A trans-spliced leader sequence on actin mRNA in *C. elegans*. *Cell* 49, 753–761. [https://doi.org/10.1016/0092-8674\(87\)90613-1](https://doi.org/10.1016/0092-8674(87)90613-1)
- Krogan, N.J., Kim, M., Tong, A., Golshani, A., Cagney, G., Canadien, V., Richards, D.P., Beattie, B.K., Emili, A., Boone, C., Shilatifard, A., Buratowski, S., Greenblatt, J., 2003. Methylation of histone H3 by Set2 in *Saccharomyces cerevisiae* is linked to transcriptional elongation by RNA polymerase II. *Mol. Cell. Biol.* 23, 4207–4218. <https://doi.org/10.1128/MCB.23.12.4207-4218.2003>
- Kumar, S., Vo, A.D., Qin, F., Li, H., 2016. Comparative assessment of methods for the fusion transcripts detection from RNA-Seq data. *Sci. Rep.* 6, 21597. <https://doi.org/10.1038/srep21597>
- Kurzrock, R., Kantarjian, H.M., Druker, B.J., Talpaz, M., 2003. Philadelphia Chromosome Positive Leukemias: From Basic Mechanisms to Molecular Therapeutics. *Ann. Intern. Med.* 138, 819–830. <https://doi.org/10.7326/0003-4819-138-10-200305200-00010>
- Lackford, B., Yao, C., Charles, G.M., Weng, L., Zheng, X., Choi, E.-A., Xie, X., Wan, J., Xing, Y., Freudenberg, J.M., Yang, P., Jothi, R., Hu, G., Shi, Y., 2014. Fip1 regulates mRNA alternative polyadenylation to promote stem cell self-renewal. *EMBO J.* 33, 878–889. <https://doi.org/10.1002/embj.201386537>
- Lai, J., An, J., Seim, I., Walpole, C., Hoffman, A., Moya, L., Srinivasan, S., Perry-Keene, J.L., Wang, C., Lehman, M.L., Nelson, C.C., Clements, J.A., Batra, J., Australian Prostate Cancer Bioresource, 2015. Fusion transcript loci share many genomic features with non-fusion loci. *BMC Genomics* 16, 1021. <https://doi.org/10.1186/s12864-015-2235-4>
- Lambert, M.-P., Terrone, S., Giraud, G., Benoit-Pilven, C., Cluet, D., Combaret, V., Mortreux, F., Auboeuf, D., Bourgeois, C.F., 2018. The RNA helicase DDX17 controls the transcriptional activity of REST and the expression of proneural microRNAs in neuronal differentiation. *Nucleic Acids Res.* 46, 7686–7700. <https://doi.org/10.1093/nar/gky545>
- Lamm, G.M., Nicol, S.M., Fuller-Pace, F.V., Lamond, A.I., 1996. p72: a human nuclear DEAD box protein highly related to p68. *Nucleic Acids Res.* 24, 3739–3747.
- Laurent, F.-X., Sureau, A., Klein, A.F., Trouslard, F., Gasnier, E., Furling, D., 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. <https://doi.org/10.1093/nar/gkr1228>
- Lee, C.-G., 2002. RH70, a Bidirectional RNA Helicase, Co-purifies with U1snRNP\*. *J. Biol. Chem.* 277, 39679–39683. <https://doi.org/10.1074/jbc.C200337200>

- Lee, Y., Rio, D.C., 2015. Mechanisms and Regulation of Alternative Pre-mRNA Splicing. *Annu. Rev. Biochem.* 84, 291–323. <https://doi.org/10.1146/annurev-biochem-060614-034316>
- Lei, Q., Li, C., Zuo, Z., Huang, C., Cheng, H., Zhou, R., 2016. Evolutionary Insights into RNA trans-Splicing in Vertebrates. *Genome Biol. Evol.* 8, 562–577. <https://doi.org/10.1093/gbe/evw025>
- Lemieux, B., Blanchette, M., Monette, A., Mouland, A.J., Wellinger, R.J., Chabot, B., 2015. A Function for the hnRNP A1/A2 Proteins in Transcription Elongation. *PLoS ONE* 10, e0126654. <https://doi.org/10.1371/journal.pone.0126654>
- Li, C., Ge, L., Li, P., Wang, Y., Sun, M., Huang, L., Ishag, H., Di, D., Shen, Z., Fan, W., Mao, X., 2013. The DEAD-box RNA helicase DDX5 acts as a positive regulator of Japanese encephalitis virus replication by binding to viral 3' UTR. *Antiviral Res.* 100, 487–499. <https://doi.org/10.1016/j.antiviral.2013.09.002>
- Li, H., Wang, J., Mor, G., Sklar, J., 2008. A Neoplastic Gene Fusion Mimics Trans-Splicing of RNAs in Normal Human Cells. *Science* 321, 1357–1361. <https://doi.org/10.1126/science.1156725>
- Li, K., Zhao, G., Yuan, H., Zhang, J., Li, Q., Gong, D., Lin, P., 2022. Upregulated expression of DDX5 predicts recurrence and poor prognosis in breast cancer. *Pathol. - Res. Pract.* 229, 153736. <https://doi.org/10.1016/j.prp.2021.153736>
- Li, W., You, B., Hoque, M., Zheng, D., Luo, W., Ji, Z., Park, J.Y., Gunderson, S.I., Kalsotra, A., Manley, J.L., Tian, B., 2015. Systematic Profiling of Poly(A)+ Transcripts Modulated by Core 3' End Processing and Splicing Factors Reveals Regulatory Rules of Alternative Cleavage and Polyadenylation. *PLOS Genet.* 11, e1005166. <https://doi.org/10.1371/journal.pgen.1005166>
- Li, X., Zhao, L., Jiang, H., Wang, W., 2009. Short Homologous Sequences Are Strongly Associated with the Generation of Chimeric RNAs in Eukaryotes. *J. Mol. Evol.* 68, 56–65. <https://doi.org/10.1007/s00239-008-9187-0>
- Licatalosi, D.D., Mele, A., Fak, J.J., Ule, J., Kayikci, M., Chi, S.W., Clark, T.A., Schweitzer, A.C., Blume, J.E., Wang, X., Darnell, J.C., Darnell, R.B., 2008. HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* 456, 464–469. <https://doi.org/10.1038/nature07488>
- Lin, C., Yang, L., Yang, J.J., Huang, Y., 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. <https://doi.org/10.1128/MCB.25.17.7484-7493.2005>
- Lin, S., Tian, L., Shen, H., Gu, Y., Li, J.-L., Chen, Z., Sun, X., James You, M., Wu, L., 2013. DDX5 is a positive regulator of oncogenic NOTCH1 signaling in T cell acute lymphoblastic leukemia. *Oncogene* 32, 4845–4853. <https://doi.org/10.1038/onc.2012.482>
- Linder, P., Jankowsky, E., 2011. From unwinding to clamping — the DEAD box RNA helicase family. *Nat. Rev. Mol. Cell Biol.* 12, 505–516. <https://doi.org/10.1038/nrm3154>

- Linke, R., Limmer, M., Juranek, S.A., Heine, A., Paeschke, K., 2021. The Relevance of G-Quadruplexes for DNA Repair. *Int. J. Mol. Sci.* 22, 12599. <https://doi.org/10.3390/ijms222212599>
- Liu, C., Wang, L., Jiang, Q., Zhang, J., Zhu, L., Lin, L., Jiang, H., Lin, D., Xiao, Y., Fang, W., Guo, S., 2019. Hepatoma-Derived Growth Factor and DDX5 Promote Carcinogenesis and Progression of Endometrial Cancer by Activating  $\beta$ -Catenin. *Front. Oncol.* 9, 211. <https://doi.org/10.3389/fonc.2019.00211>
- Liu, F., Putnam, A., Jankowsky, E., 2008. ATP hydrolysis is required for DEAD-box protein recycling but not for duplex unwinding. *Proc. Natl. Acad. Sci.* 105, 20209–20214. <https://doi.org/10.1073/pnas.0811115106>
- Liu, X., Tesfai, J., Evrard, Y.A., Dent, S.Y.R., Martinez, E., 2003. c-Myc Transformation Domain Recruits the Human STAGA Complex and Requires TRRAP and GCN5 Acetylase Activity for Transcription Activation \*. *J. Biol. Chem.* 278, 20405–20412. <https://doi.org/10.1074/jbc.M211795200>
- 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. <https://doi.org/10.1128/MCB.22.15.5443-5450.2002>
- Lode, H.N., Ehlert, K., Huber, S., Troschke-Meurer, S., Siebert, N., Zumpe, M., Loibner, H., Ladenstein, R., 2023. Long-term, continuous infusion of single-agent dinutuximab beta for relapsed/refractory neuroblastoma: an open-label, single-arm, Phase 2 study. *Br. J. Cancer* 1–7. <https://doi.org/10.1038/s41416-023-02457-x>
- London, W. b., Castleberry, R. p., Matthay, K. k., Look, A. t., Seeger, R. c., Shimada, H., Thorner, P., Brodeur, G., Maris, J. m., Reynolds, C. p., Cohn, S. l., 2005. Evidence for an Age Cutoff Greater Than 365 Days for Neuroblastoma Risk Group Stratification in the Children's Oncology Group. *J. Clin. Oncol.* 23, 6459–6465. <https://doi.org/10.1200/JCO.2005.05.571>
- Lorgeoux, R.-P., Pan, Q., Le Duff, Y., Liang, C., 2013. DDX17 promotes the production of infectious HIV-1 particles through modulating viral RNA packaging and translation frameshift. *Virology* 443, 384–392. <https://doi.org/10.1016/j.virol.2013.05.026>
- Lou, H., Gagel, R.F., Berget, S.M., 1996. An intron enhancer recognized by splicing factors activates polyadenylation. *Genes Dev.* 10, 208–219. <https://doi.org/10.1101/gad.10.2.208>
- Lu, M., Boschetti, C., Tunnacliffe, A., 2015. Long Term Aggresome Accumulation Leads to DNA Damage, p53-dependent Cell Cycle Arrest, and Steric Interference in Mitosis\*. *J. Biol. Chem.* 290, 27986–28000. <https://doi.org/10.1074/jbc.M115.676437>
- Lu, W.-T., Hawley, B.R., Skalka, G.L., Baldock, R.A., Smith, E.M., Bader, A.S., Malewicz, M., Watts, F.Z., Wilczynska, A., Bushell, M., 2018. Drosha drives the formation of DNA:RNA hybrids around DNA break sites to facilitate DNA repair. *Nat. Commun.* 9, 532. <https://doi.org/10.1038/s41467-018-02893-x>
- Lu, Y., Zhang, Y., Xiang, X., Sharma, M., Liu, K., Wei, J., Shao, D., Li, B., Tong, G., Olszewski, M.A., Ma, Z., Qiu, Y., 2020. Notch signaling contributes to the expression of

