

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

LES RÔLES DE FAM172A DANS LES MÉCANISMES D'ÉPISSAGE  
ALTERNATIF CO-TRANSCRIPTIONNEL DANS LE CONTEXTE DU  
SYNDROME CHARGE

THÈSE

PRÉSENTÉE

COMME EXIGENCE PARTIELLE

AU DOCTORAT EN BIOCHIMIE

PAR

SEPHORA SALLIS

OCTOBRE 2024

UNIVERSITÉ DU QUÉBEC À MONTRÉAL  
Service des bibliothèques

Avertissement

La diffusion de cette thèse se fait dans le respect des droits de son auteur, qui a signé le formulaire *Autorisation de reproduire et de diffuser un travail de recherche de cycles supérieurs* (SDU-522 – Rév.12-2023). Cette autorisation stipule que «conformément à l'article 11 du Règlement no 8 des études de cycles supérieurs, [l'auteur] concède à l'Université du Québec à Montréal une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de [son] travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, [l'auteur] autorise l'Université du Québec à Montréal à reproduire, diffuser, prêter, distribuer ou vendre des copies de [son] travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris l'Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de [la] part [de l'auteur] à [ses] droits moraux ni à [ses] droits de propriété intellectuelle. Sauf entente contraire, [l'auteur] conserve la liberté de diffuser et de commercialiser ou non ce travail dont [il] possède un exemplaire.»

## REMERCIEMENTS

Je remercie tout d'abord mon directeur de thèse, le Pr Nicolas Pilon, pour son accueil au sein de son laboratoire, pour son soutien et sa confiance au quotidien. Je tiens également à remercier tous mes collègues au cours de ces cinq dernières années au sein du laboratoire Pilon : Tatiana Cardinal, Elizabeth Leduc, Baptiste Charrier, Rodolphe Soret, Grégoire Bonnamour, Sandrine Girard, Félix-Antoine Bérubé Simard, Nejia Lassoued, Benoît Grondin, Sanaa Turk, Ouliana Souchkova, Shérine Ali, Mohammed Reza, Lise Rivollet, Marie Lefèvre, Catherine Bélanger, Fathia Azouz, Mostafa Esmael et Dhananjaya D. Certains d'entre eux sont devenu(e)s de très bon ami(e)s à qui je tiens beaucoup. Je remercie également le Centre d'Excellence en Recherche sur les Maladies Orphelines – Fondation Courtois (CERMO-FC) et ses différentes plateformes de bio-imagerie, de transcriptomique, de transgénése et de protéomique. Sans oublier l'aide que j'ai eu de Denis Flipo pour les expériences de BiFC en temp réel et de Geneviève Bourret pour ses conseils sur la préparation de librairie. J'aimerais remercier mon ancien directeur de recherche Thierry Forné pour toutes les connaissances qu'il m'a transmis et que j'utilise encore aujourd'hui. Je remercie énormément toute ma famille, et principalement ma mère Rachel Sallis et mon conjoint Joan Lacroix, pour leur soutien inconditionnel. Je remercie également énormément Akri, Lotfi et Nadia.

## TABLE DES MATIÈRES

REMERCIEMENTS .....	ii
LISTE DES FIGURES.....	vi
LISTE DES TABLEAUX.....	ix
LISTE DES ABRÉVIATIONS, DES SIGLES ET DES ACRONYMES .....	x
RÉSUMÉ .....	xiii
CHAPITRE I REVUE DE LA LITTÉRATURE.....	1
1.1 Le syndrome CHARGE : une neurocristopathie .....	1
1.1.1 Le syndrome CHARGE .....	1
1.1.2 Les cellules de la crête neurale .....	3
1.1.3 Les causes génétiques .....	11
1.1.3.1 Le gène <i>CHD7</i> .....	11
1.1.3.2 <i>CHD7</i> , un remodeleur de la chromatine .....	12
1.1.3.3 Patron d'expression de <i>CHD7</i> .....	17
1.1.3.4 Les autres causes génétiques .....	18
1.1.4 Modèles d'étude du Syndrome CHARGE .....	19
1.2 Le gène <i>FAM172A</i> , nouveau gène responsable du syndrome CHARGE .....	20
1.2.1 Le modèle de souris <i>Toupee</i> .....	20
1.2.2 La protéine FAM172A.....	22
1.2.3 Interaction avec Argonaute-2.....	27
1.2.3.1 Le domaine ARB2.....	27
1.2.3.2 Les protéines AGO .....	28
1.2.3.3 Les fonctions nucléaires d'AGO2 .....	29
1.2.3.4 Le complexe FAM172A-CHD7-AGO2 .....	32
1.3 Mécanismes moléculaires mis en cause dans le syndrome CHARGE.....	34
1.3.1 La transcription .....	34
1.3.1.1 Transcription par l'ARN polymérase II.....	34
1.3.1.2 Les régions promotrices.....	38
1.3.2 L'épissage des pré-ARN messagers.....	39
1.3.2.1 Le spliceosome et la réaction d'épissage.....	41
1.3.2.2 Les étapes d'assemblage du spliceosome.....	42
1.3.2.3 Épissage alternatif.....	45
1.3.3 Les mécanismes de régulation de l'épissage alternatif co-transcriptionnel..	46
1.3.3.1 Les séquences et protéines régulatrices .....	47
1.3.3.2 Les corps nucléaires.....	48

1.3.3.3	La vitesse d'élongation de l'ARN polymérase II .....	51
1.3.3.4	Les facteurs épigénétiques.....	53
1.4	Objectifs du projet de recherche.....	54
CHAPITRE II ARTICLE I: « CHARGE SYNDROME-ASSOCIATED PROTEINS FAM172A AND CHD7 ARE RECRUITED AT ACTIVE PROMOTER REGIONS AND PARTICIPATE TO SPLICING CONDENSATES ORGANIZATION IN NEURAL CREST CELLS » .....		
2.1	Résumé .....	58
2.2	Abstract.....	59
2.3	Introduction .....	60
2.4	Materials and Methods .....	63
2.5	Results .....	69
2.5.1	FAM172A is recruited at CHD7-occupied promoters in NCCs.....	69
2.5.2	FAM172A and CHD7 co-occupied promoters are enriched in H3K4me3...	72
2.5.3	Direct targets of FAM172A and CHD7 correlate better with misspliced genes in NCC from corresponding mutants.....	74
2.5.4	FAM172A and CHD7 participate to nuclear speckles organization .....	83
2.6	Discussion.....	89
CHAPITRE III ARTICLE II: «THE CHARGE SYNDROME-ASSOCIATED PROTEIN FAM172A CONTROLS AGO2 NUCLEAR IMPORT» .....		
3.1	Résumé .....	92
3.2	Abstract.....	93
3.3	Introduction .....	94
3.4	Materials and methods.....	97
3.5	Results .....	104
3.5.1	FAM172A influences AGO2 nuclear localization .....	104
3.5.2	FAM172A directly interacts with AGO2 .....	110
3.5.3	The classical bipartite NLS of FAM172A is required for AGO2 nuclear import.....	115
3.5.4	The status of FAM172A phosphorylation influences AGO2 nuclear import.....	121
3.5.5	AGO2 overexpression can functionally compensate for the loss of FAM172A <i>ex vivo</i> .....	127
3.6	Discussion.....	129
3.7	Acknowledgements .....	136

3.8 Tables .....	137
CHAPITRE IV DISCUSSION GÉNÉRALE .....	142
4.1 L'import nucléaire d'AGO2 .....	142
4.2 La protéine AGO2 dans le contexte du syndrome CHARGE .....	147
4.3 Le rôle régulateur de la phosphorylation de FAM172A.....	149
4.4 Distribution génomique de FAM172A et CHD7 dans les CCN .....	151
4.4.1 Spécificité des cellules de la crête neurale.....	152
4.4.2 Le complexe FAM172A-CHD7-AGO2 .....	153
4.5 Les cibles directes de FAM172A et CHD7 dans les cellules de la crête neurale.....	155
4.5.1 Régulation de l'épissage alternatif.....	155
4.5.2 FAM172A régule des gènes impliqués dans la prolifération, la différenciation et la migration des CCN .....	157
4.5.3 Bivalence des promoteurs .....	158
4.6 Le rôle potentiel de FAM172A et CHD7 dans l'état de pause de l'ARN polymérase II.....	160
4.7 Les autres rôles potentiels de FAM172A .....	165
4.7.1 Rôle dans la résolution des R-loops.....	165
4.7.2 FAM172A et les histones.....	168
CONCLUSION .....	169
ANNEXE A .....	172
APPENDICE A.....	183
BIBLIOGRAPHIE .....	184

## LISTE DES FIGURES

Figure	Page
1.1	Représentation des trois feuillets embryonnaires au cours du développement chez les mammifères ..... 4
1.2	Formation du tube neural et des cellules de la crête neurale au cours du développement des vertébrés ..... 6
1.3	Transition épithélio-mésenchymateuse des cellules de la crête neurale ..... 9
1.4	Diversité des cellules de la crête neurale ..... 10
1.5	Organisation de la chromatine..... 13
1.6	Schéma des domaines de la protéine CHD7 ..... 15
1.7	Schéma des domaines de la protéine FAM172A ..... 23
1.8	Les mécanismes du transport nucléaire classique ..... 26
1.9	Le rôle potentiel du complexe FAM172A-CHD7-AGO2 dans la régulation de l'épissage alternatif co-transcriptionnel ..... 31
1.10	Les différentes étapes de la transcription par l'ARN polymérase II ..... 35
1.11	Les étapes de l'activation de la transcription chez les mammifères..... 37
1.12	La réaction d'épissage ..... 43
1.13	Les étapes d'assemblage du splicéosome chez les mammifères..... 46
1.14	Schéma représentant des corps nucléaires au niveau des sites de transcription chez les mammifères ..... 49
2.1	FAM172A is recruited at CHD7-occupied promoters ..... 70
2.2	FAM172A is recruited at CHD7-occupied promoters (Supplemental)..... 71

2.3	FAM172A and CHD7 co-occupied promoters are enriched in trimethylated H3K4. ....	73
2.4	FAM172A and CHD7 co-occupied promoters are enriched in trimethylated H3K4 (Supplemental) .....	76
2.5	Direct targets of FAM172A and CHD7 correlate better with misspliced genes in NCC from corresponding mutants .....	78
2.6	Direct targets of FAM172A and CHD7 correlate better with misspliced genes in NCC from corresponding mutants (Supplemental).....	80
2.7	Gene ontology analysis of dysregulated genes in <i>Chd7<sup>Gt/+</sup></i> NCC (Supplemental) .....	82
2.8	FAM172A and CHD7 participate to nuclear speckles organization.....	85
2.9	FAM172A and CHD7 participate to nuclear speckles organization (Supplemental) .....	87
3.1	FAM172A influences AGO2 nuclear localization.....	106
3.2	FAM172A influences the nuclear localization of AGO2, but not AGO1 (Supplemental) .....	108
3.3	FAM172A directly interacts with AGO2.....	112
3.4	Subcellular distribution of <sub>FLAG</sub> AGO2 deletion mutants and E229Q-mutated <sub>MYC</sub> FAM172A (Supplemental).....	114
3.5	The NLS of FAM172A is required for AGO2 nuclear import .....	117
3.6	FAM172A-AGO2 interaction is direct and specific (Supplemental).....	119
3.7	Sequence of CK2-NLS loop and predicted 3D structure of FAM172A (Supplemental) .....	120
3.8	CK2 phosphorylates FAM172A.....	123
3.9	AGO2 nuclear import depends on the status of FAM172A phosphorylation .....	124
3.10	Phosphomimetic <sub>MYC</sub> FAM172A stimulates the nuclear entry of <sub>FLAG</sub> AGO2 (Supplemental) .....	126



3.11	AGO2 overexpression can functionally compensate for the loss of FAM172A <i>ex vivo</i> (Supplemental) .....	128
3.12	MYCFAM172A interacts with various endogenous importins (Supplemental) .....	131
3.13	FAM172A can homodimerize (Supplemental) .....	133
3.14	FAM172A also interacts with AGO1 and NCL (Supplemental) .....	135
4.1	L'import nucléaire d'AGO2 médié par FAM172A .....	146
4.2	Les protéines SR collaborent avec le complexe 7SK pour libérer l'ARN polymérase II de l'état de pause en aval du promoteur .....	161
4.3	La technique de CUT&RUN identifie différents types de fragments au niveau des promoteurs selon la taille. ....	164
4.4	Le potentiel mode d'action de FAM172A dans la régulation de l'expression et l'épissage alternatif des gènes. ....	167
A.1	<i>Ago2</i> is implicated in CHARGE Syndrome pathogenesis and genetically interacts with <i>Fam172a<sup>Trp</sup></i> .....	172
A.2	NCC-specific <i>Ago2</i> knockout phenocopies CHARGE syndrome in mice ...	174
A.3	AGO2 fulfills a non-canonical function relevant for CHARGE syndrome pathogenesis .....	176
A.4	AGO2 and FAM172A coregulate alternative splicing <i>in vivo</i> .....	177
A.5	R-loop-associated-foci are increased in <i>Fam172a<sup>Trp/Trp</sup></i> embryos .....	178

## LISTE DES TABLEAUX

Tableau	Page
1.1 Classification des signaux de localisation nucléaire .....	25
3.1 List of primers used.....	137
3.2 List of antibodies used.....	140
A.1 List of FAM172A interactors.....	179

## LISTE DES ABRÉVIATIONS, DES SIGLES ET DES ACRONYMES

ADN	Acide DésoxyriboNucléique
AGO1/2	Argonaute 1/2
Arb1/2	Argonaute binding protein 1/2
ARB2	Argonaute binding domain 2
ARN	Acide RiboNucléique
ARNm	ARN messenger
ARNmi	MicroARN
ARNsi	petit ARN interférent (en anglais siRNA; small interfering RNA)
ATP	Adénosine TriPhosphate
BAF/PBAF	Facteur Associé à BRG1 ou BRG1, Polybromo (en anglais : Polybromo, BRG1-associated factor)
BiFC	Bimolecular Fluorescence Complementation
BMP	Protéines Morphogénique des Os (en anglais : Bone Morphogenic Protein)
CCN	Cellule de la Crête Neurale (en anglais NCC ; Neural Crest Cell)
CHD	Chromodomain Helicase DNA-binding domain
CHD7	Chromodomain Helicase DNA-binding protein 7
ChIP	ImmunoPrécipitation de la Chromatine (en anglais : Chromatin Immunoprecipitation)

CK2	Protéine Kinase 2
CTD	Carboxyl-Terminal Domain
ESC	Cellule Souche Embryonnaire (en anglais : Embryonic Stem Cell)
ESE	Exonic Splicing Enhancer
ESS	Exonic Splicing Silencer
FGF	Facteur de Croissance de Fibroblastes (en anglais : Fibroblast Growth Factor)
HDAC	Déacétylase d’Histone (en anglais : Histone Deacetylase)
hnRNP	heterogeneous nuclear RiboNucleicProtein
IDR	Région Intrinsèquement Désordonnée (en anglais : Intrinsically Disordered Regions)
ISE	Intronic Splicing Enhancer
ISS	Intronic Splicing Silencer
NLS	Signal de Localisation Nucléaire (en anglais : Nuclear Localization Signal)
NPC	Complexe des Pores Nucléaire (en anglais : Nuclear Pore Complexes)
PAZ	Piwi-Argonaute-Zwille
PIC	Complexe de Pré-Initiation de la transcription (en anglais : Pre-initiation complex)
RBP	RNA-binding protein
RISC	RNA-induced silencing complexe
RNAPII	ARN Polymérase II
SEC	Super Complexe d’Elongation (en anglais : Super elongation complex)
SN	Speckle Nucléaire (en anglais NS ; Nuclear Speckles)

snRNP	small nuclear RiboNucleoProtein
SRSF	Serine arginine-rich Splicing Factor (ou protéines SR)
SRRM2	Serine/Arginine Repetitive Matrix protein 2
SWI/SNF	SWItch/Sucrose Non-Fermentable
TEM	Transition Épithélio-Mésenchymateuse
TN	Tube Neural
TSS	Site d'Initiation de la Transcription (en anglais : Transcription Start Site)
UTR	Untranslated transcribed region

## RÉSUMÉ

Le syndrome CHARGE est une maladie rare qui touche environ 1 naissance sur 10 000 et se caractérise par une combinaison variable de malformations qui sont à l'origine de son acronyme CHARGE. Cette maladie fait partie de la famille des neurocristopathies, soit des maladies causées par des anomalies des cellules de la crête neurale (CCN). La cause principale associée à ce syndrome est une mutation hétérozygote du gène *CHD7* (*Chromodomain Helicase DNA-binding protein 7*), codant pour un remodeleur de la chromatine qui intervient notamment dans la régulation de gènes essentiels à la formation et à la migration des CCN. Les autres causes menant à ce syndrome sont des mutations sur d'autres facteurs de la chromatine et/ou facteurs d'épissage. L'un de ces facteurs est la protéine FAM172A (*Family With Sequence Similarity 172 Member A*). Notre modèle suggère que la protéine FAM172A régulerait certains événements clés d'épissage alternatif co-transcriptionnel. Cette régulation aurait lieu à travers une interaction avec les protéines CHD7 et AGO2 (Argonaute 2) au sein d'un complexe à l'interface chromatine-splicéosome. Mais les mécanismes précis par lesquels elle régule ce processus sont encore méconnus.

Dans le but de mieux comprendre le rôle de ce complexe à l'interface chromatine-splicéosome, nous avons utilisé des méthodes génomiques et transcriptomiques pour déterminer la localisation génomique de FAM172A et CHD7, ainsi que leurs cibles directes au sein de CCN contrôles. Nous avons trouvé que 95% des régions occupées par FAM172A le sont aussi pour CHD7 et que ces régions corrélaient principalement avec des gènes dont le promoteur est actif (H3K4me3+) et ayant des défauts d'épissage alternatif dans des CCN mutées pour *Fam172a* ou *Chd7*. De plus, nous avons mis en évidence une désorganisation des corps nucléaires associés à l'épissage dans des tissus embryonnaires *in vivo*. Ces résultats suggèrent que FAM172A et CHD7 participent à la régulation de l'épissage alternatif à travers l'organisation et/ou le recrutement de facteurs d'épissage aux promoteurs. Nous discutons ensuite d'un potentiel modèle moléculaire dans lequel ce recrutement joue un rôle dans la régulation de l'état de pause de l'ARN polymérase II au promoteur et de la résolution de *R-loops*.

Dans un second temps, nous nous sommes intéressés à l'interaction entre les protéines FAM172A et AGO2, connues pour réguler la voie des microARNs dans le cytoplasme et l'épissage alternatif co-transcriptionnel dans le noyau. Nous avons mis en évidence que FAM172A peut lier AGO2 de manière directe, et par ce biais, participe activement à son import nucléaire grâce à son signal de localisation nucléaire et de la voie classique des importines  $\alpha/\beta$ . De plus, cette translocation nucléaire est stimulée lorsque FAM172A est phosphorylée par la Caséine Kinase 2, et diminuée par un variant de FAM172A mimant une mutation identifiée chez des patients CHARGE. Ces résultats

suggèrent que la fine régulation de l'import nucléaire d'AGO2 est un autre moyen pour FAM172A de réguler les mécanismes d'épissage alternatif co-transcriptionnel. Ces données mettent également en avant l'importance des rôles nucléaires non-canoniques d'AGO2. Nous discutons ensuite de l'implication d'une déplétion d'AGO2 dans un contexte du syndrome CHARGE.

Mots clés : Syndrome CHARGE; cellules de la crête neurale; épissage alternatif cotranscriptionnel; FAM172A; CHD7; AGO2.

## CHAPITRE I

### REVUE DE LA LITTÉRATURE

#### 1.1 Le syndrome CHARGE : une neurocristopathie

Le syndrome CHARGE est une maladie génétique rare qui touche environ une naissance sur 10 000 et dont la cause principale est une mutation du gène *CHD7* (*Chromodomain Helicase DNA-binding protein 7*). Cette maladie fait partie de la grande famille des neurocristopathies, soit des maladies ayant des problèmes avec les cellules de la crête neurale.

##### 1.1.1 Le syndrome CHARGE

Le syndrome CHARGE est une maladie génétique extrêmement complexe et rare, qui touche environ 1 naissance sur 10 000 dans le monde et jusqu'à 1 naissance sur 8 500 au Canada (Hsu et al., 2014; Issekutz et al., 2005). Une combinaison de malformations congénitales surreprésentées chez les patients a donné au syndrome son acronyme CHARGE. Il s'agit d'anomalies aux yeux (*Coloboma*), de problèmes cardiaques (*Heart problems*), d'une obstruction des voies nasales (*Atresia of the choanae*), d'un retard de croissance et de développement (*Retarded growth and development*), d'anomalies génitales (*Genital anomalies*) et d'anomalies aux oreilles (*Ear abnormalities*) (Pagon et al., 1981). Le colobome oculaire et les autres malformations de l'œil comme la microphthalmie touchent environ 80 % des patients atteints du syndrome CHARGE (Blake et al., 1998). Le colobome se caractérise par l'absence de fermeture ou l'absence de tissu d'une ou plusieurs structures de la partie ventrale de



l'œil. Cette anomalie peut toucher différents types de tissus tels que la rétine, l'iris ou la choroïde et peut concerner un seul œil (colobome unilatéral) ou les deux yeux (colobome bilatéral). Certains colobomes sont visibles et peuvent être diagnostiqués facilement, alors que d'autres nécessitent des examens approfondis de l'œil. Les malformations cardiaques concernent 75-85% des patients et peuvent toucher plusieurs parties du cœur (Jongmans et al., 2006). L'atrésie de la choane, dont l'incidence est d'environ 50- 60%, est une obstruction à l'arrière du passage nasal qui interfère avec la respiration (Blake & Prasad, 2006). A la naissance, cette malformation est diagnostiquée à la suite de problèmes respiratoires et peut dans certains cas nécessiter une intervention chirurgicale. D'autres obstructions peuvent aussi avoir lieu au niveau du larynx, de la trachée ou de l'œsophage (Hsu et al., 2014; Morgan et al., 1993). Le retard de croissance et de développement concerne environ 70-80% des cas, il peut inclure un retard mental et/ou des troubles du langage, de la communication et du comportement. Les anomalies génitales touchent environ 50-60% des patients et se présente sous la forme d'hypoplasie génitale ou d'hypogonadisme, ce qui peut provoquer un retard ou une absence de puberté à l'adolescence (Hsu et al., 2014). Les malformations de l'oreille sont également très fréquentes et peuvent concerner les structures externes, comme les structures internes de l'oreille. Ces dernières se traduisent par un niveau variable de surdité et/ou des troubles de l'équilibre (Hsu et al., 2014). D'autres malformations peuvent s'ajouter à cet acronyme, en particulier d'autres anomalies crâniocfaciales, comme la fente labio-palatine, des malformations digestives et rénales, des troubles endocriniens, un déficit immunitaire et un dysfonctionnement de certains nerfs crâniens impliqués dans les fonctions de succion, de déglutition et de la respiration (Hsu et al., 2014).

Les problèmes médicaux et physiques qui découlent de ce syndrome varient beaucoup d'un enfant à l'autre. En effet, les patients présentent rarement la même combinaison d'anomalies, ni la même pénétrance, ce qui rend le diagnostic du syndrome très difficile. Malgré une évolution au fil des années, les critères d'inclusion des patients atteints du

syndrome CHARGE sont divisés en deux catégories : les critères « majeurs » et les critères « mineurs » (Hale et al., 2016; van Ravenswaaij-Arts et al., 2015). Les premiers regroupent le colobome, l'atrésie des choanes, les anomalies des oreilles et des nerfs crâniens. Les seconds regroupent principalement les anomalies crâniofaciales, cardiaques, génitales, le retard mental et le retard de développement (Blake et al., 1998; Hsu et al., 2014). Plusieurs de ces anomalies sont potentiellement léthales, expliquant pourquoi environ 30% des enfants atteints meurent avant l'âge de 5 ans. Il n'y a actuellement aucune possibilité de guérison pour le syndrome CHARGE. Les options thérapeutiques disponibles sont en effet restreintes au traitement individuel des anomalies les plus sévères afin de favoriser la survie de l'enfant atteint. Les patients les moins sévèrement atteints peuvent vivre presque normalement.

Cette maladie est connue pour faire partie de la grande famille des neurocristopathies, soit des maladies ayant des problèmes avec les cellules de la crête neurale (Bérubé-Simard & Pilon, 2019).

### 1.1.2 Les cellules de la crête neurale

Les cellules de la crête neurale (CCN) sont une population de cellules multipotentes présentes pendant le développement embryonnaire et spécifiques aux vertébrés. La crête neurale a été décrite pour la première fois chez l'embryon de poulet par Wilhem His en 1868 (Huang & Saint-Jeannet, 2004). Ces cellules voient le jour très tôt pendant l'embryogenèse (à partir de quelques semaines chez l'homme, et quelques jours chez la souris). Elles sont générées depuis la région dorsale du tube neural en développement et vont migrer le long de l'embryon pour donner naissance à une grande variété de types cellulaires chez l'adulte. Ces cellules vont subir plusieurs processus clé qui mèneront à leur formation en CCN : l'induction, la spécification, la délamination, la migration et enfin la différenciation. Des défauts dans le développement des CCN peuvent entraîner les malformations congénitales associées aux neurocristopathies.

L'organisation moléculaire et génétique du développement des CCN a été très bien conservé au cours de l'évolution chez tous les vertébrés (Aybar & Mayor, 2002; York & McCauley, 2020). Tout commence chez l'embryon avec l'étape de la gastrulation, le processus au cours duquel les cellules de l'embryon s'organisent pour former trois feuillet embryonnaires qui donneront l'ensemble des tissus et organes. Les trois feuillet sont les suivants : l'ectoderme, le mésoderme et l'endoderme (Figure 1.1) (Technau & Scholz, 2003). Chacun de ces feuillet sera à l'origine de différents types cellulaires chez l'adulte. L'ectoderme est le feuillet supérieur, il donne naissance principalement au système nerveux, à l'épiderme et aux cellules pigmentaires (Holland, 2005). Le mésoderme est le feuillet intermédiaire, il génère les reins, les gonades, os, tissus du cœur et les cellules hématopoïétiques (Davidson & Zon, 2000). Enfin, l'endoderme est le feuillet inférieur, il est à l'origine des cellules épithéliales et divers organes appartenant principalement aux systèmes digestifs et respiratoires (Technau & Scholz, 2003).

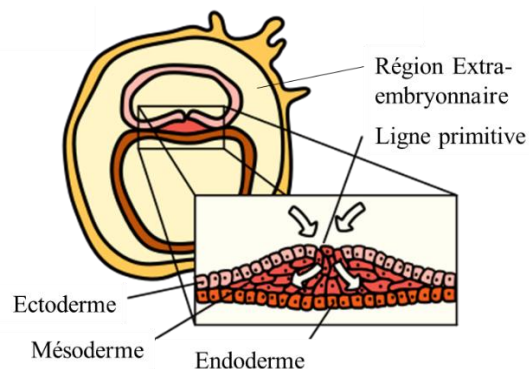


Figure 1.1 Représentation des trois feuillet embryonnaires au cours du développement chez les mammifères : l'ectoderme, le mésoderme et l'endoderme, qui donneront naissance à tous les types cellulaires et à tous les tissus de l'organisme ("Ligne primitive," 2023)

Dans le but de former le tube neural et les crêtes neurales, l'ectoderme s'épaissit au niveau de sa région dorsale pour former la plaque neurale (ou ectoderme neurale), pendant que l'ectoderme adjacent à cette plaque devient l'ectoderme non-neurale

(Figure 1.2). Cette étape du développement embryonnaire est contrôlée par une signalisation moléculaire complexe et plus précisément par différents signaux provenant du mésoderme. Il s'agit de l'étape d'induction (Knecht & Bronner-Fraser, 2002; Pla & Monsoro-Burq, 2018). Chez les espèces *Xenopus* (grenouille) et *Zebrafish* (poissons), l'ectoderme produit les facteurs de croissance BMP (en anglais *Bone Morphogenic Protein*) qui favorisent la différenciation cellulaire de l'ectoderme vers l'épiderme, en défaveur du tissu nerveux. Une partie du mésoderme, appelée « centre organisateur », sécrète des antagonistes qui vont se lier aux protéines BMP et vont empêcher leur signalisation. Il s'agit de la *noggin*, la *chordine* et la *follistatine* (Harland, 2000; Wilson & Hemmati-Brivanlou, 1997). Cette sécrétion d'inhibiteurs se fait de manière à obtenir un gradient de BMP : un haut niveau de signalisation BMP promeut la différenciation en épiderme et inhibe la différenciation en tissu neural, alors qu'un faible niveau de signalisation BMP induit la différenciation en tissu neural. De façon intéressante, le niveau intermédiaire de signalisation BMP de part et d'autre de la plaque neurale va permettre la spécification des bordures neurales. Les bordures neurales deviendront ultimement les crêtes neurales (Pla & Monsoro-Burq, 2018; Plouhinec et al., 2014; Wu et al., 2011). Cette signalisation est plus complexe encore. En effet, les voies de signalisation FGF (en anglais *Fibroblast Growth Factor*) et WNT (*Wingless-related*) sont nécessaires pour l'induction de la plaque neurale et pour les bordures de cette plaque. La voie de signalisation de l'acide rétinoïque peut également jouer un rôle dans la spécification des crêtes neurales (Pla & Monsoro-Burq, 2018). Malgré la grande similarité dans le développement des CCN chez les vertébrés, ces mécanismes de régulation peuvent être différents d'une espèce à l'autre (Huang & Saint-Jeannet, 2004). Par exemple, comme décrit plus haut, les inhibiteurs *noggin* et *chordine* sont essentiels dans l'induction de la plaque neurale chez la grenouille et le poisson. En revanche, des études menées sur des embryons de souris mutés pour le gène *Chordin* ou *noggin* ont démontré un développement normal de la plaque neurale mais des anomalies plus tardives au niveau du cerveau antérieur.

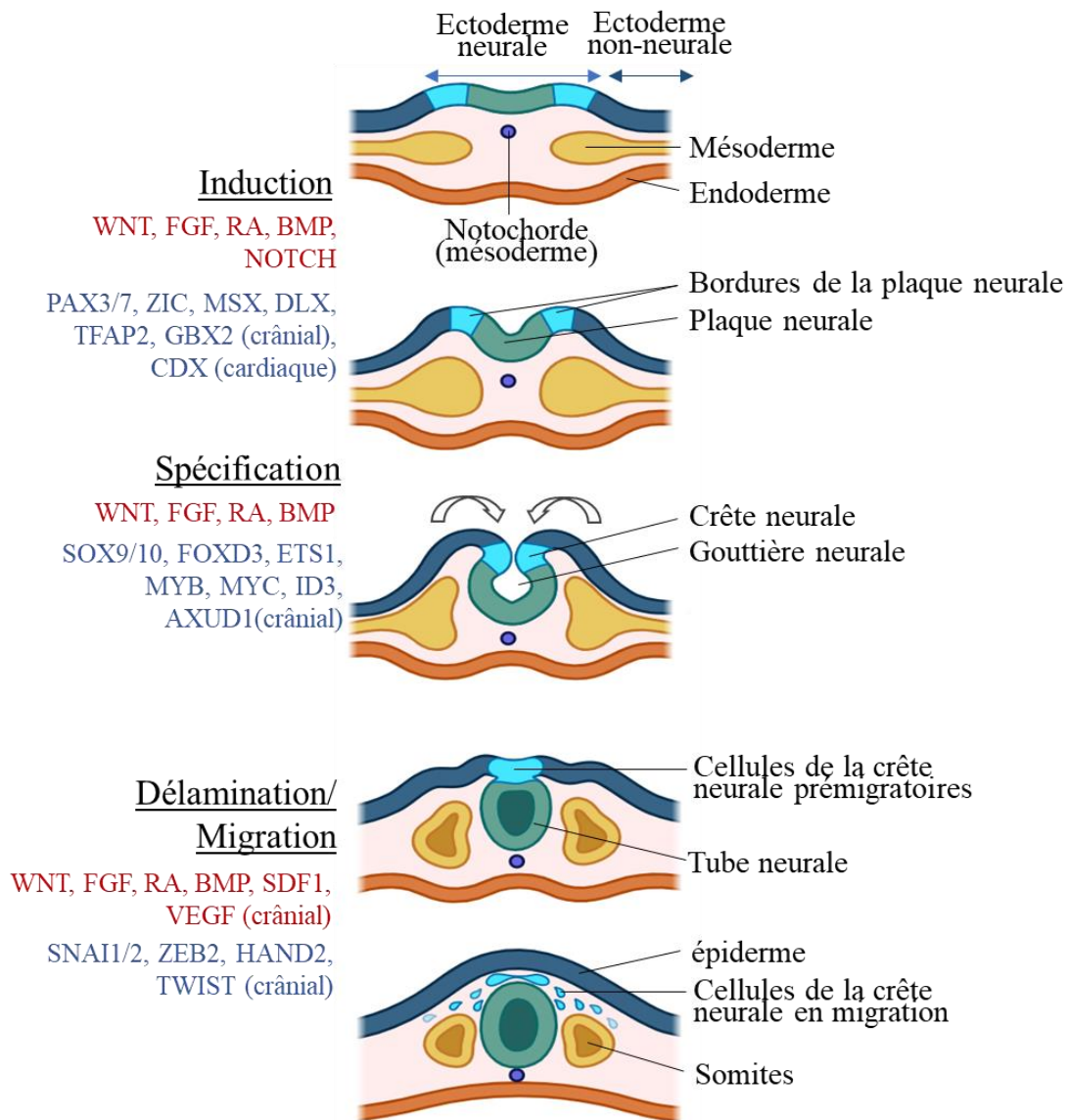


Figure 1.2 Formation du tube neurale et des cellules de la crête neurale au cours du développement des vertébrés. Schéma représentant la formation du tube neurale et des cellules de la crête neurale. Le tube neural se forme à la suite de l'induction et de la spécification de l'ectoderme neurale (ou plaque neurale). Les bordures de la plaque neurale se replient sur eux-mêmes pour former la crête neurale. À la suite de la délamination (transition épithéliale mésenchymateuse), les cellules de la crête neurale vont se détacher et migrer dans tout l'embryon. Ces étapes sont contrôlées par un réseau de gènes régulateurs. Les facteurs de transcription sont représentés en bleu et les voies de signalisation sont représentés en rouge. Figure créée à l'aide de *Biorender* et inspirée de *Pilon, 2021*.

Ces anomalies concernent également l'oreille interne/externe, le nombre de somites, l'organisation du pharynx et du système cardio-vasculaire (Bachiller et al., 2000; Huang & Saint-Jeannet, 2004). De plus, les voies de signalisation WNT et FGF restent essentielles à l'induction de la plaque neurale et de la crête neurale chez toutes les espèces, y compris la souris (Brault et al., 2001). Ces résultats mettent en évidence un autre mécanisme d'action des voies WNT/FGF dans l'induction, qui est indépendant des facteurs de croissance BMP (Huang & Saint-Jeannet, 2004; Pla & Monsoro-Burq, 2018).

Au fur et à mesure de l'induction, les bordures de la plaque neurale s'élèvent et provoquent l'invagination de l'ectoderme au niveau de la plaque neurale, ce qui va provoquer l'apparition d'une gouttière neurale (Figure 1.2). Les bords de cette gouttière finiront par se rejoindre pour former le tube neural (TN) de l'embryon. Au même moment, l'extrémité de ces bords qu'on appelle la crête neurale (CN) se détachent progressivement pour former les cellules de la crête neurale en migration (Figure 1.2). Chez la souris, ces cellules commencent à se détacher avant même la fermeture complète du tube neural contrairement à d'autres espèces où la fermeture complète du tube neural est nécessaire (Huang & Saint-Jeannet, 2004; Kulesa et al., 2004). Pendant cette étape qu'on appelle la délamination, les cellules prémigratoires vont subir une transition épithélio-mésenchymateuse (TEM) (Figure 1.3). Plus précisément, les cellules reçoivent des signaux qui induisent une diminution de l'expression des cadhérines (molécules importantes pour les jonctions adhérentes entre cellules) et se détachent de leur tissu épithélial (Simões-Costa & Bronner, 2015). Les E-cadhérines (épithéliales) sont remplacées par des N-cadhérines (neurale) (Scarpa et al., 2015) alors que les jonctions serrées sont remplacées par des jonctions Gap (Scarpa et al., 2015). Par la suite, la voie de signalisation WNT active l'expression de protéines responsables de la réorganisation du cytosquelette d'actine nécessaire à la future migration des CCN. Cependant, des expériences de microscopie en temps réel chez le poulet ont montré que les cellules prémigratoires acquièrent une plus grande motilité avant même la

diminution des adhésions cellule-cellule (Ahlstrom & Erickson, 2009; Szabó & Mayor, 2018). En effet, la délamination reste une étape fortement régulée par différents signaux moléculaires. Les cellules acquièrent leurs propriétés migratoires grâce à un ensemble de facteurs de transcription, incluant notamment les gènes : *Snail/Slug* (*Snail family transcriptional repressor*), *Foxd3* (*Forkhead Box D3*), *Sox8*, *Sox9* et *Sox10* (*Sex determining region Y-Box 8/9/10*) (Szabó & Mayor, 2018). Les cellules détachées vont migrer tout le long de l'embryon pour donner naissance à diverses lignées cellulaires.

Toutes les cellules de la crête neurale en migration vont se déplacer entre l'épiderme et le mésoderme le long de l'embryon. Les cellules migrent à partir de la région dorsale du tube neural (où les CCN se détachent) vers la région ventrale de l'embryon (Figure 1.4). Cependant, la migration peut différer selon l'espèce, selon la temporalité de leur migration et selon la position le long de la crête neurale. En effet, les CCN se détachent tout le long du tube neural, de la partie antérieure de l'embryon jusqu'à la partie postérieure. De ce fait, plusieurs catégories de CCN ont été décrites selon leur localisation d'origine : crâniale, vagale, troncale et sacrale (Figure 1.4). Chaque catégorie va engendrer des dérivées cellulaires distincts (Pilon, 2021) (Figure 1.4). Les cellules crânielles (ou céphaliques) donneront naissance aux cartilages (chondroblastes), os, muscles, tendons et tissus conjonctifs et nerveux de la région faciale et du crâne. Tandis que les CCN vagales, incluant aussi les cardiaques, génèreront le tissu nerveux entérique et certains tissus du cœur. Les CCN troncales et sacrales donnent naissance, entre autres, aux neurones et cellules gliales des ganglions sympathiques et de la racine dorsale. Et enfin, les mélanocytes et les cellules de Schwann sont originaires des CCN provenant de toutes les régions (Pilon, 2021; Szabó & Mayor, 2018). Pendant leur migration, les cellules vont entamer progressivement un changement d'identité selon leur environnement. Tout d'abord les cellules migrent ensemble d'un flux commun, puis plusieurs flux migratoires vont avoir lieu dépendamment de différents marqueurs tel que les facteurs HOX (*Homeobox*), ainsi que les facteurs EPHs (*Ephrins*) et leurs récepteurs (Szabó & Mayor, 2018).

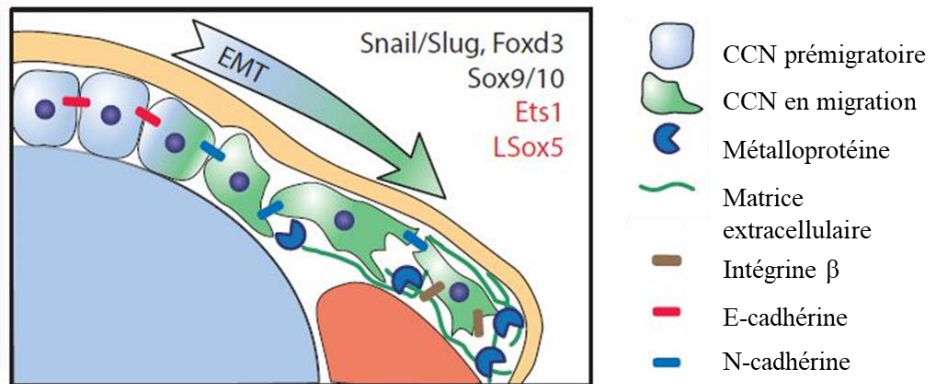


Figure 1.3 Transition épithélio-mésenchymateuse des cellules de la crête neurale. Schéma représentant la transition épithélio-mésenchymateuse des cellules de la crête neurale. Les facteurs de transcriptions en rouge sont spécifiques des CCN crâniens. CCN= Cellules de la crête neurale, EMT= transition épithélio-mésenchymateuse (TEM). Schéma adapté de *Szabó et Mayor, 2018*.

Des problèmes survenant lors de ces étapes de formation des CCN (induction, spécification, migration et différenciation) peuvent engendrer des anomalies dans le développement de l'embryon, représentées par une grande variété de maladies génétiques chez l'homme nommées les neurocristopathies. Le syndrome CHARGE fait partie de la famille des neurocristopathies, touchant particulièrement les CCN crâniens et cardiaques. Ainsi, l'étude approfondie de ces cellules est essentielle pour mieux comprendre les mécanismes de développement embryonnaire et les pathologies associées. D'autres maladies rares, tels que le syndrome de Kallmann ou le syndrome de Kabuki, appartiennent également aux neurocristopathies et partagent de nombreuses anomalies avec le syndrome CHARGE. Cependant, ils ont des causes génétiques différentes (Bérubé-Simard & Pilon, 2019; Cheon & Ko, 2015; Dodé et al., 2003; Hardelin, 2001)



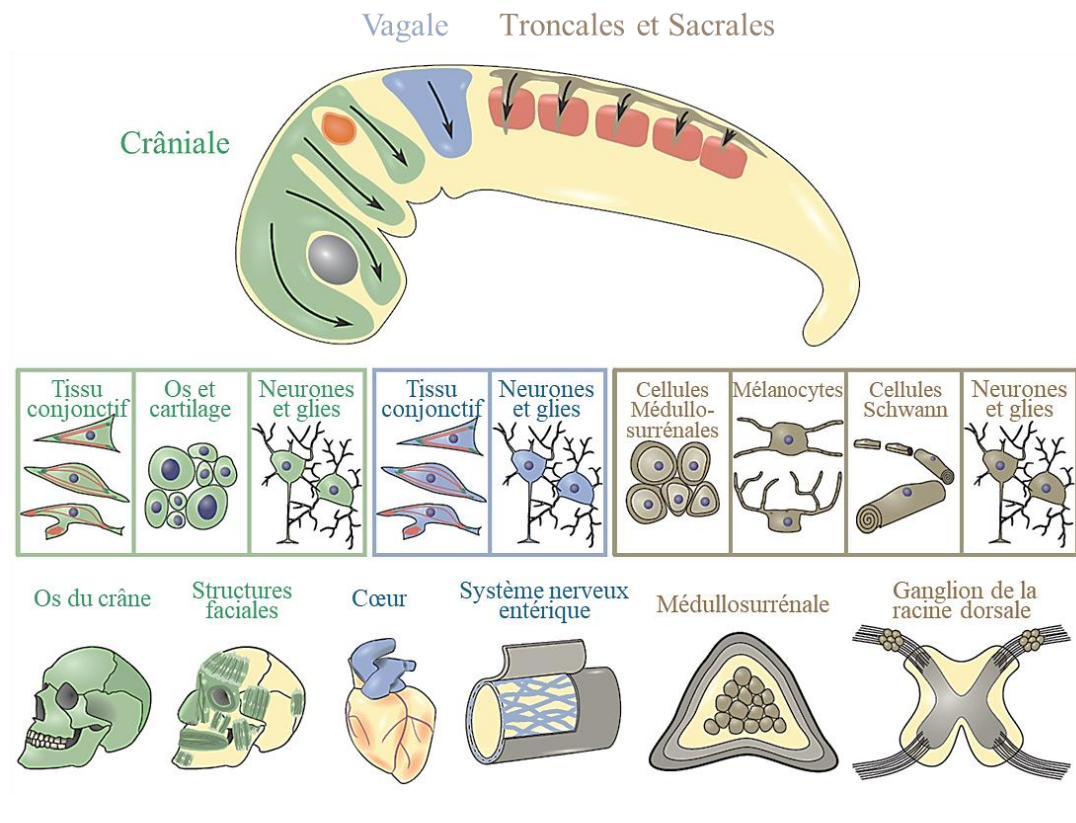


Figure 1.4 Diversité des cellules de la crête neurale. Les cellules de la crête neurale se détachent du tube neural en formation et migrent le long de l'embryon. Il existe plusieurs types de cellules de la crête neurale selon leur localisation d'origine : les crânielles, les vagales et les troncales (incluant les sacrales). En exemples, quelques cellules progénitrices et structures dérivées des différents types de cellules de la crête neurale. Schéma adapté de *Szabó et Mayor, 2018*.

### 1.1.3 Les causes génétiques

#### 1.1.3.1 Le gène *CHD7*

La cause principale associée au syndrome CHARGE est une mutation hétérozygote du gène *CHD7* (*Chromodomain Helicase DNA-binding protein 7*) (~70% des cas) (Vissers et al., 2004a). Chez l'homme, le gène *CHD7* est situé sur le chromosome 8q12.2 et contient 42 exons. Il code pour une enzyme capable de moduler la compaction de la chromatine sous la dépendance de l'ATP (Adénosine triphosphate). L'isoforme principale de *CHD7* situé dans le noyau contient 2 997 acides aminés (335 kDa) traduits à partir de 38 exons. Les mutations qui causent ce syndrome surviennent généralement *de novo*, c'est-à-dire que les mutations se produisent de manière spontanée et ne sont pas héritées des parents. Dans 3% des cas, les patients atteints du syndrome CHARGE transmettent leur mutation autosomale dominante à leur descendance (Vissers et al., 2004a). Ce faible taux de transmission peut s'expliquer par les problèmes de fertilité et de puberté chez les patients. Plusieurs types de mutations ont été répertoriés pour ce gène, les principaux étant les types de mutations « faux-sens », « non-sens » et les insertions/délétions (Janssen et al., 2012). Les premières sont des mutations ponctuelles qui induisent le changement d'un acide aminé. Lorsque le nouvel acide aminé s'avère être un codon stop, il s'agit d'une mutation « non-sens » qui provoque un arrêt prématuré de la traduction. Les insertions et délétions, même d'un seul nucléotide, créent un décalage du cadre de lecture lors de la traduction jusqu'à l'insertion d'un codon stop. Ces mutations engendrent une protéine tronquée, incomplète, potentiellement instable et souvent non-fonctionnelle. Une mutation du gène *CHD7* a aussi été retrouvée chez certains patients atteints du syndrome de Kallmann ; ainsi que dans des cas d'autisme (O'Roak et al., 2012), un phénotype également associé à des patients atteints du syndrome CHARGE (Marcos et al., 2014).

### 1.1.3.2 CHD7, un remodeleur de la chromatine

Dans le noyau cellulaire, le génome est compacté sous forme de chromatine. Un octamère d'histones autour duquel s'enroule, sur un tour et trois quart, un fragment d'ADN (Acide désoxyribonucléique) de 146 paires de bases forme un nucléosome, l'unité de base de la chromatine (Figure 1.5A). Les nucléosomes sont formés de deux hétérodimères d'histone 2A (H2A) - histone 2B (H2B) et de deux hétérodimères d'histone 3 (H3) -histone 4 (H4). Les histones, des protéines chargées positivement, se lient très facilement à l'ADN qui est chargé négativement. Lorsqu'une région du génome est enroulée autour de nucléosomes de manière très compacte, elle est inaccessible aux différents facteurs nucléaires, on parle de chromatine fermée ou « d'hétérochromatine ». Ces régions connaissent alors un très faible niveau de transcription, contrairement à la chromatine ouverte ou « euchromatine » qui sont les régions du génome entre deux nucléosomes, très peu compactées et accessibles à ces mêmes facteurs (Figure 1.5B). Ainsi, la disposition des nucléosomes sur le génome influence l'expression des gènes, celle-ci doit être finement régulée tout au long du développement et de la différenciation cellulaire. Chaque histone possède également une « queue » d'environ 20 acides aminés à son extrémité N-terminale, qui, lorsqu'assemblé en nucléosome, dépasse du complexe octamérique. Ces queues d'histones contiennent environ un tiers des charges positives des histones grâce à la présence de nombreux acides aminés basiques (lysine et arginine) (Bertin & Mangenot, 2008). Elles sont également sujettes à de nombreuses modifications post-traductionnelles médiées par diverses enzymes de modification. Les modifications post-traductionnelles les plus étudiées sont l'acétylation de lysines, la méthylation de lysines et d'arginines, la phosphorylation de sérines et thréonines et l'ubiquitinylation de lysines (Bertin & Mangenot, 2008). Par le biais de ces modifications, des facteurs de la chromatine, nommés les écrivains (en anglais *writers* ; méthyltransférases et acétyltransférases d'histone), les effaceurs (en anglais *erasers*; déméthylases et désacétylases d'histone) et les lecteurs (en anglais *readers*) sont capables d'influencer

la compaction de la chromatine. Les remodeleurs de la chromatine sont capables de moduler activement cette compaction grâce à une activité dépendante de l'ATP.

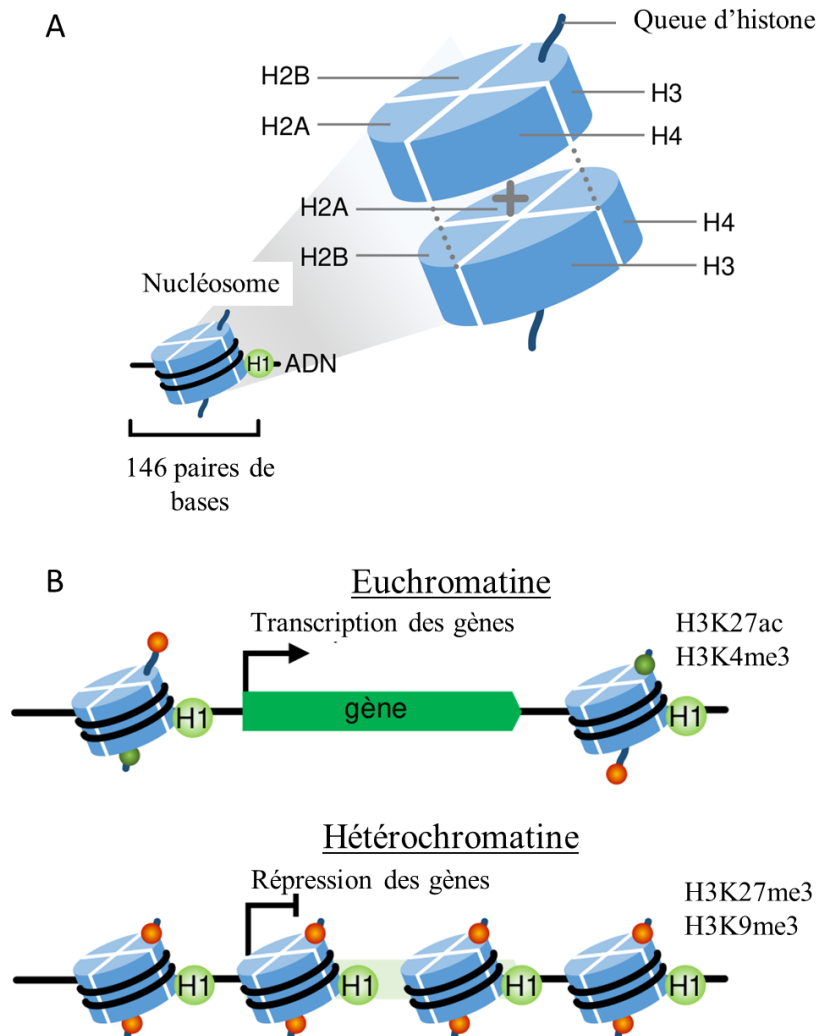


Figure 1.5 Organisation de la chromatine. (A) Schéma représentant un nucléosome : un octamère d'histones autour duquel est enroulé la double hélice d'ADN. (B) Représentation d'une chromatine ouverte (ou Euchromatine) transcriptionnellement active et enrichies en H3K4me3 et H3K27ac. (C) Représentation d'une chromatine fermée (ou Hétérochromatine) transcriptionnellement inactive et enrichies en H3K9me3 et H3K27me3. ("Contrôle de la transcription," Pierre Baduel, 2021)

Les remodeleurs de la chromatine sont répartis en 3 familles principales basées sur leurs similarités de structures et de fonctions : SWI/SNF (*SWItch/Sucrose Non-*

*Fermentable*), ISWI (*imitation-switch*) et CHD (*chromodomain helicase DNA-binding*) (Muhammad et al., 2023). Les remodeleurs SWI/SNF sont capables d'éliminer ou de déplacer les nucléosomes le long de la molécule d'ADN, alors que les membres de la famille ISWI assemblent les nucléosomes ou les espacent les uns des autres. Enfin, les membres de la famille CHD peuvent assembler les nucléosomes, éliminer des histones ou les déplacer le long de la molécule d'ADN (Basson & van Ravenswaaij-Arts, 2015; Muhammad et al., 2023). Les membres de la famille CHD, au nombre de 9 (CHD1 à CHD9), sont divisés en 3 sous-groupes selon leur composition en domaines protéiques (Woodage et al., 1997). CHD7 appartient à la sous-famille III contenant deux chromodomaines (en N-terminale), un grand domaine SNF2-hélicase dont l'activité est dépendante de l'ATP, un domaine SANT (*Swi3, Ada2, NCoR, and TFIIB*)-SLIDE-like, ainsi que des domaines BRK (*Brahma and Kismet*) (en C-terminal) (Figure 1.6) (Delmas et al., 1993; Micucci et al., 2015). La fonction principale de CHD7 est de moduler la compaction de la chromatine, et plus précisément de faire « glisser » les nucléosomes (Bouazoune & Kingston, 2012). Cette action permet d'ouvrir la chromatine afin de rendre les gènes et les séquences régulatrices associées accessibles aux facteurs de transcription. Ou au contraire, elle permet de fermer la chromatine afin de rendre certains loci inaccessibles à ces mêmes facteurs. Les chromodomaines permettent à CHD7 d'être recruté au niveau de la chromatine, par exemple au niveau de certaines modifications post-traductionnelles de l'histone 3, telles que la méthylation de la lysine 4 (H3K4me) et l'acétylation de la lysine 27 (H3K27ac) (Schnetz et al., 2009a), de l'ADN nucléosomale ou de l'ARN (Acide ribonucléique) (Akhtar et al., 2000). Ensuite, l'activité hélicase de son domaine SNF2 à travers l'hydrolyse de l'ATP permet le « glissement » des nucléosomes (Delmas et al., 1993). Des études ont par ailleurs montré que le domaine SANT est capable de lier l'ADN nu entre les nucléosomes, et qu'une liaison à la fois à celui-ci et aux nucléosomes est kalli. Malgré que CHD7 ne soit pas recruté à ces régions, le domaine BRK est connu pour interagir avec les facteurs CTCF (*CCCTC-binding factor*) chez d'autres CHD (Ishihara et al., 2006). Ces changements de conformation de chromatine sont aussi importants

pour la régulation de la transcription que pour la réplication et les mécanismes de réparations de l'ADN, ce qui rend CHD7 très important au cours du développement (Alendar & Berns, 2021).

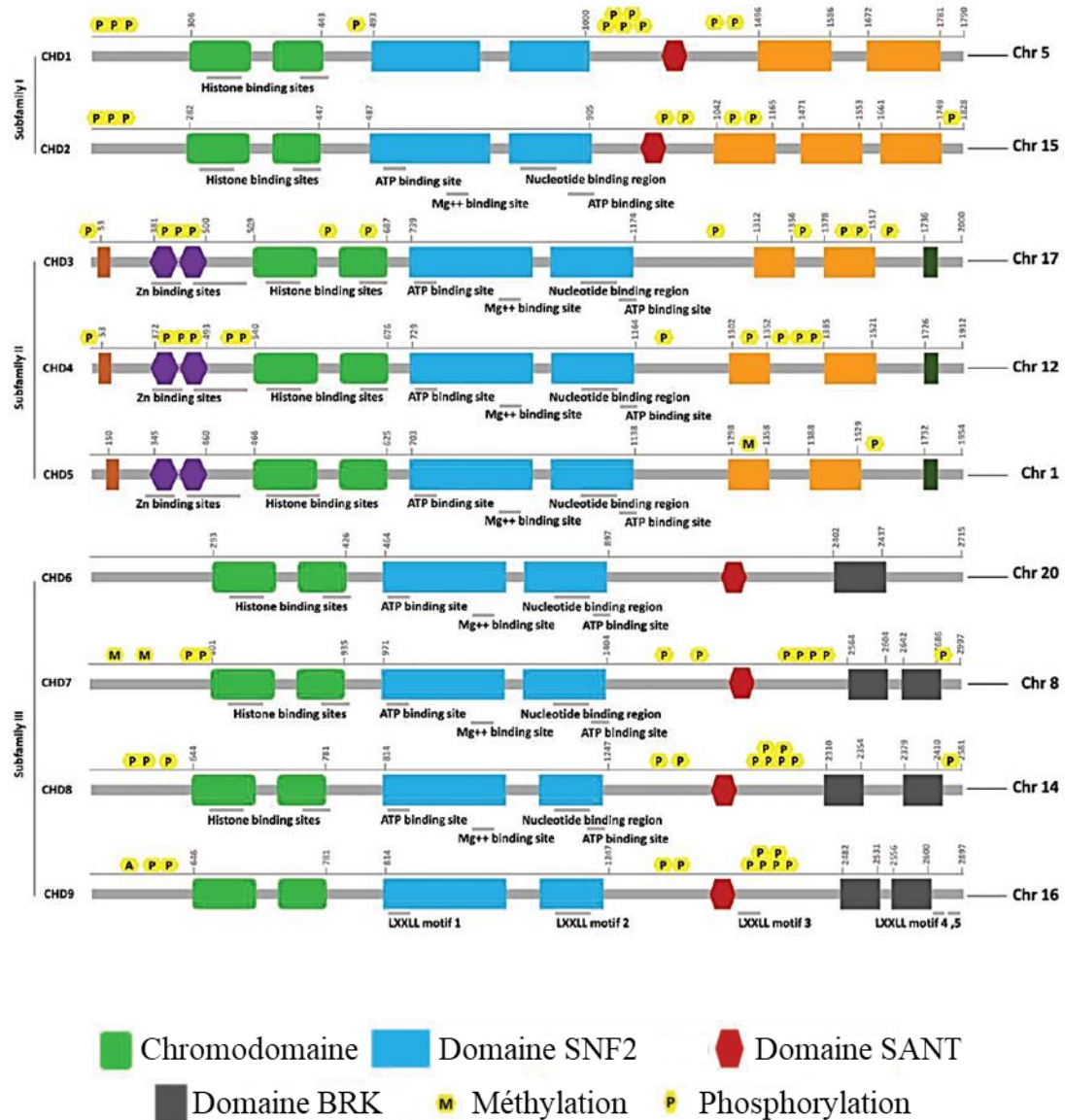


Figure 1.6 Schéma représentant les différents domaines des protéines CHD chez l'homme et ses sites de phosphorylation et méthylation. Schéma adapté de *Muhammad et al, 2023*.

Certaines régions du génome, dites séquences régulatrices, sont nécessaires pour réguler l'expression des gènes au cours du développement et de la différenciation cellulaire. Il s'agit, entre autres, de séquences promotrices situées juste en amont du site d'initiation de la transcription +1 (TSS; en anglais *Transcription Start Site*) et de séquences amplificatrices (ou *enhancers*) situées à distance du TSS. Ces séquences recrutent des facteurs de transcription et permettent d'initier (promoteurs) ou de stimuler (amplificateurs) la transcription des gènes associés. A un temps donné, certaines de ces séquences seront activées, alors que d'autres seront réprimées. Des expériences d'immunoprécipitation de la chromatine suivi d'un séquençage à haut débit (ChIP-seq; en anglais *Chromatin Immunoprecipitation followed by sequencing*) ont montré que CHD7 est préférentiellement recruté aux régions amplificatrices enrichies en monométhylation de la lysine 4 (H3K4me1) et en acétylation de la lysine 27 de l'histone 3 (H3K27ac), de même qu'aux régions promotrices enrichies en triméthylation de la lysine 4 (H3K4me3) (Feng et al., 2017; Schnetz et al., 2009a). La proportion entre ces deux types de régions varie selon le type cellulaire étudié. De plus, CHD7 peut s'associer au complexe BAF/PBAF (*Polybromo, BRG1-associated factor*), appartenant à la famille des remodeleur de la chromatine SWI/SNF dans les CCN pour médier cette régulation (Bajpai et al., 2010). Il est également connu pour interagir avec d'autres facteurs tel que CHD8, ou encore SMAD1 (*Mothers against decapentaplegic member 1*) de la voie de signalisation BMP dans le tissu cardiaque (Batsukh et al., 2010; Y. Liu et al., 2014). Au niveau des régions amplificatrices, son recrutement corrèle avec celui du facteur p300, ainsi que de plusieurs facteurs de pluripotence dans les cellules souches embryonnaires (ESCs) (Schnetz et al., 2010). Dans des cellules souches neurales, CHD7 s'avère également être un cofacteur transcriptionnel important de SOX2 (*Sex determining region Y -Box 2*) pour la régulation des gènes (Engelen et al., 2011). Les remodeleurs de la chromatine peuvent également influencer directement la transcription et la maturation des ARN (Gompers et al., 2017; Murawska & Brehm, 2011). Par exemple, CHD7 est impliqué dans la régulation de l'épissage alternatif de certains gènes durant le développement (Allemand et al., 2016; Bélanger et al., 2018,

2022). Il peut notamment interagir physiquement avec l'ARN messager naissant et le spliceosome, et plus précisément le complexe snRNP U2 (*small nuclear RiboNucleoProtein U2*) (Allemand et al., 2016; G Hendrickson et al., 2016). Ainsi, le remodelleur de la chromatine CHD7 joue un rôle important dans l'ouverture de la chromatine au niveau de séquences régulatrices, de même que dans la régulation de l'épissage alternatif afin de promouvoir la bonne expression des gènes à des stades spécifiques du développement.

#### 1.1.3.3 Patron d'expression de *CHD7*

L'expression du gène *CHD7* commence très tôt au cours du développement. Par exemple, dans des tissus embryonnaires de souris, *Chd7* est exprimé dans l'ectoderme neural, dans le TN, puis dans les CCN et ses dérivés (Fujita et al., 2014; Gage et al., 2015; J. Gao et al., 2024). Son expression a également été détectée dans les tissus associés au syndrome CHARGE. Il s'agit notamment de cellules précurseurs du système nerveux central, de l'oreille interne, de l'œil, du cœur, du rein, des gonades et de certains épithéliums et ganglions (Bosman et al., 2005; Feng et al., 2017; J. Gao et al., 2024; Krueger & Morris, 2022; N. C. Reddy et al., 2021; S. Yan et al., 2020). Chez l'adulte, l'expression de *Chd7* diminue, mais elle est néanmoins maintenue dans certains tissus comme le cerveau (Feng et al., 2013). Dans les études *in vitro*, les modèles cellulaires les plus utilisés pour étudier CHD7 sont les ESCs, les cellules de la crête neurale spécialisées, ainsi que leur différenciation en divers types de précurseurs (Bajpai et al., 2010; Schnetz et al., 2010). Son expression est très importante au cours du développement. En effet, les mutations du gène *Chd7* dans les modèles de souris transgéniques sont léthales à l'état homozygote : un seul allèle muté suffit à générer les phénotypes associées au syndrome CHARGE (Bosman et al., 2005). L'étude Bajpai et al a montré que CHD7 est essentiel à l'activation de facteurs de transcription (*Sox9*, *Twist* ou *Snail2*) requis pour la spécialisation, la migration, et



la différenciation des CCN (Bajpai et al., 2010; Pauli et al., 2017). Cependant, il n'est pas essentiel à l'induction et à la spécification des bordures de la crête neurale (*Msx1*, *Zic* ou *Pax3*) (Pauli et al., 2017). Les potentiels défauts engendrés par la diminution de *CHD7* surviendraient après la fermeture du tube neural, ce qui est cohérent avec les anomalies observées chez les patients CHARGE.

#### 1.1.3.4 Les autres causes génétiques

La mutation du gène *CHD7* est impliquée dans le syndrome CHARGE dans uniquement ~70% des cas, d'où l'intérêt d'identifier d'autres gènes candidats. Au cours de ces dernières années, d'autres gènes associés au syndrome CHARGE codant également pour des facteurs de la chromatine ou des facteurs d'épissage ont été identifiés (Bérubé-Simard & Pilon, 2019) ; Il s'agit : du gène *PUF60* (*Poly(U)-binding splicing factor*) codant pour un facteur d'épissage, du gène *RERE* (*Arginine-glutamic acid dipeptide repeats*) codant pour un répresseur co-transcriptionnel interagissant avec des déacétylases d'histone (HDAC; en anglais *Histone Deacetylase*) et des gènes *EP300* (*E1A binding protein P300, acétyltransférase*), *KMT2D* (*Lysine méthyltransférase 2D*) et *KDM6A* (*Lysine déméthylase 6A*) codants pour des facteurs de la chromatine (Jordan et al., 2018; Moccia et al., 2018a). Les gènes *SEMA3E* (*Semaphorin 3E*) et *EFTUD2* (*Elongation Factor Tu GTP Binding Domain Containing 2*) font également partie des gènes identifiés, malgré qu'ils soient déjà associés à d'autres syndromes congénitaux (Bérubé-Simard & Pilon, 2019). Enfin, le laboratoire du Pr Nicolas Pilon a récemment identifié le gène *FAM172A* (*Family With Sequence Similarity 172 Member A*) comme étant un nouveau gène responsable du syndrome. De manière complémentaire, des mutations dans *FAM172A* (pE228Q et pR306\*) ont été identifiées chez 2 familles de patients CHARGE sans mutations sur le gène *CHD7*. Malgré la découverte de tous ces nouveaux gènes associés, les mécanismes pathogéniques sous-jacents communs à tous ces facteurs restent encore très flou.

#### 1.1.4 Modèles d'étude du Syndrome CHARGE

Il existe plusieurs modèles d'études pour le syndrome CHARGE. Les plus populaires sont notamment des modèles de souris et de poisson-zèbres (*Danio rerio*) avec des mutations sur le gène *Chd7*. L'un de ses modèles est une lignée de souris générée par *gene-trapping* qui a fait l'objet de plusieurs études : le modèle *Chd7<sup>Gt</sup>* (*Gene Trap*) (Adams et al., 2007; Hurd et al., 2007). L'allèle en question est létale à l'état homozygote, les embryons de souris ne survivent pas au-delà du jour embryonnaire 10,5 (Hurd et al., 2007). Un allèle mutant suffit à engendrer les anomalies associées au syndrome CHARGE, qui se présentent sous la forme de malformations crâniofaciales, d'un retard de croissance, de problèmes cardiaques et cérébraux, ainsi que d'anomalies aux oreilles internes. Plus précisément, des problèmes au niveau des canaux semicirculaires de l'oreille interne provoquent un comportement de *circling* chez la souris, c'est-à-dire qu'elle tourne très rapidement sur elle-même (Adams et al., 2007). Ce modèle de souris sera utilisé dans la partie II.

