



Caractérisation et analyse fonctionnelle des snoARN orphelins par la mise au point d'une méthode étudiant les interactions snoARN-ARN cibles

Dylan Moulis

► To cite this version:

Dylan Moulis. Caractérisation et analyse fonctionnelle des snoARN orphelins par la mise au point d'une méthode étudiant les interactions snoARN-ARN cibles. Biologie moléculaire. Université Paul Sabatier - Toulouse III, 2023. Français. <NNT : 2023TOU30150>. <tel-04331824>

HAL Id: tel-04331824

<https://theses.hal.science/tel-04331824v1>

Submitted on 8 Dec 2023

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

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



HAL Authorization



THÈSE

**En vue de l'obtention du
DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE
Délivré par l'Université Toulouse 3 - Paul Sabatier**

**Présentée et soutenue par
Dylan MOULIS**

Le 18 juillet 2023

**Caractérisation et analyse fonctionnelle des snoARN orphelins
par la mise au point d'une méthode étudiant les interactions
snoARN-ARN cibles.**

Ecole doctorale : **BSB - Biologie, Santé, Biotechnologies**

Spécialité : **GENETIQUE MOLECULAIRE**

Unité de recherche :

MCD - Molecular, Cellular and Developmental Biology Unit

Thèse dirigée par

Tamas KISS et Beata Erika JADY

Jury

M. Clément CARRÉ, Rapporteur

Mme Séverine MASSENET, Rapporteur

Mme Cécile BOUSQUET-ANTONELLI, Rapporteur

M. Stéphane LABIALLE, Examineur

Mme Séverine BOULON, Examinatrice

Mme Kerstin BYSTRICKY, Examinatrice

M. Tamas KISS, Co-directeur de thèse

Mme Beata JADY, Co-directrice de thèse

REMERCIEMENTS

[Chaque voyage débute par un choix]. Le choix de faire un doctorat, n'a jamais vraiment été clair pour moi. Ma fâcheuse tendance à surligner les inconvénients a failli m'écarter de ce choix, qui au départ, me semblait être une certitude. Merci à toutes les personnes de m'avoir conforté dans ce choix (coucou Alix). Parce qu'aujourd'hui, prenant conscience que ce voyage s'achève, je ne peux qu'affirmer que ce choix, était sans aucun doute le meilleur.

Dans un premier temps, je tiens à grandement remercier **Cécile Bousquet-Antonelli**, **Clément Carré**, et **Séverine Massenet**, pour avoir accepté d'être mes rapporteurs/rapportrices de thèse, d'avoir pris le temps d'évaluer mon manuscrit ainsi que ma soutenance. Un grand merci également à **Séverine Boulon** et **Stéphane Labialle**, qui se sont adonnés au rôle d'examinateurs/examinatrices pour jauger mes travaux de thèse, ainsi qu'à **Kerstin Bystricky** d'avoir accepté de présider le jury lors de la soutenance de mon doctorat. J'ai adoré l'échange qui a donné suite à ma présentation, merci pour la qualité de vos interrogations ainsi que pour votre bienveillance, qui ont permis une discussion scientifique très intéressante.

Je tiens une nouvelle fois à remercier **Stéphane Labialle**, ainsi que **David Umlauf**, qui ont accepté de faire partie de mon comité de thèse, comité qui m'a accompagné et soutenu tout au long de mon doctorat.

Après avoir vagabondé entre les ARN des bactéries et des végétaux, un grand merci à **Beata Jady** et **Tamas Kiss** de m'avoir donné l'opportunité de poursuivre mon cursus sur ces molécules, mais cette fois-ci avec les cellules humaines comme modèle d'étude. Merci à **Tamas** de m'avoir permis d'intégrer son équipe de recherche, et à **Beata** pour m'avoir encadré tout au long de ces quasi quatre années de doctorat. Au sein de l'équipe Kiss, j'ai pu faire la rencontre du célèbre duo composé de **Sylvain Egloff** et **Patrice Vitali**. Outre toutes les discussions scientifiques et coachings en biologie moléculaire qu'ils m'ont apporté, je les remercie chaleureusement pour leur côté humain exceptionnel, et ce bien malgré le fait qu'ils aient essayé de se débarrasser de moi à moult reprises (je m'en souviendrai de ce semi-marathon). Merci à **Patrice** pour ta bonne humeur constante, et de m'avoir fait découvrir les bons côtés de l'enseignement. Merci à **Sylvain** de m'avoir éduqué sur le ballon rond et le sport auto, même si tu restes complètement à la masse niveau pronos. Ce fut également à chaque fois un plaisir de manipuler à tes côtés, au labo. Merci d'avoir toujours réussi à mettre une ambiance musicale, même sans enceinte ou smartphone (y'a peut-être quelque chose à creuser niveau carrière d'ailleurs, je ne sais pas). Un énorme merci, enfin, à la dernière fine lame de cette équipe, **Amandine Ketele**. Merci **Amandine** pour ton allégresse, et toutes ces discussions,

scientifiques ou non. Merci également de m'avoir grandement aidé pour la réalisation de mon projet doctoral. Promis, un jour, j'arriverai à trouver un clonage qui te résistera.

Je tiens également à remercier l'équipe de **Jerôme Cavaillé**, l'équipe « jumelle » de celle des Kiss. Merci **Jerôme** pour cette collaboration et toutes ces discussions autour du SNORD115, on arrivera à la trouver, cette foutue cible. Un grand merci également à **Virginie** pour tout ton travail autour des souris, sans quoi ce projet SNORD115 n'aurait pas été possible. Tu t'es relayée avec **Jade** pour me former à ces aspects de la science. Merci à toutes les deux. Un merci à **Marie-line** également, d'avoir pris le temps de répondre à toutes mes interrogations, ainsi que de m'avoir aidé à garder la ligne avec toutes ces pâtisseries. Vous êtes une super équipe, au plaisir de vous recroiser, nous avons d'ailleurs quelques CRISPR burgers de retard !

Malgré le fait d'être entouré de ces deux équipes de choc, je n'ai pas pu m'empêcher de me balader dans les couloirs (un peu trop peut-être ?) de l'ex LBME. J'ai pu découvrir les personnes formidables qui composent les équipes **Henry-Henras**, **Plisson-Chastang**, et **Gleizes**. Merci **Célia**, **Simon** et **Anthony**, de m'avoir apporté à plusieurs reprises votre expertise sur mon projet, merci également pour ces dégustations de grands vins, ces bières du Lot et ces coachings footing. Je tiens à remercier naturellement le laboratoire MCD, ainsi que toutes les personnes qui font tourner cette unité. Merci également au LMGM, nos voisins de palier. Un merci tout particulier à **Marie-Pierre** et **Philou**, qui m'ont formé tout en accentuant mon désir de travailler dans la biologie. Merci de m'avoir encouragé à me diriger vers cette fabuleuse aventure scientifique et humaine qu'est le doctorat.

Il y a toujours un peu d'appréhension et de doutes quand l'on se dirige vers l'inconnu, une nouvelle unité, une nouvelle équipe, un nouveau projet. Mais qu'est-ce que ces sentiments se sont si rapidement évaporés lorsque j'ai franchi les portes du laboratoire. Vous êtes une sacrée bande de doctorants, merci de m'avoir accompagné tout au long de ces trois ans et demi.

Merci à **Alex**, **Cyril** et **Côme**, de m'avoir embarqué dans le monde de la grimpe. Une rouge flashée valait bien au moins une poulie.

Merci à tous pour ces séminaires au Dubliners, où ont eu lieu de nombreux fous rires, levés de coudes, et qui nous ont fait apprécier cette fabuleuse eau de l'upsi. C'est par ailleurs entre deux pintes (il lui fallait au moins ça) pour que **Laura** accepte de devenir ma Marraine. Merci **Laura** pour toute ton aide au niveau scientifique et administratif (bordel au secours) ainsi que pour toutes ces discussions sur la vie (au soleil évidemment) entre deux manip. Même si aujourd'hui, étant enfin adoubé à mon tour, un trait peut être tiré sur ce lien Marraine-Filleul,

je sais qu'il laisse place à une amitié et que l'on continuera à se recroiser. Je te supporte à fond et te souhaite pleins de réussites dans tes nouveaux projets.

Qui parle d'adoubement, parle rédaction de ce Manuscrit. J'ai eu la chance, durant ce long exercice, d'avoir partagé un bureau avec **Nathalie**. Merci **Nath** de m'avoir soutenu et supporté pendant ces longs, très loooongs (ou trop courts) mois de rédaction. Félicitations à nous deux, on l'a fait, finalement ! Outre cela, merci pour ces dialogues, fou rires et délires durant ces années de doctorat. Pleins de réussites pour ce nouveau projet sur les terres de Shakespeare.

Un grand merci également à **Sarah**, pour tous ces débats et discussions sur la vie, qui permettaient notamment d'optimiser les temps d'incubations. Merci de m'avoir fait comprendre qu'un Northern était mieux réussi s'il était fait avant l'aube. Je ne l'aurai jamais cru mais tu as (presque) réussi à me convertir à la Bretagne et aux films d'horreur.

Merci à **Marie**, tu vois finalement on l'a réussi ce 16, malgré le fait qu'on se soit fait appâter par les ravitaillements. Merci à **Manon** pour tous tes coachings et les fous rires, Merci à **Paulo** et **Anaïs** pour toutes ces discussions et pronos footballistiques, Merci à **Carine** et **Sophie**, pour ces BBF sans fin, Merci à **Olivier** pour ces English Breakfast et ces négociations avec Catherine de Médicis, Merci à **Alix** et **Claudie** de m'avoir accompagné voir Charles Leclerc en vrai, Merci à **Ramtin**, **Charlotte**, **Ali**, **Emilie**, **Maxime**, **Sophian**, **Yassin**, **Hussein**, **Baptiste**, **Romane**, **Clair**, **Julie**, **Joyce**, et tous ceux que j'ai oublié.

Merci aux copainings qui m'ont suivi depuis la L3 et le Master, **Papalexis**, **Dana**, **Joseph**, **Clément**, **Ludo**, **Ahmed**, **Valentin** et **Alix**. Heureux pour tous ces moments passés ensemble et qui continuent encore aujourd'hui. Merci à mes amis de longue date de m'avoir épaulé tout au long de ces années. Je n'aurais jamais pu tenir sans ces week-ends improvisés avec vous, qui me permettaient de m'évader. Un gros merci à **Adrien**, **Lucas**, **Léo**, **Frank**, **Mélissa**, **Thibault**, **Clément** et **Antho**.

Merci à ma famille, mes sœurs **Ambre** et **Jade** et mes parents, pour leur support et soutien durant ces longues études. Merci à la belle famille, **Marie**, **Bruno**, **Mathéo**, **Antoine**, **François** et **Colette** pour les bons moments et les encouragements.

Pour finir, un merci tout particulier à **Hélène**, pour m'avoir supporté, dans les deux sens du terme, tout au long de ce doctorat. Merci de partager ma vie.

TABLE DES MATIÈRES

REMERCIEMENTS	1
TABLE DES MATIÈRES	5
LISTE DES ABRÉVIATIONS	11
INTRODUCTION	14
I. LES PETITS ARN NUCLEOLAIRES	15
I. 1. TERRITOIRES NUCLEAIRES HEBERGEANT LES SNOARN	15
I. 1. 1. Le nucléole	15
I. 1. 2. Les corps de Cajal	17
I. 2. LES ARN DE LA CLASSE H/ACA	18
I. 2. 1. Les pseudouridylations	18
I. 2. 2. Les caractéristiques des ARN H/ACA	20
I. 2. 3. La Dyskérine	21
I. 2. 4. Reconnaissance de l'ARN substrat par les RNP H/ACA	23
I. 2. 5. Conformation structurale des RNP H/ACA	24
I. 2. 6. Biogenèse et transport intranucléaire des RNP H/ACA	27
I. 3. LES ARN DE LA CLASSE C/D	30
I. 3. 1. Les méthylations en 2'-O-ribose	30
I. 3. 2. Caractéristiques des RNP C/D	31
I. 3. 3 La Fibrillarine	33
I. 3. 4. Reconnaissance de l'ARN substrat par les RNP C/D	36
I. 3. 5. Conformation structurale des RNP C/D	36
I. 3. 6. Biogenèse et transport intranucléaire des RNP C/D	38
II. FONCTIONS CELLULAIRES DES RNP H/ACA ET C/D	42
II. 1 LES RNP H/ACA ET C/D, GUIDES DE MODIFICATIONS POST-TRANSCRIPTIONNELLES	42
II. 1. 1. Les méthylations sur les ARN de transfert	42
II. 1. 2. Les snoARN guident des acétylations	43
II. 2. ROLES DES SNORNP DANS LA MATURATION DES ARNr ET DANS LE MAINTIEN DES TELOMERES	45

II. 2. 1. Implication dans le chaperonnage et le clivage des ARNr	45
II. 2. 2. L'ARN de la Télomérase	50
II. 3. AUTRES FONCTIONS ASSOCIEES AUX SNOARN	51
II. 3. 1. Le concept de snoARN orphelin	51
II. 3. 2. Les SnoARN associés aux maladies	53
II. 3. 2. 1 Le syndrome Labrune	53
II. 3. 2. 2. Le syndrome d'Alazami	54
II. 3. 2. 3. Le syndrome de Prader-Willi	55
II. 3. 2. 4. SnoARN et Cancers	60
II. 3. 3. Fonctions non-canoniques associés aux snoARN	64
II. 3. 3. 1. SnoARN et ARNm	64
II. 3. 3. 2. SnoARN et virus	68
II. 3. 3. 3. Formes alternatives de maturation des snoARN	69
III. TECHNIQUES D'ETUDES DES CIBLES ET MODIFICATIONS DES SNOARNP	71
III. 1. METHODES DE DETECTION DES MODIFICATIONS POST-TRANSCRIPTIONNELLES GUIDEES PAR LES SNOARNP	71
III. 1. 1. Méthodes de détection des pseudouridines	71
III. 1. 1. 1. La méthode au CMCT	71
III. 1. 1. 2. Autres composés chimiques permettant l'identification des pseudouridines	74
III. 1. 1. 3. Utilisation de la Spectrométrie de masse pour l'identification des pseudouridines	74
III. 1. 1. 4. Les méthodes permettant de détecter les pseudouridines au niveau de l'ensemble du transcriptome	76
III. 1. 1. 5. Utilisation de Nanopores	81
III. 1. 2. Méthodes de détection des méthylations en 2'-O-ribose	82
III. 1. 2. 1. Techniques à faible débit basées sur des concentrations de dNTP limitantes	84
III. 1. 2. 2. Le 2'-OMe-Seq	85
III. 1. 2. 3 Le RiboMeth-seq	85
III. 1. 2. 4. RibOxi-Seq et Nm-Seq	87
III. 1. 2. 5. Détection des méthylations en 2'-O-ribose par Nanopores	90
III. 2. METHODES <i>IN SILICO</i> PERMETTANT DE PREDIRE LES INTERACTIONS ARN-ARN	91
III. 2. 1. Algorithmes de prédiction des cibles des snoARN H/ACA et C/D	91
III. 2. 1. 1. Prédiction des ARN cibles des snoARN H/ACA	91

III. 2. 1. 2. Prédiction des ARN cibles des snoARN C/D	92
III. 2. 2. Utilisations annexes de l'informatique dans l'étude des snoARN	93
III. 2. 2. 1. Recherche de nouveaux snoARN putatifs dans le génome	93
III. 2. 2. 2. Les bases de données sur les snoARN	94
III. 3. METHODES <i>IN VIVO</i> PERMETTANT D'ETUDIER LES INTERACTIONS SNOARN-ARN	94
III. 3. 1. Méthodes utilisant une protéine « appât »	94
III. 3. 1. 1. Le CLASH	94
III. 3. 1. 2. Le MARIO	96
III. 3. 2. Méthodes basées sur l'utilisation du psoralène	98
III. 3. 2. 1. Les méthodes LIGR-Seq, PARIS, SPLASH et COMRADES	98
III. 3. 2. 2. Avantages & Limites des méthodes ayant recourt au psoralène	101

RÉSULTATS **103**

I. MISE AU POINT D'UNE METHODE PERMETTANT DE CO-PURIFIER UN SNOARN AVEC SA CIBLE	104
I. 1. MISE AU POINT DE LA METHODE	104
I. 1. 1. Méthodes développées dans la littérature	104
I. 1. 2. Les différentes étapes composant la méthode développée	105
I. 2. ARN ETUDIES POUR LA MISE AU POINT DE LA METHODE	105
I. 2. 1. SCARNA10	105
I. 2. 2. SNORA55	109
I. 2. 3. Le snARN 7SK	111
I. 3. MISE AU POINT DES ETAPES COMPOSANT LA METHODE	114
I. 3. 1. Crosslink ARN-ARN <i>in vivo</i>	114
I. 3. 2. Purification d'un snoARN d'intérêt : utilisation des MagIC Beads	117
I. 3. 3. Validation intermédiaire de la méthode PIKA par RT-qPCR	120
I. 3. 4. Digestion des duplex ARN – ARN	120
I. 3. 5. Derniers tests de mise au point par séquençage haut-débit	123
I. 3. 6. Recherche des ARN interagissant avec 7SK par la méthode PIKA	123
I. 3. 7. Etude des interactions des ARN C/D avec le transcriptome	129
II. ETUDE DE L'INTERACTION DES ARN SNORA55 ET L'ARNr 18S	133

II. 1. GENERATION DE LA LIGNEE DEpletee POUR SNORA55 ET ETUDE DE L'INTERACTION SNORA55 – ARNr 18S.	133
II. 2. DETECTION DES PSEUDOURIDINES DE L'ARNr 18S PAR L'ETUDE D'UNE LIGNEE KO SNORA55	133
III. RECHERCHE DES CIBLES DE SNORD115, UN SNOARN IMPLIQUE DANS LE SYNDROME DE PRADER WILLI	137
III. 1. LE SNORD115, UN ARN SPECIFIQUEMENT EXPRIME DANS LE CERVEAU	137
III. 2. ETUDE DE L'INTERACTION SNORD115 – ARNm 5-HT2C	139
III. 2. 1. Mise au point de la purification de SNORD115	139
III. 2. 2. Obtention de cellules murines exprimant SNORD115	139
III. 2. 3 Etude de l'interaction physique entre SNORD115 et 5-HT2C	144
IV. ETUDE MECANISTIQUE ET FONCTIONNELLES DES SNOARN H/ACA	146
IV. 1. DECOUVERTE DE NOUVEAUX SNOARN H/ACA PUTATIFS	146
IV. 1. 1. Elaboration d'une liste de potentiels candidats	146
IV. 1. 2. Test d'accumulation des snoARN H/ACA putatifs dans les cellules humaines	148
IV. 1. 3. Le cas du snoARN putatif H/ACA 186	150
IV. 2. MECANISME DE PSEUDOURIDYLATION DE DEUX URIDINES ADJACENTES PAR DES SNOARN H/ACA HUMAINS	152
IV. 2. 1. Contexte des travaux effectués	152
IV. 2. 2. Guide RNA acrobatics: positioning consecutive uridines for pseudo- uridylation by H/ACA	154
<u>DISCUSSION</u>	1689
La méthode PIKA : Le bilan	170
Interactions de l'ARN 7SK avec le transcriptome	173
SNORD115, le snoARN orphelin par excellence	174
PIKA et la différenciation neuronale, des outils d'études de l'ARN SNORD116	175
Une interaction entre SNORA55 et l'ARNr 18S : Oui, mais comment ?	177
L'identification de nouveaux snoARN H/ACA : la fin d'une quête ?	178
ARN cibles et fonctions associées aux snoARN H/ACA 29, 120 et 186	179
Le snoARN H/ACA 186, un snoARN pas comme les autres	181

MATERIELS ET METHODES	183
Procédures générales	184
Extraction des ARN cellulaires	185
Crosslink au psoralène	185
Purification des ARN d'intérêts	185
Fragmentation, Ligation, et réversion du Crosslink des ARN	186
Méthodes de détection des ARN	186
Génération de la lignée SNORA55 KO	187
Cartographie de l'ARNr 18S au CMCT	188
Différentiation neuronale des cellules murines P19	188
Dissociation et isolement de cellules neuronales à partir de cerveaux murins	188
Approche Bio-informatiques	189
ANNEXES	191
RÉFÉRENCES BIBLIOGRAPHIQUES	195
RÉSUMÉS	223

LISTE DES **ABRÉVIATIONS**

ACE : antigène arcinoembryonnaire	GC : composant granulaire
ADAR : adenosine deaminase acting on	HPLC : chromatographie en phase liquide
ADN : acide désoxyribonucléique	haute performance
ADNc : ADN complémentaire	ICR : centre de l’empreinte
ARN : acide ribonucléique	ITS : Internal transcribed spacer
ARNds : ARN double brin / ARN bicaténaire	KD : Knockdown
ARNm : ARN messenger	KO : Knockout
ARNt : ARN de transfert	LAM : leucémie aiguë myéloblastique
CLASH : cross-linking ligation and	LIGR-Seq : Ligation of interacting RNA
sequencing of hybrids	followed by high-throughput Sequencing
CLIP : Cross-linking and immunoprecipitation	LSU : grande sous-unité
CMC : Carbodiimide	MARIO : mapping RNA interactome <i>in vivo</i>
CMCT : (N-cyclohexyl-N’-β-(4-methylmor-	MS : Spectrométrie de masse
pholinium) ethylcarbodiimide p-tosylate)	MVS : méthyl vinyl sulfone
CMV : Cytomégalo virus	NSCLC : cancer du poumon non à petites
COMRADES : cross-linking of matched	cellules
RNAs and deep sequencing	NGS : Séquençage nouvelle génération
CTD : carboxy-terminal domain	NLS : Nuclear localization signals
CTE : carboxy-terminal extension	NMD : Nonsense-Mediated Decay
DKC1 : Dyskérine	NoLS : Nucleolar localization signals
DFC : composant dense fibrillaire	NOR : Nucleolar Organizer Region
dNTP : désoxyribonucléotide triphosphate	nt : nucléotide
EBV : Epstein-Barr virus	NTE : amino terminal-extension
FBL : Fibrillarine	PARIS : Psoralen analysis of RNA interactions
FC : centre fibrillaire	and structures
GAR : glycine–arginine-rich	PC : chromatine périnucléolaire
	PCR : Réaction en chaîne par polymérase

PIKA : Psoralen <i>in vivo</i> crosslink RNA-RNA	SNORA : petit ARN nucléolaire à boîtes H/ACA
Pré-ARN ^m : précurseur de l'ARN messenger	SPLASH : Sequencing of psoralen crosslinked, ligated, And selected hybrids
Pré-ARN ^r : précurseur de l'ARN ribosomique	SSD : Shq1-Specific Domain
PUA : pseudouridine synthase and archaeosine transglycosylase domain	SSU : petite sous-unité
PUS : Pseudouridine synthases	TruB : pseudouridine synthase domain
PWS : Syndrome de Prader-Willi	T4-PNK : T4-Polynucléotide kinase
qPCR : PCR quantitative	UV : Ultraviolet
RAIN : RNA-protein Association and Interaction Networks	5-HT _{2C} : Récepteur à la sérotonine 2C
RISE : RNA interactome from sequencing experiments	Ψ : Pseudouridine
RNP : Ribonucléoparticule	
RTL-p : reverse transcription at low dNTP concentrations followed by PCR	
SAM : S-adénosylméthionine	
scaARN : petit ARN spécifique des corps de Cajal	
scaRNP : petite RNP spécifique des corps de Cajal	
snARN : ARN de l'épissage	
snRNP : RNP de l'épissage	
snoARN : petits ARN nucléolaires	
snoRNP : petites RNP nucléolaires	
SNORD : petit ARN nucléolaire à boîtes C/D	

INTRODUCTION

I. Les petits ARN nucléolaires

Les petits ARN nucléolaires (snoARN) sont une abondante famille d'ARN non-codants, majoritairement impliqués dans les modifications post-transcriptionnelles des ARN ribosomiques (ARNr) et de l'épissage (snARN). Ils s'associent avec des protéines spécifiques, formant des complexes ribonucléoprotéiques (snoRNP). Dans ces complexes, le snoARN a un rôle de guide en sélectionnant les nucléotides à modifier, et les protéines associées vont modifier chimiquement l'ARN sélectionné.

Les snoARN, s'accumulant dans les nucléoles et corps de Cajal des cellules, sont largement répandus chez les eucaryotes et les archées (Watkins *et al*, 2000a), mais absent chez les bactéries (Gaspin *et al*, 2000). Entre les différentes espèces, les snoARN possèdent des fonctions et des motifs conservés, permettant de les classer en deux grands groupes : les snoARN H/ACA et les snoARN C/D, respectivement associés aux pseudouridylation et méthylations en 2'-O-ribose (Kiss, 2002).

I. 1. Territoires nucléaires hébergeant les snoARN

I. 1. 1. Le nucléole

Le nucléole est le lieu de la synthèse et de la maturation des ARN ribosomiques de la cellule. C'est également au sein de ce compartiment nucléaire que se déroule l'assemblage des ARNr avec les protéines ribosomales. Le nucléole est le compartiment le plus volumineux du noyau des eucaryotes, et, à l'instar des autres territoires nucléaires, n'est pas délimité par une membrane. Le nombre de nucléoles présent dans le noyau varie selon le cycle cellulaire (Pederson, 2011; Lafontaine *et al*, 2021).

Le nucléole a été découvert au cours des années 1835-36 par microscopie à fond clair, et il aura fallu attendre plus d'un siècle pour découvrir son rôle primordial dans la synthèse des ribosomes (Lafontaine *et al*, 2021).

Les nucléoles s'organisent autour de régions chromosomiques bien particulières, appelées NOR (Nucleolar Organizer Region). Ces régions contiennent les ADN ribosomiques, nécessaires à la synthèse des ARNr par l'ARN polymérase I. Les nucléoles sont décomposés en trois sous-région distincte : le centre fibrillaire (FC), le composant dense fibrillaire (DFC) et composant granulaire (GC) (Huang, 2002) (**Figure 1**). Au sein d'un même GC, plusieurs FC sont présents, chacun entouré par un DFC, formant ainsi plusieurs unités fonctionnelles. Les

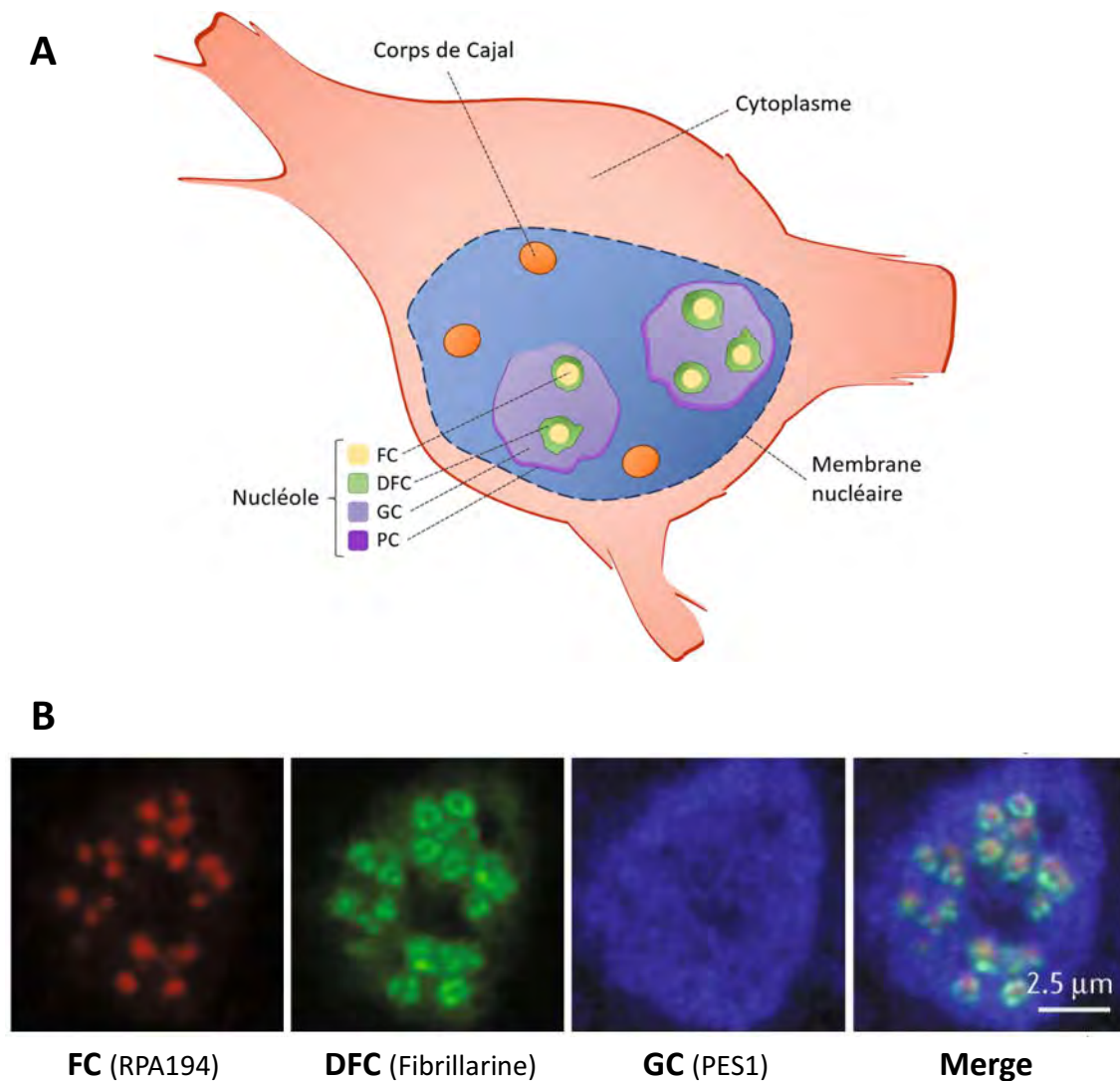


Figure 1 : Organisation des nucléoles

(A) Représentation schématique d'une cellule eucaryote. Le noyau est délimité par la membrane nucléaire, en bleu foncé. Les nucléoles sont composés des trois sous-compartiments FC, DFC et GC, et de la chromatine périnucléolaire (PC). Les corps de Cajal sont également représentés. (B) Immunofluorescence en microscopie confocale d'un nucléole. Le FC (rouge) est détecté via la sous-unité RPA194 de l'ARN polymérase I, le DFC (vert) par la Fibrillarine, et le GC via PES1 (Pescadillo ribosomal biogenesis factor 1 protein). Issue de (Lafontaine *et al.*, 2021).

ADNr sont transcrits à l'interface du FC et du DFC (Koberna *et al*, 2002). Les ARN néosynthétisés s'associent au sein du DFC avec les premières protéines ribosomales et des facteurs d'assemblages. C'est également dans cette région que les ARNr sont la cible de modifications nucléotidiques post-transcriptionnelles, majoritairement guidés par les snoRNP. Ces modifications sont nécessaires pour assurer une fonction correcte des futurs ribosomes. Les ARNr finissent leur maturation et subissent les dernières étapes de clivages dans le GC, région où l'assemblage des ribosomes continue (Lafontaine *et al*, 2021).

I. 1. 2. Les corps de Cajal

Les corps de Cajal sont impliqués dans la maturation des ARN de l'épissage et des extrémités 3' des ARNm des histones. Ils hébergent également des étapes nécessaires à l'assemblage de la télomérase. Les corps de Cajal sont des territoires nucléaires de forme sphérique, dépourvue de membrane (Morris, 2008). Ils ont été découverts en 1903 dans le noyau des cellules neuronales. Ces territoires, de par la forte présence de filaments de coiline, étaient premièrement nommés « coiled body ». Elles ont finalement pris le nom, une centaine d'année plus tard, du scientifique les ayant découverts (Gall *et al*, 1999).

Les corps de Cajal ont une structure très dynamique, et ne sont pas visible dans toutes les cellules. Les filaments de coiline sont par ailleurs nécessaires pour une conformation structurelle correcte de ces territoires nucléaires.

Si la coiline est absente, des résidus de corps de Cajal sont toujours visibles, mais ces territoires résiduels ne regroupent plus les différents processus de maturation. En ce sens, les snRNP, les complexes contenant les ARNm des histones et les snoRNP sont chacun concentré dans des structures résiduelles différentes (Jády *et al*, 2003).

Suite à leur synthèse par l'ARN polymérase II et un bref passage dans le cytoplasme, les snRNP de l'épissage achèvent leur maturation dans les corps de Cajal. Les snARN sont les cibles de pseudouridylation et de méthylations en 2'-O-ribose guidées par une classe de snoRNP spécialisée nommés scaRNP (Small CAjal bodies specific RNP). Les scaRNP sont des complexes ARN/protéines très similaires aux snoRNP, leur appellation faisant référence au territoire dans lequel ils guident ces modifications post-transcriptionnelle (Kiss, 2004). L'ARN de l'épissage U6 est quant à lui synthétisé par l'ARN polymérase III, et sa maturation a lieu exclusivement dans le noyau. Contrairement aux autres snARN de l'épissage, l'ajout des pseudouridylation et des 2'-O- méthylations a lieu dans les nucléoles (Didychuk, Butcher and Brow, 2018).

I. 2. Les ARN de la classe H/ACA

La classe des snoARN H/ACA, en s'associant avec les protéines NHP2, GAR1, NOP10 et la Dyskérine, forment un complexe ribonucléoprotéique catalysant les pseudouridylation sur les ARN cellulaires.

I. 2. 1. Les pseudouridylation

La réaction de pseudouridylation consiste à l'isomérisation d'une uridine, aboutissant à une pseudouridine (ψ). Les pseudouridines sont très largement répandues chez les ARN non codants, particulièrement sur les ARNr, snARN, ARNt, à tel point qu'elles ont premièrement été considérées comme un cinquième ribonucléotide (Davis & Allen, 1957). Les pseudouridines sont toutefois retrouvées sur d'autres ARN cellulaires, tel que les ARNm (Carlile *et al*, 2014).

Les pseudouridines permettent d'augmenter la stabilité des ARN sur lesquels elles sont présentes, de faciliter leurs repliements, et d'influer sur leurs interactions avec des protéines (Dimitrova *et al*, 2019). En outre, la présence de pseudouridines sur les ARNm pourraient influencer les mécanismes d'épissage, de traduction, ou de localisation qui leur sont associés (Boo & Kim, 2020).

Chez les procaryotes, les réactions de pseudouridylation sont réalisées par des pseudouridine synthases (PUS, pour « stand-alone pseudouridine synthases »), qui agissent de manière totalement autonome (Wrzesinski *et al*, 1995). Ce phénomène, conservé chez les eucaryotes, implique une dizaine d'enzymes notées de Pus1 à Pus10 (13 chez l'humain), qui reconnaissent la séquence et la structure secondaire de l'ARN portant l'uridine à modifier (Hamma & Ferré-D'Amaré, 2006). A cela, chez les archées et les eucaryotes, un second mécanisme s'ajoute : la sélection de l'uridine à modifier est médiée par un ARN guide, complexé avec une pseudouridine synthase (DKC1) catalysant la réaction.

L'isomérisation d'une uridine est la résultante de différentes étapes. Premièrement, la liaison azote-carbone (N1 – C1) qui relie l'uridine au ribose est cassée (**Figure 2A**). L'uridine, libre, subit une rotation de 180° selon l'axe de la liaison de l'azote 3 – carbone 6. Une nouvelle liaison est alors établie entre le carbone 5 et le carbone 1, reliant de nouveau l'uridine au ribose (Yu & Meier, 2014). Cette liaison carbone - carbone est plus stable que la liaison azote - carbone des uridines, et il y a désormais un groupe donneur d'hydrogène supplémentaire au niveau de l'azote 1, à la base de la pseudouridine. Ces différences permettent un meilleur

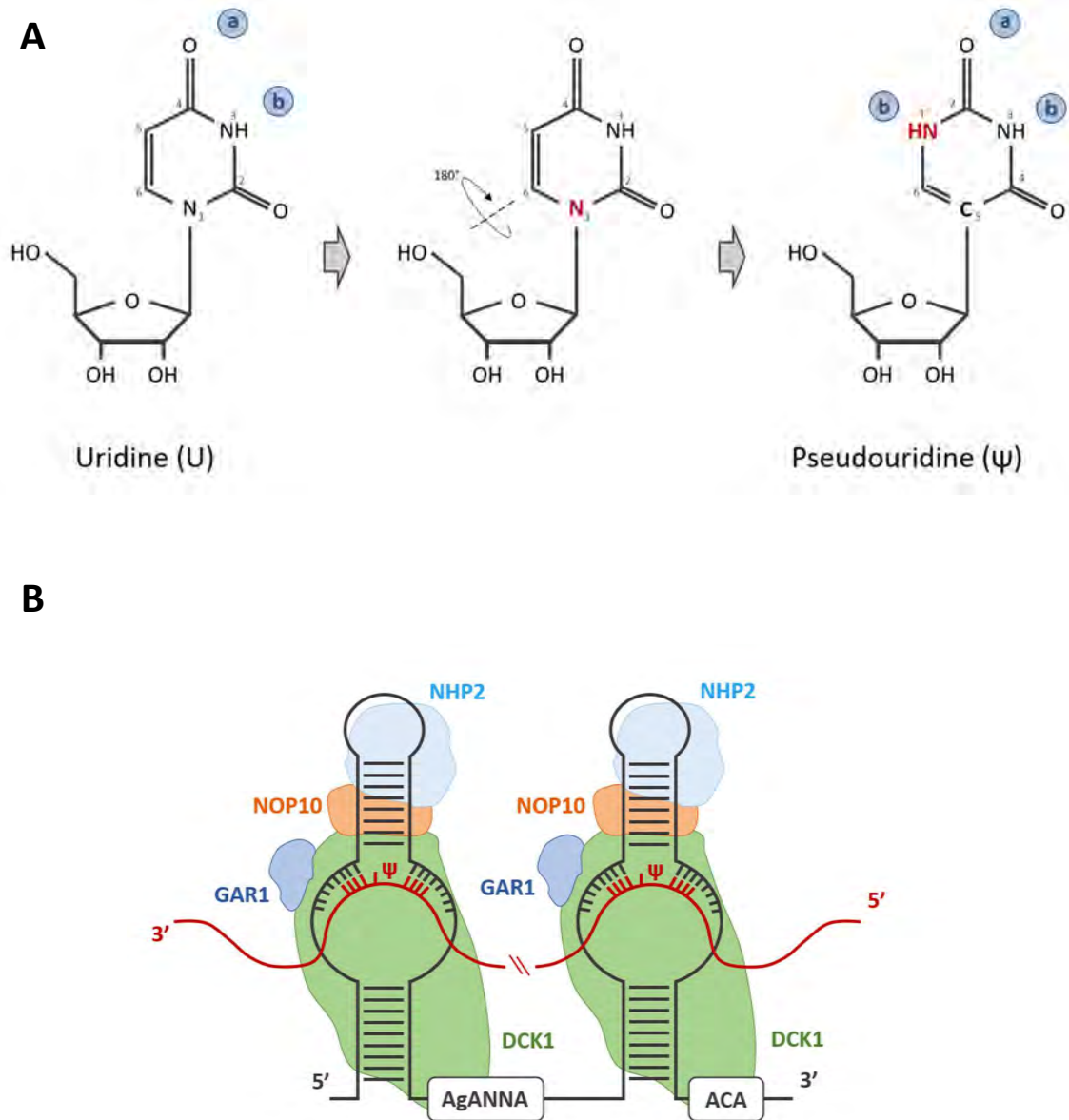


Figure 2 : Structure et fonction des snoRNP H/ACA

(A) Représentation schématique de l’isomérisation d’une Uridine en Pseudouridine. La liaison N1 – C¹, atomes faisant parties respectivement de l’uridine et du ribose, est cassée. L’uridine subit une rotation de 180° selon l’axe N3 – C6. Une nouvelle liaison est établie entre le C5 de l’Uridine et le C¹ du ribose. N1 de la pseudouridine possède un groupe donneur d’hydrogène supplémentaire, qui est représenté par (b). (a) représente un groupe accepteur d’hydrogène. Adaptée de (Ge & Yu, 2013). (B) Représentation schématique de l’organisation structurale d’une snoRNP H/ACA, et de son interaction avec son ARN cible. La structure du snoARN H/ACA suit le modèle en “hairpin-hinge-hairpin-tail”. Chaque tige-boucle est associée aux quatre protéines cœurs. La pseudouridine issue de l’isomérisation de l’uridine par la Dyskérine (DCK1) est représentée par le symbole (Ψ).

empilement des bases, rigidifient les liaisons phosphodiester, et rendent les liaisons $\psi - A$ plus stables au niveau thermodynamique que les liaisons $U - A$ (Charette & Gray, 2000).