inflammatory cytokines induced by highly pathogenic porcine reproductive and respiratory syndrome virus (HP-PRRSV) infection in porcine alveolar macrophages. *Dev. Comp. Immunol.* 108, 103690. <https://doi.org/10.1016/j.dci.2020.103690>

Luo, H., Zhu, G., Eshelman, M.A., Fung, T.K., Lai, Q., Wang, F., Zeisig, B.B., Lesperance, J., Ma, X., Chen, S., Cesari, N., Cogle, C., Chen, B., Xu, B., Yang, F.-C., So, C.W.E., Qiu, Y., Xu, M., Huang, S., 2022. HOTTIP-dependent R-loop formation regulates CTCF boundary activity and TAD integrity in leukemia. *Mol. Cell* 82, 833-851.e11. <https://doi.org/10.1016/j.molcel.2022.01.014>

Ma, L., Yang, S., Zhao, W., Tang, Z., Zhang, T., Li, K., 2012. Identification and analysis of pig chimeric mRNAs using RNA sequencing data. *BMC Genomics* 13, 429. <https://doi.org/10.1186/1471-2164-13-429>

Ma, L., Zhao, X., Wang, S., Zheng, Y., Yang, S., Hou, Y., Zou, B., Dong, L., 2020. Decreased expression of DEAD-Box helicase 5 inhibits esophageal squamous cell carcinomas by regulating endoplasmic reticulum stress and autophagy. *Biochem. Biophys. Res. Commun.* 533, 1449–1456. <https://doi.org/10.1016/j.bbrc.2020.10.026>

Macheret, M., Halazonetis, T.D., 2018. Intragenic origins due to short G1 phases underlie oncogene-induced DNA replication stress. *Nature* 555, 112–116. <https://doi.org/10.1038/nature25507>

Mah, S., Nelson, M.R., DeLisi, L.E., Reneland, R.H., Markward, N., James, M.R., Nyholt, D.R., Hayward, N., Handoko, H., Mowry, B., Kammerer, S., Braun, A., 2006. Identification of the semaphorin receptor PLXNA2 as a candidate for susceptibility to schizophrenia. *Mol. Psychiatry* 11, 471–478. <https://doi.org/10.1038/sj.mp.4001785>

Maher, C.A., Kumar-Sinha, C., Cao, X., Kalyana-Sundaram, S., Han, B., Jing, X., Sam, L., Barrette, T., Palanisamy, N., Chinnaiyan, A.M., 2009. Transcriptome sequencing to detect gene fusions in cancer. *Nature* 458, 97–101. <https://doi.org/10.1038/nature07638>

Marasco, L.E., Kornblihtt, A.R., 2023. The physiology of alternative splicing. *Nat. Rev. Mol. Cell Biol.* 24, 242–254. <https://doi.org/10.1038/s41580-022-00545-z>

Markus, M.A., Heinrich, B., Raitskin, O., Adams, D.J., Mangs, H., Goy, C., Lodomery, M., Sperling, R., Stamm, S., Morris, B.J., 2006. WT1 interacts with the splicing protein RBM4 and regulates its ability to modulate alternative splicing in vivo. *Exp. Cell Res.* 312, 3379–3388. <https://doi.org/10.1016/j.yexcr.2006.07.008>

Maselli, R.A., Vázquez, J., Schruppf, L., Arredondo, J., Lara, M., Strober, J.B., Pytel, P., Wollmann, R.L., Ferns, M., 2018. Presynaptic congenital myasthenic syndrome with altered synaptic vesicle homeostasis linked to compound heterozygous sequence variants in RPH3A. *Mol. Genet. Genomic Med.* 6, 434–440. <https://doi.org/10.1002/mgg3.370>

Mata, M. de la, Alonso, C.R., Kadener, S., Fededa, J.P., Blaustein, M., Pelisch, F., Cramer, P., Bentley, D., Kornblihtt, A.R., 2003. A Slow RNA Polymerase II Affects Alternative Splicing In Vivo. *Mol. Cell* 12, 525–532. <https://doi.org/10.1016/j.molcel.2003.08.001>

- Matthay, K.K., Maris, J.M., Schleiermacher, G., Nakagawara, A., Mackall, C.L., Diller, L., Weiss, W.A., 2016. Neuroblastoma. *Nat. Rev. Dis. Primer* 2, 16078. <https://doi.org/10.1038/nrdp.2016.78>
- Mayr, C., Bartel, D.P., 2009. Widespread Shortening of 3'UTRs by Alternative Cleavage and Polyadenylation Activates Oncogenes in Cancer Cells. *Cell* 138, 673–684. <https://doi.org/10.1016/j.cell.2009.06.016>
- McMahon, S.B., Buskirk, H.A.V., Dugan, K.A., Copeland, T.D., Cole, M.D., 1998. The Novel ATM-Related Protein TRRAP Is an Essential Cofactor for the c-Myc and E2F Oncoproteins. *Cell* 94, 363–374. [https://doi.org/10.1016/S0092-8674\(00\)81479-8](https://doi.org/10.1016/S0092-8674(00)81479-8)
- Meers, M.P., Henriques, T., Lavender, C.A., McKay, D.J., Strahl, B.D., Duronio, R.J., Adelman, K., Matera, A.G., 2017. Histone gene replacement reveals a post-transcriptional role for H3K36 in maintaining metazoan transcriptome fidelity. *eLife* 6, e23249. <https://doi.org/10.7554/eLife.23249>
- Mendoza-Figueroa, M.S., Tatomer, D.C., Wilusz, J.E., 2020. The Integrator Complex in Transcription and Development. *Trends Biochem. Sci.* 45, 923–934. <https://doi.org/10.1016/j.tibs.2020.07.004>
- Misteli, T., 2020. The self-organizing genome: Principles of genome architecture and function. *Cell* 183, 28–45. <https://doi.org/10.1016/j.cell.2020.09.014>
- Misteli, T., Spector, D.L., 1999. RNA polymerase II targets pre-mRNA splicing factors to transcription sites in vivo. *Mol. Cell* 3, 697–705. [https://doi.org/10.1016/s1097-2765\(01\)80002-2](https://doi.org/10.1016/s1097-2765(01)80002-2)
- Mitschka, S., Mayr, C., 2022. Context-specific regulation and function of mRNA alternative polyadenylation. *Nat. Rev. Mol. Cell Biol.* 23, 779–796. <https://doi.org/10.1038/s41580-022-00507-5>
- Mooney, S.M., Grande, J.P., Salisbury, J.L., Janknecht, R., 2010. Sumoylation of p68 and p72 RNA helicases affects protein stability and transactivation potential. *Biochemistry* 49, 1–10. <https://doi.org/10.1021/bi901263m>
- Morgan, M., Shiekhhattar, R., Shilatifard, A., Lauberth, S.M., 2022. It's a DoG-eat-DoG world—altered transcriptional mechanisms drive downstream-of-gene (DoG) transcript production. *Mol. Cell* 82, 1981–1991. <https://doi.org/10.1016/j.molcel.2022.04.008>
- Morimachi, M., Hirabayashi, K., Takanashi, Y., Kawanishi, A., Saika, T., Ueyama, Y., Nakagohri, T., Nakamura, N., Suzuki, H., Kagawa, T., 2021. Low expression of DDX5 is associated with poor prognosis in patients with pancreatic ductal adenocarcinoma. *J. Clin. Pathol.* 74, 741–745. <https://doi.org/10.1136/jclinpath-2020-207002>
- Mossé, Y.P., Deyell, R.J., Berthold, F., Nagakawara, A., Ambros, P.F., Monclair, T., Cohn, S.L., Pearson, A.D., London, W.B., Matthay, K.K., 2014. Neuroblastoma in older children, adolescents and young adults: A report from the International Neuroblastoma Risk Group project. *Pediatr. Blood Cancer* 61, 627–635. <https://doi.org/10.1002/pbc.24777>
- Mossé, Y.P., Laudenslager, M., Longo, L., Cole, K.A., Wood, A., Attiyeh, E.F., Laquaglia, M.J., Sennett, R., Lynch, J.E., Perri, P., Laureys, G., Speleman, F., Kim, C., Hou, C.,