## 1.2 Le gène *FAM172A*, nouveau gène responsable du syndrome CHARGE

### 1.2.1 Le modèle de souris *Toupee*

Dans le but d'identifier de nouveaux gènes impliqués dans le développement des CCN et d'étudier les maladies associées, le laboratoire du Pr Nicolas Pilon a généré des modèles de souris transgéniques à l'image de plusieurs neurocristopathies. Pour ce faire, une mutagenèse par insertion aléatoire d'un transgène dans des souris à fond FVB (albinos) a été effectuée. Le transgène exprime le minigène de la tyrosinase permettant la restauration de la pigmentation dans les lignées devenues transgéniques. Les mélanocytes étant des dérivés des CCN, un défaut dans cette pigmentation restaurée distingue les modèles dont un gène important pour les CCN a été touché (Pilon, 2016). L'un de ces modèles, nommé *Toupee* en référence à l'absence de pigmentation sur le dessus de sa tête comme un « toupet », récapitule parfaitement les phénotypes associés au syndrome CHARGE. En effet, les principales anomalies observées chez les souris *Toupee* homozygotes (*Tp/Tp*) sont le colobome, un retard de croissance, la fente labio-palatine, un défaut aux nerfs crâniens, et des malformations cardiaques et génitales (Bélanger et al., 2018). On observe également un comportement de *circling* comme précédemment décrit chez les souris *Chd7<sup>Gt/+</sup>*.

Le séquençage du génome des souris *Toupee* a mis en évidence l'insertion du transgène dans le dernier intron du gène *Fam172a* situé sur le chromosome 13 et non sur le locus du gène *Chd7*. Ce gène semble être exprimé de manière ubiquitaire pendant le développement (e10,5 à e15,5), avec un léger enrichissement au niveau du tissu crânial (Bélanger et al., 2018). Cependant, l'insertion du transgène dans une région non-codante du gène chez les souris *Toupee* homozygotes (*Fam172a<sup>Tp/Tp</sup>*) déstabilise le transcrite, laissant une expression résiduelle d'environ 15%. De plus, une copie de l'allèle *Fam172a<sup>Tp</sup>* interagit génétiquement avec un allèle muté de *Chd7*, aggravant les

phénotypes observés, et confirmant l'implication des deux gènes dans les mécanismes pathologiques menant au syndrome CHARGE (Bélanger et al., 2018).

Pour mieux comprendre les mécanismes cellulaires et moléculaires mis en cause dans ce modèle d'étude, les souris ont été croisées avec un autre modèle dont les cellules de la crête neurale contiennent un rapporteur fluorescent rouge (RFP; en anglais *Red fluorescent protein*). Cette particularité permet de trier et récupérer spécifiquement les cellules fluorescentes à partir d'embryons de souris à différents stades de développement. Ce rapporteur est sous le contrôle du promoteur du gène *Gata4* (*GATA binding protein 4*) distinguant ainsi les cellules de la crête neurale des autres types cellulaires (Pilon et al., 2008). Des analyses d'imagerie en temps réel effectuées sur ces embryons *Fam172a<sup>Tp/Tp</sup>; pGata4-RFP* ont montré des défauts dans la prolifération, la survie et la migration des CCN (Bélanger et al., 2018). Au niveau moléculaire, le séquençage du transcriptome de ces mêmes CCN obtenues à partir d'embryons au jour embryonnaire 10,5 (e10,5) a révélé la dérégulation de la transcription et de l'épissage alternatif de gènes associés à leur développement (Bélanger et al., 2018). Des études de validations des défauts d'épissage dans une cohorte de patients atteints du syndrome CHARGE présentant une mutation dans le gène *CHD7*, dans le gène *FAM172A* ou dans un gène inconnu, ont permis de proposer l'hypothèse que la dérégulation de l'épissage alternatif co-transcriptionnelle serait le mécanisme pathogénique commun pour tous les cas du syndrome CHARGE (Bélanger et al., 2018; Bérubé-Simard & Pilon, 2019). De plus, un traitement *in vitro* à la rapamycine est capable de corriger les défauts d'épissage alternatif observés. La rapamycine est une petite molécule qui inhibe la transcription des gènes ribosomiaux à travers la voie de signalisation mTOR (*mechanistic target of rapamycin*), elle permet ainsi indirectement une meilleure disponibilité des facteurs d'épissages pour les autres pré-ARN messagers (Munding et al., 2013).

### 1.2.2 La protéine FAM172A

Chez l'homme, le gène *FAM172A* est situé sur le chromosome 15q15 et code pour une protéine très peu étudiée. D'après la base de données (Ensembl), *FAM172A* possède deux isoformes principaux (Figure 1.7). Dans des embryons murins, l'isoforme le plus exprimé possède 371 acides aminés (42 KDa) traduits à partir de 10 exons. Secondairement, un isoforme plus long est également exprimé possédant 417 acides aminés (48 KDa) traduits à partir de 11 exons. Des analyses bio-informatiques de la séquence protéique de *FAM172A* prédisent de nombreux domaines fonctionnels, tel qu'un grand domaine homologue à la protéine Arb2 (*Argonaute binding protein-2*) chez la levure, un signal de localisation au noyau (NLS; en anglais *Nuclear Localization Signal*), des sites consensus de phosphorylation par la Caséine Kinase CK2 et un motif sérine hydrolase (*esterase-like*) (Figure 1.7). Deux autres domaines plus spécifiques à l'isoforme long de *FAM172A* ont également été prédits, il s'agit d'un peptide signal et un domaine de rétention au réticulum endoplasmique (RE). Ces motifs sont très bien conservés dans plusieurs espèces, tout comme la majorité de la protéine *FAM172A* qui montre un taux de 93% de similarité entre les orthologues murin et humain.

Afin d'étudier les potentiels rôles que pourrait avoir la protéine *FAM172A* dans le développement des CCN, plusieurs analyses ont été effectuées. Des expériences en spectrométrie de masse (LC-MS/MS), entre autres dans la lignée cellulaire Neuro2A (généré à partir d'un neuroblastome cérébral murin, d'origine crête neurale) a révélé que la protéine *FAM172A* pouvait interagir physiquement avec de nombreuses protéines impliquées dans le remodelage de la chromatine, la régulation de la transcription et de l'épissage alternatif (Bélanger et al., 2018, 2022). En plus des résultats obtenus avec le modèle de souris *Toupee*, ces données renforcent l'idée que *FAM172A* joue un rôle dans la transcription et la régulation de l'épissage alternatif au niveau de la chromatine, et ce probablement avec CHD7. En effet, *FAM172A* est

également capable d'interagir physiquement avec la protéine CHD7 (Bélanger et al., 2018). Néanmoins, le rôle précis de FAM172A dans ces mécanismes moléculaires reste encore à déterminer.

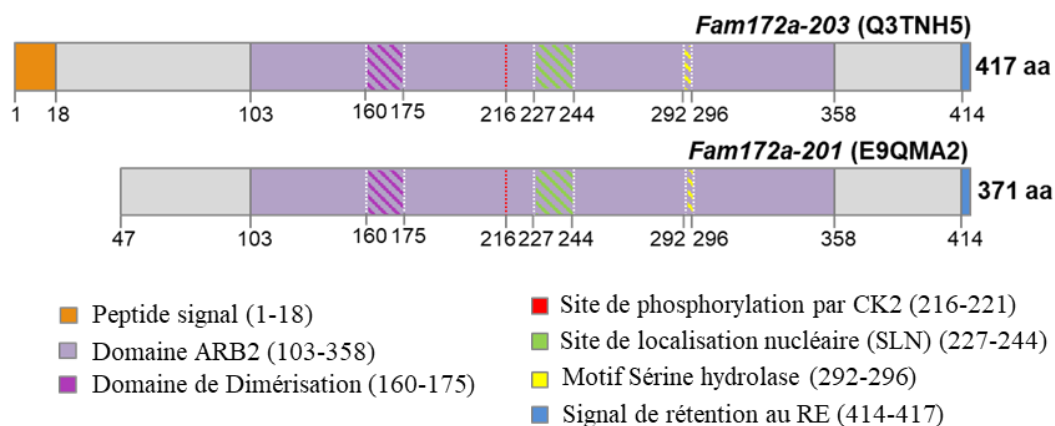


Figure 1.7 Schéma représentant les différents domaines des deux isoformes principaux de FAM172A.

L'activité hydrolase de FAM172A, via la sérine en position 294 (Figure 1.7), a été mise en évidence grâce à une réaction *in vitro* utilisant la sonde fluorescente TAMRA-FP. Cette molécule se lie de manière covalente à FAM172A lorsqu'elle est utilisée comme substrat et la rend visible en fluorescence Cy3 après migration sur gel SDS-PAGE (Bélanger et al., 2018). Cependant, la ou les cible(s) de cette activité hydrolase ne sont pas encore connue(s).

Concernant les sites de phosphorylation prédits par PhosphoSitePlus, un peptide contenant plusieurs sérines successives ( $_{216}\text{S-S-S-D}$ ) possède le plus grand indice de confiance (Figure 1.7). Il s'agit d'une séquence consensus pour la Caséine Kinase CK2 qui est [S/T-X-X-D/E] (avec X n'importe quel acide aminé), et plus rarement [S/T-D/E] (Litchfield, 2003). La kinase CK2 est très bien conservée au cours de l'évolution, elle est importante entre autres pour les mécanismes de survie cellulaire (prolifération, apoptose, cancer, etc.) (Litchfield, 2003; Pinna & Meggio, 1997). Néanmoins, le

nombre d'études portant sur le rôle de CK2 dans les processus de développement et les maladies associées n'a cessé de croître ces dernières années, ce qui rend la phosphorylation de FAM172A par cette kinase très intéressante à étudier (Borgo et al., 2021; Halloran et al., 2022).

Les outils bioinformatiques prédisent également un signal de localisation nucléaire (NLS) dans la séquence protéique de FAM172A, un domaine riche en acides aminés basiques qui lui permettrait d'être transloqué vers le noyau cellulaire (Figure 1.7). En effet, les protéines ne peuvent pas diffuser librement entre le cytoplasme et le noyau, elles nécessitent l'aide d'une famille de protéines appelée les importines (de type karyophérine) pour traverser les complexes des pores nucléaires (NPC). Il existe plusieurs types de NLS : des séquences dites « classiques » reconnues par la voie canonique des importines, constituée d'une sous-unité  $\alpha$  (1 à 7) et une sous-unité  $\beta$ 1, et des séquences « non-classiques » reconnues par d'autres types d'importines  $\beta$  (Tableau 1.1) (Lu et al., 2021; Wing et al., 2022). De plus, les séquences classiques peuvent être monopartites ou bipartites (Tableau 1.1). Le NLS prédit de FAM172A semble faire partie de cette deuxième catégorie (Bélanger et al., 2018). Concernant cette voie canonique de translocation, l'importine  $\beta$ 1 assure les interactions avec les facteurs Nup (Nucléoporines) du NPC pendant que l'importine  $\alpha$  reconnaît le NLS et sert de protéine adaptatrice entre la protéine contenant le NLS et l'importine  $\beta$ 1 (Figure 1.8). On remarque que l'importine  $\alpha$  participe également à la translocation en interagissant par exemple avec Nup153 (Lu et al., 2021). Ce complexe est ensuite transloqué jusqu'au noyau où il subit un changement de conformation en interagissant avec les protéines Ran (*Ras-related nuclear protein*) couplées à un GTP (Guanosine triphosphate). Ce changement dissocie le complexe NLS/importine  $\alpha$ /importine  $\beta$ 1/Ran en un complexe NLS /importine  $\alpha$  d'une part et un complexe importine  $\beta$ 1/RanGTP d'une autre part. L'importine  $\beta$ 1 sera ensuite renvoyée au cytoplasme où elle se dissocie de la protéine Ran grâce à l'hydrolyse du GTP en GDP (Guanosine diphosphate).

L'hydrolyse du GTP fournit l'énergie nécessaire au cycle d'import des protéines (Figure 1.8) (Lu et al., 2021).

Enfin, pour mieux comprendre les fonctions de FAM172A, son grand domaine homologue à la protéine Arb2 initialement identifié chez la levure a fait l'objet de plusieurs investigations. Ce domaine permet à FAM172A d'interagir physiquement avec la protéine Argonaute 2 (AGO2) (Bélanger et al., 2018) qui sera décrite plus en détails dans la section suivante.

Catégorie		Séquence	Récepteur de transport
Signal de localisation nucléaire (NLS) classique	NLS MP (Monopartite)	<b>PKLKRQ</b>	Importine $\alpha/\beta$
		<b>RPRK</b>	
		<b>RRARRPRG</b>	
	NLS BP (Bipartite)	<b>GKRKLITSEEERSPAKRGRKS</b>	
		<b>KGKKGRTQKEKKAARARSKGKN</b>	
		<b>RKRCAAGVGGGPAGCPAPGSTPLKKPRR</b>	
		<b>RKPVTAQERQREEREEKRRRRQERAKEREKRRQERER</b>	
	Signal de localisation nucléaire (NLS) non-classique	PY- NLS	
<b>TLLLRETMNNLGVSDHAVLSRKTPQPY</b>			
<b>PGKMDKGEHRQERDRPY</b>			
Autre NLS non-classique		<b>GKKKKGKPGKRREQRKKRRRT</b>	
		<b>SANKVTKNKSNSSPYLNKRKGKPGPDS</b>	
		<b>VHSHKKKKIPTSPTFTTPKTLTLRRQPKYPRKSAPRRNKLDHY</b>	

Tableau 1.1 Classification des signaux de localisation nucléaire. Tableau adapté de *Lu et al, 2021*.



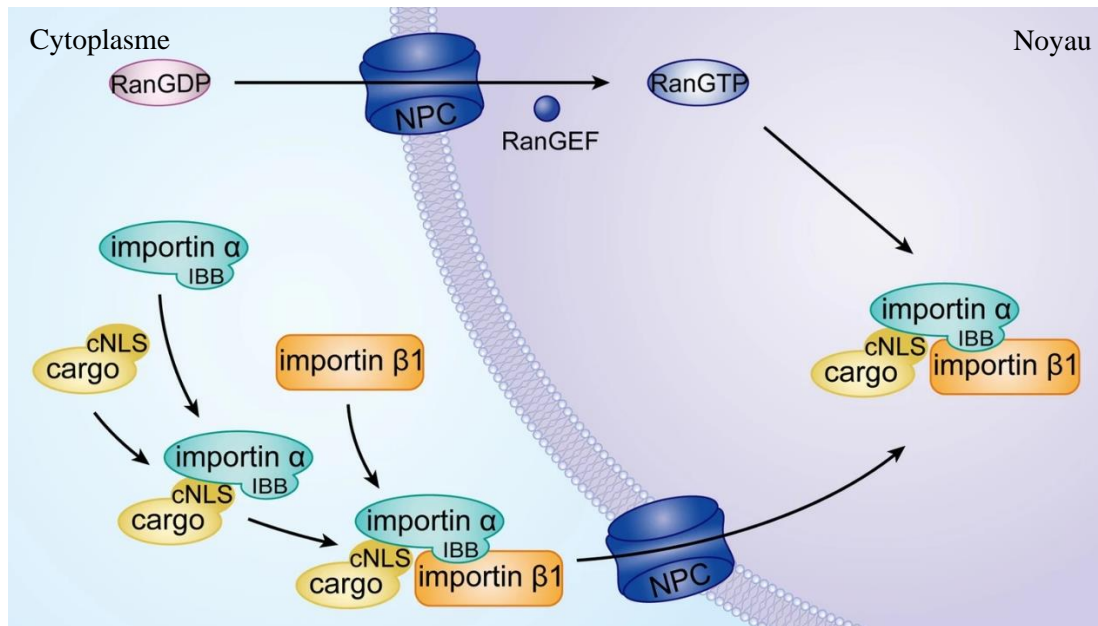


Figure 1.8 Les mécanismes du transport nucléaire classique grâce aux importine  $\alpha/\beta$ 1 chez les mammifères. Schéma adapté de *Lu et al, 2021*.

### 1.2.3 Interaction avec Argonaute-2

#### 1.2.3.1 Le domaine ARB2

Les protéines Arb1 (*Argonaute binding protein-1*) et Arb2 (*Argonaute binding protein-2*) ont été identifiées initialement chez *Schizosaccharomyces pombe* lors d'une expérience de spectrométrie de masse cherchant à identifier les interacteurs de la protéine Argonaute (Ago1) (Buker et al., 2007). Cette même étude a montré que le complexe formé d'Ago1, Arb1 et Arb2 chez *S.pombe* est important pour la génération de petits ARN interférents (ARNsi) et la génération d'hétérochromatine au niveau des centromères à travers la méthylation de la lysine 9 de l'histone 3 (H3K9me). Depuis, d'autres protéines avec des domaines homologues à Arb2 ont été identifiées chez d'autres espèces. Le domaine ARB2 le plus étudié aujourd'hui se trouve sur le facteur de la chromatine Hda1 chez *Saccharomyces cerevisiae*. Hda1 contient deux domaines principaux : Un domaine catalytique HDAC qui élimine les groupements acétyles sur les histones, et un domaine ARB2 (Lee et al., 2009). La déplétion de ce dernier n'a aucun impact sur l'activité catalytique de la protéine Hda1, suggérant que ce domaine jouerait plutôt un rôle de recrutement à la chromatine. En effet, une étude effectuée chez *S. cerevisiae* utilisant la cristallographie a démontré que le domaine ARB2, sous la forme d'homodimère, est capable de lier directement les nucléosomes *in vitro* (Shen et al., 2016). Plus précisément, le recrutement de Hda1 a lieu au niveau de gènes transcriptionnellement très actifs, favorisé par une interaction physique entre l'ARN Polymérase II (ARNPII) et le domaine ARB2 (Ha et al., 2019).

FAM172A est l'orthologue de Arb2 chez les mammifères. Au moyen de ce domaine homologue, nous avons pu émettre des hypothèses sur la ou les fonction(s) de FAM172A. Pour le moment, des études de co-immunoprécipitation dans des extraits cellulaires de Neuro2A et dans un tissu murin associé aux CCN ont confirmé que notre

protéine est capable d'interagir physiquement avec la protéine Argonaute-2 (AGO2), mais curieusement pas avec la protéine Argonaute-1 (AGO1) (Bélangier et al., 2018).

### 1.2.3.2 Les protéines AGO

Il existe huit protéines de la famille Argonaute dans le génome humain, quatre de la sous-famille AGO (AGO1-4) et quatre de la sous-famille PIWI (PIWIL1-4) (Sasaki et al., 2003). Les protéines Argonautes sont nommées ainsi d'après le phénotype de mutants *argonauts* chez *Arabidopsis thaliana*, qui sont eux-mêmes nommés à cause de leur ressemblance avec les tentacules de la pieuvre *Argonauta argo* (Bohmer et al., 1998). Les protéines AGO sont capables de réguler l'expression des gènes grâce à leur association avec des petits ARN non-codants (20 à 25 nucléotides). Cette régulation peut avoir lieu au niveau post-transcriptionnel dans le cytoplasme (en anglais *post-transcriptional gene silencing* ou *RNA interference*) ou au niveau transcriptionnel dans le noyau (en anglais TGA ; *transcriptional gene activation* ou TGS ; *transcriptional gene silencing*). Dans le cytoplasme, les protéines AGO appartiennent à un grand complexe nommé RISC (*RNA-induced silencing complexe*) et se lie à différentes classes de petits ARN non-codants, notamment les microARN (ARNmi) et les ARNsi. Ces petits ARN guident le complexe RISC à leur ARN messager (ARNm) cible par complémentarité de séquences, permettant ainsi l'inhibition de la traduction par séquestration ou dégradation des ARNm. Environ 60% des gènes humains codant pour des protéines seraient régulés par des microARN, et ce, dû à l'existence d'une séquence complémentaire au ARNmi située dans le 3'UTR (*Untranslated transcribed region*) de l'ARNm cible appelée MRE (*MicroRNA Response Element*) (Friedman et al., 2009). Chez les mammifères, toutes les protéines AGO sont capables de se lier avec la même affinité aux petits ARN et de participer à la voie des microARN par inhibition de la traduction (Su et al., 2009a). Cependant, seules AGO2 et AGO3 possèdent une activité catalytique qui dégradent l'ARNm cible lorsqu'associée à un petit ARN parfaitement complémentaire à la cible (J. Liu et al., 2004; Meister et al., 2004; Park et

al., 2017, 2020). Pour exercer cette fonction, les membres de la famille AGO possèdent trois domaines bien conservés au cours de l'évolution : Le domaine PAZ (*Piwi-Argonaute-Zwille*) qui reconnaît et se fixe à l'extrémité 3' de l'ARN guide, le domaine MID qui permet les interactions entre protéines, et le domaine PIWI qui adopte une conformation secondaire ressemblant à une RNase H et constitue le domaine catalytique de certaines protéines AGO (Song et al., 2003; K. S. Yan et al., 2003).

De retour au modèle de souris *Toupee*, les défauts moléculaires observés semblent être restreints aux mécanismes ayant lieu au noyau : la transcription et l'épissage alternatif. En effet, en utilisant un rapporteur luciférase dont l'ARNm possède des séquences MRE reconnues par *let-7*, un ARNmi connu pour s'associer à AGO2, une expérience a montré que les souris *Fam172a<sup>TP/TP</sup>* n'ont pas de défauts au niveau de la voie des microARN (Bélanger et al., 2018).

### 1.2.3.3 Les fonctions nucléaires d'AGO2

Chez l'homme, au-delà de cette fonction canonique dans le cytoplasme, plusieurs études ont montré qu'AGO1/2 joue également des rôles importants dans le noyau cellulaire. En effet, les ARNmi/ARNsi associés à AGO2 peuvent guider le complexe jusqu'à la chromatine au niveau de plusieurs types d'ARN naissants non-codant, et ainsi participer à la régulation de la transcription, de la régulation de l'épissage alternatif, de la réparation de l'ADN et du maintien de l'intégrité du génome (Alló et al., 2009; Ameyar-Zazoua et al., 2012; Gutbrod & Martienssen, 2020; Nazer et al., 2022). Le recrutement nucléaire utilise de petits ARN transcrits en sens ou antisens à partir de régions promotrices de gènes, ou encore de long ARN non-codant (ARNlnc) transcrit en antisens à partir de régions intragéniques. A noter qu'AGO2 peut également être guidé jusqu'au pré-ARNm en cours de transcription, ou s'y associer directement à travers des séquences riches en guanines (Nazer et al., 2022). Le premier mécanisme a été démontré avec l'ARN du promoteur du gène *cyclooxygenase 2 (cox-2)* où AGO2

favorise la transcription en augmentant le niveau de triméthylation de la lysine 4 de l'histone (H3K4me3) (Matsui et al., 2013). D'autres exemples montrent qu'AGO2 peut également réprimer les gènes dans des cellules cancéreuses, toujours en interagissant avec l'ARN des régions promotrices (Nazer et al., 2022; Younger & Corey, 2011). Une autre manière de réguler la transcription est d'agir sur la position des nucléosomes autour du TSS. En effet, AGO2 est guidé jusqu'à de petits ARN transcrits à partir du TSS et coopère avec le complexe SWI/SNF pour assurer le bon positionnement du nucléosome en position +1 de la transcription (Carissimi et al., 2015). CHD7 interagit également avec plusieurs composants du complexe SWI/SNF dans les CCN (Bajpai et al., 2010). En s'appuyant sur les différents rôles que joue AGO2 dans le noyau, l'une de ses fonctions nucléaires bien conservées de la levure jusqu'à l'homme semble être le recrutement de facteurs de la chromatine au niveau de séquences cibles, tel que les méthyltransférases d'histone ou le complexe SWI/SNF comme cité précédemment (Batsché & Ameyar-Zazoua, 2015). Pour rappel chez *S.pombe*, la protéine Ago1 stimule la formation d'hétérochromatine grâce à la triméthylation de H3K9 à travers le recrutement de Clr4 (orthologue de SUV39H1 chez les mammifères) (Buker et al., 2007). Un mécanisme similaire a également été décrit pour réguler l'épissage alternatif de certains gènes chez l'homme. En effet, AGO2 est guidé jusqu'à des ARN transcrits en antisens à partir de régions codantes pour des exons alternatifs, où elle favorise le recrutement de SUV39H1 et EHMT2 et la triméthylation de H3K9 (Figure 1.9A) (Ameyar-Zazoua et al., 2012; Batsché & Ameyar-Zazoua, 2015). L'augmentation de cette marque d'histone provoque ainsi la compaction de la chromatine et le ralentissement de la RNAPII aux niveaux d'exons alternatifs. Par ce biais, AGO2 est capable de favoriser l'inclusion de ces exons alternatifs au sein de plusieurs gènes tel que le gène *CD44* (Figure 1.9A) (Ameyar-Zazoua et al., 2012). Ce mécanisme sera décrit plus en détails dans la partie 1.3.3.

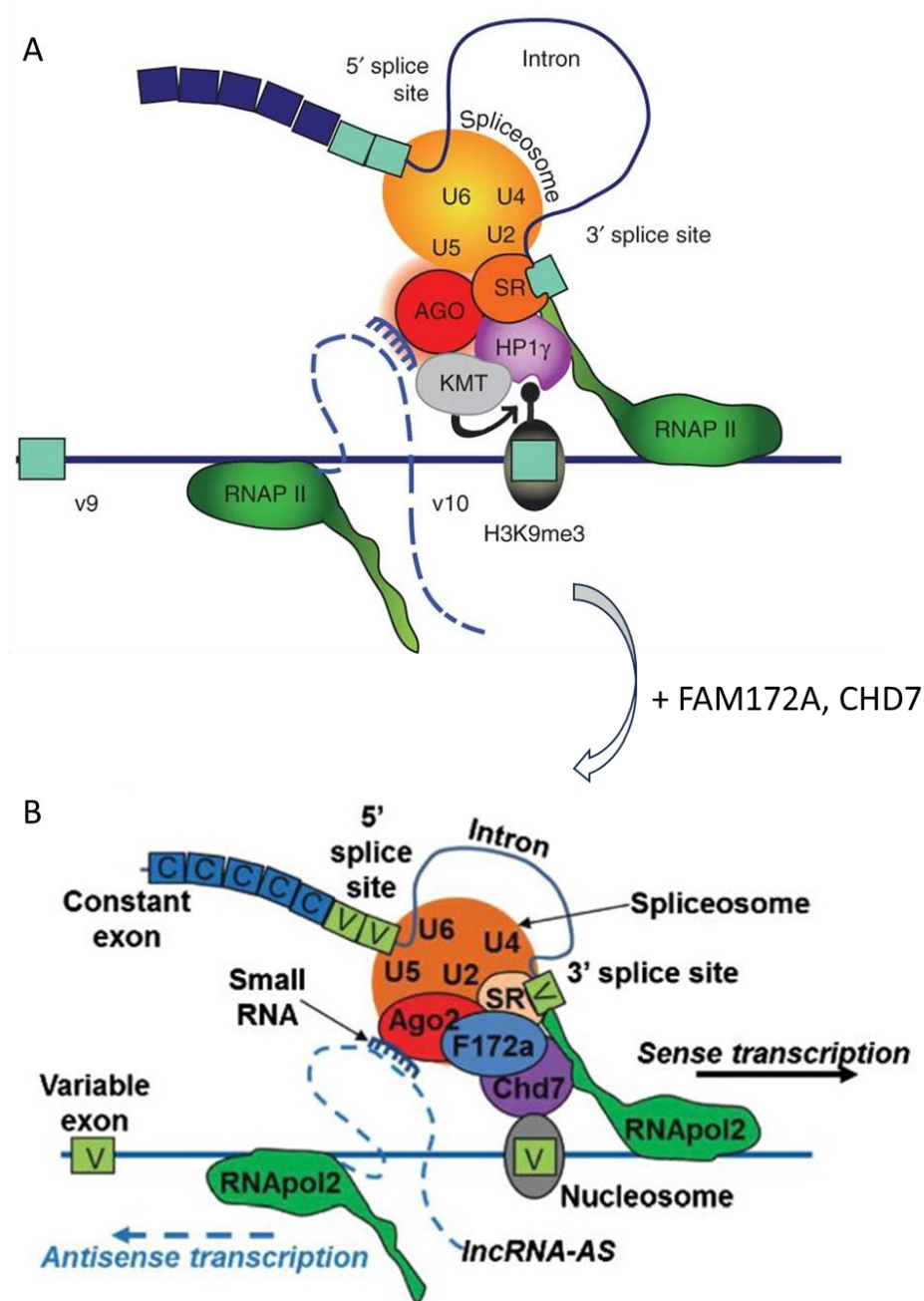


Figure 1.9 Le rôle potentiel du complexe FAM172A-CHD7-AGO2 dans la régulation de l'épissage alternatif co-transcriptionnel. (A) Rôle nucléaire de la protéine AGO2 dans la régulation de l'épissage alternatif co-transcriptionnel grâce à des petits ARN chez l'homme. (B) Le rôle potentiel des protéines FAM172A et CHD7 dans la régulation de l'épissage alternatif médiée par AGO2. Schéma adapté de Ameyar-Zazoua *et al*, 2012 et Bélanger *et al*, 2018.

Au vu de ses différents rôles dans le cytoplasme et le noyau, AGO2 se doit d'être très dynamique entre ces deux compartiments (Kalantari et al., 2016; Ohrt et al., 2008; Schraivogel & Meister, 2014). En effet, les protéines AGO2 sont réparties entre le cytoplasme et le noyau selon les besoins de la cellule. Il est d'ailleurs connu que cette répartition varie beaucoup d'un type cellulaire à l'autre. Par exemple, AGO2 sera majoritairement nucléaire dans des cultures de cellules primaires, alors qu'elle sera plutôt cytoplasmique dans des lignées cellulaires immortalisées (Sarshad et al., 2018; Sharma et al., 2016). Les conditions qui influent également sur le ratio noyau/cytoplasme sont notamment le stress cellulaire, la différenciation et l'activation de la sénescence (Benhamed et al., 2012; Castanotto et al., 2018; Rentschler et al., 2018; Sarshad et al., 2018). Les mécanismes qui permettent à AGO2 de voyager entre les deux compartiments sont encore flous car elle ne possède pas de NLS ou de signal d'export nucléaire (NES; en anglais *Nuclear Export Signal*). Les protéines TNRC6 (*Trinucleotide repeat containing 6*) fortement associées à AGO2 et au complexe RISC possèdent les deux domaines : NLS et NES. Pourtant son implication est limitée à l'export d'AGO2 vers le cytoplasme (Nishi et al., 2013; Schraivogel et al., 2015a). Des études effectuées dans le but de déterminer les mécanismes d'import d'AGO2 au noyau ont désigné l'importin-8 (appartenant aux importines de type  $\beta$ ), alors que d'autres études ont déterminé que plusieurs importines étaient impliquées de manière redondante (Schraivogel et al., 2015a; Weinmann et al., 2009). D'autre part, aucune protéine contenant un NLS, classique ou non-classique, reconnu par les importines et participant à l'import nucléaire d'AGO2 n'a été décrite à ce jour.

#### 1.2.3.4 Le complexe FAM172A-CHD7-AGO2

Les expériences de co-immunoprécipitation effectuées sur des extraits cellulaires de Neuro2A montrent que FAM172A et AGO2 se trouvent dans un grand complexe nucléaire accompagné de la protéine CHD7. Ces mêmes expériences montrent également que l'interaction entre AGO2 et CHD7 est renforcée par la surexpression de

FAM172A, ce qui suggère que les trois protéines appartiennent à un même complexe nucléaire (Bélanger et al., 2018). Plus encore, FAM172A pourrait potentiellement être le pont entre les deux, ou être responsable du recrutement de l'une d'elles (Figure 1.9B). Un traitement à la DNase ou à la RNase seule n'a pas d'impact sur l'interaction physique entre FAM172A et AGO2, en revanche, un traitement avec les deux enzymes diminue de moitié leur interaction (Bélanger et al., 2018). Cela suggère qu'une partie de l'interaction FAM172A-AGO2 serait indirecte et stabilisée par les deux types d'acides nucléiques comme à l'interface entre le spliceosome et la chromatine (Bélanger et al., 2018). Néanmoins, l'autre partie du complexe ne dépend ni de l'ARN, ni de l'ADN, ouvrant la possibilité que les deux protéines pourraient également interagir de manière directe, renforçant ainsi l'hypothèse que FAM172A pourrait participer au recrutement d'AGO2 à l'interface chromatine-spliceosome. Notre modèle d'étude actuel décrit que le complexe FAM172A-AGO2-CHD7 participerait aux mécanismes de régulation de l'épissage alternatif co-transcriptionnel (Figure 1.9B) (Bérubé-Simard & Pilon, 2019). Des expériences de CHIP et d'immunoprécipitation de l'ARN suivi d'analyses quantitatives appuient notre hypothèse en confirmant l'enrichissement de FAM172A et de CHD7 au niveau d'un exon alternatif du gène *Cd44*, une cible d'épissage connu d'AGO2 (Figure 1.9) (Ameyar-Zazoua et al., 2012; Bélanger et al., 2018). La même observation a été effectuée au niveau d'un exon alternatif du gène *LEF1* dans un contexte de différenciation et de détermination de cellules sexuelles mâles au cours du développement (Bélanger et al., 2022). L'interaction entre les deux protéines FAM172A et AGO2 est déstabilisée lorsqu'une version de FAM172A mimant un variant trouvé chez des patients CHARGE est utilisée. Ce variant contient une mutation ponctuelle pE229Q (ou pE228Q chez l'homme) qui modifie l'acide glutamique se trouvant dans le domaine ARB2 de FAM172A. Ce défaut d'interaction pourrait expliquer la pathogénicité de ce variant dans le contexte du syndrome CHARGE. C'est pourquoi, il est intéressant d'approfondir l'étude des mécanismes moléculaires par lesquels FAM172A, AGO2 et CHD7 régulent les décisions d'épissage alternatif co-transcriptionnel.



### 1.3 Mécanismes moléculaires mis en cause dans le syndrome CHARGE

Dans le but de générer les protéines nécessaires à son fonctionnement à partir de son génome, la cellule va utiliser un intermédiaire appelé l'ARN messager (ARNm). Dans le noyau, c'est l'ARN polymérase II (ARNPII) qui va synthétiser ces ARNm lors de la transcription. Ces derniers subiront ensuite la maturation pour devenir des ARNm fonctionnels et seront exportés dans le cytoplasme pour être traduits en protéines.

#### 1.3.1 La transcription

##### 1.3.1.1 Transcription par l'ARN polymérase II

La transcription est le processus par lequel la cellule synthétise une séquence d'ARN à partir d'une séquence d'ADN. La transcription est effectuée par trois enzymes appelées les ARN polymérases, chacune d'elle va transcrire une catégorie différente de gènes (Sentenac, 1985). L'ARN polymérase I (ARNPI) transcrit les ARN ribosomiques (ARNr), alors que l'ARN polymérase III (ARNPIII) transcrit les ARN de transfert (ARNt) et certains ARNr, tous deux nécessaires à la traduction de protéines. Quant à la RNAPII, elle transcrit les ARN messagers et une grande partie des ARN non-codants. Le processus de transcription est commun à toutes les ARN polymérases et se déroule en plusieurs étapes : l'initiation, l'élongation et la terminaison (Figure 1.10). En revanche, ces polymérases diffèrent dans leurs associations protéiques et leurs mécanismes de régulation (Cramer, 2019). Chez l'homme, la RNAPII est constituée de 12 sous-unités RPB1 à RPB12 (*RNA Polymérase II subunit*). Elle fait également partie d'un large complexe composé de facteurs généraux de transcription et de nombreuses autres protéines capables de lier l'ADN et l'ARN. Les facteurs généraux, tels que TFIIA, TFIIB, TFIID, TFIIE, TFIIIF et TFIIH, sont eux-mêmes composés de plusieurs protéines et sous-unités et interviennent aux différentes étapes de la transcription (Schier & Taatjes, 2020).

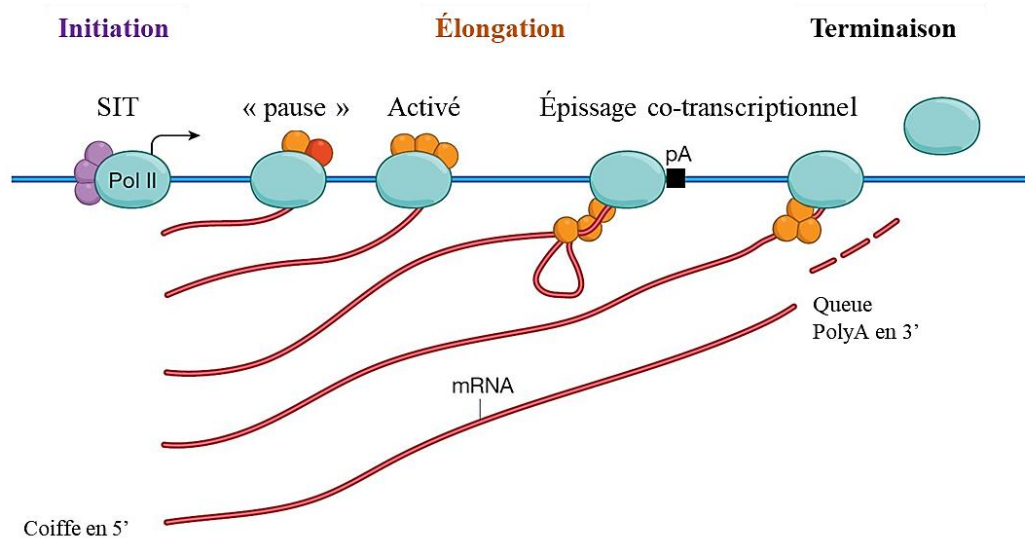


Figure 1.10 Les différentes étapes de la transcription par l'ARN polymérase II : l'initiation, l'élongation et la terminaison. Schéma adapté de *Cramer et al, 2019*.

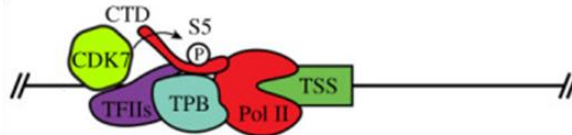
Afin d'initier la transcription, le complexe TFIID et ses sous-unités reconnaissent une séquence promotrice située en amont du site d'initiation de la transcription +1 (TSS) (Cramer, 2019; Louder et al., 2016). Une fois le promoteur reconnu, les facteurs TFIIA, TFIIB, TFIIE, TFIIIF et TFIIH sont recrutés à leur tour et ancrent la RNAPII au promoteur. Ils forment le complexe de pré-initiation de la transcription (PIC; en anglais *Pre-Initiation Complex*) (Schier & Taatjes, 2020). Le rôle principal de ce complexe est d'ouvrir la double hélice d'ADN grâce à la fonction ATP-dépendante de l'enzyme XPB (sous-unité de TFIIH) (Cramer, 2019; Egly & Coin, 2011). Ainsi, le brin matrice (simple brin) se retrouve dans le site actif de la RNAPII, qui initie la synthèse d'ARN de quelques nucléotides (~15-30 nucléotides). Une fois que l'ARN naissant atteint une taille suffisante, de nouveaux facteurs sont recrutés et se forme alors le complexe d'élongation de la transcription. La grande sous-unité de la RNAPII possède un domaine appelé CTD (*carboxyl-terminal domain*) constitué d'une répétition de 7 acides aminés (52 répétitions du peptide 'TSPTSPS' chez l'homme) (Kim et al., 1994). Les fonctions de la RNAPII et la coordination entre transcription et maturation de

l'ARN sont fortement régulées par l'état de phosphorylation de ce domaine CTD. Par exemple, un complexe coactivateur appartenant au PIC, nommée le médiateur, est important pour le passage de l'initiation à l'élongation en stimulant la phosphorylation du domaine CTD par la kinase CDK7 (sous-unité de TFIIF) (Figure 1.11A) (Kornberg, 2005). Au début de l'élongation, et dépendamment du contexte, la RNAPII marque une pause et doit attendre un signal pour poursuivre la transcription (Figures 1.10 et 1.11B). En effet, la RNAPII peut être régulée négativement par certains facteurs d'élongation qui provoquent l'arrêt de la synthèse d'ARN à environ 30-50 nucléotides en aval du TSS (Abuhashem et al., 2022). LA RNAPII a alors besoin d'une signalisation positive de la part d'autres facteurs d'élongation pour poursuivre. Dans le cas contraire, la transcription est interrompue de manière précoce (Abuhashem et al., 2022). Certains complexes tels que NELF (*negative elongation factor*) et DSIF (*DRB sensitivity inducing factor*) sont connus pour favoriser cette pause en interagissant directement avec la RNAPII et en empêchant la signalisation positive par les facteurs d'élongation (Figure 1.11B) (Abuhashem et al., 2022; Vos et al., 2018; Yamaguchi et al., 2013). Afin de quitter l'état de pause et de reprendre l'élongation, la sous-unité CDK9 du facteur p-TEFb phosphoryle le domaine CTD de la RNAPII, ainsi que NELF et DSIF, ce qui provoque la dissociation du complexe ARNPII-NELF-DSIF et la poursuite de la transcription (Figure 1.11C) (Bacon & D'Orso, 2019). Le facteur p-TEFb s'avère également être un facteur essentiel favorisant la transcription tout au long de l'élongation (Abuhashem et al., 2022).

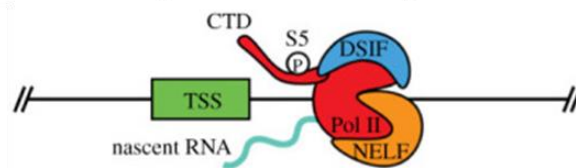
Un hybride ADN-ARN qui correspond au brin d'ADN matrice et à l'ARN naissant se forme au niveau du site actif de la polymérase (Gnatt et al., 2001). En additionnant le second simple brin d'ADN antisens, on obtient une structure appelée une R-loop. Cette structure en forme de « bulle » et se limitant à quelques nucléotides au site de synthèse de l'ARN va suivre la RNAPII tout le long de la transcription (Niehrs & Luke, 2020; Sidorenkov et al., 1998). Il est important pour la cellule que des protéines spécialisées

interviennent pour résoudre les R-loop et permettre l'avancement de la RNAPII le long du gène (Niehrs & Luke, 2020).

**A** Complexe de pré-initiation de la transcription (CPI)



**B** Complexe en état de « pause »



**C** Complexe actif en cours d'élongation

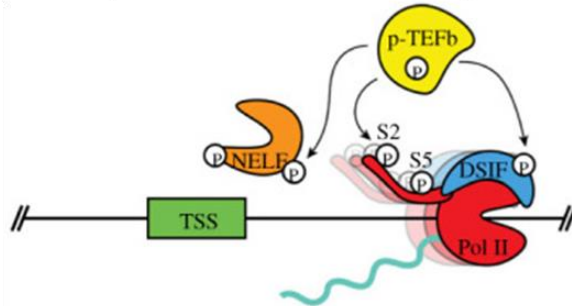


Figure 1.11 Les étapes de l'activation de la transcription chez les mammifères. (A) Le complexe de pré-initiation de la transcription est recruté au niveau des sites d'initiation de la transcription. La kinase CDK7 phosphoryle le domaine CTD de la RNAPII pour initier la transcription. (B) Les complexes DSIF et NELF interagissent avec la RNAPII pour arrêter la transcription au bout de 30-50 nucléotides après le SIT. À ce stade la RNAPII est en état de « pause ». (C) Le facteur p-TEFb phosphoryle le complexe ARNPII-DSIF-NELF pour le dissocier et permet à la RNAPII de poursuivre l'élongation. Schéma adapté de *Abuhashem et al*, 2022.

Au cours de cette élongation, le domaine CTD de la RNAPII continue de recruter de nombreux autres acteurs de la transcription ainsi que des remodeleurs de la chromatine et des facteurs nécessaires à la maturation des ARNm (Conaway et al., 2000; Shilatifard, 1998). En effet, les modifications apportées aux pré-ARNm pour qu'ils deviennent matures ont lieu au fur et à mesure de la transcription. Finalement, la RNAPII rencontre des éléments régulateurs sur la séquence d'ADN appelés terminateur et signal de polyadénylation. Ceux-ci sont reconnus par le complexe CPSF (*cleavage and polyadenylation specificity factor*) et signale à la RNAPII de terminer la synthèse d'ARN.

### 1.3.1.2 Les régions promotrices

Il existe plusieurs types de séquences consensus associées aux promoteurs. Par exemple, la boîte TATA a été découverte en 1979, et elle est encore très étudiée aujourd'hui (Goldberg, 1979). C'est la sous-unité TBP (*TATA-box binding protein*) du complexe TFIID qui reconnaît la boîte TATA située de -25 à -30 nucléotides en amont du TSS (Sainsbury et al., 2015). Chez l'homme, uniquement 3.5% des gènes possèdent une boîte TATA correspondant parfaitement aux séquences consensus tel que TATAAA ou très similaire (Vo ngoc et al., 2017). Néanmoins, 28% des gènes ont également des séquences consensus dégénérées ressemblant assez aux boîtes TATA pour être reconnu par TBP (Vo ngoc et al., 2017). De plus, les boîtes TATA se trouvent préférentiellement au niveau de gènes tissu-spécifiques régulés au cours de la différenciation (Cramer, 2019). Il existe d'autres éléments régulateurs reconnus par les sous-unités TAF de TFIID, parmi eux : l'initiateur (Inr ; *Initiator*) (Smale & Baltimore, 1989), les éléments BRE<sup>u</sup>/BRE<sup>d</sup> (Lagrange et al., 1998), le motif TCT et les îlots CpGs (Bird, 1987; Deaton & Bird, 2011; Vo ngoc et al., 2017). A noter que certains d'entre eux sont spécifiques d'une catégorie de gènes en particulier. Par exemple, les motifs TCT sont préférentiellement retrouvés en amont des gènes ribosomiaux (Perry, 2005). Les éléments BRE<sup>u/d</sup> sont des séquences retrouvées directement en amont et en aval de

la boîte TATA et sont reconnus par les TAF uniquement si la boîte TATA est reconnue par TBP (Deng & Roberts, 2005). Ainsi, elle renforce et régule la transcription uniquement en synergie avec un autre élément régulateur. Et finalement, les promoteurs riches en îlots CpG (séquences riches en répétitions de CG) se trouvent préférentiellement en amont des gènes de ménages, c'est-à-dire des gènes que l'on retrouve dans tous les types cellulaires (Angeloni & Bogdanovic, 2021; Saxonov et al., 2006). Il existe de nombreux mécanismes permettant de réguler la transcription des gènes au niveau de ces régions. Premièrement, les facteurs de transcription sont un élément régulateur important de la transcription (Ptashne & Gann, 1997). Il en existe plus de 2 000 chez l'homme (Cramer, 2019). Ces protéines possèdent 2 domaines caractéristiques : un domaine de liaison directe à l'ADN et un domaine d'interaction avec d'autres protéines. Ces facteurs peuvent avoir pour but d'activer ou de réprimer la transcription. Deuxièmement, afin d'être reconnues par les facteurs de transcription, les régions promotrices doivent être accessibles grâce à une conformation ouverte de la chromatine (euchromatine, voir partie 1.2.2) (Lorch & Kornberg, 2017). Le complexe PIC est en partie responsable du recrutement de remodelleurs de la chromatine afin d'ouvrir la chromatine. Plusieurs régulations épigénétiques interviennent pour ouvrir cette chromatine au niveau des promoteurs. Par exemple, les marques épigénétiques H3K4me3 et H3K27me3 corrélaient respectivement avec l'activation ou la répression de la transcription (Piunti & Shilatifard, 2016). De plus, les promoteurs riches en îlots CpG peuvent prévenir la compaction de la chromatine et favoriser le recrutement de la RNAPII, mais ils peuvent aussi être méthylés et au contraire inhiber la transcription lorsque nécessaire (Cramer, 2019; Deaton & Bird, 2011).

### 1.3.2 L'épissage des pré-ARN messagers

La maturation d'un pré-ARNm consiste en trois étapes importantes : l'ajout d'une coiffe, l'ajout d'une queue poly(A), et l'épissage (Parent & Bisaiillon, 2006). L'ajout de la coiffe en 5' et de la queue poly(A) en 3' du pré-ARNm sont des modifications

importantes, elles permettent notamment de différencier les ARN de type « messenger » des autres types d'ARN (Parent & Bisailon, 2006). Elles sont également essentielles à sa stabilité, à sa reconnaissance et à son exportation hors du noyau (McKendrick et al., 2001; Proudfoot et al., 2002). Au même moment, certaines séquences non-codantes présentes dans le pré-ARNm seront excisées et éliminées pendant le processus d'épissage (Jurica & Moore, 2002; Tari, 2018). Nous savons que les ARN messagers ne sont pas des copies conformes de l'ADN. Dans les régions transcrites du génome, l'ADN est formé d'une alternance entre les séquences informationnelles appelées les Exons (EXpressed regiON) et de séquences silencieuses appelées les Introns (INTragenic RegiON) (Tari, 2018). Uniquement les premières vont former l'ARNm mature, contrairement aux secondes qui seront excisées et éliminées de la séquence finale au cours du processus appelé l'épissage ; plus précisément, le processus qui élimine les introns et joint les exons les uns aux autres est appelé l'épissage. Il concerne au moins 95 % des ARN et s'effectue de façon coordonnée avec la transcription (Wilkinson et al., 2020). Ce processus constitue également un préalable à l'exportation des ARN messagers dans le cytoplasme, siège de la synthèse protéique. De plus, certains ARN dit « non-codants », c'est-à-dire qui ne seront pas traduits en protéines, subissent également l'épissage. Il s'agit, entre autres, des ARN ribosomiques et des ARN de transferts. Bien que tous les précurseurs d'ARN soient organisés de la même manière, tous les introns ne sont pas éliminés de la même manière. Les introns appartenant aux précurseurs d'ARN ribosomiques, d'ARN de transfert, ainsi que les pré-ARNm de mitochondries et de chloroplastes, sont dits « autocatalytiques » (Lambowitz & Zimmerly, 2011; Steitz & Steitz, 1993). En effet, l'élimination de ce type d'introns s'effectue en l'absence de protéines et par une action catalytique propre au pré-ARN (Lambowitz & Zimmerly, 2011). En opposition, les pré-ARNm nécessitent un grand nombre de protéines et d'unités ribonucléoprotéiques pour éliminer les introns. Ces protéines et unités forment un complexe appelé le spliceosome (Brody & Abelson, 1985; Wilkinson et al., 2020).

### 1.3.2.1 Le splicéosome et la réaction d'épissage

Le splicéosome requiert cinq ribonucléoprotéines principales, les snRNP U1, U2, U4, U5 et U6 (*small nuclear ribonucleoproteins*), chacune constituée d'une molécule de ARNsn, complexée avec sept protéines Sm ou apparentées et plus de 150 autres protéines non-snRNP (Kastner et al., 2019; R. Reddy & Busch, 1988; Ule & Blencowe, 2019). Les protéines Sm sont communes à tous les snRNP et s'organisent en anneau heptamérique autour de l'ARNsn en se liant à des séquences spécifiques (Bringmann & Lührmann, 1986; Wilkinson et al., 2020). Les protéines SR (ou SRSF, *serine arginine-rich splicing factor*) sont un autre composant important du splicéosome (Krainer et al., 1991; Zahler et al., 1992). Elles sont au nombre de 12 chez l'homme et interviennent à différentes étapes de l'épissage (Manley & Krainer, 2010). Ces protéines sont nommées ainsi en raison de leur domaine RS situé en C-terminal de leurs séquences. Celui-ci est composé d'un nombre variable des acides aminés sérine et arginine et permet d'établir des interactions avec d'autres domaine SR sur d'autres protéines (Zahler et al., 1992). Ces dernières peuvent être des protéine SR ou non-SR, comme les protéines U2AF35/65 (*U2 auxillary factor* ; 35 et 65 kD) et CAPER qui sont importantes dans la reconnaissance des sites d'épissage (Graveley, 2000; Tari, 2018). Les protéines SR contiennent également d'autres domaines importants, tels qu'un ou plusieurs domaine RRM (*RNA recognition motif*) ou RRMH (*RNA recognition motif homologous*) (Krainer et al., 1991). Ce motif est constitué de deux hélices  $\alpha$  et d'un feuillet  $\beta$  formant un cœur hydrophile et permettant l'interaction directe avec l'ARN (Tari, 2018; Zahler et al., 1992). De plus, seul SRSF7 contient un domaine en doigt de zinc de type CCHC qui lui permet également de se lier directement à l'ARN simple brin (Y. Wang et al., 2021).

Afin d'identifier les jonctions exon-intron, le splicéosome reconnaît trois déterminants sur la séquence intronique du pré-ARN messager : le site donneur d'épissage en 5', le site accepteur d'épissage en 3' et le site (ou point) de branchement 30 nucléotides en



amont du site accepteur (Breathnach et al., 1978). L'étude poussée de ces jonctions exons-introns a révélé des séquences consensus dans ces sites d'épissage conservées au cours de l'évolution (Breathnach et al., 1978). Nous retrouvons les dinucléotides GU pour les sites donneurs et AG pour les sites accepteurs ainsi que le nucléotide A au niveau du site de branchement (Figure 1.12A) (Breathnach et al., 1978; Shapiro & Senapathy, 1987). Néanmoins, dans 1% des cas, nous retrouvons plutôt les dinucléotides AT et AC pour les sites donneurs et accepteurs (Tari, 2018). Ce type de jonction exon-intron est reconnu par un splicéosome différent composé d'un ARNsn U12 au lieu de l'ARNsn U2 habituel (Tari, 2018). L'épissage des pré-ARNm se décompose en deux réactions de transestérification successives (transfert de groupement phosphodiester) (Figure 1.12B) (Padgett et al., 1984; Wahl et al., 2009). D'abord, le résidu Adénosine (A) du site de branchement engage une attaque nucléophile vers un phosphate du site donneur en 5' de l'intron. Cette réaction crée un premier exon avec une extrémité OH libre et un second exon toujours rattaché à l'intron sous forme de lasso (Moore & Sharp, 1993). S'en suit une autre réaction de transestérification consistant en l'attaque nucléophile par l'extrémité OH de l'exon envers le site accepteur en 3' de l'intron (Moore & Sharp, 1993). Ainsi, l'intron est libéré et les deux exons sont réunis. Ces réactions s'effectuent au sein du splicéosome, qui reconnaît les séquences consensus présentes sur les introns et permet le recrutement des ARN et protéines nécessaires aux réactions d'épissage (Wahl et al., 2009; Wilkinson et al., 2020).

### 1.3.2.2 Les étapes d'assemblage du splicéosome

Avant l'assemblage du splicéosome, on retrouve certaines protéines telles que CBP80 ou les protéines SR directement liées à l'ARN formant un complexe nommé le complexe H (Tari, 2018). Les protéines SR interviennent dans la reconnaissance des sites consensus d'épissage sur le pré-ARN messager. Ensuite, l'assemblage du splicéosome s'initie avec la formation du complexe E (ou complexe d'engagement)

(Wilkinson et al., 2020). Celui-ci est composé principalement de l'unité snRNP U1 au niveau du site d'épissage en 5' (Wong et al., 2018; Zhuang & Weiner, 1986).

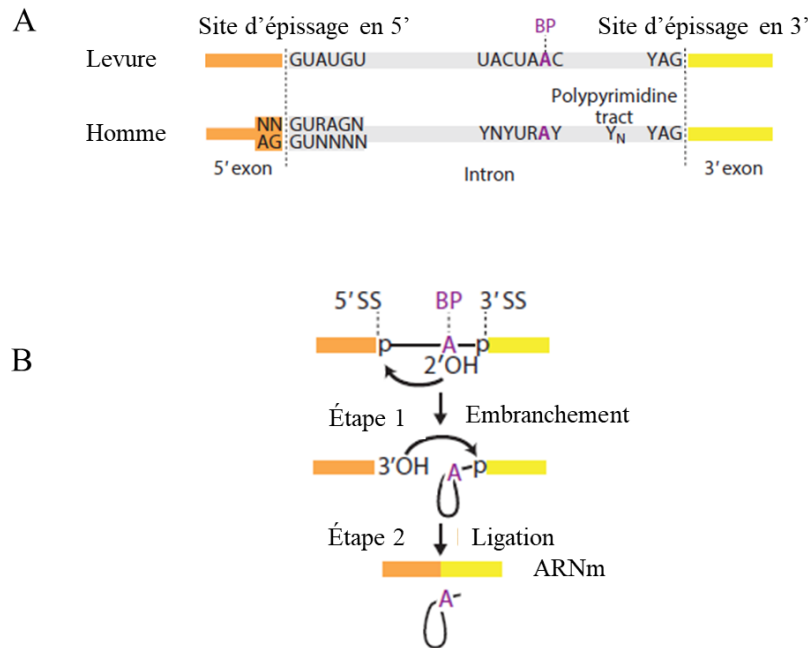


Figure 1.12 La réaction d'épissage. (A) Le splicéosome reconnaît certaines séquences spécifiques au niveau des jonctions intron-exon : le site donneur d'épissage en 5', le site accepteur d'épissage en 3' et le site (ou point) de branchement 30 nucléotides en amont du site accepteur. (B) Les deux réactions de transestérification successives de la réaction d'épissage. Schéma adapté de *Wilkinson et al., 2020*.

En effet, une partie de la séquence de l'ARNsn U1 est très conservée et permet la reconnaissance du site donneur d'épissage (Wilkinson et al., 2020; Wong et al., 2018). Certaines protéines présentes à l'intérieur ou à l'extérieur du complexe vont ensuite renforcer cette reconnaissance par des interactions protéine-protéine, telles que SRSF2 et SF2 (Guiro & O'Reilly, 2015; Muto et al., 2004). Le complexe E est également constitué de la protéine SF1 (*splicing factor 1*) qui reconnaît le site de branchement, et des protéines U2AF65 et 35 qui reconnaissent les sites de branchement et d'épissage en 3' (Berglund et al., 1998; Zamore et al., 1992). Deuxièmement, les protéines auxiliaires U2AF recrutent l'unité snRNP U2 au point de branchement grâce à leur domaine SR et aux interactions protéine-protéine (Kistler & Guthrie, 2001; Wilkinson

et al., 2020). Cette interaction est dépendante de l'ATP, elle déplace le facteur SF1 et modifie la conformation du complexe qui permet le rapprochement des exons de part et d'autre de l'intron. Cette nouvelle conformation du splicéosome est appelée le complexe A. Au sein de ce complexe, la séquence du snRNA U2 se lie spécifiquement au point de branchement (Krämer et al., 1999). De plus, les protéines SF3A, SF3B et PRP5 (*Pre-mRNA-processing ATP-dependent RNA helicase 5*) appartenant à la snRNP U2 vont faciliter et stabiliser l'interaction ARN-ARN entre le snARN et le pré-ARNm (Wilkinson et al., 2020; X. Zhang et al., 2024). Cette liaison permet une modification de la conformation de l'ARN et permet d'isoler l'adénosine (A) du point de branchement hors de la double hélice d'ARN. Par la suite, les unités snRNP U4, U5 et U6, qui se sont d'abord associées entre elles dans le noyau, vont rejoindre le splicéosome pour former le complexe B (Agafonov et al., 2016; Nguyen et al., 2015). Elles sont composées d'un grand nombre de protéines et provoquent un réarrangement structural lorsque snRNP U6 se lie à la fois au site donneur d'épissage en 5' et à snRNP U2 (Will & Lührmann, 2011). L'unité snRNP U6 libère snRNP U1 et prend sa place au site donneur d'épissage en 5' (ce changement marque le passage de pré-complexe B à complexe B). Au même moment, l'unité snRNP U6 se dissocie de snRNP U4 en faveur d'une interaction avec l'unité snRNP U2 (Bringmann et al., 1984; Will & Lührmann, 2011). Lorsque cette nouvelle interaction est formée, le site de liaison de snARN U6 devient actif (Will & Lührmann, 2011). Quant à snRNP U5, elle interagit à la fois avec les deux exons de part et d'autre de l'intron et provoque un rapprochement entre les deux sites : donneur en 5' et accepteur en 3' (Agafonov et al., 2016). Cette conformation conduit au complexe B actif ou complexe B\* (Wahl et al., 2009; Will & Lührmann, 2011). L'ensemble mène à la première réaction de transestérification de l'épissage. Finalement, il y a formation du complexe C également sous forme active où est réalisée la deuxième réaction de transestérification. Les deux exons sont alors joints l'un à l'autre par ligation dans le complexe post-splicéosomale et l'intron est libéré du site actif, ainsi que les snRNP (Wilkinson et al., 2020). De plus, plusieurs hélicases ATP-dépendante jouent un rôle important dans le remodelage de ces

complexes à chaque étape (Bourgeois et al., 2016). Il est indispensable que l'épissage s'effectue avec une précision absolue, puisque l'addition ou la perte d'un seul nucléotide au niveau d'une jonction d'épissage (jonctions introns-exons) pourrait entraîner la traduction erronée de l'ARNm obtenue.

### 1.3.2.3 Épissage alternatif

Il est important de différencier l'épissage dit « constitutif » de l'épissage dit « alternatif ». En effet, les séquences consensus présentes au niveau des sites d'épissage ne sont parfois pas reconnues par les facteurs d'épissage. Ceci mène à des ARNm matures différents selon leur composition en exons, il s'agit alors de l'épissage alternatif (Pan et al., 2008; E. T. Wang et al., 2008). Chez l'homme, environ 90% des gènes subissent de l'épissage alternatif dans un ou plusieurs types cellulaires (Dujardin et al., 2016; E. T. Wang et al., 2008). Il existe plusieurs types d'épissage alternatif notamment l'exclusion d'exon, la rétention d'intron, ou l'utilisation d'un site donneur/accepteur d'épissage alternatif (Sugnet et al., 2004). Ces différences dans l'épissage des pré-ARNm peuvent être tissu-spécifiques. Par exemple, les cellules nerveuses peuvent avoir des besoins différents en transcrits par rapport aux autres cellules pour un même gène (Ule & Blencowe, 2019). Il arrive également que plusieurs ARNm alternatifs soient présents en même temps dans un même type cellulaire mais en quantité variable (Ule & Blencowe, 2019).



### 1.3.3.1 Les séquences et protéines régulatrices

Les sites d'épissage sont souvent désignés comme « forts » ou « faibles » dépendamment d'un épissage constitutif ou d'un épissage alternatif (Sugnet et al., 2004). Pour ces derniers, en plus des séquences consensus nécessaires à l'assemblage du splicéosome, il existe des séquences régulatrices en *cis* (c'est-à-dire une séquence sur le même pré-ARN messenger) qui vont stimuler ou inhiber l'épissage. Ces séquences peuvent se trouver sur les introns et les exons. Les séquences agissant dans le but d'activer l'épissage sont appelées ISE (*Intronic Splicing Enhancer*) ou ESE (*Exonic Splicing Enhancer*). Alors que les séquences agissant dans le but d'inhiber l'épissage sont appelées ISS (*Intronic Splicing Silencer*) ou ESS (*Exonic Splicing Silencer*) (Dujardin et al., 2016).

Plusieurs familles de protéines ont été identifiées pour participer à la régulation de l'épissage alternatif. Il s'agit principalement des protéines SR et des hnRNP (*heterogeneous nuclear ribonucleoprotein*) et plusieurs autres protéines capables de lier l'ARN, appelées RBP (*RNA-binding proteins*) (M. Chen & Manley, 2009). Certaines peuvent être exclusives à un certain type cellulaire, comme les protéines NOVA qui sont spécifiquement enrichies dans les neurones du système nerveux central (Ule et al., 2005). Comme décrit précédemment, les protéines SR font partie du splicéosome et sont importantes pour la reconnaissance des sites d'épissage. Elles sont essentielles pour la régulation de l'épissage alternatif (Z. Zhou & Fu, 2013). En effet, les protéines SR peuvent lier directement les séquences activatrices ESE et ISE afin de promouvoir l'épissage d'un exon en particulier, notamment en interagissant avec les snRNP U1 et les protéines U2AF dans le complexe splicéosomale E (Ule & Blencowe, 2019). De plus, lorsque leur domaine RS est phosphorylé, les protéines SR sont capables de lier l'ARN double brin ce qui favorise l'interaction entre le splicéosome et le pré-ARNm (Shen & Green, 2006). Néanmoins, certaines études décrivent que les protéines SR peuvent aussi réprimer l'épissage si elles lient des séquences introniques

en aval de l'exon en question (Ule & Blencowe, 2019; X. Zhou et al., 2014). Parmi les protéines régulatrices, nous retrouvons également les hnRNP au nombre de 27 (hnRNP A à hnRNP U). Contrairement aux protéines SR, les hnRNP lient préférentiellement les ISS et les ESS sur le pré-ARNm et sont donc considérées comme des répresseurs de l'épissage. Pour ce faire, elles possèdent des domaines d'interaction directs à l'ARN, tels que les domaines RRM, RGG ou KH (*hnRNP K-Homology*). Une fois liées aux séquences régulatrices, les protéines hnRNP empêchent la fixation des constituants du spliceosome ou peuvent séquestrer les sites d'épissage en formant des boucles avec l'ARN (Martinez-Contreras et al., 2006; Ule & Blencowe, 2019). Les hnRNP jouent aussi un rôle dans l'activation de la transcription, le remodelage de chromatine et sur la stabilité et le transport des ARN ainsi que leur transduction. Il peut y avoir des effets de compétition ou des effets coopératifs entre ces deux familles de protéines. Par exemple, si les deux sites régulateurs se trouvent très proches dans l'espace, ce sera le ratio entre les protéines SR et hnRNP qui déterminera quelle effet, activateur ou répresseur, sera dominant. Ce ratio peut alors varier selon le type cellulaire et les conditions physiologiques.

#### 1.3.3.2 Les corps nucléaires

Durant l'interphase, la chromatine est décondensée dans le noyau cellulaire, pourtant les gènes et les séquences régulatrices associées doivent rester facilement accessible aux nombreux facteurs, complexes et interactions selon les besoins de la cellule. C'est pourquoi, il existe plusieurs niveaux d'organisation de la chromatine, comme les territoires chromosomiques et les compartiments transcriptionnellement actifs (A) et inactifs (B). Néanmoins, la chromatine n'occupe pas tout l'espace dans le noyau. En effet, l'espace entre les territoires chromosomiques correspond au compartiment interchromatinien dans lequel sont concentrés les complexes de transcription et de maturation de l'ARN. L'une des manières d'organiser ce compartiment est de regrouper les séquences et les facteurs protéiques dont la fonction cellulaire est

commune à proximité dans l'espace, dans des structures appelées des « granules interchromatiniens » ou encore des « corps nucléaires ». Ainsi, les corps nucléaires sont des organelles liquides sans membrane régulant de multiples fonctions cellulaires importantes. Il existe une grande variété de corps nucléaires, tels que les speckles nucléaires (SN), les nucléoles, les usines à transcription, les *paraspeckles* et les corps de Cajal (Lesne et al., 2019). Chaque type de corps nucléaire possède une ou plusieurs fonctions spécialisées. Par exemple, les nucléoles sont des corps nucléaires spécialisés dans la transcription des ARN ribosomiques (ARNr) et contiennent des concentrations très élevées d'ARNPI et de facteurs de transcription associés (nucléoline). Les NS sont un autre exemple de corps nucléaires, ils sont spécialisés dans la maturation de l'ARNm et sont fortement enrichis en facteurs d'épissage (Lesne et al., 2019). Les granules sont définis par l'accumulation de facteurs protéiques spécifiques et, à l'exception du nucléole qui est visible en microscopie optique, ils sont souvent identifiés à l'aide de techniques de marquage utilisant des anticorps spécifiques, comme l'anticorps SC35 dans le cas des SN.