I. 2. 2. Les caractéristiques des ARN H/ACA

Les ARN H/ACA, composants essentiels d'un des deux mécanismes de pseudouridylation chez les archées et les eucaryotes, sont d'une longueur généralement comprise entre 60 et 150 nt, et caractérisés par une structure secondaire en "hairpin-hinge-hairpin-tail" (**Figure 2B**). Les deux éléments en tiges-boucles bordent une séquence nucléotidique particulière, extrêmement conservée au cours de l'évolution : cette succession de nucléotide est nommée « boîte H », et a pour séquence consensus AgAnnA. Les ARN H/ACA contiennent une seconde séquence caractéristique et conservée, composée de seulement trois nucléotides, qui lui donneront par ailleurs son nom : boîte ACA. Cette boîte ACA est située après la seconde tige-boucle, à exactement trois nucléotides en amont de l'extrémité 3' de la molécule d'ARN (Kiss, 2002). De manière moins fréquente, il arrive que cette boîte ACA soit de séquence AUA (Balakin *et al*, 1996).

Les scaARN H/ACA possèdent une région nucléotidique conservée supplémentaire sur la partie terminale de l'une ou de chaque tige-boucle, de séquence consensus ugAG, appelée boîte CAB. Cette boîte CAB est reconnue par TCAB1 (aussi connue sous le nom de WDR79), une protéine cœur des scaARN H/ACA. En interagissant avec des protéines des corps de Cajal, TCAB1 permet la rétention des scaARN H/ACA au sein de ces territoires nucléaires (Tycowski *et al*, 2009).

Il existe cependant quelques alternatives à cette structure secondaire canonique. Pour commencer, la longueur des structures en tiges-boucles est variable selon les ARN H/ACA (Meier, 2005). De plus, certains organismes, comme les trypanosomes, présentent des ARN H/ACA ne possédant qu'une seule tige-boucle (Uliel *et al*, 2004). Fait qui a été également observé chez le protiste *Euglena gracilis* (Russell *et al*, 2004) et couramment retrouvé chez les archées. Accessoirement, l'archée *Pyrobaculum* possède également des ARN H/ACA possédant une structure secondaire avec la présence de trois tiges-boucles (Bernick *et al*, 2012).

L'interaction physique entre l'ARN H/ACA et l'ARN substrat se fait par appariement direct des paires de bases. La sélection du nucléotide cible est possible grâce à une boucle interne située entre la base et la tête de chacune des structures en tiges-boucles. Au vu des fonctions de cette boucle interne, elle est nommée « poche de pseudouridylation ».

Chacune des structures en épingles des ARN H/ACA sont associées à quatre protéines cœurs, GAR1, NHP2, NOP10 et la Dyskérine (Watkins *et al*, 1998), créant la particule ribonucléoprotéique (RNP) H/ACA. Cette ribonucléoparticule est alors composée d'un ARN H/ACA et de deux lots des quatre protéines cœurs. Ces protéines conservées possèdent de nombreux orthologues chez les eucaryotes et les archées (**Figure 3A**). La protéine humaine GAR1, de 22kDa, est composée de deux domaines riches en glycine-arginine (GAR domain) (Girard *et al*, 1992). NHP2 est une protéine de 17,2 kDa. NOP10 est la plus petite protéine du complexe, avec 7,7 kDa. La Dyskérine (DKC1) est une pseudouridine synthase de 57,6 kDa, qui permet à elle seule l'activité catalytique de la RNP H/ACA. Les protéines NHP2, NOP10 et GAR1 vont participer à la stabilisation, à la structuration de la RNP H/ACA, ainsi qu'aux interactions avec d'autres molécules.

Différentes expériences de déplétion ont déterminé que les homologues de ces protéines cœurs chez la levure étaient nécessaires à la viabilité de la cellule (Henras *et al*, 1998; Dez *et al*, 2001a). Cependant, GAR1 serait dispensable pour la stabilité de la snoRNP, contrairement aux trois autres. L'assemblage d'une RNP H/ACA avec ses protéines cœurs est très stable, signifiant que la formation d'une nouvelle RNP nécessitera une nouvelle synthèse d'un ARN H/ACA et un nouvel assemblage de la RNP (Wang & Meier, 2004).

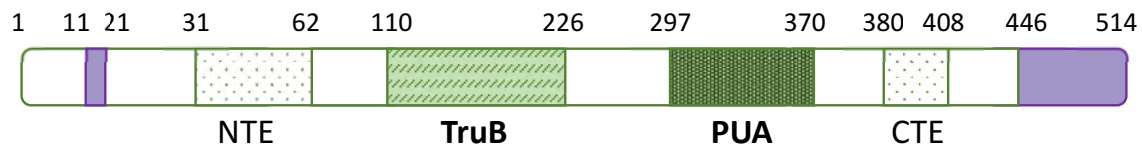
I. 2. 3. La Dyskérine

La Dyskérine est une protéine humaine codée par le gène *DKC1*, faisant partie des pseudouridine synthases. C'est une famille d'enzymes catalysant la réaction d'isomérisation d'une uridine (Koonin, 1996). La Dyskérine, fortement conservée évolutivement, possède de nombreux orthologues au sein d'une multitude d'espèces.

aCbf5 et Nap57, des orthologues de la Dyskérine, chez la levure et le rat, respectivement, ont été découvertes au début des années 90 lors d'études sur les protéines nucléaires et leurs fonctions associées (Meier & Blobel, 1990, 1994). Il faudra attendre 1998 pour que la protéine humaine, la Dyskérine, soit identifiée durant la caractérisation de mutations chez des patients atteints de la dyskératose congénitale lié au chromosome X, une maladie du vieillissement prématuré (Heiss *et al*, 1998). Les premières informations sur la structure de la Dyskérine et de ses orthologues ont pu être résolues grâce à la cristallisation d'une enzyme orthologue complexée avec une partie d'un ARN de transfert chez la bactérie (Hoang & Ferré-D'Amaré, 2001).

A

Protéines	Humain	Archée	Levure	Poids Moléculaire (Humain)
	Dyskérine (DKC1)	aCbf5	pCbf5	57,6 kDa
	GAR1	aGar1	pGar1	22 kDa
	NHP2	aL7Ae	pL7Ae	17,2 kDa
	NOP10	aNop10	pNop10	7,7 kDa

B**Figure 3 : Protéines cœurs de la snoRNP H/ACA**

(A) Nomenclature des orthologues des protéines cœurs de la snoRNP H/ACA. (B) Représentation schématique des principaux domaines de la Dyskérine. Les délimitations des domaines en nombre d'acide aminés sont indiquées par rapport à la protéine humaine. TruB (pseudouridine synthase domain), PUA (pseudouridine synthase and archaeosine transglycosylase domain), NTE (amino terminal-extension), CTE (carboxy-terminal extensions). Les régions représentées en violet sont des domaines NLS/NoLS (nuclear and nucleolar localization signals). Adaptée de (Garus and Autexier, 2021).

La Dyskérine possède au niveau de sa partie carboxy-terminale (C-terminale) un domaine catalytique de 116 acides aminés nommé TruB (**Figure 3B**). Ce domaine, ainsi que la région fixatrice à l'ARN substrat qu'il contient, sont caractéristiques d'une des cinq sous-familles des pseudouridines synthase. Cette sous-famille est nommée TruB, et la Dyskérine en fait donc partie. Ce domaine comporte un acide aspartique, qui semble crucial à son activité enzymatique (Huang *et al*, 1998). La région fixatrice à l'ARN s'appelle le motif « thumb », et interagit avec l'ARN substrat comprenant l'uridine à modifier (Garus & Autexier, 2021).

La Dyskérine possède également un domaine PUA (pseudouridine synthase and archaeosine transglycosylase) de 73 acides aminés, très riche en résidus hydrophobes. Ce domaine PUA permet la fixation de la Dyskérine à la molécule d'ARN H/ACA. Contrairement à la grande majorité des pseudouridines synthases, la fixation d'un ARN guide à la Dyskérine est nécessaire à la fonction catalytique de la Dyskérine, ainsi que pour permettre la reconnaissance de l'ARN cible (Garus & Autexier, 2021).

I. 2. 4. Reconnaissance de l'ARN substrat par les RNP H/ACA

L'ARN substrat, contenant l'uridine à modifier, est sélectionné via un appariement direct des paires de bases par la RNP H/ACA. L'ARN H/ACA est structuré de manière à laisser apparaître deux séquences de 9 à 13 nucléotides, situés de part et d'autre des brins opposés à la boucle interne de la poche de pseudouridylation. Une partie de ces nucléotides, appelés éléments antisens, reconnaissent les séquences bordant l'uridine à modifier de l'ARN cible. L'interaction entre l'élément antisens et l'ARN substrat engendre une réorganisation de la conformation de la poche de pseudouridylation, permettant de placer l'uridine à modifier à proximité de la fente catalytique de la Dyskérine. Entre les deux côtés de la boucle interne, seuls deux nucléotides de l'ARN substrat ne seront pas appariés, dont l'uridine à modifier (5' – UN – 3'). La distance entre l'uridine modifiée et les boîtes H ou ACA de l'ARN guide est invariablement de 13 à 16 nucléotides (Kiss *et al*, 2010). L'interaction par paire de base entre l'ARN guide et l'ARN substrat est un événement spontané dans lequel les protéines cœurs ne sont pas impliquées (Jin *et al*, 2007; Wu & Feigon, 2007).

La structure de ribonucléoparticule H/ACA est assemblée de manière à avoir d'un côté l'ARN guide, et les protéines cœurs de l'autre, rendant la poche catalytique facilement accessible de l'extérieur du complexe. Les nucléotides qui composent la poche de pseudouridylation ne sont en contact avec aCbf5, homologue de la Dyskérine chez l'archée,

seulement via leur squelette sucre/phosphate. Cela permet de laisser une grande flexibilité aux bases azotées, leur permettant de s'apparier avec un potentiel ARN substrat (Liang *et al*, 2009).

Le motif « thumb » de la Dyskérine, très riche en arginine, interagit avec l'ARN cible lorsque ce dernier lui est présenté par l'ARN guide, et verrouille l'uridine à modifier (Li & Ye, 2006; Duan *et al*, 2009). Lors de l'absence du substrat, le motif « thumb » adopte une conformation ouverte en se liant à la protéine cœur GAR1. L'interaction avec GAR1 permettrait de stabiliser cette conformation ouverte, permettant ainsi le relargage du substrat. En ce sens, il a été démontré chez l'archée, que des mutations ponctuelles du motif « thumb », ou de la protéine GAR1, n'affectent pas l'accumulation de la snoRNP H/ACA. En revanche, ces mutations rendaient impossible la pseudouridylation de plusieurs uridines successives par une même ribonucléoparticule (Duan *et al*, 2009). La réorganisation de la poche de pseudouridylation et des protéines cœurs lors de l'isomérisation d'une uridine laisse penser que cette réaction ne peut cibler qu'une seule uridine à la fois. Cependant, chez d'autres organismes que l'homme, des snoARN semblent être impliqués dans la synthèse de pseudouridines adjacentes, par des mécanismes encore mal compris. Ces modifications seraient réalisées de manière séquentielle, et la seconde serait possible dans le cas où l'ARN substrat resterait lié à la snoRNP. Ces phénomènes pourraient être dus à des variants structurels ou grâce à un réarrangement conformationnel de la snoRNP (Schattner *et al*, 2004a; Deryusheva & Gall, 2013; Majumder *et al*, 2020).

I. 2. 5. Conformation structurelle des RNP H/ACA

Les premières données sur la conformation de la ribonucléoparticule H/ACA ont été obtenues grâce à un assemblage *in vitro* de protéines cœurs archéales recombinantes avec un snoARN H/ACA synthétique. Ces expériences ont relevé que les protéines aCbf5 et aL7Ae se lient à l'ARN H/ACA indépendamment des protéines aNop10 et aGar1 (Charpentier *et al*, 2005). Les deux adénines de la boîte ACA de l'ARN guide se fixent au domaine PUA de aCbf5. Le domaine catalytique de aCbf5 se décompose en deux sous parties, D1 et D2, de part et d'autre de la fissure du site actif. Le domaine D1 est en contact avec la poche de pseudouridylation ainsi qu'avec la tige la surplombant (Henras *et al*, 2017). De son côté, aL7Ae se fixe à l'ARN H/ACA grâce au motif « kink-turn », situé au-dessus de la poche de pseudouridylation.

Ce motif particulier, également nommé « k-turn » ou « GA motif », est caractérisé par un renflement asymétrique de trois nucléotides, induisant la pliure de l'axe de l'hélice d'ARN

d'environ 120° (Klein *et al*, 2001; Qu *et al*, 2011). Des paires de bases non canoniques A-G sont présentes en aval de ce renflement (**Figure 4**).

Le motif k-turn n'étant pas conservé chez les snoARN H/ACA des eucaryotes, le recrutement de NHP2, l'homologue de aL7Ae, se fait par interaction protéine/protéine. *In vivo*, NHP2 se fixe sur la protéine NOP10, contrairement à aL7Ae, qui, en absence de l'ARN H/ACA, ne peut se lier aux autres protéines de la particule (Wang & Meier, 2004). Les protéines aCbf5, aNop10 et aGar1 interagissent entre elles dans la snoRNP H/ACA, et, en revanche, peuvent s'associer en l'absence de l'ARN guide (Rashid *et al*, 2006). La cristallisation d'une snoRNP H/ACA chez l'archée a révélé que la liaison entre aNop10 et le domaine PUS de aCbf5 contribuerait à l'organisation structurale ainsi qu'à la stabilité du site catalytique de aCbf5. La protéine aGar1 interagit exclusivement avec la sous-partie D2 du domaine catalytique de aCbf5 (Rashid *et al*, 2006). C'est par ailleurs la seule protéine cœur du complexe H/ACA à ne pas être en contact direct avec l'ARN guide. Ces caractéristiques structurales ont été confirmées par la cristallisation d'une snoRNP H/ACA chez la levure *Saccharomyces cerevisiae* (Li *et al.*, 2011).

Au regard de la forte conservation des snoRNP H/ACA entre les différentes espèces, il est considéré que la structure de ces complexes ribonucléoprotéiques soit sensiblement équivalente chez l'être humain. Allant en ce sens, de récentes études en Cryo-EM de la télomérase, contenant une RNP H/ACA, ont permis d'avoir la confirmation de ces suppositions tout en affinant certaines interactions (Nguyen *et al*, 2018; Ghanim *et al*, 2021). La Dyskérine se lie à la base de la tige-boucle ainsi qu'au niveau de la boîte H ou ACA, de manière à placer son site catalytique au niveau de la poche de pseudouridylation (**Figure 5**). La Dyskérine, NOP10 et NHP2 sont assemblées de manière à créer une plateforme commune permettant l'arrimage de la partie de l'ARN guide, située au-dessus de la poche de pseudouridylation, tandis que GAR1 n'interagit effectivement pas avec l'ARN guide. En supplément, chez la télomérase, NHP2 se lie aussi spécifiquement avec une uridine de la boucle terminale du snoARN H/ACA, appelée motif BIO. Cette liaison est probablement une capacité eucaryote restante de la liaison de NHP2 au motif en k-turn existant chez les archées (Ketele *et al*, 2016). Cependant, le motif BIO semble facultatif car il n'est pas présent chez tous les snoARN H/ACA. Il est probable que ce motif puisse conférer une stabilité supplémentaire à la protéine NHP2 lorsque la seconde tige-boucle de l'ARN H/ACA n'est pas canonique, ce qui est le cas pour la télomérase. En effet, une tige-boucle trop courte ou trop longue ne pourra pas s'arrimer

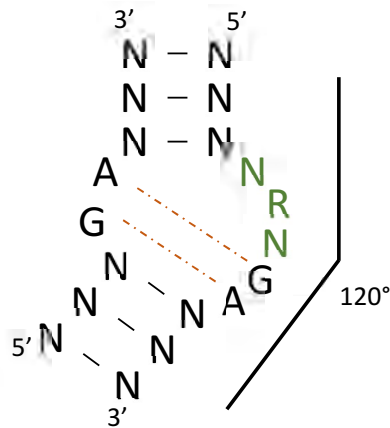


Figure 4 : Représentation schématique d'un motif en Kink-turn

Les trois nucléotides composant le renflement sont indiqués en vert. Les liens entre les paires de bases non-canoniques A – G sont illustrés en orange.

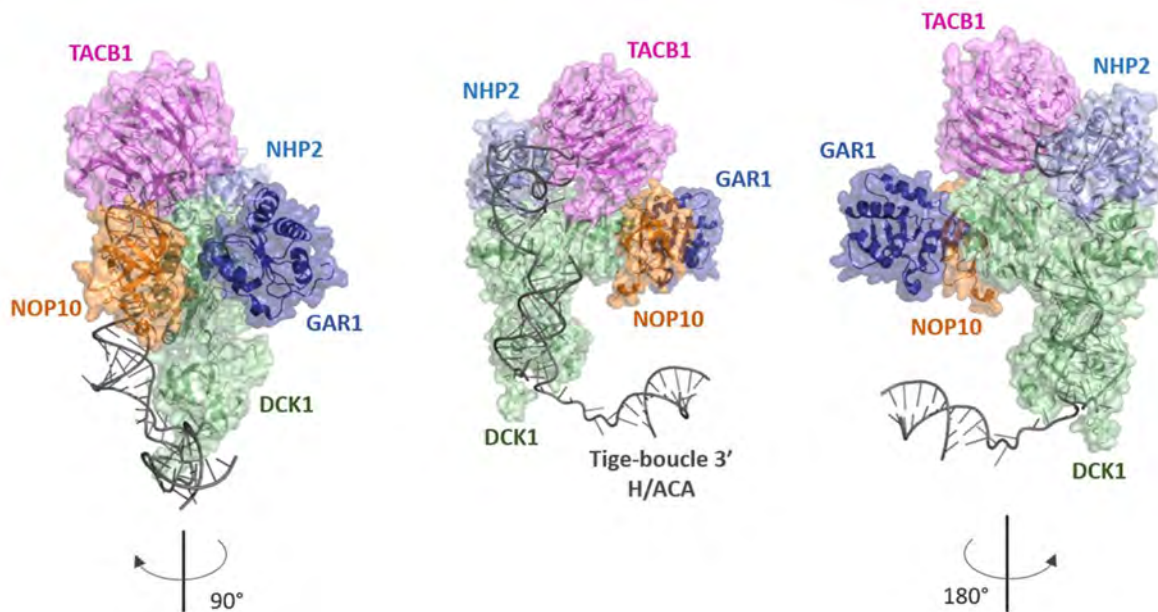


Figure 5 : Assemblage structural entre la tige-boucle 3' du domaine H/ACA de la Télomérase et les protéines cœurs

Vues avant (centre), de trois-quarts (gauche) et de dos (droite) de la tige-boucle 3' de la Télomérase humaine assemblée avec les protéines cœurs H/ACA (PDB : 7BGB)

à la plateforme commune de liaison à l'ARN, DKC1-NOP10-NHP2, engendrant ainsi une instabilité de la snoRNP (Ketele *et al*, 2016).

I. 2. 6. Biogenèse et transport intranucléaire des RNP H/ACA

La majorité des snoARN H/ACA humains sont issus de la maturation de régions introniques des ARN messagers, transcrites par l'ARN polymérase II. Les précurseurs des snoARN sont générés par l'étape d'épissage des pré-ARNm (Ooi *et al*, 1998). Les introns hôtes des snoARN sont pris en charge par différents complexes protéiques, permettant la reconnaissance, la maturation et l'adressage des snoARN H/ACA. A ce jour, chez l'humain, seul le snoARN H/ACA composant la télomérase est connu pour être transcrit sous son propre promoteur (Mitchell *et al*, 1999a).

SHQ1 et NAF1, deux protéines chaperonnes, sont cruciales dans la biogenèse de la RNP (**Figure 6**). La Dyskérine n'étant pas stable seule, il a été proposé, que, dès sa synthèse, elle se lierait avec SHQ1 dans le cytoplasme afin d'éviter son agrégation ou sa dégradation (Grozdanov *et al*, 2009). Cette étape de liaison entre SHQ1 et une Dyskérine dépourvue d'ARN H/ACA et des autres protéines cœurs est une étape antérieure à la transcription du snoARN par la polymérase II. Par ailleurs, SHQ1 se lie, via son domaine SSD (Shq1-Specific Domain), à la partie de la Dyskérine prévue pour interagir avec les ARN. L'interaction SHQ1 – Dyskérine permet aussi de prévenir l'association non spécifique de la Dyskérine avec d'autres ARN cellulaires (Walbott *et al*, 2011). En revanche, cela rend la dissociation de SHQ1 nécessaire afin de permettre l'association de la Dyskérine à un ARN guide. En effet, lorsque la Dyskérine entre en contact avec l'ARN H/ACA, la protéine SHQ1 ne fait plus partie du complexe (Massenet *et al*, 2016a). Les protéines cœurs, NHP2 et NOP10 sont recrutées pour former le pré-complexe DKC1-SHQ1-NHP2-NOP10. NAF1 s'ajoute à ce complexe en interagissant avec la Dyskérine. Tout comme SHQ1, NAF1, par son interaction avec la Dyskérine, permet de la stabiliser (Darzacq *et al*, 2006). Parallèlement, NAF1 se lie au domaine carboxy-terminal (CTD) de la polymérase II, et interagit avec certains facteurs de transcription associés à la polymérase II, Spt16, Tfg1 et Sub1 (Yang *et al*, 2005). Via ces interactions, NAF1 participe à amener les pré-complexes DKC1-NHP2-NOP10 au voisinage des gènes ayant une activité transcriptionnelle très forte (Ballarino *et al*, 2005; Darzacq *et al*, 2006).

La reconnaissance des précurseurs des ARN guides introniques ainsi que le recrutement des protéines cœurs ont alors lieu directement sur le pré-ARN hôte nouvellement synthétisé (Darzacq *et al*, 2006; Richard *et al*, 2006). Le snoARN H/ACA naissant, interagit directement

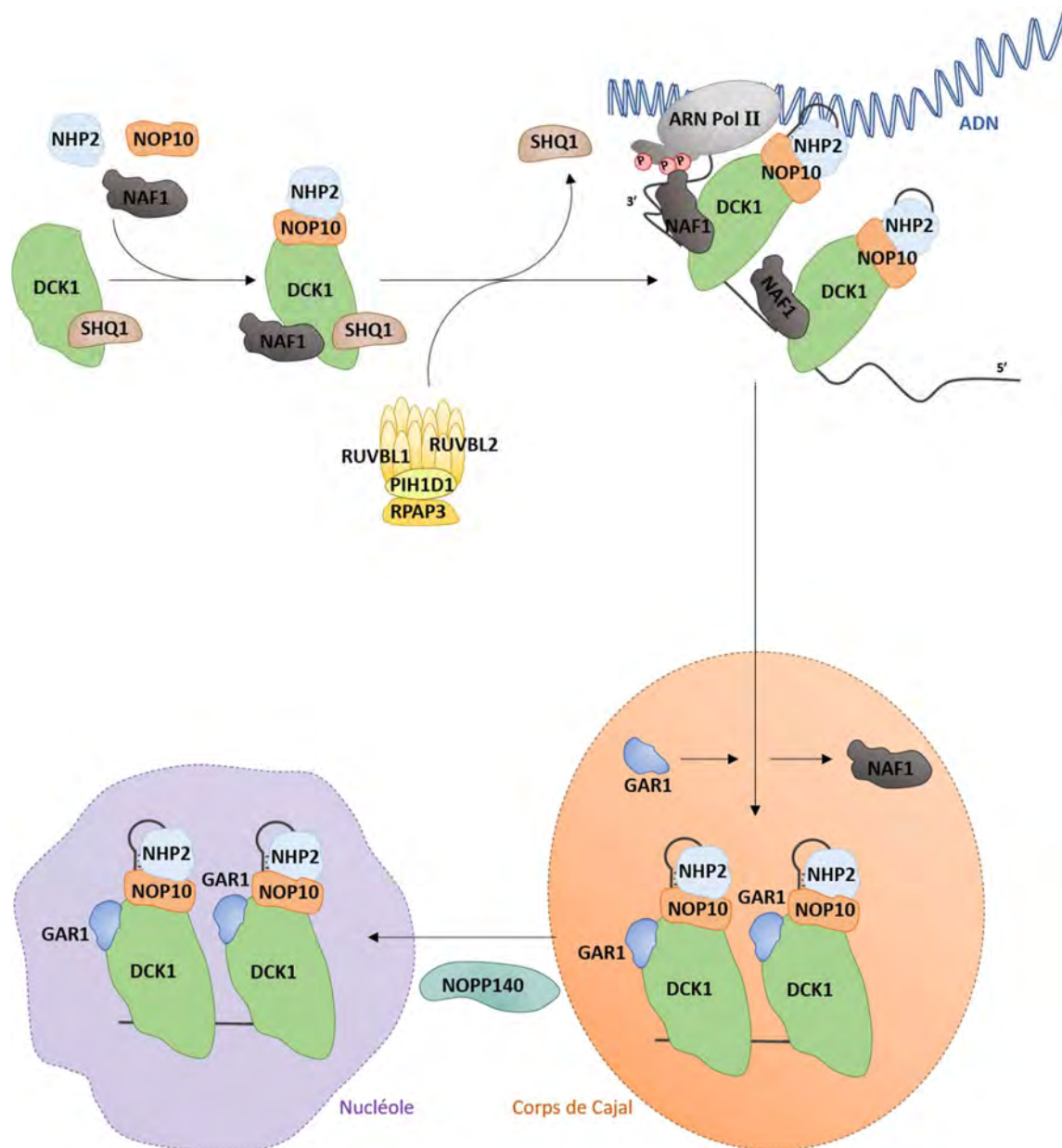


Figure 6 : Biogenèse et transport intranucléaire des snoRNP H/ACA

SHQ1 se lie à la Dyskérine afin d'éviter des interactions prématurées. Les protéines cœurs NOP10 et NHP2, ainsi que NAF1 s'ajoutent à la Dyskérine et SHQ1, formant le pré-complexe DCK1-SHQ1-NHP2-NOP10-NAF1. SHQ1 est dissocié du complexe par l'action du complexe R2TP, composé notamment des ATPases RUVBL1 et RUVBL2. L'assemblage de ce pré-complexe au snoARN H/ACA se déroule dans le même temps que la synthèse transcriptionnelle de ce snoARN. Après débranchement de l'intron excisé, les extrémités 5' et 3' sont dégradées respectivement par l'endoribonucléase XRN2 et l'exosome. GAR1 prend la place de NAF1, donnant lieu à la snoRNP fonctionnelle. Les scaRNP sont retenus dans les corps de Cajal, et les snoRNP sont acheminées en direction des nucléoles, via NOPP140. Adaptée de (Massenet *et al*, 2016b).

avec le complexe DKC1-NHP2-NOP10-NAF1, formant ainsi la pré-RNP H/ACA inactive, dans laquelle seule la protéine GAR1 est manquante. Cet assemblage entre l'ARN H/ACA et ses protéines cœurs est crucial afin d'éviter la dégradation exonucléolytique du snoARN (Kufel *et al*, 2000).

La dissociation entre SHQ1 et la Dyskérine nécessite l'action de RUVBL1 et RUVBL2 (Machado-Pinilla *et al*, 2012). Ces deux ATPases sont apportées par le complexe HSP90/R2TP, largement impliqué dans la biogenèse des RNP H/ACA et C/D, même si sa fonction moléculaire n'est pas entièrement élucidé (Boulon *et al*, 2008a; Zhao *et al*, 2008). HSP90 (Heat Shock Protein 90) est une famille de protéines chaperonnes ubiquitaires qui favorisent en général le repliement moléculaire de facteurs de signalisation en condition de chocs thermiques (Schopf *et al*, 2017). R2TP, co-chaperonne de HSP90, est composée d'un hétéro-hexamère d'ATPases RUVBL1/RUVBL2, et d'un hétérodimère de RPAP3 et PIH1D1, qui agiraient comme régulateurs et adaptateurs des ATPases (Maurizy *et al*, 2018). Les conditions du recrutement du complexe R2TP sur le lieu de la biogenèse des RNP H/ACA n'ont pas encore été totalement démontrées. Le recrutement via les interactions NUFIP (protéine adaptateur)-NHP2, Dyskérine - RUVBL1/2 et Dyskérine - PIH1D1 sont les hypothèses les plus probables (Boulon *et al*, 2008a; Machado-Pinilla *et al*, 2012). Les protéines du complexe R2TP possèdent un domaine CS, leur permettant de s'associer à HSP90. De manière intéressante, SHQ1 possède également un domaine CS, mais ne s'associe pas avec HSP90 (Godin *et al*, 2009). Les ATPases sont impliquées dans la stabilisation des facteurs d'assemblages ainsi que dans leur dissociation avec les snoRNP H/ACA, favorisant ainsi la maturation de ces derniers (Ewens *et al*, 2016).

La dernière étape de maturation de la snoRNP H/ACA, est le remplacement de NAF1 par GAR1, rendant la particule active et fonctionnelle. GAR1, absent dans les régions où les gènes hôtes des ARN H/ACA sont transcrites, est présent en grande quantité au niveau des nucléoles et des corps de Cajal (Darzacq *et al*, 2006). Cela soutient l'hypothèse que l'échange entre NAF1 et GAR1 se ferait au sein de ces territoires nucléaires (Li *et al*, 2011). Le domaine principal de GAR1 possède de fortes homologues avec le domaine de NAF1 interagissant avec la Dyskérine (Leulliot *et al*, 2007), créant ainsi un phénomène de compétitivité entre les deux protéines. Cela pourrait expliquer que ces deux protéines ne peuvent interagir de manière mutuelle avec la Dyskérine (Darzacq *et al*, 2006). Des études ont également mis en avant le rôle potentiel du complexe de Survie des MotoNeurones (SMN) dans le remplacement de NAF1 par GAR1. En effet, le complexe SMN est présent dans les corps de Cajal, et certaines des protéines le composant interagissent directement avec GAR1 (Whitehead *et al*, 2002).

Lorsque le snoARN H/ACA intronique est dans le complexe composé des protéines DKC1-NHP2-NOP10-NAF1, ses extrémités 5' et 3' sont dégradées pour aboutir à la séquence 5'-3' mature de l'ARN. L'extrémité en 5' est dégradée par l'exoribonucléase XRN2, tandis que la dégradation en 3' requiert le complexe NEXT ainsi que l'exosome nucléaire (Lubas *et al*, 2015).

Des études chez l'humain proposent que la polymérase poly(A) PAPD5 pourrait être impliquée dans la maturation 3' de l'ARN H/ACA au sein des corps de Cajal. L'ajout de queues oligoA permettrait d'éviter la dégradation en 3' des pré-ARN H/ACA. La ribonucléase PARN éliminerait par la suite ces queues oligoA ainsi que les nucléotides introniques restant, donnant l'extrémité 3' mature de l'ARN H/ACA (Berndt *et al*, 2012).

I. 3. Les ARN de la classe C/D

Les snoARN C/D, en s'associant avec les protéines NHPX, NOP56, NOP58 et la Fibrillarine, forment un complexe ribonucléoprotéique catalysant des méthylations en 2'-O-ribose sur les ARN cellulaires.

I. 3. 1. Les méthylations en 2'-O-ribose

Les méthylations en 2'-O-ribose font partie, avec les pseudouridylations, des modifications post-transcriptionnelles les plus répandues sur les ARN cellulaires. En outre d'être présentes en grande quantité sur les ARNr et les snARN, les méthylations en 2'-O-ribose sont retrouvées sur les ARNt, les ARNm ainsi que sur les petits ARN régulateurs, tel que les microARN ou ARNsi (Tycowski *et al*, 1998; Darzacq *et al*, 2002; Dai *et al*, 2017).

En revanche, cette modification chimique est peu présente chez les procaryotes : il en est dénombré généralement moins de 5 sur les ARNr, selon les espèces observées. A titre d'exemple, toujours sur les ARNr, 55 méthylations en 2'-O-ribose sont présentes chez la levure, et plus de 110 sont référencées chez l'être humain (Taoka *et al*, 2018; Krogh & Nielsen, 2019). A l'instar des pseudouridines, les méthylations en 2'-O-ribose permettent d'augmenter la stabilité de l'ARN qui les porte, ainsi que de faciliter son repliement. Ces modifications protègent également l'ARN contre les nucléases, et influent sur son interaction avec des protéines (Höfler & Carlomagno, 2020). L'abondance des méthylations en 2'-O-ribose sur les ARNr et les snARN laisse penser qu'elles incarnent un rôle majeur dans les fonctions des ribosomes et du spliceosome. En ce sens, la levure n'est pas capable de survivre après la déplétion de l'ensemble de ses méthylations en 2'-O-ribose (Tollervey *et al*, 1993). Chez

l'humain, l'absence d'une ou plusieurs de ces modifications chimiques sont généralement associées à des pathologies (pour revue : (Dimitrova *et al*, 2019).

Chez les bactéries, ces modifications sont réalisées par des méthyltransférases autonome, tandis que la grande majorité des 2'-O-ribose méthylations chez les eucaryotes sont associées à l'action des snoRNP C/D (Motorin & Helm, 2011).

Cette modification chimique consiste à transférer un groupement méthyl (CH_3) du donneur SAM sur le ribose d'un nucléoside. Plus précisément, le groupe CH_3 sera placé sur l'oxygène du groupe hydroxyle qui est situé sur le carbone en 2' du ribose (**Figure 7A**). Cette modification est retrouvée sur les différents nucléotides existants, ainsi que sur les pseudouridines (Ayadi *et al*, 2019). D'un point de vue purement moléculaire, l'ajout de ce groupement méthyl au ribose permet de rigidifier le nucléoside, et donc la structure hélicoïdale de l'ARN. Cette modification chimique peut inhiber les interactions de l'ARN avec les protéines, le groupement hydroxyle en 2' n'étant plus disponible (Treger & Westhof, 2001).

I. 3. 2. Caractéristiques des RNP C/D

Chez les eucaryotes, les méthylations en 2'-O-ribose sont guidées par les snoARN C/D. Ces petits ARN nucléolaires, d'une longueur généralement comprise entre 60 et 200 nucléotides, ont une structure secondaire en tige-boucle irrégulière. Ils sont caractérisés par la présence de deux motifs conservés, les boîtes C et D, qui ont comme séquence consensus 5'-RUGAUGA-3' (R étant une purine) et 5'-CUGA-3', respectivement (Baldini *et al*, 2021a). La boîte C est située proche de l'extrémité 5' de la molécule d'ARN, tandis que la boîte D est présente en amont de l'extrémité 3' de l'ARN (**Figure 7B**). Ces deux extrémités, proches spatialement, s'apparient sur une courte séquence, et donne ce que l'on appelle la tige terminale. Cette tige terminale forme le motif C/D, composé des boîtes C et D, qui est caractérisé par une structure particulière en k-turn, nécessaire au recrutement des protéines cœurs de la snoRNP C/D (**Figure 7C**). Les snoARN C/D, possèdent généralement au centre de la molécule, des copies imparfaites des boîtes C et D, nommées C' et D'. En général, ces séquences sont moins bien conservées, voir dégénérées (Kiss-László *et al*, 1998). Les éléments antisens qui permettent la reconnaissance de l'ARN substrat comportant le nucléoside à méthyler sont situés en amont des boîtes D et D'. Ces éléments antisens reconnaissent en général entre une dizaine et vingtaine de nucléotides appartenant à l'ARN substrat. Le nucléotide cible de la modification est invariablement celui faisant face au cinquième nucléotide placé en amont des boîtes D ou D' (Cavaillé *et al*, 1996; Kiss-László *et al*, 1998). L'ARN C/D s'associe avec quatre protéines

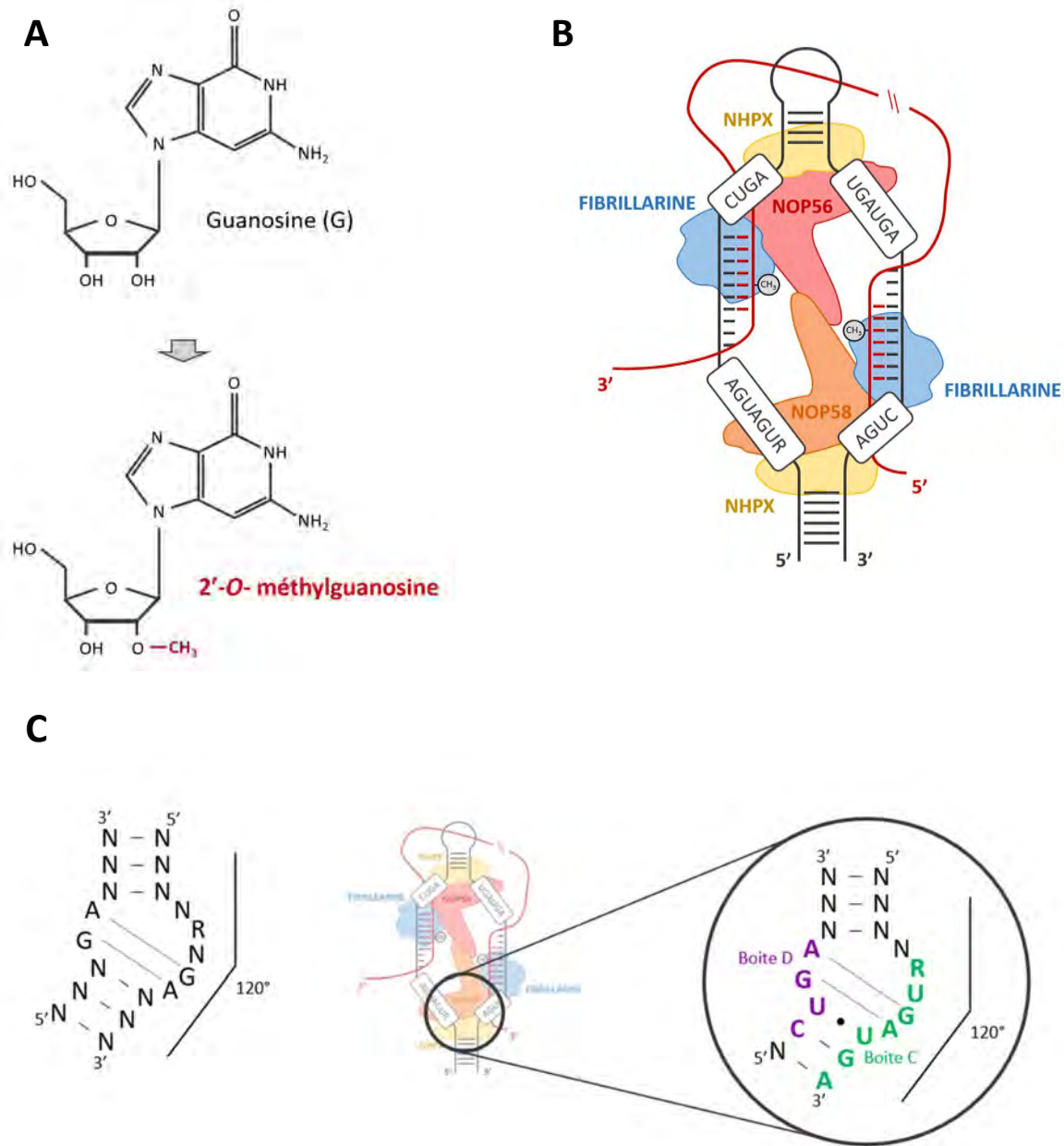


Figure 7 : Structure et fonction des snoRNP C/D

(A) Schéma d'une méthylation en 2'-O-ribose sur une guanosine. Ajout d'un groupement méthyl (CH_3) sur l'oxygène du groupe hydroxyle situé sur le C^2 du ribose. (B) Illustration de l'organisation structurale d'une snoRNP C/D, et de son interaction avec son ARN cible. Les protéines cœurs sont représentées. La méthylation du ribonucléotide cible par la Fibrillarine est représentée par la notation (CH^3). (C) Représentation schématique du motif en Kink-turn formé par les boîtes C (vert) et D (violet).

cœurs, NHPX, NOP56, NOP58 et la Fibrillarine, pour former la snoRNP C/D fonctionnelle. Ces protéines sont conservées entre les différents règnes, leurs homologues respectifs sont consignés dans le tableau de la **Figure 8A**.

NHPX est une protéine de 14,1 kDa, composée d'une région reconnaissant les motifs en kink-turn. Une première copie de la protéine s'associe alors au niveau des boîtes C/D de l'ARN, qui forment ce motif particulier. NHPX descend évolutivement de la protéine archéale composant les snoRNP C/D, L7Ae. L7Ae se lie, à l'instar de la protéine eucaryote, au motif k-turn formé par les boîtes C et D. En revanche, chez les archées, les boîtes C'et D' s'organisent en un motif dit de k-loop, structure sur laquelle se lie une seconde copie de L7Ae. Chez l'humain, ce motif de k-loop est absent, mais une seconde copie de la protéine est présente au niveau des boîtes C'et D'. Il est supposé que cette interaction soit assistée par des interactions protéines-protéines (Watkins *et al*, 2000b; Gagnon *et al*, 2010).

NOP56 et NOP58 sont deux protéines de 66,6 et 59,5 kDa, respectivement, et leurs séquences possèdent de fortes homologues (Gautier *et al*, 1997). Ces deux protéines vont chacune lier une copie de la Fibrillarine au sein de la snoRNP C/D (Aittaleb *et al*, 2003). NOP56 s'associe également aux boîtes C'/D', tandis que NOP58 se fixe aux boîtes C/D.