- Hakonarson, H., Torkamani, A., Schork, N.J., Brodeur, G.M., Tonini, G.P., Rappaport, E., Devoto, M., Maris, J.M., 2008. Identification of ALK as a major familial neuroblastoma predisposition gene. *Nature* 455, 930–935. <https://doi.org/10.1038/nature07261>
- Muniz, L., Nicolas, E., Trouche, D., 2021. RNA polymerase II speed: a key player in controlling and adapting transcriptome composition. *EMBO J.* 40, e105740. <https://doi.org/10.15252/embj.2020105740>
- Nacu, S., Yuan, W., Kan, Z., Bhatt, D., Rivers, C.S., Stinson, J., Peters, B.A., Modrusan, Z., Jung, K., Seshagiri, S., Wu, T.D., 2011. Deep RNA sequencing analysis of readthrough gene fusions in human prostate adenocarcinoma and reference samples. *BMC Med. Genomics* 4, 11. <https://doi.org/10.1186/1755-8794-4-11>
- Nagaike, T., Logan, C., Hotta, I., Rozenblatt-Rosen, O., Meyerson, M., Manley, J.L., 2011. Transcriptional Activators Enhance Polyadenylation of mRNA Precursors. *Mol. Cell* 41, 409–418. <https://doi.org/10.1016/j.molcel.2011.01.022>
- Nanavaty, V., Abrash, E.W., Hong, C., Park, S., Fink, E.E., Li, Z., Sweet, T.J., Bhasin, J.M., Singuri, S., Lee, B.H., Hwang, T.H., Ting, A.H., 2020. DNA Methylation Regulates Alternative Polyadenylation via CTCF and the Cohesin Complex. *Mol. Cell* 78, 752–764.e6. <https://doi.org/10.1016/j.molcel.2020.03.024>
- Nemeroff, M.E., Barabino, S.M.L., Li, Y., Keller, W., Krug, R.M., 1998. Influenza Virus NS1 Protein Interacts with the Cellular 30 kDa Subunit of CPSF and Inhibits 3' End Formation of Cellular Pre-mRNAs. *Mol. Cell* 1, 991–1000. [https://doi.org/10.1016/S1097-2765\(00\)80099-4](https://doi.org/10.1016/S1097-2765(00)80099-4)
- Neubauer, G., King, A., Rappsilber, J., Calvio, C., Watson, M., Ajuh, P., Sleeman, J., Lamond, A., Mann, M., 1998. Mass spectrometry and EST-database searching allows characterization of the multi-protein spliceosome complex. *Nat. Genet.* 20, 46–50. <https://doi.org/10.1038/1700>
- Ngo, T.D., Partin, A.C., Nam, Y., 2019. RNA Specificity and Autoregulation of DDX17, a Modulator of MicroRNA Biogenesis. *Cell Rep.* 29, 4024–4035.e5. <https://doi.org/10.1016/j.celrep.2019.11.059>
- Nicol, S.M., Bray, S.E., Black, H.D., Lorimore, S.A., Wright, E.G., Lane, D.P., Meek, D.W., Coates, P.J., Fuller-Pace, F.V., 2013. The RNA helicase p68 (DDX5) is selectively required for the induction of p53-dependent p21 expression and cell cycle arrest after DNA damage. *Oncogene* 32, 3461. <https://doi.org/10.1038/onc.2012.426>
- Noah, D.L., Twu, K.Y., Krug, R.M., 2003. Cellular antiviral responses against influenza A virus are countered at the posttranscriptional level by the viral NS1A protein via its binding to a cellular protein required for the 3' end processing of cellular pre-mRNAs. *Virology* 307, 386–395. [https://doi.org/10.1016/S0042-6822\(02\)00127-7](https://doi.org/10.1016/S0042-6822(02)00127-7)
- Oktaba, K., Zhang, W., Lotz, T.S., Jun, D.J., Lemke, S.B., Ng, S.P., Esposito, E., Levine, M., Hilgers, V., 2015. ELAV Links Paused Pol II to Alternative Polyadenylation in the *Drosophila* Nervous System. *Mol. Cell* 57, 341–348. <https://doi.org/10.1016/j.molcel.2014.11.024>

Olesnicky, E.C., Killian, D.J., Garcia, E., Morton, M.C., Rathjen, A.R., Sola, I.E., Gavis, E.R., 2013. Extensive Use of RNA-Binding Proteins in *Drosophila* Sensory Neuron Dendrite Morphogenesis. *G3 GenesGenomesGenetics* 4, 297–306.

<https://doi.org/10.1534/g3.113.009795>

Otto, T., Horn, S., Brockmann, M., Eilers, U., Schüttrumpf, L., Popov, N., Kenney, A.M., Schulte, J.H., Beijersbergen, R., Christiansen, H., Berwanger, B., Eilers, M., 2009.

Stabilization of N-Myc Is a Critical Function of Aurora A in Human Neuroblastoma. *Cancer Cell* 15, 67–78. <https://doi.org/10.1016/j.ccr.2008.12.005>

Padmanabhan, K., Robles, M.S., Westerling, T., Weitz, C.J., 2012. Feedback Regulation of Transcriptional Termination by the Mammalian Circadian Clock PERIOD Complex. *Science* 337, 599–602. <https://doi.org/10.1126/science.1221592>

<https://doi.org/10.1126/science.1221592>

Paine, I., Posey, J.E., Grochowski, C.M., Jhangiani, S.N., Rosenheck, S., Kleyner, R., Marmorale, T., Yoon, M., Wang, K., Robison, R., Cappuccio, G., Pinelli, M., Magli, A., Akdemir, Z.C., Hui, J., Yeung, W.L., Wong, B.K.Y., Ortega, L., Bekheirnia, M.R., Bierhals, T., Hempel, M., Johannsen, J., Santer, R., Aktas, D., Alikasifoglu, M., Bozdogan, S., Aydin, H., Karaca, E., Bayram, Y., Ityel, H., Dorschner, M., White, J.J., Wilichowski, E., Wortmann, S.B., Casella, E.B., Kitajima, J.P., Kok, F., Monteiro, F., Muzny, D.M., Bamshad, M., Gibbs, R.A., Sutton, V.R., Esch, H.V., Brunetti-Pierri, N., Hildebrandt, F., Brautbar, A., Veyver, I.B.V. den, Glass, I., Lessel, D., Lyon, G.J., Lupski, J.R., 2019. Paralog Studies Augment Gene Discovery: DDX and DHX Genes. *Am. J. Hum. Genet.* 105, 302–316. <https://doi.org/10.1016/j.ajhg.2019.06.001>

Pan, L., Huang, X., Liu, Z.-X., Ye, Y., Li, R., Zhang, J., Wu, G., Bai, R., Zhuang, L., Wei, L., Li, M., Zheng, Y., Su, J., Deng, J., Deng, S., Zeng, L., Zhang, S., Wu, C., Che, X., Wang, C., Chen, R., Lin, D., Zheng, J., 2021. Inflammatory cytokine-regulated tRNA-derived fragment tRF-21 suppresses pancreatic ductal adenocarcinoma progression. *J. Clin. Invest.* 131, e148130. <https://doi.org/10.1172/JCI148130>

Papadopoulos, D., Solvie, D., Baluapuri, A., Endres, T., Ha, S.A., Herold, S., Kalb, J., Giansanti, C., Schülein-Völk, C., Ade, C.P., Schneider, C., Gaballa, A., Vos, S., Fischer, U., Dobbstein, M., Wolf, E., Eilers, M., 2022. MYCN recruits the nuclear exosome complex to RNA polymerase II to prevent transcription-replication conflicts. *Mol. Cell* 82, 159–176.e12. <https://doi.org/10.1016/j.molcel.2021.11.002>

Papadopoulos, D., Uhl, L., Ha, S.A., Eilers, M., 2023. Beyond gene expression: how MYC relieves transcription stress. *Trends Cancer* 9, 805–816.

<https://doi.org/10.1016/j.trecan.2023.06.008>

Pérez Cañadillas, J.M., Varani, G., 2003. Recognition of GU-rich polyadenylation regulatory elements by human CstF-64 protein. *EMBO J.* 22, 2821–2830.