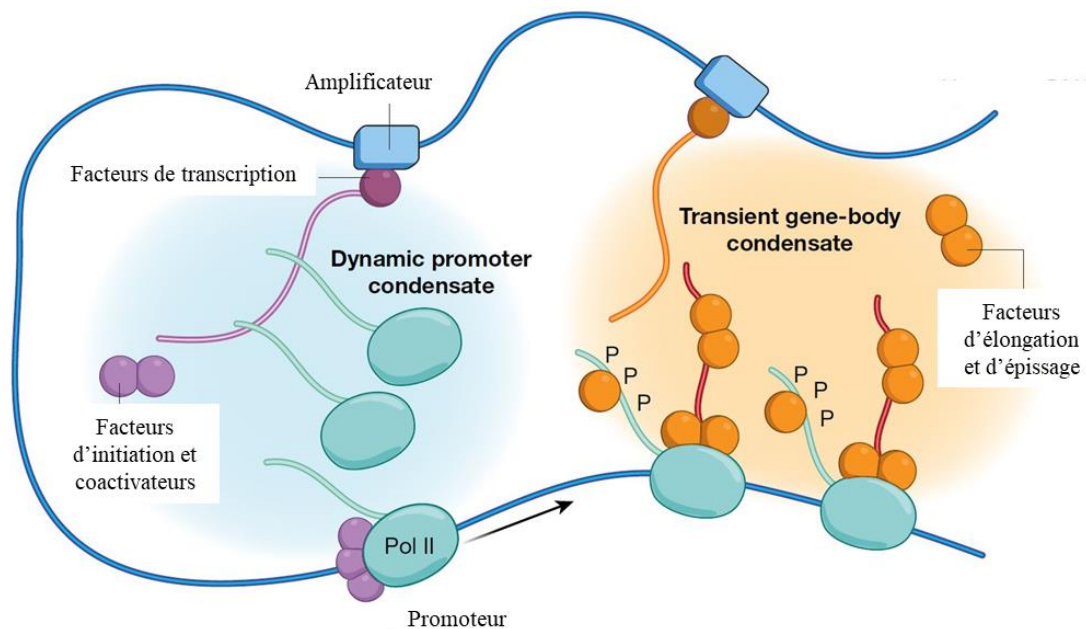




Figure 1.14 Schéma représentant des corps nucléaires au niveau des sites de transcription chez les mammifères. Les corps nucléaires sont enrichies en facteurs de transcription et en facteurs d'épissage. Les étapes d'assemblage du spliceosome chez les mammifères. Schéma adapté de *Cramer et al, 2019*.

Les NS jouent un rôle important dans la régulation de l'épissage alternatif co-transcriptionnelle. En effet, ce sont des condensats fortement enrichis en complexe splicéosomale, de même qu'en facteurs d'épissage. Plusieurs études montrent que les NS agissent comme des sites d'assemblage, de modification et de stockage de ces facteurs proches des sites de transcription (Galganski et al., 2017). Les protéines qui les composent sont également présentes dans le nucléoplasme mais en quantité variable selon la protéine ; de plus, les échanges entre les deux compartiments est constant et très dynamique (Phair & Misteli, 2000). A l'état d'équilibre, la taille et l'aspect des NS résulte de la dynamique des protéines arrivantes et sortantes du corps nucléaire. Les protéines qui composent les SN, au nombre d'environ 360, ont été identifiées par une analyse de spectrométrie de masse (Mintz et al., 1999; Saitoh et al., 2004). Les protéines SR sont les protéines les plus abondantes des SN. Le modèle actuel d'autoassemblage ou de recrutement des protéines au sein de ces corps nucléaires se fait par l'intermédiaire des domaines SR et d'interactions protéine-protéine. Par ailleurs, outre les facteurs d'épissage, des études plus récentes ont montré que les NS se composent également de facteurs de transcription, de plusieurs sous-unités de la RNAPII, d'autres facteurs se liant à l'ADN/ARN et enfin de long ARN non-codants (le plus abondant étant *MALATI*) (Galganski et al., 2017). Ces données démontrent que les NS peuvent également être associés au processus de transcription (Galganski et al., 2017; Ilık & Aktaş, 2021). Par exemple, il a été montré que la proximité spatiale des NS avec les sites de transcription corrèle directement avec une amplification de la transcription (Y. Chen et al., 2018; L. Zhang et al., 2021). La localisation exacte des NS par rapport aux sites de transcription n'est pas encore bien définie, mais les modèles actuels placent l'ARN naissant à la périphérie des NS, alors que les ARN pré-messagers en cours d'épissage sont placés à l'intérieur des NS. De plus, de nombreux facteurs

d'épissage présents au sein des NS comme SRSF1 et SRSF2 sont recrutés par le domaine CTD de la RNAPII et/ou l'ARN naissant et sont importants pour l'élongation de l'ARN (Ilik & Aktaş, 2021; Lin et al., 2008).

Comme mentionné précédemment, les corps nucléaires ne sont pas séparés du nucléoplasme par des membranes, il s'agit plutôt d'une séparation de phase de type liquide-liquide (LLPS; en anglais *liquid-liquid phase separation*). Les protéines intrinsèquement désordonnées, dont la structure tridimensionnelle est flexible, sont essentielles à la formation et au maintien de ces LLPS grâce à leur grande capacité à former des interactions avec d'autres protéines. De plus, les régions intrinsèquement désordonnées (IDR) de ces protéines sont souvent la cible de protéases et de protéines de modification tel que la phosphorylation, ce qui permet de moduler et réguler l'aspect des corps nucléaires. L'équipe *Aktas et al.* a récemment démontré que les protéines désordonnées SRRM2 (*Serine/arginine repetitive matrix protein 2*) et SON sont nécessaires à la formation des NS dans les cellules de mammifères (Ilik et al., 2020; Ilik & Aktaş, 2021). Cette même étude a révélé que l'anticorps SC35, le marqueur le plus utilisé pour identifier les NS et généré initialement à l'aide d'un extrait de splicéosome humain, reconnaît majoritairement la protéine SRRM2 (Ilik et al., 2020). Cette dernière est une large protéine possédant une IDR, et dont la taille varie entre 1 000 et 3 500 acides aminés selon l'espèce. SRRM2 est capable de réguler activement l'épissage alternatif grâce au recrutement de facteurs d'épissage dans les NS et à son rôle dans le complexe splicéosomale B actif (Xu et al., 2022).

### 1.3.3.3 La vitesse d'élongation de l'ARN polymérase II

Nous savons que les mécanismes d'épissage se font de manière co-transcriptionnelle. C'est pourquoi le recrutement des facteurs d'épissage se fait conjointement au recrutement des facteurs d'initiation et d'élongation. Par exemple, la RNAPII est capable de recruter les protéines SR directement sur l'ARN naissant grâce à son

domaine CTD. De plus, la RNAPII effectue l'élongation de l'ARN avec une vitesse variable dépendamment des besoins de la cellule. Cette cinétique joue un rôle important sur la quantité d'ARNm qui sera transcrite pour un gène. Il a également été montré que la vitesse d'élongation de la RNAPII peut également influencer la reconnaissance des exons alternatifs et de ses séquences régulatrices (de la Mata et al., 2003). En effet, une transcription lente par la RNAPII va promouvoir le recrutement du spliceosome et l'inclusion d'un exon dont les sites d'épissage sont faibles, alors qu'une transcription rapide va au contraire tendre à exclure ce même exon (de la Mata et al., 2003; Muniz et al., 2021). Il s'agit du modèle de la « fenêtre d'opportunité » au cours de laquelle une transcription lente laisse une plus grande chance aux séquences régulatrices sur l'ARN naissant d'être reconnues par les facteurs d'épissage ou encore à la RNAPII de recruter les protéines SR. Néanmoins, des études utilisant des versions mutantes de la RNAPII ont montré qu'une vitesse plus lente ou plus rapide de la transcription à l'échelle du génome entier avait les mêmes conséquences sur l'épissage alternatif. En effet, les mutations génèrent toutes deux des inclusions d'exons pour certains gènes ainsi que des exclusions pour d'autres gènes (N. Fong et al., 2014; Muniz et al., 2021). Ces résultats démontrent que malgré un effet de la cinétique de transcription sur l'épissage alternatif, les mécanismes sous-jacents semblent plus complexes qu'un simple modèle de « fenêtre d'opportunité ».

De nombreuses études portent sur les éléments qui régulent la cinétique de la RNAPII. Les facteurs de transcription sont essentiellement en cause dans cette régulation, mais des obstacles sur la chromatine peuvent également induire des pauses lors de l'élongation et ainsi influencer la cinétique de transcription. Parmi ces obstacles, on peut constater la présence de nucléosomes, de protéines liées à l'ADN ou d'erreurs lors de l'intégration des nucléotides (Muniz et al., 2021). En plus de son rôle essentiel dans la régulation de l'état de « pause » de la RNAPII au promoteur, le super complexe d'élongation (SEC; en anglais *Super Elongation Complex*) contenant le facteur P-TEFb influence positivement la vitesse d'élongation (Liang et al., 2018). D'autres facteurs

régulant cette cinétique ont également été identifiés, tels que le facteur Spt5 (*Suppressor of Ty 5* ; appartenant au complexe DSIF) ou le facteur PafC1 (*Polymerase-associated factor 1 Complex*) (Muniz et al., 2021).

#### 1.3.3.4 Les facteurs épigénétiques

L'architecture de la chromatine et les modifications post-traductionnelles des histones sont d'autres acteurs de la régulation de l'épissage alternatif. Il a d'ailleurs été montré que les exons sont plus riches en nucléosomes par rapport aux introns lors de la transcription. Les modifications d'histones peuvent influencer la reconnaissance des sites d'épissage en recrutant, de manière directe ou indirecte, les facteurs d'épissage. Par exemple, la triméthylation de la lysine 4 de l'histone 3 (H3K4me3) interagit avec le remodelleur de la chromatine CHD1 et favorise l'épissage en recrutant le complexe snRNP U2 sur le pré-ARNm (Sims et al., 2007; Tai et al., 2003). Le niveau de compaction de la chromatine peut également influencer la reconnaissance des exons alternatifs en régulant la vitesse d'élongation de la RNAPII (Muniz et al., 2021). Des études ont montré qu'une augmentation du niveau de di-méthylation de la lysine 9 de l'histone 3 (H3K9Me2), de même que le niveau de triméthylation de la lysine 27 (H3K27me3) corrèle avec une baisse de la vitesse de la transcription par la RNAPII et davantage d'inclusion d'exons (Schor et al., 2013). La protéine HP1 $\gamma$  (*heterochromatin protein 1 isoform  $\gamma$* ) qui est un facteur d'hétérochromatine recruté par les di-méthylation et triméthylation de H3K9, joue un rôle dans cette compaction (Saint-André et al., 2011). En effet, parmi les trois isoformes des protéines HP1 (HP1 $\alpha$ , HP1 $\beta$  et HP1 $\gamma$ ), HP1 $\gamma$  est le seul présent dans l'euchromatine, guidé par AGO2 et ses microARN au niveau d'exons alternatifs afin de ralentir l'ARNm et de promouvoir l'inclusion des exons (Figure 1.9) (Ameyar-Zazoua et al., 2012; Saint-André et al., 2011). De plus, la méthylation de l'ADN ainsi que l'acétylation des histones, modulée par les HDAC, peuvent aussi influencer l'épissage alternatif en agissant sur la vitesse d'élongation de la RNAPII (Muniz et al., 2021).

#### 1.4 Objectifs du projet de recherche

Le syndrome CHARGE est une maladie génétique rare qui touche environ 1 naissance sur 10 000 et dont la cause principale est une mutation du gène CHD7. Cette maladie fait partie de la grande famille des neurocristopathies, soit des maladies ayant des problèmes avec les cellules de la crête neurale, mais les mécanismes moléculaires impliqués demeurent méconnus. Le laboratoire du Pr Nicolas Pilon a récemment mis en évidence le gène *FAM172A* comme étant un nouveau gène candidat pour ce syndrome. Les premières études sur la protéine FAM172A ont révélé qu'elle appartenait à un complexe à l'interface entre la chromatine et le spliceosome où elle régulerait certains événements clés d'épissage alternatif co-transcriptionnel, et ce avec les protéines CHD7 et AGO2 (Bélanger et al., 2018). De plus, les défauts d'épissage alternatif semblent être une cause commune à plusieurs cas du syndrome CHARGE quel que soit le gène mis en cause (Bélanger et al., 2018; Bérubé-Simard & Pilon, 2019). L'objectif général du projet est de mieux comprendre les rôles que pourrait jouer FAM172A dans ces mécanismes, et ultimement mieux comprendre les mécanismes moléculaires mis en cause dans le syndrome CHARGE.

Afin de répondre à cet objectif, nous nous concentrerons d'abord sur le complexe formé de FAM172A et CHD7 à la chromatine. Le rôle de CHD7 dans la régulation de l'expression de gènes associés au développement à travers le remodelage de la chromatine est maintenant bien connu. En revanche, son implication dans la régulation de l'épissage alternatif reste encore mal définie. Étudier le fonctionnement de la protéine FAM172A à la chromatine, un autre gène causatif du syndrome CHARGE, pourrait nous éclairer davantage sur l'implication de CHD7 dans les mécanismes d'épissage alternatif co-transcriptionnel. Les objectifs de cette première partie sont :

- Localiser les protéines FAM172A et CHD7 sur l'ensemble du génome.
- Identifier les cibles directes de FAM172A et CHD7.

- Étudier le lien entre ces protéines et la régulation de l'épissage alternatif co-transcriptionnel.

Nous nous sommes ensuite intéressés à l'interaction entre les protéines FAM172A et AGO2. Cette interaction est déstabilisée lorsqu'une version de FAM172A mimant un variant présent chez des patients CHARGE est utilisée, ce qui rend cette interaction pertinente à étudier dans un contexte pathologique. Connaissant le mécanisme d'action d'AGO2 dans la régulation de l'épissage alternatif de certains gènes (Ameyar-Zazoua et al., 2012) et ayant observé les mêmes dérégulations chez le modèle *Toupee* (les gènes *Cd44*, *Col5a3*, *Mical2* et *Ift74*) (Bélanger et al., 2018), nous pourrions envisager que FAM172A agit sur la régulation de l'épissage alternatif à travers AGO2. À ce jour, la protéine FAM172A ne semble pas posséder de domaine de liaison à l'ADN ou à l'ARN. En revanche, des outils bio-informatiques lui prédisent un NLS qui lui permettrait d'être reconnu par des importines et d'être transloqué au noyau. De plus, les facteurs impliqués dans l'import nucléaire d'AGO2 sont encore inconnus, celle-ci étant très dynamique et voyageant constamment entre le cytoplasme et le noyau. L'une de nos hypothèses serait que FAM172A participe activement à l'import d'AGO2 au noyau. Les objectifs de cette partie sont :

- Tester l'interaction directe entre FAM172A et AGO2.
- Déterminer l'implication de FAM172A dans l'import nucléaire d'AGO2.
- Étudier le rôle régulateur de la phosphorylation de FAM172A.

## CHAPITRE II

## ARTICLE I

CHARGE SYNDROME-ASSOCIATED PROTEINS FAM172A AND CHD7 ARE  
RECRUITED AT ACTIVE PROMOTER REGIONS AND PARTICIPATE TO  
SPLICING CONDENSATES ORGANIZATION IN NEURAL CREST CELLS

Sephora Sallis<sup>1,2</sup>, Félix-Antoine Bérubé-Simard<sup>1,4</sup>, Sandrine Girard<sup>1,2</sup>, Elizabeth  
Leduc<sup>1,2</sup> and Nicolas Pilon<sup>1,2,3</sup>.

<sup>1</sup> Molecular Genetics of Development Laboratory, Department of Biological Sciences,  
Université du Québec à Montréal, Montreal H3C 3P8, Quebec, Canada.

<sup>2</sup> Centre d'Excellence en Recherche sur les Maladies Orphelines – Fondation Courtois,  
Université du Québec à Montréal, Montreal H2X 3Y7, Canada.

<sup>3</sup> Department of Pediatrics, Université de Montréal, Montreal H3T 1C5, Canada.

<sup>4</sup> Present address : Biopterre, Sainte-Anne-de-la-Pocatière, QC, Canada.

Manuscrit en préparation pour soumission à la revue *EMBO journal*

### Contributions :

Sephora Sallis: Design expérimental, expérimentation/collection des données, analyses bioinformatiques, figures, recherche bibliographique et rédaction.

Félix-Antoine Bérubé-Simard: Expérimentation et collection des données (RNA-Seq Chd7Gt/+).

Sandrine Girard: Expérimentation et collection des données (génotypage et aide immunofluorescence).

Elizabeth Leduc: Expérimentation (préparation des tissus).

Nicolas Pilon: Conception du projet, supervision, recherche bibliographique et rédaction.



## 2.1 Résumé

Le syndrome CHARGE est une maladie rare associée au développement des cellules de la crête neurale (CCN) et principalement causée par une mutation dans le gène *CHD7* codant pour un remodeleur de la chromatine. Une mutation dans le gène *FAM172A* est une autre cause menant à ce syndrome. Les études effectuées précédemment sur des modèles de souris et sur des cellules dérivées de patients CHARGE, ont permis de proposer que les protéines FAM172A et CHD7 coopèrent à l'interface chromatine-splicéosome afin de réguler la transcription et l'épissage alternatif de plusieurs gènes importants pour les CCN. Dans le but de tester ce modèle, nous avons utilisé la méthode de CUT&RUN pour déterminer leur localisation génomique au sein de CCN sauvages. Nous avons trouvé que 95% des régions occupées par FAM172A le sont aussi pour CHD7 et que ces régions sont principalement des promoteurs de gènes actifs (H3K4me3+). En croisant ces données avec des données de transcriptome provenant de CCN murins mutées pour *Fam172a* et *Chd7*, nous avons mis en évidence une meilleure corrélation entre les cibles directes communes de FAM172A et CHD7 et les défauts d'épissage alternatif. Finalement, nous avons identifié une désorganisation des corps nucléaires associés à l'épissage au sein de tissus embryonnaires *in vivo* mutés pour les deux gènes. Ces résultats suggèrent que FAM172A et CHD7 participent à la régulation de l'épissage alternatif à travers l'organisation et/ou le recrutement de facteurs d'épissage aux promoteurs.

Mots-clés : Syndrome CHARGE ; cellules de la crête neurale ; épissage alternatif ; transcription ; FAM172A ; CHD7

## 2.2 Abstract

CHARGE syndrome is a rare congenital disorder of neural crest cell (NCC) development mainly caused by mutation of the chromatin remodeler-coding gene *CHD7*. Mutation of the co-transcriptional regulator-coding gene *FAM172A* is another cause. Based on our work with mouse models and human cells from CHARGE patients, we previously proposed that *FAM172A* and *CHD7* cooperate at the chromatin-spliceosome interface for activating the transcription and alternative splicing of many genes important for NCCs. To further test this model and better understand how these two proteins work together, we decided to assess *FAM172A* and *CHD7* genomic distribution specifically in NCCs. Using CUT&RUN, we first found that 95% of the chromatin regions occupied by *FAM172A* are also occupied by *CHD7*, and that the vast majority of these sites correspond to active promoter regions (H3K4m3+) in WT NCCs. Then we intersected these genomic data with transcriptomic data and found a better correlation between co-bound genes and misspliced genes in *Fam172a*- and *Chd7*-mutant NCCs. Finally, we uncovered *in vivo* disorganization of splicing-associated condensates in NCCs and non-NCCs from both mutants. Altogether, these results suggest that *FAM172A* and *CHD7* can participate to alternative splicing regulation through splicing factors organization and/or recruitment.

Keywords: CHARGE syndrome, neural crest cells, Alternative splicing, transcription, *FAM172A*, *CHD7*

### 2.3 Introduction

Neural Crest Cells (NCCs) are transient multipotent progenitor cells that arise from the neural crest in vertebrates and migrate to various regions of the embryo in order to generate very distinct cell types (Sato et al., 2019; Takahashi et al., 2013). Abnormal expression, migration, differentiation, or death of NCC during embryonic development is associated with multiple disorders called neurocristopathies (Etchevers et al., 2006). CHARGE syndrome is a rare neurocristopathy that affects ~ 1/10 000 newborns worldwide (Hsu et al., 2014) and is characterized by complex clinical presentation that mainly impacts cranial and cardiac NCC derivatives (Pauli et al., 2017; Soldatov et al., 2019). The acronym CHARGE stands for the commonly associated phenotypes: Coloboma of the eye, Hear defects, Atresia of choanae, Retardation of growth/development, Genital abnormalities, and Ear anomalies.

Heterozygous mutation of *CHD7* (*Chromodomain Helicase DNA Binding Protein-7*) is the main cause of CHARGE syndrome (Vissers et al., 2004b). *CHD7* encodes a member of the CHD family, a group of ATP-dependent chromatin-remodeling factors that alter chromatin organization by rearranging the position of nucleosomes. Previous studies demonstrated that *CHD7* is necessary for the differentiation of multiple cell lines and tissues derived from NCCs (Bajpai et al., 2010; Chai et al., 2018; Feng et al., 2017; J. Gao et al., 2024; Schnetz et al., 2009b, 2010), among them, human NCCs generated from CHARGE syndrome patients iPSCs harboring a *CHD7* mutation (Okuno et al., 2017). *CHD7* is known to regulate expression levels of genes important for NCC development by modulating access to enhancer and promoter regions (Feng et al., 2017; Schnetz et al., 2009b). More recently, *CHD7* has also been implicated in alternative splicing regulation, through interaction with the U2 snRNP (*U2 small nuclear ribonucleoprotein*) component of the spliceosome and pre-mRNA (Allemand et al., 2016; Bélanger et al., 2022; G Hendrickson et al., 2016). *CHD7* can collaborate

with other chromatin factors such as PARP1 (*poly(ADP ribose) polymerase 1*) and several components of the SWI/SNF (*SWItch/Sucrose Non-Fermentable*) complex (Bajpai et al., 2010; Feng et al., 2017), both also known to regulate co-transcriptional alternative splicing by recruiting or interacting with splicing factors (Batsché et al., 2006; Matveeva et al., 2016). However, the mechanism by which CHD7 regulates alternative splicing remains unclear.

Rare variants in many other genes encoding chromatin and/or splicing factors have also been recently associated with CHARGE syndrome, such as *PUF60*, *EP300*, *RERE*, *KMT2D* and *KDM6A*. One of these additional player is *FAM172A* (Bélanger et al., 2018; Moccia et al., 2018a). Taking advantage of a transgenic mouse model called *Toupee* (*Fam172a<sup>Tp/Tp</sup>*), we previously validated the candidacy of *FAM172A* (*Family with Sequence Similarity 172 Member A*) as new cause for CHARGE syndrome (Bélanger et al., 2018). In this model, mutagenic transgene sequences are inserted in the last intron of *Fam172a* generating a hypomorphic allele that negatively affects almost every aspect of NCC ontology and recapitulates features of CHARGE syndrome (Bélanger et al., 2018; Hale et al., 2016). In addition, the *Toupee* allele was found to genetically interact with a gene-trap mutant allele of *Chd7* (*Chd7<sup>Gt</sup>*) (Bélanger et al., 2018; Hurd et al., 2007). Interestingly, *FAM172A* contains a large ARB2 (*Argonaute binding protein-2*) domain originally described in yeast (Buker et al., 2007). In *S. cerevisiae*, the ARB2 domain can bind directly to nucleosomes and is recruited preferentially at active gene regions by interacting with the RNA polymerase II (Ha et al., 2019; Shen et al., 2016). Moreover, our previous mechanistic studies allowed us to propose a model whereby *FAM172A* helps to stabilize the chromatin-spliceosome interface, notably by interacting with CHD7 in order to regulate transcription and alternative splicing of many genes important for NCC development (Bélanger et al., 2018; Bérubé-Simard & Pilon, 2019). These studies also suggested that perturbation of chromatin-mediated alternative splicing is a general pathological mechanism for CHARGE syndrome, regardless of the involved gene defect (Bélanger et al., 2018,

2022; Bérubé-Simard & Pilon, 2019). Further investigation of FAM172A and CHD7 roles in this molecular process might give us a better understanding of the pathological mechanisms leading to CHARGE syndrome.

To better understand how these two proteins work together, we first analyzed the genomic distribution of FAM172A and CHD7 in FACS-recovered NCCs from e10.5 embryos and intersected these data with transcriptomic data from corresponding mutants. In WT NCCs, FAM172A is recruited at CHD7-occupied active promoter regions (H3K4me3+) where they participate to alternative splicing regulation. In addition, both proteins are implicated in splicing factors organization into nuclear speckles *in vivo*.

## 2.4 Materials and Methods

### Animals

All experimental procedures involving mice were approved by the institutional ethics committee of the University of Quebec at Montreal (CIPA #650) in accordance with the biomedical research guidelines of the Canadian Council of Animal Care. The *Fam172a<sup>Tp/+</sup>* (FVB/N background) and the *Gata4p*-RFP (FVB/N background) transgenic mouse lines were generated as previously described (Bélanger et al., 2018; Pilon et al., 2008). The *Chd7<sup>Gt/+</sup>* transgenic mouse line was obtained from Donna Martin (*Department of Pediatrics, The University of Michigan*). Embryos were all staged by considering noon of the day of plug detection as embryonic day (e) 0.5 and were generated by natural mating. Control and mutant embryos were obtained from the same litter and processed in parallel.

### FACS and nuclei isolation

NCCs were obtained from E10.5 *Gata4p*-RFP embryos dissociated in EMEM containing 1.3 mg/ml Dispase II (Life Technologies #17105-041), 0.4 mg/ml collagenase (Sigma #C2674) and 0.1 mg/ml DNase I (Sigma #DN25), for 30 min at 37°C with gentle agitation. The resulting cell suspension was filtered through a 70µm cell strainer (Fisherbrand) and subjected to FACS (BD FACSJazz cell sorter) (*Fluorescence-activated cell sorting*) to recover specifically fluorescent NCCs. Cells were kept on ice during all steps. For CUT&RUN experiment, cells were centrifuged (600g, 5min, 4°C) and resuspended in a nuclear extraction buffer (20mM HEPES pH8, 10mM KCl, 0.1% triton X-100, 20% glycerol, 0.5mM Spermidine, 1X Roche complete protease inhibitors). After 10min-incubation on ice, nuclei were isolated by

centrifugation (600g, 5min, 4°C), resuspended in nuclear extraction buffer and stored at -80°C.

## CUT&RUN

CUT&RUN experiments were performed using CUTANA™ ChIC/CUT&RUN kit (Epiccypher) in accordance with the manufacturer's protocol. Briefly, 100 000 nuclei (in nuclear extraction buffer: 20mM HEPES pH8, 10mM KCl, 0.1% triton X-100, 20% glycerol, 0.5mM Spermidine, 1X Roche complete protease inhibitors) were directly incubated for 20 min at room temperature with ConA beads, and the supernatant was removed using a magnetic rack. Then, bound-nuclei were incubated overnight at 4°C with 1µg of primary antibodies: anti-FAM172A, anti-CHD7, anti-H3K4m3, anti-H3K27m3, in bead activation buffer. Nuclei were homogenized frequently to avoid clumping. After incubation, bound-nuclei were washed twice and incubated for 20 min at room temperature with pAG-MNase and gentle agitation. Unbound pAG-MNase were removed and nuclei were washed twice before a 2h-incubation at 4°C with 100mM calcium chloride to active enzymes. All incubations were performed with gentle agitation on nutator and a 30° inclination angle. Finally, the digestion was halted by the addition of Stop buffer and nuclei were incubated for 30 min at 37°C to facilitate the release of digested DNA fragments. The supernatant was collected, and DNA was extracted using DNA cleanup columns from the CUTANA™ kit. CUT&RUN libraries were then generated using the NEBNext Ultra DNA library Prep Kit Illumina (NEB E7645, index E6440S) in accordance with the manufacturer's protocol. Finally, libraries were sequenced using NovaSeq6000 (Illumina) (100pb, paired-end, 10 million reads per sample) at CHU de Québec-Université Laval.

## ATAC-seq

50 000 nuclei were incubated for 30 min at 37°C with Tn5 transposase-loaded (Diagenode, C01070012) diluted in Tagmentation buffer (Diagenode, C01019043) in Thermomixer with gentle agitation. Then fragmented DNA was purified using DNA cleanup columns from CUTANA™ kit and libraries were performed using reagents from NEBNext Ultra DNA library Prep Kit Illumina (NEB E7645, index Nextera) in accordance with the manufacturer's protocol. ATAC-seq libraries were sequence at CHU de Québec-Université Laval with NovaSeq6000 (Illumina) (100pb, paired-end, 50 million reads per sample).

## Bioinformatic analysis

Quality of CUT&RUN and ATAC-seq reads were assessed with FASTQC and adapters were removed using Trimmomatic. For CUT&RUN, reads were aligned to Mouse reference genome GRCm38/mm10 using Bowtie2 (Min:10, Max:700, -very sensitive-, -end-to-end-, -dovetail-) and visualized using Easseq (genome browser and heatmaps). Peaks were called using SEARC (Meers, Tenenbaum, et al., 2019). To eliminate any remaining false positive, Easseq were used to quantify CUT&RUN enrichment at those peaks ( $\pm$  500pb, normalized to reads per million) and peaks below threshold (1.5) were discarded (or over threshold for IgG samples). Only peaks identified in both biological replicates (N=2) using SEARC were selected. For the extended list of FAM172A and CHD7, peaks called using SEARC and MACS2 were added and proceeded as described above. Then annotation of peaks and gene ontology analysis were performed using HOMER et Enrichr algorithms (Kuleshov et al., 2016), respectively. For ATAC-seq, reads were aligned to Mouse reference genome GRCm38/mm10 using Bowtie2.



## RNA-seq

RNA extraction from FACS-recovered E10.5 *Chd7<sup>Gt/+</sup>* NCCs was performed using the RNeasy Plus Purification kit (Qiagen) in accordance with the manufacturer's protocol. Ribosomal RNA-depleted libraries were generated and sequenced (100pb, paired-end, 50 million reads) at Genome Québec Innovation Centre.

## Immunofluorescence

For cryosections, embryos were harvested at E10.5 and fixed in 4% paraformaldehyde for 2 hours, washed three times in PBS, incubated overnight in 30% sucrose solution (0.1 M phosphate buffer, pH 7.4) and then embedded in OCT compound (Fisher Healthcare 4585). Cross-sections of 15-  $\mu\text{m}$  were prepared using a Leica CM1950 cryostat (Leica microsystems) and stored at  $-80^{\circ}\text{C}$ . For immunofluorescence, sections were washed three times in PBS, incubated in target retrieval solution (DAKO, pH 6) for 20 min at  $95^{\circ}\text{C}$ , and washed again three times. Then, sections were incubated in blocking solution (10% FBS, 0.5% triton X-100 diluted in PBS) for 1h at room temperature, followed by overnight incubation at  $4^{\circ}\text{C}$  with primary antibodies: Rat anti-SOX10 (1:200, mébimabs), Mouse anti-SC35 (1:200, abcam #11826) and Rabbit anti-SON (1:200, Sigma #HPA023535). After three washes in blocking solution, sections were incubated for 1h at room temperature with secondary antibody: Alexa Fluor 488 Anti-Rat IgG (1:500; Jackson ImmunoResearch 712-545-150), Alexa Fluor 594 Anti-Rabbit IgG (1:500; Jackson ImmunoResearch 711-585-152), Alexa Fluor 647 Anti-Mouse IgG (1:500; Jackson ImmunoResearch 715-605-150). Sections were washed again with PBS and stained for 10min at room temperature with DAPI diluted in PBS ( $5\mu\text{g/ml}$ ; 40,6-diamidino-2-phenylindole). Immunofluorescence acquisitions were performed using a Nikon A1 laser scanning confocal microscope and analyzed using ImageJ. Nuclear speckles features (mean fluorescence intensity, number and size)

were quantified specifically in nuclei using ROI manager, the *cell counter* and *measure* functions.

### Statistical analysis

Graphical data are represented as the mean  $\pm$  SD, with the number of independent experiments (N) and/or independent biological replicates (n) specified in the figure and/or legend when relevant. Significant differences between samples were determined using the GraphPad Prism software version 6.0, with 0.01 as statistical significance cut-off. Selected statistical tests are indicated in figure legends.

### Cell culture et transfection

Neuro2A neuroblastoma cells were grown in EMEM medium supplemented with either 10% FBS and 1% penicillin/streptomycin (all products from Wisent). Cells were maintained at 37 °C and 5% CO<sub>2</sub> in a humidified atmosphere. Transfections were performed using GeneJuice reagents (Millipore-Sigma #70967) in accordance with the manufacturer's instructions.

### Co-immunoprecipitation

Neuro2A whole cell extract was obtained from a confluent 100mm-plate resuspended in lysis buffer (20 mM Tris pH 8.0, 25mM NaCl, 1 mM EGTA, 1.5 mM MgCl<sub>2</sub>, 1 mM DTT, 1% Triton TX-100, 10% Glycerol, 1X Roche complete protease inhibitors) and incubated for 1h at 4°C. Extracts were sonicated on ice (2 x 20sec, 20% amplitude) and incubated overnight at 4°C with primary antibody: Mouse anti-MYC (1:500, abcam #9132), Mouse anti-SC35 (1:500, abcam #11826). Antibody-bound complexes were incubated for 1h at room temperature with protein G-coupled magnetic beads (Dynabeads protein G, ThermoFisher), washed three times and resuspended in lysis

buffer supplemented with either: DNase, RNase, both or lysis buffer for control samples. Complexes were then incubated for 30 min at 37°C, washed three times and analyzed by western blot.

#### Western blot

Co-immunoprecipitated complexes were separated on SDS-PAGE (Bio-Rad Laboratories) and transferred on PVDF membranes. After 1h-incubation at room temperature in blocking solution TBST (Tris-buffered saline-Tween 20, 5% non-fat milk), membranes were incubated overnight at 4°C with primary anti-bodies : Mouse anti-MYC (1:500, abcam #9132), Mouse anti-SC35 (1:500, abcam #11826) and washed three times. Finally, membranes were incubated for 1h at room temperature with secondary Anti-Mouse IgG HRP (1:2000, Cell signaling 7076). Membrane detection was performed using Immobilon Western HRP substrate (Millipore-Sigma) and captured using Fusion FX imaging system (Vilber).

## 2.5 Results

### 2.5.1 FAM172A is recruited at CHD7-occupied promoters in NCCs

To investigate FAM172A and CHD7 role at the chromatin-spliceosome interface, we analyzed their genomic distribution in e10.5 WT NCCs collected by FACS owing to the G4-RFP reporter (Pilon et al., 2008).

Using the CUT&RUN method followed by NGS (Skene & Henikoff, 2017), we first identified 3 242 peaks enriched with FAM172A. Among them, 2 598 (80 %) peaks were localized around Transcription Start Sites (TSS) (Figures 2.1A, D and Fig 2.2A). We also identified 15 035 peaks enriched with CHD7 and only 7 036 (46.8 %) of them were localized around TSS (Figure 2.1B, E). Remarkably, we found that the vast majority (95 %, 3 068 of 3 242) of chromatin regions occupied by FAM172A in NCCs are co-occupied by CHD7 (Figure 2.1C). These shared regions were predominantly associated with TSS/promoters (81.5 %) (Figure 2.1F), in contrast to regions only occupied by CHD7 (Figure 2.1E). We assume the large part of intergenic and intronic regions (47.4 %) bound specifically by CHD7 is associated with enhancer elements, consistent with existing literatures (Schnetz et al., 2009a; Stathopoulou et al., 2023) (Figure 2.1E). As expected, Gene Ontology (GO) analysis of genes co-occupied by FAM172A and CHD7 shows an enrichment in phenotypes associated with CHARGE syndrome, such as embryonic growth retardation, abnormal tissue morphology and cleft palate (Figure 2.1G). It also shows an enrichment in several biological processes such as chromatin remodeling, regulation of transcription and DNA damage repair (Figure 2.2B). A representative example of the CUT&RUN data is shown in Fig 2.1H where several peaks were identified at *Rara* (*retinoic acid receptor  $\alpha$* ) locus, coding for an important factor for craniofacial differentiation. A shared peak is localized at the TSS (yellow area), while a CHD7-specific peak is localized 1.5kb upstream the TSS (grey area) (Figure 2.1H). Therefore, FAM172A is mainly recruited at CHD7-occupied

promoters of genes related to NCC development. These results also suggest a divided role for CHD7 in NCCs: (i) with FAM172A at TSS, (ii) without FAM172A at other regulatory elements, most likely enhancers.

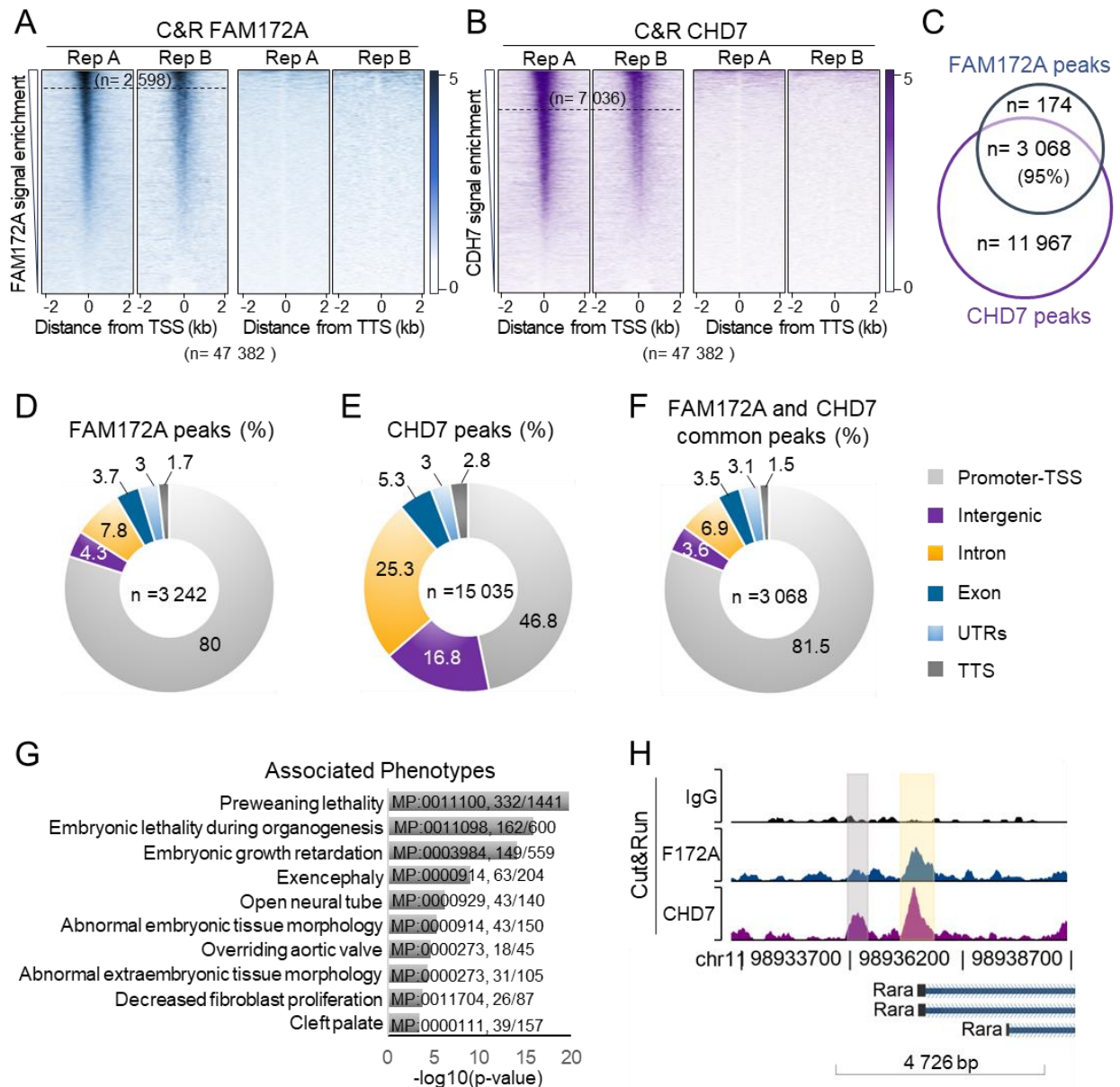


Figure 2.1 FAM172A is recruited at CHD7-occupied promoters. A, B) Heatmaps of CUT&RUN alignment data for (A) FAM172A and (B) CHD7 in WT NCCs recovered by FACS from e10.5 G4-RFP embryos. Peak signals are centred ( $\pm 2$ kb) on Transcription Start Sites (TSS) or Transcription Termination Sites (TTS). n=transcripts. C) Overlap between FAM172A and CHD7 peaks. Only peaks found in all biological replicates using SEARC are represented. D-F) Relative proportions of HOMER-annotated peaks for (D) FAM172A-, (E) CHD7- occupied regions and (F) FAM172A/CHD7 co-occupied regions. n=peaks. G) Gene ontology analysis showing enriched associated mammalian phenotypes for FAM172A/CHD7 common peaks. H) Genomic browser view of *Rara* locus using Easseq, with CUT&RUN data for IgG (control), FAM172A and CHD7 in WT NCCs. Grey area shows a specific peak to CHD7, yellow area shows a common peak to FAM172A and CHD7.

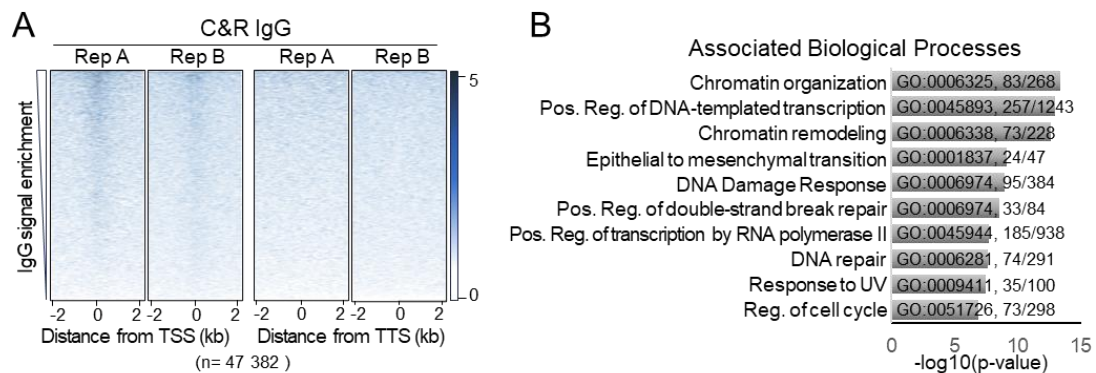


Figure 2.2 (SUPPLEMENTAL) FAM172A is recruited at CHD7-occupied promoters. A) Heatmaps of CUT&RUN alignment data for IgG in WT NCCs recovered by FACS from e10.5 G4-RFP embryos. Peak signals are centred ( $\pm 2$ kb) on Transcription Start Sites (TSS) or Transcription Termination Sites (TTS). n=genes. B) Gene ontology analysis showing enriched biological processes for FAM172A/CHD7 common peaks.

### 2.5.2 FAM172A and CHD7 co-occupied promoters are enriched in H3K4me3

Focusing on FAM172A and CHD7 shared role at TSS/promoters, we explored the transcriptional state of the associated genes by verifying the status of trimethylation of histone 3- Lysine 4 (H3K4me3, active transcription) and Lysine 27 (H3K27me3, repressed transcription).

To this end, CUT&RUN experiments were performed on WT NCCs targeting these histone marks. We identified 6 367 peaks enriched with H3K4 trimethylation and 24 281 peaks with H3K27 trimethylation (Figure 2.4A, B). Peak annotation localized 6 239 (98 %) and 3 091 (12.7 %) of these peaks around TSS, respectively (Figure 2.4A, B). Peaks enriched with FAM172A, CHD7 and at least one of histone marks were selected (n= 2 942) (Figure 2.3A). Then we quantified CUT&RUN signal levels using a ratio  $\log_2[\text{H3K4me3}/\text{H3K27me3}]$  to separate these peaks into transcriptionally active-, poised- (bivalent, with both marks) or repressed promoters (Figure 2.3B). We found that promoters co-occupied by FAM172A and CHD7 are mainly either transcriptionally active (87.3 % of shared peaks; 2 568 of 2 942; ratio  $\geq 1$ ) or poised for transcription (10 % of shared peaks; 294 of 2 942;  $-1 < \text{ratio} < 1$ ) (Figure 2.3C). To confirm this result, we decided to compare the transcriptional state distribution of FAM172A/CHD7-co-bound promoters to a set of promoters from Ensembl database (Dreos et al., 2017; Martin et al., 2023). Similarly, from this new set we selected only promoters enriched with at least one of histone marks (n= 8 701) (Figure 2.3D), and we quantified the same CUT&RUN ratio  $\log_2[\text{H3K4me3}/\text{H3K27me3}]$  (Figure 2.3E). This distribution is considered the standard distribution of active (52.9 %; 4 601 of 8 701), poised (23,8 %; 2 069 of 8 701), and repressed (23.3 %; 2 031 of 8 701) promoters in NCCs, and it is significantly different from the FAM172A/CHD7-co-bound promoter distribution (Figure 2.3F). Moreover, we assessed the chromatin accessibility at these regions using ATAC-seq method and confirmed a strong association between FAM172A, CHD7 and open chromatin (Figure 2.3A).

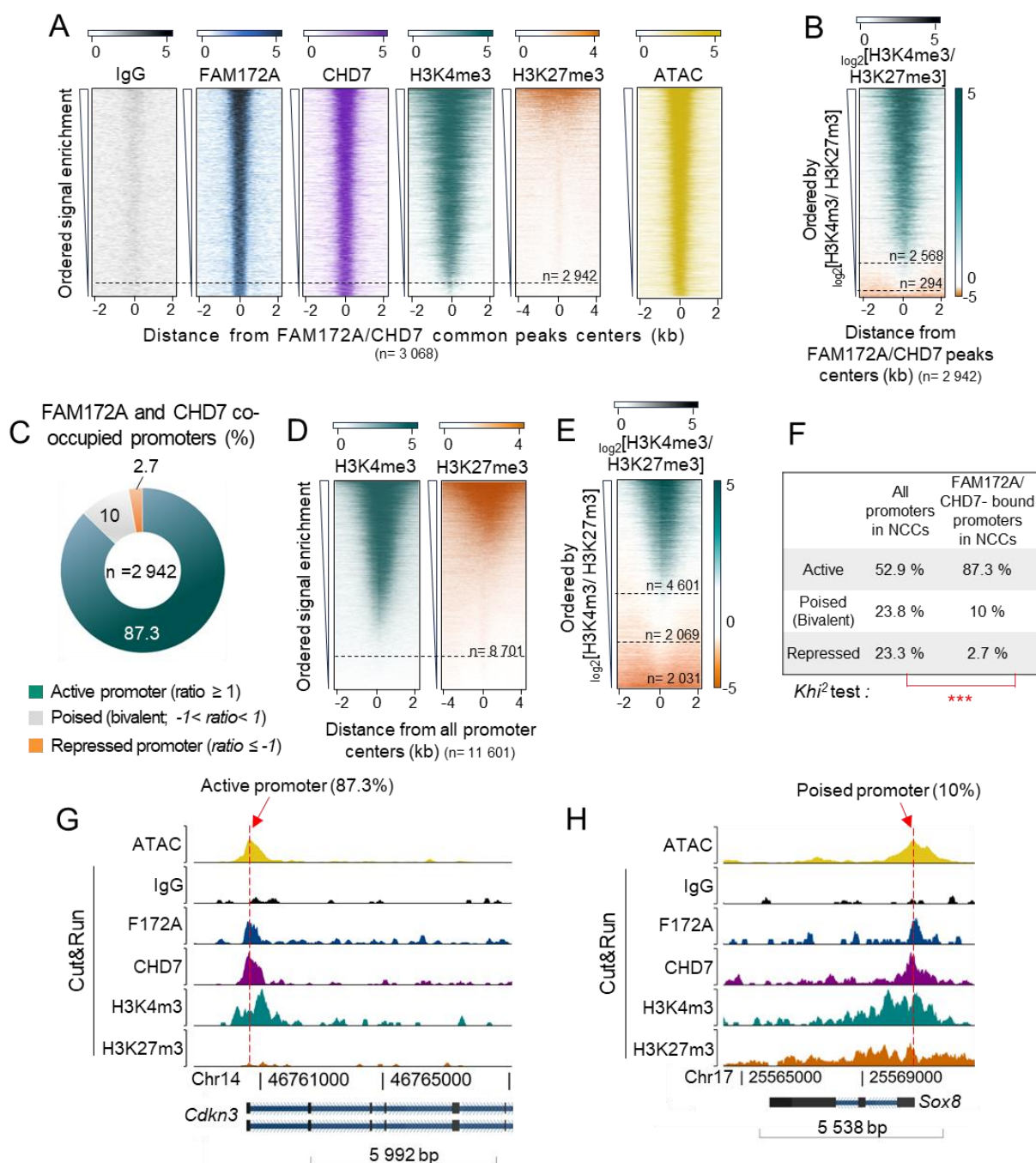


Figure 2.3 FAM172A and CHD7 co-occupied promoters are enriched in trimethylated H3K4. A, D) Heatmaps of CUT&RUN alignment data for IgG (control), FAM172A, CHD7, H3K4Me3 and H3K27me3 in WT NCCs recovered by FACS from e10.5 G4-RFP embryos. Peak signals are centred ( $\pm 2$ kb or 4kb) on either A) FAM172A/CHD7 common peaks or D)



set of promoters from Ensembl database (n=promoters). B, E) Heatmaps of CUT&RUN data for the ratio  $\log_2[\text{H3K4me3+}/\text{H3K27me3+}]$  in WT NCCs. Peak signals are centred ( $\pm 2\text{kb}$ ) on either B) FAM172A/CHD7 common peaks (n=peaks) or E) set of promoters from Ensembl database (n=promoters). C) Relative proportions of commonly bound promoters by FAM172A/CHD7 stratified in active ( $\log_2[\text{H3K4me3+}/\text{H3K27me3+}] > 1$ ), bivalent ( $-1 < \log_2[\text{H3K4me3+}/\text{H3K27me3+}] < 1$ ) or repressed ( $\log_2[\text{H3K4me3+}/\text{H3K27me3+}] < -1$ ) promoters. F) Transcriptional state distribution of FAM172A/CHD7 commonly bound promoters compared to all promoters in NCCs. G, H) Genomic browser view of (G) *Cdkn3* and (H) *Sox8* loci using Easseq, with CUT&RUN data for IgG (control), FAM172A, CHD7, H3K4me3 and H3K27me3 from WT NCCs. Promoters/TSS are represented by a dashed red line.

Representative examples of CUT&RUN and ATAC-seq data are shown in Fig 2.3G, H using *Cdkn3* (*Cyclin-dependent kinase inhibitor 3*) and *Sox8* (*SRY-box transcription factor 8*) loci.

We conclude that promoters co-occupied by FAM172A and CHD7 are predominantly active and enriched in trimethylated H3K4.

### 2.5.3 Potential direct targets of FAM172A and CHD7 are mainly misspliced genes in NCCs from corresponding mutants

To define FAM172A and CHD7 direct transcriptional targets, we intersected the CUT&RUN genomic data with our previously published transcriptomic data obtained from NCCs of *Fam172a<sup>Tp/Tp</sup>* (vs WT) mouse embryos.

We found ~20 % (803 of 4 343) of dysregulated genes in *Fam172a<sup>Tp/Tp</sup>* are bound by FAM172A and CHD7 in NCCs, stratified in upregulated (48 %, 410 of 803), misspliced (27 %, 229 of 803) and downregulated (25 %, 217 of 803) genes (Figure 2.5A). To minimize bias from the peak calling and the peak annotation in our results, we performed the same analysis using an extended list of FAM172A/CHD7 co-occupied genes and confirmed the same distribution (Figure 2.6A). In both cases, we

observed the same proportion of dysregulated events (Figures 2.5A and 2.6A). Surprisingly, we discovered that genes with active promoters ( $\log_2[\text{H3K4me3+}/\text{H3K27me3+}] > 1$ ) and co-occupied by FAM172A/CHD7 are mainly upregulated (54.3 %) in *Fam172a*<sup>Tp/Tp</sup> NCCs, in contrast to genes with poised promoters ( $-1 < \log_2[\text{H3K4me3+}/\text{H3K27me3+}] > 1$ ) where genes are mostly downregulated (54 %) (Figures 2.5B and 2.6B, C). Moreover, gene ontology analysis on FAM172A direct targets revealed different biological process enrichment depending on the transcriptional dysregulation (Figure 2.6D). Downregulated genes are enriched in NCC development-related processes, misspliced genes are enriched in cell signaling and migration-related processes, and finally, upregulated genes are enriched in cell cycle and stress-related processes (Figure 2.6D).

To complete our transcriptomic analysis, we next studied the transcriptomic profile of *Chd7*<sup>Gt/+</sup> e10.5 NCCs (gene trap allele of *Chd7*) again collected by FACS owing to the G4-RFP reporter. Differential analysis of gene expression levels identified 206 downregulated and 98 upregulated genes (*P-value* cutoff of 0.05) (Figure 2.5C). We found overlap between both transcriptomes with 31 % (95 of 304) misexpressed genes from *Chd7*<sup>Gt/+</sup> NCC dataset also affected in *Fam172a*<sup>Tp/Tp</sup> NCCs (Figure 2.5D). NCC development-associated genes tend to be downregulated in both mouse models (Figure 2.5E). More interestingly, further analysis on *Chd7*<sup>Gt/+</sup> NCC transcriptome revealed 267 misspliced genes, almost as much as misexpressed genes (only 4 genes). From the different rMATS categories, affected splicing events were mainly skipped exons (48 %) or retained introns (30.4 %) (Figure 2.5F). In addition, 31% (83 of 267) of the *Chd7*<sup>Gt/+</sup> misspliced genes are shared with *Fam172a*<sup>Tp/Tp</sup> (Figure 2.5G) and the same splicing categories were affected (Bélanger et al., 2018). Gene ontology analysis of these *Chd7*<sup>Gt/+</sup> NCC misexpressed and misspliced genes showed an enrichment in phenotypes associated with CHARGE syndrome, such as decreased body size, cleft palate, abnormal retina, nasal bone, brain and cerebellum morphology (Figure 2.7A, B).

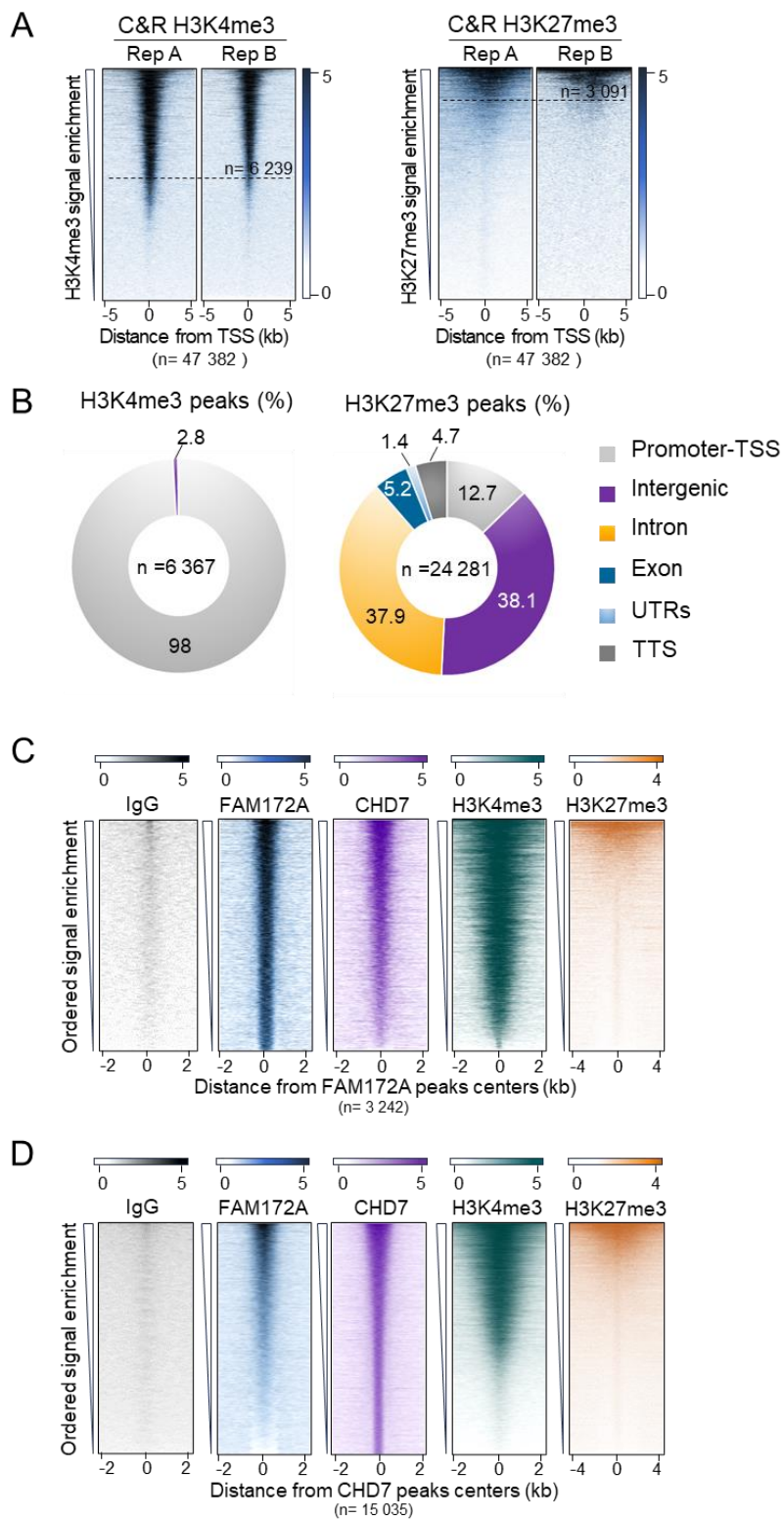


Figure 2.4 (SUPPLEMENTAL) FAM172A and CHD7 co-occupied promoters are enriched in trimethylated H3K4. A) Heatmaps of CUT&RUN alignment data for H3K4Me3 and H3K27me3 in WT NCCs recovered by FACS from e10.5 G4-RFP embryos. Peak signals are centred ( $\pm$  5kb) on Transcription Start Sites (TSS). n=transcripts. B) Relative proportions of HOMER-annotated peaks for H3K4Me3 (left panel) or H3K27me3 (right panel) regions. C, D) Heatmaps of CUT&RUN alignment data for IgG control, FAM172A, CHD7, H3K4Me3 and H3K27me3 in WT NCCs recovered by FACS from e10.5 G4-RFP embryos. Peak signals are centred ( $\pm$  2kb or 4kb) on either C) FAM172A peaks or D) CHD7 peaks. n=peaks.

Finally, we intersected the CUT&RUN data with the new *Chd7*<sup>Gt/+</sup> transcriptome data. We found that 18 % (104 of 570) of dysregulated genes in *Chd7*<sup>Gt/+</sup> are bound by FAM172A and CHD7 in WT NCCs, stratified in misspliced (68 %, 71 of 104), downregulated (20.2 %, 21 of 104) and upregulated (11.5 %, 12 of 104) genes (Figure 2.5H). Genes co-occupied by FAM172 and CHD7 are mainly misspliced in *Chd7*<sup>Gt/+</sup> NCCs (Figure 2.5H). We observed a similar result when we intersected the CUT&RUN data and commonly dysregulated genes from both transcriptomes (Figure 2.5H, blue number).

Altogether, potential direct targets of FAM172A and CHD7 are mainly misspliced genes in both *Fam172a*<sup>Tp/Tp</sup> and *Chd7*<sup>Gt/+</sup> NCCs as exemplified in Figure 2.5I and in accordance with our previous findings (Bélanger et al., 2018, 2022). However, we do not exclude a role for FAM172A and/or CHD7 in direct regulation of expression levels as well.

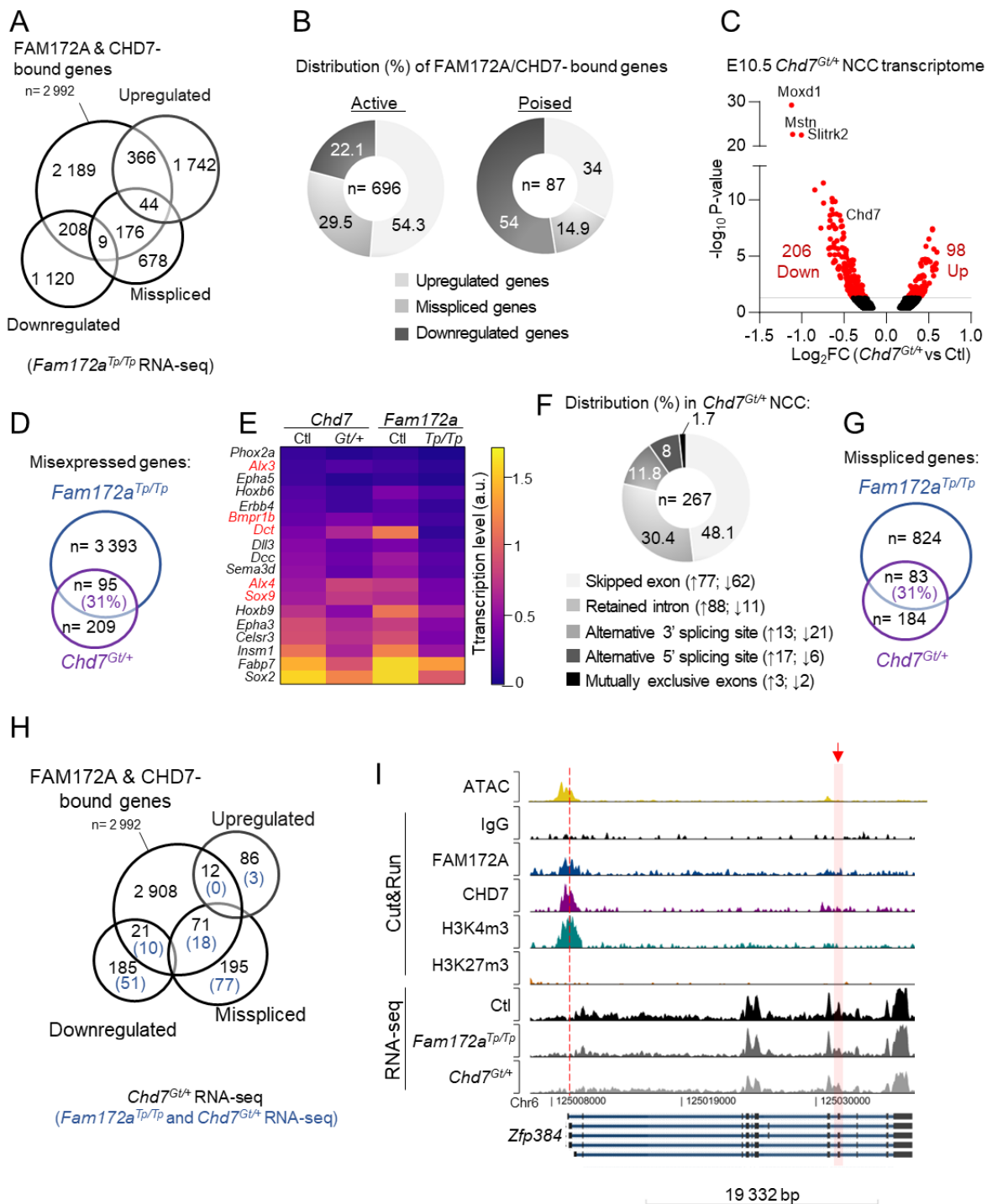
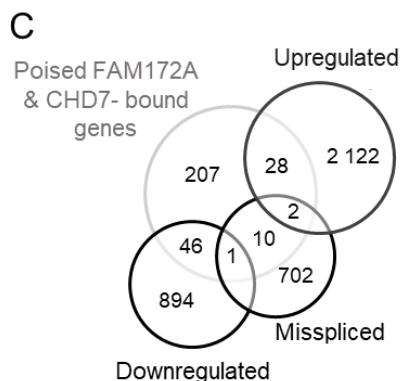
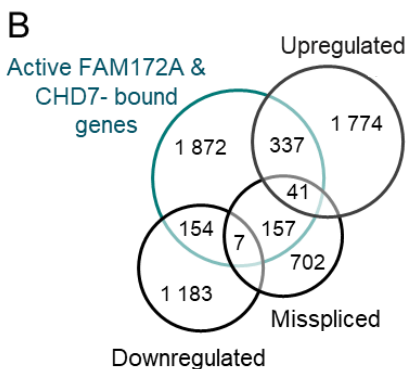
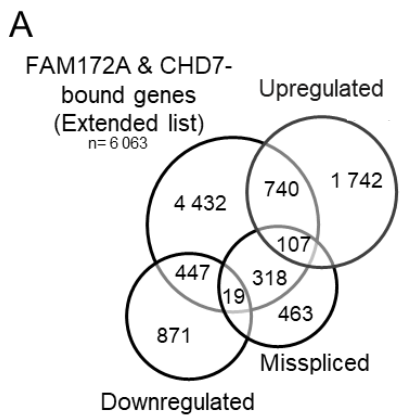
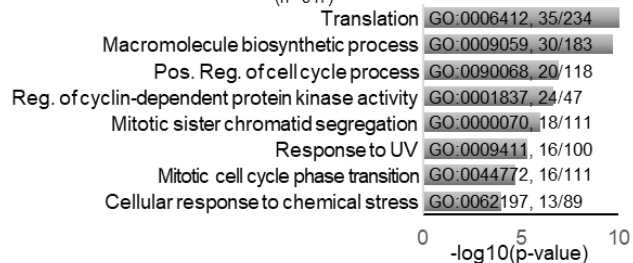


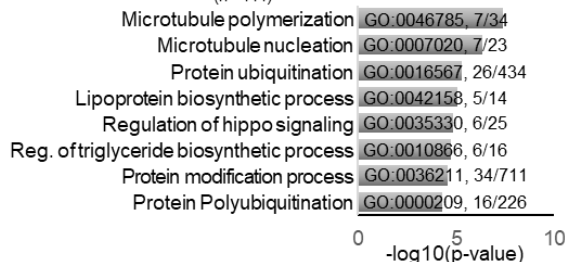
Figure 2.5 Potential direct targets of FAM172A and CHD7 are mainly misspliced genes in NCCs from corresponding mutants. A) Overlap between FAM172A/CHD7 common CUT&RUN targets in WT NCCs and dysregulated genes in *Fam172a*<sup>Tp/Tp</sup> NCC RNA-seq. B) Statistical analysis of FAM172A/CHD7 commonly bound promoter distribution (active or poised) into upregulated, misspliced or downregulated genes from *Fam172a*<sup>Tp/Tp</sup> RNA-seq. n=promoters. C) Volcano plot summarizing a RNA-seq-based analysis of differential gene expression levels in e10.5 NCCs (WT; G4-RFP vs. *Chd7*<sup>Gt/+</sup>; G4-RFP). Only genes with a p-value below 0.05 are displayed. Misexpressed genes are represented in red. D) Overlap between misexpressed genes in *Fam172a*<sup>Tp/Tp</sup> NCCs and misexpressed genes in *Chd7*<sup>Gt/+</sup> NCCs. E) Heatmap of misexpressed genes with known role in NCC development. Downregulated genes in both *Fam172a*<sup>Tp/Tp</sup> and *Chd7*<sup>Gt/+</sup> NCCs are represented in black, and upregulated genes in *Chd7*<sup>Gt/+</sup> NCCs and downregulated in *Fam172a*<sup>Tp/Tp</sup> NCCs are represented in red. F) Pie chart showing the distribution of the 267 differentially modulated alternative splicing events (P < 0.05; variation in inclusion level ≥0.1) in *Chd7*<sup>Gt/+</sup> e10.5 NCCs. Upward- and downward-pointing arrows indicate splicing events that are over- and underrepresented in *Chd7*<sup>Gt/+</sup> e10.5 NCCs, respectively. G) Overlap between misspliced genes in *Fam172a*<sup>Tp/Tp</sup> NCCs and misspliced genes in *Chd7*<sup>Gt/+</sup> NCCs. H) Overlap between FAM172A/CHD7 common CUT&RUN targets in WT NCC and dysregulated genes in *Chd7*<sup>Gt/+</sup> NCCs (black numbers) or dysregulated genes in both *Fam172a*<sup>Tp/Tp</sup> and *Chd7*<sup>Gt/+</sup> NCCs (blue numbers). I) Genomic browser view of *Zfp384* locus using Easseq, with CUT&RUN data for IgG (control), FAM172A, CHD7, H3K4me3 and H3K27me3 from WT NCCs shown in top panels, and transcriptome data for WT, *Fam172a*- and *Chd7*-mutant NCCs shown in bottom panels. Promoters/TSS are represented by a dashed red line. A misregulated splicing event is represented by a red arrow and area.