La Fibrillarine, présente en deux copies au sein de la snoRNP C/D, est une protéine de 33,7 kDa, qui est la méthyltransférase responsable à elle seule la réaction de 2'-O-méthylation sur les nucléosides cibles de la snoRNP C/D.

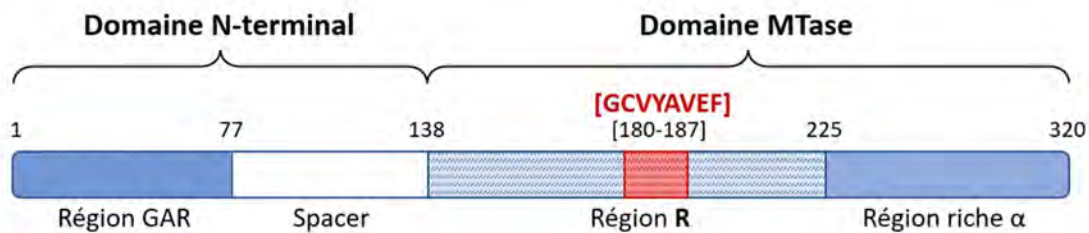
I. 3. 3 La Fibrillarine

La Fibrillarine est une protéine de la classe des méthyltransférases, extrêmement bien conservée au cours de l'évolution. Elle transfère un groupement méthyl du donneur S-adénosylméthionine (SAM) sur le ribose d'un nucléotide. Chez les eucaryotes, la Fibrillarine est une protéine nucléolaire, préférentiellement localisée dans le composant fibrillaire dense (FDC). C'est une protéine essentielle à la viabilité des organismes (de Carcer & Medina, 1999; Newton *et al*, 2003).

La Fibrillarine eucaryote est composée de deux domaines distincts : le domaine N-terminal, et le domaine MTase, responsable de la fonction catalytique de la Fibrillarine (**Figure 8B**). Le domaine MTase est composé de deux régions : une région située au cœur de la protéine, nommée région « R », et la région C-terminale. Ce domaine montre une très forte conservation évolutive entre les différentes espèces. Ce domaine est spécialisé dans les interactions protéines

A

Protéines	Humain	Archée	Levure	Poids Moléculaire (Humain)
	Fibrillarine (FBL)	aFibrillarine	pNop1	33,7 kDa
	NHPX	aL7Ae	pSnu13	14,1 kDa
	NOP56	aNop5	pNop56	66,6 kDa
	NOP58	aNop5	pNop58	59,5 kDa

B**Figure 8 : Protéines cœurs de la snoRNP C/D**

(A) Nomenclature des orthologues des protéines cœurs de la snoRNP C/D. (B) Représentation schématique des principaux domaines de la Fibrillarine. Les délimitations des domaines en nombre d'acides aminés sont indiquées par rapport à la protéine végétale. Le domaine MTase est composé de la région R et de la région α, le domaine N-terminal est composé d'une région GAR et d'un spacer. La région en rouge représente le motif permettant l'interaction de la Fibrillarine avec les ARN. Adaptée de (Rakitina *et al*, 2011a).

– ARN. Il permet donc à la Fibrillarine d’interagir physiquement avec ses ARN cibles. En ce sens, la région R possède le motif GCVYAVCF, caractéristique des interactions protéines – ARN. Cependant, cette séquence d’acide aminé n’est pas nécessaire à la liaison de la Fibrillarine aux ARN (Aris & Blobel, 1991), indiquant que d’autres motifs supplémentaires, remplissant également ce rôle, sont présents au sein du domaine MTase (Rakitina *et al*, 2011b). La partie C-terminale de la Fibrillarine est composée de sept feuillets Beta flanqués de sept hélices alpha. Cette région est nécessaire à son adressage vers les corps de Cajal (Snaar *et al*, 2000; Levitskiï *et al*, 2004) et permet à la Fibrillarine d’interagir physiquement avec NOP56 ou NOP58 (Lechertier *et al*, 2009). Le domaine MTase contient les motifs de liaisons au SAM, permettant l’activité méthyltransférase de la Fibrillarine (Rakitina *et al*, 2011b).

Le domaine N-terminal contient une région GAR, riche en arginine et en glycine. Les arginines de cette région sont les substrats de PRMT1 et PRMT7, deux méthyltransférases (Tang *et al*, 1998; Zurita-Lopez *et al*, 2012). La région GAR héberge alors de nombreuses diméthylations asymétriques au niveau des arginines, faisant de la Fibrillarine une protéine hyperméthylée (Yagoub *et al*, 2015). Même si le rôle de ces méthylations ne sont pas encore clairement établis, elles semblent être impliquées dans la régulation des interactions du domaine N-terminal de la Fibrillarine avec des ARN, ou avec d’autres protéines (McBride & Silver, 2001). Ces arginines méthylées semblent également contribuer à la localisation nucléaire de la Fibrillarine. Cependant, les motifs de fixation à l’ARN du domaine MTase seraient nécessaire à son acheminement dans le nucléole (Snaar *et al*, 2000).

Contrairement à la Dyskérine des snoRNP H/ACA, la Fibrillarine est capable de catalyser des méthylations de manière indépendante de son intégration dans le complexe snoRNP C/D. Par exemple, la méthylation de la Glutamine 105 sur les histones H2A par la Fibrillarine fut la première marque épigénétique découverte ayant lieu dans le nucléole. Cette méthylation empêche la liaison du FACT (facilitates chromatin transcription complex), induisant un effet épigénétique au niveau des promoteurs actifs de l’ARN polymérase I (Tessarz *et al*, 2014; Rodriguez-Corona *et al*, 2015)

La Fibrillarine peut également interagir avec des protéines virales. Elle se lie, via la région GAR, à la protéine ORF3, codée par le virus de la rosette de l’arachide, du genre *Umbravirus*. Cette liaison permet à la protéine virale d’être transportée entre les corps de Cajal et les nucléoles, mécanisme faisant partie intégrante du cycle de réplication du virus (Kim *et al*, 2007). Également, la protéine Tat du HIV, affectant la maturation des ARNr, interagit avec le

complexe U3 et la Fibrillarine (Ponti *et al*, 2008). Toutefois, le rôle des interactions entre la Fibrillarine et les protéines virales reste indéterminé.

I. 3. 4. Reconnaissance de l'ARN substrat par les RNP C/D

L'ARN substrat, contenant le ribose à méthyler, est sélectionné via un appariement direct des paires de bases par la snoRNP C/D. La reconnaissance de l'ARN cible se fait via des éléments antisens, situés en amont des boîtes D et D' de l'ARN guide. Le nombre de nucléotide reconnu par les éléments antisens varie d'une petite dizaine à une vingtaine de nucléotides. S'il est pris en compte seulement les éléments antisens situés en amont des boîtes D chez l'humain, ce nombre de nucléotides impliqués dans l'appariement varie de 8 à 19, avec une moyenne de 12 (Krogh *et al*, 2016). L'interaction physique entre l'ARN guide et l'ARN cible se fait sans renflement et n'accepte que très peu de mésappariement. De cette manière, l'ARN C/D guide la Fibrillarine sur le ribose du nucléoside à modifier, qui est invariablement celui associé au cinquième nucléotide en amont de la boîte D ou D' de l'ARN guide (Cavaillé *et al*, 1996).

Cependant, chez la levure, par un potentiel réarrangement de Nop56 et de Nop1, les snoRNP C/D U18 et snR48 sont capables de méthyler plusieurs nucléosides lors d'une même interaction snoARN/ARN substrat (van Nues & Watkins, 2017). De plus, snR13 dirige une méthylation sur le nucléoside apparié avec le sixième nucléotide de l'ARN guide. En revanche, ces méthylations impliquent les boîtes C' et D', dégénérées, de ces snoARN (van Nues & Watkins, 2017). L'idée que les snoARN pourraient, par une réorganisation structurale et un décalage sur l'ARN substrat, réaliser des modifications sur des éléments adjacents, commence à émerger.

I. 3. 5. Conformation structurale des RNP C/D

Les protéines cœurs qui composent la snoRNP C/D possédant de très fortes homologues entre les différents règnes, la structure de la snoRNP humaine a été déduite à partir d'études réalisées chez les archées et les levures, moins difficiles à mettre en place. En ce sens, la plupart des études structurales par cristallographie ont été faites sur les snoRNP C/D d'archée (Hamma & Ferré-D'Amaré, 2004; Moore *et al*, 2004). Des études plus récentes ont analysées par cryo-EM la structure du snoARN C/D U3 complexé avec la petite sous unité ribosomale chez la levure (Sun *et al*, 2017). Cependant, la structure de NHPX, complexée avec les snRNA U4, U5 et U6, a été étudiée chez l'humain (Vidovic *et al*, 2000). Les snoARN C/D présentent une forte flexibilité structurale au niveau des boîtes C/D et C'/D'. Lors de l'association d'une copie de

Snu13 avec les boîtes C/D et de la seconde copie avec les boîtes C'/D', la structure de l'ARN est modifiée, laissant apparaître respectivement, un motif en k-turn et k-loop (Vidovic *et al*, 2000; TURNER *et al*, 2005). Ces interactions sont nécessaires pour la structure de la snoRNP C/D et permettent la fixation des autres protéines cœurs (Marmier-Gourrier *et al*, 2003). Les protéines Nop56 et Nop58 se lient aux boîtes C' et C, respectivement. C'est au niveau de ces deux protéines qu'il est observé le plus de différences conformationnelles entre les snoRNP. Nop56 et Nop58, possèdent un domaine « coiled-coil », en supplément des domaines N et C terminaux (NTD et CTD, respectivement). Les éléments connectant ces trois domaines forment des charnières permettant la rotation des domaines NTD et CTD de chacune des protéines. Nop58 interagit avec la Fibrillarine ainsi qu'avec les nucléotides en amont de la boîte D, tandis que Nop56 se lie à une seconde copie de la Fibrillarine et avec la séquence précédant la boîte D' (Aittaleb *et al*, 2003). Ce sont les domaines NTD des protéines Nop56/58 qui se lient à la Fibrillarine, grâce à une grande flexibilité de l'articulation située entre les domaines NTD et coiled-coil. Cette flexibilité permet d'influer sur le positionnement dans l'espace de la Fibrillarine, permettant aux zones catalytiques de ces méthyltransférases de se retrouver au niveau du ou des nucléosides à modifier. Les domaines coiled-coil des deux protéines Nop56 et Nop58 interagissent ensemble (**Figure 7B**) (Aittaleb *et al*, 2003).

Chez l'archée, Nop5 se dimérise, de manière à remplir les rôles de ses homologues Nop56 et Nop58. Les domaines C-terminaux des protéines Nop5 servent d'éléments de liaison à l'ARN guide, au niveau des boîtes C/D et C'/D' (Lin *et al*, 2011). Nop5/56/58 pourraient alors participer à l'interaction de l'ARN guide avec l'ARN substrat (Granneman *et al*, 2009). La charnière située entre le domaine CTD et coiled-coil de Nop5/56/58 semble être beaucoup moins flexible que celle présente au niveau du NTD, rendant la conformation spatiale du CTD relativement stable selon les snoRNP observées (Aittaleb *et al*, 2003).

De manière intéressante, l'observation au microscope électronique de snoRNP de l'archée *Methanocaldococcus jannaschii* reconstituée *in vitro* a révélé des snoRNP complexée entre elles, de manière à former une paire (Bleichert *et al*, 2009). Ces di-RNP contiennent quatre copies de chaque protéine cœur (Fibrillarine, NOP5 et L7Ae) et deux snoARN C/D.

Les di-RNP pourraient être composées de deux copies de la même snoRNP, ou de deux RNP différentes (Lin *et al*, 2011). La tige des snoARN C/D contenant les motifs C/D serait cruciale à la formation de ces di-RNP, expliquant potentiellement le fait que toutes les snoRNP ne peuvent pas s'assembler en dimère (Yip *et al*, 2016). Également, une étude chez *Pyrococcus furiosus* montre qu'une snoRNP, considérée comme inactive, pourrait être fonctionnelle en

conformation de di-NRP. En effet, la snoRNP est inactive car, dans sa conformation, la Fibrillarine est trop éloignée de l'ARN substrat. En conformation de di-RNP, une Fibrillarine de la snoRNP 1 se retrouverait alors assez proche de l'ARN ciblé par la snoRNP 2, rendant possible le phénomène de méthylation (Xue *et al*, 2010).

I. 3. 6. Biogenèse et transport intranucléaire des RNP C/D

A l'instar des snoARN H/ACA, la quasi-totalité des snoARN C/D humains sont issus de la maturation des régions introniques des ARN messagers, transcrites par l'ARN polymérase II. Seuls les SNORD3, 118 et 13, respectivement U3, U8 et U13, sont à ce jour connus pour être transcrits via leur propre promoteur (Ismael *et al*, 2016).

Les facteurs d'assemblage impliqués dans la biogenèse sont conservés de la levure à l'homme. L'assemblage des snoRNP C/D nécessite l'intervention d'un grand complexe protéique qui permet le recrutement des protéines cœurs NHPX et NOP58. Le complexe R2TP agit comme un chaperon aidant à la formation de cet assemblage.

Le complexe R2TP est composé des ATPases AAA+ RUVBL1 et RUVBL2, et des adaptateurs PIH1D1 et RPAP3. Les ATPases du complexe se fixent à un hétérodimère composé de NUFIP et ZNHIT3 (Boulon *et al.*, 2008)(**Figure 9**). Les protéines cœurs NHPX et NOP58 sont probablement recrutées par NUFIP et les ATPases, respectivement. Le chargement en ATP de RUVBL1/2 permet l'association de ZNHIT6 et engendre la libération des régulateurs des ATPases du système R2TP, PIH1D1 et RPAP3 (Bragantini *et al*, 2016). Les protéines cœurs NHPX et NOP58 présentes dans le complexe protéique reconnaissent les séquences des snoARN C/D lors de leur transcription. L'intégration du snoARN dans le complexe protéique, ainsi que le recrutement de la Fibrillarine, initie la libération de la protéine ZNHIT3. Ensuite vient l'addition du second set de protéines cœurs, NOP56, NHPX et la Fibrillarine. Cet assemblage initie le départ de NUFIP et ZNHIT6. L'absence de NUFIP, puis le relargage de RUVBL1/2, entraînent la rotation du site catalytique de la Fibrillarine, rendant ainsi la particule C/D active.

Les protéines ZNHIT3 et ZNHIT6 sont caractérisées par la présence d'un motif « HIT ». Il s'agit d'un domaine en doigt de Zinc, composé de sept cystéines et d'une histidine. Ce motif serait impliqué dans la stabilité des snoARN, particulièrement en condition de stress (Bragantini *et al*, 2016). ZNHIT6 est nécessaire pour l'accumulation des snoARN C/D chez la

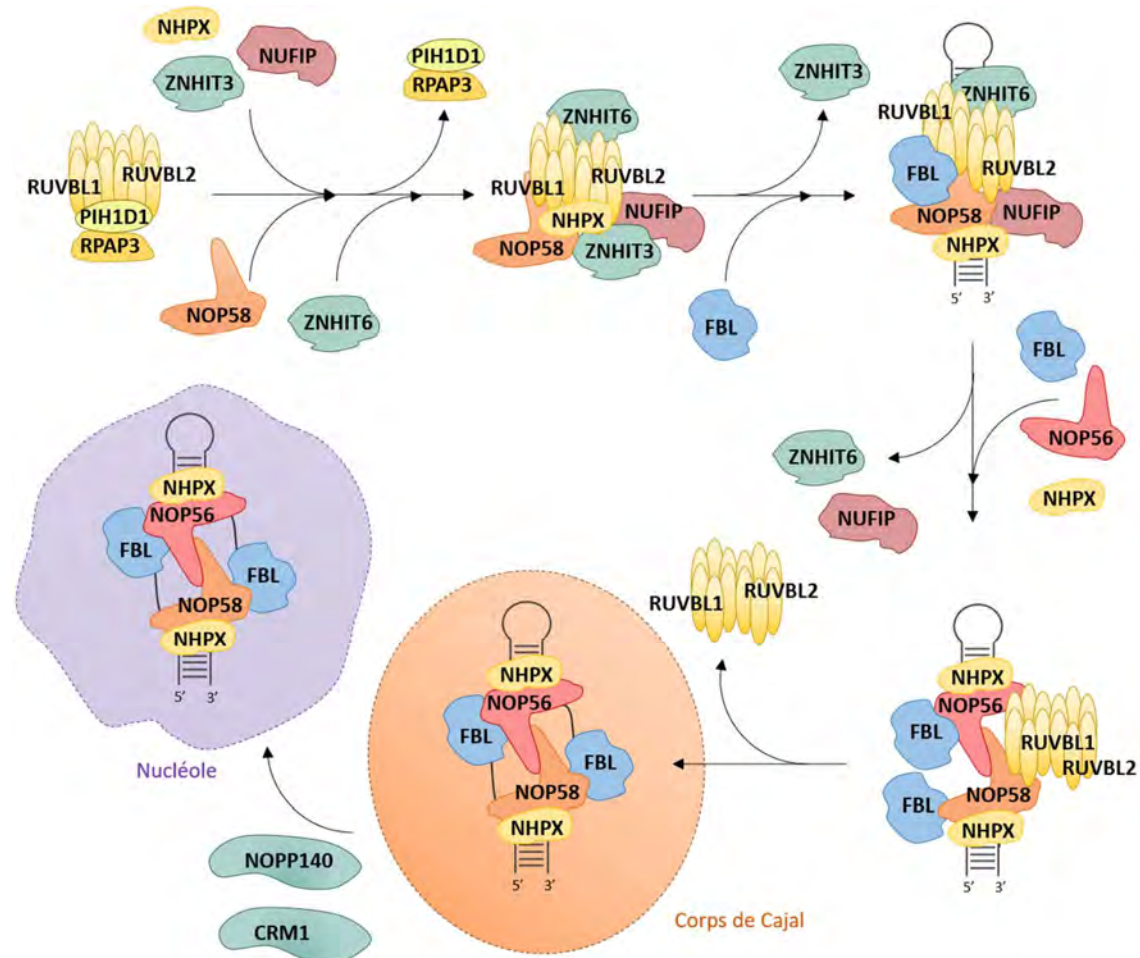


Figure 9 : Biogenèse et transport intranucléaire des snoRNP C/D

NUFIP et ZNHIT3 se fixent au complexe R2TP, composé des ATPases RUVBL1 et RUVBL2 et de ses adaptateurs PIH1D1 et RPAP3. Le chargement en ATP de RUVBL1/2 induit la dissociation des adaptateurs des ATPases PIH1D1 et RPAP3 ainsi que le recrutement de ZNHIT6. Les protéines cœurs NOP58 et NHPX sont recrutées et se fixent au complexe précédemment formé. NOP58 et NHPX permettent l'intégration du snoARN C/D dans le complexe lorsque ce dernier est en train d'être transcrit. Cette étape, en supplément de l'ajout de la Fibrillarine (FBL) dans le complexe, engendre le départ de ZNHIT3. L'arrivée du second set de protéines cœurs engendre la dissociation de ZNHIT6 et de NUFIP, puis le relargage des ATPases du complexe R2TP. Les extrémités 5' et 3' du snoARN sont digérées par des exonucléases. Les scaRNP sont retenues dans les corps de Cajal, et les snoRNP sont acheminées en direction des nucléoles, via NOPP140 et CRM1. Adaptée de (Massenet *et al*, 2016b) et de (Baldini *et al*, 2021a)

levure, mais son rôle précis dans l'assemblage des snoRNP C/D n'a pas encore été élucidé (Paul *et al*, 2019). De récents travaux ont cependant mis en évidence que ZNHIT6 (Bcd1) assiste l'incorporation de Nop58 lors de la biogenèse de la snoRNP. Également, la protéine C12orf45 est connue pour interagir avec NOP58 et l'ATPase RUVBL1 lors de l'assemblage, mais les mécanismes impliqués restent imprécis (Bragantini *et al*, 2016).

Dès le débranchement des introns lors du phénomène d'épissage, les régions flanquant des snoARN C/D sont ciblées par la dégradation. Les extrémités 5' et 3' du snoARN C/D sont formées par les exonucléases XRN1/2 et par l'exosome aidé par le complexe NEXT, respectivement (Massenet *et al*, 2016a). L'assemblage avec les protéines cœurs durant la biogenèse permet de replier le snoARN de manière à former une extrémité double brin, le protégeant des exonucléases (Falaleeva *et al*, 2017).

D'autres facteurs seraient nécessaires à la maturation des snoRNP C/D. Il a notamment été observé que la position des gènes des snoARN situés à une quarantaine de nucléotides en amont du point de branchement 3' lors de l'épissage serait une condition favorable pour leur biogenèse et leur maturation (Hirose & Steitz, 2001). L'assemblage des snoARN situés dans cette position est facilité par le facteur d'épissage Aquarius (IBP160), une protéine interagissant avec l'ARN de l'épissage U2, mettant en avant un lien entre l'épissage et la biogenèse des snoRNP C/D (Hirose *et al*, 2006).

Les snoRNP C/D sont exclusivement nucléaires. Après leur biogenèse dans le nucléoplasme, les snoARN sont acheminés dans les nucléoles. Chez les eucaryotes supérieurs, les snoRNP transitent dans les corps de Cajal avant de se retrouver dans les nucléoles (Machyna *et al*, 2014) (**Figure 9**). Leur maturation a entièrement lieu dans le noyau, à l'inverse des snARN U1, U2, U4 et U5, qui transitent dans le cytoplasme au cours de leur maturation (Kiss, 2004).

Les différentes études sur le transport nucléaire des snoRNP indiquent que les motifs C et D sont nécessaires et suffisants pour leur localisation nucléolaire (Narayanan *et al*, 1999; Watkins *et al*, 2002). Cependant, les acteurs et mécanismes impliqués dans le transport des snoRNP C/D du nucléoplasme aux autres territoires nucléaires sont encore inconnus ou mal compris. La déplétion de NOPP140, une phosphoprotéine, engendre une perte de la localisation des snoRNP dans les nucléoles et les corps de Cajal (Meier & Blobel, 1992). Il est alors proposé que NOPP140 serait impliquée dans le transport des snoRNP C/D entre ces deux territoires nucléaires (Isaac *et al*, 1998).

Le transport nucléaire de SNORD3 (snoARN U3) a été caractérisé, mais il s'agit d'un snoARN transcrit à partir de son propre promoteur, à l'inverse de la majorité des snoARN C/D. Ces études ont notamment démontré l'implication de TGS1 et CRM1 dans la maturation et l'assemblage de SNORD3, respectivement. Les domaines C-terminaux de la Dyskérine et des protéines NOP56/58 ont des régions très riches en méthylations, qui s'apparentent à des signaux de localisation nucléolaire (NoLS). TGS1 possède, en sa partie C-terminale, un domaine de fixation aux domaines C-terminaux de DKC1, et NOP56/58. Il est proposé que la fixation de TGS1 masquerait ces signaux NoLS (Pradet-Balade *et al.*, 2011). CRM1, une protéine du pré-complexe SNORD3, impliquée dans le transport nucléo-cytoplasmique, déplacerait TGS1 et engendrerait la relocalisation nucléolaire de SNORD3 via les domaines NoLS. Ce mécanisme a été confirmé *in-vitro* (Pradet-Balade *et al.*, 2011). CRM1 étant également nécessaire à la localisation nucléolaire des snoRNP C/D intronique, et TGS1 interagissant avec les snoRNP, il y a des chances que les mécanismes régulant le transport nucléaire des snoRNP introniques soient similaires à ceux qui régissent SNORD3 (Boulon *et al.*, 2004).

II. Fonctions cellulaires des RNP H/ACA et C/D

II. 1 Les RNP H/ACA et C/D, guides de modifications post-transcriptionnelles

Les fonctions, pouvant être considérées comme canoniques des snoRNP et scaRNP H/ACA et C/D, sont de guider, respectivement, les pseudouridylations et les méthylations en 2'-O-ribose sur les ARN cellulaires. Cependant, ces RNP peuvent être associés à d'autres mécanismes et/ou fonctions au sein des cellules.

II. 1. 1. Les méthylations sur les ARN de transfert

Les ARNr et les snARN ne sont pas les seuls ARN cellulaires ciblés par les snoARN. Il est connu, depuis plus d'une vingtaine d'année, que les ARNt des archées peuvent être la cible de 2'-O-méthylations guidées par des snoARN. Il aura cependant fallu attendre 2019 pour découvrir un cas similaire chez l'être humain (Clouet-d'Orval *et al*, 2005; Vitali & Kiss, 2019).

Les ARNt sont les ARN cellulaires présentant les plus grands nombres de modifications chimiques, ainsi que les plus variées (Boccaletto *et al*, 2018). Les plus abondantes, à l'instar des autres ARN cellulaires, sont les pseudouridylations ainsi que les méthylations en 2'-O-ribose. Ces modifications, outre leur importance pour la maturation des ARNt, permettent un repliement correct de l'ARN et faciliteraient les interactions codons / anticodons (Duechler *et al*, 2016). Les modifications chimiques présentes sur un certain type d'ARNt peuvent varier selon les individus, ce qui permet d'avoir un impact sur l'efficacité et la fidélité de la traduction de certains gènes, et réguler des processus cellulaires (Zhang & Pan, 2022).

Chez les sARN, homologues des snoARN chez les archées, plus d'une vingtaine possédant les motifs C/D présentent des éléments antisens complémentaires avec des ARNt (Dennis *et al*, 2001). Ces sRNP sont capables de reconnaître et de modifier les ARNt cibles, en ajoutant des méthylations en 2'-O-ribose. Ces mécanismes ont été vérifiés par des sRNP C/D assemblées *in vitro* et par des sRNP endogènes obtenus via des extraits cellulaires (d'Orval *et al*, 2001; RENALIER *et al*, 2005).

Chez l'homme, le seul exemple de méthylation sur des ARNt est celle catalysée par le SNORD97 et son homologue scaARN SCARNA97 (Vitali & Kiss, 2019). Ces deux ARN aux boîtes C/D, possèdent des éléments antisens de séquences similaires, et guident une méthylation 2'-O-ribose sur la cytidine C₃₄ de l'ARNt^{MET} (CAT) (**Figure 10**). Des expériences de déplétions

ont démontré que ces deux snoRNP sont nécessaires et coopèrent pour réaliser cette modification chimique sur la cytidine de l'ARNt. Cette méthylation permet d'éviter le clivage endonucléotidique de l'ARNt par l'angiogénine lors de conditions de stress (Vitali & Kiss, 2019). En effet, les stress cellulaires engendrent un clivage des ARNt au niveau des séquences anticodons pour générer des fragments d'ARNt (tRFs) qui agissent comme des ARN régulateurs. SNORD97 et SCARNA97 pourraient alors intervenir, par ce mécanisme, dans le contrôle de la réponse au stress, l'adaptation de la traduction, et dans la dégradation des ARNm (Ivanov *et al*, 2011; Goodarzi *et al*, 2015; Vitali & Kiss, 2019).

II. 1. 2. Les snoARN guident des acétylations

Deux cytidines acétylées sont présentes sur l'ARNr18S dans la région en amont de l'extrémité 3'. Les réactions d'acétylations consistent en l'ajout d'un groupe acétyle (-CO-CH₃) en position N4 des cytosines (**Figure 11**). Ces cytidines acétylées (ac⁴C) sont situées sur les hélices 34 et 45 de l'ARNr 18S (Sharma *et al*, 2015). Ces deux acétylations sont catalysées chez la levure et l'humain par les acétyltransférases Kre33 et NAT10, respectivement (Ito *et al*, 2014; Sharma *et al*, 2015). De manière intéressante, il a été démontré que ces acétylations, en supplément de l'acétyltransférase, requièrent la présence d'un snoRNP. Chez la levure, les snoARN C/D snR4 et snR45 possèdent des séquences complémentaires bipartite avec l'ARNr 18S en amont des boîtes C et C', leur permettant de s'apparier de part et d'autre de ces nucléotides acétylés. SnR4 se lie au niveau du nucléotide ac⁴C₁₂₈₀ située sur l'hélice 34, et snR45 au niveau de l'ac⁴C₁₇₇₃ sur l'hélice 45. Ces appariements sont dépendants de l'activité hélicase putative de Kre33. (Sharma *et al*, 2017). SnR4 et snR45 sont des snoRNP dont Nop1 fait bien partie des protéines cœurs qui leurs sont associées, bien que, en raison d'un manque de complémentarité, aucune méthylation en 2'-O-ribose ne leur ont été associées. En revanche, snR4 et snR45 sont également directement liées à Kre33, et guident cette acétyltransférase vers les sites à modifier sur l'ARNr 18S, grâce à une forte complémentarité. Cependant, ces complémentarités impliquent des nucléotides autres que les éléments antisens traditionnels, situés en amont des boîtes D et D'. L'activité de Kre33 est alors nécessaire pour établir ces paires de bases non-conventionnelles. En ce sens, une déplétion de snR4 et snR45 mène aux pertes des acétylations aux positions 1280 et 1773 de l'ARNr18S, tandis que les acétylations dirigées par Kre33 sur les ARNt restent inchangées. Ces snoARN C/D favoriseraient alors l'interaction de Kre33 avec les hélices 34 et 45 de l'ARNr 18S, permettant l'acétylation des cytosines correspondantes (Sharma *et al*, 2017).

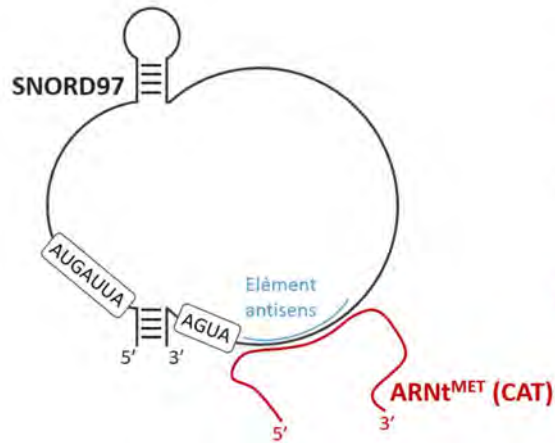
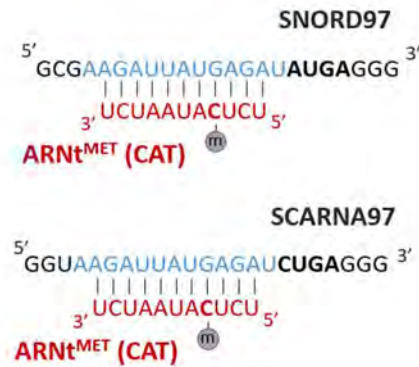
A**B**

Figure 10 : Interactions entre les éléments antisens des ARN SNORD97 et SCARNA97 avec l'ARNt^{MET} (CAT). Adaptée de (Vitali & Kiss, 2019)

(A) Représentation schématique de SNORD97 (noir) en interaction avec l'ARNt^{MET} (CAT) (rouge). Les boîtes C et D, respectivement en 5' et 3' de SNORD97 sont encadrées. La position de l'élément antisens est surlignée en bleu. (B) Appariements prédits entre SNORD97/SCARNA97 et l'ARNt^{MET} (CAT). Les boîtes D, situées en aval des éléments antisens (en bleu), sont en gras. Le résidu de l'ARNt méthylé est spécifié.

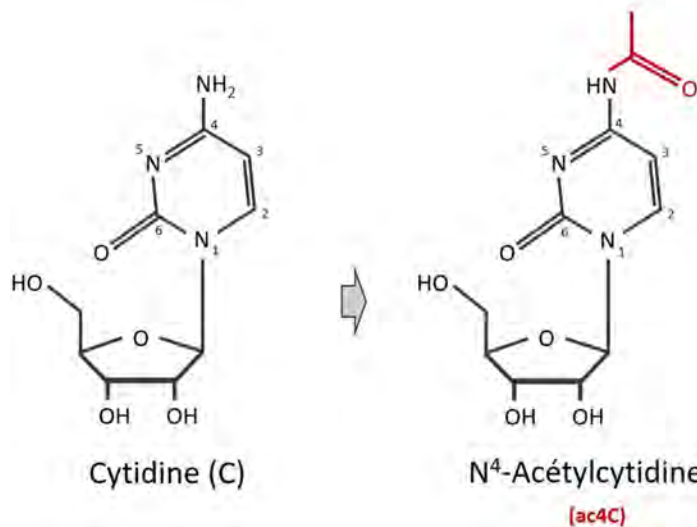


Figure 11 : La N⁴-Acétylcytidine

Représentation schématique d'une acétylation (ajout d'un groupement -CO-CH₃) au niveau du quatrième carbone d'une cytidine. Les ac4C sont des modifications chimiques conservées qui sont retrouvées au niveau des ARNm, ARNt et ARNr. Les ac4C présentes sur les ARNr et ARNt permettent d'augmenter la fidélité du mécanisme de traduction.

Chez l'humain, SNORD13 présente des appariements sur l'ARNr 18S semblables à snR45. Deux séquences de SNORD13, qui ne sont pas des régions antisens canoniques, s'apparient avec l'ARNr18S (**Figure 12**). L'inactivation de SNORD13 chez les cellules humaines engendre la perte de l'acétylation correspondante à celle ciblée par snR45 chez la levure, $\text{ac}^4\text{C}_{1842}$ (Bortolin-Cavaillé *et al*, 2022). Cependant, SNORD113 et l' $\text{ac}^4\text{C}_{1842}$ ne sont pas associés à des défauts de biogenèse des ribosomes ou de croissance cellulaire. Aucune anomalie du développement n'a également été observée lors du développement embryonnaire chez le poisson zèbre.

En revanche, SNORD13 reste un ARN conservé évolutivement. Trois snoARN orthologues de SNORD113 sont présents chez la plante *Arabidopsis thaliana* (snoR105, 108 et 146), mais ils ne sont pas encore associés à des acétylations (Kim *et al*, 2010). SNORD13 est également présent chez la drosophile *Drosophila melanogaster*, même s'il présente une extrémité 5' tronquée, indiquant alors l'absence d'un des deux appariements avec l'ARN cible (**Figure 12**, Appariement 1). Les acétylations dirigées via SNORD13 et NAT10, l'acétyltransférase humaine, ne pourraient alors n'impliquer qu'un seul élément antisens (Bortolin-Cavaillé *et al*, 2022).

II. 2. Rôles des snoRNP dans la maturation des ARNr et dans le maintien des télomères

II. 2. 1. Implication dans le chaperonnage et le clivage des ARNr

En supplément de leurs rôles de guide des pseudouridylations et des méthylations en 2'-O-ribose, les snoARN sont impliqués dans d'autres événements cellulaires. En ce sens, SNORD3, SNORD118, SNORD13 et SNORD14 (respectivement U3, U8, U13 et U14) sont des snoRNP aux motifs C/D, jouant un rôle de chaperonne et participant au clivage des ARN ribosomiques lors de leur maturation. Chez les eucaryotes, les ribosomes matures sont le résultat d'un assemblage entre deux sous-unités bien distinctes. La grande sous unité 60S contient les ARNr 28S, 5,8S et 5S, tandis que la petite sous unité 40S héberge l'ARNr 18S. Les ARNr 28S, 18S et 5,8S sont issus de la maturation d'un long précurseur (pré-ARNr 47S) transcrit par l'ARN polymérase I (**Figure 13**)(Aubert *et al*, 2018; Ojha *et al*, 2020).

Les SNORD3 et SNORD8 se sont révélés être essentiels à la synthèse de la petite sous unité du ribosome (SSU) et de la grande sous unité ribosomale (LSU), respectivement.

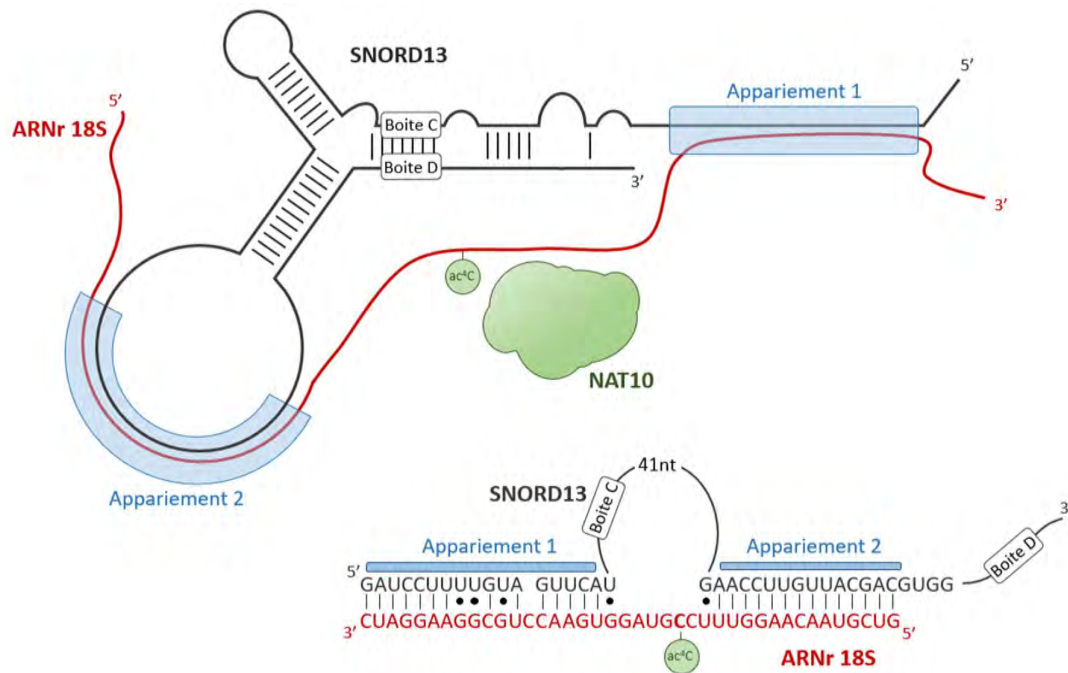


Figure 12 : Représentation schématique de la formation d'une ac⁴C sur l'ARNr 18S via une interaction avec SNORD13.

Les régions d'appariements entre SNORD13 (noir) et l'ARNr 18S (rouge) sont encadrées en bleu. Le résidu cible de l'acétylation (ac⁴C, en vert) est en gras. Adaptée de (Bortolin-Cavaillé *et al*, 2022; Thalalla Gamage *et al*, 2022).

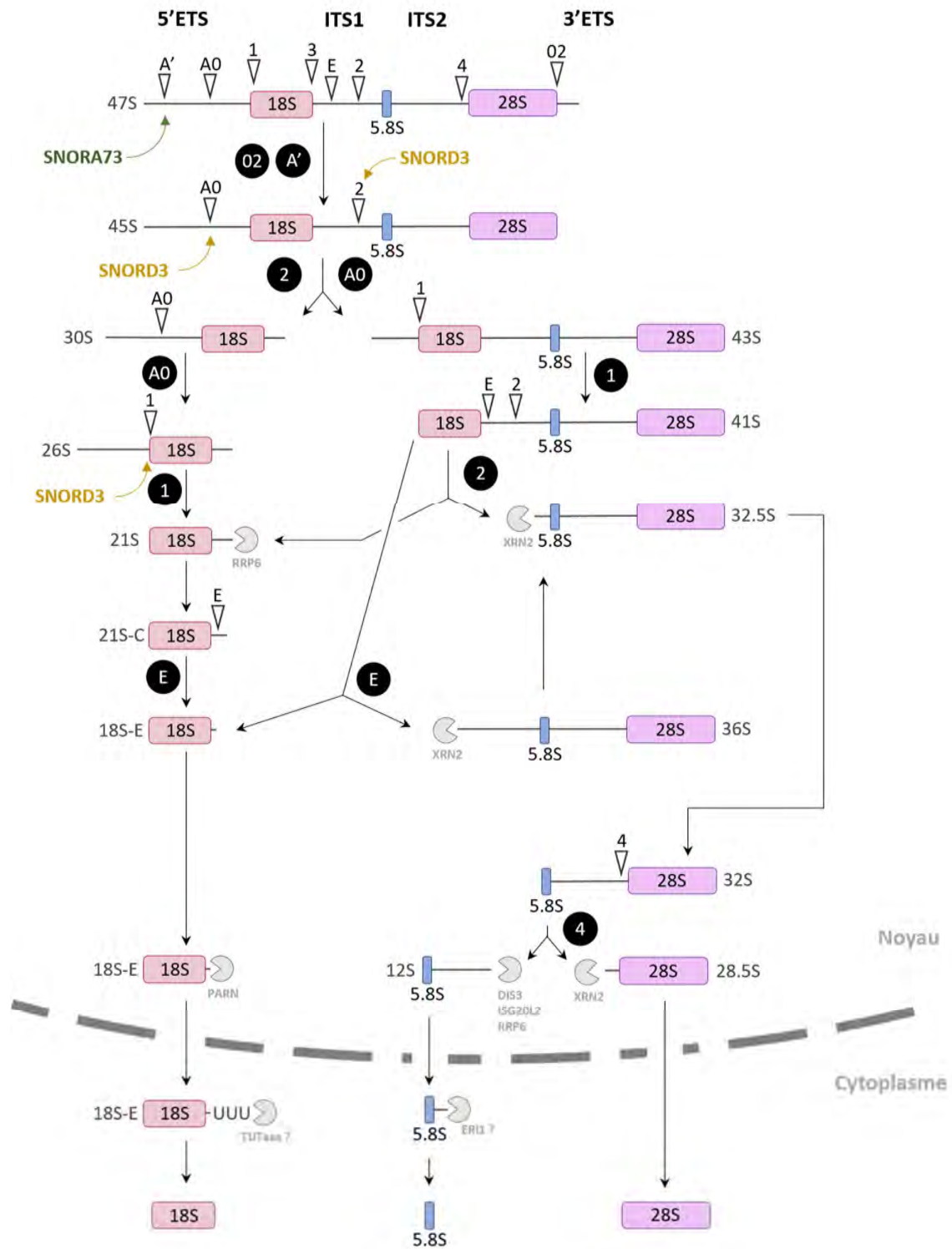


Figure 13 : Maturation des ARNr.