<https://doi.org/10.1093/emboj/cdg259>

Perry, R.B.-T., Doron-Mandel, E., Iavnilovitch, E., Rishal, I., Dagan, S.Y., Tsoory, M., Coppola, G., McDonald, M.K., Gomes, C., Geschwind, D.H., Twiss, J.L., Yaron, A., Fainzilber, M., 2012. Subcellular Knockout of Importin  $\beta$ 1 Perturbs Axonal Retrograde Signaling. *Neuron* 75, 294–305. <https://doi.org/10.1016/j.neuron.2012.05.033>

- Pezone, A., Zuchegna, C., Tramontano, A., Romano, A., Russo, G., de Rosa, M., Vinciguerra, M., Porcellini, A., Gottesman, M.E., Avvedimento, E.V., 2019. RNA Stabilizes Transcription-Dependent Chromatin Loops Induced By Nuclear Hormones. *Sci. Rep.* 9, 3925. <https://doi.org/10.1038/s41598-019-40123-6>
- Polenkowski, M., Allister, A.B., Burbano de Lara, S., Pierce, A., Geary, B., El Bounkari, O., Wiehlmann, L., Hoffmann, A., Whetton, A.D., Tamura, T., Tran, D.D.H., 2023. THOC5 complexes with DDX5, DDX17, and CDK12 to regulate R loop structures and transcription elongation rate. *iScience* 26, 105784. <https://doi.org/10.1016/j.isci.2022.105784>
- Qin, F., Song, Y., Zhang, Y., Facemire, L., Frierson, H., Li, H., 2016. Role of CTCF in Regulating SLC45A3-ELK4 Chimeric RNA. *PLOS ONE* 11, e0150382. <https://doi.org/10.1371/journal.pone.0150382>
- Qin, F., Song, Z., Babiceanu, M., Song, Y., Facemire, L., Singh, R., Adli, M., Li, H., 2015. Discovery of CTCF-Sensitive Cis-Spliced Fusion RNAs between Adjacent Genes in Human Prostate Cells. *PLOS Genet.* 11, e1005001. <https://doi.org/10.1371/journal.pgen.1005001>
- Qin, F., Zhang, Y., Liu, J., Li, H., 2017. SLC45A3-ELK4 functions as a long non-coding chimeric RNA. *Cancer Lett.* 404, 53–61. <https://doi.org/10.1016/j.canlet.2017.07.007>
- Raab-Graham, K.F., Haddick, P.C.G., Jan, Y.N., Jan, L.Y., 2006. Activity- and mTOR-Dependent Suppression of Kv1.1 Channel mRNA Translation in Dendrites. *Science* 314, 144–148. <https://doi.org/10.1126/science.1131693>
- Rahman, M.M., Bagdassarian, E., Ali, M.A.M., McFadden, G., 2017. Identification of host DEAD-box RNA helicases that regulate cellular tropism of oncolytic Myxoma virus in human cancer cells. *Sci. Rep.* 7, 15710. <https://doi.org/10.1038/s41598-017-15941-1>
- Rambout, X., Dequiedt, F., Maquat, L.E., 2018. Beyond Transcription: Roles of Transcription Factors in Pre-mRNA Splicing. *Chem. Rev.* 118, 4339–4364. <https://doi.org/10.1021/acs.chemrev.7b00470>
- Ramírez, F., Bhardwaj, V., Arrigoni, L., Lam, K.C., Grüning, B.A., Villaveces, J., Habermann, B., Akhtar, A., Manke, T., 2018. High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nat. Commun.* 9, 189. <https://doi.org/10.1038/s41467-017-02525-w>
- Rang, F.J., Kloosterman, W.P., de Ridder, J., 2018. From squiggle to basepair: computational approaches for improving nanopore sequencing read accuracy. *Genome Biol.* 19, 1–11. <https://doi.org/10.1186/s13059-018-1462-9>
- Ray-Soni, A., Bellecourt, M.J., Landick, R., 2016. Mechanisms of Bacterial Transcription Termination: All Good Things Must End. *Annu. Rev. Biochem.* 85, 319–347. <https://doi.org/10.1146/annurev-biochem-060815-014844>
- Reimer, K.A., Mimoso, C.A., Adelman, K., Neugebauer, K.M., 2021. Co-transcriptional splicing regulates 3' end cleavage during mammalian erythropoiesis. *Mol. Cell* 81, 998-1012.e7. <https://doi.org/10.1016/j.molcel.2020.12.018>



Richards, M.W., Burgess, S.G., Poon, E., Carstensen, A., Eilers, M., Chesler, L., Bayliss, R., 2016. Structural basis of N-Myc binding by Aurora-A and its destabilization by kinase inhibitors. *Proc. Natl. Acad. Sci.* 113, 13726–13731.

<https://doi.org/10.1073/pnas.1610626113>

Rickman, D.S., Pflueger, D., Moss, B., VanDoren, V.E., Chen, C.X., de la Taille, A., Kuefer, R., Tewari, A.K., Setlur, S.R., Demichelis, F., Rubin, M.A., 2009. SLC45A3-ELK4 Is a Novel and Frequent Erythroblast Transformation–Specific Fusion Transcript in Prostate Cancer. *Cancer Res.* 69, 2734–2738. <https://doi.org/10.1158/0008-5472.CAN-08-4926>

Robson-Dixon, N.D., Garcia-Blanco, M.A., 2004. MAZ elements alter transcription elongation and silencing of the fibroblast growth factor receptor 2 exon IIIb. *J. Biol. Chem.* 279, 29075–29084. <https://doi.org/10.1074/jbc.M312747200>

Roeschert, I., Poon, E., Henssen, A.G., Dorado Garcia, H., Gatti, M., Giansanti, C., Jamin, Y., Ade, C.P., Gallant, P., Schülein-Völk, C., Beli, P., Richards, M., Rosenfeldt, M., Altmeyer, M., Anderson, J., Eggert, A., Dobbelstein, M., Bayliss, R., Chesler, L., Büchel, G., Eilers, M., 2021. Combined inhibition of Aurora-A and ATR kinases results in regression of MYCN-amplified neuroblastoma. *Nat. Cancer* 2, 312–326. <https://doi.org/10.1038/s43018-020-00171-8>

Rosa-Mercado, N.A., Zimmer, J.T., Apostolidi, M., Rinehart, J., Simon, M.D., Steitz, J.A., 2021. Hyperosmotic stress alters the RNA polymerase II interactome and induces readthrough transcription despite widespread transcriptional repression. *Mol. Cell* 81, 502–513.e4. <https://doi.org/10.1016/j.molcel.2020.12.002>

Rutkowski, A.J., Erhard, F., L'Hernault, A., Bonfert, T., Schilhabel, M., Crump, C., Rosenstiel, P., Efstathiou, S., Zimmer, R., Friedel, C.C., Dölken, L., 2015. Widespread disruption of host transcription termination in HSV-1 infection. *Nat. Commun.* 6, 7126. <https://doi.org/10.1038/ncomms8126>

Saldi, T., Riemondy, K., Erickson, B., Bentley, D.L., 2021. Alternative RNA structures formed during transcription depend on elongation rate and modify RNA processing. *Mol. Cell* 81, 1789–1801.e5. <https://doi.org/10.1016/j.molcel.2021.01.040>

Sandberg, R., Neilson, J.R., Sarma, A., Sharp, P.A., Burge, C.B., 2008. Proliferating cells express mRNAs with shortened 3' UTRs and fewer microRNA target sites. *Science* 320, 1643–1647. <https://doi.org/10.1126/science.1155390>

Sanz, L.A., Hartono, S.R., Lim, Y.W., Steyaert, S., Rajpurkar, A., Ginno, P.A., Xu, X., Chédin, F., 2016. Prevalent, Dynamic, and Conserved R-Loop Structures Associate with Specific Epigenomic Signatures in Mammals. *Mol. Cell* 63, 167–178. <https://doi.org/10.1016/j.molcel.2016.05.032>

Sarkar, M., Khare, V., Guturi, K.K.N., Das, N., Ghosh, M.K., 2015. The DEAD box protein p68: a crucial regulator of AKT/FOXO3a signaling axis in oncogenesis. *Oncogene* 34, 5843–5856. <https://doi.org/10.1038/onc.2015.42>

Schäfer, P., Tüting, C., Schönemann, L., Kühn, U., Treiber, T., Treiber, N., Ihling, C., Graber, A., Keller, W., Meister, G., Sinz, A., Wahle, E., 2018. Reconstitution of

mammalian cleavage factor II involved in 3' processing of mRNA precursors. *RNA* 24, 1721–1737. <https://doi.org/10.1261/rna.068056.118>

Schaller, M.A., Neupane, R., Rudd, B.D., Kunkel, S.L., Kallal, L.E., Lincoln, P., Lowe, J.B., Man, Y., Lukacs, N.W., 2007. Notch ligand Delta-like 4 regulates disease pathogenesis during respiratory viral infections by modulating Th2 cytokines. *J. Exp. Med.* 204, 2925–2934. <https://doi.org/10.1084/jem.20070661>

Schmucker, D., Clemens, J.C., Shu, H., Worby, C.A., Xiao, J., Muda, M., Dixon, J.E., Zipursky, S.L., 2000. Drosophila Dscam Is an Axon Guidance Receptor Exhibiting Extraordinary Molecular Diversity. *Cell* 101, 671–684. [https://doi.org/10.1016/S0092-8674\(00\)80878-8](https://doi.org/10.1016/S0092-8674(00)80878-8)