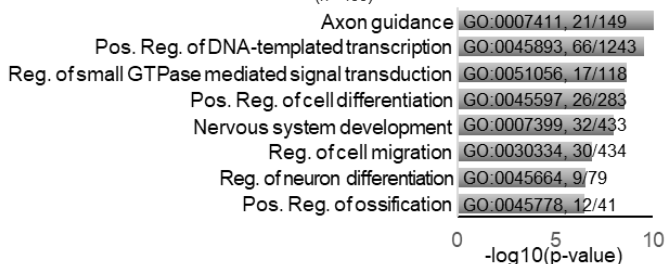
**D** Associated biological processes  
Genes bound by FAM172A/CHD7 and upregulated in *Fam172a*<sup>TP/TP</sup>  
(n=847)



Genes bound by FAM172A/CHD7 and misspliced in *Fam172a*<sup>TP/TP</sup>  
(n=444)



Genes bound by FAM172A/CHD7 and downregulated in *Fam172a*<sup>TP/TP</sup>  
(n=466)



**E** % of genes bound by FAM172A & CHD7 and :

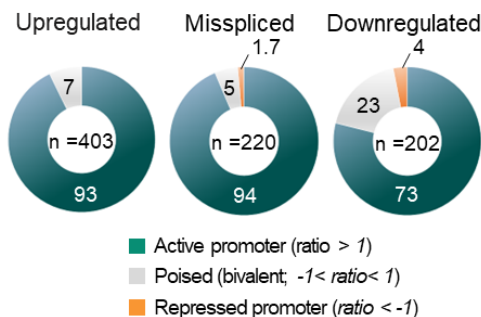


Figure 2.6 (SUPPLEMENTAL) Potential direct targets of FAM172A and CHD7 are mainly misspliced genes in NCCs from corresponding mutants. A) Overlap between FAM172A/CHD7 common CUT&RUN targets (extended list, n=peaks) in WT NCCs and dysregulated genes in *Fam172a<sup>Tr/Tr</sup>* NCC RNA-seq. B, C) Overlap between (B) FAM172A, CHD7 and H3K4Me3 (active) common CUT&RUN targets or (C) FAM172A, CHD7 and H3K27Me3 (poised) common CUT&RUN targets in WT NCCs and dysregulated genes in *Fam172a<sup>Tr/Tr</sup>* RNA-seq. D) Gene ontology analysis showing enriched biological processes for genes with their promoter/TSS co-occupied by FAM172A/CHD7 in WT NCCs and showing dysregulation in transcriptomic data from *Fam172a<sup>Tr/Tr</sup>* NCCs, stratified in upregulated, misspliced and downregulated genes (mutant vs WT). E) Distribution of FAM172A/CHD7 commonly bound genes and dysregulated in *Fam172a<sup>Tr/Tr</sup>* RNA-seq distribution into active ( $\log_2[\text{H3K4me3+}/\text{H3K27me3+}] > 1$ ), bivalent ( $-1 < \log_2[\text{H3K4me3+}/\text{H3K27me3+}] > 1$ ) or repressed ( $\log_2[\text{H3K4me3+}/\text{H3K27me3+}] < -1$ ) promoters. n=promoters.



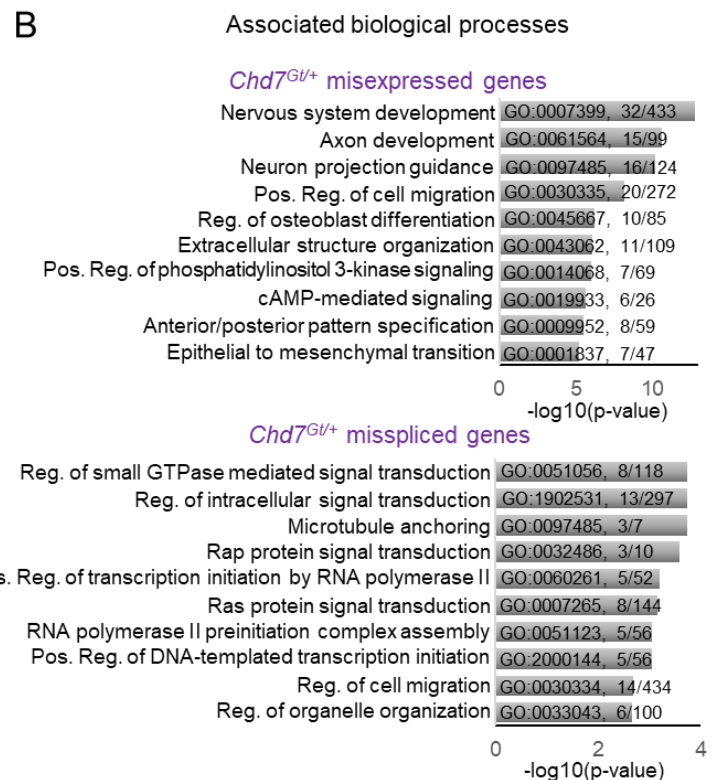
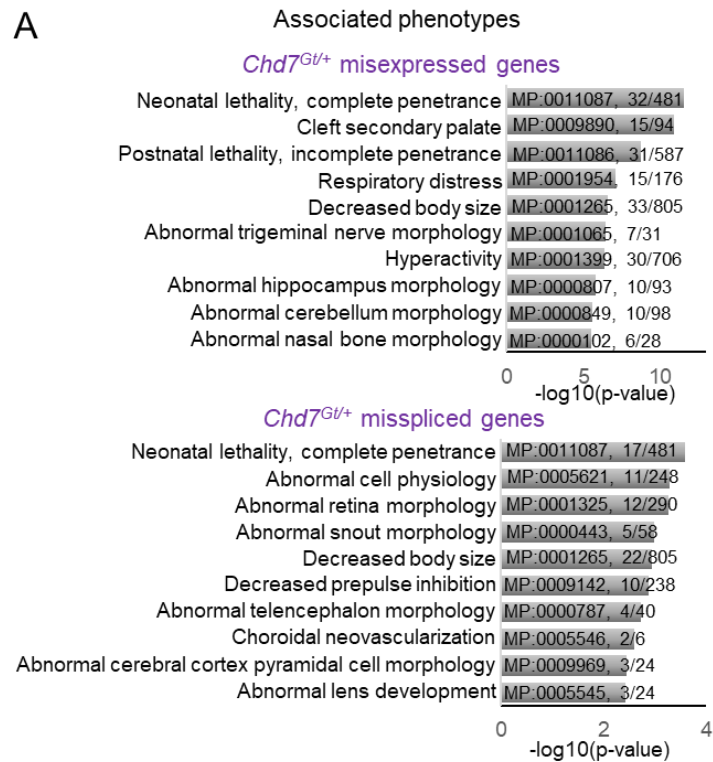


Figure 2.7 (SUPPLEMENTAL) Gene ontology analysis of dysregulated genes in *Chd7<sup>Gt/+</sup>* NCC. A) Gene ontology analysis showing enriched mammalian phenotypes associated with misexpressed and misspliced genes in transcriptomic data of *Chd7<sup>Gt/+</sup>* e10.5 NCCs. B) Gene ontology analysis showing enriched biological processes for misexpressed and misspliced genes in transcriptomic data of *Chd7<sup>Gt/+</sup>* e10.5 NCCs.

#### 2.5.4 FAM172A and CHD7 participate to nuclear speckles organization

For this section, we focused our investigation on FAM172A and CHD7 role in alternative splicing regulation. Using the previously published interactome of FAM172A (Bélanger et al., 2018), several interactions between FAM172A and splicing factors, such as SR proteins (*Serine Arginine rich proteins*) and HnRNPs (*Heterogeneous nuclear ribonucleoproteins*) were identified (Bélanger et al., 2018). These splicing factors are known to be localized at Nuclear Speckles (NS) in mammalian nuclei. To evaluate the impact of FAM172A and CHD7 on splicing factors, we decided to study NS organization specifically in NCCs of *Fam172a<sup>Tp/Tp</sup>* and *Chd7<sup>Gt/+</sup>* e10.5 mouse embryos (SOX10-positive cells) (Figures 2.8A, B and 2.9A).

Using confocal microscopy and the SC35-antibody (a well-known marker for NS that recognizes specifically SRRM2) (Fu & Maniatis, 1990), we observed a decrease in the SC35-mean fluorescence intensity in nuclei of *Fam172a<sup>Tp/Tp</sup>* and *Chd7<sup>Gt/+</sup>* NCCs compared to control (Figure 2.8C, D). Quantifying other features of NS, we also observed a decrease in size and number of NS-associated foci (Figure 2.8A and 2.9B, C). To determine if this observation is caused by either a decrease in SRRM2 expression or a defect in NS organization, we used *Fam172a<sup>Tp/Tp</sup>* and *Chd7<sup>Gt/+</sup>* NCC transcriptomes to analyze the SRRM2 expression level, along with other splicing factors known to be localized in NS (Figure 2.8E). SRRM2 expression level is not affected in both *Fam172a<sup>Tp/Tp</sup>* and *Chd7<sup>Gt/+</sup>* NCCs compared to control suggesting a defect in NS organization (Figure 2.8E). However, expression level of some splicing

factors such as SRSF12 or FUS is decreased, but only in *Fam172a*<sup>Tp/Tp</sup> NCCs (Figure 2.8E), so it cannot explain the decrease of SC35- fluorescence intensity in both mouse models. Interestingly, some studies identified SRRM2 and SON (another well known marker for NS) proteins as essential and sufficient for NS formation (Ilik et al., 2020; Ilik & Aktaş, 2021). To confirm whether only SRRM2 or all factors within NS could be disorganized, we sought to do the same immunofluorescence experiment with an antibody against SON (Figure 2.9D, E). The mean fluorescence intensity associated with SON is also decreased in *Fam172a*<sup>Tp/Tp</sup> NCC nuclei (Figure 2.9D, E), while its expression level is unchanged in both mouse models (Figure 2.8E). These results suggest a general disorganization of NS.

Using co-immunoprecipitation in Neuro2a cells, we validated a physical interaction between FAM172A and SC35 (SRRM2) (Figure 2.8F). This interaction is destabilized with a DNase treatment, but not a RNase treatment, pointing to an indirect interaction at the chromatin level. Intriguingly, the FAM172A-SC35 profile of interaction changes when FAM172A is depleted of its serine hydrolase activity (S294A) (Bélanger et al., 2018) (Figure 2.8F, lower panel).

In conclusion, the lack of FAM172A or CHD7 disrupt splicing factors organization into nuclear speckles. Given the genomic localisation of both proteins at TSS and the physical interaction with NS at the chromatin level, we hypothesize that FAM172A and CHD7 could act on spliceosome factors recruitment at specific gene promoters.

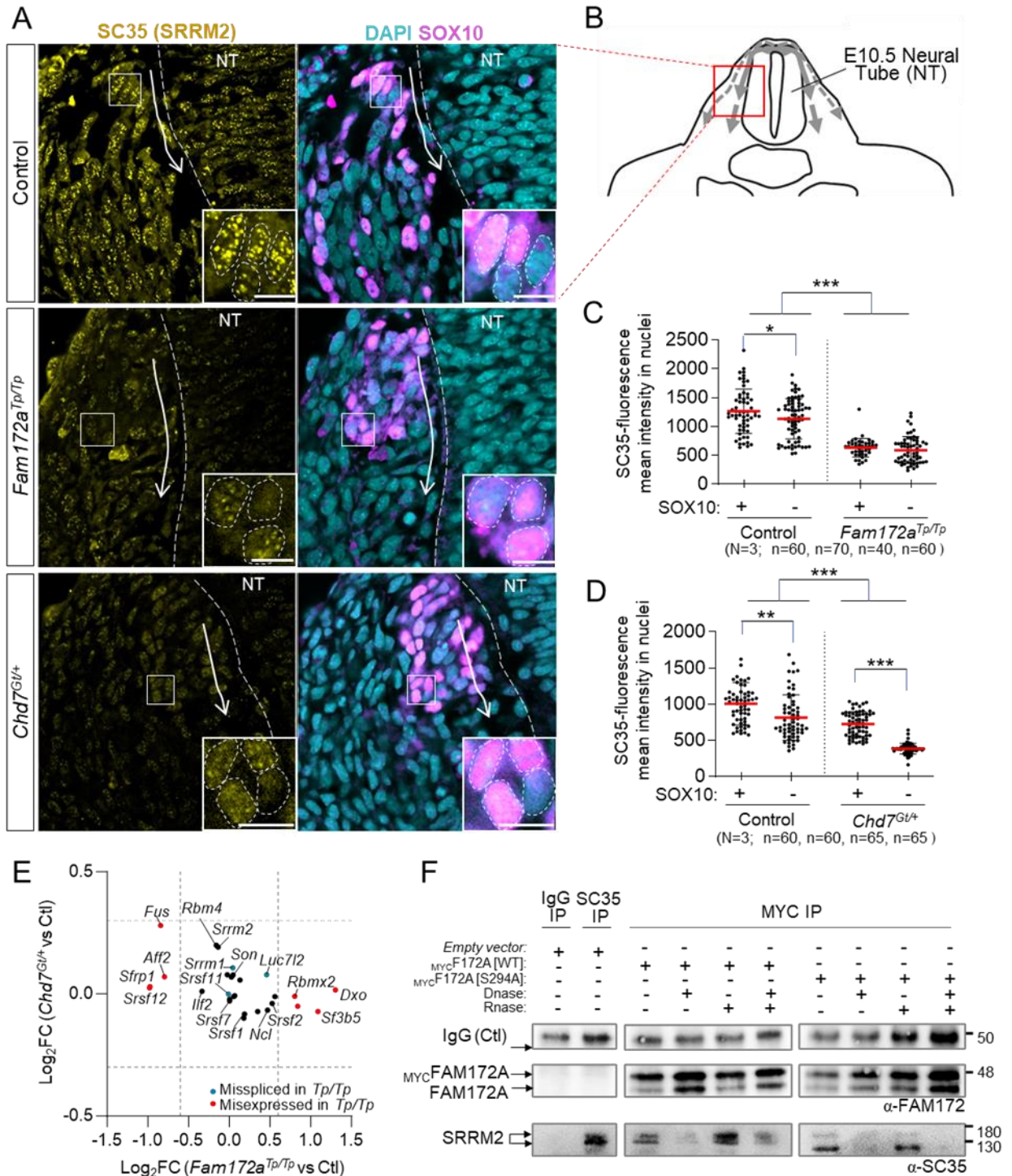


Figure 2.8 FAM172A and CHD7 participate to nuclear speckles organization. A) Immunofluorescence analysis of nuclear speckles in cross sections of control, *Fam172a*<sup>Tp/Tp</sup> and *Chd7*<sup>Gt/+</sup> embryos using an antibody against SC35. Nuclei are stained with DAPI and neural crest cells are identified by the SOX10-positive staining. B) Representative illustration of the immunofluorescence area in a cross-section of E10.5 embryos. C, D) Quantification of the mean fluorescence intensity of anti-SC35 staining in (C) *Fam172a*<sup>Tp/Tp</sup> and (D) *Chd7*<sup>Gt/+</sup> nuclei of SOX10-positive and SOX10-negative cells, using images such as those displayed in A. White dashed lines in zoomed-in views in A delineate measured areas for nucleus. E) Graphical representation of NS-associated splicing factor expression levels using transcriptomic data of e10.5 *Fam172a*<sup>Tp/Tp</sup> and *Chd7*<sup>Gt/+</sup> NCCs. Misspliced genes in *Fam172a*<sup>Tp/Tp</sup> NCCs are represented in blue and misexpressed genes *Fam172a*<sup>Tp/Tp</sup> NCCs are represented in red. F) Co-immunoprecipitation of transfected MYC-FAM172A (WT or S294A) and SC35 in Neuro2a cells using MYC as bait. IgG and SC35 were used as bait for negative and positive control, respectively. Neuro2A were treated with DNase, RNase or both. N=1. Scale bar= 10µm. \* $P \leq 0.01$ , \*\* $P \leq 0.001$  and \*\*\* $P \leq 0.0001$ ; Student's *t*-test.

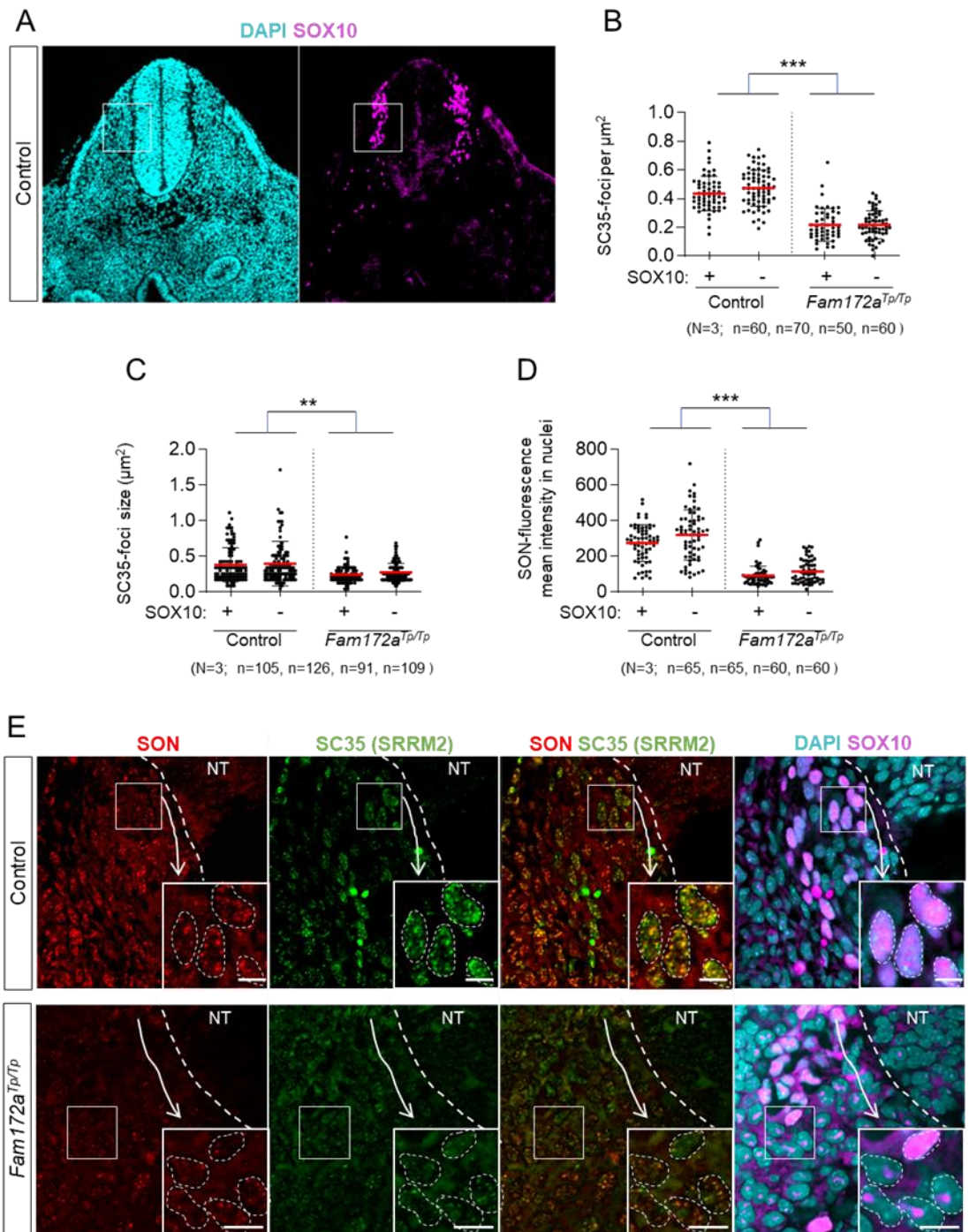


Figure 2.9 (SUPPLEMENTAL) FAM172A and CHD7 participate to nuclear speckles organization. A) Immunofluorescence analysis of neural crest cell localization in sections of E10.5 control embryos using an antibody against SOX10. Nuclei are stained with DAPI. B, C) Quantification of SC35-associated foci (B) number or (C) size in *Fam172a*<sup>Tp/Tp</sup> nuclei of SOX10-positive and SOX10-negative cells, using images such as those displayed in Figure 2.8. D) Quantification of the mean fluorescence intensity of anti-SON staining in *Fam172a*<sup>Tp/Tp</sup> nuclei of SOX10-positive and SOX10-negative cells, using images such as those displayed in E. White dashed lines in zoomed-in views in E delineate measured areas for nucleus. E) Immunofluorescence analysis of nuclear speckles in sections of e10.5 control and *Fam172a*<sup>Tp/Tp</sup> embryos using antibodies against SON and SC35. Nuclei are stained with DAPI and neural crest cells are identified by the SOX10-positive staining. Scale bar= 10µm. \*\* $P \leq 0.001$  and \*\*\* $P \leq 0.0001$ ; Student's *t*-test.



## 2.6 Discussion

Based on our previous work, we proposed that FAM172A and CHD7 cooperate at the chromatin-spliceosome interface for activating the transcription and alternative splicing of many genes important for NCCs. To better understand how these two proteins work together, we decided to study their genomic distribution specifically in NCCs. We found that FAM172A and CHD7 are mainly recruited at active promoter regions (H3K4me3+) in WT NCCs. However, CHD7 is also found at intronic and intergenic regions (47%) suggesting a dual role for CHD7 in NCCs: one with FAM172A at promoter regions and another one without FAM172A outside promoters. These non-promoter regions are most likely enhancers, which is consistent with existing literatures (He et al., 2016; N. C. Reddy et al., 2021; Schnetz et al., 2009b). To define FAM172A and CHD7 direct transcriptional targets, we intersected the CUT&RUN results with the previously published (Bélanger et al., 2018) and newly obtained transcriptomic data from NCCs of *Fam172a*<sup>Tp/Tp</sup> and *Chd7*<sup>Gt/+</sup> (vs WT) mice. The common direct targets of FAM172A and CHD7 correlate better with alternative splicing dysregulation in both *Fam172a*<sup>Tp/Tp</sup> and *Chd7*<sup>Gt/+</sup> NCCs. We also observed a large fraction of downregulated targets for both mutants and upregulated targets only from *Fam172a*<sup>Tp/Tp</sup> NCCs, also suggesting a role for FAM172A and/or CHD7 directly in regulating transcription levels as well. Finally, we shed light on nuclear speckles disorganization in embryonic tissue from *Fam172a*<sup>Tp/Tp</sup> and *Chd7*<sup>Gt/+</sup> (vs WT) mice. Altogether, these results suggest that FAM172A and CHD7 could participate to splicing factors recruitment of spliceosome organization at promoter regions in order to regulate alternative splicing and expression level of genes important for NCCs development.



## CHAPITRE III

### ARTICLE II

#### THE CHARGE SYNDROME-ASSOCIATED PROTEIN FAM172A CONTROLS AGO2 NUCLEAR IMPORT

Sephora Sallis<sup>1,2</sup>, Félix-Antoine Bérubé-Simard<sup>1,4</sup>, Benoit Grondin<sup>1,2</sup>, Elizabeth Leduc<sup>1,2</sup>, Fatiha Azouz<sup>1,2</sup>, Catherine Bélanger<sup>1,4</sup> and Nicolas Pilon<sup>1,2,3</sup>.

<sup>1</sup> Molecular Genetics of Development Laboratory, Department of Biological Sciences, Université du Québec à Montréal, Montreal H3C 3P8, Quebec, Canada.

<sup>2</sup> Centre d'Excellence en Recherche sur les Maladies Orphelines – Fondation Courtois, Université du Québec à Montréal, Montreal H2X 3Y7, Canada.

<sup>3</sup> Department of Pediatrics, Université de Montréal, Montreal H3T 1C5, Canada.

<sup>4</sup> Present address : Biopterre, Sainte-Anne-de-la-Pocatière, QC, Canada.

Contributions :

Sephora Sallis: Design expérimental, expérimentation/collection des données, figures, recherche bibliographique et rédaction.

\*Félix-Antoine Bérubé-Simard: Design expérimental, expérimentation et collection des données (Figures 3.3A-C, 3.11A-B et 3.14).

\*Benoit Grondin: Design expérimental, expérimentation et collection des données (Fig.3.3A-C, 3.12A-D et 3.13A).

Elizabeth Leduc: Expérimentation (aide pour le génotypage et les cultures primaires)

Fatiha Azouz: Expérimentation (Figure 3.1F).

Catherine Bélanger: Expérimentation (Figure 3.7A-B).

Nicolas Pilon: Conception du projet, supervision, recherche bibliographique et rédaction.

\*Contributions équivalentes

### 3.1 Résumé

Le syndrome CHARGE est une maladie génétique rare associée aux cellules de la crête neurale, et dont la cause principale est une mutation du gène *CHD7*. Les autres causes menant à ce syndrome sont des mutations sur d'autres facteurs de la chromatine et/ou facteurs d'épissage. L'un de ces facteurs est la protéine FAM172A, qui a précédemment été retrouvée dans un complexe à l'interface entre la chromatine et le spliceosome avec les protéines CHD7 et AGO2, une protéine capable de lier les petits ARN non codants. Portant notre attention sur l'interaction entre FAM172A et AGO2, nous avons mis en évidence que FAM172A peut lier AGO2 de manière directe, et par ce biais, participe activement à son import nucléaire. Nous avons démontré que ce rôle de FAM172A repose principalement sur son signal de localisation nucléaire bipartite reconnu par la voie classique des importines  $\alpha/\beta$ . Nous avons également démontré que cette translocation nucléaire est stimulée par la phosphorylation de FAM172A par la Caséine Protéine 2, et diminuée par un variant de FAM172A mimant une mutation identifiée chez des patients CHARGE. Ces données mettent en avant l'importance des rôles nucléaire non-canoniques d'AGO2 dans le contexte du syndrome CHARGE.

Mots clés : Syndrome CHARGE ; translocation nucléaire ; épissage alternatif ; AGO2 ; FAM172A ; CK2

### 3.2 Abstract

CHARGE syndrome is a neural crest-related disorder mainly caused by mutation of the chromatin remodeler-coding gene *CHD7*. Alternative causes include mutation of other chromatin and/or splicing factors. One of these additional players is the poorly characterized FAM172A, which we previously found in a complex with CHD7 and the small RNA-binding protein AGO2 at the chromatin-spliceosome interface. Focusing on the FAM172A-AGO2 interplay, we now report that FAM172A is a direct binding partner of AGO2 and, as such, one of the long sought-after regulators of AGO2 nuclear import. We show that this FAM172A function mainly relies on its classical bipartite nuclear localization signal and associated canonical importin- $\alpha/\beta$  pathway, being enhanced by CK2-induced phosphorylation and abrogated by a CHARGE syndrome-associated missense mutation. Overall, this study thus strengthens the notion that non-canonical nuclear functions of AGO2 and associated regulatory mechanisms might be clinically relevant.

Keywords: CHARGE syndrome; nucleocytoplasmic shuttling; alternative splicing; AGO2; FAM172A; CK2

### 3.3 Introduction

CHARGE syndrome is a severe multi-organ developmental disorder mainly affecting derivatives of cranial and cardiac neural crest cells (NCCs) (Pauli et al., 2017; Pilon, 2021). Most of these anomalies are included in the CHARGE acronym: Coloboma of the eye, Hear defects, Atresia of choanae, Retardation of growth/development, Genital abnormalities, and Ear anomalies. While heterozygous mutation of *CHD7* (*Chromodomain Helicase DNA Binding Protein-7*) is recognized as the main cause of this syndromic disorder (Vissers et al., 2004a), rare variants in many other genes encoding chromatin and/or splicing factors have also been recently associated, including *PUF60*, *EP300*, *RERE*, *KMT2D*, *KDM6A* and *FAM172A* (Bélanger et al., 2018; Moccia et al., 2018b).

We previously validated the candidacy of *FAM172A* (*Family With Sequence Similarity 172 Member A*) as CHARGE syndrome-associated gene using a mouse model issued from a forward genetic screen (Pilon, 2016). In this model called *Toupee*, mutagenic transgene sequences are inserted in the last intron of *Fam172a* (Bélanger et al., 2018), generating a hypomorphic allele that negatively affects almost every aspect of NCC ontology (Bélanger et al., 2018). Accordingly, homozygous *Toupee* animals (*Fam172a<sup>TP/TP</sup>*) phenocopy both “major” (*e.g.*, coloboma, cleft palate and vestibular hypoplasia) and “minor” (*e.g.*, retarded growth, cardiac malformations, cranial nerve defects and genital anomalies) features of CHARGE syndrome (Bélanger et al., 2018; Hale et al., 2016). Moreover, the *Toupee* allele was found to genetically interact with a gene-trap mutant allele of the main CHARGE syndrome-associated gene *Chd7* (Bélanger et al., 2018; Hurd et al., 2007). Our mechanistic studies further allowed us to propose a model whereby FAM172A – which contains a large ARB2 domain (Argonaute Binding protein 2) split in two halves by a classical bipartite nuclear localization signal (NLS) – helps to stabilize the chromatin-spliceosome interface as

part of a complex that includes CHD7 and AGO2 (Bélanger et al., 2018). These studies also suggested that perturbation of chromatin-mediated alternative splicing is a general pathological mechanism for CHARGE syndrome, regardless of involved gene defect (Bélanger et al., 2018, 2022; Bérubé-Simard & Pilon, 2019).

AGO2 (Argonaute-2) is a small RNA-binding protein best known as major component of the RNA-induced silencing complex (RISC), which orchestrates post-transcriptional gene silencing in the cytoplasm (Hutvagner & Simard, 2008; J. Liu et al., 2004). Yet, increasing evidence suggests that AGO2 also fulfills important functions in the nucleus (Nazer et al., 2022), notably for regulating alternative splicing (Ameyar-Zazoua et al., 2012; Batsché & Ameyar-Zazoua, 2015; Chu et al., 2021; Kalantari et al., 2016; Meng et al., 2022; Taliaferro et al., 2013; Tarallo et al., 2017). In agreement with these canonical and non-canonical functions, AGO2 has been reported to shuttle between the cytosol and the nucleus of mammalian cells (Kalantari et al., 2016; Ohrt et al., 2011; Schraivogel & Meister, 2014). The relative subcellular distribution of AGO2 is also known to vary as a function of cell types and culture conditions, being notably generally increased in the nucleus of primary cell cultures as opposed to most immortalized cell lines (Sarshad et al., 2018; Sharma et al., 2016). Other conditions that increase AGO2 nuclear localization include cellular stress (Castanotto et al., 2018; Rentschler et al., 2018), senescence (Benhamed et al., 2012; Rentschler et al., 2018) and differentiation (Sarshad et al., 2018). What is less clear is how AGO2 shuttles between both compartments. AGO2 does not contain a NLS or a nuclear export signal (NES). RISC-associated TNRC6 proteins (Trinucleotide repeat containing 6) contain both a NLS and a NES, thereby making them good candidates as intermediary factors. However, such a role appears limited to AGO2 export only, also involving the major nuclear export factor CRM1 (Chromosomal Maintenance 1) (Nishi et al., 2013; Schraivogel et al., 2015b). As of today, the molecular mechanism of AGO2 nuclear import thus remains largely unknown. Although initial studies reported a role for importin-8 (Weinmann et al., 2009), follow up work revealed that many importins are

in fact redundantly involved (Schraivogel et al., 2015b), and no NLS-containing protein is currently known to act as intermediary cargo for these importins.

Here, we report that the NLS-containing FAM172A directly interacts with AGO2 and thereby regulates AGO2 nuclear import at least in part via the canonical importin- $\alpha/\beta$  route, a process that we further found to be influenced by the status of CK2-induced phosphorylation of FAM172A.

### 3.4 Materials and methods

#### Animals

All experimental procedures involving mice were approved by the institutional ethics committee of the Université du Québec à Montréal (CIPA protocols #650) in accordance with the biomedical research guidelines of the Canadian Council of Animal Care (CCAC). The *Fam172a*<sup>Tp/+</sup> (FVB/N background; also known as *Toupee*<sup>Tg/+</sup>) mouse line was as previously described (Bélanger et al., 2018). Embryos were generated by natural mating and staged by considering noon of the day of vaginal plug detection as e0.5. Control and mutant embryos were obtained from the same litter and processed in parallel. Primers used for genotyping are listed in Table 3.1.

#### DNA constructs and recombinant proteins

MYC-, MBP- and HA-tagged versions of the 371aa-long isoform of murine FAM172A (WT and E229Q-mutated) were as previously described (Bélanger et al., 2018). Vectors containing human *AGO1* (pcDNAM-human eIF2C1-WT-Myc; gift from Dr. K. Ui-Tei) (Doi et al., 2003) and *AGO2* ORFs (p3XFLAG-myc-CMV-AGO2, gift from Dr. E. Chan) (Lian et al., 2009) were obtained from Addgene (#50360 and #21538, respectively). For BiFC assays, full-length ORFs of *Fam172* and *AGO2* were subcloned in expression vectors bearing either the N-terminal (pCDNA3-1-173Venus) or the C-terminal half (pCDNA3-174-259Venus) of Venus, which were both generously provided by Dr. S. Merabet (Institut de génomique fonctionnelle de Lyon) (Hudry et al., 2011). NLS-mutated FAM172A (R227Q, R228Q, R241Q and R242Q), phosphodead FAM172A (D219Q, D222Q), phosphomimic FAM172A (S215D, S216D, S217D, S218D and T222D), FLAG-tagged versions of both AGO1 and AGO2 as well as NLS-containing FLAGAGO2 were all generated using the Gibson assembly



method, as previously described (Gibson, 2011; Gibson et al., 2009). Briefly, relevant PCR amplicons (see Table 3.1 for primers list) and digested plasmid (ratio 10:1) were added to the 1X Gibson master mix (5% PEG-8000, 100 mM Tris-HCl, pH 7.5, 10 mM MgCl<sub>2</sub>, 10 mM DTT, 200 μM each of the four dNTPs, and 1 mM NAD) and then combined to 0.01U of T5 EXO, 0.06U of Phusion polymerase and 10U/μl Taq ligase for a 1h-long incubation at 50°C. All DNA constructs were verified via Sanger sequencing.

#### Recombinant proteins and *in vitro* kinase assay

Recombinant MBP, <sub>MBP</sub>FAM172A and <sub>MBP</sub>FAM172A[E229Q] proteins were produced in BL21 bacteria via IPTG (0.3 mM) induction of pMAL-c5X constructs, purified using amylose affinity chromatography and eluted in column buffer, as previously described (Bélanger et al., 2018). Recombinant His-tagged versions of AGO1 and AGO2 proteins were purchased from Sino Biological (#11225-H07B #11079-H07B, respectively). Recombinant CK2 was purchased from New-England Biolabs (P6010S) and used to phosphorylate recombinant FAM172A (with MBP tag released by factor Xa-mediated cleavage) in sodium phosphate buffer supplemented with 50mM KCl, 10mM MgCl<sub>2</sub> and 200μM ATP (pH 7.5) for 1h at 30°C. Samples were then immediately processed for mass spectrometry analysis by the Proteomics Discovery Platform of the Institut de Recherches Cliniques de Montréal (IRCM).

#### Cell culture and transfection

Adherent murine cells were grown in EMEM (N2a, MEFs) or DMEM (NIH3T3, R1) medium supplemented with either 10% (N2a, MEFs and NIH3T3) or 20% (R1) FBS, and 1% penicillin/streptomycin (all products from Wisent). MEFs were generated from e10.5 embryo heads dissociated in EMEM containing 1.3 mg/ml Dispase II (Life Technologies #17105-041), 0.4 mg/ml collagenase (Sigma #C2674) and 0.1 mg/ml

DNase I (Sigma #DN25), for 30 min at 37°C with gentle agitation. The resulting cell suspension was filtered through a 70µm cell strainer (Fisherbrand) and then plated on gelatin-coated coverslips for immunofluorescence or gelatin-coated plates for protein extraction. Human LCLs derived from a FAM172A[E228Q]-associated CHARGE family (Bélanger et al., 2018) were grown in suspension in RPMI medium supplemented with 10% FBS and 1% penicillin/streptomycin (all products from Wisent). All cell cultures were maintained at 37°C and 5% CO<sub>2</sub> in a humidified atmosphere. All transfections were performed using GeneJuice reagents (Millipore-Sigma #70967) in accordance with the manufacturer's instructions (0.1 µg of DNA per cm<sup>2</sup> of cultured cells, and 1µg of DNA for 3µl of GeneJuice reagent).

#### Immunofluorescence analysis of cultured cells

Cells plated on coverslips were transfected with the appropriate plasmid for 48h, fixed in 4% paraformaldehyde (diluted in PBS) for 15 minutes on ice and washed three times in PBS. Cells were then incubated in blocking solution (10% FBS and 0.1% TritonX-100 diluted in PBS) for 1h at room temperature, followed by overnight incubation at 4°C with primary antibodies of interest diluted in blocking solution (see Table 3.2 for antibodies list). After three more washes in blocking solution, cells were incubated with relevant secondary antibodies conjugated with AlexaFluor 488, 594 or 647 for 1h at room temperature (see Table 3.2 for antibodies list), washed again three times and finally counterstained with 5µg/ml DAPI (40,6-diamidino-2-phenylindole) for 10 min. Images were acquired using a Nikon A1 laser scanning confocal microscope with Plan Apo λ 60x1.40 objective, and analyzed using ImageJ. Mean fluorescence intensity in the nucleus and the cytoplasm was quantified using the *measure* function and these values were used to evaluate the relative abundance of proteins of interest in each cell compartment, which was expressed in log<sub>2</sub> scale (log<sub>2</sub>[N:C] ratio).

## Cell fractionation

Cell fractionation of MEFs grown at high density (200 000 cells/cm<sup>2</sup>) and exposed to 2ng/ml Leptomycin B (Cell Signaling Technology #9676) during 4h was performed using the REAP method (Suzuki et al., 2010). Briefly, cells were collected by trypsinization, gently centrifuged and washed twice with PBS. Cell pellets were subsequently resuspended in ice-cold 0.1% NP-40 diluted in PBS and triturated 5 times. One third of this extract was kept as the “total fraction”, while the remaining was centrifuged to pellet the nuclei. The supernatant was then kept as the “cytoplasmic fraction” while pelleted nuclei were washed once with 0.05% NP-40 diluted in PBS, and directly resuspended in standard Laemmli loading buffer to constitute the “nuclear fraction”. Preparation of both the “total fraction” and the “nuclear fraction” was finally completed by sonication (Qsonica, 30% amplitude, 2x 30 seconds).

## Co-Immunoprecipitation in cells and *in vitro*

Co-IP assays in cells were performed using two confluent 100mm plates, 48h after transfection with relevant expression vectors. Whole cell extracts were prepared in a defined lysis buffer with moderate ionic strength (50 mM Tris pH 8.0, 100mM NaCl, 1 mM EDTA, 1% Triton TX-100, 1X Roche Complete protease inhibitors) during 1h at 4°C, after which samples were centrifuged to remove insoluble material. To enhance co-IP of endogenous importins and FAM172A dimers, cells were crosslinked before lysis during 10 min with 0.75% formaldehyde, which was then quenched during 5 min using 0.125M glycine, in accordance with previous studies that used this approach to capture presumably transient protein-protein interactions (Hall and Struhl, 2002; Hoffman et al., 2015; Mohammed et al., 2013). *In vitro* co-IP assays were performed using 300ng of purified recombinant proteins diluted in a low-salt buffer (20mM Tris pH 8.0, 25mM NaCl, 1mM EGTA, 1mM EDTA, 1.5mM MgCl<sub>2</sub>, 1mM DTT, 1%

Triton TX-100, 10% Glycerol, 1X Roche Complete protease inhibitors) previously described to enhance FAM172A-AGO2 interaction (Bélanger et al., 2018). For both cellular and *in vitro* assays, 10% of protein mixture was kept aside as input material and the rest was incubated with primary antibody of interest overnight at 4°C (see Table 3.2 for antibodies list). Antibody-bound complexes were then isolated using protein G-coupled magnetic beads (Dynabeads protein G, ThermoFisher) and analyzed either by western blot (see below) or mass spectrometry (for assessing FAM172A phosphorylation) (Bélanger et al., 2018).

#### Western blot

Protein samples were separated on SDS-PAGE mini-gels using the Mini-Protean system (Bio-Rad Laboratories) and then transferred to PVDF membranes. Membranes were blocked at room temperature for 1h in TBST (Tris-buffered saline-Tween 20) containing 5% non-fat milk, followed by overnight incubation at 4°C with primary antibody of interest (see Table 3.2 for antibodies list). After three washes with TBST, membranes were subsequently incubated with relevant HRP-conjugated secondary antibodies for 1h at room temperature (see Table 3.2 for antibodies list). Chemiluminescent detection was finally performed using Immobilon Western HRP substrate (Millipore-Sigma) and captured using the Fusion FX imaging system (Vilber). When necessary, relative protein levels were quantified using the *gel analyzer* function in ImageJ.

#### Bimolecular Fluorescence Complementation (BiFC) assays in cells

Efficiency of BiFC assays in 6-well plates was assessed 24h post-transfection by calculating the percentage of Venus(YFP)-positive cells and mean fluorescence intensity using a BD Accuri C6 flow cytometer (BD Biosciences). After reaching the desired number of detected events (30,000 or 100,000 cells), remaining cells were used

to verify abundance of N-Venus- and C-Venus-tagged proteins by anti-GFP western blot (see below). Protein abundance was then used for normalization of fluorescence intensity, which also took into account auto-fluorescence in non-transfected cells as background value to be subtracted for all samples. Subcellular distribution of BiFC signal was assessed 48h post-transfection in unfixed cells plated in 35mm  $\mu$ -dishes (Ibidi #81156), after 5-min incubation with 5 $\mu$ g/ml Hoechst 33342. For live imaging, cells were plated on 8-well chamber  $\mu$ -slides (Ibidi #80826), exposed to either 2ng/ml Leptomycin B (Cell Signaling Technology #9676), 5 $\mu$ M ivermectin (Selleckchem #S1351), or vehicle only (same volume of 100% ethanol or DMSO), and monitored for an additional 5h under normal growing conditions, with image acquisition every 20 min. For CK2 inhibition, cells were treated with 20 $\mu$ M 4,5,6,7-tetrabromobenzotriazole (TBB) (Sigma #T0826) 2h prior to the start of monitoring. All images were acquired using a Nikon A1 confocal microscope and analyzed using ImageJ, as described above.

#### RNA extraction and RT-qPCR

RNA from e10.5 MEFs was extracted using the RNeasy Plus Purification Kit (Qiagen) in accordance with the manufacturer's protocol. cDNAs were then generated using 50 ng of total RNA and Superscript II reverse transcriptase (ThermoFischer Scientific). qPCR experiments were finally carried out using the Ssofast EvaGreen Supermix and C1000 Touch thermal cycler (Bio-Rad Laboratories), with *Psm2* gene used for normalization of absolute expression levels and constant exons of the gene of interest (*Cd44*) used for normalization of alternative splicing events (see Table 3.1 for primers list).

## Statistical analysis

Graphical data are represented as the mean  $\pm$  SD, with the number of independent experiments (N) and/or independent biological replicates (n) specified in the figure and/or legend when relevant. Significant differences between samples were determined using the GraphPad Prism software version 6.0, with 0.05 as statistical significance cut-off. Selected statistical tests are indicated in figure legends.

## 3.5 Results

### 3.5.1 FAM172A influences AGO2 nuclear localization

To determine if FAM172A plays a role in the nucleocytoplasmic shuttling of AGO2, we first compared the subcellular distribution of AGO2 in mouse embryonic fibroblasts (MEFs) freshly derived from *Fam172a*<sup>Tp/Tp</sup> and wild-type (WT) e10.5 embryo heads – a cellular model that previously proved useful for highlighting the nuclear colocalization of endogenous FAM172A and AGO2 proteins via immunofluorescence (Bélanger et al., 2018). As noted in this prior work, AGO2 is enriched in the nucleus of WT MEFs (Figure 3.1A-B), this enrichment being however lower when MEFs are plated at high density (Figure 3.2A-B). Under both growth conditions (moderate and high density), we found that AGO2 is depleted from the nucleus of *Fam172a*<sup>Tp/Tp</sup> MEFs, thereby becoming instead markedly enriched in the cytoplasm (Figure 3.1A-B and 3.2A-B). A similar outcome was not seen for AGO1, which remains enriched in the nucleus of MEFs despite the loss of FAM172A (Figure 3.2E-F). Using western blot and fractionation assays in MEFs grown at high density, we further confirmed that the loss of FAM172A specifically affects AGO2 nuclear localization and not overall protein levels (Figure 3.1C-D and Figure 3.2C-D), an outcome that is best evidenced when limiting AGO2 nuclear export using the CRM1 inhibitor Leptomycin B (LMB) (Q. Sun et al., 2013) (Figure 3.1C-D). Other immunofluorescence data revealed that the converse is not true: the predominantly nuclear localization of FAM172A is not altered by the presence of high levels of transfected FLAGAGO2 in the cytoplasm of control MEFs (Figure 3.2G).

To strengthen our observations, we next sought to rescue AGO2 nuclear depletion in *Fam172a*<sup>Tp/Tp</sup> MEFs via transfection of a FAM172A-expressing vector. To this end, we first reviewed the different isoforms of FAM172A, which have been updated since our

previous work (Bélanger et al., 2018). The Ensembl genome browser now predicts three major isoforms in mice (Figure 3.1E), all three being presumably detectable with the commercially-available polyclonal antibody used in our laboratory (Abcam ab121364). Using this antibody to probe embryo extracts at different developmental stages (e9.5 to e15.5) by western blot, we discovered that the 371aa-long isoform is the most predominant, whereas the 417aa- and 307aa-long isoforms are weakly expressed and undetectable, respectively (Figure 3.1F). The same pattern was observed in several murine cell lines such as Neuro2a (N2a) neuroblasts, NIH 3T3 fibroblasts and R1 embryonic stem cells (Figure 3.2H). Accordingly, we found that a MYC-tagged version of the 371aa isoform can largely recapitulate the subcellular distribution of endogenous FAM172A in MEFs (Figure 3.1G), except for the ER that is most likely targeted by the signal peptide-containing 417aa-long isoform (Bélanger et al., 2018). It should be noted that immunofluorescence was our best option for these analyses, due to very low transfection efficiency in MEFs – which notably prevented us from using cell fractionation and western blot that require large amounts of cells. Most importantly, when this MYC-tagged version of the 371aa-long isoform (hereinafter referred to as MYCFAM172A) was transiently transfected in *Fam172a*<sup>TP/TP</sup> MEFs, it could re-establish the predominantly nuclear localization of AGO2 (Figure 3.1H-I) – albeit not up to levels seen in WT MEFs (Figure 3.1B).

Hence, we conclude that FAM172A is actively involved in the regulation of AGO2 nuclear localization.



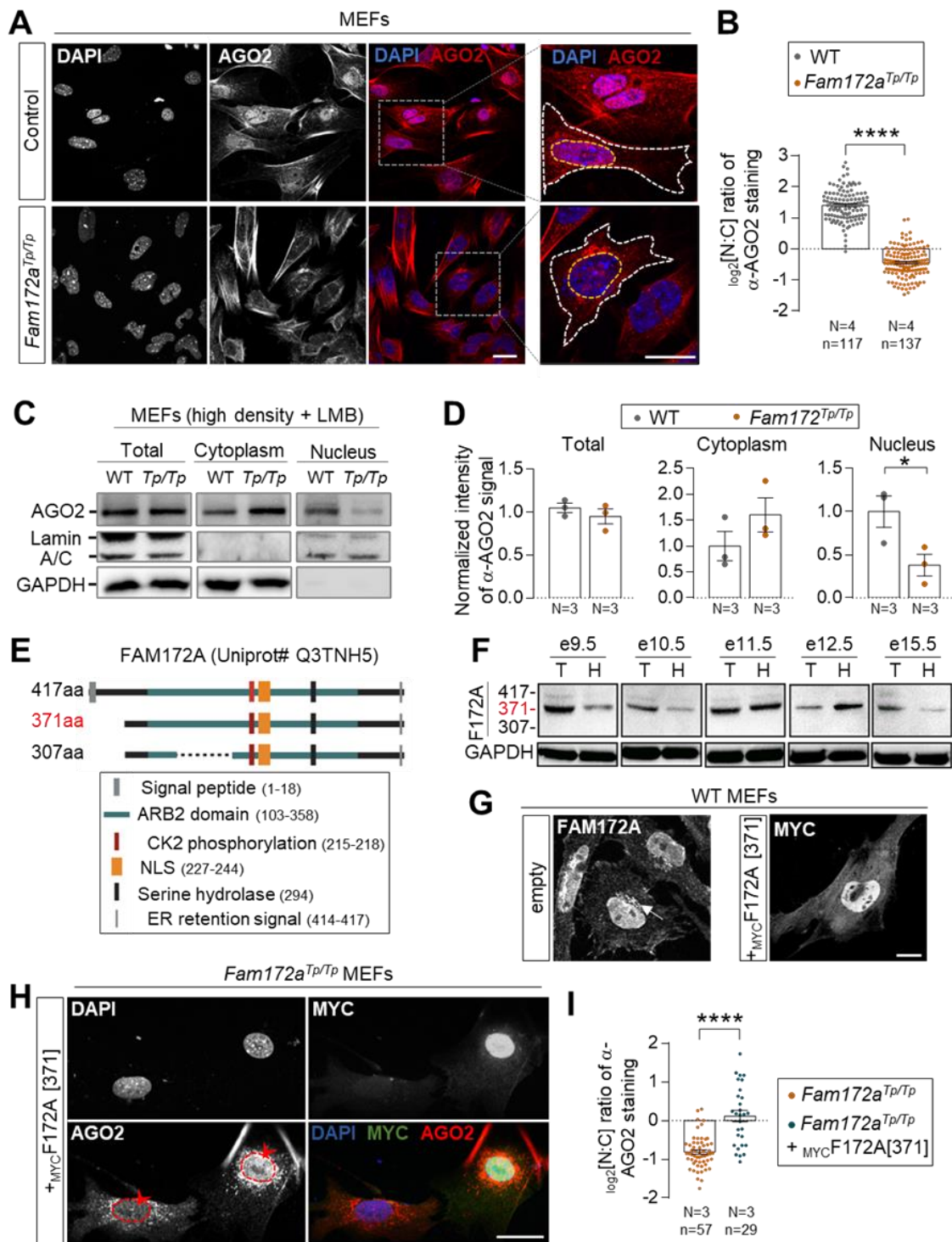


Figure 3.1 FAM172A influences AGO2 nuclear localization (A) Immunofluorescence analysis of AGO2 distribution in WT and *Fam172a*<sup>Tp/Tp</sup> e10.5 MEFs at moderate density (40 000 cells/cm<sup>2</sup>), with nuclei stained using DAPI. (B) Quantification of relative fluorescence intensity of anti-AGO2 staining in nucleus and cytoplasm (N:C ratio, expressed in log 2 scale) using images such as those displayed in A. Yellow and white dashed lines in zoomed-in views in A delineate measured areas for nucleus and cytoplasm, respectively. (C) Western blot analysis of AGO2 protein levels in cytoplasmic and nuclear fractions of WT and *Fam172a*<sup>Tp/Tp</sup> e10.5 MEFs grown at high density (200 000 cells/cm<sup>2</sup>) and exposed to Leptomycin B showing a specific decrease in the nucleus of *Fam172a*<sup>Tp/Tp</sup> cells (N=3). GAPDH and Lamin A/C are used as loading control for cytoplasmic and nuclear fractions, respectively. (D) Quantitative analysis of relative anti-AGO2 signals in mutant vs WT cells (with average in WT set at 1) after normalization with relevant loading control (GAPDH for total and cytoplasmic fractions; Lamin A/C for nuclear fraction), as determined via densitometry (ImageJ) using images such as those displayed in C. (E) Diagram of mouse FAM172A isoforms based on Ensembl/Uniprot databases. (F) Western blot analysis of FAM172A isoforms in trunk (T) or head (H) extracts from embryos at indicated developmental stages, using GAPDH as loading control (N=3). (G) Immunofluorescence staining of endogenous FAM172A and transfected MYC-tagged version of the 371aa-long isoform (MYCF172A[371]) in e10.5 MEFs, with nuclei stained using DAPI (N=3). The arrow in the left panel points the endoplasmic reticulum. (H) Immunofluorescence analysis of AGO2 distribution in *Fam172a*<sup>Tp/Tp</sup> e10.5 MEFs transfected or not with MYCF172A[371]. Red arrowheads compare fluorescence intensity in nuclei of transfected and non-transfected cells. (I) Quantification of relative fluorescence intensity of anti-AGO2 staining in nucleus and cytoplasm (N:C ratio, expressed in log 2 scale) using images such as those displayed in H. Scale bar, 20µm. N= number of biological replicates, n= number of cells. \**P* ≤ 0.05 and \*\*\*\**P* ≤ 0.0001; Student's *t*-test. (Further data can be found in Figure S1).

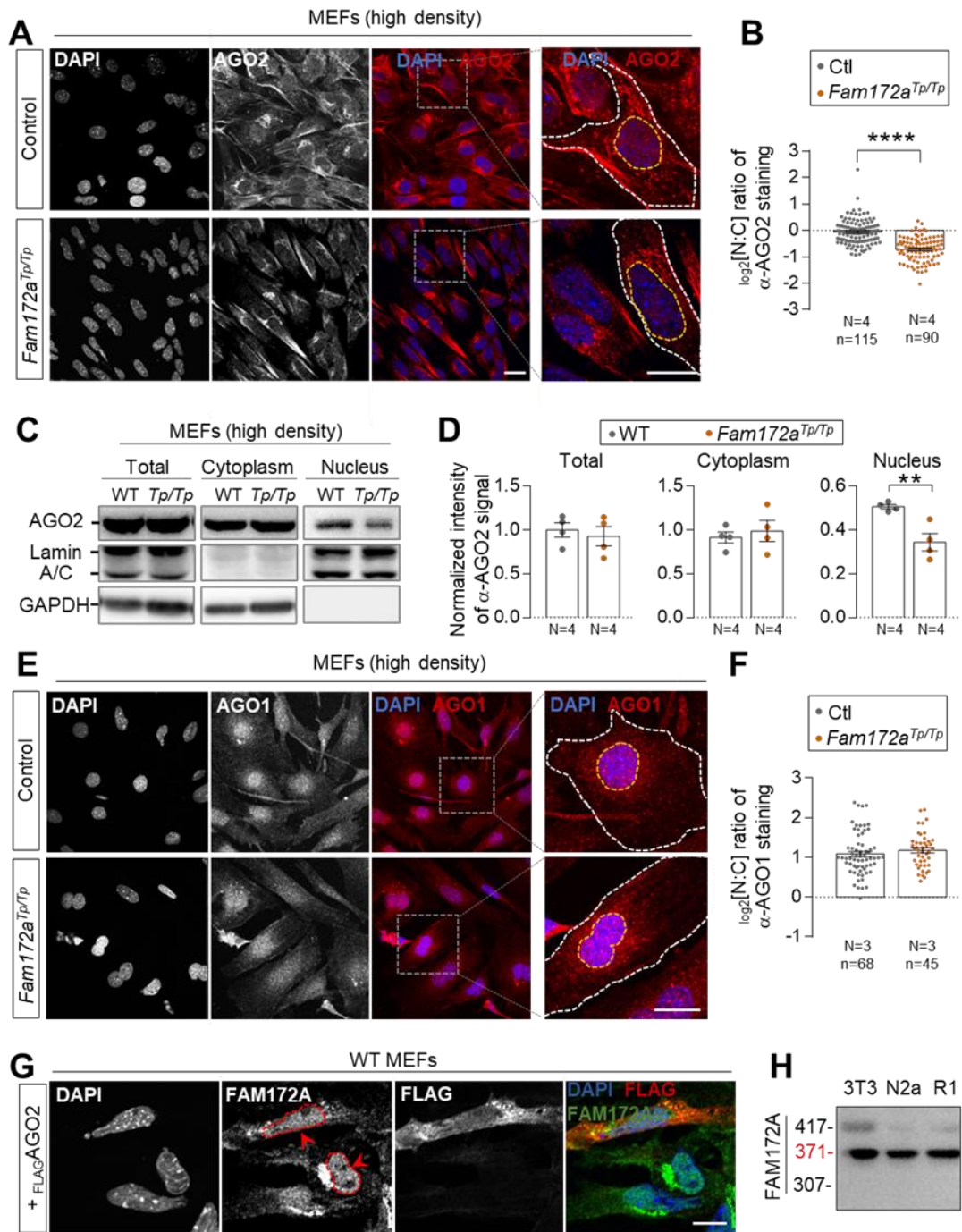


Figure 3.2 (SUPPLEMENTAL) FAM172A influences the nuclear localization of AGO2, but not AGO1. (A) Immunofluorescence analysis of AGO2 distribution in WT and *Fam172<sup>Tr/Tr</sup>* e10.5 MEFs at high density (200 000 cells/cm<sup>2</sup>), with nuclei stained using DAPI. (B) Quantification of relative fluorescence intensity for anti-AGO2 staining in nucleus and cytoplasm (N:C ratio, expressed in log 2 scale) using images such as those displayed in A. Yellow and white dashed lines in zoomed-in views in A delineate areas of measurement for nucleus and cytoplasm, respectively. (C) Western blot analysis of AGO2 protein levels in nuclear and cytoplasmic fractions of WT and *Fam172<sup>Tr/Tr</sup>* e10.5 MEFs grown at high density (200 000 cells/cm<sup>2</sup>) showing a specific decrease in the nucleus of *Fam172<sup>Tr/Tr</sup>* cells (N=3). Lamin A/C and GAPDH are used as loading control for nuclear and cytoplasmic fractions, respectively. (D) Quantitative analysis of relative anti-AGO2 signals in mutant vs WT cells after normalization with relevant loading control (GAPDH for total and cytoplasmic fractions; Lamin A/C for nuclear fraction), as determined via densitometry (ImageJ) using images such as those displayed in C. (E) Immunofluorescence analysis of AGO1 distribution in WT and *Fam172<sup>Tr/Tr</sup>* e10.5 MEFs at high density (200 000 cells/cm<sup>2</sup>), with nuclei stained using DAPI. (F) Quantification of relative fluorescence intensity for anti-AGO1 staining in nucleus and cytoplasm (N:C ratio, expressed in log 2 scale) using images such as those displayed in E. (G) Immunofluorescence analysis of FAM172A distribution in WT e10.5 MEFs transfected or not with <sub>FLAG</sub>AGO2 (N=3). Red arrowheads compare fluorescence intensity in nuclei of transfected and non-transfected cells. (H) Western blot analysis of FAM172A isoforms in NIH 3T3 fibroblasts, Neuro2a (N2a) neuroblasts, and R1 embryonic stem cells (N=3). Scale bar, 20µm. N= number of biological replicates, n= number of cells. \*\**P* ≤ 0.01, \*\*\*\**P* ≤ 0.0001; Student's *t*-test.

### 3.5.2 FAM172A directly interacts with AGO2

Co-immunoprecipitation (co-IP) assays in multiple cell types/tissues previously showed that FAM172A and AGO2 from both endogenous and exogenous sources can be found in a same complex, but not when using a version of murine FAM172A mimicking the CHARGE syndrome-associated human variant pE228Q (pE229Q in mice) (Bélanger et al., 2018). To formally determine if the FAM172A-AGO2 interaction is direct, we performed *in vitro* co-IP assays using purified<sub>MBP</sub>FAM172A and<sub>His</sub>AGO2 recombinant proteins. Analysis of both WT and E229Q-mutated<sub>MBP</sub>FAM172A in this context revealed that the FAM172A-AGO2 interaction is direct and further confirmed the prominent role played by the NLS-located E229 residue of FAM172A, although its mutation did not completely abrogate the interaction with AGO2 in these fairly permissive conditions (Figure 3.3A). To narrow down the interaction interface in AGO2, we also performed co-IP assays using whole lysates of N2a cells transfected with<sub>MYC</sub>FAM172A and deletion constructs of<sub>FLAG</sub>AGO2. This analysis identified the PAZ domain-containing N-terminal half of AGO2 (amino acids 1-480) as being essential for mediating the interaction with FAM172A, with very little contribution by the PIWI-containing C-terminal half (amino acids 478-860) (Figure 3.3B-C). Consistent with the requirement of this interaction for AGO2 nuclear localization, double-immunofluorescence staining of<sub>MYC</sub>FAM172A-transfected N2a cells showed that nuclear signal intensity for co-transfected PAZ domain-containing N-terminal half is similar to full-length<sub>FLAG</sub>AGO2, being in both cases higher than for the PIWI-containing C-terminal half (Figure 3.4A-B). However, both truncated versions of<sub>FLAG</sub>AGO2 gave widely spread out data values in this assay and hence were not found to be statistically different from each other (Figure 3.4B). This suggests that the minimal contribution of the PIWI-containing half of AGO2 to FAM172A binding might nonetheless be important for efficient nuclear import.

To further evaluate the functional relevance of the direct FAM172A-AGO2 interaction for AGO2 nuclear localization, we then reiterated the rescue experiment in *Fam172a*<sup>Tp/Tp</sup> MEFs. Strikingly, in contrast to WT<sub>MYC</sub>FAM172A (Figure 3.1H-I), we found that E229Q-mutated<sub>MYC</sub>FAM172A is unable to correct the abnormally enriched cytoplasmic distribution of AGO2 in these cells (Figure 3.3D-E). Of note, this lack of effect is not due to a reduced ability of E229Q-mutated<sub>MYC</sub>FAM172A to reach the nucleus (Figure 3.3D). Quantitative analyses in N2a cells even suggest that the E229Q mutation increases<sub>MYC</sub>FAM172A nuclear localization (Figure 3.4C-D).

Collectively, these data thus clearly show that AGO2 nuclear localization relies on a direct interaction between FAM172A and AGO2, which is chiefly mediated by the E229 residue of FAM172A and the PAZ domain-containing half of AGO2. Importantly, this mechanism is most likely at play in human cells as well, as evidenced by the reduced nucleus:cytoplasm ratio of AGO2 in a lymphoblastoid cell line (LCL) derived from a CHARGE syndrome child bearing the FAM172A variant E228Q (Figure 3.3F-G).

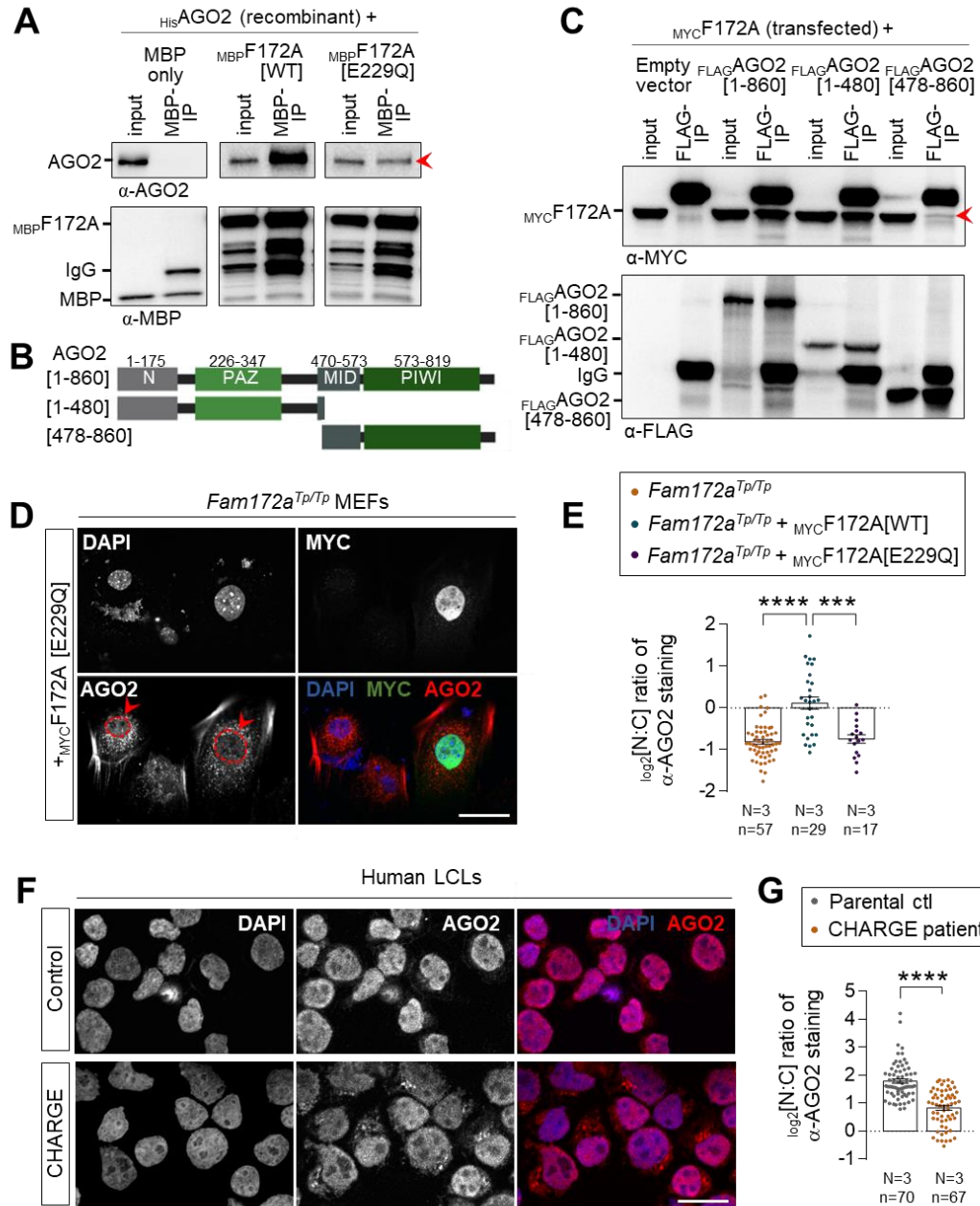


Figure 3.3 FAM172A directly interacts with AGO2. (A) *In vitro* co-IP of recombinant MBP-FAM172A (WT vs E229Q versions) and His-AGO2 using MBP as bait (N=3). MBP alone was used as negative control. The red arrowhead points to reduced amount of co-immunoprecipitated His-AGO2 when using the E229Q version of FAM172A. (B-C) Co-IP of transfected MYC-FAM172A and FLAG-AGO2 (full-length vs indicated N-term and C-term truncations; see panel B) in N2a cells using FLAG as bait (N=3). The red arrowhead in panel C points to reduced amount of co-immunoprecipitated MYC-FAM172A when using the C-term half of AGO2. (D) Immunofluorescence analysis of AGO2 distribution in *Fam172a<sup>Tr/Tr</sup>* e10.5 MEFs transfected or not with E229Q-mutated MYC-FAM172A, with nuclei stained using DAPI. Red arrowheads compare relative fluorescence intensity in nuclei of transfected and non-transfected cells. (E) Quantification of relative fluorescence intensity of anti-AGO2 staining in nucleus and cytoplasm (N:C ratio, expressed in log<sub>2</sub> scale) using images such as those displayed in D. Data for *Fam172a<sup>Tr/Tr</sup>* and *Fam172a<sup>Tr/Tr</sup> + MYC-FAM172A[WT]* conditions are the same as initially displayed in Figure 1I (both assays were performed at the same time), being duplicated here for comparison purposes only. (F) Immunofluorescence analysis of AGO2 distribution in human LCLs derived from an individual with FAM172A[E228Q]-associated CHARGE syndrome and parental control, with nuclei stained using DAPI (N=3 technical replicates). (G) Quantification of relative fluorescence intensity of anti-AGO2 staining in nucleus and cytoplasm (N:C ratio, expressed in log<sub>2</sub> scale) using images such as those displayed in F. Scale bar, 20µm. N= number of biological replicates, n= number of cells. \*\*\* $P \leq 0.001$  and \*\*\*\* $P \leq 0.0001$ ; Student's *t*-test. (Further data can be found in Figure S2).