Représentation schématique de la maturation des ARNr issus du long transcrit par la polymérase I. Les snoRNP étant impliquées dans les sites de clivages via leur rôle de chaperon y sont représentées. Adaptée de (Aubert *et al*, 2018)

SNORD3 est présent dans la globalité des eucaryotes observés, tandis que SNORD118 est présent seulement chez les vertébrés (Tyc & Steitz, 1989; Myslinski *et al*, 1990).

L'ARN SNORD3, en supplément des protéines cœurs des snoARN C/D, s'associe aux protéines Sof1p, Mpp10, Lcp5p, Imp3p, Imp4p et U3-55k. Chacune de ces protéines se sont retrouvées être nécessaire pour la maturation de l'ARNr 18S, à l'instar du snoARN SNORD3 et de ses protéine cœurs (Lukowiak *et al*, 2000). SNORD3 joue un rôle critique dans les étapes de clivages du pré-ARNr 18S en agissant comme un chaperon. Tandis que la région en 3' de l'ARN SNORD3 porte les motifs consensus C et D, la région en 5' du snoARN contient trois séquences qui s'associent au pré-ARNr 18S (Phipps, Charette and Baserga, 2011). Une de ces séquences s'hybride avec une région au cœur de l'ARNr 18S, tandis que les deux autres interagissent avec l'extrémité 5' ETS (External Transcribed Spacer) de l'ARNr pré-ribosomique (**Figure 14C**). Ces interactions empêchent la formation d'une structure tertiaire en pseudonœud du pré ARNr 18S, structure nécessaire à l'activité de l'ARN 18S mature. Ce mécanisme permettrait de séquestrer l'extrémité 5' du pré-ARNr 18S en jouant le rôle de chaperon, jusqu'aux dernières étapes de maturation de l'ARNr (Phipps, Charette and Baserga, 2011). Par ailleurs, une déplétion de SNORD3 engendre une accumulation du pré-ARNr 35S, due à une absence de clivages aux sites A0, A1 et A2 (Phipps, Charette and Baserga, 2011). Cependant, même si requise, la présence de SNORD3 n'est pas suffisante pour réaliser les clivages du pré-ARNr (Ojha *et al*, 2020).

A l'instar de SNORD3, SNORD13 et SNORD14 participent également à la maturation de l'ARNr 18S durant la synthèse des ribosomes. Comme vu précédemment, deux régions de SNORD13 ciblent l'ARNr 18S pour y guider deux acétylations, qui sont nécessaires à la maturation de l'ARNr. L'ARN SNORD14, quant à lui, s'apparie également sur deux régions de l'ARNr 18S, localisées au niveau de l'extrémité 5' de l'ARNr (**Figure 14A**). Une de ces interactions est impliquée dans le chaperonnage de l'ARNr, tandis que la seconde réalise la fonction canonique en guidant une méthylation en 2'-O-ribose à la position 462 de l'ARNr 18S. Cette méthylation a lieu lors de l'étape 40S du précurseur de l'ARNr 18S. Il n'est pas encore connu si cette méthylation est requise pour la maturation de l'ARNr 18S, ou s'il s'agit-là d'une fonction indépendante (Dunbar & Baserga, 1998). Dans tous les cas, la déplétion de SNORD14 engendre des défauts de maturation de l'ARNr 18S (Li *et al*, 1990).

La maturation de l'ARNr 18S au cours de la synthèse des ribosomes semble impliquer d'autres snoRNP, comme SNORD22 (U22) et SNORA73 (U17) (Tycowski *et al*, 1994; Enright *et al*, 1996). De manière intéressante, SNORA73 est une snoRNP présentant les motifs H/ACA.

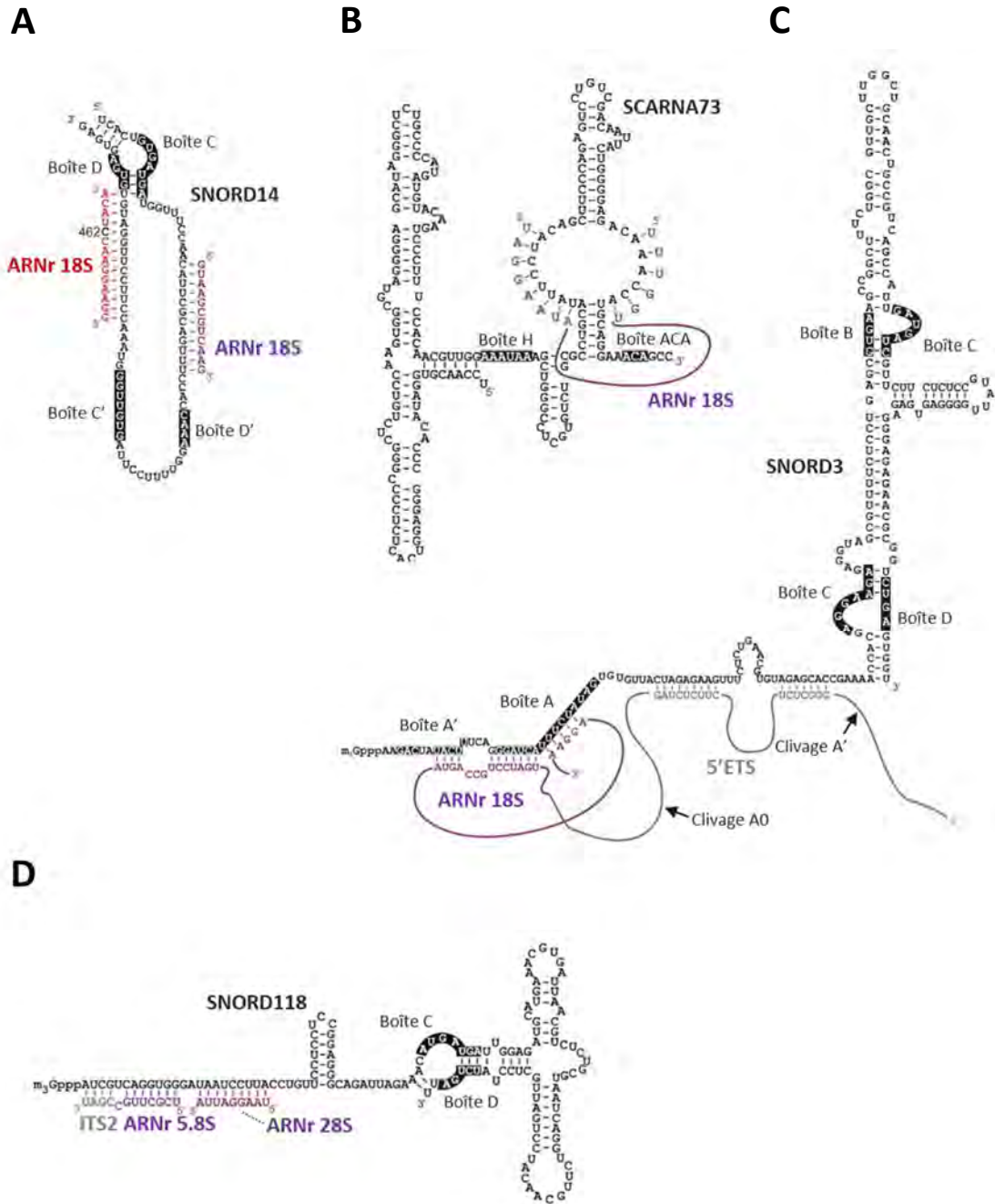


Figure 14 : Interactions entre les ARNr et les snoARN jouant un rôle de chaperon lors de la maturation des pré-ARNr.

Représentation schématique des appariements des SNORD3, 14, 118 et du SNORA73 avec les pré-ARNr, respectivement dans les encadrés C, A, D et B. Les interactions ayant l'ARN cible en violet sont associées à une fonction de chaperon, tandis que celle en rouge est associée à une méthylation en 2'-O-ribose. Adaptée de (Watkins & Bohnsack, 2012).

De plus, les deux interactions entre l'ARN SNORA73 et l'ARNr 18S impliquent, à défaut des élément antisens, les nucléotides situés sur la partie inférieure de la poche de pseudouridylation du snoARN (**Figure 14B**). En ce sens, aucune pseudouridylation n'est catalysée lors de ces interactions (Enright *et al*, 1996). Ces interactions sont nécessaires au clivage du site A' du pré-ARNr 47S, aboutissant à la formation de l'ARNr 18S (Enright *et al*, 1996).

SNORD118 est en revanche requis pour la maturation des ARNr 28S et 5,8S, deux ARN avec lesquels il interagit. En effet, l'ARN SNORD118 s'apparie sur le pré-ARNr 60S au niveau de la région centrale de l'ARN 28S, au niveau de l'extrémité 3' de l'ITS2 (Internal Transcribed Spacer), et sur l'extrémité 5' de l'ARNr 5,8S (**Figure 14D**) (Michot *et al*, 1999). A défaut de catalyser des modifications chimiques, les interactions entre l'ARN SNORD118 et le pré-ARNr sont nécessaires à la maturation des ARNr 28S et 5,8S. Des expériences de mutations de la région 5' de l'ARN SNORD118 ont en ce sens montré une inhibition de la maturation du pré-ARNr (Peculis, 1997). Également, déplétion de SNORD118 chez le Xénope aboutit à une grande sous-unité ribosomique immature. Ces faits ont également été observés chez la souris et l'humain (Srivastava *et al*, 2010; Langhendries *et al*, 2016).

Une partie des snoRNP connue est alors impliquée dans le processus de maturation des pré-ARNr. Lors de la synthèse des ARNr, ces snoRNP assurent un meilleur repliement de ces derniers, jouent un rôle de chaperon en facilitant les clivages, et les empêchent d'interagir de manière non-spécifique avec d'autres molécules (Lafontaine, 2015).

II. 2. 2. L'ARN de la Télomérase

La télomérase est une enzyme spécialisée dans la maintenance de l'ADN situé aux extrémités des chromosomes (Greider & Blackburn, 1989).

En effet, lors de la réplication de l'ADN, une amorce d'ARN doit s'hybrider sur le brin opposé, afin de fournir une extrémité 3' -OH libre, permettant ainsi à l'ADN polymérase d'initier la réplication. Cependant, l'élimination ultérieure de cette amorce d'ARN entraîne la perte d'une vingtaine de nucléotides à la fin de chaque chromosome, et ce, à chaque cycle de réplication. Généralement, les pertes répétées d'ADN télomérique rendent les télomères trop courts et non-fonctionnels, dirigeant les cellules vers la sénescence (Armanios & Blackburn, 2012). La télomérase, grâce à une activité rétrotranscriptase, ajoute des séquences répétées d'ADN télomérique aux extrémités des chromosomes en utilisant une région de l'ARN non-codant la composant comme modèle (template) afin de contrer ce processus. Fait intéressant,

l'ARN de la télomérase possède un domaine H/ACA et s'associe avec les protéines cœurs des snoRNP H/ACA au sein de la télomérase. L'ARN de la télomérase possède une coiffe m⁷G, issue de la transcription d'un gène indépendant par l'ARN polymérase II. La biogenèse de l'ARN H/ACA de la Télomérase diffère alors des snoARN H/ACA issus des régions introniques. La maturation l'extrémité 5' de l'ARN de la Télomérase est réalisée par la méthylation de TGS1, donnant une coiffe m^{2,2,7}G (Pradet-Balade *et al.*, 2011).

L'ARN de la télomérase est composé de différents domaines : le domaine « pseudoknot », contenant la séquence template pour la synthèse des séquences répétées d'ADN télomérique, le domaine CR4/5, et un domaine H/ACA (**Figure 15**)(Chen *et al.*, 2000). Le domaine H/ACA de la télomérase recrute les protéines cœurs DKC1, NOP10, NHP2, GAR1 et TCAB1 (Dez *et al.*, 2001b; Nguyen *et al.*, 2018). Le domaine H/ACA possède également les boîtes CAB et BIO. Le motif CAB, reconnu par la protéine cœur TCAB1, est positionné sur la boucle terminale de la structure en tige-boucle du domaine H/ACA. En ce sens, la télomérase est localisée dans les corps de Cajal (Jády *et al.*, 2004). Le domaine H/ACA de l'ARN de la télomérase n'est associé à aucune fonction de guide de pseudouridylation, et n'est pas directement nécessaire à l'activité enzymatique de la télomérase. En effet, la position des poches de pseudouridylation ne permet pas de placer les nucléotides d'un ARN apparié au niveau du site catalytique. Cependant, le domaine H/ACA joue un rôle crucial dans l'accumulation de l'ARN de la télomérase *in vivo*, ainsi que dans sa localisation cellulaire et la maturation de son extrémité 3'. Et la localisation de l'ARN de la télomérase dans les corps de Cajal est nécessaire à sa fonction (Mitchell *et al.*, 1999b; Fu & Collins, 2003; Jády *et al.*, 2004).

II. 3. Autres fonctions associées aux snoARN

II. 3. 1. Le concept de snoARN orphelin

L'étude des génomes a permis d'identifier un nombre important de snoARN. Ces ARN non codants ont, comme vu précédemment, des structures et des motifs fortement conservés au cours de l'évolution. Il a alors été possible de déterminer, par approches bio-informatiques, les séquences des éléments antisens de ces snoARN, afin d'élucider les ARNr et snARN qui en sont les cibles. Malgré cela, plus d'un tiers des snoARN connus à ce jour ne possèdent pas encore de cibles proposées ou découvertes (Lestrade & Weber, 2006).

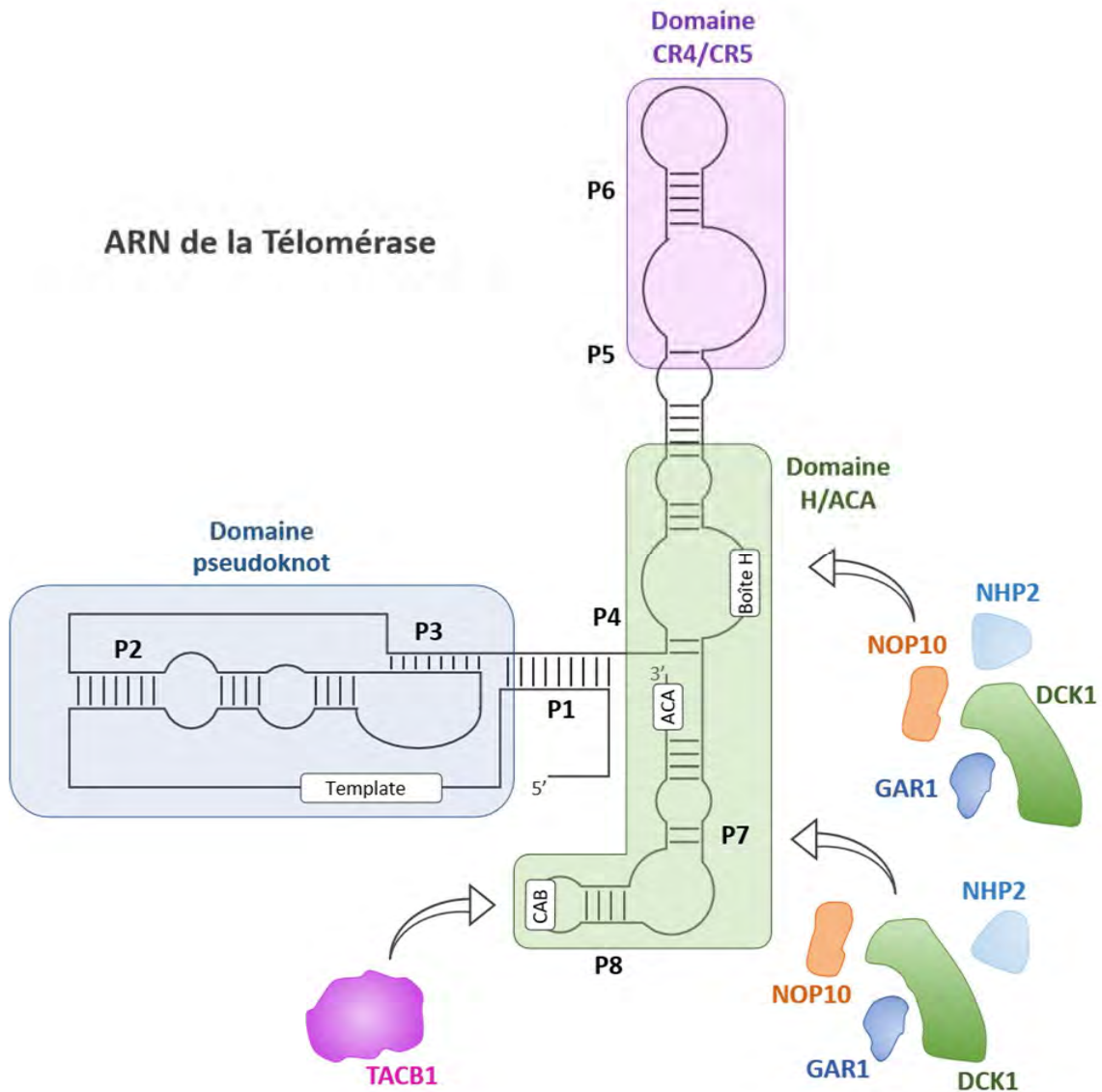


Figure 15 : Structure secondaire de l'ARN de la Télomérase humaine.

Représentation schématique de l'ARN de la Télomérase humaine. Les domaines pseudoknot, H/ACA et CR4/CR5 sont respectivement encadrés en bleu, vert et violet. Les boîtes consensus de la région H/ACA sont représentées. La protéine TACB1, à l'image des scaARN, s'associe au motif CAB située sur la région apicale de la tige-boucle 3' du domaine H/ACA, permettant la localisation de l'ARN de la Télomérase dans les corps de Cajal.

De là découle ce concept de « snoARN orphelin », qui désigne ces ARN, dont leurs séquences et leurs structures sont connues, mais dont la communauté scientifique s’efforce toujours d’en déterminer les cibles.

Le fait que les approches bio-informatiques ne révèlent aucun candidat pour ces snoARN orphelins parmi les ARNr et snARN, laisse présager qu’ils puissent cibler d’autres types d’ARN cellulaire. Dans cette optique, d’autres familles du transcriptome sont connues pour être la cible de pseudouridylation et de méthylations en 2’-O-ribose. En effet, de récentes études ont notamment mis en évidence que les ARNm sont porteurs de ces modifications post-transcriptionnelles. Cela peut par ailleurs correspondre avec le fait que certains snoARN orphelins soient peu abondants dans les cellules. Parallèlement, la faible abondance de certains snoARN, et le fait que certains sont exprimés seulement dans des tissus spécifiques, constituent des limites supplémentaires dans la recherche de leur ARN cible.

Il est aussi envisageable que les snoARN orphelins n’assurent pas de pseudouridylation ou de méthylations en 2’-O-ribose, mais qu’ils soient impliqués dans de nouveaux rôles ou mécanismes non-canoniques au sein des cellules.

II. 3. 2. Les SnoARN associés aux maladies

II. 3. 2. 1 Le syndrome Labrune

La leucoencéphalopathie avec kystes et calcifications (LCC), également nommée « syndrome Labrune » après le premier report de cette pathologie par Labrune en 1996, est une maladie génétique autosomique récessive rare (Labrune *et al*, 1996). Les principales caractéristiques d’imagerie cérébrales des patients laissent apparaître des calcifications cérébrales étendues, la formation de kystes parenchymateux et une leucodystrophie (atteintes de la substance blanche du système nerveux central) (Joshi *et al*, 2021). Ces symptômes sont généralement causés par des lésions cérébrales ischémiques et hémorragiques (microangiopathie cérébrale). Ces symptômes sont associés à un déclin des capacités cognitives et motrices des patients (Livingston *et al*, 2014).

Cette pathologie est classée via les observations cliniques, et malheureusement les facteurs de causes exacts ne sont pas déterminés. Il n’existe à ce jour aucun moyen de soigner la maladie, seulement des traitements permettant d’alléger les symptômes (McFadden and Baserga, 2022).

Assez récemment, il a été découvert que les patients atteints du syndrome Labrune possèdent des délétions et des mutations au niveau du gène codant pour SNORD118 (U8). L'étude de ces mutations montrent des défauts de maturation, d'expression et d'assemblage des protéines cœurs du SNORD118 (Jenkinson *et al*, 2016). Comme vu précédemment, ce snoARN découvert en 1985 est impliqué dans la maturation des ARNr 28S et 5,8S lors de la synthèse des ribosomes (Reddy *et al*, 1985; Peculis, 1997). L'observation de lignées de poissons zèbres déplétées de SNORD118 montre des perturbations du développement du système nerveux central et de la biogenèse des ARNr, ainsi qu'une activation de la voie de signalisation p53 (Badrock *et al*, 2020). La stabilisation de p53 est par ailleurs un marqueur de la réponse au stress nucléolaire (Farley & Baserga, 2016). Ces poissons zèbres mutants ont été sauvés par l'ajout du précurseur humain du snoARN SNORD118, indiquant une conservation des mécanismes. L'activation de cette voie de signalisation, induite par l'absence du SNORD118 et par conséquent, un défaut de production nucléolaire de ribosomes, pourrait être alors une cause du syndrome de Labrune (Badrock *et al*, 2020).

II. 3. 2. 2. Le syndrome d'Alazami

Le syndrome d'Alazami est une rare pathologie autosomique récessive, caractérisée par des troubles neurodéveloppementaux. Les premiers cas ont été décrits par Alazami et son équipe en 2012, dans une famille consanguine atteinte de nanisme primordial et présentant de sévères déficiences intellectuelles. Le syndrome d'Alazami est causé par une perte de fonction du gène LARP7 due à des mutations bi-alléliques (Alazami *et al*, 2012).

Le gène LARP7 code pour une protéine membre de la famille LARP (La Related Proteins). De par son association avec l'ARN non codant 7SK, LARP7 est impliquée dans la régulation négative de l'élongation de la transcription par la polymérase II (Egloff *et al*, 2018). LARP7 étant nécessaire à l'accumulation et à la stabilisation du snARN 7SK, une perte de fonction de LARP7 engendre une déplétion de 7SK. L'absence de 7SK tend vers un relargage du facteur d'élongation de la transcription P-TEFb, mécanisme pouvant expliquer les phénotypes neurodéveloppementaux et le nanisme primordial caractéristiques du syndrome d'Alazami (Alazami *et al*, 2012).

Cependant, la perte de 7SK seule ne serait pas suffisante pour expliquer l'apparition du syndrome d'Alazami (Dai *et al*, 2014). Fait intéressant, une étude récente a montré qu'une déplétion de LARP7 engendre une baisse du niveau de 2'-O-ribose méthylations sur l'ARN de l'épissage U6 (Hasler *et al*, 2020). Le snARN U6 est un des ARN de l'épissage les mieux

conservés, et joue un rôle primordial dans l'épissage des ARN (Didychuk, Butcher and Brow, 2018). Le snARN U6 est connu pour être porteur de nombreuses méthylations en 2'-O-ribose, catalysées par les SNORD 7, 8, 9, 10, 67 et 94 (Tycowski *et al*, 1998; Ganot *et al*, 1999; Hüttenhofer *et al*, 2001; Vitali *et al*, 2003). Le snARN U6, ainsi que les snoARN C/D impliqués dans les modifications chimiques du snARN U6 sont retrouvés lors d'expériences de purification de la protéine LARP7. Le module caractéristique « La » des protéines LARP se fixe au snARN U6, tandis que la région C-terminale RRM2 interagit avec les SNORD modifiant U6 (**Figure 16A**). En ce sens, le motif LAB, nécessaire à la fixation de LARP7, est particulièrement présent au niveau des ARN SNORD 7, 8, 9, 10, 67 et 94 (**Figure 16B**). Il est proposé que LARP7 facilite les réactions de méthylations sur le snARN U6 en se liant à la fois à l'ARN cible (U6) et aux ARN catalysant la réaction (SNORD) (Hasler *et al*, 2020; Wang *et al*, 2020). En ce sens, une méthode permettant de cartographier les méthylations en 2'-O-ribose sur l'ensemble du transcriptome, a révélé des taux de ces modifications très faible chez les patients atteints du syndrome d'Alazami (Hasler *et al*, 2020).

Les modifications chimiques post-transcriptionnelles au niveau des ARN de l'épissage sont impliquées dans la biogenèse des snRNP ainsi que dans l'assemblage des spliceosomes. Une absence de ces modifications engendrée par une perte de fonction de LARP7 tendrait alors vers une baisse de la fidélité et de l'efficacité de l'épissage des pré-ARNm (Bohnsack & Sloan, 2018). Ce phénomène pourrait alors expliquer les symptômes observés lors du syndrome d'Alazami (Hasler *et al*, 2020; Wang *et al*, 2020).

II. 3. 2. 3. Le syndrome de Prader-Willi

Le syndrome de Prader-Willi (PWS) est une rare maladie génétique qui a été décrite pour la première fois en 1956. Elle touche environ 1/15000 nouveau-né, induit de nombreux symptômes, et est actuellement incurable (Chung *et al*, 2020). Dès la naissance, le syndrome de Prader-Willi se caractérise par une sévère hypotonie et des difficultés à s'alimenter. Au cours de l'enfance, les patients développent une hyperphagie, amenant généralement à une obésité morbide si non contrôlée. Ils sont également atteints d'un retard de croissance et du développement du langage, ainsi que d'une déficience cognitive. Les hommes atteints du symptôme de Prader-Willi sont atteints d'hypogonadisme, et les femmes d'une hypoplasie génitale, associés dans la plupart des cas à une perte de la fertilité. Les individus touchés développent aussi des troubles du sommeil et des diabètes de type II (Cassidy *et al*, 2012).

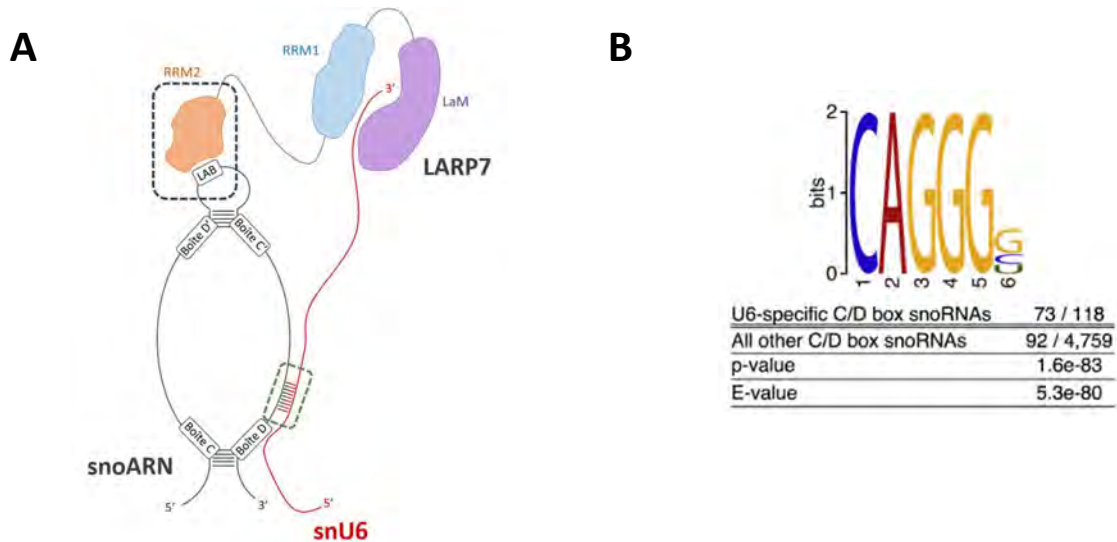


Figure 16 : Interaction entre LARP7, snU6 et les snoARN modifiant snU6

(A) Représentation schématique de snU6 et des snoARN le modifiant en interaction avec la protéine LARP7. Les interactions entre snU6/snoARN et LARP7/snoARN sont encadrées respectivement en vert et en bleu. LARP7 est composée de trois domaines : LaM (violet), RRM1 (bleu) et RRM2 (orange). LARP7 s'associe au snoARN via la région C-terminale du domaine RRM2 et à snU6 par le module La, composé des domaines LaM et RRM1. Adaptée de (Hasler *et al*, 2020). (B) Motif, nommé boîte LAB, fortement enrichi chez les snoARN interagissant avec l'ARN de l'épissage U6. Tirée de (Hasler *et al*, 2020).

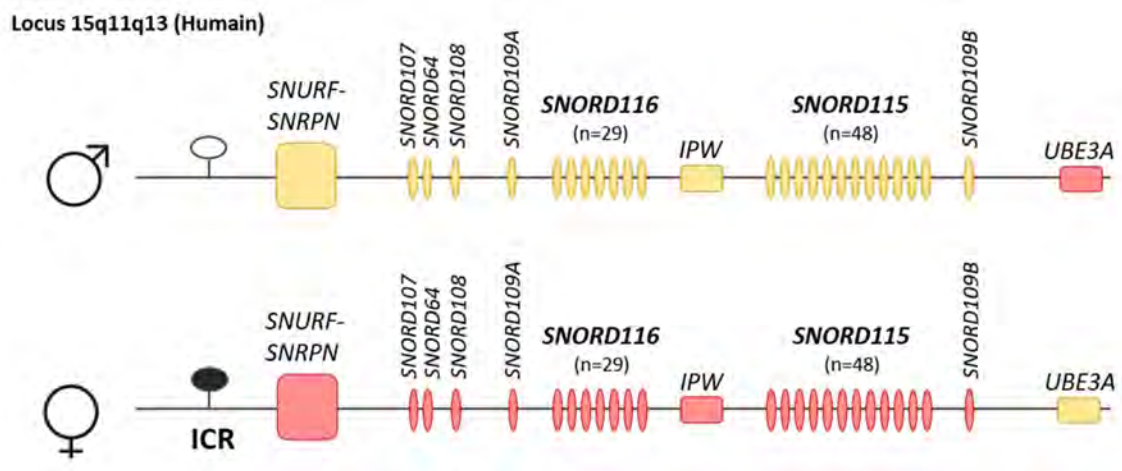


Figure 17 : Le domaine SNURF-SNRPN chez l'humain

Les gènes exprimés sont représentés en jaune, et ceux qui sont silencieux en rouge. Une méthylation de la région du centre de l'empreinte (ICR) engendre une expression mono-allélique du locus correspondant. Au niveau du domaine SNURF-SNRPN, l'ICR situé sur le chromosome hérité de la mère (♀) est méthylé, contrairement au chromosome issu du père (♂). Adaptée de (Cavaillé, 2017).

Le syndrome PWS est associé à la perte du domaine génomique SNURF-SNRPN, également nommé domaine du syndrome PWS, situé sur la région chromosomique 15q11q13 soumis à l’empreinte parentale (**Figure 17**). L’empreinte génomique parentale est un mécanisme épigénétique amenant à l’expression d’un seul allèle en fonction de l’origine parentale des allèles. Le domaine SNURF-SNRPN contient notamment les séquences codantes des ARN SNORD115 (HBII-52) et SNORD116 (HBII-85) (Cavaillé, 2017). Fait intéressant, les séquences de ces deux snoARN sont présentes en de nombreuses copies, qui possèdent de très fortes homologies. SNORD116 est par exemple présent en 29 copies, et SNORD115 en 48. L’expression des SNORD115 et 116 est restreinte aux allèles du chromosome paternel (Cavaillé *et al*, 2000). La forte expression de ces deux snoARN dans les différentes régions du cerveau, ainsi que la perte du domaine SNURF-SNRPN chez les patients développant le syndrome PWS, mettent en avant les implications potentielles des SNORD115 et SNORD116 dans le syndrome de Prader-Willi. De plus, certains patients développent le syndrome PWS en n’ayant seulement des micro-délétions du domaine SNURF-SNRPN, et ces délétions contiennent à chaque fois les séquences du SNORD116 (Sahoo *et al*, 2008; de Smith *et al*, 2009; Duker *et al*, 2010; Tan *et al*, 2020).

Sachant cela, les SNORD115 et SNORD116 sont deux snoARN C/D très activement étudiés par la communauté scientifique. Cependant, même si plusieurs pistes ont été suggérées, leurs cibles, ainsi que les mécanismes associés, restent mal connus. SNORD115/116 sont alors considérés comme orphelins.

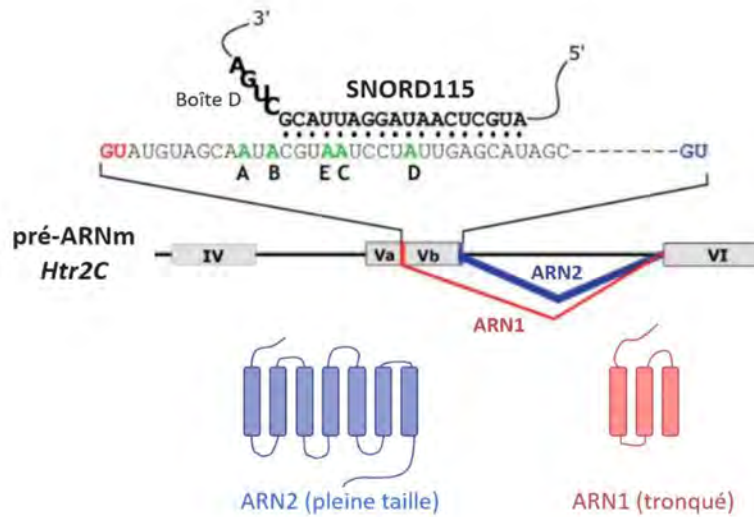
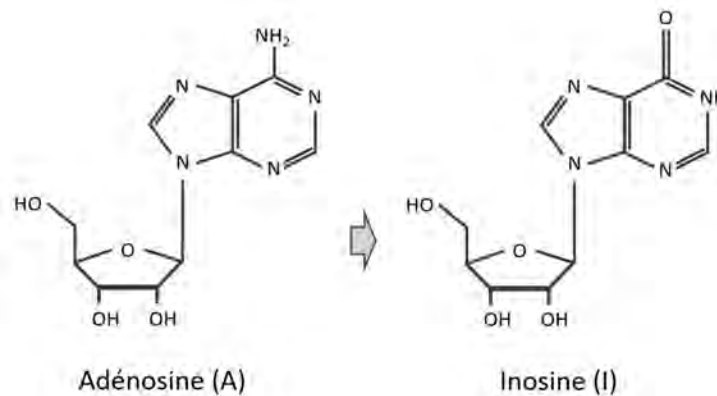
SNORD115 est un ARN d’une longueur de 82 nucléotides, et la délétion de l’ensemble de ses copies du domaine SNURF-SNRPN n’est pas suffisante pour engendrer le syndrome PWS (Runte *et al*, 2005). De manière intéressante, SNORD115 possède une complémentarité parfaite de 18 nucléotides avec l’exon V de l’ARNm du récepteur à la sérotonine 2C (5-HT_{2C}) (Cavaillé *et al*, 2000). Le récepteur 5-HT_{2C} est une protéine transmembranaire, couplée aux protéines G, se liant au neurotransmetteur qu’est la sérotonine. Les protéines G sont une famille de protéines qui permettent la transduction du signal lors d’activation de voies cellulaires (Duc *et al*, 2015). Le récepteur à la sérotonine 2C est particulièrement abondant dans les cellules épithéliales situées au niveau des plexus choroïdes, structures dans lesquelles le liquide cébrospinal est sécrété, chez l’homme et la souris (Abramowski & Staufenbiel, 1995). Les voies de signalisations des récepteurs 5-HT_{2C} régulent de nombreux mécanismes impliqués dans la satiété, les cycles du sommeil, l’anxiété ou les troubles neurodégénératifs (De Deurwaerdère *et al*, 2013; Palacios *et al*, 2017).

Le pré-ARNm du récepteur 5-HT2C est soumis à l'épissage alternatif. Notamment, l'épissage alternatif de l'exon V amène à la synthèse d'un isoforme non-fonctionnelle du récepteur à la sérotonine 2C, dépourvu de la seconde boucle interne ainsi qu'un quatrième domaine incomplet (**Figure 18A**) (Raabe *et al*, 2019). Également, le pré-ARNm du récepteur 5-HT2C est soumis au mécanisme d'édition. L'édition est un mécanisme post-transcriptionnel qui permet, entre autres, de modifier une base adénosine A en une base inosine I. L'inosine, résultant d'une désamination de l'adénosine (**Figure 18B**), est reconnue comme une guanosine lors de processus comme l'épissage ou la traduction, et peut former des appariements Watson-Crick avec les cytosines. La modification des adénosines en inosines permet donc de modifier l'information génétique des ARNm et d'avoir un impact sur la traduction (Srinivasan *et al*, 2021). ADAR1 et ADAR2, sont, chez les mammifères, les enzymes responsables de la désamination des adénosines en inosines. Le récepteur à la sérotonine 2C possède cinq sites, nommés A, B, C, D et E, au niveau des quelles les adénosines subissent une modification en inosines. Notamment, trois de ces sites, C, D et E, sont situées sur la zone de complémentarité entre SNORD115. Le site C se retrouve quant à lui exactement au niveau du cinquième nucléotide en amont de la boîte D lors de l'appariement de l'exon V de l'ARNm avec le SNORD115 (Hebras *et al*, 2020).

L'hypothèse proposée pour le rôle du SNORD115 est qu'il guiderait une méthylation en 2'-O-ribose au niveau du site C de l'édition sur l'exon V de l'ARNm du récepteur 5-HT2C. Cette méthylation inhiberait le mécanisme d'édition, ce qui impacterait directement l'épissage de l'ARNm du récepteur 5-HT2C, notamment en favorisant l'inclusion de l'exon V. Ce processus aboutirait à la synthèse d'un récepteur à la sérotonine 2C fonctionnel. En revanche, la délétion du SNORD115 orienterait l'excision d'une partie de l'exon V lors de l'épissage, donnant un récepteur 5-HT2C peu ou non-fonctionnel, pouvant expliquer les symptômes observés chez les patients du syndrome de Prader-Willi (**Figure 18A**).

Allant dans ce sens, des études *in vivo* ont démontré que l'ARN SNORD115 réduisait l'édition d'ARN substrat nucléolaire portant la séquence de l'ARNm 5-HT2C (Vitali *et al*, 2005), et qu'il orientait l'inclusion de l'exon V lors de l'épissage (Kishore & Stamm, 2006). Cependant, une étude *in vivo* chez des souris KO SNORD115 a montré que la perte du SNORD115 ne semble pas avoir d'effet sur l'épissage du pré-ARNm du récepteur 5-HT2C, même si un léger changement au niveau de l'édition a été observé (Hebras *et al*, 2020).

SNORD116 est un snoARN C/D de 97-98nt, qui semble être fortement impliqué dans le syndrome PWS. Comme vu précédemment, SNORD116 est absent dans la globalité des

A**B****Figure 18 : Epissage alternatif du pré-ARNm *Htr2c***

(A) Illustration de l'interaction prédite entre SNORD115 et l'exon V du pré-ARNm du récepteur à la sérotonine 2C. Les sites d'édition A → I, nommés A, B, E, C, D apparaissent en vert. L'épissage alternatif du pré-ARNm *Htr2C* donne lieu soit à l'ARN1 (rouge) aboutissant à un récepteur tronqué, soit à l'ARN2 pleine taille, formant un récepteur fonctionnel. Adaptée de (Hebras *et al.*, 2020). (B) Schéma de la désamination d'une adénosine en inosine, réaction catalysée par l'enzyme ADAR (adenosine deaminases that act on RNA).

patients atteints du symptôme, même ceux possédant seulement des micros-délétions du domaine SNURF-SNRPM (Sahoo *et al*, 2008; de Smith *et al*, 2009; Duker *et al*, 2010; Tan *et al*, 2020). Dans cette optique, différents KO des copies du SNORD116 chez la souris ont induit de nombreux symptômes associés au syndrome de Prader-Willi (Polex-Wolf *et al*, 2018; Adhikari *et al*, 2019). De plus, une déplétion ou une surexpression de SNORD116 dans des cellules engendrent une dérégulation de nombreux gènes (Falaleeva *et al*, 2015; Pace *et al*, 2020). Récemment, une étude a mis en lumière de potentiels appariements conservés de SNORD116 sur les ARNm Dgkk, Nlgn3 et Rsbn11 (Baldini *et al*, 2021b). Une baisse de l'expression de SNORD116 dans des cellules humaines a engendré un changement significatif de l'expression de ces trois ARNm, dont deux d'entre eux (Dgkk et Nlgn3) sont impliqués dans des mécanismes cérébraux. Cependant, il reste encore à valider les interactions physiques entre SNORD116 et ces ARNm.