Sessa, G., Gómez-González, B., Silva, S., Pérez-Calero, C., Beaupere, R., Barroso, S., Martineau, S., Martin, C., Ehlén, Å., Martínez, J.S., Lombard, B., Loew, D., Vagner, S., Aguilera, A., Carreira, A., 2021. BRCA2 promotes DNA-RNA hybrid resolution by DDX5 helicase at DNA breaks to facilitate their repair. *EMBO J.* 40, e106018. <https://doi.org/10.15252/embj.2020106018>

Shamseldin, H.E., Rajab, A., Alhashem, A., Shaheen, R., Al-Shidi, T., Alamro, R., Al Harassi, S., Alkuraya, F.S., 2013. Mutations in DDX59 Implicate RNA Helicase in the Pathogenesis of Orofaciodigital Syndrome. *Am. J. Hum. Genet.* 93, 555–560. <https://doi.org/10.1016/j.ajhg.2013.07.012>

Shi, Y., Giammartino, D.C.D., Taylor, D., Sarkeshik, A., Rice, W.J., Yates, J.R., Frank, J., Manley, J.L., 2009. Molecular Architecture of the Human Pre-mRNA 3' Processing Complex. *Mol. Cell* 33, 365–376. <https://doi.org/10.1016/j.molcel.2008.12.028>

Shi, Y., Yuan, J., Rraklli, V., Maxymovitz, E., Cipullo, M., Liu, M., Li, S., Westerlund, I., Bedoya-Reina, O.C., Bullova, P., Rorbach, J., Juhlin, C.C., Stenman, A., Larsson, C., Kogner, P., O'Sullivan, M.J., Schlisio, S., Holmberg, J., 2021. Aberrant splicing in neuroblastoma generates RNA-fusion transcripts and provides vulnerability to spliceosome inhibitors. *Nucleic Acids Res.* 49, 2509–2521. <https://doi.org/10.1093/nar/gkab054>

Shukla, S., Kavak, E., Gregory, M., Imashimizu, M., Shutinoski, B., Kashlev, M., Oberdoerffer, P., Sandberg, R., Oberdoerffer, S., 2011. CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature* 479, 74–79. <https://doi.org/10.1038/nature10442>

Singh, S., Qin, F., Kumar, S., Elfman, J., Lin, E., Pham, L.-P., Yang, A., Li, H., 2020. The landscape of chimeric RNAs in non-diseased tissues and cells. *Nucleic Acids Res.* 48, 1764–1778. <https://doi.org/10.1093/nar/gkz1223>

Sithole, N., Williams, C.A., Vaughan, A.M., Kenyon, J.C., Lever, A.M.L., 2018. DDX17 Specifically, and Independently of DDX5, Controls Use of the HIV A4/5 Splice Acceptor Cluster and Is Essential for Efficient Replication of HIV. *J. Mol. Biol.* 430, 3111–3128. <https://doi.org/10.1016/j.jmb.2018.06.052>

Skourti-Stathaki, K., Proudfoot, N.J., Gromak, N., 2011. Human Senataxin Resolves RNA/DNA Hybrids Formed at Transcriptional Pause Sites to Promote Xrn2-Dependent Termination. *Mol. Cell* 42, 794–805. <https://doi.org/10.1016/j.molcel.2011.04.026>

Smith, V., Foster, J., 2018. High-Risk Neuroblastoma Treatment Review. *Children* 5. <https://doi.org/10.3390/children5090114>

Snijders Blok, L., Madsen, E., Juusola, J., Gilissen, C., Baralle, D., Reijnders, M.R.F., Venselaar, H., Helsmoortel, C., Cho, M.T., Hoischen, A., Vissers, L.E.L.M., Koemans, T.S., Wissink-Lindhout, W., Eichler, E.E., Romano, C., Van Esch, H., Stumpel, C., Vreeburg, M., Smeets, E., Oberndorff, K., van Bon, B.W.M., Shaw, M., Gecz, J., Haan, E., Bienek, M., Jensen, C., Loeys, B.L., Van Dijck, A., Innes, A.M., Racher, H., Vermeer, S., Di Donato, N., Rump, A., Tatton-Brown, K., Parker, M.J., Henderson, A., Lynch, S.A., Fryer, A., Ross, A., Vasudevan, P., Kini, U., Newbury-Ecob, R., Chandler, K., Male, A., Dijkstra, S., Schieving, J., Giltay, J., van Gassen, K.L.I., Schuurs-Hoeijmakers, J., Tan, P.L., Padiaditakis, I., Haas, S.A., Retterer, K., Reed, P., Monaghan, K.G., Haverfield, E., Natowicz, M., Myers, A., Kruer, M.C., Stein, Q., Strauss, K.A., Brigatti, K.W., Keating, K., Burton, B.K., Kim, K.H., Charrow, J., Norman, J., Foster-Barber, A., Kline, A.D., Kimball, A., Zackai, E., Harr, M., Fox, J., McLaughlin, J., Lindstrom, K., Haude, K.M., van Roozendaal, K., Brunner, H., Chung, W.K., Kooy, R.F., Pfundt, R., Kalscheuer, V., Mehta, S.G., Katsanis, N., Kleefstra, T., 2015. Mutations in DDX3X Are a Common Cause of Unexplained Intellectual Disability with Gender-Specific Effects on Wnt Signaling. *Am. J. Hum. Genet.* 97, 343–352. <https://doi.org/10.1016/j.ajhg.2015.07.004>

Sormanni, P., Aprile, F.A., Vendruscolo, M., 2015. The CamSol Method of Rational Design of Protein Mutants with Enhanced Solubility. *J. Mol. Biol.* 427, 478–490. <https://doi.org/10.1016/j.jmb.2014.09.026>

Spies, N., Burge, C.B., Bartel, D.P., 2013. 3' UTR-isoform choice has limited influence on the stability and translational efficiency of most mRNAs in mouse fibroblasts. *Genome Res.* 23, 2078–2090. <https://doi.org/10.1101/gr.156919.113>

Sun, Y., Bell, J.L., Carter, D., Gherardi, S., Poulos, R.C., Milazzo, G., Wong, J.W.H., Al-Awar, R., Tee, A.E., Liu, P.Y., Liu, B., Atmadibrata, B., Wong, M., Trahair, T., Zhao, Q., Shohet, J.M., Haupt, Y., Schulte, J.H., Brown, P.J., Arrowsmith, C.H., Vedadi, M., MacKenzie, K.L., Hüttelmaier, S., Perini, G., Marshall, G.M., Braithwaite, A., Liu, T., 2015. WDR5 Supports an N-Myc Transcriptional Complex That Drives a Protumorigenic Gene Expression Signature in Neuroblastoma. *Cancer Res.* 75, 5143–5154. <https://doi.org/10.1158/0008-5472.CAN-15-0423>

Sun, Y., Li, H., 2022. Chimeric RNAs Discovered by RNA Sequencing and Their Roles in Cancer and Rare Genetic Diseases. *Genes* 13, 741. <https://doi.org/10.3390/genes13050741>

Suthapot, P., Xiao, T., Felsenfeld, G., Hongeng, S., Wongtrakoongate, P., 2022. The RNA helicases DDX5 and DDX17 facilitate neural differentiation of human pluripotent stem cells NTERA2. *Life Sci.* 291, 120298. <https://doi.org/10.1016/j.lfs.2021.120298>

Suto, F., Ito, K., Uemura, M., Shimizu, M., Shinkawa, Y., Sanbo, M., Shinoda, T., Tsuboi, M., Takashima, S., Yagi, T., Fujisawa, H., 2005. Plexin-A4 Mediates Axon-Repulsive

Activities of Both Secreted and Transmembrane Semaphorins and Plays Roles in Nerve Fiber Guidance. *J. Neurosci.* 25, 3628–3637. <https://doi.org/10.1523/JNEUROSCI.4480-04.2005>

Sutton, R.E., Boothroyd, J.C., 1986. Evidence for Trans splicing in trypanosomes. *Cell* 47, 527–535. [https://doi.org/10.1016/0092-8674\(86\)90617-3](https://doi.org/10.1016/0092-8674(86)90617-3)

Takagaki, Y., Seipelt, R.L., Peterson, M.L., Manley, J.L., 1996. The Polyadenylation Factor CstF-64 Regulates Alternative Processing of IgM Heavy Chain Pre-mRNA during B Cell Differentiation. *Cell* 87, 941–952. [https://doi.org/10.1016/S0092-8674\(00\)82000-0](https://doi.org/10.1016/S0092-8674(00)82000-0)

Tang, P., Yang, Y., Li, G., Huang, L., Wen, M., Ruan, W., Guo, X., Zhang, C., Zuo, X., Luo, D., Xu, Y., Fu, X.-D., Zhou, Y., 2022. Alternative polyadenylation by sequential activation of distal and proximal PolyA sites. *Nat. Struct. Mol. Biol.* 29, 21–31. <https://doi.org/10.1038/s41594-021-00709-z>