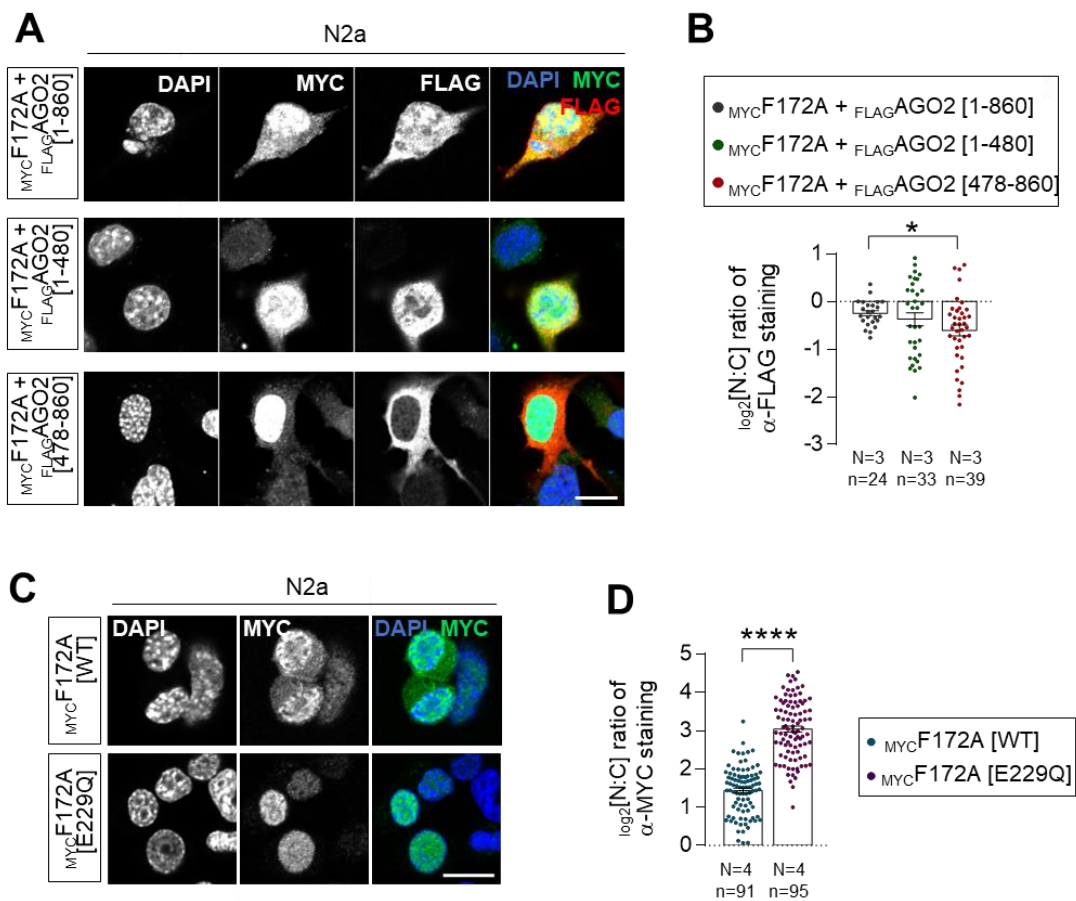


Figure 3.4 (SUPPLEMENTAL) Subcellular distribution of  $_{FLAG}AGO2$  deletion mutants and E229Q-mutated  $_{MYC}FAM172A$ . (A) Immunofluorescence analysis of the subcellular distribution of full-length and indicated deletion mutants of  $_{FLAG}AGO2$  proteins in N2a cells co-transfected with  $_{MYC}FAM172A$ , with nuclei stained using DAPI. Red arrowheads compare relative fluorescence intensity in nuclei. (B) Quantification of relative fluorescence intensity for anti-FLAG staining in nucleus and cytoplasm (N:C ratio, expressed in log<sub>2</sub> scale) using images such as those displayed in A. (C) Immunofluorescence analysis of the subcellular distribution of WT and E229Q-mutated  $_{MYC}FAM172A$  proteins in transfected N2a cells, with nuclei stained using DAPI. Red arrowheads compare relative fluorescence intensity in cytoplasm. (D) Quantification of relative fluorescence intensity for anti-MYC staining in nucleus and cytoplasm (N:C ratio, expressed in log<sub>2</sub> scale) using images such as those displayed in C. Scale bar, 20 $\mu$ m. N= number of biological replicates, n= number of cells. \* $P \leq 0.05$ , \*\*\*\* $P \leq 0.0001$ ; Student's  $t$ -test.

### 3.5.3 The classical bipartite NLS of FAM172A is required for AGO2 nuclear import

To determine if FAM172A either plays an active role in AGO2 nuclear import or only prevents nuclear AGO2 to leave the nucleus, we turned to Bimolecular Fluorescence Complementation (BiFC) assays (Kodama & Hu, 2012). To this end, we generated a series of fusion proteins where the N-terminal half of Venus (a brighter variant of YFP; Yellow Fluorescent Protein) is located on either end of AGO2 and the C-terminal half of Venus is located on either end of FAM172A (Figure 3.5A). Upon transfection in N2a cells, 3 out of 4 tested combinations allowed to reconstitute a fluorescent Venus protein (Figure 3.6A). The combination where both AGO2 and FAM172A have their respective Venus half at their N-terminal end ( $N\text{-VenusAGO2}$  and  $C\text{-VenusFAM172A}$ ) gave a more robust fluorescent signal and was thus selected for further studies. Using flow cytometry, we confirmed the specificity of this BiFC signal using an E229Q-mutated version of  $C\text{-VenusFAM172A}$  as negative control (Figure 3.6B) and by competition with an excess of  $\text{FLAGAGO2}$  (Figure 3.6C). Using immunofluorescence, we also verified that the non-fluorescent fragments fused to AGO2 and FAM172A do not alter their respective subcellular distribution in N2a cells (Figure 3.6D). Initial observations of the subcellular distribution of the BiFC signal generated by the  $N\text{-VenusAGO2-C-VenusFAM172A}$  complex in living N2a and NIH 3T3 cells showed an enrichment in the cytoplasm after 48h of culture (Figure 3.6A,E). This outcome was somehow surprising considering our previous co-IP experiments using N2a cell fractions (cytoplasm *vs* nucleus) that detected an AGO2-FAM172A complex in the nuclear fraction only (Bélanger et al., 2018). We interpret this apparent discrepancy to the methods used, with BiFC being able to capture and lock transient interactions that cannot be detected using co-IP (Kodama & Hu, 2012).

The cytoplasmic enrichment of the BiFC signal supported the hypothesis that AGO2 and FAM172A initiate their interaction in the cytosol and then together enter the

nucleus using the classical bipartite NLS of FAM172A (K/R-K/R-X<sub>10-12</sub>-K/R<sub>3/5</sub>, Fig.S4A) – thereby also predicting a role for the canonical importin- $\alpha/\beta$  route (Wing et al., 2022). To test this model, we devised an experiment based on time-lapse imaging and LMB-induced blockage of nuclear export, also comparing WT and NLS-mutated versions of *C-Venus*FAM172A. In preparation for this experiment, we carefully point-mutated basic residues in each lysine/arginine-rich half of the bipartite NLS sequence of FAM172A (Figure 3.7A) in order to impair its nuclear import without affecting its capacity to interact with AGO2 (Figure 3.5B-D). N2a cells co-transfected with *N-Venus*AGO2 and either WT or NLS-mutated versions of *C-Venus*FAM172A were then followed under a confocal microscope during 5h after addition of LMB. Remarkably, LMB addition led to a constant increase of the BiFC signal in the nucleus over time, but only when using WT *C-Venus*FAM172A (Figure 3.5E-F). With NLS-mutated *C-Venus*FAM172A, the BiFC signal already appeared more enriched in the cytoplasm before LMB addition and remained as such after LMB addition (Figure 3.5E-F). To further validate the involvement of the canonical importin- $\alpha/\beta$  route, we reiterated the BiFC assay using *N-Venus*AGO2 and the WT version of *C-Venus*FAM172A in LMB-treated N2a cells, but now in the additional presence of ivermectin – a small molecule that inhibits importin- $\alpha/\beta$  complexes by specifically targeting importin- $\alpha$  (Wagstaff et al., 2012). From 3h post-treatment onward, we found that ivermectin blunted the otherwise steady time-dependent increase of BiFC signal in the nucleus of N2a cells treated with LMB only (Figure 3.5G-H).

Altogether these data are consistent with the notion that FAM172A-mediated nuclear import of AGO2 involves an interaction between the classical bipartite NLS of FAM172A and importin- $\alpha/\beta$  complexes, as further demonstrated by the inability of our NLS-mutated version of *MYC*FAM172A to re-establish AGO2 nuclear enrichment in *Fam172a*<sup>Tp/Tp</sup> MEFs (Figure 3.5I-J).

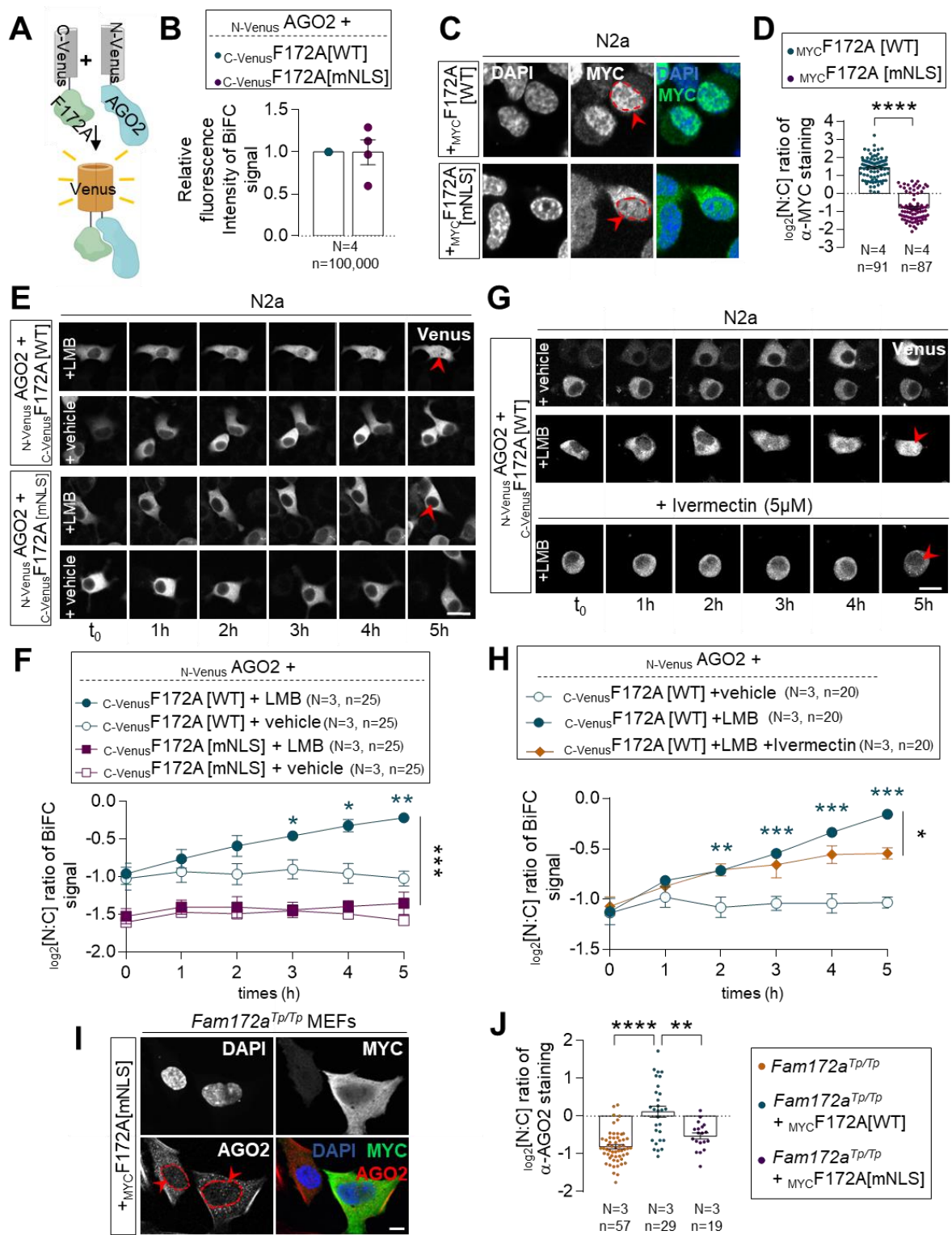


Figure 3.5 The NLS of FAM172A is required for AGO2 nuclear import. (A) Diagram of the BiFC assay based on structural complementation between two non-fluorescent N-terminal and C-terminal halves of a Yellow Fluorescent Protein (Venus) that are respectively fused to AGO2 (N-VenusAGO2) and FAM172A (C-VenusF172A). (B) Comparison of BiFC fluorescence intensity between  $N\text{-VenusAGO2-C-VenusF172A[WT]}$  and  $N\text{-VenusAGO2-C-VenusF172A[mNLS]}$ . Mean fluorescence intensity was determined using flow cytometry 24h after transfection in N2a cells. (C) Immunofluorescence analysis of MYC-tagged WT and NLS-mutated FAM172A proteins in transfected N2a cells, with nuclei stained using DAPI. Red arrowheads compare relative fluorescence intensity in the nucleus for WT and NLS-mutated FAM172A. (D) Quantification of relative fluorescence intensity of anti-MYC staining in nucleus and cytoplasm (N:C ratio, expressed in log 2 scale) using images such as those displayed in C. (E) Five-hour-long time-lapse recordings of BiFC signal generated using  $N\text{-VenusAGO2}$  and  $C\text{-VenusF172A}$  (comparing WT and mNLS versions) in N2a cells treated with Leptomycin B (LMB) or vehicle (ethanol) only. Red arrowheads compare relative fluorescence intensity in the nucleus. (F) Quantification of relative BiFC signal intensity in nucleus and cytoplasm (N:C ratio, expressed in log 2 scale) using images such as those displayed in E. G) Five-hour-long time-lapse recordings of BiFC signal generated using  $N\text{-VenusAGO2}$  and  $C\text{-VenusF172A}$  in N2a cells treated with Leptomycin B (LMB) with or without ivermectin, or vehicle (ethanol) only. Red arrowheads compare relative fluorescence intensity in the nucleus. (H) Quantification of relative BiFC signal intensity in nucleus and cytoplasm (N:C ratio, expressed in log 2 scale) using images such as those displayed in G. (I) Immunofluorescence analysis of AGO2 distribution in *Fam172a*<sup>Tp/Tp</sup> e10.5 MEFs transfected or not with MYCF172A[mNLS], with nuclei stained using DAPI. Red arrowheads compare relative fluorescence intensity in nuclei of transfected and non-transfected cells. (J) Quantification of relative fluorescence intensity of anti-AGO2 staining in nucleus and cytoplasm (N:C ratio, expressed in log 2 scale) using images such as those displayed in I. Data for *Fam172a*<sup>Tp/Tp</sup> and *Fam172a*<sup>Tp/Tp</sup> + MYCF172A[WT] conditions are the same as initially displayed in Figure 1I (both assays were performed at the same time), being duplicated here for comparison purposes only. Scale bar, 20 $\mu$ m (E, G). N= number of biological replicates, n= number of cells. \* $P \leq 0.05$ , \*\* $P \leq 0.01$ , \*\*\* $P \leq 0.001$  and \*\*\*\* $P \leq 0.0001$ ; Student's *t*-test.

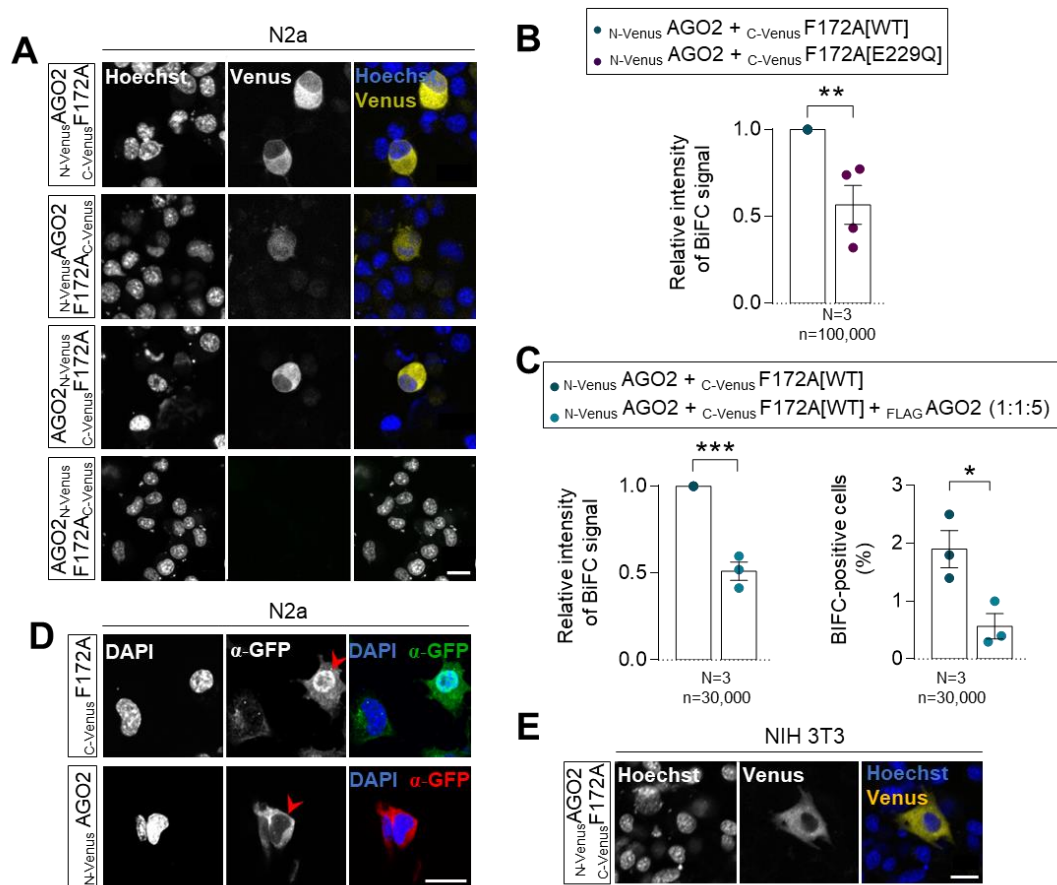


Figure 3.6 (SUPPLEMENTAL) FAM172A-AGO2 interaction is direct and specific. (A) Distribution of AGO2-FAM172A BiFC signal in living N2a cells 48h after transfection with indicated combinations of N-Venus-tagged AGO2 and C-Venus-tagged FAM172A constructs, with nuclei stained using Hoechst (N=3). (B-C) Comparison of overall BiFC fluorescence intensity between  $N\text{-Venus-AGO2-C-Venus-F172A[WT]}$  and  $N\text{-Venus-AGO2-C-Venus-F172A[E229Q]}$  (B) or after competition with a 5-time excess of  $\text{FLAG-AGO2}$  (C; also displaying percentage of positive cells in right panel). Mean fluorescence intensity and percentage of positive cells were determined using flow cytometry 24h after transfection in N2a cells. (D) Immunofluorescence analysis of  $C\text{-Venus-FAM172A}$  and  $N\text{-Venus-AGO2}$  distribution in transfected N2a using a GFP antibody that recognize both halves of Venus, with nuclei stained using DAPI (N=3). Red arrowheads point to transfected cells. (E) Distribution of  $N\text{-Venus-AGO2-C-Venus-F172A}$  BiFC signal in living NIH3T3 cells 48h after transfection, with nuclei stained using Hoechst (N=3). Scale bar, 20 $\mu\text{m}$ . N= number of biological replicates, n= number of cells. \* $P \leq 0.05$ , \*\* $P \leq 0.01$  and \*\*\* $P \leq 0.001$ ; Student's *t*-test.

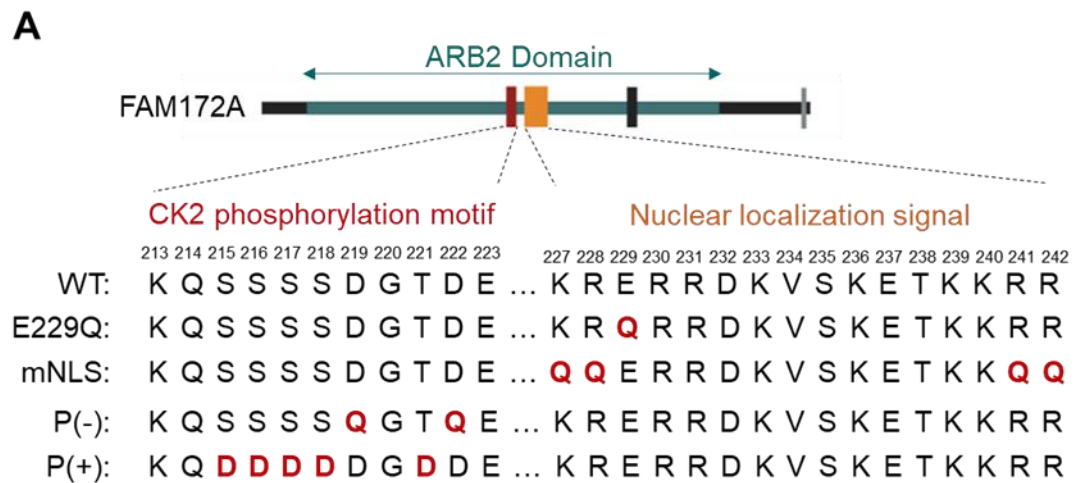
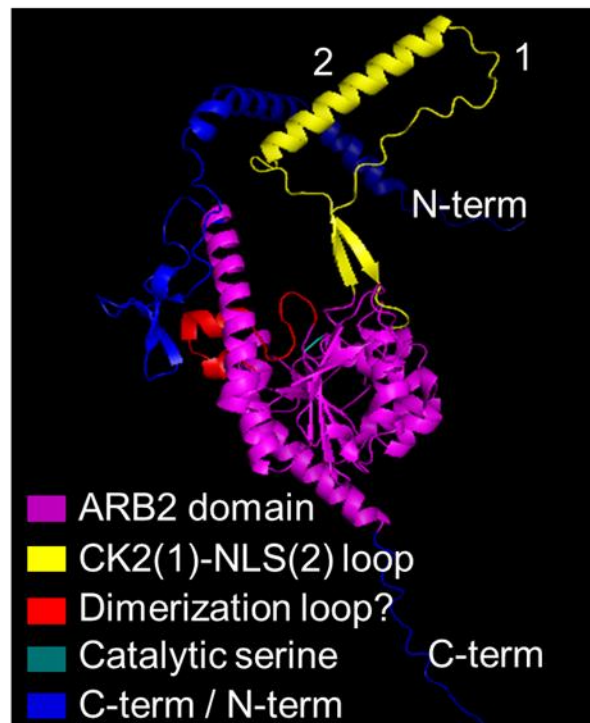
**B**

Figure 3.7 (SUPPLEMENTAL) Sequence of CK2-NLS loop and predicted 3D structure of FAM172A. (A) Diagram of all mutated versions of FAM172A used in this study. (B) PyMOL visualization of AlphaFold-predicted 3D structure of FAM172A.

#### 3.5.4 The status of FAM172A phosphorylation influences AGO2 nuclear import

The FAM172A interactome in N2a cells is known to include the catalytic subunit alpha of Casein kinase II (CK2) (Bélanger et al., 2018). Moreover, we noted that FAM172A bears two juxtaposed consensus-like sequences for CK2 phosphorylation (S/T-X-X-D/E) close to its NLS (Figure 3.7A), and an *in vitro* assay combined to mass spectrometry analysis confirmed that this motif can indeed be directly phosphorylated by CK2 (Figure 3.8A-D). Publicly available high-throughput mass spectrometry data ([www.phosphosite.org](http://www.phosphosite.org)) and our own mass spectrometry analysis of N2 cells (Figure 3.8E-F) also confirmed that this serine-rich motif (S215-S-S-D-S-S-D in human; S215-S-S-S-D-G-T-D in mouse) can be phosphorylated in a natural cellular context. These data further revealed some variance in the identity of the phosphorylated residue, although S217 and S218 appear preferentially targeted in murine FAM172A – representing 87% (including 5% of double-phosphorylation) and 100% of relevant phosphorylated peptides *in vitro* (Figure 3.8A-D) and in N2a cells (Figure 3.8E-F), respectively.

Given that proximity of a CK2 phosphorylation site was previously reported to influence NLS recognition by importins (Hübner et al., 1997; Jans et al., 2000; Xiao et al., 1997), we thus wondered if the status of CK2-induced phosphorylation of FAM172A might play a regulatory role in the nuclear translocation of AGO2. To evaluate this possibility, we again took advantage of our BiFC-based live-cell imaging system to now monitor the nuclear import of  $N\text{-VenusAGO2-C-VenusFAM172A}$  complexes in LMB-treated N2a cells also exposed to the CK2 inhibitor 4,5,6,7-tetrabromobenzotriazole (TBB) (Duncan et al., 2008). This inhibitor significantly reduced the nuclear import of the BiFC complex compared to N2a cells treated with LMB only (Figure 3.9A-B). To complement this analysis, we also generated phosphodead (P-) and phosphomimetic (P+) versions of murine FAM172A, in which either the aspartate residues D220/D223 were replaced by glutamines (Litchfield, 2003;



Lussier et al., 2014) or the upstream serine/threonine residues were replaced by aspartates (H.-L. Sun et al., 2016), respectively (Figure 3.7A). These new constructs were then tested for their ability to mediate AGO2 nuclear import, once again using the BiFC system in living N2a cells. With the phosphodead  $C\text{-VenusFAM172A}$ , we found that the BiFC signal generated in conjunction with  $N\text{-VenusAGO2}$  remains enriched in the cytoplasm regardless of the presence of LMB or not (Figure 3.9C-D, Movie S5) – just like when we used the NLS-mutated  $C\text{-VenusFAM172A}$  (Figure 3.5E-F). In stark contrast, the BiFC signal appears greatly enriched in the nucleus when using the phosphomimetic  $C\text{-VenusFAM172A}$  (Figure 3.9C-D), with both basal and LMB-induced levels being greater than for WT  $C\text{-VenusFAM172A}$  (Figure 3.5E-F). A similar outcome was observed when assessing the subcellular distribution of  $\text{FLAGAGO2}$  in N2a cells co-transfected with  $\text{MYCFAM172A[P+]}$ , which was found to be more efficient than  $\text{MYCFAM172A[WT]}$  at increasing the nuclear localization of  $\text{FLAGAGO2}$  (Figure 3.10).

Of note, our BiFC assays further revealed that both phosphodead and phosphomimetic versions of  $C\text{-VenusFAM172A}$  specifically affect the cytoplasmic-*vs*-nuclear distribution of the  $N\text{-VenusAGO2}$ – $C\text{-VenusFAM172A}$  BiFC signal, not overall intensity (Figure 3.9E). Analysis of the subcellular distribution of phosphodead and phosphomimetic versions of  $\text{MYCFAM172A}$  also suggests that CK2-induced phosphorylation of FAM172A is stimulatory but apparently not essential for its own entry in the nucleus (Figure 3.9F-G).

Nonetheless, this series of experiments strongly suggests that CK2-induced phosphorylation of FAM172A is regulatory important for FAM172A-mediated nuclear import of AGO2.

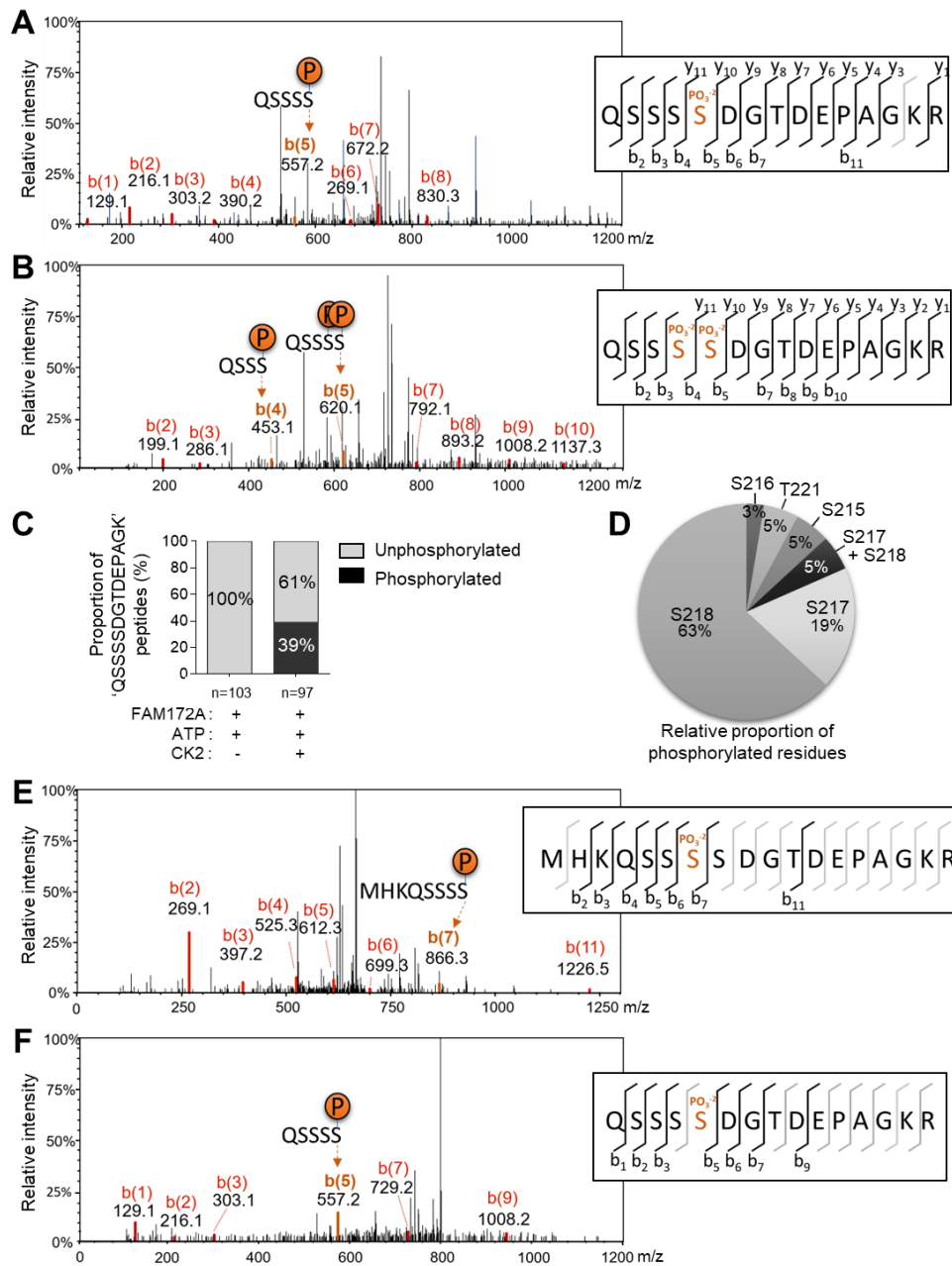


Figure 3.8 CK2 phosphorylates FAM172A. (A-B) Representative CID-MS/MS spectra of most frequently phosphorylated peptides after *in vitro* co-incubation of recombinant FAM172A with recombinant CK2 and ATP. (C-D) Proportion of overall phosphorylation (C) and relative frequency of phosphorylated residues (D) after *in vitro* CK2-induced phosphorylation. (E-F) Representative CID-MS/MS spectra of phosphorylated FAM172A peptides after anti-MYC immunoprecipitation from MYC-FAM172A-transfected N2a cells.

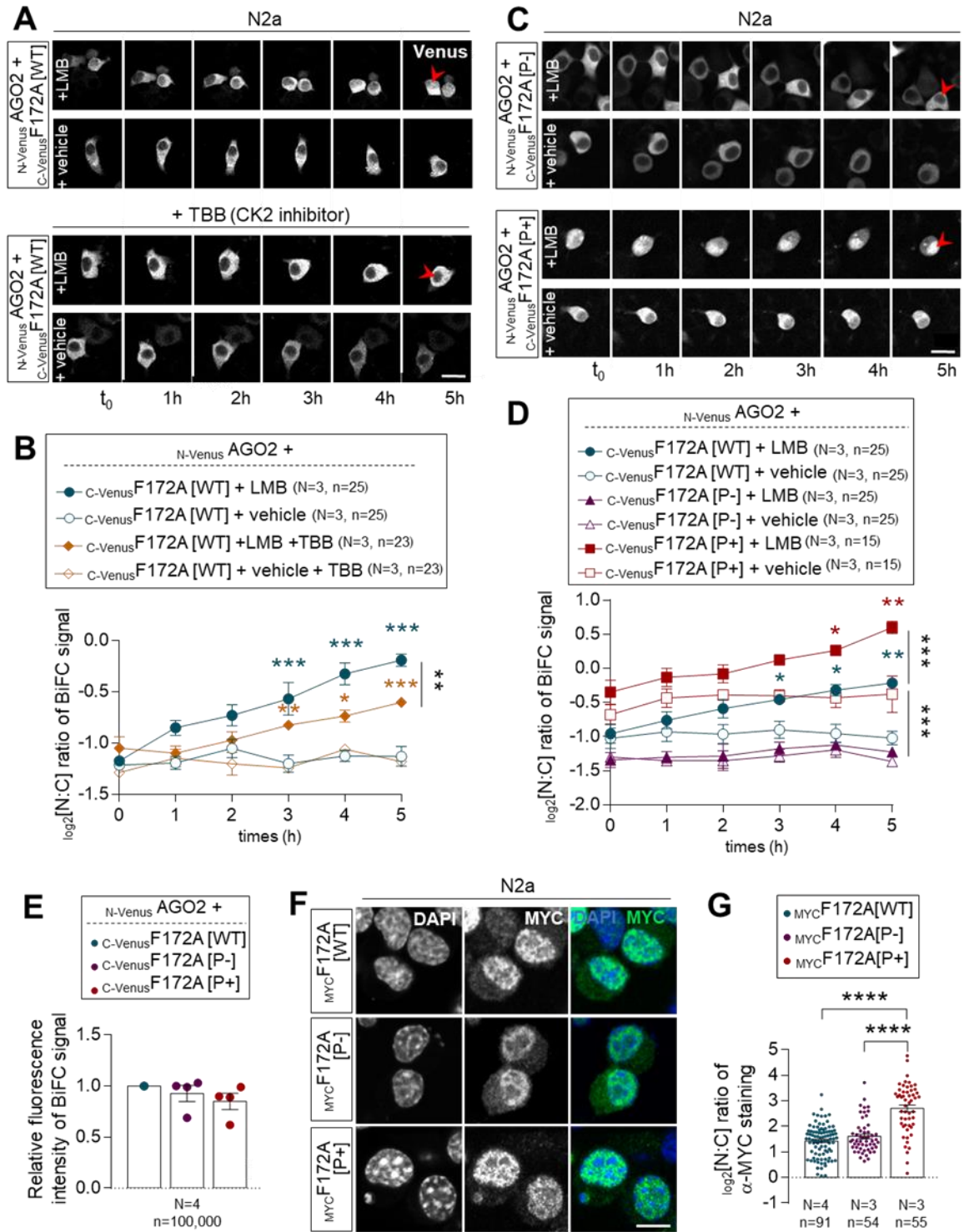


Figure 3.9 AGO2 nuclear import depends on the status of FAM172A phosphorylation. (A) Five-hour-long time-lapse recordings of BiFC signal generated using  $N$ -VenusAGO2 and  $C$ -VenusF172A in N2a cells treated with Leptomycin B (LMB) or vehicle (ethanol) only, with or without 4,5,6,7-tetrabromobenzotriazole (TBB). Red arrowheads compare relative fluorescence intensity in the nucleus. (B) Quantification of relative BiFC signal intensity in nucleus and cytoplasm (N:C ratio, expressed in log 2 scale) using images such as those displayed in A. (C) Five-hour-long time-lapse recordings of BiFC signal generated using  $N$ -VenusAGO2 and  $C$ -VenusF172A (comparing phosphodead [P-] and phosphomimetic [P+] versions) in N2a cells treated with Leptomycin B (LMB) or vehicle (ethanol) only. Red arrowheads compare relative fluorescence intensity in the nucleus. (D) Quantification of relative BiFC signal intensity in nucleus and cytoplasm (N:C ratio, expressed in log 2 scale) using images such as those displayed in C. Data for  $C$ -VenusF172A[WT] are the same as initially displayed in Figure 3F (both assays were performed at the same time), being duplicated here for comparison purposes only. (E) Comparison of overall BiFC fluorescence intensity between  $N$ -VenusAGO2- $C$ -VenusF172A[WT],  $N$ -VenusAGO2- $C$ -VenusF172A[P-] and  $N$ -VenusAGO2- $C$ -VenusF172A[P+]. Mean fluorescence intensity was determined using flow cytometry 24h after transfection in N2a cells. (F) Immunofluorescence analysis of the distribution of MYC-tagged WT, phosphodead (P-) and phosphomimetic (P+) versions of murine FAM172A protein in transfected N2a cells, with nuclei stained using DAPI. (G) Quantification of relative fluorescence intensity of anti-MYC staining in nucleus and cytoplasm (N:C ratio, expressed in log 2 scale) using images such as those displayed in F. Data for the MYCF172A[WT] condition are the same as initially displayed in Figure 3D (both assays were performed at the same time), being duplicated here for comparison purposes only. Scale bar, 20 $\mu$ m (A, C) and 10 $\mu$ m (F). N= number of biological replicates, n= number of cells. \* $P \leq 0.05$ , \*\* $P \leq 0.01$ , \*\*\* $P \leq 0.001$  and \*\*\*\* $P \leq 0.0001$  ; Student's  $t$ -test.

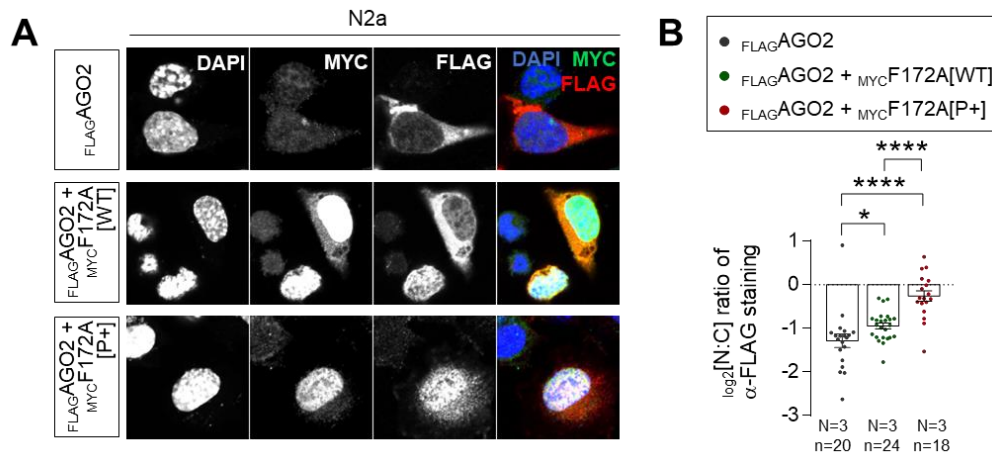


Figure 3.10 (SUPPLEMENTAL) Phosphomimetic  $MYC$ FAM172A stimulates the nuclear entry of  $FLAG$ AGO2. (A) Immunofluorescence analysis of full-length  $FLAG$ AGO2 distribution as a function of the presence of WT and phosphomimetic versions of  $MYC$ FAM172A in transfected N2a, with nuclei stained using DAPI. Red arrowheads compare relative fluorescence intensity in nuclei (N=3). (B) Quantification of relative fluorescence intensity for anti-FLAG staining in nucleus and cytoplasm (N:C ratio, expressed in log 2 scale) using images such as those displayed in A. Scale bar, 20 $\mu$ m. N= number of biological replicates, n= number of cells. \* $P \leq 0.05$  and \*\*\*\* $P \leq 0.0001$ ; Student's  $t$ -test.

### 3.5.5 AGO2 overexpression can functionally compensate for the loss of FAM172A *ex vivo*

In a final set of experiments, we tested the functional relevance of FAM172A-mediated nuclear import of AGO2 via rescue experiments in freshly dissociated *Fam172a<sup>Tr/Tr</sup>* MEFS, using proliferation rate and alternative splicing of *Cd44* as cellular and molecular readouts, respectively (Figure 3.11A-B). To this end, *Fam172a<sup>Tr/Tr</sup>* MEFS were transfected with either <sub>MYC</sub>FAM172A (WT, E229Q-mutant or NLS-mutant) or <sub>FLAG</sub>AGO2 (WT or NLS-containing), and then recovered by FACS owing to coexpressed GFP. As previously observed (Bélanger et al., 2018), transfection of WT <sub>MYC</sub>FAM172A fully rescued the otherwise decreased proportion of proliferating (*i.e.*, Ki67-positive via immunofluorescence) *Fam172a<sup>Tr/Tr</sup>* MEFS compared to WT MEFs (Figure 3.11A). In contrast, both E229Q-mutated and NLS-mutated versions were totally unable to do so (Figure 3.11A), consistent with their incapacity to rescue AGO2 nuclear import in these cells (Figures 3.3D-E and 3.5G-H). Interestingly, overexpression of full-length WT <sub>FLAG</sub>AGO2 (which increases overall levels in both cytoplasm and nucleus; Figure 3.2G) partially rescued while another version engineered to contain a strong NLS (3 copies of SV40 monopartite NLS; Figure 3.11C) fully rescued the proliferation rate of *Fam172a<sup>Tr/Tr</sup>* MEFS (Figure 3.11A). Moreover, congruent with their relative ability to interact with FAM172A (Figure 3.3C) and enter the nucleus (Figure 3.4A-B), the N-terminal half of <sub>FLAG</sub>AGO2 (PAZ domain-containing; amino acids 1-480) was almost as effective as full-length <sub>FLAG</sub>AGO2 in this assay, whereas the C-terminal half (PIWI domain-containing; amino acids 478-860) had virtually no impact (Figure 3.11A). Similar results were obtained when assessing alternative splicing of *Cd44*, a known splicing target of AGO2 (Ameyar-Zazoua et al., 2012) that we also found to be dysregulated by the loss of either of the CHARGE syndrome-associated proteins FAM172A or CHD7 (Bélanger et al., 2018). RT-qPCR analyses revealed higher levels of variable exons 8/9 in *Cd44* transcripts from *Fam172a<sup>Tr/Tr</sup>* MEFS, these abnormal levels being almost fully rescued (*i.e.*, back to

normal levels in WT MEFs) upon transfection of WT  $_{MYC}FAM172A$  or NLS-containing  $_{FLAG}AGO2$  (Figure 3.11B). Altogether, these data are thus in agreement with a general model suggesting that proper FAM172A-mediated nuclear import of AGO2 might be necessary to avoid developing CHARGE syndrome-associated cellular and molecular anomalies.

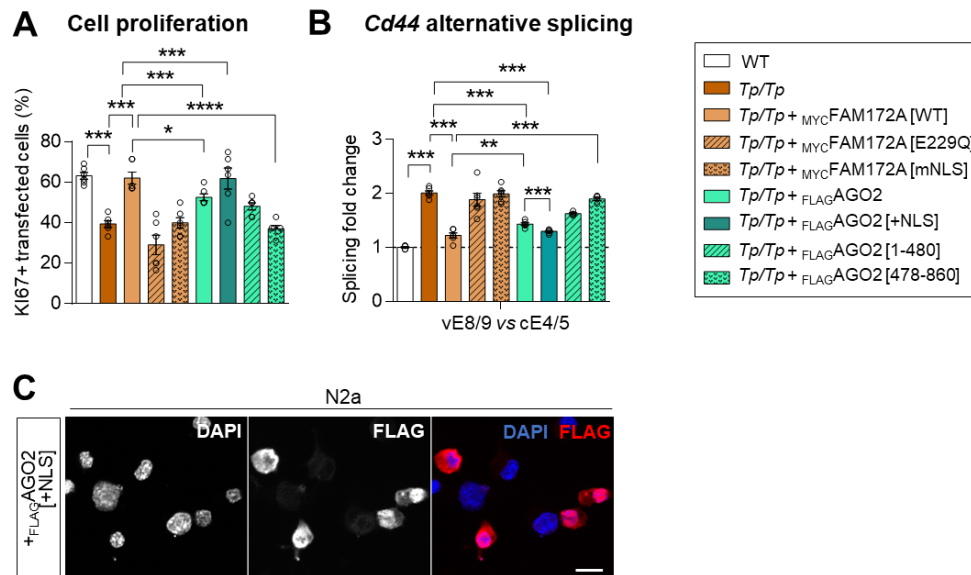


Figure 3.11 (SUPPLEMENTAL) AGO2 overexpression can functionally compensate for the loss of FAM172A *ex vivo*. (A-B) Quantitative analysis of cell proliferation (A) and alternative splicing of *Cd44* (B) in e10.5 *Fam172a*<sup>*Tp/Tp*</sup> MEFs, after cotransfection with a GFP expression vector and either  $_{MYC}FAM172A$  (WT, E229Q-mutant or NLS-mutant) or  $_{FLAG}AGO2$  (WT, NLS-containing, and deletion mutants) expression vectors. Cell proliferation is expressed in % of GFP-positive cells based on double-immunofluorescence staining of GFP and the proliferation marker Ki67. Alternative splicing of *Cd44* was analyzed by RT-qPCR after FACS-mediated recovery of GFP-positive cells. Expression levels of variable exons 8/9 (vE8/9) were normalized against levels of a constant exons 4/5 (cE4/5), and splicing fold change was then determined by comparison to the reference value (dashed line) obtained with WT control embryos (N=5 per condition). (C) Immunofluorescence analysis of the distribution of NLS-containing  $_{FLAG}AGO2$  in transfected N2a cells, with nuclei stained using DAPI. Scale bar, 20 $\mu$ m. \* $P \leq 0.05$ , \*\* $P \leq 0.01$ , \*\*\* $P \leq 0.001$  and \*\*\*\* $P \leq 0.0001$ ; Student's *t*-test.

### 3.6 Discussion

To follow up on our prior work and learn more about the poorly characterized CHARGE syndrome-associated protein FAM172A, we here focused on its interaction with the small RNA-binding protein AGO2 – which is known to play key gene regulatory roles in the nucleus beyond its canonical function in post-transcriptional gene silencing in the cytoplasm (Nazer et al., 2022). Previously published models predicted a central role for AGO2 at the chromatin-spliceosome interface (Ameyar-Zazoua et al., 2012), together with FAM172A that appears required for stabilizing protein-protein interactions (Bélanger et al., 2018; Bérubé-Simard & Pilon, 2019). Our new data generated using various complementary approaches in murine and patient-derived cells now suggest that FAM172A is in addition a regulator of AGO2 nuclear import, involving both the nuclear localization signal and CK2 phosphorylation motif of FAM172A.

Although AGO2 has long been known to shuttle between the cytosol and nucleus of mammalian cells (Kalantari et al., 2016; Ohrt et al., 2008; Schraivogel & Meister, 2014), we still know very little about the underlying mechanisms. Previous studies suggested that many different NLS-binding importins are likely mediating AGO2 nuclear entry (Schraivogel et al., 2015a; Weinmann et al., 2009) but no NLS-containing proteins were previously identified as intermediary factors for the NLS-free AGO2. Such a role has been tested for members of the well-known AGO2-binding protein family TNRC6, which contain both a NLS and a NES (Nishi et al., 2013), with results being consistent with a role in AGO2 nuclear export but not import (Nishi et al., 2013; Schraivogel et al., 2015a). Notably, knockdown of importin- $\beta$  was shown to specifically impair the non-canonical importin- $\alpha$ -independent nuclear import of TNRC6 proteins without perturbing AGO2 nuclear import (Schraivogel et al., 2015a). In contrast, we now present several lines of evidence demonstrating a key role for



FAM172A in this process, with BiFC and rescue data from single cells involving its classical bipartite NLS and associated canonical importin- $\alpha/\beta$  route (Figure 3.5). Yet, we do not think that FAM172A is the only factor involved in AGO2 nuclear import. The fact that there are still appreciable levels of AGO2 in the nucleus of *Fam172a*<sup>Tg/Tg</sup> MEFs argues in that favor (Figure 3.1A-D) – although this might also be partly explained by the residual expression of *Fam172a* (~15% of WT levels) in these hypomorphic mutant cells (Bélangier et al., 2018). The existence of other regulators of AGO2 nuclear import would explain, at least in part, why siRNA-mediated knockdown of importin- $\beta$  did not affect AGO2 nuclear import in prior work (Schraivogel et al., 2015), while our FAM172A-centered studies based on NLS mutagenesis and ivermectin treatment did point to a role for the canonical importin- $\alpha/\beta$  route (Figure 3.5). Another factor potentially contributing to this apparent discrepancy might be that the preferred nuclear import pathway for AGO2 varies as a function of cell types (HEK293 in Schraivogel et al. 2015 vs MEFs and N2a in the current study).

FAM172A itself might also contribute to the diversity of import routes employed by AGO2 – a possibility that would help to explain the rather slow impact of ivermectin in BiFC assays (Figure 3.5G-H). In line with this, close examination of the linker region that connects each lysine/arginine-rich half of the classical bipartite NLS of FAM172A revealed the additional presence of a K-V-S-K-E-T-K motif (Figure 3.7A) that matches the consensus valine/isoleucine-dependent binding sequence (K-V/I-X-K-X<sub>1-2</sub>-K/H/R) for Kap121, the yeast ortholog of importin-5 (Kobayashi & Matsuura, 2013; Wing et al., 2022). Based on this observation, we tested and confirmed that endogenous importin-5 could be specifically co-immunoprecipitated with MYC-FAM172A in N2 cells (Figure 3.12A-B). Yet, we think that these co-IP data should be interpreted with caution, not only because of the rather permissive buffer conditions used (100 mM NaCl  $\pm$  acute crosslink with 0.75% formaldehyde), but also because the FAM172A interactome includes dozens of nuclear proteins that collectively employ various classes of NLS (Bélangier et al., 2018, 2022). This latter point most likely explains the

otherwise intriguing finding that FAM172A can co-immunoprecipitate importin- $\alpha$ 1/ $\beta$ 1 even when using our NLS-mutated version (Figure 3.12C-D) that markedly perturb FAM172A-mediated nuclear import of AGO2 in single-cell assays (Figure 3.5E-F,I-J). In these co-IP assays, we detected a negative effect of the NLS mutation on importin- $\alpha$ 1 recruitment only once (Figure 3.12E).

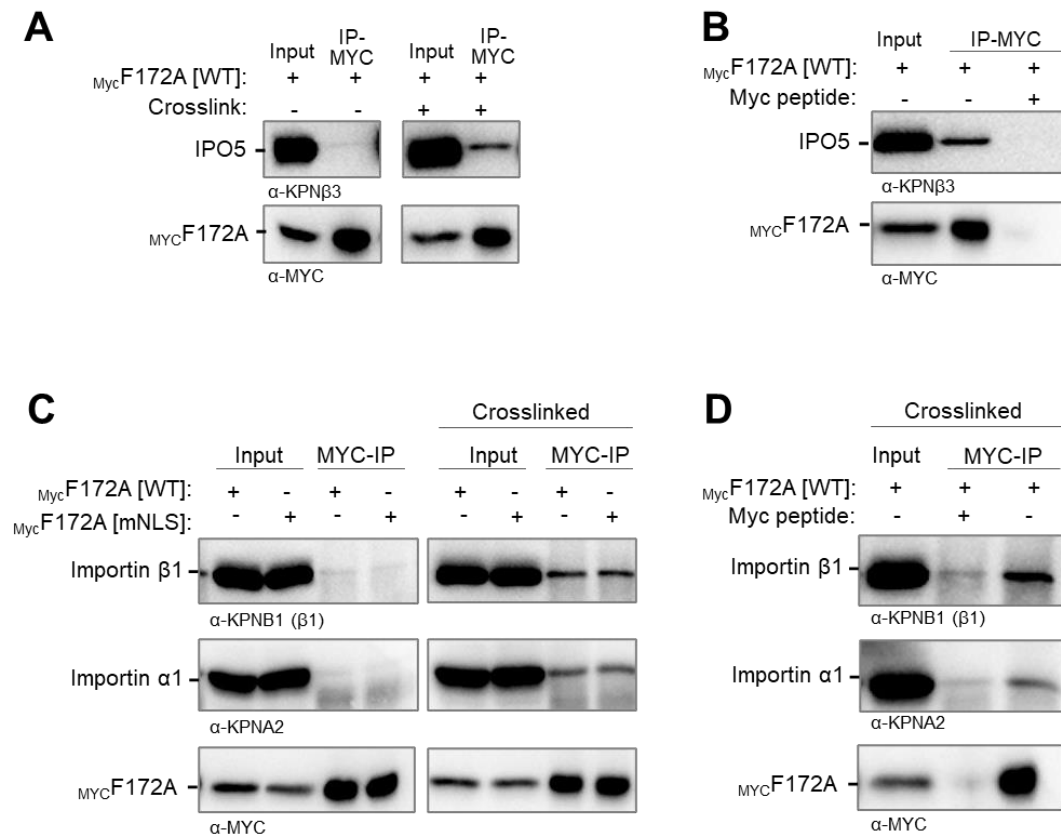


Figure 3.12 (SUPPLEMENTAL) MYCFAM172A interacts with various endogenous importins. (A, B, C, D) Co-IP of transfected MYCFAM172A with endogenous importin-5 (A, B) or endogenous importin- $\beta$ 1 and importin- $\alpha$ 1 (C, D) in N2a cells, using MYC as bait (N = 3). (B, D) Binding was increased by short-term formaldehyde-induced crosslink, and an excess of MYC peptide was used to control specificity (B, D). (E) Impact of the NLS mutation on importin- $\alpha$ 1 recruitment could be evidenced once, in the absence of crosslink.

A better understanding of the molecular underpinnings of FAM172A-mediated nuclear import of AGO2 will definitely require additional studies that are beyond the scope of the current study. An important question to address is how FAM172A can interact with both importins and AGO2 given that the E229 residue important for AGO2 interaction (Figure 3.3) is located within the N-terminal end of the bipartite NLS of FAM172A (Figure 3.7A). A closely related question to address is how CK2-mediated phosphorylation of FAM172A can enhance AGO2 nuclear import. The AlphaFold-predicted 3D structure of FAM172A (with its NLS and juxtaposed CK2 motif both protruding away from the globular core; Figure 3.7B) suggests the possibility that the “CK2-NLS loop” might enable interactions with both AGO2 and importins after phosphorylation-induced conformation changes. Alternatively, other experimental data suggest a potential role for FAM172A homodimerization. Indeed, based on the presence of a conserved sub-region of the ARB2 domain (Figure 3.7B) that was previously shown to mediate homodimerization of the yeast proteins Hda1 (Shen et al., 2016) and Clr3 (Job et al., 2016), we tested and validated the possibility that FAM172A can similarly form homodimers using co-IP (Figure 3.13A) and BiFC (Figure 3.13B-C) assays. Moreover, we found that ivermectin markedly decreased the nuclear enrichment of BiFC signal from FAM172A-FAM172A complexes (Figure 3.13B-C), with a kinetic profile similar to what was observed for AGO2-FAM172A complexes (Figure 3.5G-H). As the putative FAM172A homodimerization interface is away from the CK2-NLS loop (Figure 3.7B), this suggests that FAM172A homodimers could use an NLS-containing loop to interact with importins on one end and another NLS-containing loop to interact with AGO2 on the other end. Clearly, testing both of the above potential mechanisms will represent exciting lines of future research.

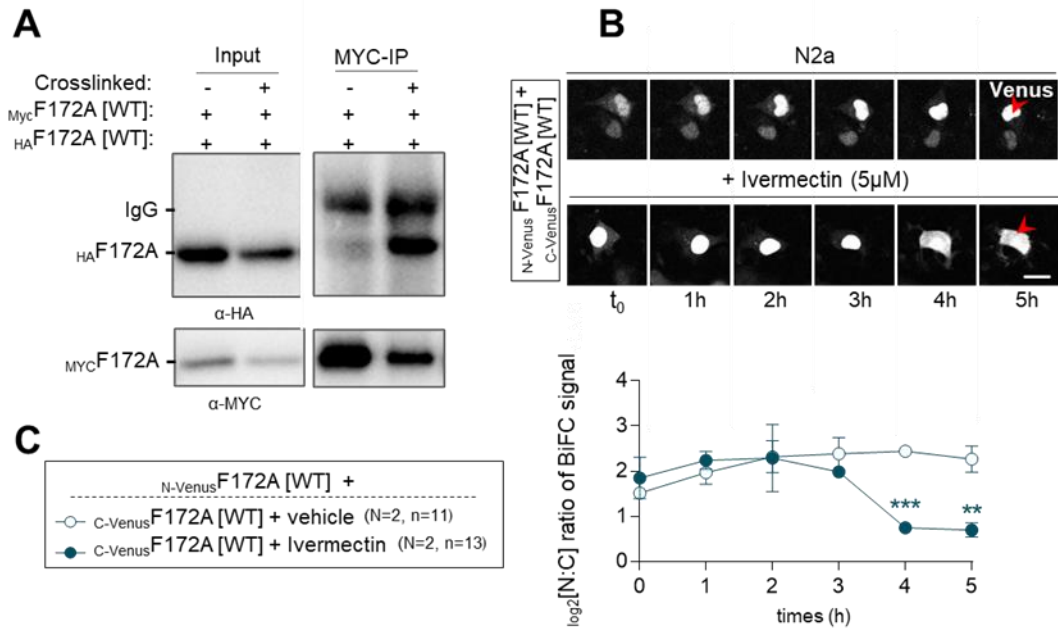


Figure 3.13 (SUPPLEMENTAL) FAM172A can homodimerize. (A) Co-IP of transfected  $MYC_{F172A}$  with  $HA_{F172A}$  in N2a cells, using MYC as bait (N=3). Binding was increased by short-term formaldehyde-induced crosslink, and an excess of MYC peptide was used to control specificity. (B) Five-hour-long time-lapse recordings of BiFC signal generated using  $N_{-Venus}F172A$  and  $C_{-Venus}F172A$  in N2a cells treated with ivermectin or vehicle (ethanol) only. Red arrowheads compare relative fluorescence intensity in the cytoplasm. (C) Quantification of relative BiFC signal intensity in nucleus and cytoplasm (N:C ratio, expressed in log 2 scale) using images such as those displayed in B. Scale bar, 20 $\mu$ m. N= number of biological replicates, n= number of cells. \*\* $P \leq 0.01$ , \*\*\* $P \leq 0.001$ ; Student's *t*-test.

Another intriguing aspect of current study is the fact that the loss of FAM172A in MEFs decreases nuclear levels of AGO2 (Figures 3.1A-D and 3.2A-D) but not those of the closely related AGO1 (Figure 3.2E-F). These observations are consistent with our previous co-IP experiments where we failed to detect an interaction between transfected <sub>MYC</sub>FAM172A and endogenous AGO1 in COS7 cells, despite having used exact same conditions that worked for endogenous AGO2 (Bélanger et al., 2018). However, new data show that FAM172A can interact with AGO1 under certain conditions, like when using recombinant proteins *in vitro* (Figure 3.14A) or upon overexpression in N2a cells (Figure 3.14B). A plausible explanation for all these observations might be that FAM172A simply has a greater affinity for AGO2 than for AGO1, as suggested by the decreased ability of <sub>FLAG</sub>AGO1 to co-immunoprecipitate <sub>MYC</sub>FAM172A in transfected N2A cells compared to <sub>FLAG</sub>AGO2 (Figure 3.14B).

Interestingly, our data suggesting a regulatory role for CK2 in FAM172A-mediated AGO2 nuclear import (Figure 3.9) provide a potential mechanistic link to explain why and how different cellular stresses can increase AGO2 translocation in the nucleus (Castanotto et al., 2018; Rentschler et al., 2018). Indeed, CK2 activity is known to be enhanced by multiple cellular stressors (Ampofo et al., 2013; Sayed et al., 2000; Watabe & Nakaki, 2011). Moreover, it is noteworthy that about half of the proteins (43%; 29/67 proteins) composing the stress-induced response complex (SIRC) associated with AGO2 in HEK293 cells (Castanotto et al., 2018) are also included in the published FAM172A interactome from N2a (Bélanger et al., 2018) and/or PGR9E11 cells (Bélanger et al., 2022). One protein identified as being of special significance in the SIRC is Nucleolin (Castanotto et al., 2018), which we further validated as interactor for both FAM172A and AGO2 in N2a cells using co-IP (Figure 3.14C). All these observations further make sense when considering the particular vulnerability of NCCs to cellular stress in general, along with the extensive overlap between neurocrisopathies and spliceosomopathies (Beauchamp et al., 2020; Bérubé-

Simard & Pilon, 2019; Pilon, 2021). Stress-induced and FAM172A-mediated nuclear import of AGO2 might be a way for this unique, multipotent and highly migratory cell population to defend against cellular stressors.

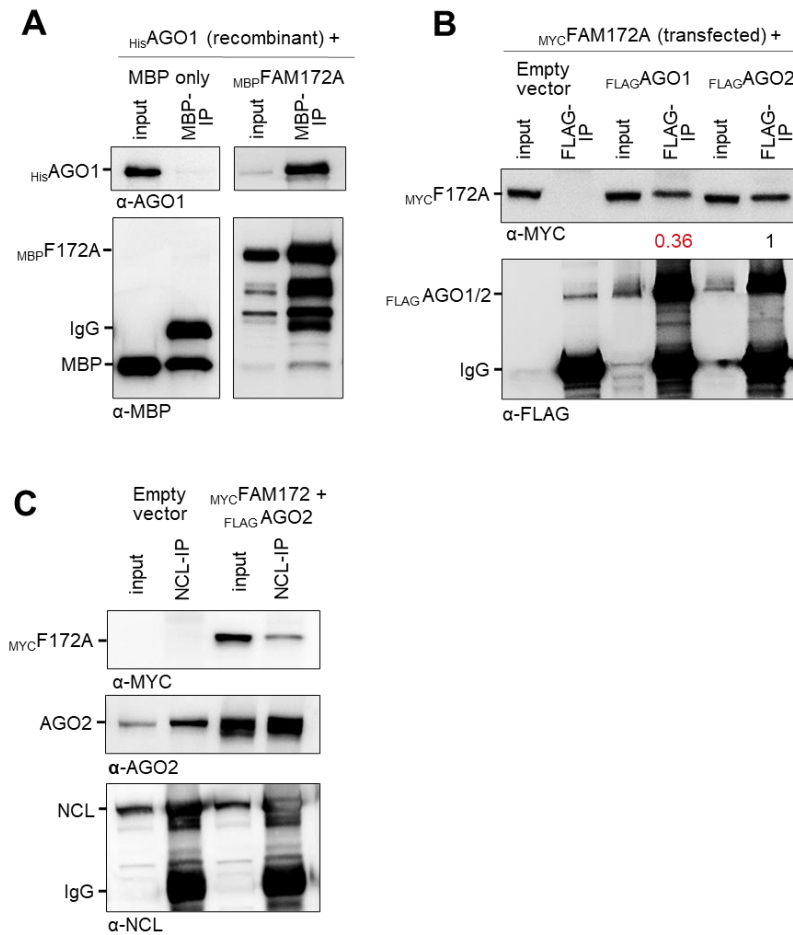


Figure 3.14 (SUPPLEMENTAL) FAM172A also interacts with AGO1 and NCL. (A) *In vitro* co-IP of recombinant *MBP*FAM172A and *His*AGO1, using MBP as bait (N=3). MBP alone was used as negative control. (B) Co-IP of transfected *MYC*FAM172A and either *FLAG*AGO1 or *FLAG*AGO2 in N2a cells, using FLAG as bait (N=3). Numbers between immunoblots refer to the amount of co-immunoprecipitated *MYC*FAM172A after normalization for amount of immunoprecipitated *FLAG*AGO proteins, as determined via densitometry. (C) Co-IP of transfected *MYC*FAM172A and *FLAG*AGO2 with endogenous Nucleolin (NCL) in N2a cells, using NCL as bait (N=3)

### 3.7 Acknowledgements

The authors thank the Cellular analyses and Imaging core (CERMO-FC, UQAM) for assistance with confocal imaging and flow cytometry; the Proteomics Discovery Platform (IRCM) for the mass spectrometry analyses; Dr. S. Merabet and J. Reboulet (Institut de génomique fonctionnelle de Lyon) for providing BiFC constructs and technical advice; Dr. J. W. Belmont and P. Hernandez (Baylor College of Medicine) for providing the human lymphoblastoid cell lines; and Dr. K. Borden and B. Culjkovic-Kraljacic (Institut de recherche en immunologie et cancer) for the gift of importin antibodies. This work was supported by grants from the Canadian Institutes of Health Research to NP (grant #PJT-152933 and OGB-183998). NP was also supported by the Fonds de la recherche du Québec – Santé (FRQS Senior Research Scholar) and by the UQAM Research Chair on Rare Genetic Diseases, while EL and CB were supported by doctoral scholarships from the Natural Sciences and Engineering Research Council (NSERC) and FRQS, respectively, and both SS and FA were supported by scholarships from the CERMO-FC.