L'étude SNORD115/116 ainsi que la détection de leurs cibles *in vivo* permettrait de mieux comprendre le syndrome de Prader-Willi, ainsi que les mécanismes et fonctions associés aux snoARN orphelins.

II. 3. 2. 4. *SnoARN et Cancers*

Lors de développement de cancers, il a été remarqué que l'expression des snoARN subissait de grandes variations. La grande majorité des niveaux d'expressions des snoARN dépendant du taux de transcription de leur gène hôte, les expressions des snoARN introniques se retrouvent logiquement impactées lors d'un dérèglement de l'expression génique et transcriptomique des cellules.

De manière intéressante, les changements des niveaux d'expressions des snoARN dans les cancers semblent s'effectuer dans la plupart des cas dès les stades précoces de ces maladies. De part de leur stabilité et de la possibilité de mesurer leur taux facilement chez les patients, les snoARN apparaissent désormais comme étant des biomarqueurs attrayant et potentiellement efficaces dans le dépistage des stades précoces des cancers (Liao *et al*, 2010).

Dans cette optique, le SNORD88C montre une forte dérégulation de son expression dès les stades I et II des cancers du poumon non à petites cellules (NSCLC). Les cancers NSCLC représentent plus de 80% des cancers des poumons et sont l'un des cancer les plus mortels dans le monde (Bray *et al*, 2018). L'efficacité du diagnostic basé sur l'expression de l'ARN SNORD88C s'est révélé être moins efficace que le dépistage basé sur la quantification de l'antigène carcinoembryonnaire (ACE), un biomarqueur traditionnellement utilisé pour dépister

les cancers NSCLC (Wang *et al*, 2023). Cependant, une utilisation en tandem de la quantification de l'ARN SNORD88C et de l'ACE montre un pouvoir de détection supérieur à l'utilisation seule de détection d'ACE. Cette étude met alors en avant le fait que les snoARN auraient un rôle potentiel à jouer dans le dépistage des stades précoces dans les cancers, en tant que biomarqueurs efficaces (Wang *et al*, 2023).

D'un autre côté, les modifications post-transcriptionnelles, notamment les méthylations, ont été reconnues comme étant des facteurs impliqués dans les mécanismes de régulations, dans la progression des cancers, et la formation de métastases (Yang *et al*, 2021). Également, suite aux nombreux indices qui tendent à montrer que les ARNm soient ciblés par les snoARN, il n'est alors pas illogique de se demander si les différences d'expression des snoARN observées lors de développement de cancer, ne seraient pas qu'une conséquence, mais potentiellement impliquées dans le développement de ces maladies.

En ce sens, le potentiel biomarqueur biologique SNORD88C semble être associé au développement des cancers NSCLC. SNORD88C est le snoARN dont l'augmentation de son expression est la plus présente et intense lors des cancers d'adénocarcinome pulmonaire, du carcinome épidermoïde pulmonaire, et de NSCLC (Wang *et al*, 2023). Une déplétion de SNORD88C dans des cellules de poumons tend vers une baisse significative de la prolifération, de l'invasion et de la migration cellulaire. A contrario, lors d'une surexpression de SNORD88C, ces trois mécanismes cellulaires se retrouvent être fortement accentués, caractéristiques d'un développement de cancer. De manière intéressante, SNORD88C dirige une méthylation en 2'-O-ribose sur l'ARNr 28S sur la cytosine 3680. Les auteurs ont montré que cette méthylation serait impliquée dans la régulation positive de la traduction de la protéine SCD1, une enzyme catalysant l'étape limitant la vitesse de formation des acides gras monoinsaturés (AGMI) (Wang *et al*, 2015). La hausse de la production de la protéine SCP1 engendre alors une inhibition du mécanisme d'autophagie de la cellule, favorisant ainsi l'apparition du processus métastatique (Wang *et al*, 2023). De part ces observations, SNORD88C semble avoir le rôle d'un oncogène.

Dans le cadre des cancers, le snoARN H/ACA SNORA42 fait partie des snoARN les plus étudiés. L'ARN SNORA42 est issu de la maturation de l'intron 7 du gène KIAA090. Ce gène est localisé dans la région chromosomique (1q22), connue pour être fortement amplifiée lors de développement de cancer NSCLC (Mannoor *et al*, 2014). Également, l'observation de cas cliniques montrant qu'une forte expression de SNORA42 dans les cellules est associée à un taux faible de survie des patients. Une déplétion de cet ARN dans les cellules a engendré une inhibition de la formation des tumeurs. Les auteurs ont observé une régulation négative de la

synthèse des protéines p21 et p53, ainsi qu'un taux accru de production des cyclines E et D, ce qui pourrait expliquer les phénotypes observés lors du KO SNORA42 (Mei *et al*, 2012; Wang *et al*, 2021a). Cependant, les mécanismes reliant les différences d'expressions de SNORA42 et la voie de signalisation p53 n'ont pas encore été caractérisés. Lors d'étude *in silico*, SNORA42 est prédit pour diriger deux pseudouridylations sur l'ARN ribosomique 18S, au niveau des uridines 109 et 572 (Kiss *et al*, 2004). Même si ces deux interactions n'ont pas été vérifiées *in vivo*, la modification des ARNr pourraient expliquer les impacts observés au niveau la voie de signalisation p53 (Barbieri & Kouzarides, 2020).

D'après la base de données The Cancer Genome Atlas (TCGA), répertoriant de nombreuses données de séquençage de patients atteints de cancer, l'expression de SNORD89 est fortement augmentée lors de cancers de l'endomètre par rapport aux tissus sains. Dans les pays développés, le cancer de l'endomètre est la tumeur maligne gynécologique la plus fréquente, et le taux de survie est extrêmement faible lors d'un dépistage tardif (supérieur à 5 ans) (Abdalla *et al*, 2020; Makker *et al*, 2020). SNORD89 est un snoARN C/D situé dans la région chromosomique 2q11.2, également connu sous le nom HBII-289. Une surexpression de SNORD89 dans des cellules d'adénocarcinome de l'endomètre tend vers une augmentation de la prolifération cellulaire, une inhibition de l'apoptose, ainsi qu'une migration cellulaire accrue (Bao *et al*, 2022). Également, chez des souris « nude », la surexpression de SNORD89 engendre la formation de croissance tumorale. Après avoir démontré que la Fibrillarine faisait bien partie du complexe ribonucléoprotéique SNORD89, les auteurs ont mis en évidence une complémentarité de séquences entre les éléments antisens du snoARN et l'ARNm codant pour la protéine Bim. Une expérience RTL-p (méthode qui sera développée plus tard) a montré qu'une surexpression de SNORD89 réduit la quantité de la protéine Bim dans les cellules, avec une augmentation du taux de méthylation de l'ARNm Bim (Bao *et al*, 2022). Bim est une protéine appartenant au groupe pro-apoptotique de la famille Bcl-2, qui favorise la tumorigénèse dans différents cancers (Sur *et al*, 2019; Byerly *et al*, 2020; Tanaka *et al*, 2021). Par méthylation de l'ARNm de Bim, SNORD89 réduit le niveau d'expression de Bim, dérégulant les voies de signalisation en aval, inhibant ainsi l'apoptose. Fait remarquable, une étude en parallèle a relevé une surexpression du SNORD104 au sein du cancer de l'endomètre. La déplétion de SNORD104 engendre une baisse de l'invasion et de la migration cellulaire *in vitro*, tandis qu'une surexpression tend à augmenter le développement du cancer de l'endomètre *in vivo* (Lu *et al*, 2022). Par des méthodes d'analyses similaires, les auteurs ont démontré que SNORD104 jouerait sur le taux de méthylation de l'ARNm codant pour la protéine PARP1,

stabilisant ainsi cet ARNm et augmentant le nombre de protéine PARP1 produites. PARP1 est une protéine nucléaire impliquée dans la réparation de l'ADN et la régulation génique (Gupte *et al*, 2017).

En revanche, le rôle des snoARN dans les cancers ne s'apparenterait pas seulement à être des proto-oncogènes. Diverses études ont mis en lumière que certains snoARN pourraient incarner le rôle de gènes suppresseurs de tumeurs.

En ce sens, le gène SNORD50 est issu de la maturation intronique du cinquième intron du gène hôte SNHGS. SNORD50 est situé au niveau du chromosome 6, dans la région 6q14-15. De manière intéressante, les délétions de régions du chromosomes 6 sont généralement associées au développement de processus cancéreux (Orphanos *et al*, 1995; Dong, 2001). La délétion de la région 6q14-q22 est associée au cancer de la prostate, et similairement, la délétion de la région 6q14-27 est retrouvée chez les patients atteints de cancer du sein. Ces deux régions comprennent le gène SNORD50. Les délétions de régions chromosomiques sont des signes distinctifs de l'emplacement de gènes suppresseurs de tumeurs. Se révélant comme tel, SNORD50 est alors activement étudié dans le cadre de développement métastatiques (Dong *et al*, 2008). Diverses analyses fonctionnelles ont révélé que l'expression de SNORD50 réprime les mécanismes de survies et de proliférations cellulaires dans des lignées de cellules de cancer de la prostate. Également, la complémentation de SNORD50 dans des cellules de cancer du sein induit une diminution de la formation de colonies cellulaires. En supplément, SNORD50 voit son expression être drastiquement diminuée lors d'un développement du cancer du côlon, à l'instar des cancers de la prostate et du sein, et des analyses mettent en avant la diminution de son expression lors de la prolifération cellulaire (Dong *et al*, 2008, 2009; Pacilli *et al*, 2013). SNORD50 guide deux 2'-O-ribose méthylations sur l'ARNr 28S, au niveau des cytosines aux positions 2848 et 2863 (Kiss-László *et al*, 1996). En condition de prolifération cellulaire, les cellules présentent un faible taux de méthylation au niveau de la cytosine 2848. Une surexpression de SNORD50 engendre une augmentation du taux de méthylation de la cytosine C2848. La présence de cette modification sur les ribosomes engendre une préférence de traduction des site IRES au détriment des sites CAP par les ribosomes influençant ainsi globalement l'efficacité de la traduction des ARNm (Pacilli *et al*, 2013).

SNORD50 semble influencer donc sur la traduction dans les cellules, notamment via l'ajout d'une méthylation sur la cytosine C2848. Par ce phénomène, la réactivation du gène SNORD50 dans des lignées de cellules cancéreuses a tendance à freiner la prolifération cellulaire et la formation de métastases (Pacilli *et al*, 2013).

En supplément de son implication dans la maturation des ARNr, une fonction non canonique a été attribuée à l'ARN SNORA73. Lors d'un développement de leucémie aiguë myéloblastique (LAM), une forte régulation négative de l'expression de l'ARN SNORA73 est observée (Han *et al*, 2022). Il a premièrement été démontré que l'ARN SNORA73 est capable de se lier de manière stable à la chromatine (Schubert *et al*, 2012). Le snoARN SNORA73 possède une longue tige-boucle au niveau de son extrémité 5', qui s'associerait avec la protéine PARP1, tandis que la tige-boucle située en 3' du snoARN s'associe avec les protéines cœurs canoniques (Kim *et al*, 2019; Han *et al*, 2022). Une déplétion de SNORA73 entraîne la libération de PARP1 qui engendre une réactivité augmentée suite à des dommages à l'ADN. Cette facilitation de réponse au stress en cas d'endommagement de l'ADN initié par l'activité de PARP1 induit un blocage de la différenciation cellulaire et confère une stabilité génomique plus importante aux cellules cancéreuses.

Suite aux fortes dérégulations des snoARN lors de processus cancéreux, leur implication au sein de ces phénomènes ont alors été largement étudiées. Les snoARN apparaissent être des biomarqueurs biologiques intéressants, suite à leur forte dérégulation dès les stades précoces de ces maladies. Outre cette caractéristique, les changements d'expressions des snoARN ne s'apparentent pas seulement être une conséquence des développements cancéreux. En effet, certains snoARN apparaissent comme étant des proto-oncogènes ou des supresseurs de tumeurs. Même s'ils sont reliés à des modifications des ARN ribosomiques, les mécanismes moléculaires précis reliant les différences d'expressions des snoARN aux phénotypes observés sont encore très mal compris.

II. 3. 3. Fonctions non-canoniques associés aux snoARN

II. 3. 3. 1. SnoARN et ARNm

Une méthode basée sur le crosslink ARN/ARN, nommée le LIGR-Seq, a permis en 2016 de dévoiler des interactions physiques entre des snoARN et d'autres ARN cellulaires (Sharma *et al*, 2016). Outre des interactions déjà répertoriées qui valident l'approche technique utilisée, cette étude a mis en avant l'interaction entre le snoARN SNORD83B et les ARNm NOP14, RPS5 et SRSF3. Les potentielles modifications chimiques que pourraient impliquer ces interactions n'ont pas été démontrées, mais une hausse de l'expression de ces trois ARNm a été observée lors du retrait de SNORD83B dans les cellules humaines (Sharma *et al*, 2016).

L'ARN SNORD27 possède un élément antisens ciblant l'ARNr 18S, et est prédit pour guider la 2'-O-méthylation sur l'adénine₂₇ (Tycowski *et al*, 1996). Cependant, l'ARN

SNORD27 a été retrouvé dans des fractions nucléoplasmiques dans lesquelles la Fibrillarine était absente lors de séparation de fraction par un gradient de glycérol. Ce fut le cas pour d'autres snoARN ciblant les ARNr, comme SNORD2, SNORD7, et SNORD60 (Falaleeva *et al*, 2016). La formation de snoRNP dans lesquelles la protéine cœur Fibrillarine est absente pourrait associer ces snoARN C/D à de nouvelles fonctions non-canoniques. SNORD27 possède par ailleurs une complémentarité de 29 nucléotides, contenant les boîtes C' et D', avec le pré-ARNm du facteur de transcription E2F7. Cela renforce l'idée d'une fonction non canonique associée à SNORD27, les boîtes C' et D' étant nécessaires pour le recrutement des protéines cœurs (Marmier-Gourrier *et al*, 2003). En ce sens, aucune méthylation en 2'-O-ribose n'est détectée sur le pré-ARNm E2F7. Un Knock-Out (KO) du gène SNORD27 influence l'épissage alternatif de quatre pré-ARNm (Falaleeva *et al*, 2016). Même si les modes d'actions mécanistiques restent inconnus, SNORD27 pourrait influencer la régulation de l'épissage alternatif de certains gènes en se liant physiquement au pré-ARNm E2F7.

SNORD86 pourrait être quant à lui impliqué dans la régulation de la quantité des snoRNP C/D de la cellule. Le gène du SNORD86 est situé dans une région intronique du gène codant pour la protéine NOP56, composante essentielle des snoRNP C/D (Lykke-Andersen *et al.*, 2018). L'épissage alternatif du gène NOP56 peut donner lieu soit à : (i) un ARNm codant pour la protéine NOP56, (ii) un ARN contenant des exons et la partie intronique hébergeant SNORD86, nommé SNORD86-cSPA. Lors de l'exportation dans le cytoplasme de SNORD86-cSPA, il est ciblé par la voie de dégradation NMD (Nonsense-Mediated Decay) (Lykke-Andersen *et al*, 2014). La NMD va dégrader l'extrémité 5' l'ARN, mais seulement jusqu'à la région protégée par l'assemblage des protéines cœurs au niveau de l'ARN SNORD86. Le long ARN non codant SNORD86-cSPA est alors stabilisé, de par la présence de l'assemblage des protéines cœurs en 5', et d'une queue polyA en 3'. La fixation des protéines cœurs sur SNORD86 au niveau du pré-ARNm influence l'épissage de manière à prioriser la voie donnant le SNORD86-cSPA. L'interaction des facteurs de l'épissage avec la snoRNP SNORD86 pourrait également être impliquée dans ce phénomène (Lykke-Andersen *et al.*, 2018). La stabilisation de l'ARN SNORD86-cSPA dans le cytoplasme peut être un moyen de séquestrer l'excès des protéines cœurs des snoRNP C/D (**Figure 19**).

SNORD50A est un snoARN guidant deux méthylations en 2'-O-ribose sur l'ARNr 28S (Rebane *et al*, 2002). En supplément de ces fonctions canoniques, des études ont révélé que SNORD50A serait également impliqué dans la maturation des extrémités 3' des ARNm (Huang *et al*, 2017). En effet, la déplétion de cette snoRNP C/D dans les cellules humaines induit des

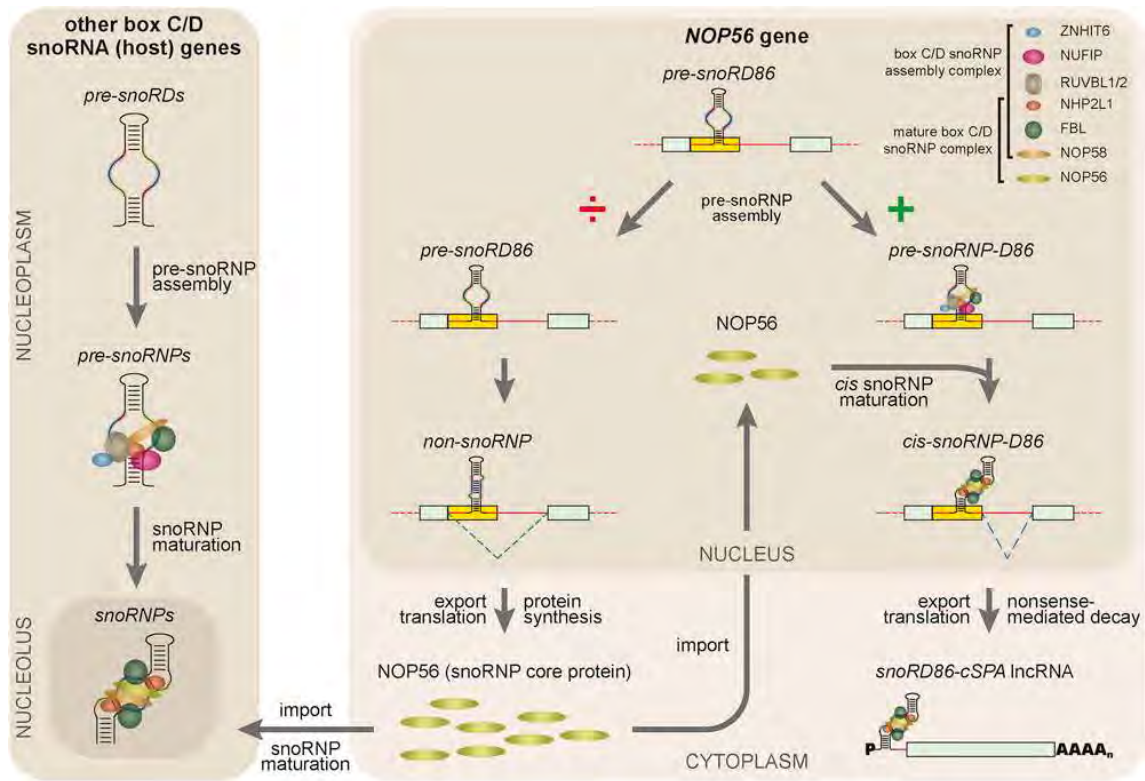


Figure 19 : Rôle de SNORD86 dans la régulation de l'expression de son gène hôte

SNORD86 est situé dans un intron de l'ARNm codant pour la protéine cœur des snoARN C/D NOP56. Le mécanisme d'épissage du pré-ARNm NOP56 peut engendrer deux ARN différents : (i) dans le cas où il n'y a pas d'assemblage de la snoRNP SNORD56, l'ARN intronique SNORD86 est excisé, et l'ARNm NOP56 est produit (ii) si la snoRNP SNORD56 est assemblée, l'intron contenant SNORD56 est retenu lors de l'excision-épissage du pré-ARNm NOP56, aboutissant à la synthèse d'un long transcrit, protégé en 5' par l'assemblage de la snoRNP SNORD86 et en 3' par une queue poly A. Ce long ARN, nommé SNORD86-cSPA est exporté dans le cytoplasme. Cette forme pourrait permettre de séquestrer les protéines cœurs des snoARN afin de réguler leur concentration cellulaire. Également, lors d'un déficit de NOP56 dans la cellule, la voie (i) sera priorisée. Issue de Lykke-Anderson *et al.*, 2018.

changements d'expressions de plusieurs gènes ainsi que leur niveau de polyadénylation. SNORD50A est capable de se lier avec le complexe CPSF (Cleavage and Polyadenylation Specificity Factor) en interagissant directement avec la protéine membre Fip1. Le complexe CPSF, en s'appariant sur les sites poly(A) des pré-ARNm (PAS) via Fip1, est impliqué dans le clivage et la polyadénylation des pré-ARNm. L'interaction entre SNORD50A et Fip1 engendre alors une compétition avec les sites (PAS), mécanisme pouvant affecter la maturation des extrémités 3' polyadénylées (Huang *et al*, 2017).

Le SNORD113-6 et AF357425, son homologue murin, sont des ARN orphelins impliqués dans la régulation d'ARNm à la fois par les voies canoniques et non canoniques des snoRNP aux motifs C/D (van Ingen *et al*, 2022). SNORD113-6 est situé sur le locus chromosomique 14q32 (12F1 chez la souris), nommé DLCK1-DIO3 et connu pour être relié aux maladies cardiovasculaires chez l'humain (Håkansson *et al*, 2019). Ce locus contient un cluster de 41 snoARN C/D, dont 9 copies de SNORD113. Les SNORD de ces régions sont considérés comme snoARN orphelins (Cavaillé *et al*, 2002). SNORD113-6 fait partie des snoARN les plus exprimés du locus 14q32, ne possède aucune complémentarité avec les ARNr, et contient une boîte C' non dégénérée (Jorjani *et al*, 2016). SNORD113-6 guide des 2'-O-méthylations sur les sites prédits des ARNm DUSP7 et EBPL, tandis que AF357425 guide ces modifications chimiques sur DUSP7 et Jag1 (van Ingen *et al*, 2022). La déplétion de AF357425 étant létale, les inhibitions partielles de ces snoRNP montrent une augmentation de la dégradation des ARNm ciblés par les méthylations en 2'-O-ribose. ITGB3 est un ARNm possédant un site putatif d'interaction avec SNORD113-6, mais aucune méthylation en 2'-O-ribose n'a été observée. Cependant, en l'absence du snoARN, il est observé que la balance entre les deux variants de cet ARNm est dérégulée, donnant l'avantage au transcrit ciblé par la NMD. Les SNORD113-6 et AF357425 ciblent donc de nombreux ARNm via deux mécanismes différents, qui sont l'épissage des pré-ARNm et les méthylations en 2'-O-ribose, augmentant la stabilité des ARNm cibles dans la cellule (van Ingen *et al*, 2022).

Autre exemple intéressant sur les potentiels rôles des modifications des snoARN sur les ARNm, SNORD32A et SNORD51, guideraient des modifications sur l'ARNr 28S ainsi qu'au niveau du site A3150 sur l'ARNm codant pour la protéine PXDN. Les nucléotides 2-O-méthylés augmenteraient la stabilité des ARNm et auraient un effet sur l'efficacité de leur traduction par les ribosomes. En effet, lors du KO de ces snoARN, le taux de méthylation de A3150 diminue, entraînant une baisse de quantité de l'ARNm Pxdn, mais en revanche une augmentation de la quantité de la protéine PXDN (Elliott *et al*, 2019). Il est prédit que la

méthylation au niveau du codon contenant A3150 pourrait alors altérer la traduction de cette protéine en bloquant les interactions stériques entre l'ARNr du ribosome et le ribo-squelette ARNt-ARNm (Choi *et al*, 2018; Hoernes *et al*, 2019).

Au vu des précédentes études, de nombreux indices semblent indiquer que les éléments antisens d'une partie des snoARN possèdent des complémentarités avec des ARNm. Cependant, aucune interaction physique en condition endogène n'a clairement été démontrée. De par son expression élevée dans les tissus cérébraux de souris, ainsi que par une complémentarité parfaite avec l'ARNm du récepteur à la sérotonine 2C, SNORD115 demeure aujourd'hui l'exemple le plus convaincant pour l'étude des interactions snoARN-ARNm. Néanmoins, l'interaction physique SNORD115-mARN 5-HT2C demeure toujours théorique.

II. 3. 3. 2. *SnoARN et virus*

A l'instar du cancer, l'infection virale des cellules entraîne une modification d'expression du transcriptome. Depuis une dizaine d'année, il est acquis que l'expression des snoARN se retrouve dérégulée lors de l'infection virale, mais les mécanismes les concernant n'ont pas été caractérisés (Saxena *et al*, 2013). Des données ont par ailleurs montré qu'une hausse de l'expression de SNORD126 faciliterait l'entrée du virus de l'hépatite C dans les cellules hôtes, en jouant sur la régulation de la voie PI3K-AKT (Qian *et al*, 2020).

Fait intéressant, un snoARN C/D encodé par le virus Epstein-Barr (EBV), nommée v-snoRNA1, a été découvert lors de l'infection de lymphocytes B (Hutzinger *et al*, 2009). 9 transcrits de 24 nucléotides (nt), v-snoRNA1^{24pp}, parfaitement identiques à la séquence de l'extrémité 3' de v-snoRNA1 ont également été trouvés lors de l'analyse du génome d'Epstein-Barr. Ce snoARN viral est conservé dans la totalité des souches du virus EBV, avec 100% d'identité. Chez un virus (*rhesus lymphocryptovirus*) possédant 65% d'identité génomique avec EBV, un snoARN possède 86% d'identité avec v-snoARN1. L'analyse *in silico* des cibles potentielles des éléments antisens de v-snoRNA1 a souligné 2 sites putatifs de méthylations sur l'ARNr 18S, et 23 sites sur l'ARNr 28S. Ces complémentarités concernent uniquement l'élément antisens situé en amont de la boîte D' de v-snoRNA1. Aucune cible potentielle n'a été trouvée en utilisant celui en amont de la boîte D (Hutzinger *et al*, 2009). De manière intéressante, les auteurs ont montré par 5' RACE PCR que les v-snoRNA1^{24pp} cliveraient l'ARNm BALF5, codant pour la sous-unité catalytique de l'ADN polymérase de l'EBV, en ayant un rôle analogue aux microARN (miARN).

II. 3. 3. 3. *Formes alternatives de maturation des snoARN*

La présence de fragments issus de snoARN, tel que les v-snoRNA1^{24pp}, n'est pas un phénomène spécifiquement retrouvé chez les virus. En effet, dans les cellules de mammifères, de nombreux fragments issus de snoARN sont retrouvés, et ne sont pas classifiés comme étant des produits de dégradation des snoARN. Ces fragments, nommés sdARN (snoRNA-Derived RNAs) sont catégorisés en fonction de leurs longueurs et de leurs origines. Ceux issus de snoARN H/ACA ont une taille comprise entre 20 et 24nt de long, et sont généralement issus de la région 3'. Les sdARN générés à partir des snoARN C/D font soit 17 à 19 nt de long, soit 26nt, et résultent principalement de la partie 5' (Taft *et al*, 2009; Wajahat *et al*, 2021). Ce mécanisme est assez fréquent, vu qu'il concernerait plus de la moitié des snoARN (Falaleeva & Stamm, 2013). Cependant, les rôles et fonctions qui leur sont associés sont pauvrement caractérisés.

Une des principales pistes qui en est ressortie est que ces snARN auraient une fonction analogue aux miARN et au piARN, à l'instar de v-snoARN1^{24pp}, en exerçant une fonction de régulation négative de l'expression génique. Les miARN matures sont des ARN non codants d'une vingtaine de nucléotides, qui après association avec une protéine argonaute, induit l'ensemble d'un complexe nucléoprotéique nommé RISC. Au sein de ce complexe, les miARN sont guidés vers l'extrémité 3' de leur ARNm cible auquel il s'apparie directement par paire de bases. Chez l'humain, cet appariement est généralement imparfait. L'interaction du miARN au sein du complexe RISC et un ARNm induit la déstabilisation de l'ARNm et régule ainsi négativement sa traduction (Saliminejad *et al*, 2019). Les piARN (PIWI-interacting RNAs) sont également une famille d'ARN non-codants, d'une longueur variant entre 24 et 31nt. Les piARN matures interagissent avec les protéines PIWI, une sous classe protéique des protéines argonautes, différente de celle contenant les protéines AGO1 et AGO2, liées aux fonctions des miARN (Liu *et al*, 2019; Santos *et al*, 2023). Les piARN sont majoritairement exprimés dans les cellules germinales chez les mammifères, et ont pour fonctions de participer à la dégradation de l'ARN des transposons ainsi qu'à la régulation épigénétique de l'expression génique (Zhong *et al*, 2015).

Les sdARN ont notamment été retrouvés lors de la purification de la protéine argonaute 2 (AGO2), protéine nécessaire à l'action des miARN (Ender *et al*, 2008). Les auteurs ont mis en évidence qu'un fragment dérivé du scaARN 15 (ACA45) était capable de cibler l'ARNm CDC2L6, régulant alors négativement son expression. Ce mécanisme nécessite, à l'instar des miARN la protéine DICER. Cependant, de manière étonnante, la protéine DROSHA, également

impliquée dans la biogenèse des miARN, ne semble pas être requise pour la production du sdARN issu du scaARN 15 (Ender *et al*, 2008; Taft *et al*, 2009). D'un autre côté, des fragments issus des snoARN SNORD63 et SNORD75 s'apparentent à des piARN. Le sdARN issu de SNORD63 a été détecté lors d'une étude sur les lymphocytes T primaires. Le sdARN, lié aux protéines PIWI, s'apparie au niveau d'une région intronique du pré-ARNm de l'interleukine -4 (IL-4). Cette liaison est impliquée dans le recrutement du complexe TRAMP, provoquant la dégradation de l'ARNm IL-4 (Zhong *et al*, 2015). Le fragment de SNORD75 se lie aux protéines PIWI, PIWIL1 et PIWIL4. Les auteurs ont démontré que cette interaction faciliterait le changement de marques répressives sur les histones (H3K27me3) en marques actives (H3K4me3), régulant ainsi positivement l'expression du gène suppresseur de tumeur TRAIL (He *et al*, 2015).

D'autres fragments, issus de scaARN, ont également été détectés. Leurs tailles et leurs fonctions différant des sdARN, ces fragments ont été nommées regRNP (Burke *et al*, 2018). Quatre regRNP, appelées individuellement mgU2-61, mgU2-19, mgU2-30 et mgU4-8 sont issus respectivement des scaARN 2, 9, 9 et 17. Il est proposé que les regRNP, en interagissant avec des snoRNP classiques, régulent les niveaux de méthylations catalysés par ces snoRNP. En ce sens, mgU4-8 possède une région de son extrémité 3' qui s'associe avec le SNORD16, connu pour guider une méthylation sur l'adénosine⁴⁸⁴ de l'ARNr 18S. Lors d'une baisse d'expression de scaARN17, dont est issu le fragment mgU4-8, une augmentation du taux de méthylation est observée au niveau du site A484 de l'ARN18S. *In vitro*, les auteurs ont démontré que mgU4-8 perturbe l'appariement de SNORD16 sur l'ARNr 18S (Burke *et al*, 2018).

Les snoARN peuvent être aussi présents sous une forme de long ARN non codants (sno-lncRNA). Ce phénomène apparaît lorsque qu'une snoRNP est assemblée avec un ARN qui possède une de ses extrémités protégées, comme une coiffe en 5' ou une queue polyA en 3'. De cette manière l'ARN sera d'un côté protégé par les conditions précédemment décrites et l'autre extrémité par l'assemblage des protéines cœurs sur le snoARN, donnant ainsi des longs ARN non codants. C'est le cas de SNORD86-cSPA, comme vu précédemment (Lykke-Andersen *et al.*, 2018), ou de SNORD116, suite à l'épissage des introns. Cet ARN reste par ailleurs stable dans le noyau (Vitali *et al*, 2010). Ces ARN semblent jouer un rôle dans la régulation de l'épissage alternatif.

III. Techniques d'études des cibles et modifications des snoRNP

III. 1. Méthodes de détection des modifications post-transcriptionnelles guidées par les snoRNP

Les snoARN interagissent physiquement avec leurs ARN cibles par appariement direct de paires de bases. Comme vu précédemment, l'appariement se fait via l'élément antisens du snoARN et les régions de l'ARN substrat encadrant l'uridine à isomériser par les snoRNP H/ACA. Également, lors d'une méthylation en 2'-O-ribose, l'élément antisens des snoARN C/D s'apparie avec la région contenant le nucléoside à modifier. Dans cette optique, la connaissance de l'emplacement dans le transcriptome, au nucléotide près, de ces nucléotides modifiés permet de déterminer et de délimiter les séquences cibles avec lesquelles les éléments antisens des snoARN pourraient être complémentaires, et s'y hybrider. Sachant cela, le développement de méthodes cartographiant les positions de modifications chimiques sur le transcriptome a constitué, et constitue, un atout majeur dans l'identification des snoARN associés à ces modifications post transcriptionnelles.

III. 1. 1. Méthodes de détection des pseudouridines

Après la découverte de la pseudouridine en 1957 via l'analyse de fractions de nucléosides par HPLC (Davis & Allen, 1957), différentes méthodes ont émergé pour étudier ce « cinquième nucléotide ». Malgré leurs abondances au sein des ARN, la détection des positions des pseudouridines s'est révélée être une tâche relativement difficile. En effet, aucun groupement caractéristique n'est rajouté lors de cette modification, qui consiste à la rotation d'un groupement chimique. Il est donc, par exemple, impossible de marquer cette modification radioactivement. En supplément, sa masse est sensiblement équivalente à celle de l'uridine (Li *et al*, 2017). Les premières techniques qui ont été mises au point permettaient d'étudier les pseudouridines seulement à un endroit précis du transcriptome. Une modernisation de ces méthodes a par la suite eu lieu, rendant possible l'étude des pseudouridines à l'échelle de l'ensemble du transcriptome.

III. 1. 1. 1. La méthode au CMCT

Après la découverte des pseudouridines, les techniques utilisées pour étudier ces modifications impliquaient généralement des digestions par nucléases et des chromatographies. Ces techniques, longues et laborieuses, ont rapidement disparu grâce à la découverte, en 1993,

de la méthode au CMCT (*N*-cyclohexyl-*N'*- β -(4-methylmorpholinium) ethylcarbodiimide p-tosylate) (Bakin & Ofengand, 1993). Le principe de cette méthode repose sur le couplage du CMCT (**Figure 20A**), un carbodiimide soluble dans l'eau, avec une extension d'amorces. Dans un milieu aqueux pH8-8.5, la partie carbodiimide (CMC) du CMCT réagit avec les guanosines, les uridines et les pseudouridines des molécules d'ARN, via une réaction d'acylation. Le CMC se lie via les atomes d'azotes en position 1 et 3 des guanosines et uridines, respectivement. A contrario, une molécule de CMC interagit avec l'azote en position 1 de la pseudouridine, et une seconde molécule se lie à l'azote en position 3 (**Figure 20C**) (Ho & Gilham, 1971). Les molécules d'ARN traitées au CMCT sont placées ensuite dans une solution alcaline (pH 10.4), engendrant un clivage des liaisons CMC-guanosines et CMC-uridines. La molécule de CMC située sur l'azote 1 des pseudouridines est également clivée. En revanche, ce n'est pas le cas de la molécule de CMC placée au niveau de l'azote 3, cette liaison n'étant pas réversible. La molécule de CMC restante, présente uniquement sur les pseudouridines, va alors permettre l'identification de cette modification post-transcriptionnelle (Ho & Gilham, 1971; Bakin & Ofengand, 1993; Bakin *et al*, 1994).

Pour cela, la seconde étape se traduit par la synthèse d'ADN complémentaire (ADNc) par une étape de transcription inverse à partir des ARN traités au CMCT. La polymérisation de la transcriptase inverse se retrouve bloquée lorsqu'elle rencontre une pseudouridine couplée au CMC. Ces produits de terminaisons précoces sont analysés par électrophorèse sur gel. En comparant avec un profil de gel issu d'une réaction sans l'ajout de CMCT, il est possible de déterminer les positions sur l'ARN où la polymérase s'est retrouvée bloquée par le CMC, ce qui traduit donc la présence de pseudouridines (Bakin & Ofengand, 1993; Bakin *et al*, 1994).

Malgré l'efficacité et la facilité de mise en œuvre de cette méthode, elle présente toutefois quelques limites. En effet, la liaison du CMC aux pseudouridines, ainsi que la suppression des résidus U et G n'ont pas une efficacité parfaite, ce qui peut engendrer des faux négatifs et faux positifs, respectivement (Zaringhalam & Papavasiliou, 2016). De plus, la méthode n'est pas assez sensible lorsque plusieurs résidus sont adjacents, tel que les motifs $\psi\psi$ ou $U\psi$. Une étape supplémentaire d'ajout de queue polyA sur l'ARN est également nécessaire afin de couvrir la séquence complète de l'ARN étudié. Dans le cas contraire, l'absence de sites de liaison des amorces pour la transcriptase inverse rend impossible la caractérisation de l'extrémité 3' de l'ARN (Ofengand *et al*, 2001).

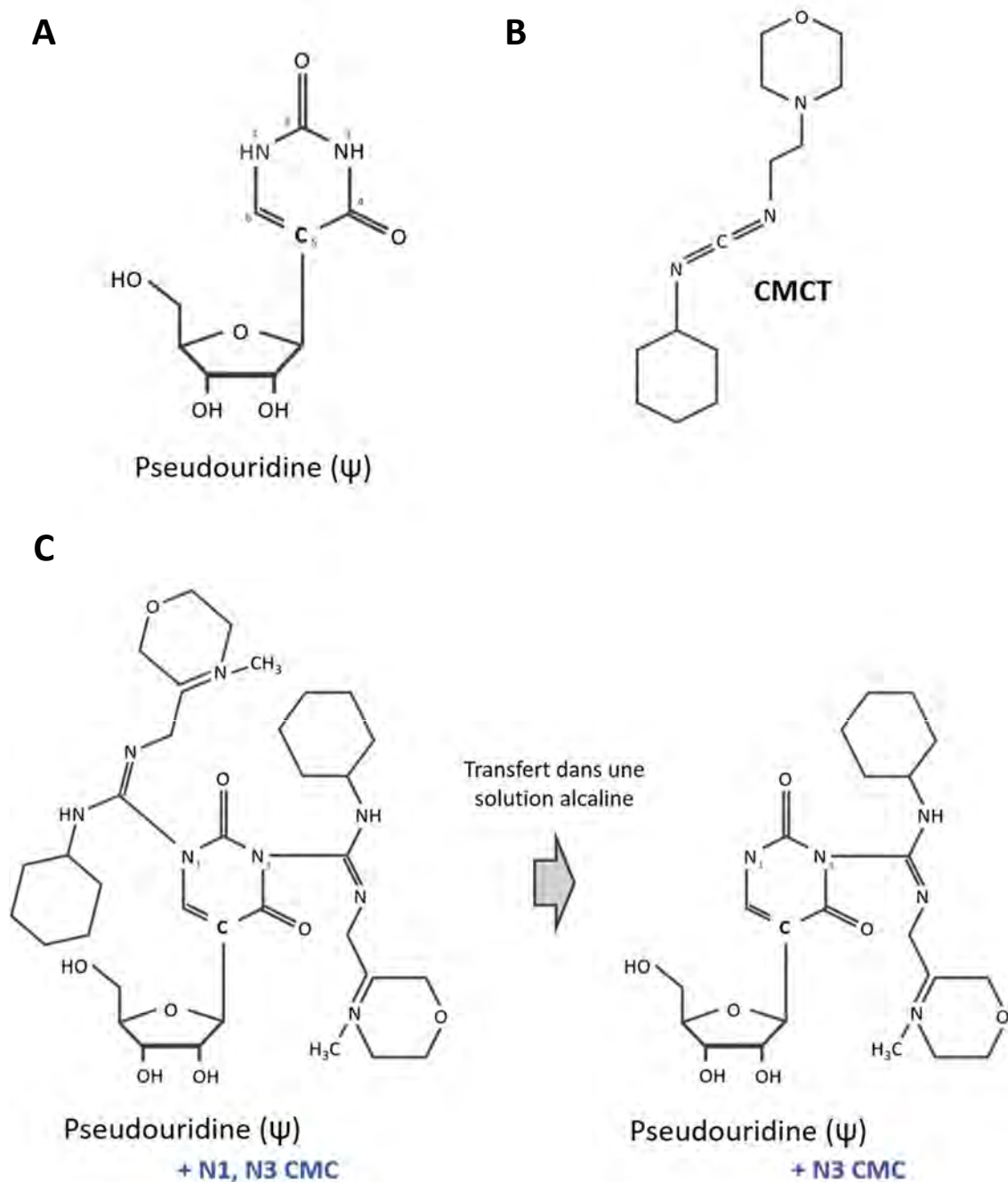


Figure 20 : Modifications des Pseudouridines par traitement au CMCT

(A) Schéma d'une Pseudouridine. (B) Schéma d'une molécule de CMCT (N-cyclohexyl-N'-β-(4-methylmorpholinium) ethylcarbodiimide p-tosylate). (C) Représentation schématique d'un traitement d'une pseudouridine au CMCT. La partie Carbodiimide (CMC) du CMCT réagit avec les azotes en position 1 et 3 de la pseudouridine. Après transfert dans une solution à pH 10.4, la liaison reliant le CMC à l'azote N1 de la pseudouridine est clivée. Le CMC restant au niveau de l'azote N3 permet l'identification de la pseudouridine. Adaptée de (Lovejoy *et al*, 2014a).