Tang, Y., Qin, F., Liu, A., Li, H., 2017. Recurrent fusion RNA DUS4L-BCAP29 in non-cancer human tissues and cells. *Oncotarget* 8, 31415–31423. <https://doi.org/10.18632/oncotarget.16329>

Terrone, S., Valat, J., Fontrodona, N., Giraud, G., Claude, J.-B., Combe, E., Lapendry, A., Polvèche, H., Ameer, L.B., Duvermy, A., Modolo, L., Bernard, P., Mortreux, F., Auboeuf, D., Bourgeois, C.F., 2022. RNA helicase-dependent gene looping impacts messenger RNA processing. *Nucleic Acids Res.* 50, 9226–9246. <https://doi.org/10.1093/nar/gkac717>

Teves, S.S., Henikoff, S., 2014. Transcription-generated torsional stress destabilizes nucleosomes. *Nat. Struct. Mol. Biol.* 21, 88–94. <https://doi.org/10.1038/nsmb.2723>

Tian, B., Graber, J.H., 2012. Signals for pre-mRNA cleavage and polyadenylation. *Wiley Interdiscip. Rev. RNA* 3, 385–396. <https://doi.org/10.1002/wrna.116>

Tolbert, V.P., Coggins, G.E., Maris, J.M., 2017. Genetic susceptibility to neuroblastoma. *Curr. Opin. Genet. Dev.* 42, 81–90. <https://doi.org/10.1016/j.gde.2017.03.008>

Tonelli, R., Purgato, S., Camerin, C., Fronza, R., Bologna, F., Alboresi, S., Franzoni, M., Corradini, R., Sforza, S., Faccini, A., Shohet, J.M., Marchelli, R., Pession, A., 2005. Anti-gene peptide nucleic acid specifically inhibits MYCN expression in human neuroblastoma cells leading to cell growth inhibition and apoptosis. *Mol. Cancer Ther.* 4, 779–786. <https://doi.org/10.1158/1535-7163.MCT-04-0213>

Tu, W.B., Helander, S., Pilstål, R., Hickman, K.A., Lourenco, C., Jurisica, I., Raught, B., Wallner, B., Sunnerhagen, M., Penn, L.Z., 2015. Myc and its interactors take shape. *Biochim. Biophys. Acta BBA - Gene Regul. Mech.*, Myc proteins in cell biology and pathology 1849, 469–483. <https://doi.org/10.1016/j.bbagr.2014.06.002>

Turunen, J.J., Niemelä, E.H., Verma, B., Frilander, M.J., 2013. The significant other: splicing by the minor spliceosome. *WIREs RNA* 4, 61–76. <https://doi.org/10.1002/wrna.1141>

- Twu, K.Y., Noah, D.L., Rao, P., Kuo, R.-L., Krug, R.M., 2006. The CPSF30 Binding Site on the NS1A Protein of Influenza A Virus Is a Potential Antiviral Target. *J. Virol.* 80, 3957–3965. <https://doi.org/10.1128/JVI.80.8.3957-3965.2006>
- Uhlmann-Schiffler, H., Rössler, O.G., Stahl, H., 2002. The mRNA of DEAD box protein p72 is alternatively translated into an 82-kDa RNA helicase. *J. Biol. Chem.* 277, 1066–1075. <https://doi.org/10.1074/jbc.M107535200>
- Ulianov, S.V., Khrameeva, E.E., Gavrilov, A.A., Flyamer, I.M., Kos, P., Mikhaleva, E.A., Penin, A.A., Logacheva, M.D., Imakaev, M.V., Chertovich, A., Gelfand, M.S., Shevelyov, Y.Y., Razin, S.V., 2016. Active chromatin and transcription play a key role in chromosome partitioning into topologically associating domains. *Genome Res.* 26, 70–84. <https://doi.org/10.1101/gr.196006.115>
- van Bokhoven, H., Celli, J., van Reeuwijk, J., Rinne, T., Glaudemans, B., van Beusekom, E., Rieu, P., Newbury-Ecob, R.A., Chiang, C., Brunner, H.G., 2005. MYCN haploinsufficiency is associated with reduced brain size and intestinal atresias in Feingold syndrome. *Nat. Genet.* 37, 465–467. <https://doi.org/10.1038/ng1546>
- Van Oss, S.B., Cucinotta, C.E., Arndt, K.M., 2017. Emerging Insights into the Roles of the Paf1 Complex in Gene Regulation. *Trends Biochem. Sci.* 42, 788–798. <https://doi.org/10.1016/j.tibs.2017.08.003>
- Varley, K.E., Gertz, J., Roberts, B.S., Davis, N.S., Bowling, K.M., Kirby, M.K., Nesmith, A.S., Oliver, P.G., Grizzle, W.E., Forero, A., Buchsbaum, D.J., LoBuglio, A.F., Myers, R.M., 2014. Recurrent read-through fusion transcripts in breast cancer. *Breast Cancer Res. Treat.* 146, 287–297. <https://doi.org/10.1007/s10549-014-3019-2>
- Varshney, D., Spiegel, J., Zyner, K., Tannahill, D., Balasubramanian, S., 2020. The regulation and functions of DNA and RNA G-quadruplexes. *Nat. Rev. Mol. Cell Biol.* 21, 459–474. <https://doi.org/10.1038/s41580-020-0236-x>
- Vijayakumar, A., Park, A., Steitz, J.A., 2022. Modulation of mRNA 3'-End Processing and Transcription Termination in Virus-Infected Cells. *Front. Immunol.* 13.
- Vilborg, A., Passarelli, M.C., Yario, T.A., Tycowski, K.T., Steitz, J.A., 2015. Widespread Inducible Transcription Downstream of Human Genes. *Mol. Cell* 59, 449–461. <https://doi.org/10.1016/j.molcel.2015.06.016>
- Vilborg, A., Sabath, N., Wiesel, Y., Nathans, J., Levy-Adam, F., Yario, T.A., Steitz, J.A., Shalgi, R., 2017. Comparative analysis reveals genomic features of stress-induced transcriptional readthrough. *Proc. Natl. Acad. Sci. U. S. A.* 114, E8362–E8371. <https://doi.org/10.1073/pnas.1711120114>
- Villarreal, O.D., Mersaoui, S.Y., Yu, Z., Masson, J.-Y., Richard, S., 2020. Genome-wide R-loop analysis defines unique roles for DDX5, XRN2, and PRMT5 in DNA/RNA hybrid resolution. *Life Sci. Alliance* 3, e202000762. <https://doi.org/10.26508/lsa.202000762>
- Wang, K., Ubriaco, G., Sutherland, L.C., 2007. RBM6-RBM5 transcription-induced chimeras are differentially expressed in tumours. *BMC Genomics* 8, 348. <https://doi.org/10.1186/1471-2164-8-348>

- Wang, R., Jiao, Z., Li, R., Yue, H., Chen, L., 2012. 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, 1116–1124. <https://doi.org/10.1093/neuonc/nos131>
- Wang, X., Hennig, T., Whisnant, A.W., Erhard, F., Prusty, B.K., Friedel, C.C., Forouzmand, E., Hu, W., Erber, L., Chen, Y., Sandri-Goldin, R.M., Dölken, L., Shi, Y., 2020. Herpes simplex virus blocks host transcription termination via the bimodal activities of ICP27. *Nat. Commun.* 11, 293. <https://doi.org/10.1038/s41467-019-14109-x>
- Wang, Z., Luo, Z., Zhou, L., Li, X., Jiang, T., Fu, E., 2015. DDX5 promotes proliferation and tumorigenesis of non-small-cell lung cancer cells by activating  $\beta$ -catenin signaling pathway. *Cancer Sci.* 106, 1303–1312. <https://doi.org/10.1111/cas.12755>
- Wei, L., Lee, S., Majumdar, S., Zhang, B., Sanfilippo, P., Joseph, B., Miura, P., Soller, M., Lai, E.C., 2020. Overlapping activities of ELAV/Hu family RNA binding proteins specify the extended neuronal 3' UTR landscape in *Drosophila*. *Mol. Cell* 80, 140-155.e6. <https://doi.org/10.1016/j.molcel.2020.09.007>
- Wei, Z., Wang, S., Xu, Y., Wang, W., Soares, F., Ahmed, M., Su, P., Wang, T., Orouji, E., Xu, X., Zeng, Y., Chen, Sujun, Liu, X., Jia, T., Liu, Z., Du, L., Wang, Y., Chen, Shaoyong, Wang, C., He, H.H., Guo, H., 2023. MYC reshapes CTCF-mediated chromatin architecture in prostate cancer. *Nat. Commun.* 14, 1787. <https://doi.org/10.1038/s41467-023-37544-3>
- Weirather, J.L., Cesare, M. de, Wang, Y., Piazza, P., Sebastiano, V., Wang, X.-J., Buck, D., Au, K.F., 2017. Comprehensive comparison of Pacific Biosciences and Oxford Nanopore Technologies and their applications to transcriptome analysis. <https://doi.org/10.12688/f1000research.10571.2>
- Weiss, W.A., Aldape, K., Mohapatra, G., Feuerstein, B.G., Bishop, J.M., 1997. Targeted expression of MYCN causes neuroblastoma in transgenic mice. *EMBO J.* 16, 2985–2995. <https://doi.org/10.1093/emboj/16.11.2985>
- Wells, J.P., White, J., Stirling, P.C., 2019. R Loops and Their Composite Cancer Connections. *Trends Cancer* 5, 619–631. <https://doi.org/10.1016/j.trecan.2019.08.006>
- Wenger, A.M., Peluso, P., Rowell, W.J., Chang, P.-C., Hall, R.J., Concepcion, G.T., Ebler, J., Functammasan, A., Kolesnikov, A., Olson, N.D., Töpfer, A., Alonge, M., Mahmoud, M., Qian, Y., Chin, C.-S., Phillippy, A.M., Schatz, M.C., Myers, G., DePristo, M.A., Ruan, J., Marschall, T., Sedlazeck, F.J., Zook, J.M., Li, H., Koren, S., Carroll, A., Rank, D.R., Hunkapiller, M.W., 2019. Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nat. Biotechnol.* 37, 1155–1162. <https://doi.org/10.1038/s41587-019-0217-9>
- West, S., Gromak, N., Proudfoot, N.J., 2004. Human 5'  $\rightarrow$  3' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. *Nature* 432, 522–525. <https://doi.org/10.1038/nature03035>
- Westermann, F., Muth, D., Benner, A., Bauer, T., Henrich, K.-O., Oberthuer, A., Brors, B., Beissbarth, T., Vandesompele, J., Pattyn, F., Hero, B., König, R., Fischer, M., Schwab, M., 2008. Distinct transcriptional MYCN/c-MYC activities are associated with spontaneous