## 3.8 Tables

Table 3.1 List of primers used in this study

	Oligos Name	Sens	Sequence 5'-3'	
Genotyping	<i>FAM172A<sup>Trp</sup></i>	F	5'-GGGAGTAAGTCCTACCAATGTTAAATC	
		R	5'-TGACCTTCTAAACAGTCCCATATCCCC	
	<i>Wild type</i>	F	5'-GAAGTGGGAACAAAACACCCTTGG	
		R	5'-GTAAGGCTCGCACTGACATAGC	
mutagenesis	Gibson for FAM172A in backbone pCDNA3.1	N-term-F	5'- CTCGGCATGGACGAGCTGTACAAGGAATTCATGAAAAAGGAC GAACCACCTT TT	
		C-term-R	5'-CACTATAGAATAGGGCCCTCTAGACTCACAGCTCCTCGTGC	
	Gibson for FAM172A in backbone pIRES2-GFP	N-term-F	5'- TTTCTGAAGAAGATCTGAGATCTCTCGAGATGAAAAAGGACG AACCACCTTTT	
		C-term-R	5'-GGATCCC GCCGGTTCGACCTGGAGGTCACAGCTCCTCGTGC	
	FAM172A [mNLS]	N-term-R	5'-CCTTTCCTGTTGCCCTGCTGGCTCATCTG	
		mNLS-F	5'- CAACAGGAAAGGAGAGATAAGGTCTCCAAGGAAACAAAGAA GCAACAGGAT TTCATGAGAAG	
		mNLS-R	5'- ATCCTGTTGCTTCTTTGTTTCCTTGGAGACCTTATCTCTCCTTT CCTGTTGC CCTGTGGCTC	
		C-term-F	5'-AAGCAACAGGATTTCTATGAGAAGTACCGCAACCC	



FAM172A[P-]	N-term-R	5'- ATGCATTTTCTGCTTTTCCACTTCTATATAGTTTTTCATTTGGGT T
	P [-] -F	5'- GAAGTGGAAAAGCAGAAAATGCATAAACAGTCATCATCTTCT CAAGGTACACA GGAGCCAGCAGGGAAG
	P [-] -R	5'- TTCCCCTCCCTGCTGGCTCCTGTGTACCTTGAGAAGATGAT GACTGTTTAT GCATTTTCTGCTTTTCCACTTC
	C-term-F	5'-GATGAGCCAGCAGGGAAGCGGGAAAGGAGAGATAAG
FAM172A[P+]	N-term-R	5'- ATGCATTTTCTGCTTTTCCACTTCTATATAGTTTTTCATTTGGGT T
	P [+] -F	5'- GAAGTGGAAAAGCAGAAAATGCATAAACAGGACGATGATGA CGATGGTGACG ATGAGCCAGCAGGGAAG
	P [+] -R	5'- TTCCCCTCCCTGCTGGCTCATCGTCACCATCGTCATCATCG TCCTGTTTAT GCATTTTCTGCTTTTCCACTTC
	C-term-F	5'-GATGAGCCAGCAGGGAAGCGGGAAAGGAGAGATAAG
FLAGAGO1[W T]	3xFLAG-F	5'- TACGACTCACTATAGGGAGACCCAATGGACTACAAAGACCAT GACG
	3xFLAG-R	5'-GGGTCCCCTCCTTGTCATCGTCATCCTTGTAATC
	Ago1-F	5'-GACGATGACAAGGAAGCGGGACCCTCG
	Ago1-R	5'- GGTGACACTATAGAATAGGGCCCTTCAAGCGAAGTACATGGT GC
	3xFLAG-F	5'- TACGACTCACTATAGGGAGACCCAATGGACTACAAAGACCAT GACG
FLAGAGO2[W T]	3xFLAG-R	5'-CTGGAATGGGTGCTTGTCATCGTCATCCTTGTAATC

		Ago2-F	5'-GACGATGACAAGCACCCATTCCAGTGGTGTAAC
		Ago2-R	5'-CTTTTTTGGATCAGCAAAGTACATGGTGCGC
	FLAGAGO2[+N LS]	+NLS-F	5'- ATGTACTTTGCTGGTGGTGATCCAAAAAAGAAGAGAAAGGTA GATCCAAAAA AGAAGAGAAAGGTAGATCCAAAAAAGAAGAGAAAGGTATGA AGGGCCCTATTCTATAGTGTACC
		+NLS-R	5'- GGTGACACTATAGAATAGGGCCCTTCATACCTTTCTCTTCTTT TTTGGATCTAC CTTCTCTTCTTTTTTGGATCTACCTTTCTCTTCTTTTTTGGATC ACCACCAGCAAAGTACAT
qPCR	<i>Cd44_vE8-9</i>	F	5'-TACCCAGTTTTTCTGGATCAGG
		R	5'-GCCATCCTGGTGGTTGTCTG
	<i>Cd44_vE13-14</i>	F	5'-TGGAAGACTTGAACAGGACAGG
		R	5'-GTTTTCGTCTTCTTCCGGCTC
	<i>Cd44_cE4-5</i>	F	5'-ACAGACCTACCCAATTCCTTCG
		R	5'-GGGTGCTCTTCTCGATGGTG

Table 3.2: List of antibodies used in this study

	Antibody Name	Dilution	Source	RRID
Primary antibody	Rabbit polyclonal Anti-FAM172A	WB (1:1000) IF (1:500) IP (2 µg/ml)	Abcam ab1213	64 AB_11127114
	Rabbit polyclonal Anti-AGO2	WB (1:1000) IF (1:200) IP (3 µg/ml)	Abcam ab32381	AB_867543
	Mouse monoclonal Anti-AGO2	WB (1:1000) IF (1:200) IP (2µg/ml)	Abcam ab57113	AB_2230916
	Rat monoclonal Anti-AGO1	WB (1:1000) IF (1:500) IP (3 µg/ml)	Sigma SAB4200084	AB_10602786
	Rabbit polyclonal Anti-Ki67	IF (1:1000)	Abcam ab15580	AB_443209
	Mouse monoclonal Anti-MYC tag	WB (1:500) IF (1:100)	In house hybridoma (9E10)	-
	Goat polyclonal Anti-MYC tag	WB (1:1000)	Abcam ab9132	AB_307033
	Mouse monoclonal Anti-FLAG tag	WB (1:1000) IP (2µg/ml)	Sigma F1804	AB_262044
	Rat monoclonal Anti-MBP tag	WB (1:1000) IP (2µg/ml)	Covance Research Products MRT-131P	AB_10720557
	Mouse monoclonal Anti-GAPDH	WB (1:3000)	Santa Cruz Biotech sc- 32233	AB_627679
	Rabbit polyclonal Anti-GFP	WB (1:3000) IF (1:500)	Abcam ab290	AB_303395
	Rabbit polyclonal Anti-lamin A/C	WB (1:1000)	Abcam ab227176	-
	Rabbit Polyclonal Anti-Nucleolin	WB (1:1000) IP (2µg/ml)	Abcam ab50279	AB_881762

Secondary antibody	Donkey Alexa Fluor 488 Anti-Rat IgG	IF (1:500)	Jackson ImmunoResearch 712-545-150	AB_2340683
	Donkey Alexa Fluor 594 Anti-Rabbit IgG	IF (1:500)	Jackson ImmunoResearch 711-585-152	AB_2340621
	Donkey Alexa Fluor 647 Anti-Mouse IgG	IF (1:500)	Jackson ImmunoResearch 715-605-150	AB_2340862
	Goat Anti-Rat IgG HRP	WB (1:10000)	Santa Cruz Sc-2032	AB_631755
	Horse Anti-Mouse IgG HRP	WB (1:2000)	Cell signaling 7076	AB_330924
	Goat Anti-Rabbit IgG HRP	WB (1:500)	Abcam 6721	AB_955447
	Rabbit Anti-Goat IgG HRP	WB (1:25000)	Bio-Rad 1721034	AB_11125144

## CHAPITRE IV

### DISSCUSION GÉNÉRALE

#### 4.1 L'import nucléaire d'AGO2

Afin d'en apprendre davantage sur le rôle potentiel de FAM172A dans la régulation de l'épissage alternatif, nous nous sommes intéressés à son interaction physique avec la protéine AGO2. Cette interaction est déstabilisée lorsqu'une version de FAM172A mimant un variant présent chez des patients CHARGE est utilisée (pE229Q), ce qui rend cette interaction intéressante à étudier. AGO2 est principalement connu pour son rôle cytoplasmique dans la voie des microARN (Nazer et al., 2022), pourtant des études plus récentes lui associent également un rôle important dans la régulation de l'épissage alternatif co-transcriptionnel dans le noyau cellulaire (Ameyar-Zazoua et al., 2012). Au vu de ses différentes fonctions, AGO2 est connu pour être très dynamique entre les deux compartiments (Kalantari et al., 2016; Ohrt et al., 2008; Schraivogel & Meister, 2014). Les mécanismes qui permettent à AGO2 de voyager entre les deux compartiments sont longtemps demeurés inconnus car elle ne possède pas de NLS ou de NES. Des études précédentes ont démontré que plusieurs importines étaient impliquées dans la translocation d'AGO2 dans le noyau, comme par exemple l'importine-8 (Schraivogel et al., 2015a; Weinmann et al., 2009). Mais aucune protéine avec un NLS, classique ou non-classique, faisant le lien entre AGO2 et les importines n'avait été décrite à ce jour. Les meilleures candidates pour ce rôle auraient été les protéines TNRC6 fortement associées à AGO2 et au complexe RISC et possédant les deux domaines : NLS et NES, pourtant leur implication est limitée à l'export d'AGO2

vers le cytoplasme (Nishi et al., 2013; Schraivogel et al., 2015a). Plus encore, la triple déplétion de *TNRC6A/B/C* favorise la localisation nucléaire d'AGO2 dans ces cellules humaines, montrant l'importance de ces protéines pour retenir ou transporter AGO2 au cytoplasme (Johnson et al., 2024). Nos résultats ont démontré que la protéine FAM172A, qui contient un NLS, participe activement à l'import nucléaire d'AGO2 de manière directe (Figure 4.1 et Article II).

Premièrement, nous avons mis en évidence un défaut dans la localisation nucléaire d'AGO2 dans des fibroblastes embryonnaires en culture provenant d'embryons *Fam172a<sup>Tp/Tp</sup>* (Figure 3.1). A noter que cette localisation nucléaire est variable selon les conditions de culture cellulaire et plus précisément selon la confluence des cellules. D'autres études ont également démontré que la translocation d'AGO2 dans le noyau et sa répartition entre les deux compartiments varie beaucoup d'un type cellulaire à l'autre. Par exemple, AGO2 sera majoritairement nucléaire dans des cultures de cellules primaires, alors qu'elle sera plutôt cytoplasmique dans des lignées cellulaires immortalisées (Sarshad et al., 2018; Sharma et al., 2016). En cohérence avec nos résultats, les conditions qui influent sur cette répartition d'AGO2 sont notamment le stress cellulaire, la différenciation et l'activation de la sénescence (Benhamed et al., 2012; Castanotto et al., 2018; Rentschler et al., 2018; Sarshad et al., 2018). Une autre étude plus récente a confirmé l'impact de la confluence cellulaire sur la localisation subcellulaire d'AGO2 en montrant une différence de cette répartition en faveur du noyau au sein de culture en deux ou trois dimensions (Johnson et al., 2024).

Deuxièmement, nous avons démontré par les méthodes de co-immunoprécipitation *in vitro* et de BiFC (*Bimolecular Fluorescence Complementation*) que les deux protéines pouvaient interagir de manière directe, et ce à travers le domaine PAZ et N-terminal de la protéine AGO2 (Figures 3.3 à 3.5). De plus, la mutation pE229Q dans le domaine ARB2 de FAM172A induit une diminution de cette interaction de 90% par la méthode de co-immunoprécipitation et de 50% par la méthode de BiFC, ce qui confirme

également l'implication de ce domaine ARB2. Cette différence de pourcentage dans nos résultats (50% contre 90%) est pertinente car elle met en évidence une des limites du BiFC : l'irréversibilité de l'interaction entre les deux moitiés de la protéine fluorescente Vénus. Cette limite favorise les interactions très dynamiques et peu stables, comme celle de FAM172A-AGO2 dans le cytoplasme, contrairement à leur interaction nucléaire très stable à l'interface chromatine-splicéosome. Nous avons montré dans une étude précédente utilisant les cellules COS7 que FAM172A n'interagissait pas avec la protéine AGO1 endogène. De manière cohérente, FAM172A ne semble pas non plus être impliquée les mécanismes de son transport nucléaire au sein de fibroblastes embryonnaires en culture (Figure 3.2E-F). Pourtant d'autres données montrent que les deux protéines peuvent interagir entre elles mais sous certaines conditions, comme une surexpression d'AGO1 dans des cellules Neuro2a ou une version recombinante d'AGO1 *in vitro* (Figure 3.14). Pour rappel, les protéines Arb1 et Arb2 ont été identifiées comme des protéines interagissant avec l'unique protéine Ago (Ago1) chez *S.pombe*. Plusieurs évènements de duplication ont eu lieu au cours de l'évolution, ce qui a donné naissance à quatre paralogues (de AGO1 à AGO4) chez l'homme. Chacun de ces paralogues possède le potentiel pour interagir avec le domaine ARB2 de FAM172A. Nous avons déjà confirmé l'interaction directe avec AGO1 et AGO2 (co-immunoprécipitation et BiFC), nous pourrions également la tester pour les protéines AGO3 et AGO4, leurs rôles nucléaires étant très peu étudiés chez l'homme.

Enfin, après avoir validé la fonctionnalité de ce domaine, nous avons démontré que FAM172A participe à la translocation nucléaire d'AGO2 grâce à son NLS et à la voie classique des importines  $\alpha/\beta$ 1 (Figures 4.1 et 3.6). Des expériences supplémentaires de co-immunoprécipitation indique également qu'il pourrait s'agir des importines de type  $\alpha$ 1 (Figure 3.13). Néanmoins, une expérience d'imagerie en temps réel inhibant les importines  $\alpha/\beta$ 1 montre que la voie classique n'est probablement pas la seule voie utilisée pour le transport nucléaire de FAM172A-AGO2 (Figure 3.6H) (Wagstaff et al., 2012). Il y a d'autres explications à cette observation : (i) d'autres protéines sont

impliquées dans la translocation de FAM172A-AGO2 au noyau, ou (ii) FAM172A participe à la translocation du complexe en utilisant un autre NLS non-classique n'impliquant pas les importines  $\alpha/\beta$ 1. Une analyse plus approfondie de la séquence protéique de FAM172A nous a permis d'identifier un potentiel NLS non-classique (acides aminés 233 à 239, Figure 3.7) reconnu par les importines-5 (appartenant à la famille des importines  $\beta$ ) (Tableau 1.1). Des résultats préliminaires ont ensuite confirmé une interaction physique entre FAM172A et l'importine-5 par co-immunoprécipitation (Figure 3.12A-B) soutenant ainsi notre deuxième hypothèse. De plus, des immunofluorescences effectuées sur des Neuro2a transfectées avec une version de FAM172A mutée sur son NLS montrent une baisse significative, mais tout de même une présence résiduelle, de la protéine dans le noyau (Figure 3.5C-D). La localisation nucléaire d'AGO2 au sein des fibroblastes embryonnaires *Fam172a<sup>TP/TP</sup>* n'est pas totalement abolie non plus. Ces observations soutiennent l'implication d'un ou de plusieurs autres facteurs dans la translocation nucléaire de FAM172A, d'AGO2 et/ou du complexe FAM172A-AGO2.

FAM172A ne semble pas contenir de domaine de liaison à l'ADN ou à l'ARN, mais pourrait-elle participer à son recrutement à l'interface chromatine-splicéosome à travers d'autres interactions protéiques ? Nous pourrions comparer le niveau de recrutement d'AGO2 au niveau d'exons alternatifs de cibles connues, dans des CCN provenant d'embryons *Toupee* comparé à des CCN sauvages (Ameyar-Zazoua et al., 2012; Bélanger et al., 2018). Pour ce faire, nous pourrions effectuer des CUT&RUN suivi de PCR quantitative (CUT&RUN-qPCR) en utilisant des anticorps reconnaissant spécifiquement la protéine AGO2. Il serait également intéressant de tester l'hypothèse inverse ; est-ce que une fois dans le noyau, AGO2 guide FAM172A jusqu'à l'interface chromatine-splicéosome grâce à ses microARN ?



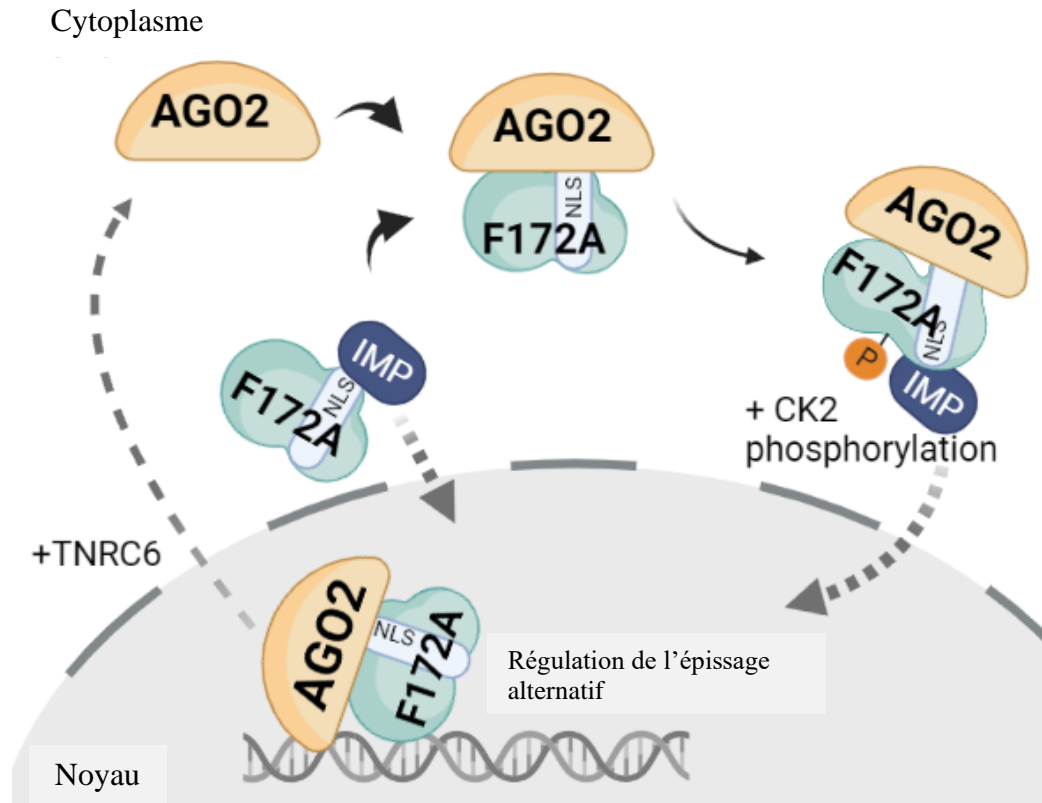


Figure 4.1 L'import nucléaire d'AGO2 médié par FAM172A. Schéma crée sur Biorender

## 4.2 La protéine AGO2 dans le contexte du syndrome CHARGE

Les résultats obtenus précédemment avec les protéines FAM172A et AGO2 suggèrent fortement que cette dernière pourrait également être impliquée dans les mécanismes pathologiques menant au syndrome CHARGE. Pour confirmer cette hypothèse, un modèle de souris KO (*Knock Out*) pour le gène *Ago2* spécifiquement dans les CCN (*Ago2<sup>cKO</sup>; Wnt1cre2*) a été généré. L'allèle est léthal à l'état homozygote, les animaux meurent à la naissance à cause d'une défaillance respiratoire (Annexe A : Figure A.1A). Néanmoins, nous avons pu observer des phénotypes associés au syndrome CHARGE chez des embryons e12,5 à e18,5 et ainsi confirmer l'implication d'AGO2 dans les mécanismes pathologiques menant au syndrome CHARGE (Annexe A : Figures .1 et A.2). Les phénotypes principaux observés sont le colobome oculaire, le retard de croissance, des défauts au niveau des oreilles externes et des nerfs crâniens, ainsi qu'une hypoplasie du thymus (Annexe A : Figures A.1 et A.2). Pour renforcer le lien avec le syndrome CHARGE, l'interaction génétique entre les allèles *Ago2<sup>cKO</sup>* et *Fam172a<sup>Tp</sup>* a été testée. Nous avons observé une aggravation des phénotypes lorsqu'un allèle *Fam172a<sup>Tp</sup>* est ajouté au modèle de souris *Ago2<sup>cKO</sup>* (Annexe A : Figure A.1). De plus, il y a un effet dose-dépendant pour le retard de croissance et le colobome oculaire selon le nombre d'allèles mutés (Annexe A : Figure A.1D-F). Ces résultats confirment le rôle commun de FAM172A et AGO2 dans la pathogénicité menant au syndrome CHARGE.

Connaissant le rôle principale d'AGO2 dans les mécanismes de répression post-transcriptionnel par la voie des microARN, nous nous sommes demandé si ce mécanisme était mis en cause chez le modèle de souris *Ago2<sup>cKO</sup>*. Pour ce faire, nous avons établi une nouvelle lignée de souris au sein desquels les CCN possèdent un rapporteur fluorescent (*Wnt1-Cre2<sup>Tg/+</sup>; Rosa26<sup>FloxSTOP-YFP/+</sup>*) et peuvent être triés par FACS. Grâce à cet outil, nous avons d'abord pu confirmer la déplétion de la protéine

AGO2 dans nos cellules d'intérêts (Annexe A : Figure A.3A). En utilisant un rapporteur luciférase dont l'ARNm possède des séquences MRE reconnues par *let-7* (un ARNm connu pour s'associer à AGO2) transfecté dans ces cellules triées mises en cultures (Johnston et al., 2010), nous avons montré que nos CCN *Ago2<sup>CKO/CKO</sup>* n'ont pas de défauts au niveau de la voie des microARNs (Annexe A : Figure A.3B). Ce résultat peut être expliqué par l'activité redondante des protéines AGO (Su et al., 2009b). En conclusion, la déplétion d'AGO2 spécifiquement dans les cellules de la crête neurale entraîne des phénotypes associés au syndrome CHARGE, et ce à travers un de ses rôles non-canoniques.

L'un de ces rôles est la régulation de l'épissage alternatif (Ameyar-Zazoua et al., 2012; Batsché & Ameyar-Zazoua, 2015; Chu et al., 2021). Afin d'identifier de potentiels défauts au niveau de ce mécanisme, nous avons analysé les événements d'épissage alternatif sur plusieurs cibles connues d'AGO2 : *Mical2*, *Ift74*, *Col5a3* et *Cd44* au sein d'embryons *Ago2<sup>CKO/CKO</sup>* comparés à des embryons contrôles (Annexe A : Figure A.4A). Nous avons en effet observé des dérégulations d'épissage alternatif sur les cibles d'AGO2 dans notre nouvelle lignée de souris *Ago2<sup>CKO/CKO</sup>*. Plus encore, lorsque nous ajoutons un allèle *Fam172a<sup>Tp</sup>* chez les souris *Ago2<sup>CKO</sup>*, nous observons une aggravation des défauts d'épissage sur certaines cibles d'AGO2 : *Mical2*, *Ift74* et *Col5a3* (Annexe A : Figure A.4A). Ces résultats démontrent encore une fois que FAM172A et AGO2 collaborent dans la régulation de l'épissage alternatif co-transcriptionnel, dont la dérégulation pourrait mener au syndrome CHARGE.

### 4.3 Le rôle régulateur de la phosphorylation de FAM172A

Nous savons que FAM172A participe activement à l'entrée d'AGO2 dans le noyau. Connaissant les rôles nucléaires de cette dernière, nous suggérons que l'un des moyens pour FAM172A de réguler l'épissage alternatif co-transcriptionnel est à travers la présence nucléaire d'AGO2, celle-ci étant régulée positivement par la phosphorylation de FAM172A par CK2 (Figure 3.9). Comme expliqué précédemment, la répartition d'AGO2 entre le noyau et le cytoplasme varie beaucoup selon les conditions physiologiques et selon le type cellulaire (Sarshad et al., 2018; Sharma et al., 2016). Ces conditions sont notamment le stress cellulaire, la différenciation et l'activation de la sénescence (Benhamed et al., 2012; Castanotto et al., 2018; Rentschler et al., 2018; Sarshad et al., 2018). Plusieurs études montrent également que la caséine kinase CK2 est régulée par des conditions de stress cellulaire (Ampofo et al., 2013; Sayed et al., 2000; Watabe & Nakaki, 2011), ce qui est cohérent avec les différences de localisation nucléaire d'AGO2 observées. Par exemple, une étude a montré qu'AGO2 est transloqué vers le noyau en étant associée au complexe de réponse au stress SIRC (en anglais *Stress-Induced Response Complex*). De nombreux facteurs appartenant à ce complexe SIRC ont également été retrouvés parmi les interacteurs de FAM172, l'un d'entre eux est la protéine nucléoline (Bélanger et al., 2022). Les interactions physiques entre FAM172A, AGO2 et nucléoline ont été confirmées par co-immunoprécipitation dans des extraits cellulaires (Figure 3.14C). La littérature associe à AGO2 des processus biologiques différents selon si elle se trouve dans le cytoplasme ou le noyau. Par exemple, Sarshad et al, 2018 a montré que la fraction cytoplasmique d'AGO2 joue un rôle notamment dans la prolifération cellulaire, alors que sa fraction nucléaire joue plutôt un rôle dans la différenciation cellulaire. Parallèlement, CK2 est également connue pour réguler des processus associés au développement (Bender et al., 2017; Götz & Montenarh, 2017; Wei et al., 2021). Nous émettons l'hypothèse que l'import nucléaire de FAM172A-AGO2 est stimulé par des conditions de stress, et/ou par une

signalisation induisant la différenciation cellulaire, ce qui favorise certains évènements d'épissage alternatif par AGO2 au niveau de la chromatine (Figure 4.1).

Un autre aspect intéressant à étudier serait la façon dont FAM172A interagit à la fois avec AGO2 et les importines, sachant que son acide aminé en position 229 important pour l'interaction avec AGO2 est situé à proximité du NLS bipartite (Figure 3.7). Des expériences d'immunofluorescence sur *MYC*FAM172A [P-] n'a montré aucun défaut concernant sa localisation nucléaire par rapport à la version sauvage, indiquant que la phosphorylation n'est pas nécessaire pour la translocation de FAM172A au noyau. Pourtant, d'après les expériences de BiFC, le complexe FAM172A-AGO2 n'est pas capable d'aller dans le noyau sans la phosphorylation par CK2 (Figure 3.9). Ces résultats suggèrent que la version non-phosphorylée de FAM172A est reconnue par les importines et peut être importée au noyau ; en revanche lorsque FAM172A interagit avec AGO2, elle n'est plus reconnue par les importines et nécessite un changement de conformation médié par la phosphorylation de CK2 (Figure 4.1). Ce mécanisme semble important pour réguler la présence d'AGO2 dans le noyau et son rôle dans l'épissage alternatif co-transcriptionnel. La version mutée de FAM172A [E229Q] est significativement plus présente au noyau de cellules Neuro2a par rapport à la version sauvage de FAM172A (Figure 3.4C-D). Il y a plusieurs explications possibles à cette observation : (i) le changement de charge dans la séquence protéique de FAM172A induit par la mutation ponctuelle au cœur du NLS influence sa reconnaissance par les importines, (ii) FAM172A n'est plus capable d'interagir avec AGO2 et donc n'est plus séquestré au cytoplasme en attendant la phosphorylation. Nous pourrions éliminer ou confirmer cette deuxième possibilité en effectuant des immunofluorescences sur des fibroblastes embryonnaires provenant d'embryons *Ago2<sup>CKO/CKO</sup>* et d'embryons sauvages, et en quantifiant le ratio de distribution [noyau/cytoplasme] de FAM172A dans ces cellules.

#### 4.4 Distribution génomique de FAM172A et CHD7 dans les CCN

Afin de mieux comprendre le rôle que pourraient jouer FAM172A, CHD7 et AGO2 au niveau de la chromatine, leur distribution sur le génome a été analysé en utilisant la méthode de CUT&RUN. Cette technique a été mis au point récemment et ne demande pas d'étape de réticulation (*Crosslink* en anglais) contrairement à la technique de ChIP (*chromatin immunoprecipitation*) (Meers, Bryson, et al., 2019; Skene et al., 2018; Skene & Henikoff, 2017). Elle a également l'avantage de pouvoir être réalisée sur très peu de cellules et produit un faible taux de bruit de fond, ce qui nous a permis de réaliser l'expérience spécifiquement sur des CCN triées à partir d'embryons e10,5 (~100 000 cellules ou noyaux) (Pilon et al., 2008). Nous identifions 15 035 pics en utilisant un anticorps contre CHD7, alors que seulement 3 242 pics ont été identifiés avec un anticorps contre FAM172A. Soutenant le modèle d'appartenance des deux protéines à un même complexe nucléaire, 95% des pics identifiés pour FAM172A sont communs avec CHD7 (Figure 2.1). Pour le moment, aucun pic n'a été identifié en utilisant un anticorps commercial contre AGO2. Une analyse d'ontologie sur les gènes associés aux pics communs de FAM172A et CHD7 a démontré un enrichissement dans les processus biologiques tels que l'organisation de la chromatine, la régulation positive de la transcription par la RNAPII et la réparation de l'ADN. Des phénotypes associés au développement des CCN ont également été révélés enrichis par notre analyse. D'après l'annotation des pics effectués lors de l'utilisation de l'outil HOMER, 80% des protéines FAM172A liées à la chromatine sont situées aux TSS des gènes, 7% se trouvent aux niveaux d'exons ou de régions UTR, et 12% se trouvent dans les régions introniques ou intragéniques. Les chiffres sont très similaires lorsque nous utilisons les pics communs aux deux protéines. En revanche, seulement 45% des protéines CHD7 se trouvent aux niveaux des TSS. Les autres pics sont associés à 42% aux régions introniques et intergéniques. Ces résultats suggèrent plusieurs rôles à CHD7, l'un de ses rôles serait en commun avec FAM172A aux promoteurs/TSS (Figure 2.1). De plus,

nous avons confirmé la corrélation entre les sites de liaison de FAM172A/CHD7, une chromatine ouverte, et la triméthylation de H3K4, une modification post-traductionnelle spécifique des promoteurs actifs (Figure 2.3).

#### 4.4.1 Spécificité des cellules de la crête neurale

Contrairement à nos résultats, la littérature situe CHD7 préférentiellement dans les régions intergéniques et plus précisément aux niveaux de régions amplificatrices (He et al., 2016; N. C. Reddy et al., 2021; Schnetz et al., 2009a, 2010). Cependant, les proportions d'annotation associées aux amplificateurs et promoteurs varient beaucoup d'une étude à l'autre et d'un type cellulaire à l'autre. Par exemple, seulement 10-15% des protéines CHD7 identifiées par ChIP-seq sont situées aux niveaux des promoteurs dans des cellules souches embryonnaires, des progéniteurs neuraux, des CCN et neuroépithéliale en culture (différenciation effectué *in vitro*), et dans la lignée cellulaire SY5Y (dérivée d'un neuroblastome chez l'homme) (Bajpai et al., 2010; Chai et al., 2018; Schnetz et al., 2009a, 2010). Cette proportion augmente jusqu'à 25% en moyenne lorsque l'étude est effectuée sur des tissus murins *in vivo* tel que le tissu cardiaque au 10,5<sup>ème</sup> jour embryonnaire (e10,5) ou le cervelet au 4<sup>ème</sup> jour post-natal (N. C. Reddy et al., 2021; S. Yan et al., 2020). De plus, une étude effectuée sur des cellules progénitrices d'oligodendrocytes (CPO) et sur des oligodendrocytes (OL) triés à partir de cerveaux murins à P7 confirme une différence de la distribution génomique de CHD7 selon le type cellulaire (Marie et al., 2018). Cette étude montre que 55% des protéines CHD7 sont situées au niveau de promoteurs dans les CPO, contre seulement 15% dans les OL (Marie et al., 2018, Figures 1 et S2). La différence de proportion observée dans nos résultats en comparaison à la littérature pourrait également être due au type cellulaire utilisé, le recrutement de CHD7 aux promoteurs étant déjà augmenté dans les tissus *in vivo*. À la lumière de ces observations, nous pourrions préciser notre analyse sur la distribution génomique de CHD7. Premièrement, afin de confirmer que les pics associés aux régions intergéniques et introniques corrélerent avec les régions

amplificatrices du génome, un CUT&RUN pourrait être effectué en utilisant des anticorps contre la monométhylation de H3K4 (H3K4me1) et l'acétylation de H3K27 (H3K27ac). Cette première marque tous les amplificateurs spécifiques du type cellulaire utilisée, alors que l'association des deux marques d'histones identifiera uniquement les amplificateurs actifs. Dans un second temps, nous pourrions étudier la raison pour laquelle la protéine CHD7 est davantage recrutée aux promoteurs dans nos CCN. FAM172A pourrait-il être impliqué dans le recrutement de CHD7 spécifiquement à ces régions ? CHD7 partage 95 % des pics de FAM172A, alors que FAM172A partage uniquement 20 % des pics de CHD7 (Figure 2.1). Cette observation suggère que FAM172A a de plus fortes probabilités d'être impliqué dans le recrutement de CHD7 plutôt que le contraire. De plus, nous savons que l'expression de *Fam172a* est augmentée d'un facteur de 2 dans un tissu crânial en comparaison au reste de l'embryon à e13,5 et e15,5 (Bélanger et al., 2018). Il serait intéressant d'analyser quantitativement l'expression de *Fam172a* spécifiquement dans nos CCN triées en comparaison à d'autres types de cellules, puis de tester le recrutement de CHD7 au niveau des promoteurs dans des CCN provenant d'embryons *Toupee*. Il serait difficile en revanche d'effectuer une analyse sur le niveau protéique de FAM172A ou CHD7 (*western blot* en anglais) au vu de la faible quantité de CCN triées récoltées. Il faudrait également confirmer un niveau d'expression inchangé de *Chd7* dans les CCN *Toupee* et sauvages, d'autant qu'une étude antérieure a montré une baisse d'expression de *FAM172A* chez des cellules dérivées de patients CHARGE possédant une mutation sur le gène *CHD7* (Bélanger et al., 2018).

#### 4.4.2 Le complexe FAM172A-CHD7-AGO2

Nous savons que le complexe formé des protéines FAM172A, CHD7 et AGO2 se trouve à l'interface entre la chromatine et le spliceosome (Bélanger et al., 2018; Bérubé-Simard & Pilon, 2019). Des expériences antérieures de ChIP-qPCR et RIP-qPCR ont également démontré un enrichissement de recrutement de ces protéines à



certain exons alternatifs en comparaison aux exons constitutifs (gènes *CD44* par exemple) (Bélangier et al., 2018, 2022). Cependant, dans notre étude, un faible pourcentage (3-5%) de pics a été associé aux exons (Figure 2.1). Aucun pic associé à FAM172A ou CHD7 n'a été identifié sur le locus du gène *Cd44*. Plusieurs raisons peuvent expliquer cette différence de résultats. Par exemple, l'étape de réticulation du ChIP est connue pour favoriser l'identification de complexes moins stables et d'interactions physiques plus éloignées dans l'espace en comparaison au CUT&RUN. Une possibilité serait que le recrutement de FAM172A-AGO2 aux exons alternatifs soit très dynamique, en cohérence avec l'import nucléaire d'AGO2 (Article II). La seconde hypothèse serait que le recrutement du complexe aux exons alternatifs serait davantage lié à l'ARN qu'à l'ADN, contrairement au recrutement aux régions promotrices. Sans réticulation, les grands complexes formés majoritairement d'ARN pourraient être déstabilisés lors de nos expériences de CUT&RUN et nous perdriions cette catégorie de pics associés aux exons. De plus, l'étude précédente soutient cette hypothèse en démontrant que les protéines FAM172A et AGO2 ont une meilleure affinité avec l'ARN que l'ADN sur les exons alternatifs du gène *Cd44* (d'un facteur de 2 à 4) (Bélangier et al., 2018). Pour confirmer ces hypothèses, nous pourrions adapter notre protocole en ajoutant une étape de réticulation (Henikoff et al., 2023; Meers, Bryson, et al., 2019), ou encore effectuer une expérience plus adaptée à la récupération des séquences d'ARN, comme le RIP-seq (*RNA Immunoprecipitation followed by Sequencing*) ou le CARIP-seq (*Chromatin-associated RNA immunoprecipitation followed by sequencing*) (Kidder, 2018). Pour pallier à une éventuelle inefficacité des anticorps contre AGO2 lors de nos expériences d'immunoprécipitation, le *greenCUT&RUN* utilisant des étiquettes de protéines fluorescentes a été récemment mis au point et pourrait également être utilisé (Koidl & Timmers, 2021).

## 4.5 Les cibles directes de FAM172A et CHD7 dans les cellules de la crête neurale

### 4.5.1 Régulation de l'épissage alternatif

Dans le but d'identifier les cibles directes de FAM172A et CHD7, nous avons croisé les données génomiques du CUT&RUN avec les données transcriptomiques provenant de CCN *Fam172a<sup>Tp/Tp</sup>* et *Chd7<sup>Gt/+</sup>* (Figure 2.5) (Bélanger et al., 2018). En croisant les quatre groupes de données, nous avons identifié seulement 18 gènes avec des défauts d'épissage alternatif, 10 gènes sous-régulés et 0 gènes sur-régulés (Figure 2.5H). Ce faible taux de corrélation avec les gènes dérégulés est probablement dû au transcriptome provenant des CCN *Chd7<sup>Gt/+</sup>* qui identifie 304 gènes dérégulés (206 sous-régulés et 98 sur-régulés) en utilisant un ratio de 1,2 et un taux de probabilité d'erreur inférieur à 0,05. Ce résultat est par ailleurs cohérent avec d'autres études transcriptomiques qui montrent également un faible taux de dérégulation significative au sein des tissus provenant d'embryons e10,5 mutés pour *Chd7*. Par exemple, une étude effectuée sur des CCN cardiaques triées (*Wnt1-Cre2*) identifie uniquement 6 gènes dérégulés comparés aux CCN sauvages (S. Yan et al., 2020). Une autre étude identifie 63 gènes dérégulés dans des otocystes *Chd7<sup>Gt/+</sup>* (J. Gao et al., 2024). Leur hypothèse est que CHD7 participe à la régulation fine du niveau d'expression des gènes plutôt que d'activer ou de réprimer totalement leur expression (S. Yan et al., 2020). En considérant que nous avons également identifié 267 gènes avec des défauts d'épissage alternatif dans le transcriptome des CCN *Chd7<sup>Gt/+</sup>*, une autre hypothèse serait que CHD7 participe davantage à la régulation de l'épissage alternatif que ce qui est actuellement décrit. De plus, l'étude *Bélanger et al.* suggère que les défauts d'épissage alternatif co-transcriptionnel pourraient être une cause commune à tous les cas du syndrome CHARGE quel que soit le gène en cause (Bélanger et al., 2018; Bérubé-Simard & Pilon, 2019).

Afin de mieux comprendre le rôle que pourraient jouer FAM172A et CHD7 dans la régulation de l'épissage alternatif, nous avons décidé d'étudier l'organisation des speckles nucléaires. Les SNs sont des condensats protéiques localisés proches des sites de transcription, constitués essentiellement de facteurs d'épissage appartenant au splicéosome et participant activement à la régulation de l'épissage alternatif (Galganski et al., 2017; Ilik & Aktaş, 2021). Selon le modèle actuel, la bonne organisation de ces facteurs en NS facilite leur recrutement aux sites de transcription et sur l'ARNm. La proximité dans l'espace des SNs avec les sites de transcription corrèle positivement avec le niveau de transcription et l'efficacité d'épissage des gènes associés (Bhat et al., 2023). Des expériences d'immunofluorescence utilisant des marqueurs de SN, les protéines SC35 (reconnait la protéine SRRM2) et SON, ont été effectuées sur des coupes d'embryons e10,5. Nos résultats *in vivo* ont démontré que la formation des SNs est perturbée dans les cellules provenant d'embryons *Fam172a<sup>Tp/Tp</sup>* et *Chd7<sup>Gt/+</sup>* comparé aux cellules provenant d'embryons sauvages (Figures 2.8 et 2.9). En effet, la fluorescence qui marque les SNs est diminuée au niveau de l'intensité, du nombre et de la taille alors que le niveau d'expression de SRRM2, SON et d'autres facteurs d'épissage est inchangé (Figures 2.8 et 2.9). Une étude a montré que la déplétion de ces deux protéines au sein d'une cellule empêche la formation des SNs (Ilik et al., 2020), ce qui rend SRRM2 et SON essentielles à leur formation (Xu et al., 2022; M. Zhang et al., 2024). Nos résultats suggèrent que la baisse de fluorescence est due à une diffusion du signal dans le nucléoplasme plutôt qu'à une baisse d'expression (Kurogi et al., 2014) et que celle-ci pourrait concerner plusieurs autres facteurs d'épissage autres que seulement SRRM2 et SON. La désorganisation de ces facteurs appartenant majoritairement au splicéosome nuirait à l'efficacité d'épissage et mènerait à de mauvaises décisions d'épissage alternatif et à un défaut de transcription. Afin de confirmer un défaut de recrutement (positif ou négatif) de ces facteurs aux sites d'épissage de gènes cibles (avec des défauts d'épissage alternatif), nous pourrions effectuer des CUT&RUN suivi de PCR quantitatifs (CUT&RUN-qPCR) dans des CCN *Fam172a<sup>Tp/Tp</sup>*, *Chd7<sup>Gt/+</sup>* ou sauvages, et en utilisant des anticorps reconnaissant

certaines sous-unités du spliceosome, de protéines SR ou de protéines hnRNP appartenant à l'interactome de FAM172A.

Il reste également à déterminer de quelle manière FAM172A et CHD7 participent à l'organisation des SNs. Nous savons que ces deux protéines sont localisées au niveau des promoteurs/TSS. Nous savons également que l'interaction entre FAM172A et SC35 (SRRM2) est indirecte et qu'elle a lieu au niveau de la chromatine (sensible au traitement DNase). De plus, la structure très organisée de FAM172A et sa localisation nucléaire très différente de celle des SNs éliminent la possibilité que FAM172A participe directement à la formation des condensats associés aux SNs par la séparation de phase liquide-liquide. Nous émettons l'hypothèse que FAM172A et CHD7 participent au recrutement de certains composants du spliceosome et à l'assemblage de celui-ci au niveau des promoteurs/TSS. Nous approfondirons davantage cette hypothèse dans la partie suivante (4.6).

#### 4.5.2 FAM172A régule des gènes impliqués dans la prolifération, la différenciation et la migration des CCN

Étant donnée le faible taux de corrélation entre le transcriptome des CCN *Chd7<sup>Gt/+</sup>* et les autres ensembles de données, nous avons décidé de continuer l'analyse en utilisant uniquement le transcriptome des CCN *Fam172a<sup>Tp/Tp</sup>*. Une analyse d'ontologie sur les cibles directes de FAM172A a montré des enrichissements différents selon la catégorie de dérégulation : cibles sous-régulées, cibles sur-régulées et cibles ayant de défauts d'épissage alternatif (Figure 2.6). Les processus biologiques enrichis associées aux gènes sous-régulées concernent le développement, nous pouvons notamment citer le développement du système nerveux, la régulation positive de la différenciation des neurones, la régulation de la migration et de l'ossification. Ces processus sont en cohérence avec plusieurs phénotypes du syndrome CHARGE. Alors que les processus biologiques associées aux cibles sur-régulées concernent la traduction, la prolifération

et la réponse au stress, nous pouvons citer la régulation positive du cycle cellulaire, la réponse cellulaire au UV et au stress chimique (Figure 2.6). La protéine FAM172A semble ainsi participer à plusieurs processus biologiques important pour la cellule, au niveau de la chromatine elle régulerait positivement les gènes développementaux et inhiberait la prolifération cellulaire. Une autre étude montre le même profil d'enrichissement pour la protéine CHD7, ce qui montre l'importance de ces processus dans le contexte du syndrome CHARGE (Marie et al., 2018). Les cibles anormalement épissées sont enrichies en processus biologiques associés à la biogénèse et à la modification des protéines (ubiquitination), à la régulation de voies de signalisation et à la régulation du cytosquelette. Ces mêmes processus biologiques sont également enrichis en utilisant les gènes anormalement épissés des CCN *Chd7<sup>Gt/+</sup>* (Figure 2.7). Il reste néanmoins à déterminer comment les défauts d'expression et les défauts d'épissage alternatif sont reliés. S'agit-il d'une anomalie dans un même mécanisme d'action de la protéine FAM172A et/ou de CHD7, ou s'agit-il de plusieurs mécanismes d'action différents ?

#### 4.5.3 Bivalence des promoteurs

En utilisant les modifications post-traductionnelles de l'histone 3, nous avons identifié les promoteurs actifs (H3K4me3), les promoteurs bivalents (H3K4me3 et H3K27me3) et les promoteurs réprimés (H3K27me3) dans les CCN à e10,5 (Figure 2.3). Les protéines FAM172A et CHD7 lient les promoteurs actifs à 87% et les promoteurs bivalents à 10%. La marque épigénétique H3K4me3 est spécifique d'une chromatine ouverte et d'une transcription active (en l'absence de H3K27me3). Son niveau d'enrichissement aux promoteurs corrèle directement avec la présence d'une RNAPII active et d'un niveau de transcription élevé (Beacon et al., 2021). Elle a également été associée à la régulation d'épissage alternatif de certains gènes (Davie et al., 2016; Khan et al., 2016). La marque épigénétique H3K27me3 est quant à elle spécifique de l'hétérochromatine facultative. Ce type d'hétérochromatine est associé au complexe

PRC (*Polycomb repressive complex*) qui réprime beaucoup de gènes tissu-spécifiques, notamment impliqués dans le développement et dans la différenciation cellulaire (Margueron & Reinberg, 2011), contrairement à l'hétérochromatine constitutive marquée par H3K9me3 qui réprime essentiellement les éléments répétés, les centromères et les télomères. Les promoteurs bivalents caractérisés par un co-enrichissement de H3K4me3 et H3K27me3 ne présentent pas de RNAPII engagée dans la transcription, mais plutôt une RNAPII prête à s'engager (Baudre et al., 2022). Cette bivalence est un moyen transitoire de réprimer les gènes afin de pouvoir rapidement les activer ou les réprimer selon les besoins développementaux de la cellule (Bernstein et al., 2006). Dans nos résultats, nous avons observé une plus grande corrélation entre les promoteurs actifs et les cibles sur-régulées de FAM172A (54%), alors que les promoteurs bivalents corrélaient mieux avec les cibles sous-régulées (54%) (Figure 2.5B). Sachant que les cibles sous-régulées de FAM172A sont essentiellement associées au développement comme discuté précédemment, il n'est pas surprenant qu'ils possèdent davantage de marques H3K27me3 au niveau de leur promoteur. Les autres gènes, comme les gènes de ménage par exemple, seront plutôt réprimés grâce à la méthylation de l'ADN au niveau de séquences riches en CG dans leur promoteur. De plus, il a été montré que les gènes réprimés par H3K27me3 possèdent des promoteurs pauvres en CG (Baudre et al., 2022). Les cibles directes de FAM172A corrélaient très peu avec les gènes réprimés uniquement par H3K27me3 (environ 2%) (Figure 2.3C). Ces résultats démontrent que FAM172A corréla principalement avec la marque épigénétique H3K4me3 (promoteurs actifs ou bivalents), suggérant qu'elle ne serait pas associée aux complexes PRC, mais qu'elle semble plutôt travailler de concert avec la machinerie transcriptionnelle quel que soit le défaut d'expression (augmenté ou diminué) et plus probablement avec la RNAPII.

#### 4.6 Le rôle potentiel de FAM172A et CHD7 dans l'état de pause de l'ARN polymérase II

Des études récentes utilisant le CUT&RUN, le CUT&Tag, le CUTAC (*Cleavage Under Targeted Accessible Chromatin*) (technique adaptée du CUT&RUN) et des anticorps contre la RNAPII ont identifié un enrichissement génomique très similaires à ceux de FAM172A et CHD7 notamment aux sites d'initiation de la transcription (Brahma & Henikoff, 2024; Henikoff et al., 2023; Miura & Chen, 2020). En revanche, FAM172A et CHD7 ne sont pas enrichies au niveau des sites de terminaison de la transcription, contrairement à la RNAPII. Ces deux protéines participeraient donc aux premières étapes de la transcription et ne suivraient pas la RNAPII tout au long du processus de transcription. Pour mieux comprendre le lien entre nos protéines et la RNAPII, nous avons étudié l'interactome de FAM172A obtenu par des expériences de Co-immunoprécipitation et de Bio-ID (anciennement publié et nouvellement obtenu grâce à une collaboration avec le Pr E.Coyaud) (Bélangier et al., 2018). Parmi les interacteurs associés à la machinerie transcriptionnelle, plusieurs protéines appartiennent au PIC et au complexe d'élongation de la transcription (Annexe A: tableau A.1). Les facteurs POLR2M (*RNA polymerase II subunit M*), et LARP7 (*Lar-related protein 7*) connus pour stabiliser l'état de pause de la RNAPII ont également été identifiés (Annexe A: tableau A.1). Ces résultats sont très pertinents, considérant l'importance du recrutement des facteurs d'épissage dans la régulation de l'état de pause de la RNAPII, ayant lieu à 30-50 nucléotide en aval du TSS (Caizzi et al., 2021). De plus, ce phénomène concerne environ 40% des gènes chez les mammifères et favorise les gènes associés au développement (Abuhashem et al., 2022; Gaertner & Zeitlinger, 2014). Le modèle actuel décrit cette « pause » comme un point de contrôle avant de poursuivre l'élongation de la transcription, le facteur p-TEFb étant un élément clé pour libérer la RNAPII de son état de pause (voir introduction 1.3.1).

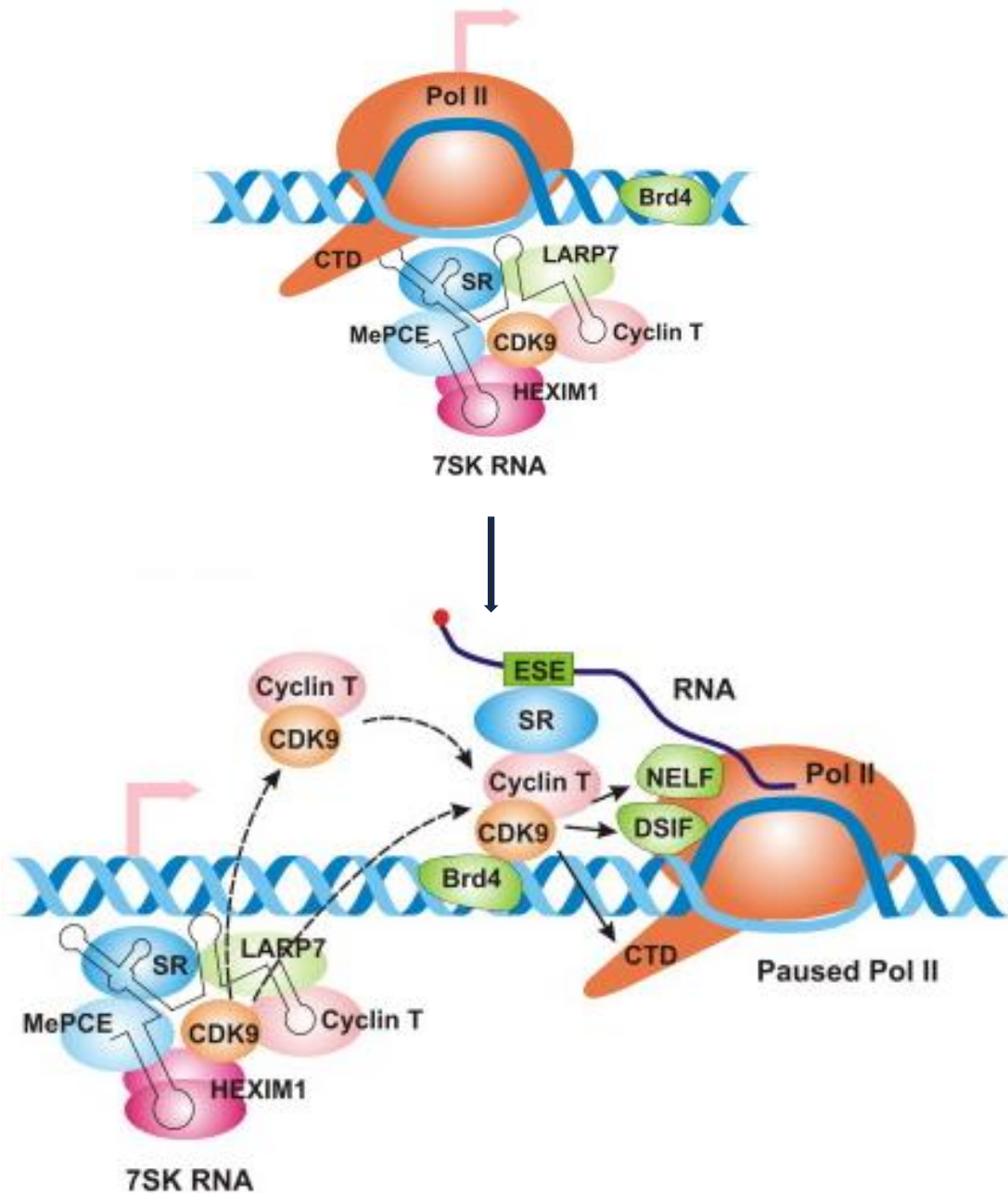


Figure 4.2 Les protéines SR collaborent avec le complexe 7SK pour libérer l'ARN polymérase II de l'état de pause au niveau en aval du promoteur. CDK9 et Cyclin T est des sous-unités de p-TEFb. Schéma adapté de *Ji et al, 2013*



Des études ont montré qu'une inhibition de l'activité de snRNP U2 engendre un défaut de recrutement de p-TEFb et augmente la durée de pause de la RNAPII en aval du promoteur (Caizzi et al., 2021; Y. W. Fong & Zhou, 2001). Une autre étude montre que la protéine SRSF2 est importante pour la phosphorylation de la RNAPII par p-TEFb (Ji et al., 2013). En effet, 60% de SRSF2 est présente aux promoteurs actifs des gènes ainsi qu'au sein de complexes nommés 7SK où p-TEFb est séquestré. Le recrutement de SRSF2 par l'ARN naissant libère p-TEFb du complexe 7SK et permet son recrutement actif au site de pause de la RNAPII (Figure 4.2). Le complexe 7SK est également composé de SRSF1 et LARP7, deux protéines identifiées dans l'interactome de FAM172A (Annexe A: tableau A.1) (Ji et al., 2013). Ces résultats démontrent qu'en plus d'être essentiel tout au long de la transcription, le recrutement de facteurs d'épissage et le bon assemblage du spliceosome au niveau des 30 à 50 premiers nucléotides d'ARN naissant où la RNAPII est en état de pause sont importants pour l'activation de l'élongation.

Étant donnée la désorganisation des SNs dans nos deux modèles de souris, notre hypothèse est que FAM172A et CHD7 régulent l'état de pause de la RNAPII à travers le recrutement de facteurs d'épissage et du bon assemblage du spliceosome au promoteur, la vitesse d'élongation par la RNAPII étant également importante pour les décisions d'épissage alternatif le long des gènes. Une étude soutient notre hypothèse en montrant que l'enrichissement de CHD7 corrèle positivement avec un temps de pause plus court de la RNAPII au promoteur et une augmentation de la vitesse d'élongation (Yao et al., 2020). Une autre étude portant sur la localisation génomique de la RNAPII et utilisant la technique de CUT&RUN a pu identifier deux types de complexes transcriptionnels au niveau des promoteurs/TSS (RNAPII inactivé ou en « pause ») grâce à la taille des fragments séquencés (Figure 4.3) (Miura & Chen, 2020). Afin de confirmer notre hypothèse, nous pourrions d'abord effectuer la même analyse sur nos données génomiques afin de déterminer si FAM172A et CHD7 sont enrichies

au niveau des TSS où la RNAPII n'est pas encore engagé (fragments <120 nucléotides) ou directement en aval du TSS lorsque la RNAPII est en état de pause (fragments >270 nucléotides) (Figure 4.3) (Miura & Chen, 2020). Deuxièmement, nous effectuerions des expériences de CUT&RUN-qPCR avec un anticorps contre la RNAPII dans des CCN *Fam172a*<sup>Tp/Tp</sup>, *Chd7*<sup>Gt/+</sup> ou sauvages pour identifier un éventuel défaut dans la durée de l'état de pause (ratio d'enrichissement TSS/TTS) ou dans la vitesse d'élongation de la transcription sur des loci de gènes cibles. Enfin, nous pourrions déterminer si la régulation de la RNAPII par FAM172A et CHD7 (s'il y a lieu) est effectué à travers p-TEFb en analysant le statut de phosphorylation de la RNAPII. Par exemple, l'utilisation d'un anticorps contre la RNAPII phosphorylée sur la deuxième sérine de son domaine CTD nous donnerait une idée de son niveau de phosphorylation par p-TEFb (ratio ARNPII-S2p/ ARNPII) (Abuhashem et al., 2022; H. Wang et al., 2023). Parmi les cibles directes de FAM172A obtenues à la suite du croisement entre les données de CUT&RUN et des données transcriptomiques des CCN *Fam172a*<sup>Tp/Tp</sup>, nous avons identifié 410 cibles sur-régulées (Figure 2.5A). Il s'agit de la catégorie la plus représentée, pourtant elle n'a aucune cible en commun avec les CCN *Chd7*<sup>Gt/+</sup> (Figure 2.5H). Cette observation et nos résultats actuels ne nous permettent pas de prédire avec certitude si FAM172A et CHD7 régulent la RNAPII de manière positive ou négative. Nous supposons que le type de régulation dépendra de la catégorie de gènes cibles : sous-régulés ou sur-régulés. Ainsi, FAM172A pourrait également participer à maintenir la RNAPII en état de pause, et ce indépendamment de CHD7.

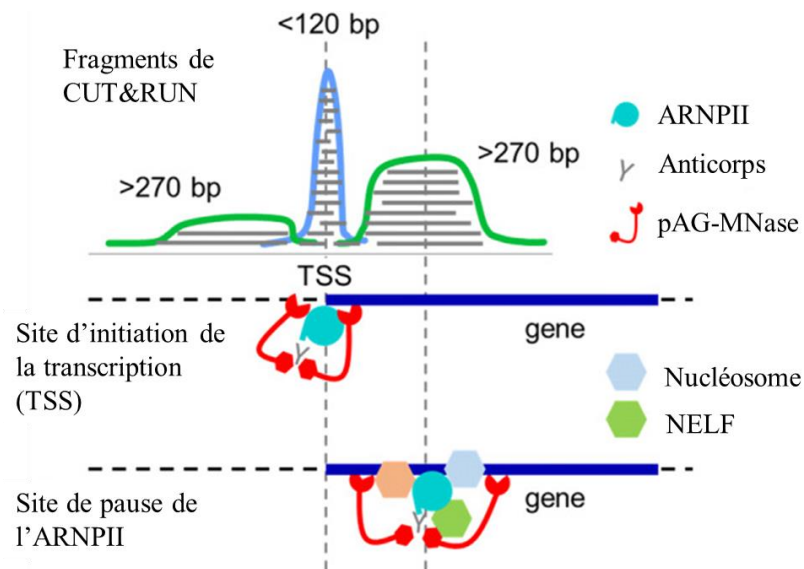


Figure 4.3 La technique de CUT&RUN identifie différents types de fragments au niveau des promoteurs selon la taille. Schéma adapté de *Miura and Chen, 2020*

## 4.7 Les autres rôles potentiels de FAM172A

### 4.7.1 Rôle dans la résolution des R-loops

Au cours de la transcription, les structures constituées d'un hybride ADN-ARN (correspondant au brin d'ADN matrice et à l'ARN naissant et d'un second brin d'ADN) sont nommées R-loops. Ces structures transitoires se forment au niveau du site actif de la RNAPII et ont besoin d'être résolues après son passage. Une résolution prématurée ou déficiente de ces R-loops par des enzymes spécifiques engendre des anomalies au niveau de plusieurs processus et étapes importantes de la transcription. Elles sont en effet connues pour réguler l'état de pause de la RNAPII aux TSS, l'élongation et la terminaison de la transcription (Niehrs & Luke, 2020). Les R-loops anormalement présentes le long des gènes peuvent également engendrer des cassures de l'ADN et de l'instabilité génomique au cours de la réplication (Niehrs & Luke, 2020; Rinaldi et al., 2020). Inversement, des anomalies dans les mécanismes associés à la transcription et à l'épissage peuvent mener à l'apparition de R-loops pathologiques. Par exemple, un défaut dans le processus d'épissage (et plus précisément dans le recrutement de facteurs d'épissage aux jonctions exon-intron) peut empêcher la résolution des R-loops dans le corps des gènes (Tam & Stirling, 2019). L'interactome de FAM172A identifie plusieurs protéines associées à la résolution des R-loops tels que TOP3B (*DNA topoisomerase 3-beta*), RNASEH2C (*Ribonuclease H2 subunit C*) et DHX9 (*ATP-dependent RNA helicase A*) (Annexe A: tableau A.1). Nous savons qu'une des manières de réguler et libérer l'ARPII de son état de pause est la bonne résolution des R-loops au niveau des TSS. Nous pouvons donc imaginer que FAM172A régule l'état de pause de la RNAPII à travers la résolution des R-loops. Afin de confirmer la présence de R-loops pathologiques dans notre contexte, nous avons effectué des immunofluorescences sur des coupes d'embryons e10,5 *Fam172a*<sup>Tr/Tr</sup> ou sauvages en utilisant des anticorps S9.6 et contre H2AX $\gamma$  qui reconnaissent respectivement les

hydrides ADN-ARN et les cassures d'ADN. Les résultats préliminaires montrent une augmentation des foci associés à S9.6 et aux cassures d'ADN dans nos cellules mutantes comparées aux cellules contrôles (Annexe A: tableau A.1). Ces résultats appuient un potentiel rôle de FAM172A dans la résolution des R-loops (Figure 4.4). Pour confirmer ce rôle, nous pourrions effectuer des expériences de DRIP-seq (*DNA-RNA immunoprecipitation followed by sequencing*) et de MapR (*Mapping Native R-Loops Genome-wide Using a Targeted Nuclease Approach*) qui localise les hydrides ADN-ARN sur le génome, puis étudier la corrélation entre l'augmentation des R-loops dans les CCN *Fam172a<sup>Tp/Tp</sup>* et la localisation génomique de FAM172A. La méthode MapR permet d'identifier les R-loops se trouvant particulièrement aux régions promotrices, contrairement au DRIP-seq qui identifie les R-loops dans les régions intergéniques (Q. Yan et al., 2019; Q. Yan & Sarma, 2020). En perspective, il serait intéressant de déterminer (i) si FAM172A participe directement la résolution des R-loops, ce qui en conséquence induit des défauts dans la durée de l'état de pause de la RNAPII, ou (ii) si l'augmentation des R-loops au niveau des TSS et des jonctions exon-intron (s'il y a lieu) est engendrée par l'augmentation de la durée de pause de la RNAPII et par les défauts d'épissage (Figure 4.4).

De plus, des rôles dans la réparation des cassures d'ADN et dans l'homéostasie des R-loops ont été décrits pour les deux interacteurs privilégiés de FAM172A, les protéines AGO2 et CHD7 (M. Gao et al., 2014; Gutbrod & Martienssen, 2020; Hu et al., 2021; Rother et al., 2020).

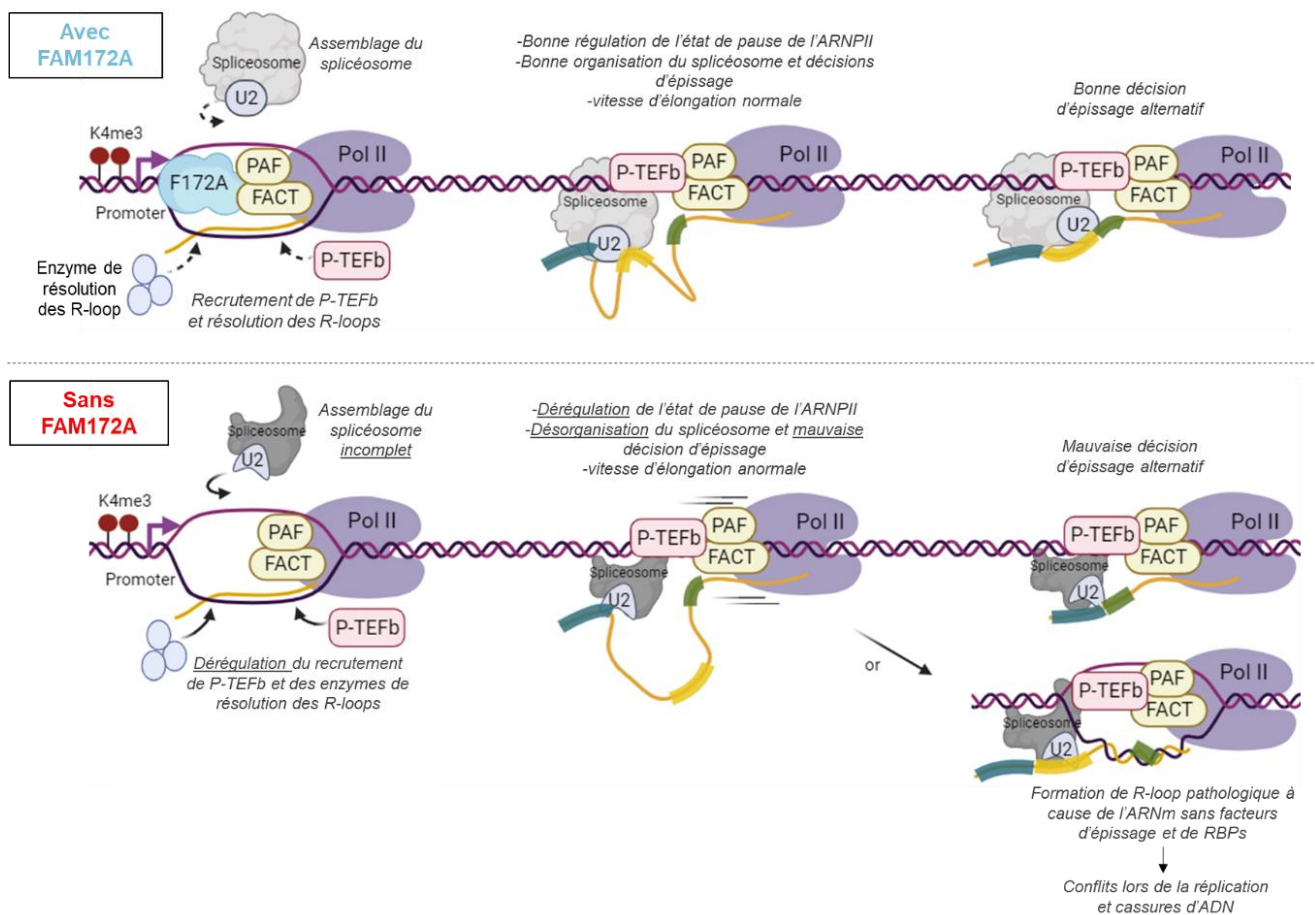


Figure 4.4 Le potentiel mode d'action de FAM172A dans la régulation de l'expression et l'épissage alternatif des gènes. Ce mode d'action dépend de la bonne régulation de l'état de pause de la RNAPII au niveau des promoteurs et de la résolution des R-loops. Schéma créé sur Biorender.