III. 1. 1. 2. Autres composés chimiques permettant l'identification des pseudouridines

En 1974, dans l'optique d'étudier les modifications des ARNt, une équipe s'est servie de la molécule 7-methoxy-4-bromomethyl coumarine. Cet agent fluorescent réagit spécifiquement avec les pseudouridines (Yang & Söll, 1974). Certaines molécules dérivées de la 7-methoxy-4-bromomethyl coumarine ont ensuite été testées, mais elles n'ont pas montré la même spécificité (Plumbridge *et al*, 1980).

Deux autres composés chimiques se sont révélés être utile pour l'étude des pseudouridines. Le premier, l'acrylonitrile, est un composé chimique qui interagit avec toutes les bases azotées de l'ARN, ainsi que les pseudouridines, dans un milieu à pH 11.5. Cependant, à pH 8.8, l'acrylonitrile cyanoéthyle (attachement d'un groupe CH₂-CH₂-CN à une molécule substrat) spécifiquement la pseudouridine au niveau de l'azote en position 1 (N-1). Du fait de l'interaction spécifique de l'acrylonitrile avec la pseudouridine, cette méthode ne nécessite pas d'étapes supplémentaires visant à retirer les liaisons entre l'acrylonitrile et les autres bases azotées, contrairement au CMCT (Yoshida & Ukita, 1968). Néanmoins, cette molécule est hautement plus dangereuse à manipuler que le CMCT.

Le second composé, le méthyl vinyl sulfone (MVS), dans un milieu à pH neutre chauffé à 70°C, réagit également avec l'azote N1 des pseudouridines. Les pseudouridines modifiées avec le MVS sont détectées plus facilement lors de la séparation chromatographique. Par contre, les uridines se retrouvent également modifiées lors de l'utilisation du MVS. A contrario du CMCT ou de l'acrylonitrile, le MVS ne peut servir de marqueur spécifique des pseudouridines (Emmrechts *et al*, 2005).

III. 1. 1. 3. Utilisation de la Spectrométrie de masse pour l'identification des pseudouridines

La spectrométrie de masse (MS) est une méthode de plus en plus répandue pour étudier et quantifier les modifications chimiques des ARN. Même si l'uridine et la pseudouridine possèdent la même masse, il est possible de les différencier par MS en analysant les ions issus de leur fragmentation (Addepalli & Limbach, 2011; Taucher *et al*, 2011).

L'approche généralement utilisée pour étudier les modifications post-transcriptionnelles des ARN est la chromatographie liquide couplée à une spectrométrie de masse à ionisation par électrobulsion (LS-ESI-MS). Pour ce faire, les ARN sont digérés par une hydrolyse enzymatique de manière à isoler les nucléosides. Les mesures de masse de ces nucléosides couplées à l'analyse par chromatographie liquide permettent d'identifier les modifications post-transcriptionnelles présentes sur une séquence d'ARN donnée. Cependant, il n'est pas possible,

via cette approche, de déterminer précisément la position du nucléoside qui est porteur de telle ou telle modification détectée (Ross *et al*, 2016). Il est possible de coupler à cette technique l'utilisation de l'agent chimique MVS. L'utilisation du MVS permet de différencier plus aisément les pseudouridines des autres nucléosides lors de l'utilisation de LS-MS. Par contre, le MVS ne permet pas de compenser les limites de l'identification précise du nucléoside modifié de cette approche (Emmerechts *et al*, 2005).

Afin de compenser ces limites, un mapping aux RNases est généralement couplé aux techniques de spectrométrie de masse. Ces enzymes permettent de digérer l'ARN d'intérêt en différents oligonucléotides, qui sont par la suite analysés par MS. La différence des masses obtenues expérimentalement de ces oligonucléotides avec leurs masses théoriques en absence de modifications permet de localiser plus précisément les modifications post-transcriptionnelles détectées (Kowalak *et al*, 1993). Cependant, ce protocole n'est pas efficace pour toutes les modifications, notamment les pseudouridines, leurs masses étant équivalentes avec celles des uridines.

A l'instar du MVS, il existe des protocoles mettant en jeu un traitement des ARN avec l'acrylonitrile ou le CMCT avant l'analyse de ces ARN par spectrométrie de masse MALDI (Matrix Assisted Laser Desorption Ionization), pour détecter les pseudouridines. Notamment, l'acrylonitrile a été utilisé pour mettre en évidence des sites de pseudouridylation sur l'ARN^{TyrII} chez *E.coli* (Mengel-Jørgensen & Kirpekar, 2002). Le traitement des ARN à l'acrylonitrile est précédé d'une hydrolyse enzymatique par les RNases A et T1. Les résultats obtenus sont ensuite comparés à des échantillons n'ayant pas subi le traitement à l'acrylonitrile. La différence de 53 Da entre les deux échantillons a permis de mettre en évidence les fragments d'ARN contenant les résidus ayant subi une pseudouridylation (Mengel-Jørgensen & Kirpekar, 2002).

Lors de l'utilisation de l'acrylonitrile couplé à la spectrométrie de masse, les résultats obtenus se sont révélés être plus pertinents qu'un simple traitement au CMCT, couplé à une extension d'amorce. En effet, lors de l'étude de l'ARNr 23S chez l'archée *Haloarcula marismortui*, la méthode CMCT + extension d'amorces a révélé les sites U1956, U1958, U2607 et U2621 comme étant pseudouridylés (Kirpekar *et al*, 2005). Cependant, lors de l'analyse du même ARN avec l'utilisation d'acrylonitrile couplé au MALDI-MS, le site U2607, s'est révélé être un faux positif (Del Campo *et al*, 2005). D'un autre côté, il est également possible d'ajouter une étape d'analyse MALDI-MS à la suite d'un traitement des ARN au CMCT. L'ajout du CMC représente une augmentation de masse de 252 Da, rendant ainsi possible de faire la différence

entre une uridine et une pseudouridine lors de l'analyse par MS. Couplés à une analyse des fragments d'ARN par HPLC (High Performance Liquid Chromatography) et un séquençage enzymatique, il est possible de déterminer la position exacte des pseudouridines détectées avec l'utilisation du CMCT et de la MS (Patteson *et al*, 2001).

L'utilisation d'un composé chimique interagissant avec les pseudouridines, couplé à la spectrométrie de masse, permet d'identifier précisément les uridines modifiées d'une séquence d'ARN donnée. Mais la tâche peut se révéler plus complexe si la pseudouridine est elle-même porteuse d'autres modifications, comme les 2'-O-méthyl pseudouridines, les 1-méthylpseudouridines (substitution de l'azote 1 par un groupe méthyl) ou les 1-méthyl-3-(3-amino-3-carboxypropyl) pseudouridines (**Figure 21**)(Rozenski *et al*, 1999). Par exemple, la modification 1-méthyl-3-(3-amino-3-carboxypropyl) s'ajoute au niveau de l'azote 3 de la pseudouridine, site nécessaire à la liaison du CMCT. Les pseudouridines portant la modification 1-méthyl-3-(3-amino-3-carboxypropyl) ne seront donc pas détectable par un traitement au CMCT couplé à une extension d'amorce et/ou au MALDI-MS. A contrario, l'acrylonitrile cyanoéthyle l'azote en position 1 des pseudouridines. Les 1-méthyl-3-(3-amino-3-carboxypropyl) pseudouridines seront alors détectées, contrairement aux 1-méthylpseudouridines.

III. 1. 1. 4. Les méthodes permettant de détecter les pseudouridines au niveau de l'ensemble du transcriptome

Les différentes méthodes récemment énoncées permettent de déterminer l'emplacement d'une pseudouridine sur la séquence d'un ARN purifié. L'avancée des technologies, notamment des techniques de séquençages, ont permis l'émergence de nouvelles méthodes capables de déterminer l'emplacement des pseudouridines, non plus à l'échelle d'un ARN, mais à l'échelle du transcriptome.

Le RBS-seq est une technique permettant de détecter l'emplacement de différents nucléosides modifiés, dont les pseudouridines, des différents ARN présents dans le transcriptome. Pour ce faire, cette technique requiert l'utilisation d'un composé chimique, le bisulfite. Les ARN totaux, après être extraits des cellules, sont traités avec du bisulfite à 50°C (Khoddami *et al*, 2019). Cette opération rajoute au niveau des pseudouridines un groupement SO₃, qui est contourné par la transcriptase inverse lors de l'étape de synthèse de l'ADNc à partir des ARN traités. Lors de l'analyse des ADNc, ces contournements vont se traduire par une délétion d'un ou deux nucléotides, révélant l'emplacement des pseudouridines (**Figure 22**).

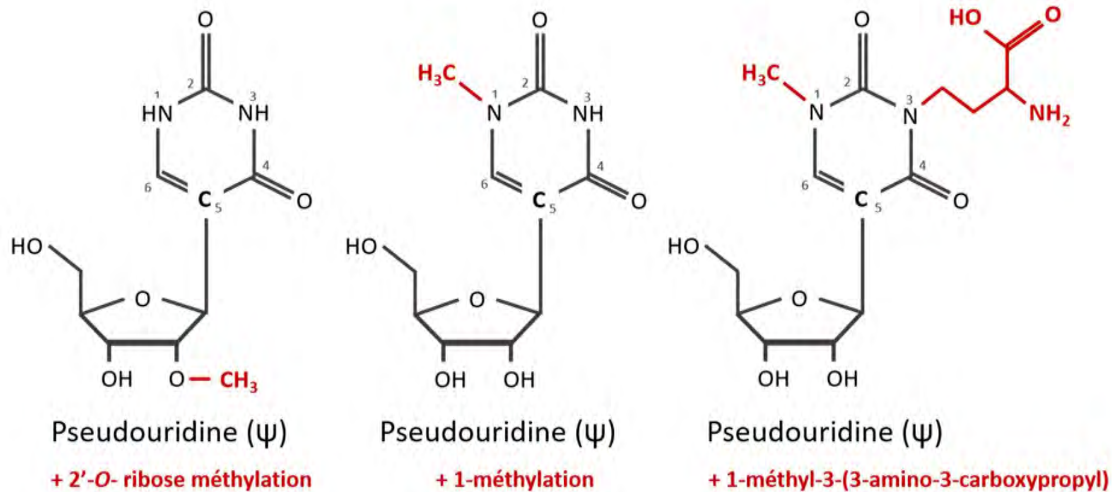


Figure 21 : Modifications chimiques portées par les Pseudouridines

Illustrations de pseudouridines modifiées. De gauche à droite : 2'-O-méthyl pseudouridine, 1-méthylpseudouridine et 1-méthyl-3-(3-amino-3-carboxypropyl) pseudouridine.

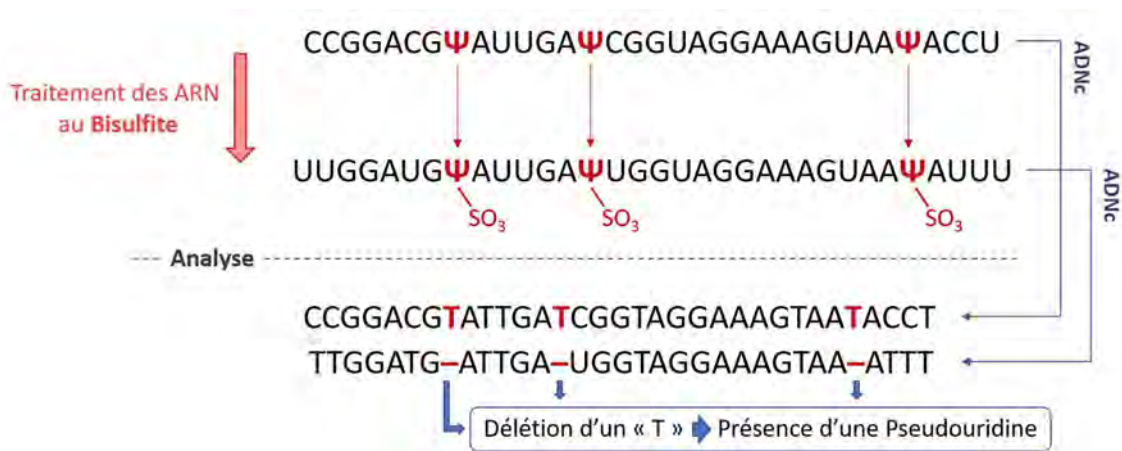


Figure 22 : Le RBS-Seq, une méthode permettant de détecter les Pseudouridines

Représentation schématique des étapes composant la technique RBS-Seq. Après extraction, les ARN cellulaires sont traités au bisulfite, ajoutant des groupements SO_3 au niveau des résidus Pseudouridine. Lors de la synthèse d'ADNc, la transcriptase inverse contourne les nucléotides possédant un groupement SO_3 , induisant des délétions dans les séquences d'ADNc. Ces délétions sont la signature de l'emplacement des Pseudouridines. Adaptée de (Khoddami *et al*, 2019a).

En parallèle du RBS-Seq, cinq méthodes couplées aux méthodes de séquençage nouvelle génération, nommées Pseudo-seq, ψ -seq, PSI-seq, CeU-seq et HydraPsiSeq ont récemment vu le jour (Lovejoy *et al*, 2014; Schwartz *et al*, 2014; Carlile *et al*, 2015; Li *et al*, 2015; Marchand *et al*, 2020). Les méthodes Pseudo-seq, ψ -seq, PSI-seq et CeU-seq reposent sur l'utilisation des arrêts de la transcriptase inverse induits par la présence de CMCT pour déterminer la présence des pseudouridines sur les ARN cellulaires (Li *et al*, 2017).

Ces quatre méthodes débutent toutes par un ajout de polyA aux extrémités 3' des ARN totaux après leur extraction des cellules (**Figure 23**). Ces ARN sont ensuite fragmentés, puis subissent un traitement au CMC couplé à une hydrolyse alcaline, de manière à modifier spécifiquement les pseudouridines. Concernant les méthodes Pseudo-seq et PSI-seq, la fragmentation est faite de façon à obtenir des fragments de tailles uniformes. Après le traitement au CMC et la synthèse d'ADNc par transcriptase inverse, les fragments possédant des pseudouridines sont alors tronqués à cause du blocage de la transcription inverse. Cela permet de les sélectionner et donc d'enrichir les échantillons en séquences contenant une pseudouridine. Le CeU-seq en revanche, utilise un CMC lié à un azide, permettant de former une liaison tri-azole stable avec une molécule de biotine par la réaction chimique « Click ». Les fragments contenant du CMC peuvent alors être sélectionnés et enrichis grâce à l'interaction biotine-streptavidine (Li *et al*, 2015). La technique ψ -seq n'utilise aucune stratégie afin d'enrichir les échantillons en fragments contenant des pseudouridines. Par contre, une méthode bio-informatique lors de l'analyse permet de réduire le bruit de fond (Schwartz *et al*, 2014). Pour chacune de ces méthodes, entre l'étape de traitement au CMC et la transcription inverse, un adaptateur est ligué au niveau des extrémités 3' des ARN. Après transcription inverse, un second adaptateur est ajouté aux extrémités 5' des ADNc lors du ψ -seq. Pour les trois autres méthodes, les ADNc sont circularisés. Ces étapes permettent d'amplifier par PCR les fragments d'ADNc obtenus, puis de les séquencer. Pour finir, ces méthodes utilisent toutes des approches bio-informatiques similaires pour révéler les sites de pseudouridylation. La comparaison entre des échantillons traités au CMCT et d'autres sans traitement permet de déterminer si les arrêts de la transcriptase inverse sont bien dus à la présence d'une pseudouridine couplée au CMCT ou à une structure secondaire de l'ARN.

La méthode HydraPsiSeq repose en revanche sur un clivage aléatoire des uridines (Marchand *et al*, 2020). Pour ce faire, deux molécules chimiques sont utilisées, l'hydrazine (N_2H_4) et l'aniline ($C_6H_5-NH_2$). L'hydrazine permet d'ouvrir le cycle de base pyrimidine des uridines, et l'aniline clive alors le site abasique de l'ARN (**Figure 24**). Après l'action de ces

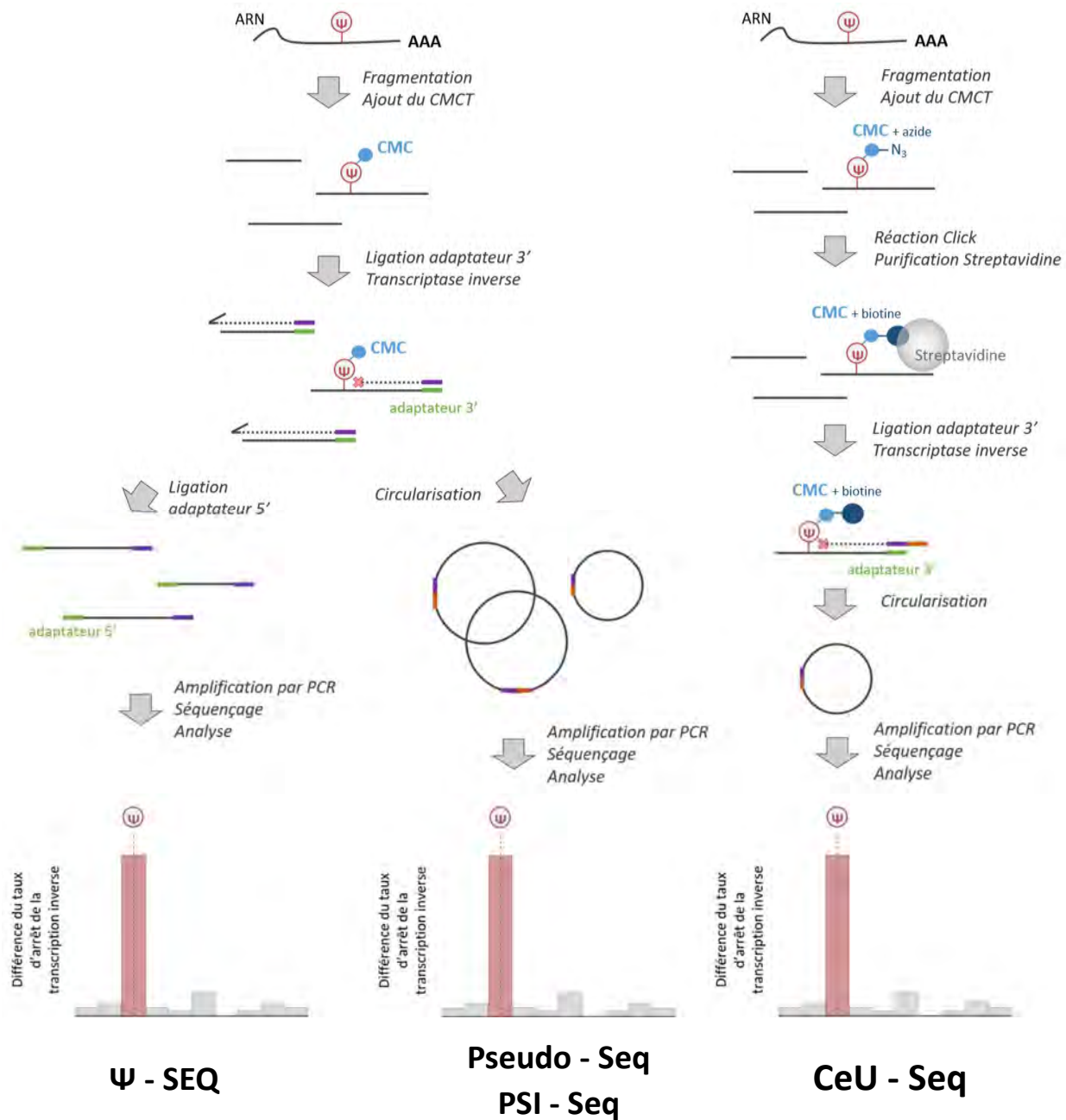


Figure 23: Méthodes permettant la cartographie des Pseudouridines au niveau de l'ensemble du transcriptome

Illustrations des différentes étapes composant les méthodes Ψ – SEQ, Pseudo-Seq, PSI-Seq, CeU-Seq. Adaptée de (Li *et al*, 2017).

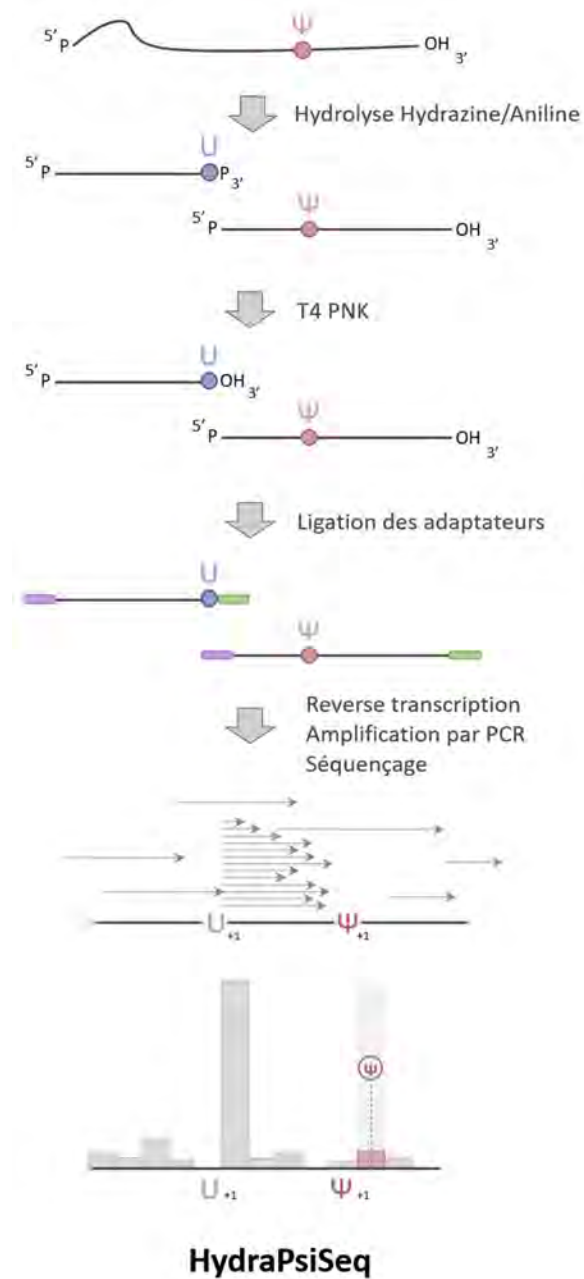


Figure 24 : Cartographie des Pseudouridines du transcriptome grâce à l'HydraPsiSeq

Représentation schématique des étapes de la méthode HydraPsiSeq. Adaptée de (Marchand *et al*, 2020)

deux molécules, les ARN sont traités avec la T4-Polynucléotide kinase (T4-PNK), ce qui permet d'enlever les résidus phosphates issus des clivages par l'aniline aux extrémités 3' des fragments. Les bases A, C, G et les pseudouridines sont résistantes à l'hydrazine et ne sont donc pas clivées (Peattie, 1979). Des adaptateurs sont ensuite ajoutés aux extrémités 3'-OH et 5'-phosphates des différents fragments. Ils permettent, à l'instar des quatre techniques précédentes, une amplification par PCR des fragments, suivie d'un séquençage nouvelle génération (Marchand *et al*, 2020). En comparant avec les séquences de références, il est possible de compter les reads qui commencent à la position N+1 par rapport aux uridines. Les A, C et G servent à quantifier le bruit de fond par leur résistance à l'hydrazine. Les uridines sur les séquences de références pour lesquelles peu de reads débute à la position N+1 expérimentalement sont donc protégées du clivage, traduisant la présence de pseudouridines.

De par l'émergence de ces nouvelles méthodes impliquant les séquençages nouvelle génération, il a été possible d'étudier et d'enrichir les connaissances sur le paysage des pseudouridines des ARN cellulaires. Tout en confirmant les sites de pseudouridines déjà connus, ces techniques ont révélé de nouvelles pseudouridines, notamment sur les ARNm (Zaringhalam & Papavasiliou, 2016). En détectant les pseudouridines avec une précision au nucléotide près, ces techniques ont aussi permis de relier la synthèse de certaines pseudouridines aux PUS et/ou snoRNP correspondantes. Par ailleurs, le CeU-seq a permis de valider expérimentalement le premier exemple de pseudouridylation sur un ARNm, au niveau de l'uridine 519 de l'ARNm EEf1A1 (Li *et al*, 2015). Cependant, la plupart de ces méthodes requièrent l'utilisation du CMC. En théorie, cela limite l'identification de deux pseudouridines adjacentes, due à l'arrêt de la transcriptase inverse sur la pseudouridine en amont de la seconde. Néanmoins, cela ne semble pas toujours vrai expérimentalement. Par exemple, le ψ -seq a révélé plusieurs pseudouridines adjacentes, comme celles en position 1052 et 1056 sur l'ARNr (Li *et al*, 2015). L'HydraPsiSeq ne nécessite pas d'enrichissement sélectif des ARN modifiés. L'HydraPsiSeq se retrouve alors être très efficace dans la détection des pseudouridines très abondantes, tel que les ARNr et ARNt, mais semble avoir une efficacité réduite lors de l'étude d'ARN moins abondant tel que les ARNm, même si cela reste possible (Marchand *et al*, 2020).

III. 1. 1. 5. Utilisation de Nanopores

L'utilisation de Nanopores pour déterminer les modifications des acides nucléiques est un concept qui a émergé il y a une quarantaine d'années. Un nanopore est une protéine transmembranaire modifiée, qui possède un canal par lequel passe un acide nucléique (**Figure 25**). Cette protéine sert de biomarqueur, permettant de détecter la composition des nucléobases

de l'acide nucléique qui le traverse (Deamer *et al*, 2016). Pour cela, ces nanopores sont insérés dans une membrane avec une tension polarisée constante, ce qui permet d'avoir un flux ionique faisant passer les acides nucléiques simple brin à travers les pores. Lors du passage de l'acide nucléique au niveau de la plus étroite des ouvertures du nanopore, le flux d'ions subit des modifications selon la taille et la forme des nucléotides présents (Clamer *et al*, 2014; Smith *et al*, 2015). En amont du nanopore, une hélicase est présente, permettant le déroulement correct de l'ARN ou de l'ADN avant son entrée dans le nanopore. L'avancée de la molécule d'acide nucléique dans le nanopore est stochastique. Pour une molécule d'ARN, les hélicases généralement utilisées permettent une avancée de 70 bases par seconde, et jusqu'à 450 par seconde pour une molécule d'ADN (Garalde *et al*, 2018; Wang *et al*, 2021b).

Cette méthode diffère des précédentes, car elle ne requiert ni d'étape de transcription inverse, ni d'amplification par PCR. Cependant, des biais peuvent intervenir lors de la détection de pseudouridines. Les modifications du flux d'ions d'une pseudouridine sont très similaires des modifications induites par une N1-méthylpseudouridine par exemple, ce qui rend difficile de distinguer les pseudouridines modifiées des pseudouridines classiques (Fleming & Burrows, 2022).

III. 1. 2. Méthodes de détection des méthylations en 2'-O-ribose

A l'instar des pseudouridines, de nombreuses méthodes ont été développées afin de cartographier les méthylations en 2'-O-ribose. Les premières 2'-O-méthylations ont pu être détectées grâce à l'utilisation de chromatographies bidimensionnelles à couche mince et par RNA fingerprinting (Motorin & Marchand, 2018). L'apparition de la spectrométrie de masse a par la suite été une révolution dans l'étude de ces modifications chimiques. La spectrométrie de masse couplée à la chromatographie en phase liquide haute performance (HPLC-MS), et la Quadrupole mass analyzer (q-MS) ont en effet largement été utilisées pour cartographier les méthylations en 2'-O-ribose des ARN cellulaires (Yang *et al*, 2016). Cependant, la MS reste une analyse globale des molécules d'ARN, qui ne cible pas spécifiquement la 2'-O-ribose méthylation. Même si la MS permet une détection précise de ces modifications et de les quantifier, elle reste une méthode coûteuse et exigeante au niveau des compétences et d'équipements nécessaires, tout en ayant des longs délais d'analyse. D'autres méthodes basées sur le séquençage à haut débit, moins coûteuses, et dont la précision d'analyse se rapproche de la MS, ont alors été spécifiquement développées afin d'étudier les méthylations en 2'-O-ribose.

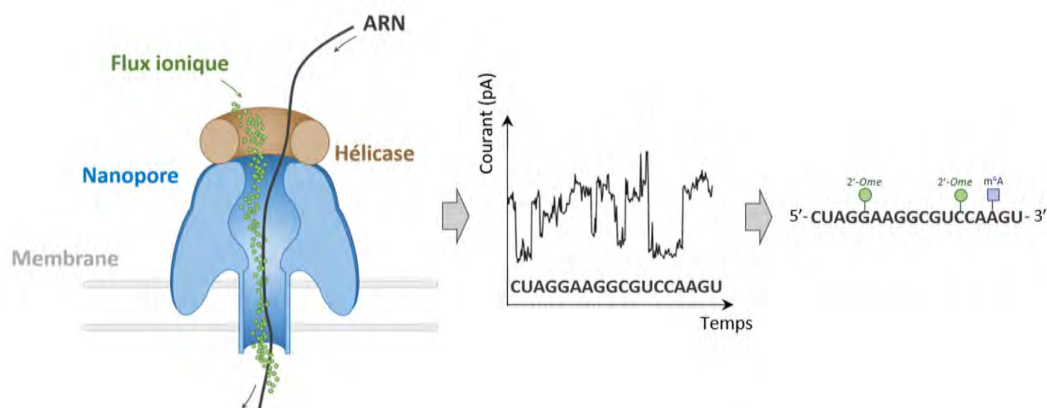


Figure 25 : Cartographie des modifications chimiques du transcriptome par technologie Nanopore.

Représentation schématique d'un Nanopore. Les Nanopores sont placés sur une membrane et plongés dans un milieu conducteur. De cette manière, un courant électrique traverse les Nanopores de part en part. Une Hélicase permet de faire passer la molécule d'ARN à cartographier au travers du Nanopore, perturbant ainsi le courant électrique. En mesurant la variation du courant électrique, il est possible de déterminer les acides ribonucléiques d'une molécule d'ARN, ainsi que les modifications chimiques présentes.

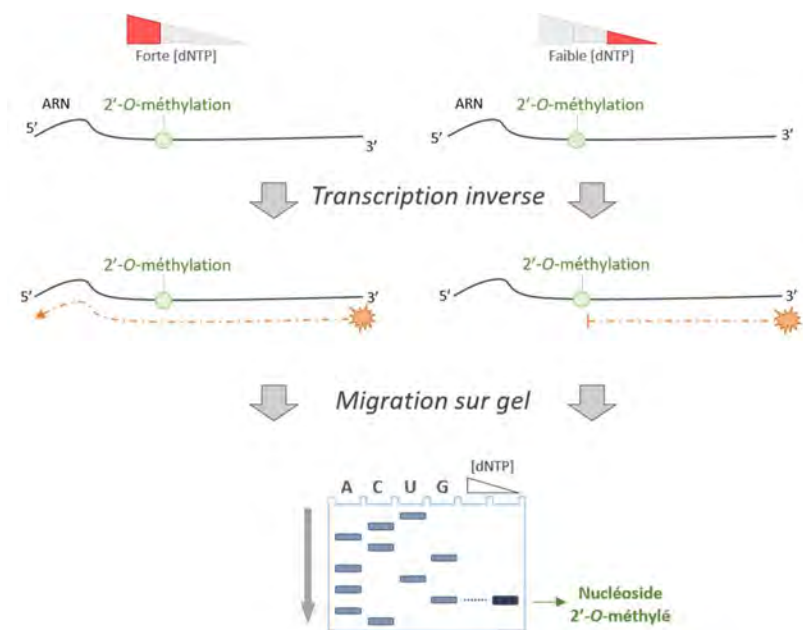


Figure 26 : Détection des méthylations en 2'-O-ribose par Transcription inverse radioactive

A faible concentration de dNTP, les 2'-O-méthylations provoquent un arrêt de l'étape d'élongation de la transcriptase inverse. Les positions des méthylations en 2'-O-ribose sur l'ARN étudiées sont déterminées lors de la comparaison des tailles des fragments rétro-transcrits en faible et forte concentration de dNTP.

III. 1. 2. 1. Techniques à faible débit basées sur des concentrations de dNTP limitantes

Fait intéressant, en condition de désoxyribonucléotide triphosphate (dNTP) limitantes, la transcription inverse se termine prématurément lorsque l'enzyme responsable de ce processus rencontre une méthylation en 2'-O-ribose (Maden *et al*, 1995). Ce mécanisme a servi de base pour le développement de différentes techniques permettant de détecter les nucléotides 2'-O-méthylés, comme notamment les méthodes de transcription inverse radioactive et la RTL-P.

La transcription inverse radioactive implique l'utilisation d'un oligonucléotide radiomarké qui possède une complémentarité avec la zone située en aval de la région d'un ARN à étudier. L'expérience est réalisée en condition de concentration de dNTP croissante. À faible concentration, l'élongation de l'ADNc par la reverse transcriptase est souvent stoppée aux nucléotides possédant une méthylation en 2'-O-ribose, résultant fréquemment à des arrêts prématurés de la synthèse (**Figure 26**). En revanche, à forte concentration de dNTP, la synthèse de l'ADNc n'est pas influencée par la présence des nucléotides méthylés en 2'-O-ribose. La comparaison de la taille des fragments d'ADNc obtenus en faible et en forte concentration de dNTP par autoradiographie révèle les positions des nucléotides sur lesquels une méthylation en 2'-O-ribose est présente (Maden, 2001). La méthode RTL-P repose sur le même mécanisme que la technique précédente, mais utilise la fluorescence à la place de la radioactivité, et couple l'étape de transcription inverse avec une amplification par qPCR (Dong *et al*, 2012). Plus récemment, un variant de l'ADN polymérase a été étudié, qui a la particularité de se retrouver bloqué lors de la présence de méthylations en 2'-O-ribose en condition normale de dNTP. Le gros avantage est de pouvoir réaliser la transcription inverse et l'amplification par q-PCR au cours d'une même réaction. (Aschenbrenner & Marx, 2016). La méthode RTL-P possède cet avantage de remplacer l'utilisation de la radioactivité par la fluorescence, moins dangereuse à manipuler, mais cela engendre une diminution de la sensibilité.

Par contre, ces deux méthodes requièrent généralement la purification des ARN étudiés, rendant plus complexe l'étude d'ARN étant moins abondant au sein des cellules. Également, de faux positifs et faux négatifs sont sensibles d'apparaître si les niveaux de méthylations des résidus sont faibles, et ces méthodes ne permettent pas l'analyse des ARN à haut débit (Maden *et al*, 1995).

Dans l'optique de contourner ces limites, de nouvelles méthodes ont émergé, en utilisant les caractéristiques des méthylations en 2'-O-ribose couplées à la technologie de séquençage, permettant alors l'analyse de ces modifications chimiques à l'échelle du transcriptome.

III. 1. 2. 2. *Le 2'-OMe-Seq*

La méthode 2'-Ome-Seq reprend la spécificité des arrêts de la transcriptase inverse tout en l'adaptant de manière à pouvoir coupler les séquençages nouvelle génération, permettant ainsi d'avoir une vision globale de la présence des méthylations en 2'-O-ribose au sein des ARN cellulaires (Incarnato *et al*, 2017). Les ARN cellulaires, après extraction, sont mis en présence d'amorces aléatoires couplés à des adaptateurs. Une étape de transcription inverse en condition faible de dNTP est alors effectuée (**Figure 27**). D'autres adaptateurs sont ajoutés au niveau des extrémités 5' des ADNc, ce qui permet leur séquençage après une étape d'amplification par PCR antérieure. La faible concentration de dNTP induit les arrêts de la transcriptase inverse un nucléotide en amont des nucléotides méthylés en 2'-O-ribose, traduisant ainsi leur présence. Le comparatif avec la même expérience faite dans des conditions de concentration normale de dNTP permet de déterminer quels arrêts de la transcription inverse sont dues aux méthylations en 2'-O-ribose. Cette condition contrôle permet en supplément de réduire drastiquement les faux positifs, en détectant quels arrêts de la transcriptase inverse sont dues à d'autres modifications que les 2'-O-méthyl, ou aux conformations structurales particulières des ARN étudiés, connues pour bloquer également l'enzyme.

Par cela, le 2'-OMe-Seq permet d'identifier les positions des méthylations en 2'-O-ribose avec une haute sensibilité et une résolution au nucléotide près (Incarnato *et al*, 2017). Cette méthode a notamment permis de révéler douze nouveaux sites de 2'-O-méthyl sur les ARNr 18S et 28S. De plus, l'utilisation de cette méthode dans des lignées Knock Down (KD) pour la Fibrillarine a mis en lumière certains sites de méthylations associés à des snoRNP putatifs (Incarnato *et al*, 2017). Cependant, cette technique n'a, pour le moment, seulement été utilisée dans le cadre de l'étude des ARNr, laissant penser qu'elle ne serait pas aussi efficace lors de l'étude des méthylations en 2'-O-ribose d'ARN moins abondants, tel que les ARNm.

III. 1. 2. 3 *Le RiboMeth-seq*

Le RiboMeth-Seq est une méthode permettant de détecter les méthylations en 2'-O-ribose en haut débit, mais ne se repose pas sur les pauses observées lors d'étapes de transcription inverse, contrairement aux techniques vues précédemment. Le RiboMeth-Seq s'appuie sur la résistance aux clivages lors d'une attaque alcaline que procure une 2'-O-méthyl à la liaison phosphodiester adjacente (Birkedal *et al*, 2015).

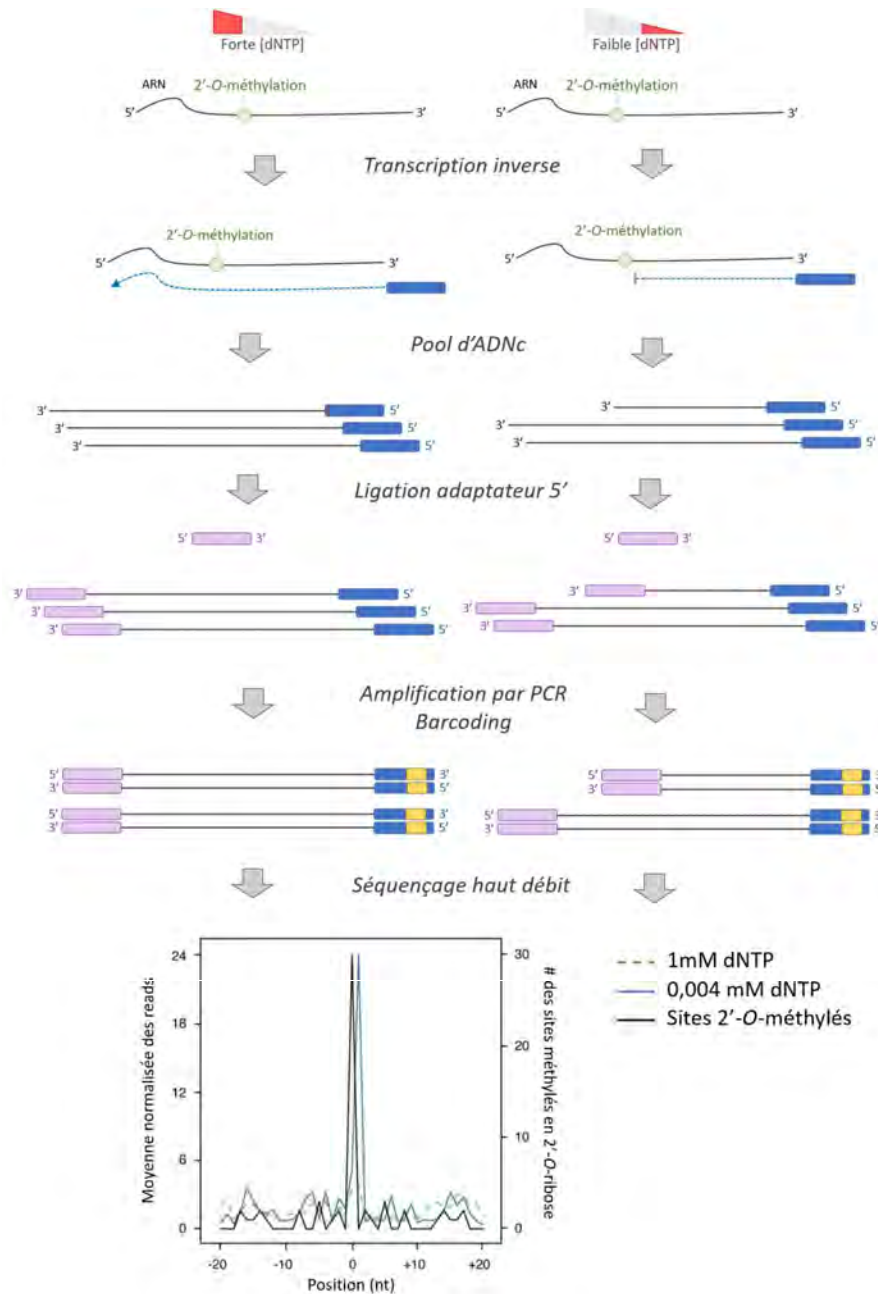


Figure 27 : Le 2'-OMe-Seq

Représentation schématique des étapes composant la méthode 2'-OMe-Seq, permettant de cartographier les méthylations en 2'-O-ribose de l'ensemble du transcriptome. Les ARN cellulaires sont rétro-transcrits en faible ou forte concentration de dNTP. Les ADNc, via ligations d'adaptateurs, sont amplifiés par PCR et séquencés. Les arrêts de transcription induits par une faible concentration de dNTP traduisent les sites méthylés.