regression or malignant progression in neuroblastomas. *Genome Biol.* 9, R150. <https://doi.org/10.1186/gb-2008-9-10-r150>

Wilkinson, M.E., Charenton, C., Nagai, K., 2020. RNA Splicing by the Spliceosome. *Annu. Rev. Biochem.* 89, 359–388. <https://doi.org/10.1146/annurev-biochem-091719-064225>

Willis, I.M., 1993. RNA polymerase III. Genes, factors and transcriptional specificity. *Eur. J. Biochem.* 212, 1–11. <https://doi.org/10.1111/j.1432-1033.1993.tb17626.x>

Wong, H.Y., Demmers, J.A.A., Bezstarosti, K., Grootegoed, J.A., Brinkmann, A.O., 2009. DNA dependent recruitment of DDX17 and other interacting proteins by the human androgen receptor. *Biochim. Biophys. Acta BBA - Proteins Proteomics* 1794, 193–198. <https://doi.org/10.1016/j.bbapap.2008.11.001>

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., Fuller-Pace, F.V., 2009. The DEAD box protein p72 regulates ER $\alpha$ -/Estrogen-dependent transcription and cell growth, and is associated with improved survival in ER $\alpha$  positive breast cancer. *Oncogene* 28, 4053–4064. <https://doi.org/10.1038/onc.2009.261>

Wu, G., Xing, Z., Tran, E.J., Yang, D., 2019. DDX5 helicase resolves G-quadruplex and is involved in MYC gene transcriptional activation. *Proc. Natl. Acad. Sci.* 116, 20453–20461. <https://doi.org/10.1073/pnas.1909047116>

Wu, J., You, Y.-Q., Ma, Y.-X., Kang, Y.-H., Wu, T., Wu, X.-J., Hu, X.-X., Meng, Q.-H., Huang, Y., Zhang, N., Pan, X.-B., 2020. DDX5-targeting fully human monoclonal autoantibody inhibits proliferation and promotes differentiation of acute promyelocytic leukemia cells by increasing ROS production. *Cell Death Dis.* 11, 1–14. <https://doi.org/10.1038/s41419-020-02759-5>

Wulfridge, P., Yan, Q., Rell, N., Doherty, J., Jacobson, S., Offley, S., Deliard, S., Feng, K., Phillips-Cremins, J.E., Gardini, A., Sarma, K., 2023. G-quadruplexes associated with R-loops promote CTCF binding. *Mol. Cell.* <https://doi.org/10.1016/j.molcel.2023.07.009>

Xing, Z., Ma, W.K., Tran, E.J., 2019. The DDX5/Dbp2 subfamily of DEAD-box RNA helicases. *Wiley Interdiscip. Rev. RNA* 10, e1519. <https://doi.org/10.1002/wrna.1519>

Xing, Z., Russon, M.P., Utturkar, S.M., Tran, E.J., 2020. The RNA helicase DDX5 supports mitochondrial function in small cell lung cancer. *J. Biol. Chem.* 295, 8988–8998. <https://doi.org/10.1074/jbc.RA120.012600>

Xu, K., Sun, S., Yan, M., Cui, J., Yang, Y., Li, W., Huang, X., Dou, L., Chen, B., Tang, W., Lan, M., Li, J., Shen, T., 2022. DDX5 and DDX17—multifaceted proteins in the regulation of tumorigenesis and tumor progression. *Front. Oncol.* 12, 943032. <https://doi.org/10.3389/fonc.2022.943032>

Xu, Z., Lodge, R., Power, C., Cohen, E.A., Hobman, T.C., 2020. The HIV-1 Accessory Protein Vpu Downregulates Peroxisome Biogenesis. *mBio* 11, 10.1128/mbio.03395-19. <https://doi.org/10.1128/mbio.03395-19>

- Xue, Y., Jia, X., Li, C., Zhang, K., Li, L., Wu, J., Yuan, J., Li, Q., 2019. DDX17 promotes hepatocellular carcinoma progression via inhibiting Klf4 transcriptional activity. *Cell Death Dis.* 10, 1–11. <https://doi.org/10.1038/s41419-019-2044-9>
- Xue, Y., Jia, X., Li, L., Dong, X., Ling, J., Yuan, J., Li, Q., 2018. DDX5 promotes hepatocellular carcinoma tumorigenesis via Akt signaling pathway. *Biochem. Biophys. Res. Commun.* 503, 2885–2891. <https://doi.org/10.1016/j.bbrc.2018.08.063>
- Yang, J., Altahan, A.M., Hu, D., Wang, Y., Cheng, P.-H., Morton, C.L., Qu, C., Nathwani, A.C., Shohet, J.M., Fotsis, T., Koster, J., Versteeg, R., Okada, H., Harris, A.L., Davidoff, A.M., 2015. The Role of Histone Demethylase KDM4B in Myc Signaling in Neuroblastoma. *JNCI J. Natl. Cancer Inst.* 107, djv080. <https://doi.org/10.1093/jnci/djv080>
- Yang, L., Lin, C., Liu, Z.-R., 2006. P68 RNA Helicase Mediates PDGF-Induced Epithelial Mesenchymal Transition by Displacing Axin from  $\beta$ -Catenin. *Cell* 127, 139–155. <https://doi.org/10.1016/j.cell.2006.08.036>
- Yang, L., Wei, M., Wang, Y., Zhang, J., Liu, S., Liu, M., Wang, S., Li, K., Dong, Z., Zhang, C., 2023. Rabphilin-3A undergoes phase separation to regulate GluN2A mobility and surface clustering. *Nat. Commun.* 14, 379. <https://doi.org/10.1038/s41467-023-36046-6>
- Yang, Q., Del Campo, M., Lambowitz, A.M., Jankowsky, E., 2007. DEAD-box proteins unwind duplexes by local strand separation. *Mol. Cell* 28, 253–263. <https://doi.org/10.1016/j.molcel.2007.08.016>
- Yang, X., Liu, Q.-L., Xu, W., Zhang, Y.-C., Yang, Y., Ju, L.-F., Chen, J., Chen, Y.-S., Li, K., Ren, J., Sun, Q., Yang, Y.-G., 2019. m6A promotes R-loop formation to facilitate transcription termination. *Cell Res.* 29, 1035–1038. <https://doi.org/10.1038/s41422-019-0235-7>
- Yanling Zhao, D., Gish, G., Braunschweig, U., Li, Y., Ni, Z., Schmitges, F.W., Zhong, G., Liu, K., Li, W., Moffat, J., Vedadi, M., Min, J., Pawson, T.J., Blencowe, B.J., Greenblatt, J.F., 2016. SMN and symmetric arginine dimethylation of RNA polymerase II C-terminal domain control termination. *Nature* 529, 48–53. <https://doi.org/10.1038/nature16469>
- Yasuda-Inoue, M., Kuroki, M., Ariumi, Y., 2013. Distinct DDX DEAD-box RNA helicases cooperate to modulate the HIV-1 Rev function. *Biochem. Biophys. Res. Commun.* 434, 803–808. <https://doi.org/10.1016/j.bbrc.2013.04.016>
- You, Z., Liu, C., Wang, C., Ling, Z., Wang, Yiduo, Wang, Yali, Zhang, M., Chen, S., Xu, B., Guan, H., Chen, M., 2019. LncRNA CCAT1 Promotes Prostate Cancer Cell Proliferation by Interacting with DDX5 and MIR-28-5P. *Mol. Cancer Ther.* 18, 2469–2479. <https://doi.org/10.1158/1535-7163.MCT-19-0095>
- Yu, X., Hu, L., Liu, X., Zhan, G., Mei, M., Wang, H., Zhang, X., Qiu, Z., Zhou, W., Yang, L., 2018. A Novel MYCN Variant Associated with Intellectual Disability Regulates Neuronal Development. *Neurosci. Bull.* 34, 854–858. <https://doi.org/10.1007/s12264-018-0236-4>