#### 4.7.2 FAM172A et les histones

Grâce aux études effectuées chez la levure sur le domaine ARB2, nous savons que ce domaine permet à la protéine Hda1 de former des homodimères afin de lier directement les nucléosomes *in vitro*, et d'être recruté par la RNAPII au niveau de gènes actifs (Ha et al., 2019; Shen et al., 2016). Pour le moment, nous avons confirmé que FAM172A est recrutée au niveau de gènes actifs (Article I), et qu'elle interagit avec plusieurs composants de la machinerie transcriptionnelle (Annexe A : Tableau A.1). Nous avons également des données de BiFC et de co-immunoprécipitation qui démontrent que FAM172A est capable de former des homodimères, et qu'ils sont localisés majoritairement dans le noyau (Figure 3.13). Finalement, FAM172A n'ayant pas de domaine de liaison à l'ADN ni à l'ARN, nous pourrions nous demander comment FAM172A est recrutée à la chromatine, est-ce à travers une interaction directe avec les nucléosomes comme chez *S. cerevisiae* ? ou est-ce à travers d'autres interactions protéiques ? Pour répondre à ces questions, nous pourrions d'abord vérifier si FAM172A est capable de lier les nucléosomes avec une expérience *in vitro*. Deuxièmement, nous pourrions utiliser l'outil bioinformatique MEME-suite pour identifier d'éventuels motifs enrichis dans les régions génomiques associées à FAM172A et CHD7, et d'identifier un éventuel mode de recrutement.

Des études ont montré que CHD7 pouvait participer au développement de certains tissus avec ou sans son activité dépendante de l'ATP (S. Yan et al., 2020). Dans la même idée, nous pourrions tester l'implication de l'activité de remodelage de la protéine CHD7 dans la régulation de l'épissage alternatif co-transcriptionnel.

## CONCLUSION

Le syndrome CHARGE est une maladie génétique extrêmement complexe et rare, qui touche environ 1 personne sur 10 000 dans le monde. Cette maladie fait partie de la famille des neurocristopathies, soit des maladies causées par des anomalies des cellules de la crête neurale. Les troubles médicaux et physiques qui découlent de ce syndrome varient beaucoup d'un enfant à l'autre, ce qui rend le diagnostic du syndrome très difficile. Dans de nombreux cas, beaucoup de tests et d'analyses sont faits, mais aucun diagnostic n'est posé. Néanmoins, une combinaison de symptômes surreprésentés chez les patients a donné au syndrome son acronyme CHARGE. Il s'agit d'anomalies aux yeux, de problèmes cardiaques, d'une obstruction des voies nasales, d'un retard de croissance et de développement, d'anomalies génitales et d'anomalies aux oreilles. Plusieurs de ces anomalies sont potentiellement léthales, expliquant pourquoi environ 30% des enfants atteints meurent avant l'âge de 5 ans. Il n'y a actuellement aucune possibilité de guérison pour le syndrome CHARGE. Les options thérapeutiques disponibles sont en effet restreintes au traitement individuel des anomalies les plus sévères afin de favoriser la survie de l'enfant atteint.

Une mutation à l'état hétérozygote dans le gène *CHD7* est actuellement la cause principale conduisant au syndrome CHARGE. Ce gène code pour un remodeleur de la chromatine qui intervient notamment dans la régulation de gènes essentiels à la formation et à la migration des cellules de la crête neurale. Les autres causes menant à ce syndrome sont des mutations sur d'autres facteurs de la chromatine et/ou facteurs d'épissage. Le laboratoire du Pr Nicolas Pilon a récemment mis en évidence le gène très peu étudié *FAM172A* comme étant un nouveau gène candidat pour ce syndrome. Le laboratoire a généré un modèle de souris qui possède une mutation sur le gène *FAM172A* et qui récapitule parfaitement les symptômes associés au syndrome CHARGE. Les premières études faites sur ce modèle ont révélé que notre protéine



d'intérêt se trouve majoritairement dans un complexe à l'interface entre la chromatine et le splicéosome avec les protéines CHD7 et AGO2, où elle semble jouer un rôle dans la régulation de l'expression et de l'épissage alternatif des gènes. Mais les mécanismes précis par lesquels elle intervient dans ces processus moléculaires restent à élucider.

Dans le but d'en apprendre davantage sur ces mécanismes, nous nous sommes d'abord intéressés à la localisation génomique de nos protéines d'intérêt. Nous avons découvert que les protéines FAM172A et CHD7 sont recrutées au niveau de promoteurs actifs associés à des gènes importants pour le développement des cellules de la crête neurale. De plus, la désorganisation des facteurs d'épissage en speckles nucléaires au sein de tissu embryonnaires *Fam172a<sup>Tr/Tr</sup>* et *Chd7<sup>Gt/+</sup>* nous a amené à l'hypothèse que nos protéines d'intérêts participent au recrutement ou à l'organisation du splicéosome au niveau de promoteurs de gènes. En s'appuyant sur la liste d'interacteurs de FAM172A, nous suggérons que ce défaut de recrutement est lié à la régulation de l'état de pause de l'ARN polymérase II et à la résolution des R-loops.

Un autre moyen pour FAM172A de réguler l'épissage alternatif co-transcriptionnel est à travers la protéine AGO2. Nous avons découvert que la protéine AGO2 interagit directement avec la protéine FAM172A grâce à un domaine d'interaction nommée ARB2 présent dans cette dernière. La protéine AGO2 est connue pour jouer plusieurs rôles importants dans la cellule, que ce soit la régulation de la voie des microARNs dans le cytoplasme ou la régulation de l'épissage alternatif dans le noyau. Il est également connu qu'AGO2 est très dynamique et se déplace constamment entre ces deux compartiments dépendamment du besoin de la cellule à un temps donné. Les mécanismes qui régulent cette dynamique et la manière dont AGO2 est importé au noyau restaient jusqu'à présent méconnus. Nous avons mis en évidence que la protéine FAM172A est la clé de ce mécanisme. En effet, elle participe activement à l'import nucléaire d'AGO2, et ce à travers son signal de localisation nucléaire bipartite reconnu par la voie classique des importines  $\alpha/\beta$ 1. Nous avons également démontré que cette

translocation nucléaire est stimulée par la phosphorylation de FAM172A par CK2, et diminuée par un variant de FAM172A mimant une mutation identifiée chez des patients CHARGE. Ces résultats de recherche ont tout récemment valu à FAM172A d'être renommée ARB2A (*ARB2 cotranscriptional regulator A*) par le comité international de nomenclature des gènes HUGO pour son rôle clé d'interacteur d'AGO2.

Ces travaux de recherche ouvrent de nouvelles portes et perspectives à la découverte d'un mécanisme pathologique lié spécifiquement au syndrome CHARGE. Toutes ces nouvelles connaissances pourront aider à l'élaboration d'un diagnostic fiable et efficace pour tous les patients, et *in fine*, au développement éventuel d'une thérapie *in utero* à base de petites molécules.

ANNEXE A

FIGURES ET TABLEAU SUPPLÉMENTAIRES

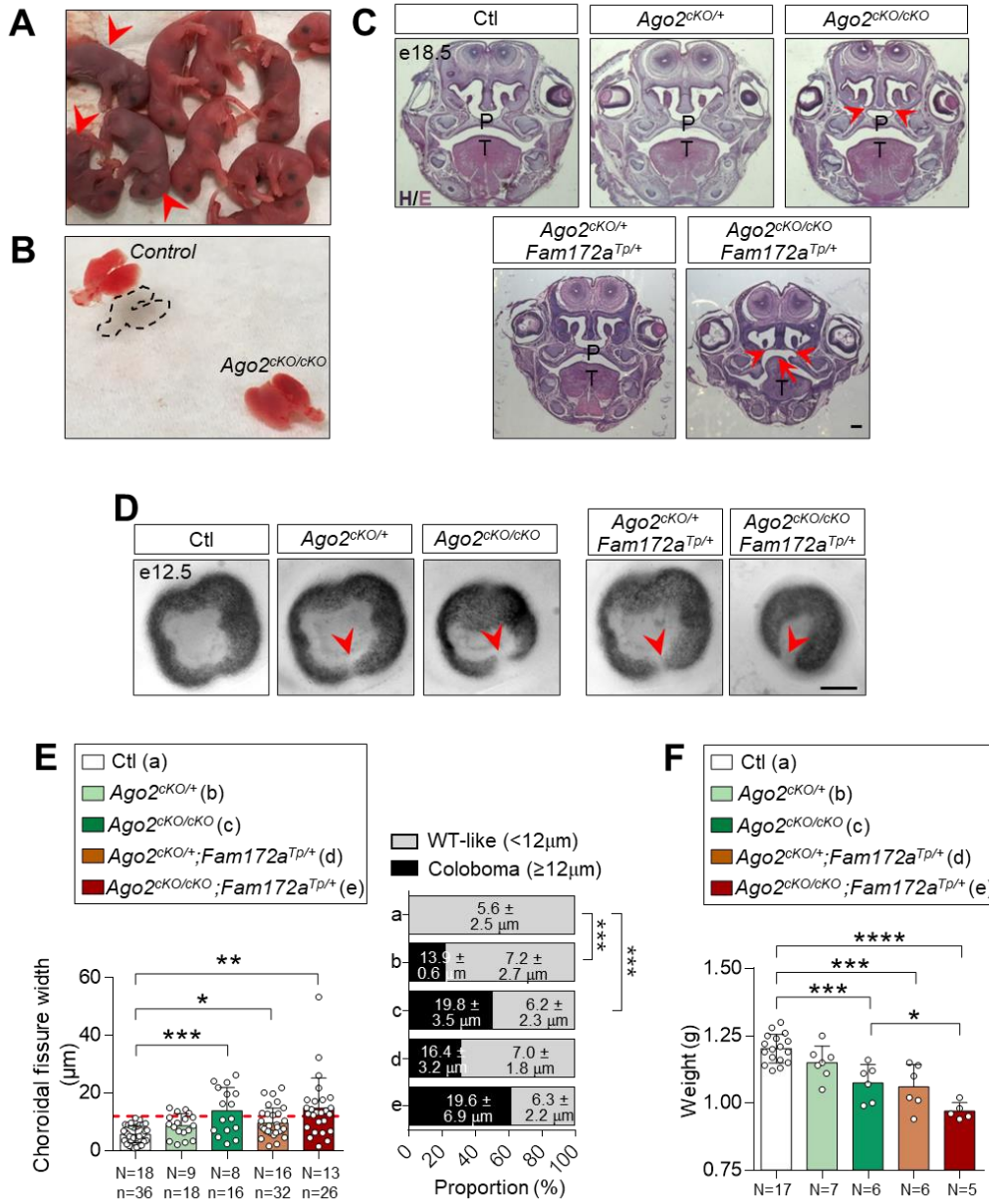


Figure A.1 *Ago2* is implicated in CHARGE Syndrome pathogenesis and genetically interacts with *Fam172a*<sup>Tp</sup>. (A-B) *Ago2*<sup>cKO/cKO</sup> newborns are cyanotic (red arrowheads in A) and lack air in their lungs (see floating test in B, with shadow of floating control lung delineated by dashed outline). (C) H&E-stained cross-sections of e18.5 fetus heads of indicated genotypes, with red arrowheads indicating blockage of nasal passages (choanal atresia) and red arrow indicating cleft palate (N=5 per genotype). P, palate, T, tongue. (D) Bright field view of e12.5 eyes of indicated genotypes, with red arrowheads pointing to unclosed choroidal fissure (coloboma). (E) Quantitative analysis of fissure width (left panel) and percentage of coloboma occurrence (right panel), using images such as those displayed in D. Red dotted line in left panel corresponds to the minimal width for coloboma (12µm; based on the maximum width observed in control embryos). The average width of choroidal fissure for coloboma and WT-like subgroups are indicated in their corresponding bar subdivisions in right panel. (F) Body weight of e18.5 fetuses of indicated genotype.

**A**

Stage	<i>Ago2</i> <sup>+/+</sup>	<i>Ago2</i> <sup>+/H</sup>	<i>Ago2</i> <sup>cKO/+</sup>	<i>Ago2</i> <sup>cKO/cKO</sup>	P value (Khi <sup>2</sup> )
e12.5 (N=62)	16	16	16	14	0.979
e15.5 (N=50)	14	13	13	10	0.868
e18.5 (N=29)	6	10	7	6	0.686
P0 (N=21)	5	6	5	5	0.986
P30 (N=103)	35	38	30	0	9.09E-08

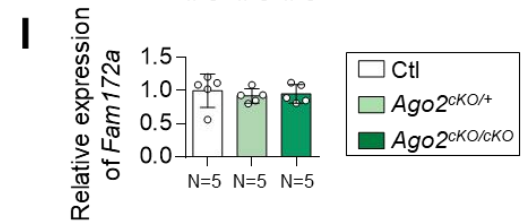
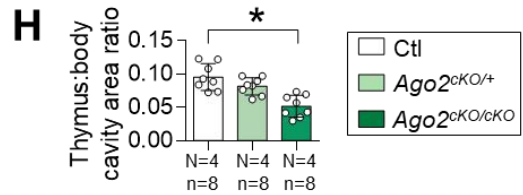
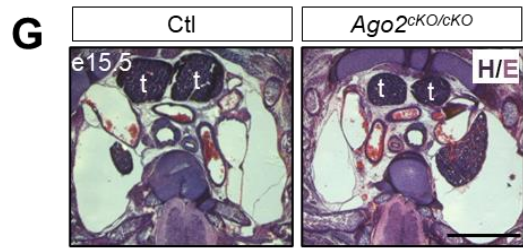
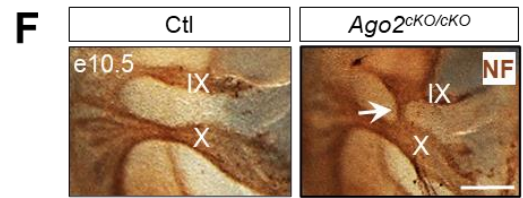
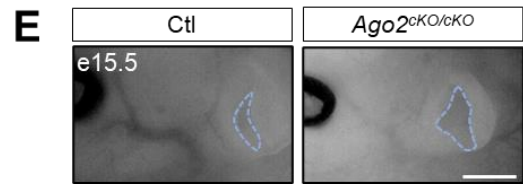
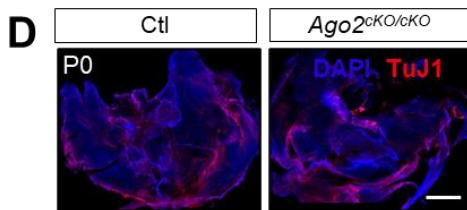
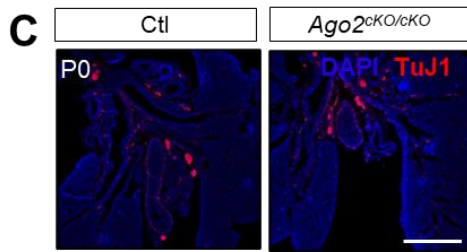
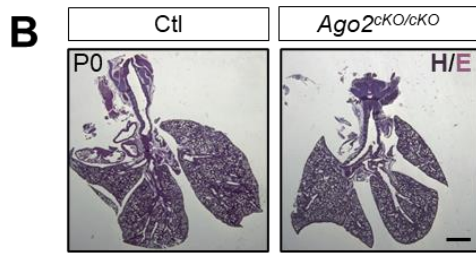


Figure A.2 NCC-specific *Ago2* knockout phenocopies CHARGE syndrome in mice. (A) Table showing the number of animals recovered from *Wnt1-Cre2<sup>Tg/+</sup>;Ago2<sup>Flox/+</sup> X Ago2<sup>Flox/Flox</sup>* crosses as a function of indicated stages and *Ago2* genotypes. (B) H&E-stained section of P0 lungs (N=3 per genotype). (C-D) Immunofluorescence analysis of neuronal class III  $\beta$ -Tubulin (TuJ1) distribution in sections of P0 lungs (C) and whole diaphragms (D), with nuclei stained using DAPI (N=3 per genotype). (E) Bright-field views of e15.5 embryos, with dashed lines delineating the ear openings (N>7 per genotype). (F) Whole-mount staining of cranial nerves in e10.5 embryos using anti-neurofilament immunohistochemistry, with arrow pointing to an abnormal connection between glossopharyngeal (IX) and vagal (X) nerves. (G) H&E-stained cross-sections of e15.5 embryos at the level of thymus (t). (H) Quantitative analysis of thymus:body cavity area ratio, using images such as those displayed in G. (I) *Psm2*-normalized RT-qPCR analysis of *Fam172a* expression in e12.5 embryo heads from *Ago2*-mutant embryos. Scale bar, 1mm. \* $P \leq 0.05$ ; Student's *t*-test.

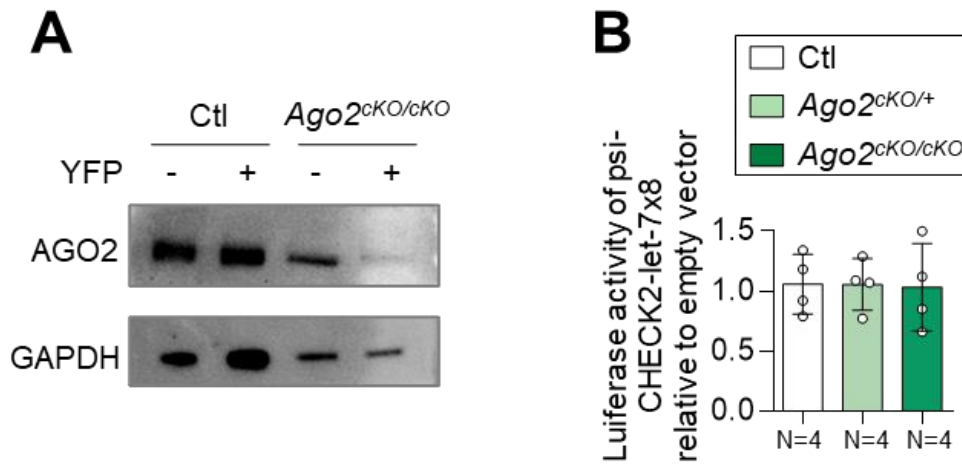


Figure A.3 AGO2 fulfills a non-canonical function relevant for CHARGE syndrome pathogenesis. A) Western blot analysis of AGO2 protein levels in FACS-recovered cells (both YFP-negative and -positive) from e10.5 control (*Wnt1-Cre2*<sup>Tg/+</sup>;*Rosa26*<sup>FloxSTOP-YFP/+</sup>) and homozygous *Ago2* mutant (*Wnt1-Cre2*<sup>Tg/+</sup>;*Rosa26*<sup>FloxSTOP-YFP/+</sup>;*Ago2*<sup>Flox/Flox</sup>) embryos, with GAPDH used as loading control. (B) Luciferase assay in YFP-positive cells recovered by FACS from e10.5 embryos (owing to the *Rosa26*<sup>FloxSTOP-YFP</sup> reporter) and transfected with either the psi-CHECK2-let-7x8 reporter or empty psi-CHECK2 vector. Scale bar, 1 mm (B) and 50  $\mu$ m (C). \* $P \leq 0.05$ , \*\*\* $P \leq 0.001$  and \*\*\*\* $P \leq 0.0001$ ; Khi2 test (E, right) or one-way ANOVA with Tukey's post-hoc test.

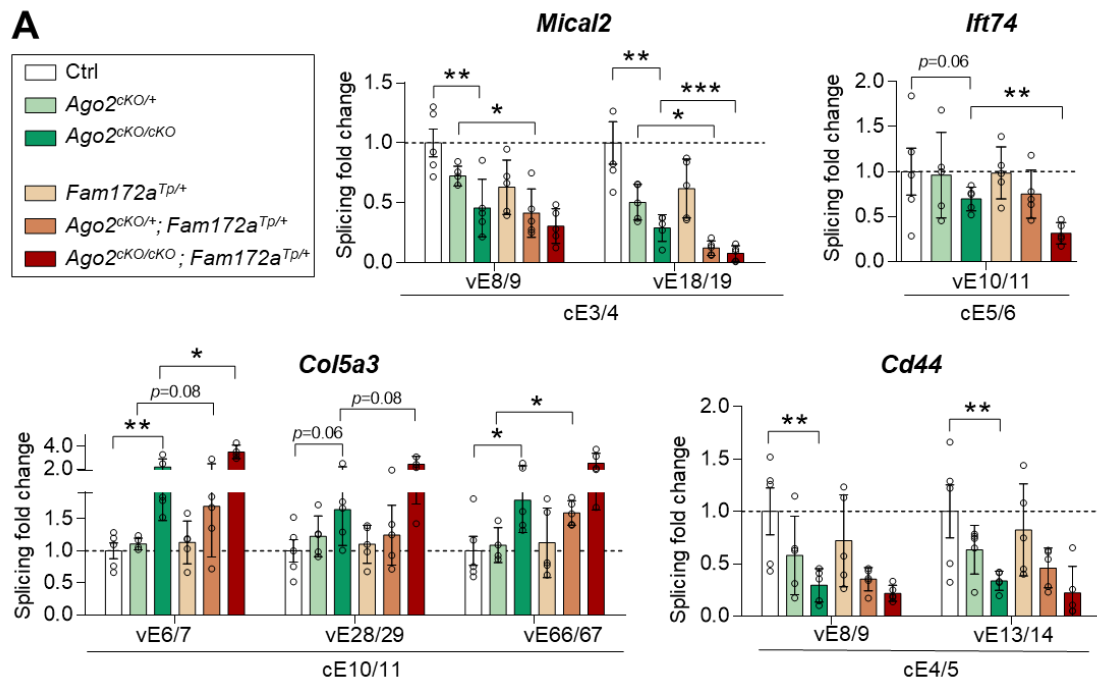


Figure A.4 AGO2 and FAM172A coregulate alternative splicing *in vivo*. (A) RT-qPCR analysis of alternative splicing events for *Mical2*, *Ifi74*, *Col5a3* and *Cd44* in e12.5 heads from an allelic series of *Ago2*<sup>cKO</sup> and *Fam172a*<sup>Tp</sup> mutant embryos. Expression levels of indicated variable exon pairs (vE) were normalized against levels of a constant exon pair (cE), and splicing fold change was then determined by comparison to the reference value (dashed line) obtained with WT control embryos (N=5 per genotype). \* $P \leq 0.05$ , \*\* $P \leq 0.01$ , \*\*\* $P \leq 0.001$ ; Student's *t*-test.



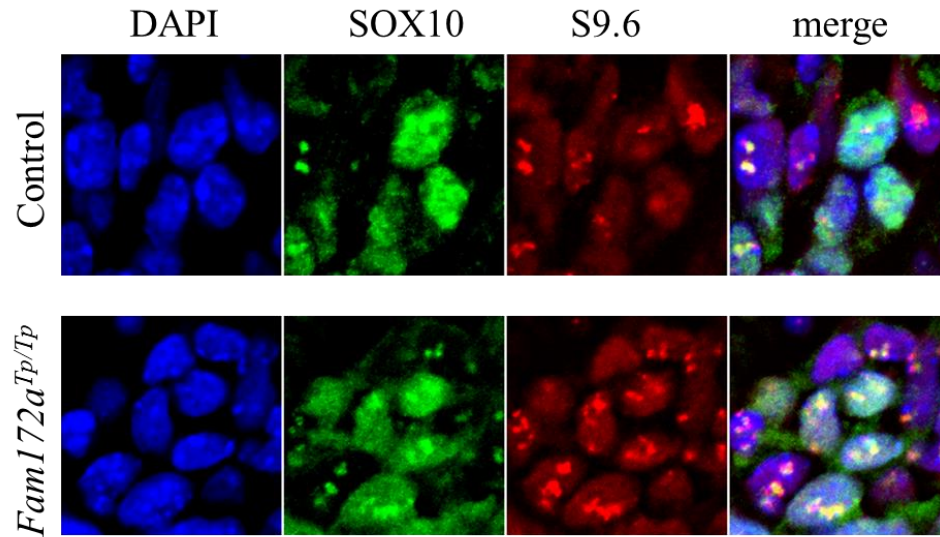


Figure A.5 R-loop-associated-foci are increased in *Fam172a<sup>Tp/Tp</sup>* embryos. Immunofluorescence analysis of R-loop in sections of E10.5 control and *Fam172a<sup>Tp/Tp</sup>* embryos using antibodies against SON and S9.6. Nuclei are stained with DAPI and neural crest cells are identified by the SOX10-positive staining.

Tableau A.1 List of FAM172A interactors.

Category	Uniprot ID	Protein name	Gene name	Main Role
Chromatin structure & composition	P10922	<i>Histone H1.0</i>	H1FO	linker histone
	P43275	<i>Histone H1.1</i>	HIST1H1A	linker histone
	P43276	<i>Histone H1.5</i>	HIST1H1B	linker histone
	P15864	<i>Histone H1.2</i>	HIST1H1C	linker histone
	P43277	<i>Histone H1.3</i>	HIST1H1D	linker histone
	P10412	<i>Histone H1.4</i>	HIST1H1E	linker histone
	Q9QZQ8	<i>Core Histone macro-H2A.1</i>	MACROH2A1	core histone
	P27661	<i>Histone H2AX</i>	H2AX	variant histone
	Q93077	<i>Histone H2A type 1-C</i>	HIST1H2AC	variant histone
	Q99878	<i>Histone H2A type 1-J</i>	HIST1H2AJ	variant histone
	Q6GSS7	<i>Histone H2A type 2-A</i>	HIST2H2AA1	variant histone
	P84228	<i>Histone H3.2</i>	HIST1H3B	core histone
	P62806	<i>Histone H4</i>	HIST1H4A	core histone
	Q3TEA8	<i>Heterochromatin protein 1-binding protein 3</i>	HP1BP3	Histone H1-like
	Q60973	<i>Histone-binding protein RBBP7</i>	RBBP7	recruits remodelers and modifiers
	Q14149	<i>MORC family CW-type zinc finger protein 3</i>	MORC3	H3K4me3 reader
	Q96ES7	<i>SAGA-associated factor 29 homolog</i>	CCDC101	H3K4me3 reader, HAT activity
	O14646	<i>Chromodomain-helicase-DNA-binding protein 1</i>	CHD1	H3K4me3 reader, chromatin remodeler; bridges U2 complex to H3K4me3
	A2AJK6	<i>Chromodomain-helicase-DNA-binding protein 7</i>	CHD7	H3K4me1/3 reader, chromatin remodeler
	Q96KQ7	<i>Histone-lysine N-methyltransferase EHMT2</i>	EHMT2	H3K9 mono- and di-methyltransferase
	Q13185	<i>Chromobox protein homolog 3</i>	CBX3	H3K9me reader
	Q69ZK6	<i>JmjC domain-containing histone demethylation protein 2C</i>	JMJD1C	H3K9 demethylase
	O95696	<i>Bromodomain-containing protein 1</i>	BRD1	H3K14 acetyltransferase
	O95619	<i>YEATS domain-containing protein 4</i>	YEATS4	H3K27ac reader
	Q96L73	<i>Histone-lysine N-methyltransferase, H3 lysine-36 and H4 lysine-20 specific</i>	NSD1	H3K36me2 dimethyltransferase
	Q99MY8	<i>Histone-lysine N-methyltransferase ASH1L</i>	ASH1L	H3K36 trimethyltransferase
	Q9Y2K7	<i>Lysine-specific demethylase 2A</i>	KDM2A	H3K36me3 demethylase
	P46100	<i>Transcriptional regulator ATRX</i>	ATRX	chromatin remodeling
	Q9NRL2	<i>Bromodomain adjacent to zinc finger domain protein 1A</i>	BAZ1A	chromatin remodeling (ISWI complex)
	Q9Z277	<i>Tyrosine-protein kinase BAZ1B</i>	BAZ1B	chromatin remodeling (ISWI complex)
	P28370	<i>Probable global transcription activator SNF2L1</i>	SMARCA1	chromatin remodeling (ISWI complex)
	O60264	<i>SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A member 5</i>	SMARCA5	chromatin remodeling (ISWI complex)

	Q12824	<i>SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily B member 1</i>	SMARCB1	chromatin remodeling (SWI/SNF complex)
	Q969G3	<i>SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily E member 1</i>	SMARCE1	chromatin remodeling (SWI/SNF complex)
	Q13547	<i>Histone deacetylase 1</i>	HDAC1	chromatin compaction
	P70288	<i>Histone deacetylase 2</i>	HDAC2	chromatin compaction
	Q8C2B3	<i>Histone deacetylase 7</i>	HDAC7	chromatin compaction
	Q9BSM1	<i>Polycomb group RING finger protein 1</i>	PCGF1	chromatin compaction
RNA Pol.II-associated machinery	P78347	<i>General transcription factor II-I</i>	GTF2I	pre-initiation complex
	O14981	<i>TATA-binding protein-associated factor 172</i>	BTAF1	pre-initiation complex
	Q9NX70	<i>Mediator of RNA polymerase II transcription subunit 29</i>	MED29	pre-initiation complex
	P19387	<i>DNA-directed RNA polymerase II subunit RPB3</i>	POLR2C	core Pol. II complex
	Q6EEV4	<i>DNA-directed RNA polymerase II subunit GRINL1A, isoforms 4/5</i>	POLR2M	stabilizes paused RNA pol.II
	Q4G0J3	<i>La-related protein 7</i>	LARP7	stabilizes paused RNA pol.II (P-TEFb inhibitor)
	Q5TA45	<i>Integrator complex subunit 11</i>	CPSF3L	promoter-proximal transcription termination
	Q6PD62	<i>RNA polymerase-associated protein CTR9 homolog</i>	CTR9	transcriptional elongation (PAF1 complex)
	Q08943	<i>FACT complex subunit SSRP1</i>	SSRP1	transcriptional elongation (FACT complex)
	P49718	<i>DNA replication licensing factor MCM5</i>	MCM5	transcriptional elongation
Transcription factors & co-factors	Q6P1N0	<i>Coiled-coil and C2 domain-containing protein 1A</i>	CC2D1A	transcription factor
	Q9Y6H1	<i>Coiled-coil-helix-coiled-coil-helix domain-containing protein 2</i>	CHCHD2	transcription factor
	Q9P031	<i>Thyroid transcription factor 1-associated protein 26</i>	CCDC59	transcription factor
	P39880	<i>Homeobox protein cut-like 1</i>	CUX1	transcription factor
	P26583	<i>High mobility group protein B2</i>	HMGB2	transcription factor
	P52952	<i>Homeobox protein Nkx-2.5</i>	NKX2-5	transcription factor
	P13056	<i>Nuclear receptor subfamily 2 group C member 1</i>	NR2C1	transcription factor
	P04150	<i>Glucocorticoid receptor</i>	NR3C1	transcription factor
	Q9Y467	<i>Sal-like protein 2</i>	SALL2	transcription factor
	Q9P016	<i>Thymocyte nuclear protein 1</i>	THYN1	transcription factor
	Q9H7S9	<i>Zinc finger protein 703; Zinc finger protein 503</i>	ZNF703	transcription factor
	Q9Y462	<i>Zinc finger protein 711</i>	ZNF711	transcription factor
	Q6DD87	<i>Zinc finger protein 787</i>	ZNF787	transcription factor
	Q6PIJ6	<i>F-box only protein 38</i>	FBXO38	co-activator
	Q15596	<i>Nuclear receptor coactivator 2</i>	NCOA2	co-activator
	Q13625	<i>Apoptosis-stimulating of p53 protein 2</i>	TP53BP2	co-activator
	Q9NPF5	<i>DNA methyltransferase 1-associated protein 1</i>	DMAP1	co-repressor/co-activator
	Q8IU81	<i>Interferon regulatory factor 2-binding protein 1</i>	IRF2BP1	co-repressor
	Q96SI1	<i>BTB/POZ domain-containing protein KCTD15</i>	KCTD15	co-repressor
	Q96RE7	<i>Nucleus accumbens-associated protein 1</i>	NACC1	co-repressor
	Q96T58	<i>Msx2-interacting protein</i>	SPEN	co-repressor
	Q04726	<i>Transducin-like enhancer protein 3</i>	TLE3	co-repressor
	Q13263	<i>Transcription intermediary factor 1-beta</i>	TRIM28	co-repressor

mRNA metabolism	Q77932	<i>Decapping and exoribonuclease protein</i>	DXO	5'-end processing
	Q43148	<i>mRNA cap guanine-N7 methyltransferase</i>	RNMT	5'-end processing
	Q6PII3	<i>Coiled-coil domain-containing protein 174</i>	CCDC174	Exon Junction Complex
	Q9Y3Y2	<i>Chromatin target of PRMT1 protein</i>	CHTOP	Exon Junction Complex (TREX subunit)
	Q15287	<i>RNA-binding protein with serine-rich domain 1</i>	RNPS1	Exon Junction Complex (ASAP subunit)
	Q9HCG8	<i>Pre-mRNA-splicing factor CWC22 homolog</i>	CWC22	Exon Junction Complex and splicing
	Q96T37	<i>Putative RNA-binding protein 15</i>	RBM15	m6A regulator
	Q9HCE5	<i>N6-adenosine-methyltransferase subunit METTL14</i>	METTL14	m6A methyltransferase
	Q9ER69	<i>Pre-mRNA-splicing regulator WTAP</i>	WTAP	m6A methyltransferase complex
	Q15717	<i>ELAV-like protein 1</i>	ELAVL1	m6A reader and splicing
	P35637	<i>RNA-binding protein FUS</i>	FUS	splicing (bridges pol.II and U1)
	Q9NQ29	<i>Putative RNA-binding protein Luc7-like 1</i>	LUC7L	splicing (U1-associated)
	Q9Y383	<i>Putative RNA-binding protein Luc7-like 2</i>	LUC7L2	splicing (U1-associated)
	Q14498	<i>RNA-binding protein 39</i>	RBM39	splicing (bridges U1 and U2)
	Q921F2	<i>TAR DNA-binding protein 43</i>	TARDBP	splicing (U1- and U2-associated; splice site selection)
	Q99020	<i>Heterogeneous nuclear ribonucleoprotein A/B</i>	HNRNPAB	splicing (U2-associated; splice site selection)
	P07910	<i>Heterogeneous nuclear ribonucleoproteins C1/C2</i>	HNRNPC	splicing (U1- and U2-associated; splice site selection) and m6A reader
	P61979	<i>Heterogeneous nuclear ribonucleoprotein K</i>	HNRNPK	splicing (U2-associated; splice site selection)
	Q7TMK9	<i>Heterogeneous nuclear ribonucleoprotein Q</i>	HNRNPQ	splicing (U2-associated; splice site selection)
	Q00839	<i>Heterogeneous nuclear ribonucleoprotein U</i>	HNRNPU	splicing (U2-associated; splice site selection) and Pol.II pre-initiation complex
	Q6PDM2	<i>Serine/arginine-rich splicing factor 1</i>	SRSF1	splicing (U1- and U2-associated; splice site selection)
	P84104	<i>Serine/arginine-rich splicing factor 3</i>	SRSF3	splicing (U1- and U2-associated; splice site selection)
	Q13247	<i>Serine/arginine-rich splicing factor 6</i>	SRSF6	splicing (U1- and U2-associated; splice site selection)
	Q8BL97	<i>Serine/arginine-rich splicing factor 7</i>	SRSF7	splicing (U1- and U2-associated; splice site selection)
	Q9JII5	<i>DAZ-associated protein 1</i>	DAZAP1	splicing (U2-associated; splice site selection)
	Q9CXY6	<i>Interleukin enhancer-binding factor 2</i>	ILF2	splicing (U2-associated)
	Q9Y388	<i>RNA-binding motif protein, X-linked 2</i>	RBMX2	splicing (U2-associated)
	Q12874	<i>Splicing factor 3A subunit 3</i>	SF3A3	splicing (U2-associated)
	Q9BWJ5	<i>Splicing factor 3B subunit 5</i>	SF3B5	splicing (U2-associated)
	Q9Y2W2	<i>WW domain-binding protein 11</i>	WBP11	splicing (U2-associated)
	P09405	<i>Nucleolin</i>	NCL	splicing (U2-associated; splice site selection)

	Q15233	<i>Non-POU domain-containing octamer-binding protein</i>	NONO	splicing (U2-associated; splice site selection), partner of SFPQ
	P17225	<i>Polypyrimidine tract-binding protein 1</i>	PTBP1	splicing (U2-associated; splice site selection)
	P23246	<i>Splicing factor, proline- and glutamine-rich</i>	SFPQ	splicing (U2-associated; splice site selection), partner of NONO
	Q9UQ35	<i>Serine/arginine repetitive matrix protein 2</i>	SRRM2	splicing (U2-associated; splice site selection)
	P62960	<i>Nuclease-sensitive element-binding protein 1</i>	YBX1	splicing (U2-associated; splice site selection)
	O75940	<i>Survival of motor neuron-related-splicing factor 30</i>	SMNDC1	splicing (U2- and U4/5/6-associated)
	Q16637	<i>Survival motor neuron protein</i>	SMN1	splicing (spliceosome assembly)
	Q9BQA1	<i>Methylosome protein 50</i>	WDR77	splicing (Sm protein methyltransferase)
	Q8WWY3	<i>U4/U6 small nuclear ribonucleoprotein Prp31</i>	PRPF31	splicing (U4/5/6-associated)
	P62304	<i>Small nuclear ribonucleoprotein E</i>	SNRPE	splicing (spliceosome assembly)
	O43143	<i>Pre-mRNA-splicing factor ATP-dependent RNA helicase DHX15</i>	DHX15	splicing (spliceosome disassembly)
	Q92841	<i>Probable ATP-dependent RNA helicase DDX17</i>	DDX17	splicing
	P98179	<i>RNA-binding protein 3</i>	RBM3	splicing
	Q9UKM9	<i>RNA-binding protein Raly</i>	RALY	splicing
	Q12996	<i>Cleavage stimulation factor subunit 3</i>	CSTF3	3'-end processing
	O94913	<i>Pre-mRNA cleavage complex 2 protein Pcf11</i>	PCF11	3'-end processing
	Q6PK04	<i>Coiled-coil domain-containing protein 137</i>	CCDC137	mRNA stability/localization
	Q9GZR7	<i>ATP-dependent RNA helicase DDX24</i>	DDX24	mRNA stability/localization
	Q3U0V1	<i>Far upstream element-binding protein 2</i>	KHSRP	mRNA stability/localization
	Q9BU76	<i>Multiple myeloma tumor-associated protein 2</i>	MMTAG2	mRNA binding, unclear role
R-Loop resolution	O95985	<i>DNA topoisomerase 3-beta-1</i>	TOP3B	Decatenation of nucleic acids at R-loop junction
	Q8TDP1	<i>Ribonuclease H2 subunit C</i>	RNASEH2C	Hydrolysis of the RNA moiety in R-loops
	O70133	<i>ATP-dependent RNA helicase A</i>	DHX9	Unwinding of the DNA-RNA hybrid
	Q8CFQ3	<i>RNA helicase aquarius</i>	AQR	Unwinding of the DNA-RNA hybrid
	O00148	<i>ATP-dependent RNA helicase DDX39A</i>	DDX39A	Unwinding of the DNA-RNA hybrid
RNAi-linked	Q8CJG0	<i>Argonaute RISC catalytic component 2</i>	AGO2	small RNA binding/processing
	Q9UPY3	<i>Endoribonuclease Dicer</i>	DICER1	small RNA binding/processing
	Q9UPQ9	<i>Trinucleotide repeat-containing gene 6B protein</i>	TNRC6B	AGO2 partner

## APPENDICE A

### AUTRES CONTRIBUTIONS SCIENTIFIQUES



ORIGINAL ARTICLE

### **NR2F1 regulates a Schwann cell precursor-vs-melanocyte cell fate switch in a mouse model of Waardenburg syndrome type IV**

Grégoire Bonnamour, Baptiste Charrier, Sephora Sallis, Elizabeth Leduc, Nicolas Pilon ✉

First published: 11 July 2022 | <https://doi.org/10.1111/pcmr.13054>

Grégoire Bonnamour and Baptiste Charrier contributed equally to this work.

## BIBLIOGRAPHIE

- Abuhashem, A., Garg, V., & Hadjantonakis, A.-K. (2022). RNA polymerase II pausing in development: Orchestrating transcription. *Open Biology*, *12*(1), 210220. <https://doi.org/10.1098/rsob.210220>
- Adams, M. E., Hurd, E. A., Beyer, L. A., Swiderski, D. L., Raphael, Y., & Martin, D. M. (2007). Defects in vestibular sensory epithelia and innervation in mice with loss of Chd7 function: Implications for human CHARGE syndrome. *The Journal of Comparative Neurology*, *504*(5), 519–532. <https://doi.org/10.1002/cne.21460>
- Agafonov, D. E., Kastner, B., Dybkov, O., Hofele, R. V., Liu, W.-T., Urlaub, H., Lührmann, R., & Stark, H. (2016). Molecular architecture of the human U4/U6.U5 tri-snRNP. *Science*, *351*(6280), 1416–1420. <https://doi.org/10.1126/science.aad2085>
- Ahlstrom, J. D., & Erickson, C. A. (2009). The neural crest epithelial-mesenchymal transition in 4D: A 'tail' of multiple non-obligatory cellular mechanisms. *Development (Cambridge, England)*, *136*(11), 1801–1812. <https://doi.org/10.1242/dev.034785>
- Akhtar, A., Zink, D., & Becker, P. B. (2000). Chromodomains are protein-RNA interaction modules. *Nature*, *407*(6802), 405–409. <https://doi.org/10.1038/35030169>
- Alendar, A., & Berns, A. (2021). Sentinels of chromatin: Chromodomain helicase DNA-binding proteins in development and disease. *Genes & Development*, *35*(21–22), 1403–1430. <https://doi.org/10.1101/gad.348897.121>
- Allemand, E., Myers, M. P., Garcia-Bernardo, J., Harel-Bellan, A., Krainer, A. R., & Muchardt, C. (2016). A Broad Set of Chromatin Factors Influences Splicing. *PLoS Genetics*, *12*(9), e1006318. <https://doi.org/10.1371/journal.pgen.1006318>
- Alló, M., Buggiano, V., Fededa, J. P., Petrillo, E., Schor, I., de la Mata, M., Agirre, E., Plass, M., Eyra, E., Elela, S. A., Klinck, R., Chabot, B., & Kornblihtt, A. R. (2009). Control of alternative splicing through siRNA-mediated transcriptional gene silencing. *Nature Structural & Molecular Biology*, *16*(7), 717–724. <https://doi.org/10.1038/nsmb.1620>

- Ameyar-Zazoua, M., Rachez, C., Souidi, M., Robin, P., Fritsch, L., Young, R., Morozova, N., Fenouil, R., Descostes, N., Andrau, J.-C., Mathieu, J., Hamiche, A., Ait-Si-Ali, S., Muchardt, C., Batsché, E., & Harel-Bellan, A. (2012). Argonaute proteins couple chromatin silencing to alternative splicing. *Nature Structural & Molecular Biology*, *19*(10), 998–1004. <https://doi.org/10.1038/nsmb.2373>
- Ampofo, E., Sokolowsky, T., Götz, C., & Montenarh, M. (2013). Functional interaction of protein kinase CK2 and activating transcription factor 4 (ATF4), a key player in the cellular stress response. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, *1833*(3), 439–451. <https://doi.org/10.1016/j.bbamcr.2012.10.025>
- Angeloni, A., & Bogdanovic, O. (2021). Sequence determinants, function, and evolution of CpG islands. *Biochemical Society Transactions*, *49*(3), 1109–1119. <https://doi.org/10.1042/BST20200695>
- Aybar, M. J., & Mayor, R. (2002). Early induction of neural crest cells: Lessons learned from frog, fish and chick. *Current Opinion in Genetics & Development*, *12*(4), 452–458. [https://doi.org/10.1016/s0959-437x\(02\)00325-8](https://doi.org/10.1016/s0959-437x(02)00325-8)
- Bachiller, D., Klingensmith, J., Kemp, C., Belo, J. A., Anderson, R. M., May, S. R., McMahon, J. A., McMahon, A. P., Harland, R. M., Rossant, J., & De Robertis, E. M. (2000). The organizer factors Chordin and Noggin are required for mouse forebrain development. *Nature*, *403*(6770), Article 6770. <https://doi.org/10.1038/35001072>
- Bacon, C. W., & D'Orso, I. (2019). CDK9: A signaling hub for transcriptional control. *Transcription*, *10*(2), 57–75. <https://doi.org/10.1080/21541264.2018.1523668>
- Bajpai, R., Chen, D. A., Rada-Iglesias, A., Zhang, J., Xiong, Y., Helms, J., Chang, C.-P., Zhao, Y., Swigut, T., & Wysocka, J. (2010). CHD7 cooperates with PBAF to control multipotent neural crest formation. *Nature*, *463*(7283), 958–962. <https://doi.org/10.1038/nature08733>
- Basson, M. A., & van Ravenswaaij-Arts, C. (2015). Functional Insights into Chromatin Remodelling from Studies on CHARGE Syndrome. *Trends in Genetics: TIG*, *31*(10), 600–611. <https://doi.org/10.1016/j.tig.2015.05.009>
- Batsché, E., & Ameyar-Zazoua, M. (2015). The influence of Argonaute proteins on alternative RNA splicing. *Wiley Interdisciplinary Reviews. RNA*, *6*(1), 141–156. <https://doi.org/10.1002/wrna.1264>



- Batsché, E., Yaniv, M., & Muchardt, C. (2006). The human SWI/SNF subunit Brm is a regulator of alternative splicing. *Nature Structural & Molecular Biology*, *13*(1), 22–29. <https://doi.org/10.1038/nsmb1030>
- Batsukh, T., Pieper, L., Koszucka, A. M., von Velsen, N., Hoyer-Fender, S., Elbracht, M., Bergman, J. E. H., Hoefsloot, L. H., & Pauli, S. (2010). CHD8 interacts with CHD7, a protein which is mutated in CHARGE syndrome. *Human Molecular Genetics*, *19*(14), 2858–2866. <https://doi.org/10.1093/hmg/ddq189>
- Baudre, L., Vallot, C., & Marsolier, J. (2022). La chromatine bivalente préserve la plasticité transcriptionnelle. *médecine/sciences*, *38*(5), Article 5. <https://doi.org/10.1051/medsci/2022050>
- Beacon, T. H., Delcuve, G. P., López, C., Nardocci, G., Kovalchuk, I., van Wijnen, A. J., & Davie, J. R. (2021). The dynamic broad epigenetic (H3K4me3, H3K27ac) domain as a mark of essential genes. *Clinical Epigenetics*, *13*(1), 138. <https://doi.org/10.1186/s13148-021-01126-1>
- Beauchamp, M.-C., Alam, S. S., Kumar, S., & Jerome-Majewska, L. A. (2020). Spliceosomopathies and neurocristopathies: Two sides of the same coin? *Developmental Dynamics*, *249*(8), 924–945. <https://doi.org/10.1002/dvdy.183>
- Bélangier, C., Bérubé-Simard, F.-A., Leduc, E., Bernas, G., Campeau, P. M., Lalani, S. R., Martin, D. M., Bielas, S., Moccia, A., Srivastava, A., Silversides, D. W., & Pilon, N. (2018). Dysregulation of cotranscriptional alternative splicing underlies CHARGE syndrome. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(4), E620–E629. <https://doi.org/10.1073/pnas.1715378115>
- Bélangier, C., Cardinal, T., Leduc, E., Viger, R. S., & Pilon, N. (2022). CHARGE syndrome-associated proteins FAM172A and CHD7 influence male sex determination and differentiation through transcriptional and alternative splicing mechanisms. *FASEB Journal: Official Publication of the Federation of American Societies for Experimental Biology*, *36*(3), e22176. <https://doi.org/10.1096/fj.202100837RR>
- Bender, M., Schwind, L., Grundmann, D., Martin, M., Klotz, M., Götz, C., Montenarh, M., & Schäfer, K.-H. (2017). Impact of protein kinase CK2 inhibitors on proliferation and differentiation of neural stem cells. *Heliyon*, *3*(6), e00318. <https://doi.org/10.1016/j.heliyon.2017.e00318>
- Benhamed, M., Herbig, U., Ye, T., Dejean, A., & Bischof, O. (2012). Senescence is an endogenous trigger for microRNA-directed transcriptional gene silencing

in human cells. *Nature Cell Biology*, *14*(3), 266–275.  
<https://doi.org/10.1038/ncb2443>

- Berglund, J. A., Abovich, N., & Rosbash, M. (1998). A cooperative interaction between U2AF65 and mBBP/SF1 facilitates branchpoint region recognition. *Genes & Development*, *12*(6), 858–867.
- Bernstein, B. E., Mikkelsen, T. S., Xie, X., Kamal, M., Huebert, D. J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., Jaenisch, R., Wagschal, A., Feil, R., Schreiber, S. L., & Lander, E. S. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell*, *125*(2), 315–326.  
<https://doi.org/10.1016/j.cell.2006.02.041>
- Bertin, A., & Mangenot, S. (2008). Structure et dynamique de la particule cœur de nucléosome. *médecine/sciences*, *24*(8–9), Article 8–9.  
<https://doi.org/10.1051/medsci/20082489715>
- Bérubé-Simard, F.-A., & Pilon, N. (2019). Molecular dissection of CHARGE syndrome highlights the vulnerability of neural crest cells to problems with alternative splicing and other transcription-related processes. *Transcription*, *10*(1), 21–28. <https://doi.org/10.1080/21541264.2018.1521213>
- Bhat, P., Chow, A., Emert, B., Ettl, O., Quinodoz, S. A., Takei, Y., Huang, W., Blanco, M. R., & Guttman, M. (2023). *3D genome organization around nuclear speckles drives mRNA splicing efficiency*.
- Bird, A. P. (1987). CpG islands as gene markers in the vertebrate nucleus. *Trends in Genetics*, *3*, 342–347. [https://doi.org/10.1016/0168-9525\(87\)90294-0](https://doi.org/10.1016/0168-9525(87)90294-0)
- Blake, K. D., Davenport, S. L., Hall, B. D., Hefner, M. A., Pagon, R. A., Williams, M. S., Lin, A. E., & Graham, J. M. (1998). CHARGE association: An update and review for the primary pediatrician. *Clinical Pediatrics*, *37*(3), 159–173.  
<https://doi.org/10.1177/000992289803700302>
- Blake, K. D., & Prasad, C. (2006). CHARGE syndrome. *Orphanet Journal of Rare Diseases*, *1*, 34. <https://doi.org/10.1186/1750-1172-1-34>
- Bohmert, K., Camus, I., Bellini, C., Bouchez, D., Caboche, M., & Benning, C. (1998). AGO1 defines a novel locus of Arabidopsis controlling leaf development. *The EMBO Journal*, *17*(1), 170–180.  
<https://doi.org/10.1093/emboj/17.1.170>
- Borgo, C., D'Amore, C., Sarno, S., Salvi, M., & Ruzzene, M. (2021). Protein kinase CK2: A potential therapeutic target for diverse human diseases. *Signal*

*Transduction and Targeted Therapy*, 6(1), 1–20.  
<https://doi.org/10.1038/s41392-021-00567-7>

- Bosman, E. A., Penn, A. C., Ambrose, J. C., Kettleborough, R., Stemple, D. L., & Steel, K. P. (2005). Multiple mutations in mouse *Chd7* provide models for CHARGE syndrome. *Human Molecular Genetics*, 14(22), 3463–3476.  
<https://doi.org/10.1093/hmg/ddi375>
- Bouazoune, K., & Kingston, R. E. (2012). Chromatin remodeling by the CHD7 protein is impaired by mutations that cause human developmental disorders. *Proceedings of the National Academy of Sciences of the United States of America*, 109(47), 19238–19243. <https://doi.org/10.1073/pnas.1213825109>
- Bourgeois, C. F., Mortreux, F., & Auboeuf, D. (2016). The multiple functions of RNA helicases as drivers and regulators of gene expression. *Nature Reviews. Molecular Cell Biology*, 17(7), 426–438. <https://doi.org/10.1038/nrm.2016.50>
- Brahma, S., & Henikoff, S. (2024). The BAF chromatin remodeler synergizes with RNA polymerase II and transcription factors to evict nucleosomes. *Nature Genetics*, 56(1), 100–111. <https://doi.org/10.1038/s41588-023-01603-8>
- Brault, V., Moore, R., Kutsch, S., Ishibashi, M., Rowitch, D. H., McMahon, A. P., Sommer, L., Boussadia, O., & Kemler, R. (2001). Inactivation of the  $\beta$ -catenin gene by Wnt1-Cre-mediated deletion results in dramatic brain malformation and failure of craniofacial development. *Development*, 128(8), 1253–1264. <https://doi.org/10.1242/dev.128.8.1253>
- Breathnach, R., Benoist, C., O'Hare, K., Gannon, F., & Chambon, P. (1978). Ovalbumin gene: Evidence for a leader sequence in mRNA and DNA sequences at the exon-intron boundaries. *Proceedings of the National Academy of Sciences*, 75(10), 4853–4857.  
<https://doi.org/10.1073/pnas.75.10.4853>
- Bringmann, P., Appel, B., Rinke, J., Reuter, R., Theissen, H., & Lührmann, R. (1984). Evidence for the existence of snRNAs U4 and U6 in a single ribonucleoprotein complex and for their association by intermolecular base pairing. *The EMBO Journal*, 3(6), 1357–1363. <https://doi.org/10.1002/j.1460-2075.1984.tb01977.x>
- Bringmann, P., & Lührmann, R. (1986). Purification of the individual snRNPs U1, U2, U5 and U4/U6 from HeLa cells and characterization of their protein constituents. *The EMBO Journal*, 5(13), 3509–3516.  
<https://doi.org/10.1002/j.1460-2075.1986.tb04676.x>

- Brody, E., & Abelson, J. (1985). The “Spliceosome”: Yeast Pre-Messenger RNA Associates with a 40S Complex in a Splicing-Dependent Reaction. *Science*, 228(4702), 963–967. <https://doi.org/10.1126/science.3890181>
- Buker, S. M., Iida, T., Bühler, M., Villén, J., Gygi, S. P., Nakayama, J.-I., & Moazed, D. (2007). Two different Argonaute complexes are required for siRNA generation and heterochromatin assembly in fission yeast. *Nature Structural & Molecular Biology*, 14(3), 200–207. <https://doi.org/10.1038/nsmb1211>
- Caizzi, L., Monteiro-Martins, S., Schwalb, B., Lysakovskaia, K., Schmitzova, J., Sawicka, A., Chen, Y., Lidschreiber, M., & Cramer, P. (2021). Efficient RNA polymerase II pause release requires U2 snRNP function. *Molecular Cell*, 81(9), 1920–1934.e9. <https://doi.org/10.1016/j.molcel.2021.02.016>
- Carissimi, C., Laudadio, I., Cipolletta, E., Gioiosa, S., Mihailovich, M., Bonaldi, T., Macino, G., & Fulci, V. (2015). ARGONAUTE2 cooperates with SWI/SNF complex to determine nucleosome occupancy at human Transcription Start Sites. *Nucleic Acids Research*, 43(3), 1498–1512. <https://doi.org/10.1093/nar/gku1387>
- Castanotto, D., Zhang, X., Alluin, J., Zhang, X., Rüger, J., Armstrong, B., Rossi, J., Riggs, A., & Stein, C. A. (2018). A stress-induced response complex (SIRC) shuttles miRNAs, siRNAs, and oligonucleotides to the nucleus. *Proceedings of the National Academy of Sciences of the United States of America*, 115(25), E5756–E5765. <https://doi.org/10.1073/pnas.1721346115>
- Chai, M., Sanosaka, T., Okuno, H., Zhou, Z., Koya, I., Banno, S., Andoh-Noda, T., Tabata, Y., Shimamura, R., Hayashi, T., Ebisawa, M., Sasagawa, Y., Nikaido, I., Okano, H., & Kohyama, J. (2018). Chromatin remodeler CHD7 regulates the stem cell identity of human neural progenitors. *Genes & Development*, 32(2), 165–180. <https://doi.org/10.1101/gad.301887.117>
- Chen, M., & Manley, J. L. (2009). Mechanisms of alternative splicing regulation: Insights from molecular and genomics approaches. *Nature Reviews. Molecular Cell Biology*, 10(11), 741–754. <https://doi.org/10.1038/nrm2777>
- Chen, Y., Zhang, Y., Wang, Y., Zhang, L., Brinkman, E. K., Adam, S. A., Goldman, R., van Steensel, B., Ma, J., & Belmont, A. S. (2018). Mapping 3D genome organization relative to nuclear compartments using TSA-Seq as a cytological ruler. *The Journal of Cell Biology*, 217(11), 4025–4048. <https://doi.org/10.1083/jcb.201807108>

- Cheon, C.-K., & Ko, J. M. (2015). Kabuki syndrome: Clinical and molecular characteristics. *Korean Journal of Pediatrics*, *58*(9), 317–324. <https://doi.org/10.3345/kjp.2015.58.9.317>
- Chu, Y., Yokota, S., Liu, J., Kilikevicius, A., Johnson, K. C., & Corey, D. R. (2021). Argonaute binding within human nuclear RNA and its impact on alternative splicing. *RNA (New York, N.Y.)*, *27*(9), 991–1003. <https://doi.org/10.1261/rna.078707.121>
- Conaway, J. W., Shilatifard, A., Dvir, A., & Conaway, R. C. (2000). Control of elongation by RNA polymerase II. *Trends in Biochemical Sciences*, *25*(8), 375–380. [https://doi.org/10.1016/s0968-0004\(00\)01615-7](https://doi.org/10.1016/s0968-0004(00)01615-7)
- Contrôle de la transcription. (2021, November 13). *Biologie cellulaire et génétique du Développement*. <https://bcgdevelop.fr/contrôle-de-la-transcription/>
- Cramer, P. (2019). Organization and regulation of gene transcription. *Nature*, *573*(7772), 45–54. <https://doi.org/10.1038/s41586-019-1517-4>
- Davidson, A. J., & Zon, L. I. (2000). Turning mesoderm into blood: The formation of hematopoietic stem cells during embryogenesis. *Current Topics in Developmental Biology*, *50*, 45–60. [https://doi.org/10.1016/s0070-2153\(00\)50003-9](https://doi.org/10.1016/s0070-2153(00)50003-9)
- Davie, J. R., Xu, W., & Delcuve, G. P. (2016). Histone H3K4 trimethylation: Dynamic interplay with pre-mRNA splicing. *Biochemistry and Cell Biology = Biochimie Et Biologie Cellulaire*, *94*(1), 1–11. <https://doi.org/10.1139/bcb-2015-0065>
- de la Mata, M., 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. *Molecular Cell*, *12*(2), 525–532. <https://doi.org/10.1016/j.molcel.2003.08.001>
- Deaton, A. M., & Bird, A. (2011). CpG islands and the regulation of transcription. *Genes & Development*, *25*(10), 1010–1022. <https://doi.org/10.1101/gad.2037511>
- Delmas, V., Stokes, D. G., & Perry, R. P. (1993). A mammalian DNA-binding protein that contains a chromodomain and an SNF2/SWI2-like helicase domain. *Proceedings of the National Academy of Sciences of the United States of America*, *90*(6), 2414–2418. <https://doi.org/10.1073/pnas.90.6.2414>

- Deng, W., & Roberts, S. G. E. (2005). A core promoter element downstream of the TATA box that is recognized by TFIIB. *Genes & Development*, *19*(20), 2418–2423. <https://doi.org/10.1101/gad.342405>
- Dodé, C., Levilliers, J., Dupont, J.-M., De Paepe, A., Le Dû, N., Soussi-Yanicostas, N., Coimbra, R. S., Delmaghani, S., Compain-Nouaille, S., Baverel, F., Pêcheux, C., Le Tessier, D., Cruaud, C., Delpech, M., Speleman, F., Vermeulen, S., Amalfitano, A., Bachelot, Y., Bouchard, P., ... Hardelin, J.-P. (2003). Loss-of-function mutations in FGFR1 cause autosomal dominant Kallmann syndrome. *Nature Genetics*, *33*(4), 463–465. <https://doi.org/10.1038/ng1122>
- Dreos, R., Ambrosini, G., Groux, R., Cavin Périer, R., & Bucher, P. (2017). The eukaryotic promoter database in its 30th year: Focus on non-vertebrate organisms. *Nucleic Acids Research*, *45*(D1), D51–D55. <https://doi.org/10.1093/nar/gkw1069>
- Dujardin, G., Daguene, É., Bernard, D. G., Flodrops, M., Durand, S., Chauveau, A., Khoury, F. E., Jossic-Corcus, C. L., & Corcos, L. (2016). L'épissage des ARN pré-messagers: Quand le spliceosome perd pied. *médecine/sciences*, *32*(12), Article 12. <https://doi.org/10.1051/medsci/20163212014>
- Duncan, J. S., Gyenis, L., Lenehan, J., Bretner, M., Graves, L. M., Haystead, T. A., & Litchfield, D. W. (2008). An unbiased evaluation of CK2 inhibitors by chemoproteomics: Characterization of inhibitor effects on CK2 and identification of novel inhibitor targets. *Molecular & Cellular Proteomics: MCP*, *7*(6), 1077–1088. <https://doi.org/10.1074/mcp.M700559-MCP200>
- Egly, J.-M., & Coin, F. (2011). A history of TFIIH: Two decades of molecular biology on a pivotal transcription/repair factor. *DNA Repair*, *10*(7), 714–721. <https://doi.org/10.1016/j.dnarep.2011.04.021>
- Engelen, E., Akinci, U., Bryne, J. C., Hou, J., Gontan, C., Moen, M., Szumska, D., Kockx, C., van Ijcken, W., Dekkers, D. H. W., Demmers, J., Rijkers, E.-J., Bhattacharya, S., Philipsen, S., Pevny, L. H., Grosveld, F. G., Rottier, R. J., Lenhard, B., & Poot, R. A. (2011). Sox2 cooperates with Chd7 to regulate genes that are mutated in human syndromes. *Nature Genetics*, *43*(6), 607–611. <https://doi.org/10.1038/ng.825>
- Etchevers, H. C., Amiel, J., & Lyonnet, S. (2006). Molecular bases of human neurocristopathies. *Advances in Experimental Medicine and Biology*, *589*, 213–234. [https://doi.org/10.1007/978-0-387-46954-6\\_14](https://doi.org/10.1007/978-0-387-46954-6_14)

- Feng, W., Kawauchi, D., Körkel-Qu, H., Deng, H., Serger, E., Sieber, L., Lieberman, J. A., Jimeno-González, S., Lambo, S., Hanna, B. S., Harim, Y., Jansen, M., Neuerburg, A., Friesen, O., Zuckermann, M., Rajendran, V., Gronych, J., Ayrault, O., Korshunov, A., ... Liu, H.-K. (2017). Chd7 is indispensable for mammalian brain development through activation of a neuronal differentiation programme. *Nature Communications*, 8. <https://doi.org/10.1038/ncomms14758>
- Feng, W., Khan, M. A., Bellvis, P., Zhu, Z., Bernhardt, O., Herold-Mende, C., & Liu, H.-K. (2013). The Chromatin Remodeler CHD7 Regulates Adult Neurogenesis via Activation of SoxC Transcription Factors. *Cell Stem Cell*, 13(1), 62–72. <https://doi.org/10.1016/j.stem.2013.05.002>
- Fong, N., Kim, H., Zhou, Y., Ji, X., Qiu, J., Saldi, T., Diener, K., Jones, K., Fu, X.-D., & Bentley, D. L. (2014). Pre-mRNA splicing is facilitated by an optimal RNA polymerase II elongation rate. *Genes & Development*, 28(23), 2663–2676. <https://doi.org/10.1101/gad.252106.114>
- Fong, Y. W., & Zhou, Q. (2001). Stimulatory effect of splicing factors on transcriptional elongation. *Nature*, 414(6866), 929–933. <https://doi.org/10.1038/414929a>
- Friedman, R. C., Farh, K. K.-H., Burge, C. B., & Bartel, D. P. (2009). Most mammalian mRNAs are conserved targets of microRNAs. *Genome Research*, 19(1), 92–105. <https://doi.org/10.1101/gr.082701.108>
- Fu, X. D., & Maniatis, T. (1990). Factor required for mammalian spliceosome assembly is localized to discrete regions in the nucleus. *Nature*, 343(6257), 437–441. <https://doi.org/10.1038/343437a0>
- Fujita, K., Ogawa, R., Kawawaki, S., & Ito, K. (2014). Roles of chromatin remodelers in maintenance mechanisms of multipotency of mouse trunk neural crest cells in the formation of neural crest-derived stem cells. *Mechanisms of Development*, 133, 126–145. <https://doi.org/10.1016/j.mod.2014.05.001>
- G Hendrickson, D., Kelley, D. R., Tenen, D., Bernstein, B., & Rinn, J. L. (2016). Widespread RNA binding by chromatin-associated proteins. *Genome Biology*, 17, 28. <https://doi.org/10.1186/s13059-016-0878-3>
- Gaertner, B., & Zeitlinger, J. (2014). RNA polymerase II pausing during development. *Development (Cambridge, England)*, 141(6), 1179–1183. <https://doi.org/10.1242/dev.088492>

- Gage, P. J., Hurd, E. A., & Martin, D. M. (2015). Mouse Models for the Dissection of CHD7 Functions in Eye Development and the Molecular Basis for Ocular Defects in CHARGE Syndrome. *Investigative Ophthalmology & Visual Science*, 56(13), 7923–7930. <https://doi.org/10.1167/iovs.15-18069>
- Galganski, L., Urbanek, M. O., & Krzyzosiak, W. J. (2017). Nuclear speckles: Molecular organization, biological function and role in disease. *Nucleic Acids Research*, 45(18), 10350–10368. <https://doi.org/10.1093/nar/gkx759>
- Gao, J., Skidmore, J. M., Cimerman, J., Ritter, K. E., Qiu, J., Wilson, L. M. Q., Raphael, Y., Kwan, K. Y., & Martin, D. M. (2024). CHD7 and SOX2 act in a common gene regulatory network during mammalian semicircular canal and cochlear development. *Proceedings of the National Academy of Sciences of the United States of America*, 121(10), e2311720121. <https://doi.org/10.1073/pnas.2311720121>
- Gao, M., Wei, W., Li, M.-M., Wu, Y.-S., Ba, Z., Jin, K.-X., Li, M.-M., Liao, Y.-Q., Adhikari, S., Chong, Z., Zhang, T., Guo, C.-X., Tang, T., Zhu, B.-T., Xu, X.-Z., Mailand, N., Yang, Y.-G., Qi, Y., & Danielsen, J. M. R. (2014). Ago2 facilitates Rad51 recruitment and DNA double-strand break repair by homologous recombination. *Cell Research*, 24(5), 532–541. <https://doi.org/10.1038/cr.2014.36>
- Gibson, D. G. (2011). Enzymatic Assembly of Overlapping DNA Fragments. *Methods in Enzymology*, 498, 349–361. <https://doi.org/10.1016/B978-0-12-385120-8.00015-2>
- Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A., & Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*, 6(5), 343–345. <https://doi.org/10.1038/nmeth.1318>
- Gnatt, A. L., Cramer, P., Fu, J., Bushnell, D. A., & Kornberg, R. D. (2001). Structural basis of transcription: An RNA polymerase II elongation complex at 3.3 Å resolution. *Science (New York, N.Y.)*, 292(5523), 1876–1882. <https://doi.org/10.1126/science.1059495>
- Goldberg, M. L. (1979). *SEQUENCE ANALYSIS OF DROSOPHILA HISTONE GENES*. - *ProQuest*. <https://www.proquest.com/openview/cfbfa1b425598a91dda317c0f79a9a6b/1.pdf?pq-origsite=gscholar&cbl=18750&diss=y>
- Gompers, A. L., Su-Feher, L., Ellegood, J., Copping, N. A., Riyadh, M. A., Stradleigh, T. W., Pride, M. C., Schaffler, M. D., Wade, A. A., Catta-Preta,