Après extraction, les ARN sont soumis à une dégradation alcaline partielle, et les fragments générés (entre 20 et 40 nt) sont ligués à des adaptateurs ARN permettant le séquençage, et copiés en ADNc (**Figure 28**). La ligation est réalisée par l'ARNt ligase et est permise grâce aux extrémités 5' hydroxyle et les extrémités 2',3' phosphate cyclique issues du clivage par l'attaque alcaline. La librairie d'ADNc est alors séquencée sans passer par l'étape d'amplification. Cette étape de séquençage permet d'enregistrer et de comptabiliser les extrémités 5' et 3' des fragments. Idéalement, cette méthode donnera un grand nombre de comptage sur les nucléosides non méthylés, et un comptage proche de zéro au niveau des sites sur lesquels une méthylation en 2'-O-ribose est présente. En effet, ces positions sont protégées lors de l'hydrolyse alcaline et aucun fragment ayant cette position en 3' ou en 5'+1 ne seront générés. Si jamais des sites sont modifiés partiellement, un nombre intermédiaire de comptage sera relevé. Cette méthode permet donc de quantifier les méthylations en 2'-O-ribose de manière absolue, ce qui n'est pas le cas du 2'-OMe-seq par exemple (Birkedal *et al*, 2015). En revanche, le 2'-OMe-seq possède une sensibilité et une spécificité plus élevée que le RiboMeth-seq, notamment à cause de la résistance au clivage par hydrolyse alcaline conférée par d'autres modifications, comme les pseudouridines (Incarnato *et al*, 2017). Le RiboMeth-seq ne possédant pas d'étape d'amplification, cette méthode est généralement réservée à l'étude d'ARN abondants dans les cellules, comme les ARNr, snARN et ARNt. Par ailleurs, comme vu précédemment, le RiboMeth-seq a révélé des taux de méthylations en 2'-O-ribose très réduits au niveau du snARN U6 chez les patients atteints du syndrome d'Alazami, aidant grandement à la compréhension de cette pathologie (Hasler *et al*, 2020).

III. 1. 2. 4. RibOxi-Seq et Nm-Seq

Les méthodes RibOxi-Seq (Zhu *et al*, 2017) et Nm-Seq (Dai *et al*, 2017) sont deux autres méthodes permettant de détecter les méthylations en 2'-O-ribose au niveau de l'ensemble du transcriptome. Ces deux méthodes exploitent la résistance des extrémités 3' des riboses possédant une méthylation en 2'-O-ribose au clivage induit par le périodate (IO_4^-). Concrètement, les ARN subissent une fragmentation aléatoire par une nucléase, donnant des fragments possédant des extrémités 5' phosphate et 3' OH. Ces fragments d'ARN subissent ensuite une oxydation au périodate (**Figure 29**). Les riboses non modifiés sont clivés et convertis en dialdéhydes. A contrario, les extrémités 3' des fragments qui possèdent un ribose avec une 2'-O-méthylation résiste à l'action du périodate. Lors de l'ajout des adaptateurs permettant un futur séquençage haut débit, les extrémités des fragments possédant un ribose non protégé, et donc converti en dialdéhyde, ne sont pas reconnues et exclues de la future

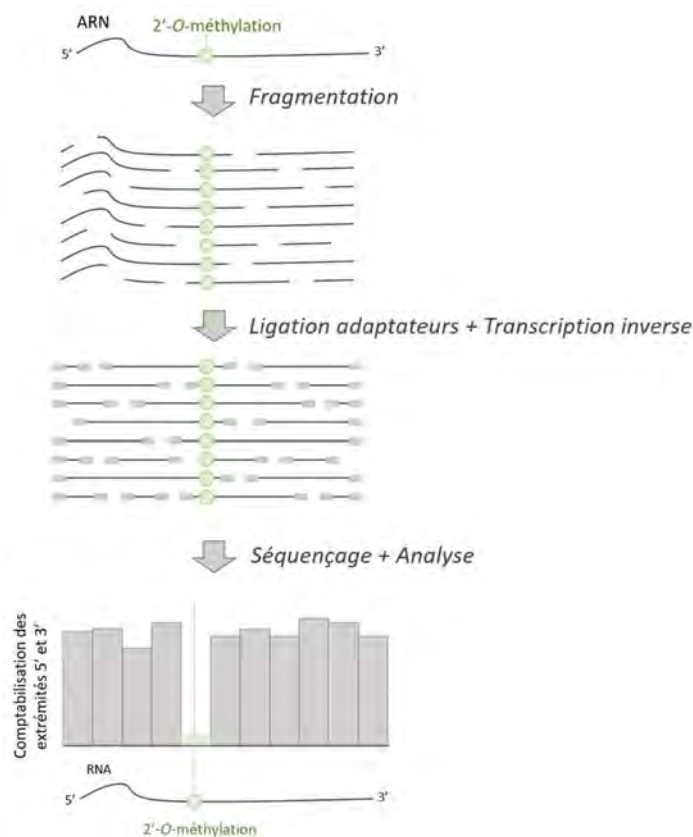


Figure 28 : Le RiboMeth-Seq

Représentation schématique de la méthode RiboMeth-Seq, permettant de cartographier les méthylations en 2'-O-ribose de l'ensemble du transcriptome. Les fragments générés lors d'une hydrolyse alcaline d'ARN cellulaires sont ligués à des adaptateurs et rétro-transcrits. Le séquençage haut-débit est effectué sans passer par une étape d'amplification. Les méthylations en 2'-O-ribose protègent les acides ribonucléiques de l'attaque alcaline. Les résidus protégés par la méthylation sont moins fréquemment retrouvés au niveau des extrémités 5' et 3', permettant ainsi leur détection.

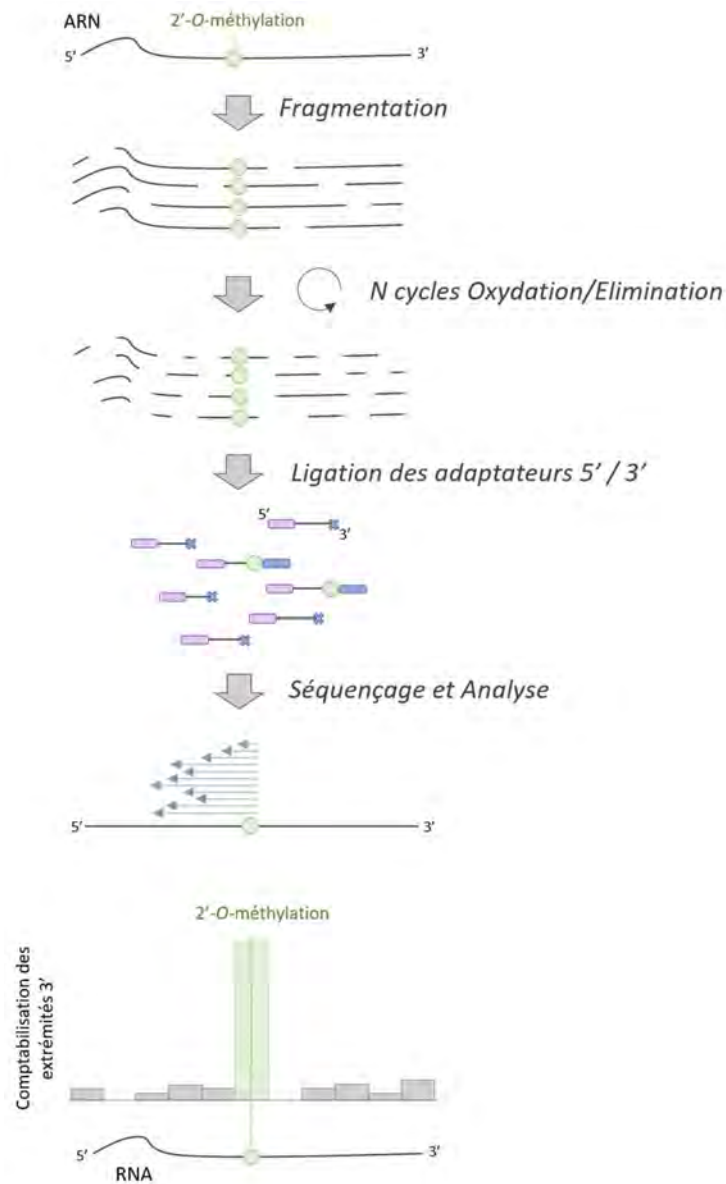


Figure 29 : RibOxi-Seq et Nm-Seq

Schéma présentant les étapes des méthodes équivalentes RibOxi-Seq et Nm-Seq, permettant de détecter les 2'-O-ribose méthylations. Après une fragmentation des ARN, plusieurs cycles d'oxydation au périodate / élimination des fragments n'ayant pas leurs riboses protégés par une 2'-O-méthylations en extrémité 3' sont effectués. En comptabilisant les extrémités 3' des fragments après une étape de séquençage, les positions des 2'-O-méthylations sont détectées.

librairie de ADNc. Un enrichissement d'extrémités 3' à une position donnée traduit donc la présence d'une méthylation en 2'-O-ribose. Le clivage par les nucléases n'est pas réellement aléatoire, et ces derniers apparaissent rarement pile au niveau des 2'-O-méthyl. Afin de pallier à ces problématiques, plusieurs cycles d'oxydation-élimination-déphosphorylation (au moins 8 en général) sont habituellement effectués sur les mêmes échantillons d'ARN. Même si une succession de ces cycles engendre une perte conséquente de matériel biologique, ces deux méthodes restent relativement sensibles dans l'identification des 2'-O-méthylations. Notamment, le Nm-seq a permis de détecter précisément ce type de modifications chimiques sur des ARN cellulaires peu abondant, tel que les ARNm (Dai *et al*, 2017).

Parmi les différentes méthodes permettant l'analyse des méthylations en 2'-O-ribose, l'analyse via la spectrométrie de masse reste la plus qualitative (Krogh & Nielsen, 2019). Cependant, ces méthodes sont laborieuses à mettre en place, et difficilement utilisables dans le cadre d'expérience biologique visant à étudier des modifications en particulier. Concernant les méthodes basées sur le séquençage à haut débit, le RibOxi-Seq et le Nm-Seq semblent particulièrement bien fonctionner malgré le fait qu'elles ne fournissent que des valeurs relatives, et seulement si la quantité des ARN étudiés est suffisante. Pour les analyses quantitatives, l'utilisation de la méthode RiboMeth-Seq semble être préférable. En effet, les résultats obtenus par cette méthode se rapprochent de ceux générés en spectrométrie de masse (Krogh & Nielsen, 2019).

III. 1. 2. 5. Détection des méthylations en 2'-O-ribose par Nanopores

A l'instar des autres modifications portées par les ARN cellulaires, les méthylations en 2'-O-ribose peuvent être détectées par l'utilisation de Nanopores. Néanmoins, les signaux produits par ces modifications chimiques semblent être moins reproductibles dans des contextes de séquences différentes (Begik *et al*, 2021).

Grâce aux différentes méthodes existantes pour étudier les pseudouridines et les méthylations en 2'-O-ribose, le paysage de ces modifications chimiques sur les ARN abondants, tel que les ARNr et snARN, commence à être richement répertorié. Grâce à l'analyse des régions dans lesquelles ces modifications chimiques ont été détectées, il a été possible de déterminer des éléments antisens de snoARN qui y sont complémentaires. En ce sens, la quasi-totalité des méthylations en 2'-O-ribose répertoriées sur les ARNr sont associées avec un snoRNP C/D (Jorjani *et al*, 2016). Cependant, plus de 50% des snoRNP C/D demeurent toujours orphelines, et n'ont pas de fonctions proposées ou associées. Sachant cela, les probabilités que

les snoARN H/ACA et C/D orphelins ciblent des ARN cellulaires beaucoup moins abondants que les ARNr et snARN sont très fortes.

En ce sens, les méthodes citées précédemment ont permis de cartographier des pseudouridines et méthylations en 2'-O-ribose sur des ARN moins abondants dans la cellule, tel que les ARNm. Malgré tout, la détection de ces modifications chimiques sur de tels ARN reste imprécise et laborieuse.

L'optimisation des méthodes actuelles, ainsi que le développement de nouvelles techniques, sont alors un enjeu crucial pour permettre d'accroître la puissance et la sensibilité de détection des modifications chimiques sur ces ARN cellulaires moins abondant. Cartographier ces ARN permettra de relier les séquences hébergeant des modifications chimiques à des éléments antisens de snoARN orphelins, apportant de nouvelles fonctions et mécanismes aux snoARN.

III. 2. Méthodes *in silico* permettant de prédire les interactions ARN-ARN

Les interactions entre les snoARN et leurs cibles étant par appariement direct de paires de bases, l'efficacité de l'utilisation de méthodes bio informatiques afin de détecter ces associations n'est plus à prouver. En effet, la bio-informatique, en outre de permettre de parcourir rapidement le transcriptome à la recherche d'une séquence particulière, permet également la synthèse de données obtenues expérimentalement par la création de banque de données.

III. 2. 1. Algorithmes de prédiction des cibles des snoARN H/ACA et C/D

III. 2. 1. 1. Prédiction des ARN cibles des snoARN H/ACA

Différents programmes bio-informatiques ont été développés dans le but d'établir une liste de candidats potentiels pouvant être les cibles d'un snoARN donné. Par exemple, les programmes snoGPS (Schattner *et al.*, 2004) et Fisher (Freyhult *et al.*, 2008) parcourent les séquences de l'ADN génomique à la recherche de séquences complémentaires aux éléments antisens d'un snoARN H/ACA donné. Cependant, outre le fait que ces programmes ne tiennent pas en compte de potentiels mésappariements, ils ne donnent pas d'informations sur la stabilité des tiges-boucles ou sur l'énergie lors des repliements des ARN, deux facteurs importants dans la prédiction des interactions snoARN-ARN (Xiao *et al.*, 2009).

En revanche, le programme RNAsnoop associe les logiciels existants pour détecter les potentiels appariements, et un algorithme calculant la thermodynamique lors des repliements des ARN couplé avec du machine learning, permettant d'être plus spécifique dans les cibles putatives déterminées pour un snoARN donné (Tafer *et al.*, 2010). Au cours de différents tests, RNAsnoop et snoGPS se sont retrouvés être plus performant que Fisher notamment lors de la reconnaissance des interactions déjà connues des snoARN H/ACA chez la levure. RNAsnoop et snoGPS détectent respectivement 43 et 41 sites sur les 44 connus, contre 16 par Fisher (Tafer *et al.*, 2010). Lors de prédictions de sites d'interactions chez l'homme, RNAsnoop se révèle être plus performant que snoGPS, notamment en classant les interactions déjà confirmées avec un meilleur score.

III. 2. 1. 2. Prédiction des ARN cibles des snoARN C/D

SnoTARGET est un programme qui a été développé dans l'optique de caractériser des snoARN impliqués dans la régulation du processus d'épissage (Bazeley *et al.*, 2008). SnoTARGET détecte des régions dans les séquences codantes pour des protéines humaines ou murines qui pourraient s'associer avec les éléments antisens d'un snoARN donné. En supplément, ce programme calcule la distance entre une séquence cible putative d'un snoARN avec les jonctions d'épissage à proximité, et donne des informations sur la rigueur de cette potentielle interaction snoARN-ARN. Ce programme fonctionnerait pour l'étude des deux classes de snoARN, mais il a surtout été utilisé pour prédire les cibles potentielles du snoARN C/D SNORD116 (Bazeley *et al.*, 2008). SnoTARGET a permis de mettre en évidence que la moitié des copies du SNORD116 humain issue des séquences répétées de la région chromosomique 15q11-13q cible préférentiellement des régions situées sur des exons, qui font partie de gènes subissant un épissage alternatif. Ces ARN cellulaires cibles sont par ailleurs classés en fonction de l'énergie de repliement du snoARN et de la cible par snoTARGET, à l'instar du programme RNAsnoop pour les snoARN H/ACA.

Le programme bio-informatique PLEXY est quant à lui spécifiquement utilisé pour prédire les cibles des snoARN C/D. En supplément de prédire les interactions snoARN-ARN par appariement de séquences, PLEXY calcule l'énergie lors de l'interaction entre un snoARN et sa cible putative (Kehr *et al.*, 2011). Cette méthode s'est révélée être plus efficace que les programmes de prédictions précédents. Par exemple, PLEXY a prédit 100% des sites d'appariement connus des snoARN sur les ARNr de la levure, contre 74% lors de l'utilisation du programme snoTARGET. De la même manière, 98,3% des sites de méthylations en 2'-O-

ribose connues sur les ARNr humains ont été retrouvés par PLEXY, contre 78,88% par snoTARGET (Kehr *et al*, 2011).

Ces différents programmes bio-informatiques développés démontrent une grande efficacité et fiabilité dans la recherche des cibles de snoARN, notamment en retrouvant avec précision les interactions déjà connues entre les snoARN et les ARNr et snARN. Ces programmes apparaissent alors comme des atouts majeurs dans la recherche des fonctions associées aux snoARN orphelins, même si les validations expérimentales demeurent et resteront nécessaires.

III. 2. 2. Utilisations annexes de l'informatique dans l'étude des snoARN

III. 2. 2. 1. Recherche de nouveaux snoARN putatifs dans le génome

En supplément des séquences consensus, les structures secondaires entre les différents snoARN H/ACA et C/D restent globalement très similaires. En tirant parti de ces informations, des programmes informatiques ont été développés permettant de fouiller dans le génome des gènes putatifs de snoARN qui contiendraient ces séquences consensus, et ayant une structure secondaire prédite analogue aux snoARN déjà connus.

Le programme SnoSeeker est composé de deux algorithmes, ACAsseeker et CDseeker, créé pour trouver de nouveaux snoARN H/ACA et C/D, respectivement. Ces algorithmes combinent une prédiction de structures secondaires des snoARN et un modèle probabiliste (Yang *et al*, 2006). Concernant ACAsseeker, le programme débute par trouver dans le génome des régions contenant les séquences consensus H et ACA. Une fois ces séquences trouvées, le programme vérifie le repliement en tige-boucle des séquences adjacentes et la formation d'une structure secondaire en hairpin-hinge-hairpin-tail. La dernière étape est la recherche de cibles putatives de ce nouvel snoARN parmi les ARNr et les snARN, qui sera classé dans la famille des orphelins, le cas échéant. CDseeker fonctionne de manière assez similaire avec les caractéristiques des snoARN C/D. SnoSeeker a notamment permis l'identification de 54 nouveaux snoARN putatifs, dont 26 orphelins, en supplément des 266 déjà connus, démontrant ainsi la fiabilité du programme. L'accumulation de 18 de ces nouveaux snoARN a ensuite été validée de manière expérimentale (Yang *et al*, 2006).

III. 2. 2. 2. Les bases de données sur les snoARN

La création de base de données rassemblant les informations sur les snoARN est également importante pour la collaboration scientifique en vue de prédire et déterminer les cibles des snoARN orphelins.

La base de données snoRNA-LBME (Lestrade & Weber, 2006) rassemble notamment les cibles, validées expérimentalement ou non, de la littérature. SnOPY donne en supplément des informations sur les locus des gènes hôtes des snoARN et de leurs orthologues (Yoshihama *et al*, 2013). RISE, en addition des cibles validées de la littérature, compile les données obtenues par des études réalisées à l'échelle du transcriptome sur les interactions ARN-ARN.

Le développement et l'utilisation de programmes informatiques jouent un grand rôle dans l'identification de candidats potentiels pouvant être la cible des snoARN orphelins. Cependant, les méthodes de prédictions informatiques nécessitent une validation expérimentale. En ce sens, de nombreuses méthodes étudiant les interactions ARN-ARN à travers le transcriptome ont été développées.

III. 3. Méthodes *in vivo* permettant d'étudier les interactions snoARN-ARN

De nombreuses méthodes permettant d'étudier les interactions ARN-ARN ont été mises en place, tel que la cristallographie par rayon X, le transfert d'énergie par résonance de fluorescence (FRET), la spectroscopie par résonance magnétique nucléaire (NMR) ou encore la purification via un ARN antisens (RAP).

Ces méthodes étant difficiles à mettre en place, et requérant généralement la connaissance des deux éléments de l'interaction, ne contribuent pas particulièrement à la recherche des cibles des snoARN orphelins. En ce sens, elles ne seront pas développées ici. Seules les méthodes permettant d'identifier les interactions snoARN-ARN sans connaître en avance les ARN cibles seront mises en avant.

III. 3. 1. Méthodes utilisant une protéine « appât »

III. 3. 1. 1. Le CLASH

La méthode CLASH (Cross-linking, ligation, and sequencing of hybrids) est une adaptation de la technique de CLIP-Seq (Kudla *et al*, 2011). Le CLASH consiste en un crosslink des interactions protéines – ARN dans la cellule, par une exposition aux UV (**Figure 30**). Les protéines d'intérêts sont ensuite spécifiquement purifiées. Par exemple, lors d'une

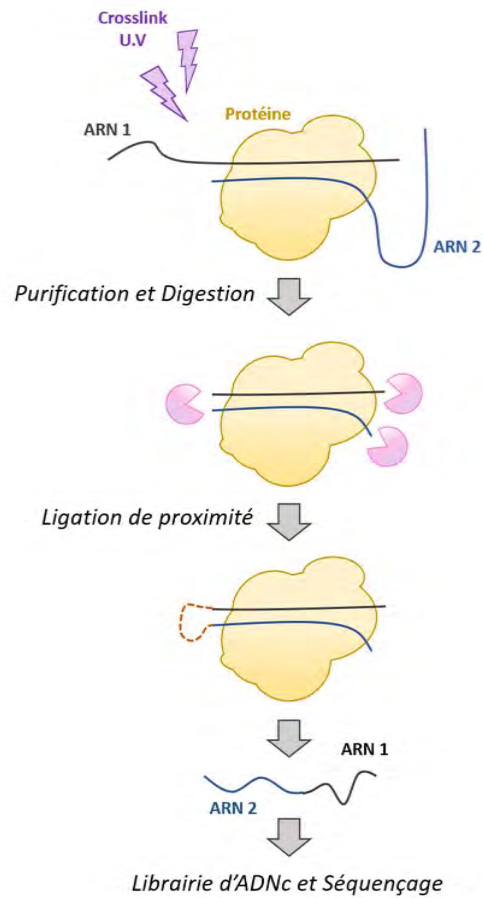


Figure 30 : Le CLASH

Représentation schématique de la méthode CLASH (Cross-linking, ligation, and sequencing of hybrids). *In vivo*, les interactions protéines – ARN sont crosslinkées par une exposition aux UV, puis la protéine d'intérêt est spécifiquement purifiée. Les ARN en interaction sont protégés de l'étape de digestion par les RNases A/T1, et subissent une ligation de proximité. Les ARN interagissant avec la protéine d'intérêt se retrouvent ligués avec les autres ARN avec lesquels ils interagissaient en parallèle, créant ainsi un ARN chimérique. Le séquençage de ces ARN chimériques permet de révéler les interactions ARN-ARN *in vivo*.

étude sur les interactions des snoARN C/D, les auteurs ont purifié les protéines Fibrillarines et NOP56 après crosslink (Dunn-Davies *et al*, 2021). Après une digestion partielle des ARN via les RNases A/T1 succédant à la purification des protéines, les ARN subissent une ligation de proximité. Dans le cas où deux ARN interagissent dans le complexe contenant la protéine, ils seront ligués ensemble, donnant ainsi un ARN chimérique, composé des deux fragments des deux ARN fusionnés. Une librairie d'ADNc est ensuite préparée à partir de ces ARN chimériques, puis séquencée à haut débit.

Le CLASH a notamment démontré son efficacité en identifiant des interactions snoARN-ARNr déjà connues chez la levure, tout en en découvrant de nouvelles. Cette méthode permet également d'analyser les structures intramoléculaires des ARN. En effet, le CLASH a permis de réviser par exemple la structure secondaire du snARN U2 (Kudla *et al*, 2011).

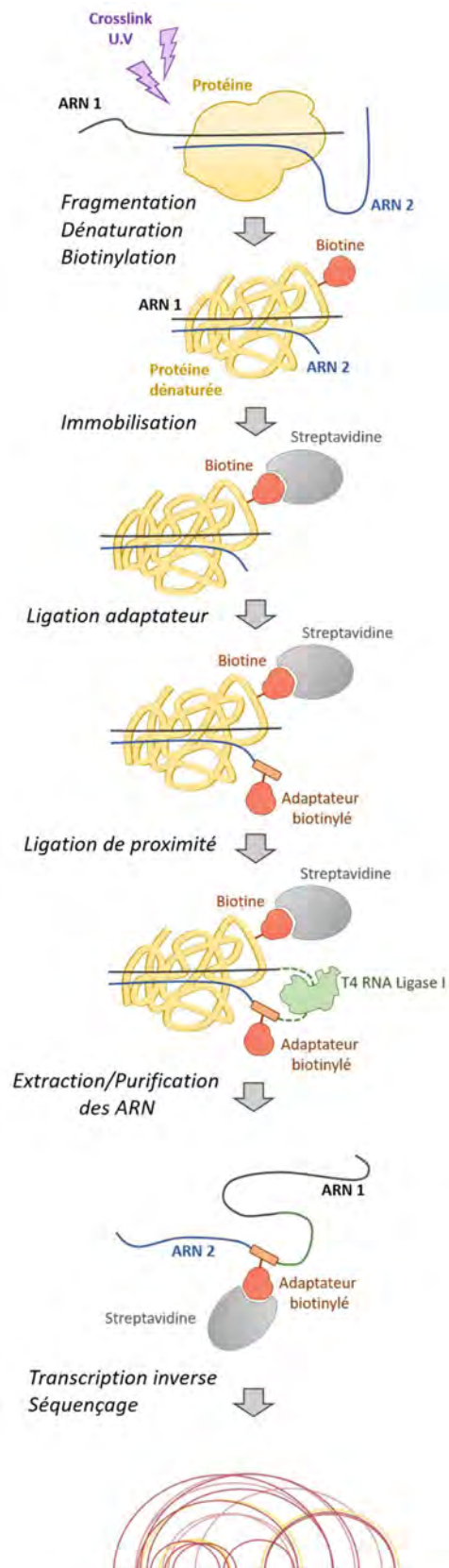
III. 3. 1. 2. Le MARIO

La méthode MARIO (MApping RNA Interactome *In vivo*) a été développée dans le but de pouvoir caractériser les interactions ARN-ARN ayant lieu à proximité de n'importe quelle protéine dans la cellule. Pour ce faire, un crosslink protéine-ARN aux UV est effectué (Nguyen *et al*, 2016). Après une fragmentation des ARN, les protéines sont dénaturées puis biotinylées (**Figure 31**). L'ajout de la biotine permet d'immobiliser les protéines sur des billes de streptavidine. Cette étape permet de réduire les ligations aléatoires entre des oligonucléotides non-crosslinkés, de changer efficacement les tampons, et d'augmenter l'efficacité des lavages afin de réduire le nombre d'interactions non-physiologiques. Un adaptateur biotinylé est ajouté sur les extrémités 5' des ARN, puis une ligation de proximité est effectuée, générant des hybrides d'ARN avec un adaptateur biotinylé ligué entre les deux ARN en interaction. La biotine sur l'adaptateur permet alors de purifier spécifiquement les ARN chimériques sous forme ARN1 – adaptateur – ARN2, qui sont par la suite séquencés (Nguyen *et al*, 2016).

En comparaison du CLASH, la méthode MARIO permet d'étudier la globalité des interactions ARN-ARN ayant lieu dans un complexe protéique grâce à une purification non-spécifique des protéines. Également, en général, la protéine appât lors de l'utilisation de la méthode CLASH est surexprimée (Helwak *et al*, 2013), phénomène pouvant influencer et biaiser les interactions ARN-ARN, ce qui n'est pas le cas lors de la méthode MARIO. En revanche, les méthodes CLASH et MARIO ne permettent pas d'étudier les interactions ARN-ARN n'étant pas dans un complexe ribonucléoprotéique.

Figure 31 : Le MARIO

Illustration de la technique MARIO.



III. 3. 2. Méthodes basées sur l'utilisation du psoralène

Ces méthodes couplent l'utilisation d'un crosslink entre les ARN cellulaires grâce à une molécule chimique, le psoralène, avec des techniques de séquençage nouvelle génération. Cela permet d'avoir une vision globale des interactions physiques entre les différents ARN cellulaires. L'utilisation de ces méthodes constitue un moyen novateur qui permettrait de déterminer les cibles potentielles des snoARN orphelins.

III. 3. 2. 1. Les méthodes LIGR-Seq, PARIS, SPLASH et COMRADES

Les méthodes LIGR-Seq (LIGation of interacting RNA followed by high-throughput Sequencing), PARIS (Psoralen Analysis of RNA Interactions and Structures), SPLASH (Sequencing of Psoralen crosslinked , Ligated, And Selected Hybrids) et COMRADES (Crosslinking Of Matched RNAs And Deep Sequencing) reposent toutes sur l'utilisation du psoralène (Aw *et al*, 2016; Lu *et al*, 2016; Sharma *et al*, 2016; Ziv *et al*, 2018).

Les psoralènes sont des molécules chimiques de la classe des furocoumarines, qui ont la capacité de s'insérer dans les acides nucléiques à double hélice (Calvet & Pederson, 1979). Lors d'une exposition aux UV à 365nm, les molécules de psoralènes intercalées entre les paires de bases subissent une réaction photochimique conduisant à la formation de liaisons transversales covalentes entre les bases pyrimidines (préférentiellement les thymine et uridine). Un des avantages du crosslink au psoralène est qu'il est réversible (Calvet & Pederson, 1979). Ces méthodes se servent du psoralène couplé à une exposition aux UV de manière à figer la globalité des interactions ARN-ARN se déroulant à un instant *t* dans la cellule (**Figure 32**). Les méthodes LIGR-Seq et PARIS utilisent l'AMT psoralène (4'-aminométhyltrioxsalen), tandis que la méthode SPLASH utilise un psoralène biotinylé. Cela permet, par l'interaction avec la streptavidine, d'enrichir les échantillons en ARN crosslinkés. L'ajout d'un détergent doux reste néanmoins nécessaire lors de l'étape de crosslink, tel que la digitonine, car la biotine n'est pas perméable dans les cellules (Aw *et al*, 2016). La méthode COMRADES utilise un psoralène couplé à un groupement azide (Psoralen-triethylene glycol azide), qui lui reste perméable dans les cellules. Ce groupement azide permet de pouvoir y rajouter une molécule de biotine par réaction Click, et permettre ainsi la purification spécifique des ARN crosslinkés (Ziv *et al*, 2018). En revanche, une exposition plus courte aux UV est requise lors de l'étape de crosslink, sous peine de dégrader les composés azide du psoralène.

Une fois l'étape du crosslink ARN-ARN réalisée, la suite des quatre méthodes consiste en une purification et une digestion des ARN cellulaires. La RNase III, hydrolysant

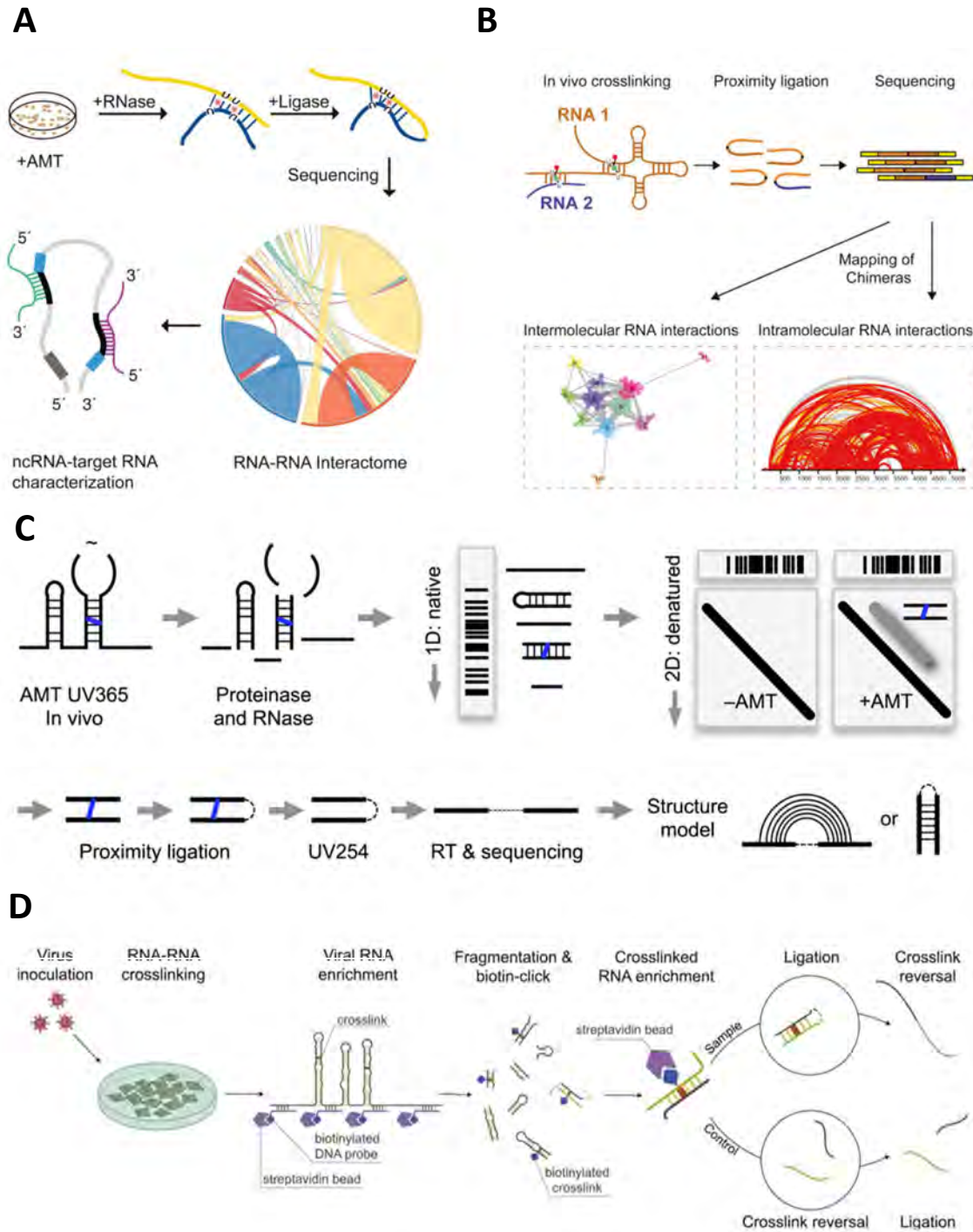


Figure 32 : Méthodes d'études des interactions ARN-ARN basées sur un crosslink *in vivo* au psoralène.

Représentation schématique des différentes méthodes d'études d'interactions ARN-ARN utilisant le psoralène. **(A)** Méthode LIGR-Seq, issue de (Sharma *et al.*, 2016). **(B)** Méthode SPLASH, issue de (Aw *et al.*, 2016). **(C)** Méthode PARIS, issue de (Lu *et al.*, 2016). **(D)** Méthode COMRADES, issue de (Ziv *et al.*, 2020).

spécifiquement les ARN bicaténaires (dsRNA), utilisée dans les méthodes PARIS et COMRADES, permet d'optimiser l'étape ultérieure de ligation de proximité. La méthode LIGR-Seq utilise la nucléase S1, dégradant les ARN simple brin, pour enrichir l'échantillon en duplex d'ARN crosslinkés. En revanche, la technique SPLASH utilise l'hydrolyse chimique pour fragmenter les ARN, et sélectionne ceux d'une taille comprise en 90 et 110nt par séparation de taille sur gel. Avant la digestion, les ARNr sont enlevés lors de l'utilisation du LIGR-Seq, réduisant ainsi le bruit de fond de la méthode (Sharma *et al*, 2016). Les étapes de digestion permettent également de préparer les extrémités des ARN à une étape de ligation. Les méthodes PARIS, SPLASH et COMRADES effectuent une ligation de proximité grâce à la T4 RNA Ligase I, tandis que le LIGR-Seq utilise la circRNA Ligase. Des ARN chimériques, composés d'au minimum deux séquences d'ARN différents, sont alors générés. Le crosslink est par la suite retiré grâce à une seconde exposition aux UV à une longueur d'onde différente (254nm), ce qui permet la préparation d'une librairie d'ADNc à partir de ces ARN chimériques, et un séquençage par méthode NGS (New Generation Sequencing).

En amont du retrait du crosslink, l'enrichissement des échantillons en ARN crosslinkés est effectué par purification avec des billes de streptavidine pour les protocoles SPLASH et COMRADES. Lors de la méthode PARIS, cette sélection est faite grâce une électrophorèse en deux dimensions (Lu *et al*, 2016).

Les méthodes LIGR-Seq, PARIS, SPLASH et COMRADES permettent d'étudier la globalité des interactions ARN-ARN *in vivo*, de manière quantitative et qualitative, sans l'utilisation d'une protéine appât. Le psoralène s'intercalant également dans les structures secondaires double brin des ARN, ces techniques donnent également des informations sur la structure des ARN *in vivo*.

Lors de l'utilisation du LIGR-Seq, de nombreuses interactions connues, telles qu'entre les ARN de l'épissage U4-U6 ou U2-U6, ont permis de démontrer la fiabilité de la méthode. Le LIGR-Seq a notamment mis en évidence l'interaction physique de la snoRNP C/D SNORD83 avec les ARNm NOP14, RPS5 et SRSF3 (Sharma *et al*, 2016). Cette découverte montre que le LIGR-Seq est capable d'explorer les interactions d'ARN cellulaires peu abondants. La méthode PARIS a mis en évidence le fait que certains duplex d'ARN peuvent se former sur de longues distances, avec des structures alternatives et avec des séquences pouvant d'apparier sur plusieurs ARN partenaires différents (Lu *et al*, 2016). Cette étude a montré aussi que l'extrémité 5' de SNORD118 (U8) interagit avec l'ARN28S au niveau de l'extrémité 3', au lieu de celle en 5' comme il était précédemment établi par des études informatiques et génétiques (Peculis,

1997; Michot *et al*, 1999). Également, une seconde édition de la méthode (PARIS-2) a mis en évidence une liaison potentielle entre SNORD118 et SNORD13 (Zhang *et al*, 2021). De nombreuses interactions connues entre les snoARN et les ARNr ont été retrouvées lors de l'utilisation de la méthode SPLASH (55 C/D et 23 H/ACA). Même si les auteurs se sont majoritairement concentrés sur l'interactome des ARNm et l'organisation structurale du transcriptome, aucune indication à propos de potentielles interactions non-canoniques entre les snoARN et ARNm n'a été relevée (Aw *et al*, 2016). La méthode COMRADES a eu pour particularité d'étudier les interactions ARN-ARN dans un contexte où les cellules humaines sont infectées par un virus (ici, virus Zika). COMRADES a révélé que le transcriptome cellulaire est très dynamique, notamment lors d'une infection. Certains ARN viraux ont la capacité d'interagir par appariement direct de paires de base avec les ARN cellulaires hôte, notamment la région 5'CS du virus Zika avec le microARN 21 humain (Ziv *et al*, 2018).

III. 3. 2. 2. Avantages & Limites des méthodes ayant recourt au psoralène

L'utilisation de ces différentes méthodes requérant l'utilisation du psoralène, outre de déterminer les interactions ARN-ARN dans la cellule, dispose de nombreux avantages. Ces méthodes permettent notamment de déterminer des interactions structurales à longue distance des ARN, comme les pseudoknots, contrairement à des méthodes reposants sur des mécanismes d'ARN footprinting couplées au séquençage haut débit (Ding *et al*, 2014). Les méthodes au psoralène permettent également l'étude des structures secondaires intramoléculaires des ARN. Il a été démontré qu'il était possible d'identifier ces structures intramoléculaires ainsi que les interactions à longue distance entre les ARN par cristallographie au rayon X ou par microscopie Cryo-EM, mais ces techniques ne permettent pas le haut débit (Feigon, 2015).