- Yu, Z., Mersaoui, S.Y., Guillon-Sert, L., Coulombe, Y., Song, J., Masson, J.-Y., Richard, S., 2020. DDX5 resolves R-loops at DNA double-strand breaks to promote DNA repair and avoid chromosomal deletions. *NAR Cancer* 2. <https://doi.org/10.1093/narcan/zcaa028>
- Yuan, H., Qin, F., Movassagh, M., Park, H., Golden, W., Xie, Z., Zhang, P., Sklar, J., Li, H., 2013. A Chimeric RNA Characteristic of Rhabdomyosarcoma in Normal Myogenesis Process. *Cancer Discov.* 3, 1394–1403. <https://doi.org/10.1158/2159-8290.CD-13-0186>
- Yuan, J., He, X., Wang, Y., 2023. G-quadruplex DNA contributes to RNA polymerase II-mediated 3D chromatin architecture. *Nucleic Acids Res.* gkad588. <https://doi.org/10.1093/nar/gkad588>
- Yuan, J., Song, Y., Pan, W., Li, Y., Xu, Y., Xie, M., Shen, Y., Zhang, N., Liu, J., Hua, H., Wang, B., An, C., Yang, M., 2020. LncRNA SLC26A4-AS1 suppresses the MRN complex-mediated DNA repair signaling and thyroid cancer metastasis by destabilizing DDX5. *Oncogene* 39, 6664–6676. <https://doi.org/10.1038/s41388-020-01460-3>
- Zeineldin, M., Patel, A.G., Dyer, M.A., 2022. Neuroblastoma: When differentiation goes awry. *Neuron* 110, 2916–2928. <https://doi.org/10.1016/j.neuron.2022.07.012>
- Zhang, H., Lam, J., Zhang, D., Lan, Y., Vermunt, M.W., Keller, C.A., Giardine, B., Hardison, R.C., Blobel, G.A., 2021. CTCF and transcription influence chromatin structure re-configuration after mitosis. *Nat. Commun.* 12, 5157. <https://doi.org/10.1038/s41467-021-25418-5>
- Zhang, H., Rigo, F., Martinson, H.G., 2015. Poly(A) Signal-Dependent Transcription Termination Occurs through a Conformational Change Mechanism that Does Not Require Cleavage at the Poly(A) Site. *Mol. Cell* 59, 437–448. <https://doi.org/10.1016/j.molcel.2015.06.008>
- Zhang, H., Shi, Z., Banigan, E.J., Kim, Y., Yu, H., Bai, X., Finkelstein, I.J., 2023. CTCF and R-loops are boundaries of cohesin-mediated DNA looping. *Mol. Cell.* <https://doi.org/10.1016/j.molcel.2023.07.006>
- Zhang, H., Xing, Z., Mani, S.K.K., Bancel, B., Durantel, D., Zoulim, F., Tran, E.J., Merle, P., Andrisani, O., 2016. RNA helicase DEAD box protein 5 regulates Polycomb repressive complex 2/Hox transcript antisense intergenic RNA function in hepatitis B virus infection and hepatocarcinogenesis. *Hepatology* 64, 1033–1048. <https://doi.org/10.1002/hep.28698>
- Zhang, H., Zhang, Y., Zhu, X., Chen, C., Zhang, C., Xia, Y., Zhao, Y., Andrisani, O., Kong, L., 2019. DEAD Box Protein 5 Inhibits Liver Tumorigenesis by Stimulating Autophagy via Interaction with p62/SQSTM1. *Hepatology* 69, 1046–1063. <https://doi.org/10.1002/hep.30300>
- Zhang, L., Li, L.X., Zhou, J.X., Harris, P.C., Calvet, J.P., Li, X., 2020. RNA helicase p68 inhibits the transcription and post-transcription of Pkd1 in ADPKD. *Theranostics* 10, 8281–8297. <https://doi.org/10.7150/thno.47315>
- Zhang, R., Shu, H., Wang, Y., Tao, T., Tu, J., Wang, C., Mergny, J.-L., Sun, X., 2023. G-Quadruplex Structures Are Key Modulators of Somatic Structural Variants in Cancers. *Cancer Res.* 83, 1234–1248. <https://doi.org/10.1158/0008-5472.CAN-22-3089>

- Zhang, Z., Tian, H., Miao, Y., Feng, X., Li, Y., Wang, H., Song, X., 2016. Upregulation of p72 enhances malignant migration and invasion of glioma cells by repressing Beclin1 expression. *Biochem. Mosc.* 81, 574–582.  
<https://doi.org/10.1134/S0006297916060031>
- Zhao, L., Zhao, Y., Liu, Q., Huang, J., Lu, Y., Ping, J., 2022. DDX5/METTL3-METTL14/YTHDF2 Axis Regulates Replication of Influenza A Virus. *Microbiol. Spectr.* 10, e01098-22. <https://doi.org/10.1128/spectrum.01098-22>
- Zhao, S., Ge, X., Wang, X., Liu, A., Guo, X., Zhou, L., Yu, K., Yang, H., 2015. The DEAD-box RNA helicase 5 positively regulates the replication of porcine reproductive and respiratory syndrome virus by interacting with viral Nsp9 in vitro. *Virus Res.* 195, 217–224. <https://doi.org/10.1016/j.virusres.2014.10.021>
- Zhao, X., Li, D., Yang, F., Lian, H., Wang, J., Wang, X., Fang, E., Song, H., Hu, A., Guo, Y., Liu, Y., Li, H., Chen, Y., Huang, K., Zheng, L., Tong, Q., 2020. Long Noncoding RNA NHEG1 Drives  $\beta$ -Catenin Transactivation and Neuroblastoma Progression through Interacting with DDX5. *Mol. Ther. J. Am. Soc. Gene Ther.* 28, 946–962.  
<https://doi.org/10.1016/j.ymthe.2019.12.013>
- Zhao, X.-F., Kohen, R., Parent, R., Duan, Y., Fisher, G.L., Korn, M.J., Ji, L., Wan, G., Jin, J., Püuschel, A.W., Dolan, D.F., Parent, J.M., Corfas, G., Murphy, G.G., Giger, R.J., 2018. PlexinA2 Forward Signaling through Rap1 GTPases Regulates Dentate Gyrus Development and Schizophrenia-like Behaviors. *Cell Rep.* 22, 456–470.  
<https://doi.org/10.1016/j.celrep.2017.12.044>
- Zhou, X., Luo, J., Mills, L., Wu, S., Pan, T., Geng, G., Zhang, J., Luo, H., Liu, C., Zhang, H., 2013. DDX5 Facilitates HIV-1 Replication as a Cellular Co-Factor of Rev. *PLOS ONE* 8, e65040. <https://doi.org/10.1371/journal.pone.0065040>
- Zhu, S., Lee, J.-S., Guo, F., Shin, J., Perez-Atayde, A.R., Kutok, J.L., Rodig, S.J., Neuberg, D.S., Helman, D., Feng, H., Stewart, R.A., Wang, W., George, R.E., Kanki, J.P., Look, A.T., 2012. Activated ALK Collaborates with MYCN in Neuroblastoma Pathogenesis. *Cancer Cell* 21, 362–373. <https://doi.org/10.1016/j.ccr.2012.02.010>
- Zhu, Y., Chen, G., Lv, F., Wang, X., Ji, X., Xu, Y., Sun, J., Wu, L., Zheng, Y.-T., Gao, G., 2011. Zinc-finger antiviral protein inhibits HIV-1 infection by selectively targeting multiply spliced viral mRNAs for degradation. *Proc. Natl. Acad. Sci.* 108, 15834–15839.  
<https://doi.org/10.1073/pnas.1101676108>
- Zhu, Y., Wang, X., Forouzmand, E., Jeong, J., Qiao, F., Sowd, G.A., Engelman, A.N., Xie, X., Hertel, K.J., Shi, Y., 2018. Molecular Mechanisms for CFIm-Mediated Regulation of mRNA Alternative Polyadenylation. *Mol. Cell* 69, 62-74.e4.  
<https://doi.org/10.1016/j.molcel.2017.11.031>
- Zimmerman, K.A., Yancopoulos, G.D., Collum, R.G., Smith, R.K., Kohl, N.E., Denis, K.A., Nau, M.M., Witte, O.N., Toran-Allerand, D., Gee, C.E., 1986. Differential expression of myc family genes during murine development. *Nature* 319, 780–783.  
<https://doi.org/10.1038/319780a0>

Zukher, I., Dujardin, G., Sousa-Luís, R., Proudfoot, N.J., 2023. Elongation roadblocks mediated by dCas9 across human genes modulate transcription and nascent RNA processing. *Nat. Struct. Mol. Biol.* 1–13. <https://doi.org/10.1038/s41594-023-01090-9>