- R., Zdilar, I., Louis, S., Kaushik, G., Mannion, B. J., Plajzer-Frick, I., Afzal, V., Visel, A., Pennacchio, L. A., Dickel, D. E., ... Nord, A. S. (2017). Germline Chd8 haploinsufficiency alters brain development in mouse. *Nature Neuroscience*, 20(8), 1062–1073. <https://doi.org/10.1038/nn.4592>
- Götz, C., & Montenarh, M. (2017). Protein kinase CK2 in development and differentiation (Review). *Biomedical Reports*, 6(2), 127–133. <https://doi.org/10.3892/br.2016.829>
- Graveley, B. R. (2000). Sorting out the complexity of SR protein functions. *RNA (New York, N.Y.)*, 6(9), 1197–1211. <https://doi.org/10.1017/s1355838200000960>
- Guiro, J., & O'Reilly, D. (2015). Insights into the U1 small nuclear ribonucleoprotein complex superfamily. *Wiley Interdisciplinary Reviews. RNA*, 6(1), 79–92. <https://doi.org/10.1002/wrna.1257>
- Gutbrod, M. J., & Martienssen, R. A. (2020). Conserved chromosomal functions of RNA interference. *Nature Reviews. Genetics*, 21(5), 311–331. <https://doi.org/10.1038/s41576-019-0203-6>
- Ha, S. D., Ham, S., Kim, M. Y., Kim, J. H., Jang, I., Lee, B. B., Lee, M. K., Hwang, J.-T., Roh, T.-Y., & Kim, T. (2019). Transcription-dependent targeting of Hda1C to hyperactive genes mediates H4-specific deacetylation in yeast. *Nature Communications*, 10(1), 4270. <https://doi.org/10.1038/s41467-019-12077-w>
- Hale, C. L., Niederriter, A. N., Green, G. E., & Martin, D. M. (2016). Atypical phenotypes associated with pathogenic CHD7 variants and a proposal for broadening CHARGE syndrome clinical diagnostic criteria. *American Journal of Medical Genetics. Part A*, 170A(2), 344–354. <https://doi.org/10.1002/ajmg.a.37435>
- Halloran, D., Pandit, V., & Nohe, A. (2022). The Role of Protein Kinase CK2 in Development and Disease Progression: A Critical Review. *Journal of Developmental Biology*, 10(3), 31. <https://doi.org/10.3390/jdb10030031>
- Hardelin, J.-P. (2001). Kallmann syndrome: Towards molecular pathogenesis. *Molecular and Cellular Endocrinology*, 179(1), 75–81. [https://doi.org/10.1016/S0303-7207\(01\)00462-2](https://doi.org/10.1016/S0303-7207(01)00462-2)
- Harland, R. (2000). Neural induction. *Current Opinion in Genetics & Development*, 10(4), 357–362. [https://doi.org/10.1016/S0959-437X\(00\)00096-4](https://doi.org/10.1016/S0959-437X(00)00096-4)

- He, D., Marie, C., Zhao, C., Kim, B., Wang, J., Deng, Y., Clavairoly, A., Frah, M., Wang, H., He, X., Hmidan, H., Jones, B. V., Witte, D., Zalc, B., Zhou, X., Choo, D. I., Martin, D. M., Parras, C., & Lu, Q. R. (2016). Chd7 cooperates with Sox10 and regulates the onset of CNS myelination and remyelination. *Nature Neuroscience*, *19*(5), 678–689. <https://doi.org/10.1038/nn.4258>
- Henikoff, S., Henikoff, J. G., Ahmad, K., Paranal, R. M., Janssens, D. H., Russell, Z. R., Szulzewsky, F., Kugel, S., & Holland, E. C. (2023). Epigenomic analysis of formalin-fixed paraffin-embedded samples by CUT&Tag. *Nature Communications*, *14*(1), 5930. <https://doi.org/10.1038/s41467-023-41666-z>
- Holland, L. Z. (2005). Non-neural ectoderm is really neural: Evolution of developmental patterning mechanisms in the non-neural ectoderm of chordates and the problem of sensory cell homologies. *Journal of Experimental Zoology. Part B, Molecular and Developmental Evolution*, *304*(4), 304–323. <https://doi.org/10.1002/jez.b.21038>
- Hsu, P., Ma, A., Wilson, M., Williams, G., Curotta, J., Munns, C. F., & Mehr, S. (2014). CHARGE syndrome: A review. *Journal of Paediatrics and Child Health*, *50*(7), 504–511. <https://doi.org/10.1111/jpc.12497>
- Hu, X., Li, Y., Zhang, T., Li, L., Chen, S., Wu, X., Li, H., Qi, B., & Chen, Z. (2021). Phosphorylation of Ago2 is required for its role in DNA double-strand break repair. *Journal of Genetics and Genomics*, *48*(4), 333–340. <https://doi.org/10.1016/j.jgg.2021.03.011>
- Huang, X., & Saint-Jeannet, J.-P. (2004). Induction of the neural crest and the opportunities of life on the edge. *Developmental Biology*, *275*(1), 1–11. <https://doi.org/10.1016/j.ydbio.2004.07.033>
- Hübner, S., Xiao, C. Y., & Jans, D. A. (1997). The protein kinase CK2 site (Ser111/112) enhances recognition of the simian virus 40 large T-antigen nuclear localization sequence by importin. *The Journal of Biological Chemistry*, *272*(27), 17191–17195. <https://doi.org/10.1074/jbc.272.27.17191>
- Hudry, B., Viala, S., Graba, Y., & Merabet, S. (2011). Visualization of protein interactions in living *Drosophila* embryos by the bimolecular fluorescence complementation assay. *BMC Biology*, *9*, 5. <https://doi.org/10.1186/1741-7007-9-5>
- Hurd, E. A., Capers, P. L., Blauwkamp, M. N., Adams, M. E., Raphael, Y., Poucher, H. K., & Martin, D. M. (2007). Loss of Chd7 function in gene-trapped reporter mice is embryonic lethal and associated with severe defects in multiple developing tissues. *Mammalian Genome: Official Journal of the*

*International Mammalian Genome Society*, 18(2), 94–104.  
<https://doi.org/10.1007/s00335-006-0107-6>

- Hutvagner, G., & Simard, M. J. (2008). Argonaute proteins: Key players in RNA silencing. *Nature Reviews. Molecular Cell Biology*, 9(1), 22–32.  
<https://doi.org/10.1038/nrm2321>
- Ilik, İ. A., Malszycki, M., Lübke, A. K., Schade, C., Meierhofer, D., & Aktaş, T. (2020). SON and SRRM2 are essential for nuclear speckle formation. *eLife*, 9, e60579. <https://doi.org/10.7554/eLife.60579>
- Ilik, İ. A., & Aktaş, T. (2021). Nuclear speckles: Dynamic hubs of gene expression regulation. *The FEBS Journal*, febs.16117. <https://doi.org/10.1111/febs.16117>
- Ishihara, K., Oshimura, M., & Nakao, M. (2006). CTCF-dependent chromatin insulator is linked to epigenetic remodeling. *Molecular Cell*, 23(5), 733–742.  
<https://doi.org/10.1016/j.molcel.2006.08.008>
- Issekutz, K. A., Graham, J. M., Prasad, C., Smith, I. M., & Blake, K. D. (2005). An epidemiological analysis of CHARGE syndrome: Preliminary results from a Canadian study. *American Journal of Medical Genetics. Part A*, 133A(3), 309–317. <https://doi.org/10.1002/ajmg.a.30560>
- Jans, D. A., Xiao, C. Y., & Lam, M. H. (2000). Nuclear targeting signal recognition: A key control point in nuclear transport? *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, 22(6), 532–544.  
[https://doi.org/10.1002/\(SICI\)1521-1878\(200006\)22:6<532::AID-BIES6>3.0.CO;2-O](https://doi.org/10.1002/(SICI)1521-1878(200006)22:6<532::AID-BIES6>3.0.CO;2-O)
- Janssen, N., Bergman, J. E. H., Swertz, M. A., Tranebjaerg, L., Lodahl, M., Schoots, J., Hofstra, R. M. W., van Ravenswaaij-Arts, C. M. A., & Hoefsloot, L. H. (2012). Mutation update on the CHD7 gene involved in CHARGE syndrome. *Human Mutation*, 33(8), 1149–1160.  
<https://doi.org/10.1002/humu.22086>
- Ji, X., 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(4), 855–868.  
<https://doi.org/10.1016/j.cell.2013.04.028>
- Johnson, K. C., Kilikevicius, A., Hofman, C., Hu, J., Liu, Y., Aguilar, S., Graswich, J., Han, Y., Wang, T., Westcott, J. M., Brekken, R. A., Peng, L., Karagkounis, G., & Corey, D. R. (2024). Nuclear localization of Argonaute 2 is affected by

- cell density and may relieve repression by microRNAs. *Nucleic Acids Research*, 52(4), 1930–1952. <https://doi.org/10.1093/nar/gkad1155>
- Johnston, M., Geoffroy, M.-C., Sobala, A., Hay, R., & Hutvagner, G. (2010). HSP90 Protein Stabilizes Unloaded Argonaute Complexes and Microscopic P-bodies in Human Cells. *Molecular Biology of the Cell*, 21(9), 1462–1469. <https://doi.org/10.1091/mbc.E09-10-0885>
- Jongmans, M. C. J., Admiraal, R. J., van der Donk, K. P., Vissers, L. E. L. M., Baas, A. F., Kapusta, L., van Hagen, J. M., Donnai, D., de Ravel, T. J., Veltman, J. A., Geurts van Kessel, A., De Vries, B. B. A., Brunner, H. G., Hoefsloot, L. H., & van Ravenswaaij, C. M. A. (2006). CHARGE syndrome: The phenotypic spectrum of mutations in the CHD7 gene. *Journal of Medical Genetics*, 43(4), 306–314. <https://doi.org/10.1136/jmg.2005.036061>
- Jordan, V. K., Fregeau, B., Ge, X., Giordano, J., Wapner, R. J., Balci, T. B., Carter, M. T., Bernat, J. A., Moccia, A. N., Srivastava, A., Martin, D. M., Bielas, S. L., Pappas, J., Svoboda, M. D., Rio, M., Boddaert, N., Cantagrel, V., Lewis, A. M., Scaglia, F., ... Scott, D. A. (2018). Genotype-phenotype correlations in individuals with pathogenic RERE variants. *Human Mutation*, 39(5), 666–675. <https://doi.org/10.1002/humu.23400>
- Jurica, M. S., & Moore, M. J. (2002). Capturing splicing complexes to study structure and mechanism. *Methods*, 28(3), 336–345. [https://doi.org/10.1016/S1046-2023\(02\)00240-2](https://doi.org/10.1016/S1046-2023(02)00240-2)
- Kalantari, R., Chiang, C.-M., & Corey, D. R. (2016). Regulation of mammalian transcription and splicing by Nuclear RNAi. *Nucleic Acids Research*, 44(2), 524–537. <https://doi.org/10.1093/nar/gkv1305>
- Kastner, B., Will, C. L., Stark, H., & Lührmann, R. (2019). Structural Insights into Nuclear pre-mRNA Splicing in Higher Eukaryotes. *Cold Spring Harbor Perspectives in Biology*, 11(11), a032417. <https://doi.org/10.1101/cshperspect.a032417>
- Khan, D. H., Gonzalez, C., Tailor, N., Hamedani, M. K., Leygue, E., & Davie, J. R. (2016). Dynamic Histone Acetylation of H3K4me3 Nucleosome Regulates MCL1 Pre-mRNA Splicing. *Journal of Cellular Physiology*, 231(10), 2196–2204. <https://doi.org/10.1002/jcp.25337>
- Kidder, B. L. (2018). CARIP-Seq and ChIP-Seq: Methods to Identify Chromatin-Associated RNAs and Protein-DNA Interactions in Embryonic Stem Cells. *Journal of Visualized Experiments : JoVE*, 135, 57481. <https://doi.org/10.3791/57481>

- Kim, Y. J., Björklund, S., Li, Y., Sayre, M. H., & Kornberg, R. D. (1994). A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. *Cell*, *77*(4), 599–608. [https://doi.org/10.1016/0092-8674\(94\)90221-6](https://doi.org/10.1016/0092-8674(94)90221-6)
- Kistler, A. L., & Guthrie, C. (2001). Deletion of MUD2, the yeast homolog of U2AF65, can bypass the requirement for Sub2, an essential spliceosomal ATPase. *Genes & Development*, *15*(1), 42–49. <https://doi.org/10.1101/gad.851601>
- Knecht, A. K., & Bronner-Fraser, M. (2002). Induction of the neural crest: A multigene process. *Nature Reviews. Genetics*, *3*(6), 453–461. <https://doi.org/10.1038/nrg819>
- Kobayashi, J., & Matsuura, Y. (2013). Structural Basis for Cell-Cycle-Dependent Nuclear Import Mediated by the Karyopherin Kap121p. *Journal of Molecular Biology*, *425*(11), 1852–1868. <https://doi.org/10.1016/j.jmb.2013.02.035>
- Kodama, Y., & Hu, C.-D. (2012). Bimolecular fluorescence complementation (BiFC): A 5-year update and future perspectives. *BioTechniques*, *53*(5), 285–298. <https://doi.org/10.2144/000113943>
- Koidl, S., & Timmers, H. T. M. (2021). greenCUT&RUN: Efficient Genomic Profiling of GFP-Tagged Transcription Factors and Chromatin Regulators. *Current Protocols*, *1*(10), e266. <https://doi.org/10.1002/cpz1.266>
- Kornberg, R. D. (2005). Mediator and the mechanism of transcriptional activation. *Trends in Biochemical Sciences*, *30*(5), 235–239. <https://doi.org/10.1016/j.tibs.2005.03.011>
- Krainer, A. R., Mayeda, A., Kozak, D., & Binns, G. (1991). Functional expression of cloned human splicing factor SF2: Homology to RNA-binding proteins, U1 70K, and Drosophila splicing regulators. *Cell*, *66*(2), 383–394. [https://doi.org/10.1016/0092-8674\(91\)90627-b](https://doi.org/10.1016/0092-8674(91)90627-b)
- Krämer, A., Grüter, P., Gröning, K., & Kastner, B. (1999). Combined Biochemical and Electron Microscopic Analyses Reveal the Architecture of the Mammalian U2 snRNP. *Journal of Cell Biology*, *145*(7), 1355–1368. <https://doi.org/10.1083/jcb.145.7.1355>
- Krueger, L. A., & Morris, A. C. (2022). Eyes on CHARGE syndrome: Roles of CHD7 in ocular development. *Frontiers in Cell and Developmental Biology*, *10*, 994412. <https://doi.org/10.3389/fcell.2022.994412>

- Kulesa, P., Ellies, D. L., & Trainor, P. A. (2004). Comparative analysis of neural crest cell death, migration, and function during vertebrate embryogenesis. *Developmental Dynamics: An Official Publication of the American Association of Anatomists*, 229(1), 14–29. <https://doi.org/10.1002/dvdy.10485>
- Kuleshov, M. V., Jones, M. R., Rouillard, A. D., Fernandez, N. F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S. L., Jagodnik, K. M., Lachmann, A., McDermott, M. G., Monteiro, C. D., Gundersen, G. W., & Ma'ayan, A. (2016). Enrichr: A comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Research*, 44(W1), W90–97. <https://doi.org/10.1093/nar/gkw377>
- Kurogi, Y., Matsuo, Y., Mihara, Y., Yagi, H., Shigaki-Miyamoto, K., Toyota, S., Azuma, Y., Igarashi, M., & Tani, T. (2014). Identification of a chemical inhibitor for nuclear speckle formation: Implications for the function of nuclear speckles in regulation of alternative pre-mRNA splicing. *Biochemical and Biophysical Research Communications*, 446(1), 119–124. <https://doi.org/10.1016/j.bbrc.2014.02.060>
- Lagrange, T., Kapanidis, A. N., Tang, H., Reinberg, D., & Ebricht, R. H. (1998). New core promoter element in RNA polymerase II-dependent transcription: Sequence-specific DNA binding by transcription factor IIB. *Genes & Development*, 12(1), 34–44. <https://doi.org/10.1101/gad.12.1.34>
- Lambowitz, A. M., & Zimmerly, S. (2011). Group II introns: Mobile ribozymes that invade DNA. *Cold Spring Harbor Perspectives in Biology*, 3(8), a003616. <https://doi.org/10.1101/cshperspect.a003616>
- Lee, J.-H., Maskos, K., & Huber, R. (2009). Structural and functional studies of the yeast class II Hda1 histone deacetylase complex. *Journal of Molecular Biology*, 391(4), 744–757. <https://doi.org/10.1016/j.jmb.2009.06.059>
- Lesne, A., Baudement, M.-O., Rebouissou, C., & Forné, T. (2019). Exploring Mammalian Genome within Phase-Separated Nuclear Bodies: Experimental Methods and Implications for Gene Expression. *Genes*, 10(12), 1049. <https://doi.org/10.3390/genes10121049>
- Lian, S. L., Li, S., Abadal, G. X., Pauley, B. A., Fritzler, M. J., & Chan, E. K. L. (2009). The C-terminal half of human Ago2 binds to multiple GW-rich regions of GW182 and requires GW182 to mediate silencing. *RNA (New York, N.Y.)*, 15(5), 804–813. <https://doi.org/10.1261/rna.1229409>
- Liang, K., Smith, E. R., Aoi, Y., Stoltz, K. L., Katagi, H., Woodfin, A. R., Rendleman, E. J., Marshall, S. A., Murray, D. C., Wang, L., Ozark, P. A., Mishra, R. K., Hashizume, R., Schiltz, G. E., & Shilatifard, A. (2018).

Targeting Processive Transcription Elongation via SEC Disruption for MYC-Induced Cancer Therapy. *Cell*, 175(3), 766-779.e17.  
<https://doi.org/10.1016/j.cell.2018.09.027>

- Ligne primitive. (2023). In *Wikipédia*.  
[https://fr.wikipedia.org/w/index.php?title=Ligne\\_primitive&oldid=200362322](https://fr.wikipedia.org/w/index.php?title=Ligne_primitive&oldid=200362322)
- Lin, S., Coutinho-Mansfield, G., Wang, D., Pandit, S., & Fu, X.-D. (2008). The splicing factor SC35 has an active role in transcriptional elongation. *Nature Structural & Molecular Biology*, 15(8), 819–826.  
<https://doi.org/10.1038/nsmb.1461>
- Litchfield, D. W. (2003). Protein kinase CK2: Structure, regulation and role in cellular decisions of life and death. *The Biochemical Journal*, 369(Pt 1), 1–15.  
<https://doi.org/10.1042/BJ20021469>
- Liu, J., Carmell, M. A., Rivas, F. V., Marsden, C. G., Thomson, J. M., Song, J.-J., Hammond, S. M., Joshua-Tor, L., & Hannon, G. J. (2004). Argonaute2 is the catalytic engine of mammalian RNAi. *Science (New York, N.Y.)*, 305(5689), 1437–1441. <https://doi.org/10.1126/science.1102513>
- Liu, Y., Harmelink, C., Peng, Y., Chen, Y., Wang, Q., & Jiao, K. (2014). CHD7 interacts with BMP R-SMADs to epigenetically regulate cardiogenesis in mice. *Human Molecular Genetics*, 23(8), 2145–2156.  
<https://doi.org/10.1093/hmg/ddt610>
- Lorch, Y., & Kornberg, R. D. (2017). Chromatin-remodeling for transcription. *Quarterly Reviews of Biophysics*, 50, e5.  
<https://doi.org/10.1017/S003358351700004X>
- Louder, R. K., He, Y., López-Blanco, J. R., Fang, J., Chacón, P., & Nogales, E. (2016). Structure of promoter-bound TFIID and model of human pre-initiation complex assembly. *Nature*, 531(7596), 604–609.  
<https://doi.org/10.1038/nature17394>
- Lu, J., Wu, T., Zhang, B., Liu, S., Song, W., Qiao, J., & Ruan, H. (2021). Types of nuclear localization signals and mechanisms of protein import into the nucleus. *Cell Communication and Signaling*, 19(1), 60.  
<https://doi.org/10.1186/s12964-021-00741-y>
- Lussier, M. P., Gu, X., Lu, W., & Roche, K. W. (2014). Casein kinase 2 phosphorylates GluA1 and regulates its surface expression. *The European Journal of Neuroscience*, 39(7), 1148–1158. <https://doi.org/10.1111/ejn.12494>

- Manley, J. L., & Krainer, A. R. (2010). A rational nomenclature for serine/arginine-rich protein splicing factors (SR proteins). *Genes & Development*, *24*(11), 1073–1074. <https://doi.org/10.1101/gad.1934910>
- Marcos, S., Sarfati, J., Leroy, C., Fouveaut, C., Parent, P., Metz, C., Wolczynski, S., Gérard, M., Bieth, E., Kurtz, F., Verier-Mine, O., Perrin, L., Archambeaud, F., Cabrol, S., Rodien, P., Hove, H., Prescott, T., Lacombe, D., Christin-Maitre, S., ... Dodé, C. (2014). The prevalence of CHD7 missense versus truncating mutations is higher in patients with Kallmann syndrome than in typical CHARGE patients. *The Journal of Clinical Endocrinology and Metabolism*, *99*(10), E2138-2143. <https://doi.org/10.1210/jc.2014-2110>
- Margueron, R., & Reinberg, D. (2011). The Polycomb complex PRC2 and its mark in life. *Nature*, *469*(7330), 343–349. <https://doi.org/10.1038/nature09784>
- Marie, C., Clavairoly, A., Frah, M., Hmidan, H., Yan, J., Zhao, C., Van Steenwinckel, J., Daveau, R., Zalc, B., Hassan, B., Thomas, J.-L., Gressens, P., Ravassard, P., Moszer, I., Martin, D. M., Lu, Q. R., & Parras, C. (2018). Oligodendrocyte precursor survival and differentiation requires chromatin remodeling by Chd7 and Chd8. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(35), E8246–E8255. <https://doi.org/10.1073/pnas.1802620115>
- Martin, F. J., Amode, M. R., Aneja, A., Austine-Orimoloye, O., Azov, A. G., Barnes, I., Becker, A., Bennett, R., Berry, A., Bhai, J., Bhurji, S. K., Bignell, A., Boddu, S., Branco Lins, P. R., Brooks, L., Ramaraju, S. B., Charkhchi, M., Cockburn, A., Da Rin Fiorretto, L., ... Flicek, P. (2023). Ensembl 2023. *Nucleic Acids Research*, *51*(D1), D933–D941. <https://doi.org/10.1093/nar/gkac958>
- Martinez-Contreras, R., Fiset, J.-F., Nasim, F. H., Madden, R., Cordeau, M., & Chabot, B. (2006). Intronic binding sites for hnRNP A/B and hnRNP F/H proteins stimulate pre-mRNA splicing. *PLoS Biology*, *4*(2), e21. <https://doi.org/10.1371/journal.pbio.0040021>
- Matsui, M., Chu, Y., Zhang, H., Gagnon, K. T., Shaikh, S., Kuchimanchi, S., Manoharan, M., Corey, D. R., & Janowski, B. A. (2013). Promoter RNA links transcriptional regulation of inflammatory pathway genes. *Nucleic Acids Research*, *41*(22), 10086–10109. <https://doi.org/10.1093/nar/gkt777>
- Matveeva, E., Maiorano, J., Zhang, Q., Eteleeb, A. M., Convertini, P., Chen, J., Infantino, V., Stamm, S., Wang, J., Rouchka, E. C., & Fondufe-Mittendorf, Y.



- N. (2016). Involvement of PARP1 in the regulation of alternative splicing. *Cell Discovery*, 2(1), 1–19. <https://doi.org/10.1038/celldisc.2015.46>
- McKendrick, L., Thompson, E., Ferreira, J., Morley, S. J., & Lewis, J. D. (2001). Interaction of Eukaryotic Translation Initiation Factor 4G with the Nuclear Cap-Binding Complex Provides a Link between Nuclear and Cytoplasmic Functions of the m7Guanosine Cap. *Molecular and Cellular Biology*, 21(11), 3632–3641. <https://doi.org/10.1128/MCB.21.11.3632-3641.2001>
- Meers, M. P., Bryson, T. D., Henikoff, J. G., & Henikoff, S. (2019). Improved CUT&RUN chromatin profiling tools. *eLife*, 8, e46314. <https://doi.org/10.7554/eLife.46314>
- Meers, M. P., Tenenbaum, D., & Henikoff, S. (2019). Peak calling by Sparse Enrichment Analysis for CUT&RUN chromatin profiling. *Epigenetics and Chromatin*, 12(1). Scopus. <https://doi.org/10.1186/s13072-019-0287-4>
- Meister, G., Landthaler, M., Patkaniowska, A., Dorsett, Y., Teng, G., & Tuschl, T. (2004). Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Molecular Cell*, 15(2), 185–197. <https://doi.org/10.1016/j.molcel.2004.07.007>
- Meng, Y., Ma, X., Li, J., Ito, H., Oracz, K., Cai, J., & Shao, C. (2022). The novel activity of Argonautes in intron splicing: A transcriptome-wide survey in plants. *Journal of Plant Physiology*, 270, 153632. <https://doi.org/10.1016/j.jplph.2022.153632>
- Micucci, J. A., Sperry, E. D., & Martin, D. M. (2015). Chromodomain helicase DNA-binding proteins in stem cells and human developmental diseases. *Stem Cells and Development*, 24(8), 917–926. <https://doi.org/10.1089/scd.2014.0544>
- Mintz, P. J., Patterson, S. D., Neuwald, A. F., Spahr, C. S., & Spector, D. L. (1999). Purification and biochemical characterization of interchromatin granule clusters. *The EMBO Journal*, 18(15), 4308–4320. <https://doi.org/10.1093/emboj/18.15.4308>
- Miura, M., & Chen, H. (2020). CUT&RUN detects distinct DNA footprints of RNA polymerase II near the transcription start sites. *Chromosome Research: An International Journal on the Molecular, Supramolecular and Evolutionary Aspects of Chromosome Biology*, 28(3–4), 381–393. <https://doi.org/10.1007/s10577-020-09643-0>

- Moccia, A., Srivastava, A., Skidmore, J. M., Bernat, J. A., Wheeler, M., Chong, J. X., Nickerson, D., Bamshad, M., Hefner, M. A., Martin, D. M., & Bielas, S. L. (2018a). Genetic analysis of CHARGE syndrome identifies overlapping molecular biology. *Genetics in Medicine: Official Journal of the American College of Medical Genetics*, *20*(9), 1022–1029. <https://doi.org/10.1038/gim.2017.233>
- Moccia, A., Srivastava, A., Skidmore, J. M., Bernat, J. A., Wheeler, M., Chong, J. X., Nickerson, D., Bamshad, M., Hefner, M. A., Martin, D. M., & Bielas, S. L. (2018b). Genetic analysis of CHARGE syndrome identifies overlapping molecular biology. *Genetics in Medicine: Official Journal of the American College of Medical Genetics*, *20*(9), 1022–1029. <https://doi.org/10.1038/gim.2017.233>
- Moore, M. J., & Sharp, P. A. (1993). Evidence for two active sites in the spliceosome provided by stereochemistry of pre-mRNA splicing. *Nature*, *365*(6444), 364–368. <https://doi.org/10.1038/365364a0>
- Morgan, D., Bailey, M., Phelps, P., Bellman, S., Grace, A., & Wyse, R. (1993). Ear-nose-throat abnormalities in the CHARGE association. *Archives of Otolaryngology--Head & Neck Surgery*, *119*(1), 49–54. <https://doi.org/10.1001/archotol.1993.01880130051006>
- Muhammad, T., Pastore, S. F., Good, K., Ausió, J., & Vincent, J. B. (2023). Chromatin gatekeeper and modifier CHD proteins in development, and in autism and other neurological disorders. *Psychiatric Genetics*, *33*(6), 213–232. <https://doi.org/10.1097/YPG.0000000000000353>
- Munding, E. M., Shiue, L., Katzman, S., Donohue, J. P., & Ares, M. (2013). Competition between pre-mRNAs for the splicing machinery drives global regulation of splicing. *Molecular Cell*, *51*(3), 338–348. <https://doi.org/10.1016/j.molcel.2013.06.012>
- Muniz, L., Nicolas, E., & Trouche, D. (2021). RNA polymerase II speed: A key player in controlling and adapting transcriptome composition. *The EMBO Journal*, *40*(15), e105740. <https://doi.org/10.15252/embj.2020105740>
- Murawska, M., & Brehm, A. (2011). CHD chromatin remodelers and the transcription cycle. *Transcription*, *2*(6), 244–253. <https://doi.org/10.4161/trns.2.6.17840>
- Muto, Y., Pomeranz Krummel, D., Oubridge, C., Hernandez, H., Robinson, C. V., Neuhaus, D., & Nagai, K. (2004). The structure and biochemical properties of

- the human spliceosomal protein U1C. *Journal of Molecular Biology*, 341(1), 185–198. <https://doi.org/10.1016/j.jmb.2004.04.078>
- Nazer, E., Gómez Acuña, L., & Kornblihtt, A. R. (2022). Seeking the truth behind the myth: Argonaute tales from “nuclearland.” *Molecular Cell*, 82(3), 503–513. <https://doi.org/10.1016/j.molcel.2021.11.005>
- Nguyen, T. H. D., Galej, W. P., Bai, X., Savva, C. G., Newman, A. J., Scheres, S. H. W., & Nagai, K. (2015). The architecture of the spliceosomal U4/U6.U5 tri-snRNP. *Nature*, 523(7558), 47–52. <https://doi.org/10.1038/nature14548>
- Niehrs, C., & Luke, B. (2020). Regulatory R-loops as facilitators of gene expression and genome stability. *Nature Reviews. Molecular Cell Biology*, 21(3), 167–178. <https://doi.org/10.1038/s41580-019-0206-3>
- Nishi, K., Nishi, A., Nagasawa, T., & Ui-Tei, K. (2013). Human TNRC6A is an Argonaute-navigator protein for microRNA-mediated gene silencing in the nucleus. *RNA*, 19(1), 17–35. <https://doi.org/10.1261/rna.034769.112>
- Ohr, T., Muetze, J., & Schwillle, P. S. and P. (2011, December 31). *Intracellular Localization and Routing of miRNA and RNAi Pathway Components*. Current Topics in Medicinal Chemistry. <http://www.eurekaselect.com/89371/article>
- Ohr, T., Mütze, J., Staroske, W., Weinmann, L., Höck, J., Crell, K., Meister, G., & Schwillle, P. (2008). Fluorescence correlation spectroscopy and fluorescence cross-correlation spectroscopy reveal the cytoplasmic origination of loaded nuclear RISC in vivo in human cells. *Nucleic Acids Research*, 36(20), 6439–6449. <https://doi.org/10.1093/nar/gkn693>
- Okuno, H., Renault Mihara, F., Ohta, S., Fukuda, K., Kurosawa, K., Akamatsu, W., Sanosaka, T., Kohyama, J., Hayashi, K., Nakajima, K., Takahashi, T., Wysocka, J., Kosaki, K., & Okano, H. (2017). CHARGE syndrome modeling using patient-iPSCs reveals defective migration of neural crest cells harboring CHD7 mutations. *eLife*, 6. <https://doi.org/10.7554/eLife.21114>
- O’Roak, B. J., Vives, L., Girirajan, S., Karakoc, E., Krumm, N., Coe, B. P., Levy, R., Ko, A., Lee, C., Smith, J. D., Turner, E. H., Stanaway, I. B., Vernot, B., Malig, M., Baker, C., Reilly, B., Akey, J. M., Borenstein, E., Rieder, M. J., ... Eichler, E. E. (2012). Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature*, 485(7397), 246. <https://doi.org/10.1038/nature10989>
- Padgett, R. A., Konarska, M. M., Grabowski, P. J., Hardy, S. F., & Sharp, P. A. (1984). Lariat RNA’s as Intermediates and Products in the Splicing of

- Messenger RNA precursors. *Science*, 225(4665), 898–903.  
<https://doi.org/10.1126/science.6206566>
- Pagon, R. A., Graham, J. M., Zonana, J., & Yong, S. L. (1981). Coloboma, congenital heart disease, and choanal atresia with multiple anomalies: CHARGE association. *The Journal of Pediatrics*, 99(2), 223–227.  
[https://doi.org/10.1016/s0022-3476\(81\)80454-4](https://doi.org/10.1016/s0022-3476(81)80454-4)
- Pan, Q., Shai, O., Lee, L. J., Frey, B. J., & Blencowe, B. J. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature Genetics*, 40(12), 1413–1415.  
<https://doi.org/10.1038/ng.259>
- Parent, A., & Bisailon, M. (2006). Synergie entre les complexes de transcription et de maturation des ARN messagers. *médecine/sciences*, 22(6–7), Article 6–7.  
<https://doi.org/10.1051/medsci/20062267626>
- Park, M. S., Phan, H.-D., Busch, F., Hinckley, S. H., Brackbill, J. A., Wysocki, V. H., & Nakanishi, K. (2017). Human Argonaute3 has slicer activity. *Nucleic Acids Research*, 45(20), 11867–11877. <https://doi.org/10.1093/nar/gkx916>
- Park, M. S., Sim, G., Kehling, A. C., & Nakanishi, K. (2020). Human Argonaute2 and Argonaute3 are catalytically activated by different lengths of guide RNA. *Proceedings of the National Academy of Sciences of the United States of America*, 117(46), 28576–28578. <https://doi.org/10.1073/pnas.2015026117>
- Pauli, S., Bajpai, R., & Borchers, A. (2017). CHARGEd with neural crest defects. *American Journal of Medical Genetics. Part C, Seminars in Medical Genetics*, 175(4), 478–486. <https://doi.org/10.1002/ajmg.c.31584>
- Perry, R. P. (2005). The architecture of mammalian ribosomal protein promoters. *BMC Evolutionary Biology*, 5, 15. <https://doi.org/10.1186/1471-2148-5-15>
- Phair, R. D., & Misteli, T. (2000). High mobility of proteins in the mammalian cell nucleus—Nature. *Nature*, 404(6778), 604–609.  
<https://doi.org/10.1038/35007077>
- Pilon, N. (2016). Pigmentation-based insertional mutagenesis is a simple and potent screening approach for identifying neurocristopathy-associated genes in mice. *Rare Diseases (Austin, Tex.)*, 4(1), e1156287.  
<https://doi.org/10.1080/21675511.2016.1156287>

- Pilon, N. (2021). Treatment and Prevention of Neurocristopathies. *Trends in Molecular Medicine*, 27(5), 451–468. <https://doi.org/10.1016/j.molmed.2021.01.009>
- Pilon, N., Raiwet, D., Viger, R. S., & Silversides, D. W. (2008). Novel pre- and post-gastrulation expression of Gata4 within cells of the inner cell mass and migratory neural crest cells. *Developmental Dynamics: An Official Publication of the American Association of Anatomists*, 237(4), 1133–1143. <https://doi.org/10.1002/dvdy.21496>
- Pinna, L. A., & Meggio, F. (1997). Protein kinase CK2 (“casein kinase-2”) and its implication in cell division and proliferation. *Progress in Cell Cycle Research*, 3, 77–97.
- Piunti, A., & Shilatifard, A. (2016). Epigenetic balance of gene expression by Polycomb and COMPASS families. *Science (New York, N.Y.)*, 352(6290), aad9780. <https://doi.org/10.1126/science.aad9780>
- Pla, P., & Monsoro-Burq, A. H. (2018). The neural border: Induction, specification and maturation of the territory that generates neural crest cells. *Developmental Biology*, 444, S36–S46. <https://doi.org/10.1016/j.ydbio.2018.05.018>
- Plouhinec, J.-L., Roche, D. D., Pegoraro, C., Figueiredo, A. L., Maczkowiak, F., Brunet, L. J., Milet, C., Vert, J.-P., Pollet, N., Harland, R. M., & Monsoro-Burq, A. H. (2014). Pax3 and Zic1 trigger the early neural crest gene regulatory network by the direct activation of multiple key neural crest specifiers. *Developmental Biology*, 386(2), 461–472. <https://doi.org/10.1016/j.ydbio.2013.12.010>
- Proudfoot, N. J., Furger, A., & Dye, M. J. (2002). Integrating mRNA Processing with Transcription. *Cell*, 108(4), 501–512. [https://doi.org/10.1016/S0092-8674\(02\)00617-7](https://doi.org/10.1016/S0092-8674(02)00617-7)
- Ptashne, M., & Gann, A. (1997). Transcriptional activation by recruitment. *Nature*, 386(6625), 569–577. <https://doi.org/10.1038/386569a0>
- Reddy, N. C., Majidi, S. P., Kong, L., Namera, M., Ferguson, C. J., Moore, M., Goncalves, T. M., Liu, H.-K., Fitzpatrick, J. A. J., Zhao, G., Yamada, T., Bonni, A., & Gabel, H. W. (2021). CHARGE syndrome protein CHD7 regulates epigenomic activation of enhancers in granule cell precursors and gyrification of the cerebellum. *Nature Communications*, 12(1), 5702. <https://doi.org/10.1038/s41467-021-25846-3>

- Reddy, R., & Busch, H. (1988). Small Nuclear RNAs: RNA Sequences, Structure, and Modifications. In M. L. Birnstiel (Ed.), *Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles* (pp. 1–37). Springer. [https://doi.org/10.1007/978-3-642-73020-7\\_1](https://doi.org/10.1007/978-3-642-73020-7_1)
- Rentschler, M., Chen, Y., Pahl, J., Soria-Martinez, L., Braumüller, H., Brenner, E., Bischof, O., Röcken, M., & Wieder, T. (2018). Nuclear Translocation of Argonaute 2 in Cytokine-Induced Senescence. *Cellular Physiology and Biochemistry: International Journal of Experimental Cellular Physiology, Biochemistry, and Pharmacology*, *51*(3), 1103–1118. <https://doi.org/10.1159/000495490>
- Rinaldi, C., Pizzul, P., Longhese, M. P., & Bonetti, D. (2020). Sensing R-Loop-Associated DNA Damage to Safeguard Genome Stability. *Frontiers in Cell and Developmental Biology*, *8*, 618157. <https://doi.org/10.3389/fcell.2020.618157>
- Rother, M. B., Pellegrino, S., Smith, R., Gatti, M., Meisenberg, C., Wiegant, W. W., Luijsterburg, M. S., Imhof, R., Downs, J. A., Vertegaal, A. C. O., Huet, S., Altmeyer, M., & van Attikum, H. (2020). CHD7 and 53BP1 regulate distinct pathways for the re-ligation of DNA double-strand breaks. *Nature Communications*, *11*(1), 5775. <https://doi.org/10.1038/s41467-020-19502-5>
- Sainsbury, S., Bernecky, C., & Cramer, P. (2015). Structural basis of transcription initiation by RNA polymerase II. *Nature Reviews. Molecular Cell Biology*, *16*(3), 129–143. <https://doi.org/10.1038/nrm3952>
- Saint-André, V., Batsché, E., Rachez, C., & Muchardt, C. (2011). Histone H3 lysine 9 trimethylation and HP1 $\gamma$  favor inclusion of alternative exons. *Nature Structural & Molecular Biology*, *18*(3), 337–344. <https://doi.org/10.1038/nsmb.1995>
- Saitoh, N., Spahr, C. S., Patterson, S. D., Bubulya, P., Neuwald, A. F., & Spector, D. L. (2004). Proteomic analysis of interchromatin granule clusters. *Molecular Biology of the Cell*, *15*(8), 3876–3890. <https://doi.org/10.1091/mbc.e04-03-0253>
- Sarshad, A. A., Juan, A. H., Muler, A. I. C., Anastasakis, D. G., Wang, X., Genzor, P., Feng, X., Tsai, P.-F., Sun, H.-W., Haase, A. D., Sartorelli, V., & Hafner, M. (2018). Argonaute-miRNA Complexes Silence Target mRNAs in the Nucleus of Mammalian Stem Cells. *Molecular Cell*, *71*(6), 1040-1050.e8. <https://doi.org/10.1016/j.molcel.2018.07.020>

- Sasaki, T., Shiohama, A., Minoshima, S., & Shimizu, N. (2003). Identification of eight members of the Argonaute family in the human genome. *Genomics*, 82(3), 323–330. [https://doi.org/10.1016/s0888-7543\(03\)00129-0](https://doi.org/10.1016/s0888-7543(03)00129-0)
- Sato, T. S., Handa, A., Priya, S., Watal, P., Becker, R. M., & Sato, Y. (2019). Neurocristopathies: Enigmatic Appearances of Neural Crest Cell-derived Abnormalities. *Radiographics: A Review Publication of the Radiological Society of North America, Inc*, 39(7), 2085–2102. <https://doi.org/10.1148/rg.2019190086>
- Saxonov, S., Berg, P., & Brutlag, D. L. (2006). A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proceedings of the National Academy of Sciences of the United States of America*, 103(5), 1412–1417. <https://doi.org/10.1073/pnas.0510310103>
- Sayed, M., Kim, S. O., Salh, B. S., Issinger, O.-G., & Pelech, S. L. (2000). Stress-induced Activation of Protein Kinase CK2 by Direct Interaction with p38 Mitogen-activated Protein Kinase \*. *Journal of Biological Chemistry*, 275(22), 16569–16573. <https://doi.org/10.1074/jbc.M000312200>
- Scarpa, E., Szabó, A., Bibonne, A., Theveneau, E., Parsons, M., & Mayor, R. (2015). Cadherin Switch during EMT in Neural Crest Cells Leads to Contact Inhibition of Locomotion via Repolarization of Forces. *Developmental Cell*, 34(4), 421–434. <https://doi.org/10.1016/j.devcel.2015.06.012>
- Schier, A. C., & Taatjes, D. J. (2020). Structure and mechanism of the RNA polymerase II transcription machinery. *Genes & Development*, 34(7–8), 465–488. <https://doi.org/10.1101/gad.335679.119>
- Schnetz, M. P., Bartels, C. F., Shastri, K., Balasubramanian, D., Zentner, G. E., Balaji, R., Zhang, X., Song, L., Wang, Z., Laframboise, T., Crawford, G. E., & Scacheri, P. C. (2009a). Genomic distribution of CHD7 on chromatin tracks H3K4 methylation patterns. *Genome Research*, 19(4), 590–601. <https://doi.org/10.1101/gr.086983.108>
- Schnetz, M. P., Bartels, C. F., Shastri, K., Balasubramanian, D., Zentner, G. E., Balaji, R., Zhang, X., Song, L., Wang, Z., Laframboise, T., Crawford, G. E., & Scacheri, P. C. (2009b). Genomic distribution of CHD7 on chromatin tracks H3K4 methylation patterns. *Genome Research*, 19(4), 590–601. <https://doi.org/10.1101/gr.086983.108>
- Schnetz, M. P., Handoko, L., Akhtar-Zaidi, B., Bartels, C. F., Pereira, C. F., Fisher, A. G., Adams, D. J., Flicek, P., Crawford, G. E., Laframboise, T., Tesar, P.,

- Wei, C.-L., & Scacheri, P. C. (2010). CHD7 targets active gene enhancer elements to modulate ES cell-specific gene expression. *PLoS Genetics*, *6*(7), e1001023. <https://doi.org/10.1371/journal.pgen.1001023>
- Schor, I. E., Fiszbein, A., Petrillo, E., & Kornblihtt, A. R. (2013). Intragenic epigenetic changes modulate NCAM alternative splicing in neuronal differentiation. *The EMBO Journal*, *32*(16), 2264–2274. <https://doi.org/10.1038/emboj.2013.167>
- Schraivogel, D., & Meister, G. (2014). Import routes and nuclear functions of Argonaute and other small RNA-silencing proteins. *Trends in Biochemical Sciences*, *39*(9), 420–431. <https://doi.org/10.1016/j.tibs.2014.07.004>
- Schraivogel, D., Schindler, S. G., Danner, J., Kremmer, E., Pfaff, J., Hannus, S., Depping, R., & Meister, G. (2015a). Importin- $\beta$  facilitates nuclear import of human GW proteins and balances cytoplasmic gene silencing protein levels. *Nucleic Acids Research*, *43*(15), 7447–7461. <https://doi.org/10.1093/nar/gkv705>
- Schraivogel, D., Schindler, S. G., Danner, J., Kremmer, E., Pfaff, J., Hannus, S., Depping, R., & Meister, G. (2015b). Importin- $\beta$  facilitates nuclear import of human GW proteins and balances cytoplasmic gene silencing protein levels. *Nucleic Acids Research*, *43*(15), 7447–7461. <https://doi.org/10.1093/nar/gkv705>
- Sentenac, A. (1985). Eukaryotic RNA polymerases. *CRC Critical Reviews in Biochemistry*, *18*(1), 31–90. <https://doi.org/10.3109/10409238509082539>
- Shapiro, M. B., & Senapathy, P. (1987). RNA splice junctions of different classes of eukaryotes: Sequence statistics and functional implications in gene expression. *Nucleic Acids Research*, *15*(17), 7155–7174. <https://doi.org/10.1093/nar/15.17.7155>
- Sharma, N. R., Wang, X., Majerciak, V., Ajiro, M., Kruhlak, M., Meyers, C., & Zheng, Z.-M. (2016). Cell Type- and Tissue Context-dependent Nuclear Distribution of Human Ago2. *The Journal of Biological Chemistry*, *291*(5), 2302–2309. <https://doi.org/10.1074/jbc.C115.695049>
- Shen, H., & Green, M. R. (2006). RS domains contact splicing signals and promote splicing by a common mechanism in yeast through humans. *Genes & Development*, *20*(13), 1755–1765. <https://doi.org/10.1101/gad.1422106>
- Shen, H., Zhu, Y., Wang, C., Yan, H., Teng, M., & Li, X. (2016). Structural and histone binding ability characterization of the ARB2 domain of a histone



- deacetylase Hda1 from *Saccharomyces cerevisiae*. *Scientific Reports*, 6. <https://doi.org/10.1038/srep33905>
- Shilatifard, A. (1998). The RNA polymerase II general elongation complex. *Biological Chemistry*, 379(1), 27–31. <https://doi.org/10.1515/bchm.1998.379.1.27>
- Sidorenkov, I., Komissarova, N., & Kashlev, M. (1998). Crucial role of the RNA:DNA hybrid in the processivity of transcription. *Molecular Cell*, 2(1), 55–64. [https://doi.org/10.1016/s1097-2765\(00\)80113-6](https://doi.org/10.1016/s1097-2765(00)80113-6)
- Simões-Costa, M., & Bronner, M. E. (2015). Establishing neural crest identity: A gene regulatory recipe. *Development (Cambridge, England)*, 142(2), 242–257. <https://doi.org/10.1242/dev.105445>
- Sims, R. J., Millhouse, S., Chen, C.-F., Lewis, B. A., Erdjument-Bromage, H., Tempst, P., Manley, J. L., & Reinberg, D. (2007). Recognition of trimethylated histone H3 lysine 4 facilitates the recruitment of transcription postinitiation factors and pre-mRNA splicing. *Molecular Cell*, 28(4), 665–676. <https://doi.org/10.1016/j.molcel.2007.11.010>
- Skene, P. J., Henikoff, J. G., & Henikoff, S. (2018). Targeted in situ genome-wide profiling with high efficiency for low cell numbers. *Nature Protocols*, 13(5), 1006–1019. <https://doi.org/10.1038/nprot.2018.015>
- Skene, P. J., & Henikoff, S. (2017). An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *eLife*, 6, e21856. <https://doi.org/10.7554/eLife.21856>
- Smale, S. T., & Baltimore, D. (1989). The “initiator” as a transcription control element. *Cell*, 57(1), 103–113. [https://doi.org/10.1016/0092-8674\(89\)90176-1](https://doi.org/10.1016/0092-8674(89)90176-1)
- Soldatov, R., Kaucka, M., Kastriiti, M. E., Petersen, J., Chontorotzea, T., Englmaier, L., Akkuratova, N., Yang, Y., Häring, M., Dyachuk, V., Bock, C., Farlik, M., Piacentino, M. L., Boismoreau, F., Hilscher, M. M., Yokota, C., Qian, X., Nilsson, M., Bronner, M. E., ... Adameyko, I. (2019). Spatiotemporal structure of cell fate decisions in murine neural crest. *Science (New York, N.Y.)*, 364(6444), eaas9536. <https://doi.org/10.1126/science.aas9536>
- Song, J.-J., Liu, J., Tolia, N. H., Schneiderman, J., Smith, S. K., Martienssen, R. A., Hannon, G. J., & Joshua-Tor, L. (2003). The crystal structure of the Argonaute2 PAZ domain reveals an RNA binding motif in RNAi effector complexes. *Nature Structural Biology*, 10(12), 1026–1032. <https://doi.org/10.1038/nsb1016>

- Stathopoulou, A., Wang, P., Thellier, C., Kelly, R. G., Zheng, D., & Scambler, P. J. (2023). CHARGE syndrome-associated CHD7 acts at ISL1-regulated enhancers to modulate second heart field gene expression. *Cardiovascular Research*, *119*(11), 2089–2105. <https://doi.org/10.1093/cvr/cvad059>
- Steitz, T. A., & Steitz, J. A. (1993). A general two-metal-ion mechanism for catalytic RNA. *Proceedings of the National Academy of Sciences*, *90*(14), 6498–6502. <https://doi.org/10.1073/pnas.90.14.6498>
- Su, H., Trombly, M. I., Chen, J., & Wang, X. (2009a). Essential and overlapping functions for mammalian Argonautes in microRNA silencing. *Genes & Development*, *23*(3), 304–317. <https://doi.org/10.1101/gad.1749809>
- Su, H., Trombly, M. I., Chen, J., & Wang, X. (2009b). Essential and overlapping functions for mammalian Argonautes in microRNA silencing. *Genes & Development*, *23*(3), 304–317. <https://doi.org/10.1101/gad.1749809>
- Sugnet, C. W., Kent, W. J., Ares, M., & Haussler, D. (2004). Transcriptome and genome conservation of alternative splicing events in humans and mice. *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, 66–77. [https://doi.org/10.1142/9789812704856\\_0007](https://doi.org/10.1142/9789812704856_0007)
- Sun, H.-L., Cui, R., Zhou, J., Teng, K.-Y., Hsiao, Y.-H., Nakanishi, K., Fassan, M., Luo, Z., Shi, G., Tili, E., Kutay, H., Lovat, F., Vicentini, C., Huang, H.-L., Wang, S.-W., Kim, T., Zanesi, N., Jeon, Y.-J., Lee, T. J., ... Croce, C. M. (2016). ERK Activation Globally Downregulates miRNAs through Phosphorylating Exportin-5. *Cancer Cell*, *30*(5), 723–736. <https://doi.org/10.1016/j.ccell.2016.10.001>
- Sun, Q., Carrasco, Y. P., Hu, Y., Guo, X., Mirzaei, H., Macmillan, J., & Chook, Y. M. (2013). Nuclear export inhibition through covalent conjugation and hydrolysis of Leptomycin B by CRM1. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(4), 1303–1308. <https://doi.org/10.1073/pnas.1217203110>
- Suzuki, K., Bose, P., Leong-Quong, R. Y., Fujita, D. J., & Riabowol, K. (2010). REAP: A two minute cell fractionation method. *BMC Research Notes*, *3*, 294. <https://doi.org/10.1186/1756-0500-3-294>
- Szabó, A., & Mayor, R. (2018). Mechanisms of Neural Crest Migration. *Annual Review of Genetics*, *52*, 43–63. <https://doi.org/10.1146/annurev-genet-120417-031559>

- Tai, H. H., Geisterfer, M., Bell, J. C., Moniwa, M., Davie, J. R., Boucher, L., & McBurney, M. W. (2003). CHD1 associates with NCoR and histone deacetylase as well as with RNA splicing proteins. *Biochemical and Biophysical Research Communications*, *308*(1), 170–176. [https://doi.org/10.1016/s0006-291x\(03\)01354-8](https://doi.org/10.1016/s0006-291x(03)01354-8)
- Takahashi, Y., Sipp, D., & Enomoto, H. (2013). Tissue Interactions in Neural Crest Cell Development and Disease. *Science*, *341*(6148), 860–863. <https://doi.org/10.1126/science.1230717>
- Taliaferro, J. M., Aspden, J. L., Bradley, T., Marwha, D., Blanchette, M., & Rio, D. C. (2013). Two new and distinct roles for *Drosophila* Argonaute-2 in the nucleus: Alternative pre-mRNA splicing and transcriptional repression. *Genes & Development*, *27*(4), 378–389. <https://doi.org/10.1101/gad.210708.112>
- Tam, A. S., & Stirling, P. C. (2019). Splicing, genome stability and disease: Splice like your genome depends on it! *Current Genetics*, *65*(4), 905–912. <https://doi.org/10.1007/s00294-019-00964-0>
- Tarallo, R., Giurato, G., Bruno, G., Ravo, M., Rizzo, F., Salvati, A., Ricciardi, L., Marchese, G., Cordella, A., Rocco, T., Gigantino, V., Pierri, B., Cimmino, G., Milanese, L., Ambrosino, C., Nyman, T. A., Nassa, G., & Weisz, A. (2017). The nuclear receptor ER $\beta$  engages AGO2 in regulation of gene transcription, RNA splicing and RISC loading. *Genome Biology*, *18*(1), 189. <https://doi.org/10.1186/s13059-017-1321-0>
- Tari, M. (2018). *Etude du rôle des facteurs d'épissage à domaines UHM dans la régulation de l'épissage alternatif*. <https://theses.hal.science/tel-03051836v1/file/2018SACLE041.pdf>
- Technau, U., & Scholz, C. B. (2003). Origin and evolution of endoderm and mesoderm. *The International Journal of Developmental Biology*, *47*(7–8), 531–539.
- Ule, J., & Blencowe, B. J. (2019). Alternative Splicing Regulatory Networks: Functions, Mechanisms, and Evolution. *Molecular Cell*, *76*(2), 329–345. <https://doi.org/10.1016/j.molcel.2019.09.017>
- Ule, J., Ule, A., Spencer, J., Williams, A., Hu, J.-S., Cline, M., Wang, H., Clark, T., Fraser, C., Ruggiu, M., Zeeberg, B. R., Kane, D., Weinstein, J. N., Blume, J., & Darnell, R. B. (2005). Nova regulates brain-specific splicing to shape the synapse. *Nature Genetics*, *37*(8), 844–852. <https://doi.org/10.1038/ng1610>

- van Ravenswaaij-Arts, C. M. A., Blake, K., Hoefsloot, L., & Verloes, A. (2015). Clinical utility gene card for: CHARGE syndrome - update 2015. *European Journal of Human Genetics: EJHG*, 23(11). <https://doi.org/10.1038/ejhg.2015.15>
- Vissers, L. E. L. M., van Ravenswaaij, C. M. A., Admiraal, R., Hurst, J. A., de Vries, B. B. A., Janssen, I. M., van der Vliet, W. A., Huys, E. H. L. P. G., de Jong, P. J., Hamel, B. C. J., Schoenmakers, E. F. P. M., Brunner, H. G., Veltman, J. A., & van Kessel, A. G. (2004a). Mutations in a new member of the chromodomain gene family cause CHARGE syndrome. *Nature Genetics*, 36(9), 955–957. <https://doi.org/10.1038/ng1407>
- Vissers, L. E. L. M., van Ravenswaaij, C. M. A., Admiraal, R., Hurst, J. A., de Vries, B. B. A., Janssen, I. M., van der Vliet, W. A., Huys, E. H. L. P. G., de Jong, P. J., Hamel, B. C. J., Schoenmakers, E. F. P. M., Brunner, H. G., Veltman, J. A., & van Kessel, A. G. (2004b). Mutations in a new member of the chromodomain gene family cause CHARGE syndrome. *Nature Genetics*, 36(9), 955–957. <https://doi.org/10.1038/ng1407>
- Vo ngoc, L., Wang, Y.-L., Kassavetis, G. A., & Kadonaga, J. T. (2017). The punctilious RNA polymerase II core promoter. *Genes & Development*, 31(13), 1289–1301. <https://doi.org/10.1101/gad.303149.117>
- Vos, S. M., Farnung, L., Urlaub, H., & Cramer, P. (2018). Structure of paused transcription complex Pol II-DSIF-NELF. *Nature*, 560(7720), 601–606. <https://doi.org/10.1038/s41586-018-0442-2>
- Wagstaff, K. M., Sivakumaran, H., Heaton, S. M., Harrich, D., & Jans, D. A. (2012). Ivermectin is a specific inhibitor of importin  $\alpha/\beta$ -mediated nuclear import able to inhibit replication of HIV-1 and dengue virus. *Biochemical Journal*, 443(Pt 3), 851–856. <https://doi.org/10.1042/BJ20120150>
- Wahl, M. C., Will, C. L., & Lührmann, R. (2009). The spliceosome: Design principles of a dynamic RNP machine. *Cell*, 136(4), 701–718. <https://doi.org/10.1016/j.cell.2009.02.009>
- Wang, E. T., Sandberg, R., Luo, S., Khrebtkova, I., Zhang, L., Mayr, C., Kingsmore, S. F., Schroth, G. P., & Burge, C. B. (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature*, 456(7221), 470–476. <https://doi.org/10.1038/nature07509>
- Wang, H., Fan, Z., Shliaha, P. V., Miele, M., Hendrickson, R. C., Jiang, X., & Helin, K. (2023). H3K4me3 regulates RNA polymerase II promoter-proximal

- pause-release. *Nature*, 615(7951), 339–348. <https://doi.org/10.1038/s41586-023-05780-8>
- Wang, Y., Yu, Y., Pang, Y., Yu, H., Zhang, W., Zhao, X., & Yu, J. (2021). The distinct roles of zinc finger CCHC-type (ZCCHC) superfamily proteins in the regulation of RNA metabolism. *RNA Biology*, 18(12), 2107–2126. <https://doi.org/10.1080/15476286.2021.1909320>
- Watabe, M., & Nakaki, T. (2011). Protein kinase CK2 regulates the formation and clearance of aggresomes in response to stress. *Journal of Cell Science*, 124(9), 1519–1532. <https://doi.org/10.1242/jcs.081778>
- Wei, H., Yang, W., Hong, H., Yan, Z., Qin, H., & Benveniste, E. N. (2021). Protein Kinase CK2 Regulates B Cell Development and Differentiation. *Journal of Immunology (Baltimore, Md.: 1950)*, 207(3), 799–808. <https://doi.org/10.4049/jimmunol.2100059>
- Weinmann, L., Höck, J., Ivacevic, T., Ohrt, T., Mütze, J., Schwille, P., Kremmer, E., Benes, V., Urlaub, H., & Meister, G. (2009). Importin 8 is a gene silencing factor that targets argonaute proteins to distinct mRNAs. *Cell*, 136(3), 496–507. <https://doi.org/10.1016/j.cell.2008.12.023>
- Wilkinson, M. E., Charenton, C., & Nagai, K. (2020). RNA Splicing by the Spliceosome. *Annual Review of Biochemistry*, 89, 359–388. <https://doi.org/10.1146/annurev-biochem-091719-064225>
- Will, C. L., & Lührmann, R. (2011). Spliceosome structure and function. *Cold Spring Harbor Perspectives in Biology*, 3(7), a003707. <https://doi.org/10.1101/cshperspect.a003707>
- Wilson, P. A., & Hemmati-Brivanlou, A. (1997). Vertebrate neural induction: Inducers, inhibitors, and a new synthesis. *Neuron*, 18(5), 699–710. [https://doi.org/10.1016/s0896-6273\(00\)80311-6](https://doi.org/10.1016/s0896-6273(00)80311-6)
- Wing, C. E., Fung, H. Y. J., & Chook, Y. M. (2022). Karyopherin-mediated nucleocytoplasmic transport. *Nature Reviews Molecular Cell Biology*, 23(5), Article 5. <https://doi.org/10.1038/s41580-021-00446-7>
- Wong, M. S., Kinney, J. B., & Krainer, A. R. (2018). Quantitative Activity Profile and Context Dependence of All Human 5' Splice Sites. *Molecular Cell*, 71(6), 1012–1026.e3. <https://doi.org/10.1016/j.molcel.2018.07.033>
- Woodage, T., Basrai, M. A., Baxevanis, A. D., Hieter, P., & Collins, F. S. (1997). Characterization of the CHD family of proteins. *Proceedings of the National*

*Academy of Sciences of the United States of America*, 94(21), 11472–11477.  
<https://doi.org/10.1073/pnas.94.21.11472>

- Wu, M. Y., Ramel, M.-C., Howell, M., & Hill, C. S. (2011). SNW1 Is a Critical Regulator of Spatial BMP Activity, Neural Plate Border Formation, and Neural Crest Specification in Vertebrate Embryos. *PLoS Biology*, 9(2), e1000593. <https://doi.org/10.1371/journal.pbio.1000593>
- Xiao, C. Y., Hübner, S., & Jans, D. A. (1997). SV40 large tumor antigen nuclear import is regulated by the double-stranded DNA-dependent protein kinase site (serine 120) flanking the nuclear localization sequence. *The Journal of Biological Chemistry*, 272(35), 22191–22198.  
<https://doi.org/10.1074/jbc.272.35.22191>
- Xu, S., Lai, S.-K., Sim, D. Y., Ang, W. S. L., Li, H. Y., & Roca, X. (2022). SRRM2 organizes splicing condensates to regulate alternative splicing. *Nucleic Acids Research*, 50(15), 8599–8614. <https://doi.org/10.1093/nar/gkac669>
- Yamaguchi, Y., Shibata, H., & Handa, H. (2013). Transcription elongation factors DSIF and NELF: Promoter-proximal pausing and beyond. *Biochimica Et Biophysica Acta*, 1829(1), 98–104.  
<https://doi.org/10.1016/j.bbagr.2012.11.007>
- Yan, K. S., Yan, S., Farooq, A., Han, A., Zeng, L., & Zhou, M.-M. (2003). Structure and conserved RNA binding of the PAZ domain. *Nature*, 426(6965), 468–474. <https://doi.org/10.1038/nature02129>
- Yan, Q., & Sarma, K. (2020). MapR: A method for identifying native R-loops genome-wide. *Current Protocols in Molecular Biology*, 130(1), e113.  
<https://doi.org/10.1002/cpmb.113>
- Yan, Q., Shields, E. J., Bonasio, R., & Sarma, K. (2019). Mapping Native R-Loops Genome-wide Using a Targeted Nuclease Approach. *Cell Reports*, 29(5), 1369–1380.e5. <https://doi.org/10.1016/j.celrep.2019.09.052>
- Yan, S., Thienthanasit, R., Chen, D., Engelen, E., Brühl, J., Crossman, D. K., Kesterson, R., Wang, Q., Bouazoune, K., & Jiao, K. (2020). CHD7 regulates cardiovascular development through ATP-dependent and -independent activities. *Proceedings of the National Academy of Sciences of the United States of America*, 117(46), 28847–28858.  
<https://doi.org/10.1073/pnas.2005222117>
- Yao, H., Hannum, D. F., Zhai, Y., Hill, S. F., Albanus, R. D., Oliveira, Lou, W., Skidmore, J. M., Sanchez, G., Saiakhova, A., Bielas, S. L., Scacheri, P.,

- Ljungman, M., Parker, S. C. J., & Martin, D. M. (2020). CHD7 promotes neural progenitor differentiation in embryonic stem cells via altered chromatin accessibility and nascent gene expression. *Scientific Reports*, *10*, 17445. <https://doi.org/10.1038/s41598-020-74537-4>
- York, J. R., & McCauley, D. W. (2020). The origin and evolution of vertebrate neural crest cells. *Open Biology*, *10*(1), 190285. <https://doi.org/10.1098/rsob.190285>
- Younger, S. T., & Corey, D. R. (2011). Transcriptional gene silencing in mammalian cells by miRNA mimics that target gene promoters. *Nucleic Acids Research*, *39*(13), 5682–5691. <https://doi.org/10.1093/nar/gkr155>
- Zahler, A. M., Lane, W. S., Stolk, J. A., & Roth, M. B. (1992). SR proteins: A conserved family of pre-mRNA splicing factors. *Genes & Development*, *6*(5), 837–847. <https://doi.org/10.1101/gad.6.5.837>
- Zamore, P. D., Patton, J. G., & Green, M. R. (1992). Cloning and domain structure of the mammalian splicing factor U2AF. *Nature*, *355*(6361), 609–614. <https://doi.org/10.1038/355609a0>
- Zhang, L., Zhang, Y., Chen, Y., Gholamalamdari, O., Wang, Y., Ma, J., & Belmont, A. S. (2021). TSA-seq reveals a largely conserved genome organization relative to nuclear speckles with small position changes tightly correlated with gene expression changes. *Genome Research*, *31*(2), 251–264. <https://doi.org/10.1101/gr.266239.120>
- Zhang, M., Gu, Z., Guo, S., Sun, Y., Ma, S., Yang, S., Guo, J., Fang, C., Shu, L., Ge, Y., Chen, Z., & Wang, B. (2024). SRRM2 phase separation drives assembly of nuclear speckle subcompartments. *Cell Reports*, *43*(3), 113827. <https://doi.org/10.1016/j.celrep.2024.113827>
- Zhang, X., Zhan, X., Bian, T., Yang, F., Li, P., Lu, Y., Xing, Z., Fan, R., Zhang, Q. C., & Shi, Y. (2024). Structural insights into branch site proofreading by human spliceosome. *Nature Structural & Molecular Biology*, 1–11. <https://doi.org/10.1038/s41594-023-01188-0>
- Zhou, X., Wu, W., Li, H., Cheng, Y., Wei, N., Zong, J., Feng, X., Xie, Z., Chen, D., Manley, J. L., Wang, H., & Feng, Y. (2014). Transcriptome analysis of alternative splicing events regulated by SRSF10 reveals position-dependent splicing modulation. *Nucleic Acids Research*, *42*(6), 4019–4030. <https://doi.org/10.1093/nar/gkt1387>

- Zhou, Z., & Fu, X.-D. (2013). Regulation of splicing by SR proteins and SR protein-specific kinases. *Chromosoma*, *122*(3), 191–207.  
<https://doi.org/10.1007/s00412-013-0407-z>
- Zhuang, Y., & Weiner, A. M. (1986). A compensatory base change in U1 snRNA suppresses a 5' splice site mutation. *Cell*, *46*(6), 827–835.  
[https://doi.org/10.1016/0092-8674\(86\)90064-4](https://doi.org/10.1016/0092-8674(86)90064-4)