Les techniques aux psoralènes permettent d'étudier les interactions du transcriptome sans l'usage de protéines particulières, mécanisme utilisé notamment lors des méthodes hiCLIP et CLASH (Helwak *et al*, 2013; Sugimoto *et al*, 2015), et avec plus de profondeur, comparé à la méthode RNP (technique basée sur la ligation de proximité des ARN), par exemple (Ramani *et al*, 2015; Lu *et al*, 2016)

Comme vu précédemment, il est également possible d'étudier les interactions ARN-ARN, leurs structures intramoléculaires et longue distance par méthodes bio-informatiques. Cependant, ces prédictions varient en précision, deviennent complexes lorsque les interactions autorisent des mésappariements, et dans tous les cas, nécessitent une validation expérimentale (Sharma *et al*, 2016). Cependant, malgré l'utilisation des techniques LIGR-Seq, PARIS, SPLASH et COMRADES, la fraction de snoARN orphelins reste toujours aussi importante.

Ces méthodes semblent avoir une limite de détection lors de l'étude d'ARN cellulaires moins abondants où leurs interactions sont transitoires, tel que les interactions de snoARN.

En ce sens, le développement de nouvelles méthodes, permettant d'étudier les interactions ARN-ARN *in-vivo*, spécifiquement dirigées sur l'étude des snoARN, constitue un axe majeur pour pouvoir déterminer les cibles des snoARN orphelins. Cet axe constitue une thématique importante de l'équipe de Tamas KISS, et qui a fait partie intégrante de mon projet de doctorat.

RÉSULTATS

I. Mise au point d'une méthode permettant de co-purifier un snoARN avec sa cible

I. 1. Mise au point de la méthode

I. 1. 1. Méthodes développées dans la littérature

Dans l'objectif de déterminer les ARN cibles des snoARN orphelins, mon projet de thèse, au sein de l'équipe de Tamas KISS, a consisté à mettre en place une méthode permettant de co-purifier les snoARN avec leurs cibles. Comme vu précédemment, différentes techniques étudiant les interactions ARN-ARN ont préalablement vu le jour dans la bibliographie (Kudla *et al*, 2011; Aw *et al*, 2016; Lu *et al*, 2016; Nguyen *et al*, 2016; Sharma *et al*, 2016; Ziv *et al*, 2018). Ces méthodes se sont retrouvées être très pertinentes, notamment lors de l'étude d'ARN qui ont une abondance relativement élevée. En effet, de nombreuses interactions déjà établies entre les snoARN et les ARNr/snARN ont été retrouvées par ces méthodes, gage de leur efficacité. En revanche, lors de l'analyse d'ARN ayant une abondance plus faible, comme étant le cas de certains snoARN orphelins, ces méthodes n'ont malheureusement pas pu établir une liste précise de potentiels ARN ciblés par ces snoARN orphelins. Le LIGR-Seq a tout de même réussi à identifier les ARNm *NOP14*, *RPS5*, et *SRSF3* comme étant ciblés par le snoARN orphelin SNORD83, mais aucune expérience complète de validation n'a été publiée ensuite, et ce après plus de sept ans (Sharma *et al*, 2016).

L'étape de crosslink, utilisée dans les méthodes précédemment citées, est cruciale pour pouvoir identifier les interactions ARN-ARN, mais elle apporte à contrario un lot d'interactions faux positives conséquent, par le biais de crosslinks non-spécifiques entre différents ARN, qui résultent de la capture de molécules d'ARN avoisinantes sans qu'il y ait d'appariements physiques entre elles. Le problème majeur qui semble ressortir lors de l'utilisation de ces différentes méthodes est qu'il est difficile d'arriver à détecter une interaction ARN-ARN ayant une faible fréquence dans la cellule parmi les interactions non-spécifiques par les crosslinks.

Le fait que les ARNr, ARN cellulaires fortement exprimés, aient été retiré lors du LIGR-Seq, a dû réduire la quantité de crosslink non spécifique, permettant ainsi aux auteurs d'établir une liste d'ARN cibles candidats pour le snoARN orphelin SNORD83B (Sharma *et al*, 2016).

Le concept de base imaginé par l'équipe d'accueil afin de contourner ces limites, implique une co-purification d'un seul snoARN et de sa ou ses cibles, et ce sans passer par

l'intermédiaire de protéines appâts, à l'inverse du CLASH ou du MARIO (Kudla *et al*, 2011; Nguyen *et al*, 2016). J'ai donc été chargé, au cours de mon doctorat, de mettre en place les différentes étapes composant cette méthode. Pour ce faire, nous nous sommes appuyés sur les méthodes utilisant le crosslink ARN-ARN via le psoralène (Aw *et al*, 2016; Lu *et al*, 2016; Sharma *et al*, 2016; Ziv *et al*, 2018), en les détournant, de manière à optimiser la méthode sur un seul snoARN.

I. 1. 2. Les différentes étapes composant la méthode développée

De manière à simplifier le discours qui va suivre, la méthode que nous essayons de mettre au point sera nommée PIKA (Psoralen In vivo Crosslink RNA-RNA). Les différentes étapes qui la composent sont représentées sur la **Figure 33**. La première étape est un crosslink ARN-ARN, *in vivo*, effectué via l'ajout d'un dérivé du psoralène, suivi d'une exposition aux UV à 365nm. Les interactions physiques entre les snoARN d'intérêts et leurs cibles étant figées, les ARN totaux sont extraits des cellules. Le snoARN d'intérêt est spécifiquement purifié grâce à des billes possédant des oligonucléotides antisens complémentaires. Grâce à l'étape de crosslink, la ou les cibles du snoARN d'intérêt seront co-purifiées. La fraction snoARN/ARN cible est soumise à la dégradation partielle par une endonucléase coupant spécifiquement dans les duplex ARN-ARN. Les fragments d'ARN générés sont liés par une ligation de proximité, permettant ainsi la synthèse d'ARN chimériques, qui sont composés d'un fragment du snoARN d'intérêt et d'un fragment de son ARN cible. Cette étape permet d'éviter de perdre les informations d'interactions du snoARN et de sa cible lors de l'inversion du crosslink. L'inversion du crosslink est par ailleurs effectuée à 254nm, et permet la génération d'une librairie d'ADNc à partir des ARN obtenus. La librairie d'ADNc est ensuite analysée par séquençage haut-débit.

La réflexion et la mise au point de ces différentes étapes sont développées dans le chapitre I/3. Avant cela, voici les ARN dont nous nous sommes servis afin de mettre au point la méthode PIKA.

I. 2. ARN étudiés pour la mise au point de la méthode

I. 2. 1. SCARNA10

SCARNA10, également connu sous le nom d'U85, est un scaARN hybride de 330nt. La scaRNP est composée à la fois des protéines cœurs des snoARN H/ACA et des snoARN C/D (Jády & Kiss, 2001). En ce sens, SCARNA10 possède les boîtes consensus H, ACA, C/D et

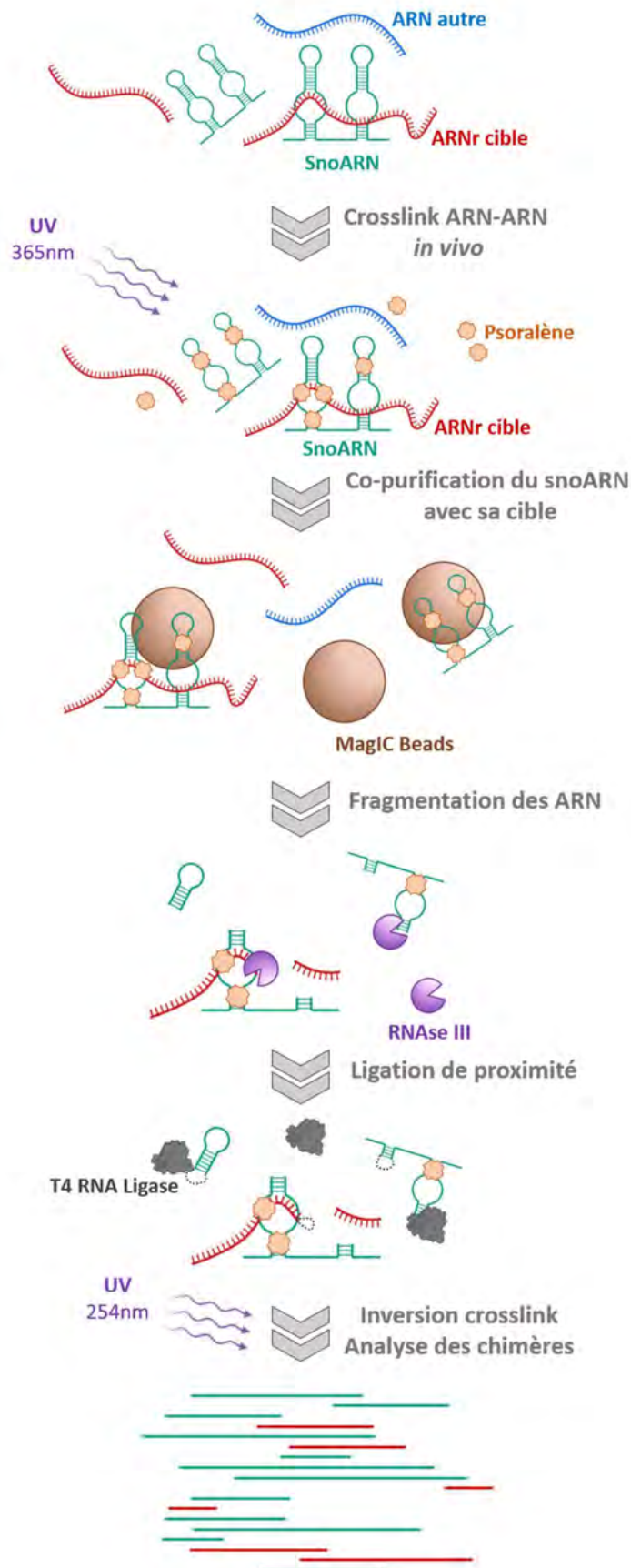


Figure 33 : Représentation schématique de la méthode PIKA

Les cellules sont incubées avec de l'AMT psoralène et soumises à une irradiation aux UV d'une longueur d'onde 365nm. Un snoARN d'intérêt est spécifiquement purifié à l'aide d'oligonucléotides complémentaires immobilisés sur des billes magnétiques.

Grâce au crosslink, la cible du snoARN se retrouve co-purifiée. Les ARN purifiés subissent une digestion partielle spécifiquement dirigée contre les parties double brin. Une ligation de proximité permet de lier les fragments du snoARN avec ceux de la cible, formant des ARN chimériques. Suite à l'inversion du crosslink, les ARN servent de template pour la génération d'une librairie d'ADNc, qui est séquencé à haut débit. L'analyse des séquences d'ARN chimériques par bio-informatique permet de déterminer une liste d'ARN cibles candidats interagissant physiquement avec le snoARN étudié.

C'/D' (**Figure 34A**). Par Northern Blot, les auteurs ont révélé la présence de SCARNA10 chez l'homme, la souris, et la drosophile, montrant une forte conservation de cet ARN au cours de l'évolution. Également, la structure du SCARNA10 de la drosophile est prédite pour être identique à celle de l'humain. Les boîtes C/D sont des éléments localisés aux extrémités de l'ARN et par conséquent, nécessaires à l'accumulation de SCARNA10 dans les cellules, contrairement aux boîtes H/ACA. En revanche, les boîtes H/ACA jouent un rôle primordial dans l'assemblage de la partie H/ACA du snoRNP. De manière intéressante, SCARNA10 guide une méthylation en 2'-O-ribose sur le site C45 de l'ARN de l'épissage U5, ainsi qu'une pseudouridylation sur le site adjacent, U46, du même ARN (**Figure 34B**). Nous avons spécifiquement choisi cet ARN en tant que contrôle positif lors de la mise en place de PIKA. Notre choix s'est porté sur le SCARNA10 car (i) Il est possible de le détecter facilement par Northern Blot dans les lignées cellulaires humaines (Jády & Kiss, 2001), ce qui traduit une forte expression de cet ARN dans les cellules (ii) SCARNA10 guide deux modifications chimiques sur un ARN cible qui a été validé (snU5), permettant ainsi de vérifier si lors de l'utilisation de PIKA une interaction SCARNA10 – snU5 est bien retrouvée (iii) Les deux modifications chimiques guidées par SCARNA10 se réalisent sur le même ARN cible. L'interaction entre un snoARN et sa cible est un événement relativement bref dans le temps. Guidant deux modifications différentes sur un même ARN, il est hautement probable que le nombre d'interactions entre SCARNA10 et l'ARN cible snU5 soit multiplié par deux dans la cellule, augmentant ainsi la probabilité de co-purifier SCARNA10 avec sa cible.

D'après la base de données RISE, l'interaction SCARNA10 – snU5 a été retrouvée dans la base de données RAIN (RNA-protein Association and Interaction Networks, Junge *et al.*, 2017), mais pas lors de l'utilisation des méthodes de crosslink au psoralène (Gong *et al.*, 2018)(**Figure 34C**). La base de données RAIN est composée d'interactions protéines – ARN et ARN – ARN venant d'interactions (i) prédites *in silico* par différents logiciels, principalement centrés sur les microARN (ii) validées expérimentalement, par notamment des expériences de CLIP, CRAC et CLASH (iii) connues dans la littérature scientifique. En revanche, les méthodes PARIS et LIGR-Seq ont relevés les interactions SCARNA10 – ARNr 5,8S et SCARNA10 – SNORA48, respectivement. Pourtant, SCARNA10 semble posséder seulement une complémentarité avec snU5 (Jády & Kiss, 2001). La base de données RAIN montre en supplément de nombreuses interactions entre SCARNA10 et d'autres snoARN et scaARN.

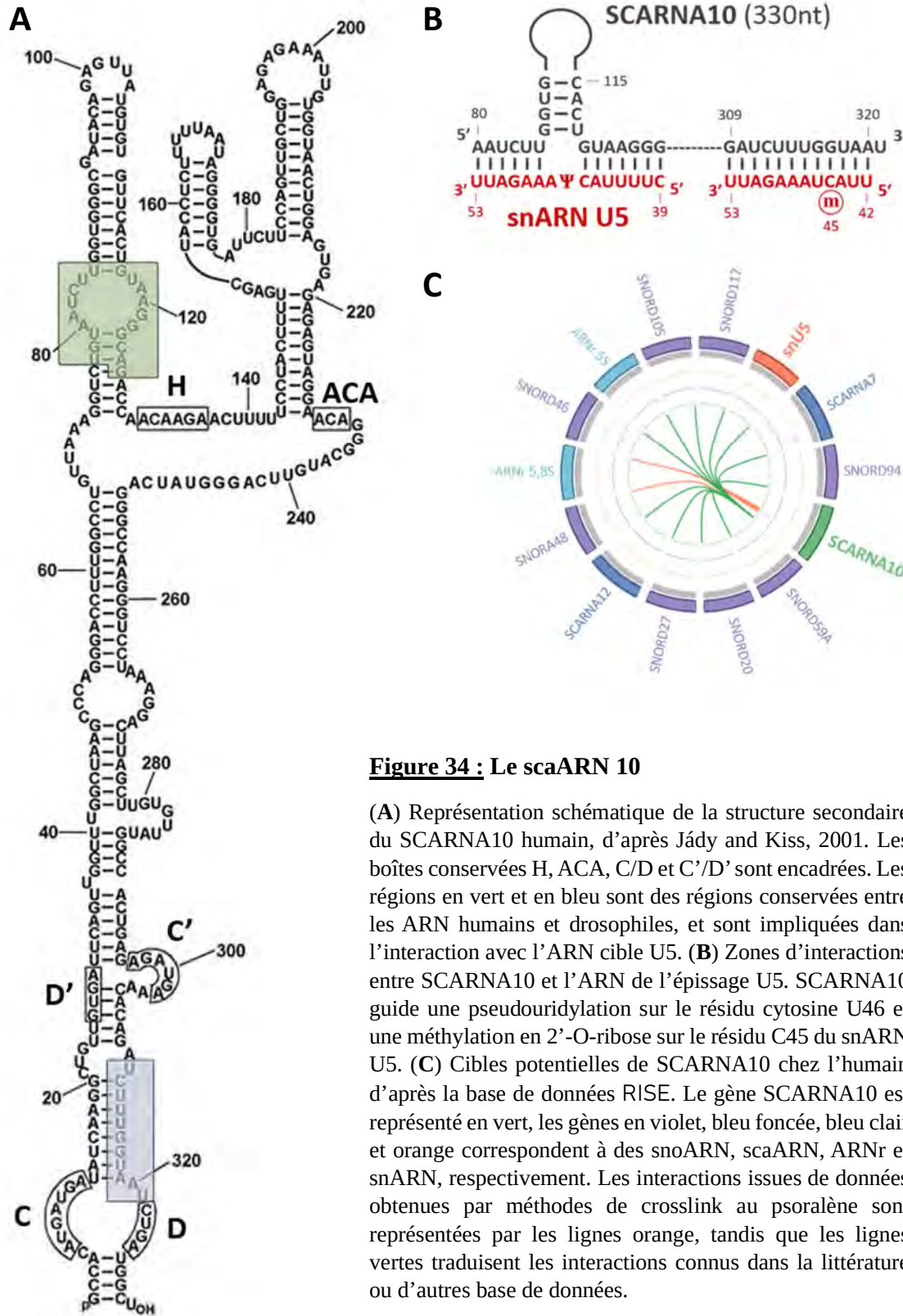


Figure 34 : Le scaARN 10

(A) Représentation schématique de la structure secondaire du SCARNA10 humain, d'après Jádý and Kiss, 2001. Les boîtes conservées H, ACA, C/D et C'/D' sont encadrées. Les régions en vert et en bleu sont des régions conservées entre les ARN humains et drosophiles, et sont impliquées dans l'interaction avec l'ARN cible U5. (B) Zones d'interactions entre SCARNA10 et l'ARN de l'épissage U5. SCARNA10 guide une pseudouridylation sur le résidu cytosine U46 et une méthylation en 2'-O-ribose sur le résidu C45 du snARN U5. (C) Cibles potentielles de SCARNA10 chez l'humain d'après la base de données RISE. Le gène SCARNA10 est représenté en vert, les gènes en violet, bleu foncée, bleu clair et orange correspondent à des snoARN, scaARN, ARNr et snARN, respectivement. Les interactions issues de données obtenues par méthodes de crosslink au psoralène sont représentées par les lignes orange, tandis que les lignes vertes traduisent les interactions connus dans la littérature ou d'autres base de données.

I. 2. 2. SNORA55

SNORA55 est un snoARN H/ACA de 137nt. Il possède les boîtes consensus H et ACA. Ces boîtes consensus, ainsi que ses éléments antisens des tiges-boucles 5' et 3' sont fortement conservées au cours de l'évolution (**Annexe 1**). D'après la base de données snoRNA-LBME-db, l'élément antisens au niveau de la tige-boucle en 5' de l'ARN est prédite pour s'apparier sur l'ARNr 18S (Lestrade & Weber, 2006). Le résidu U36 de l'ARN 18S serait le nucléotide cible de la modification chimique (**Figure 35A**).

Nous avons sélectionné ce snoARN comme ARN expérimental pour la mise au point de la méthode, principalement car (i) l'élément antisens de la tige-boucle en 5' du snoARN est prédit *in silico* pour guider une pseudouridylation sur l'ARNr 18S, interaction que nous allons pouvoir vérifier lors de la mise en place de la technique (ii) l'élément antisens de la tige-boucle en 3' est orphelin, la méthode nous permettra potentiellement d'établir une liste d'ARN cibles candidats (iii) les éléments antisens de SNORA55 sont conservés inter-espèces. Nous supposons donc que ces éléments antisens soient fonctionnels et que des fonctions leurs sont associées.

Une étude a mis en évidence que SNORA55 est surexprimé dans des xénogreffes dérivées de tissus tumoraux de patients atteints du cancer de la prostate (Crea *et al*, 2016). Également, SNORA55 s'est révélé être nécessaire lors des mécanismes de prolifération et de migration d'une lignée cellulaire issus de cancer de la prostate. Les auteurs ont mis en évidence que l'accumulation de SNORA55 est seulement localisé dans les noyaux cellulaires (Crea *et al*, 2016).

D'un autre côté, aucune méthode étudiant les interactions ARN-ARN *in vivo*, par crosslink au psoralène ou par protéine appâts, n'ont établi d'ARN candidats pouvant être la cible de SNORA55, d'après la base de donnée RISE (Gong *et al*, 2018). La base de données RAIN en revanche, évoque l'interaction entre SNORA55 et l'ARNr 18S (Junge *et al*, 2017). Par contre, l'interaction prédite est différente de celle faite par la base de données snoRNA-LBME-db. L'interaction prédite concerne cette fois-ci la tige-boucle située au niveau de la région 3' de SNORA55, qui guiderait une pseudouridylation putative sur le résidu U681 de l'ARN 18S. A première vue, cette interaction paraît moins pertinente que celle prédite par snoRNA-LBME-db, l'élément antisens au niveau de la région 3' du snoARN présentant de nombreux mésappariements avec la région de l'ARNr 18S contenant l'Uridine 681 (**Figure**

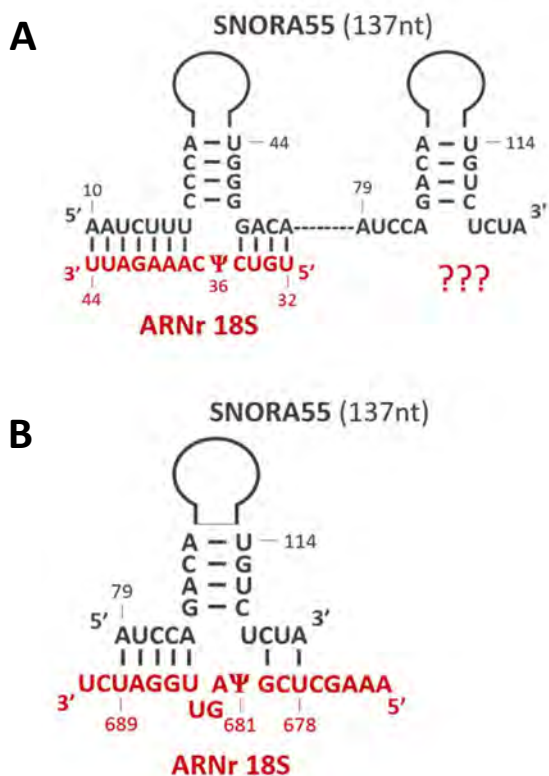


Figure 35 : Le snoARN 55

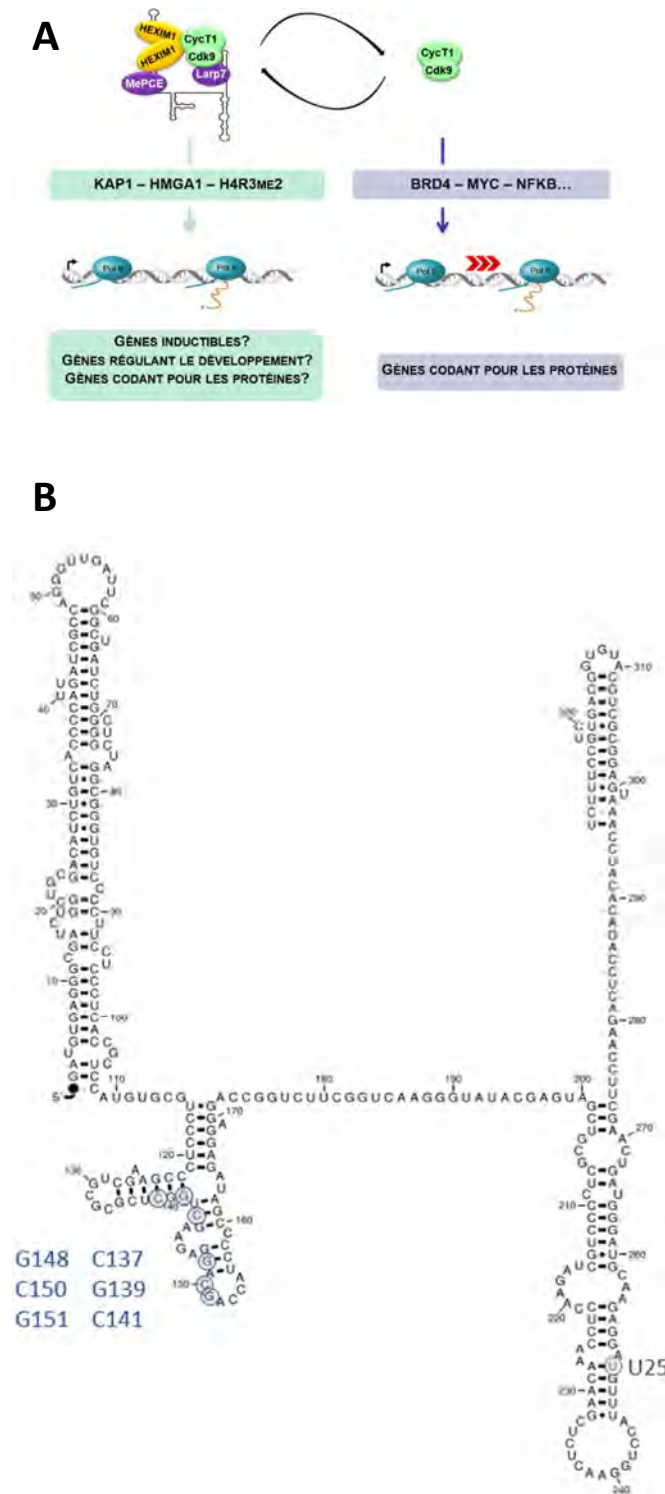
Représentation schématique des interactions prédites entre SNORA55 et l'ARNr 18S. **(A)** Interaction de la tige-boucle en 5' de SNORA55 avec l'ARNr 18S, prédite par la base de données snoRNA-LBME-db (Lestrade and Weber, 2006). **(B)** Interaction, présentant un appariement imparfait, de la tige-boucle en 3' de SNORA55 avec l'ARNr 18S, prédite par la base de données RAIN (Junge *et al.*, 2017).

35B). Cependant, cette interaction n'est pas à réfuter, il n'est pas rare que les interactions entre un snoARN et sa cible autorisent quelques mésappariements (Tafer *et al.*, 2010).

I. 2. 3. Le snARN 7SK

Le snARN 7SK est un ARN non codant de 331nt, très abondant au sein des cellules, ainsi que très conservé au cours de l'évolution. 7SK, transcrit par l'ARN polymérase III, est un régulateur de la transcription médiée par l'ARN polymérase II en contrôlant l'activité du facteur d'élongation P-TEFb (pour revue : (Egloff *et al.*, 2018)) (**Figure 36A**). 7SK est connu pour porter une pseudouridylation et six méthylations en 2'-O-ribose (**Figure 36B**). Il a été montré dans la littérature que la pseudouridine, située à la position U250, est catalysée par la snoRNP SNORA70 (Zhao *et al.*, 2016). A ce jour, c'est la seule interaction connue entre 7SK et un autre ARN cellulaire. Par ailleurs, les six 2'-O-ribose méthylations, respectivement situées aux sites C137, G139, C141, G148, C150 et G151, ne sont pas encore associées à des snoRNP C/D (Kishore *et al.*, 2013). Nous avons choisi d'étudier cet ARN pour la mise en place de la méthode PIKA, car (i) il possède une abondance plus élevée au sein des cellules que les snoARN (plus de 200.000 copies par cellule), ce qui permet de faciliter la mise au point technique des différentes étapes de PIKA (ii) l'interaction avec SNORA70 pourra potentiellement être retrouvée, confirmant ainsi cette interaction (iii) une liste de SNORD candidats associés aux méthylations en 2'-O-ribose pourra être établie (iv) L'ARN 7SK constituant le second axe de recherche de mon équipe d'accueil, des outils génétiques sont déjà disponibles et les résultats générés pourront alors être valorisés au sein de l'équipe (Studniarek *et al.*, 2021).

Étonnamment, un nombre important d'interactions entre 7SK et d'autres éléments du transcriptome apparaissent lors de l'étude de 7SK avec la base de données RISE (**Figure 37**). Aucune interaction entre 7SK et des snoARN H/ACA n'a été identifiée. En revanche, les SNORD12C, 30, 113-1, 115 et 116 apparaissent comme interagissant avec 7SK. Ces interactions ne sont détectées uniquement dans la base de données RAIN, et ne sont pas retrouvées via les méthodes de crosslink aux psoralènes et par protéines appâts. La méthode au psoralène PARIS a identifié de nombreuses interactions de 7SK avec des ARNm et des ARN non codants (Lu *et al.*, 2016; Junge *et al.*, 2017; Gong *et al.*, 2018).

**Figure 36 : Le snARN 7SK**

(A) L'ARN 7SK est représenté en noir. La snoRNP 7SK est composée, en supplément de l'ARN, des protéines MePCE, LARP7, HEXIM1, ainsi que du facteur d'élongation de la transcription ARN Pol II (P-TEFb). P-TEFb est composé de Cdk9 et de son partenaire CycT1 ou CycT2. Les formes actives et inactives de P-TEFb sont recrutées sur les promoteurs de leurs gènes cibles respectifs via association avec des facteurs de transcription spécifiques. Figure issue de (Egloff *et al.*, 2018). (B) La pseudouridine est surlignée en vert, et les méthylations en 2'-O-ribose en bleu. Structure secondaire de l'ARN 7SK issue de (Leger *et al.*, 2021).

I. 3. Mise au point des étapes composant la méthode

I. 3. 1. Crosslink ARN-ARN *in vivo*

La première étape est composée d'un crosslink ARN-ARN effectué grâce au psoralène, directement sur des cellules humaines en culture. L'activation du psoralène se fait par une irradiation aux UV d'une longueur d'onde de 365nm, permettant ainsi de figer les interactions ARN-ARN à l'instant *t in vivo* (Calvet & Pederson, 1979). La méthode COMRADES utilise un psoralène couplé à un groupement azide, permettant une purification des hybrides ARN-ARN ultérieure. L'utilisation de ce type de psoralène limite en revanche le temps d'exposition des cellules aux UV, de par la dégradation rapide du composé azide sous UV (Ziv *et al*, 2018). En ce sens, l'exposition aux UV se fait sur 10 mins pour la méthode COMRADES, contre 30 mins pour le LIGR-Seq, qui elle utilise un dérivé classique du psoralène (4'-aminomethyltrioxsalen)(Sharma *et al*, 2016). Pour la mise en place de PIKA, l'utilisation d'un psoralène couplé avec un groupement azide nous semblait moins intéressante. En effet, les interactions snoARN-ARN cible sont des événements limités dans le temps, et risque d'être très peu fréquents, notamment dans le cas où le snoARN étudié est faiblement abondant. Une réduction du temps d'exposition aux UV générera moins de crosslink ARN-ARN au sein des cellules, diminuant les chances de détection d'une interaction snoARN-ARN. Dans le cas contraire, une exposition trop longue des cellules aux UV peut engendrer des complications : (i) cela pourrait induire une dégradation des ARN, ainsi qu'entraîner des modifications chimiques de ses bases (Wurtmann & Wolin, 2009) (ii) générer une quantité importante de crosslink ARN-ARN, enrichissant les échantillons en crosslink non spécifique, pouvant donner lieu à des faux positifs. En supplément, plus les ARN forment de crosslink dans la cellule, plus ils se retrouvent associés à de larges complexes ribonucléoprotéiques.

Lors de l'étape de l'extraction des ARN cellulaires, si la quantité des larges complexes est trop importante, une grande proportion des ARN crosslinkés pourrait être perdue pendant l'étape de purification d'ARN (**Figure 38A**). Ces larges complexes, en raison d'une solubilité réduite, pourrait se retrouver dans l'interphase au lieu de la phase aqueuse lors des étapes de purification des ARN qui suivent l'étape de Crosslink ARN-ARN, engendrant une perte de matériel.

Pour notre méthode, lors d'une réaction de crosslink *in vivo*, nous avons décidé d'utiliser 2.10^7 cellules avec 0,375 µg/µL du dérivé AMT du psoralène dans un volume de 600µL. Après avoir testé plusieurs conditions de crosslink, nous avons choisi d'effectuer l'irradiation à 365nm

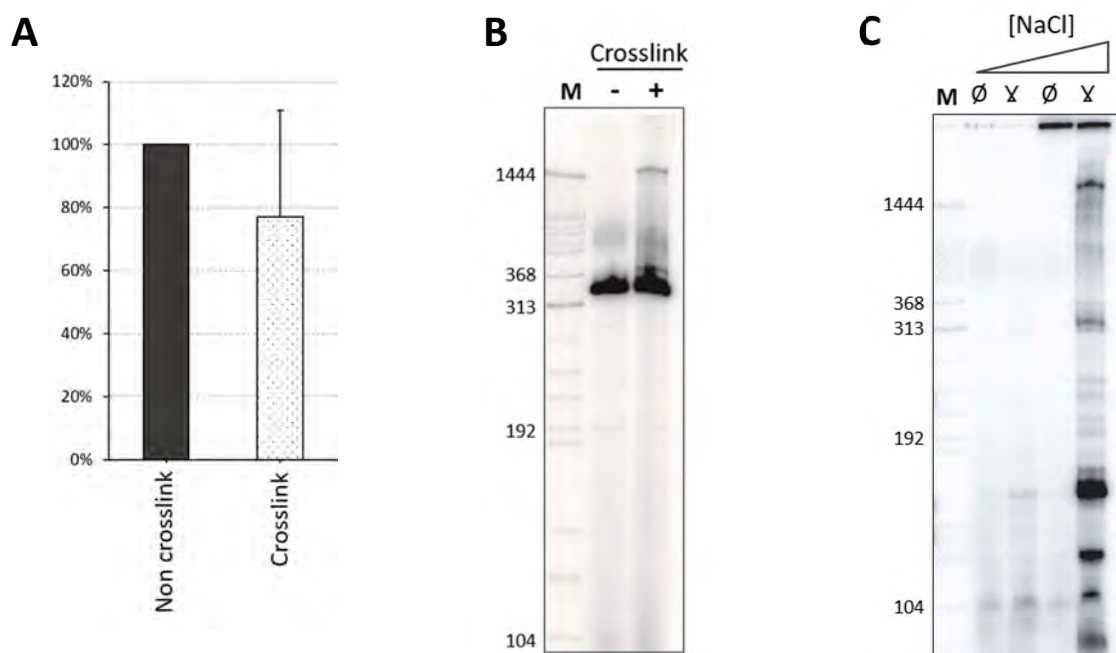


Figure 38 : Mise au point de l'étape de crosslink ARN - ARN

(A) Test de la perte de matériel (représentée en %) engendrée par l'étape de Crosslink au psoralène. Les cellules ont été crosslinkées avec ou sans 0,375µg/µL de psoralène. Les quantités des ARN 7SK, SCARNA10, snU5 et snU6 ont été analysées par qRT-PCR. (B) Test de l'efficacité du crosslink. 2.10^7 cellules sont incubées ou non avec 225 µg d'AMT psoralène et irradiées 20 mins à une longueur d'onde de 365nm. Après purification de l'ARN 7SK, les extrémités 3' des ARN sont radiomarqués au pCp, puis séparer en fonction de leur taille. (C) Test d'un crosslink avec du psoralène biotinylé. Les cellules sont incubées avec 0.01% de digitonine, puis subissent une étape de crosslink incluant un psoralène biotinylé. Une étape de purification est effectuée avec des billes de streptavidine. Les éluas sont radiomarqués au pCp. L'expérience a été réalisée avec deux conditions en NaCl différentes, 200 et 700mM. (Ø) et (Y) représentent respectivement les échantillons non-crosslink et crosslink, qui diffère sur la présence ou non de psoralène.

sur glace pendant 20 mins, avec les cellules étant à 1 cm de la lampe UV. En parallèle, des cellules subissent le même protocole, mais sans l'ajout de l'AMT psoralène, afin d'avoir un contrôle négatif de crosslink.

Pour déterminer l'efficacité de nos paramétrages de l'étape de crosslink, ainsi que de la conservation de l'intégrité des ARN, l'ARN 7SK a été purifié à partir de cellules humaines HeLa ayant subi ou non un crosslink à l'AMT psoralène, puis migré sur gel acrylamide. Un marquage au pCp a ensuite été effectué (**Figure 38B**). Ce marquage consiste en l'ajout d'une cytidine bis phosphate radiomarkée au phosphate 32 (^{32}P) à l'extrémité 3' hydroxyle des ARN présents dans l'échantillon. Cette expérience permet de visualiser la taille et l'abondance des ARN présents dans les échantillons. Dans les deux échantillons, une forte bande apparaît entre les marqueurs de taille de 313 et 368nt, correspondant à l'ARN 7SK (331nt). Une seconde bande, beaucoup plus légère, apparaît vers 192nt. Etant présente dans les deux échantillons, elle ne résulte pas de l'effet du crosslink, mais peut représenter des ARN partiellement dégradés de 7SK. Un signal diffus est observé entre 500 et 430 nt, également dans les deux échantillons. Il est indépendant des acides nucléiques, et est une signature des billes utilisées lors de la purification (données mises en avant ultérieurement, voir **Figure 39E**). En revanche, au niveau de la piste avec crosslink, des bandes supplémentaires, relativement nettes, apparaissent vers 380 et 1444nt. Ces deux bandes, à l'instar des bandes diffuses situées au-dessus de 7SK, correspondent très certainement à des complexes formés par 7SK. Nous avons donc conclu qu'une irradiation de 20 minutes des cellules dans une solution contenant 0,375 $\mu\text{g}/\mu\text{L}$ d'AMT psoralène est la condition idéale pour induire des pontages covalents entre des ARN nucléaires non-codants *in vivo*.

En parallèle, nous avons essayé de mettre en place un crosslink au psoralène biotinylé, pour enrichir les échantillons en duplex ARN-ARN, à l'instar du SPLASH (Aw *et al*, 2016). Après un traitement à la digitonine à 0,01%, les cellules HeLa ont été incubées avec du psoralène biotinylé et irradiées durant 20 mins au UV à une longueur d'onde de 364nm. Après extraction des ARN, les ARN totaux crosslinkés ou non, sont incubés avec des billes streptavidine. L'immobilisation des ARN biotinylés sur les billes a été effectuée dans un des deux tampons différents contenant soit 200 mM or 700 mM NaCl. Les ARN retenus par les billes ont été radiomarkés au pCp puis séparés sur gel acrylamide (**Figure 38C**). Après révélation, aucune bande n'apparaît dans les échantillons contrôles au sein desquels le psoralène n'a pas été ajouté (pistes Ø), et ce, indépendamment de la concentration saline. En revanche, lors du crosslink (Y) réalisé avec un taux bas de NaCl, seuls des signaux très faibles sont

observés aux alentours de 110 et 170nt. En revanche, lors d'une forte concentration en sel, de nombreuses bandes, beaucoup plus intenses, sont observées sur la piste γ .

Il semblerait qu'une augmentation de la concentration saline lors de l'incubation des ARN biotinylés avec des billes streptavidine permet d'augmenter l'accessibilité de la biotine au sein des complexes d'ARN crosslinkés.

Nous n'avons pas essayé le crosslink avec l'utilisation d'un psoralène portant un composé azide, car cette option impacte le temps d'irradiation des échantillons lors du crosslink, et donc sa performance (Ziv *et al*, 2018).

Comme vu précédemment avec la **Figure 33** nous sommes restés sur l'utilisation de l'AMT psoralène pour le développement de la technique PIKA, au détriment du psoralène biotinylé. Cela pour les raisons suivantes **(i)** Le psoralène biotinylé n'étant pas perméable dans les cellules, l'utilisation d'un détergeant doux est alors requis (Aw *et al*, 2016). Une perturbation des membranes phospholipidiques des cellules peut affecter le fonctionnement cellulaire et ainsi induire un biais lors de l'analyse. **(ii)** L'utilisation du psoralène biotinylé implique une purification supplémentaire par des billes streptavidine, étape qui s'ajoute à la purification de l'ARN d'intérêt. Les snoARN ayant une faible abondance au sein des cellules, une étape composée de deux purifications successives nous semblait moins pertinente, car cela engendre une perte non négligeable de matériel biologique.

I. 3. 2. Purification d'un snoARN d'intérêt : utilisation des MagIC Beads

La seconde étape de la méthode PIKA consiste à purifier un snoARN d'intérêt, lié à ses ARN cibles, parmi les différents duplex ARN – ARN formés lors de la précédente étape de crosslink. Généralement, la purification d'un ARN est effectuée par l'intermédiaire d'un oligo biotinylé et de billes streptavidine. Ce protocole a l'avantage d'être efficace, et ce avec un moindre coût. Au début de ma thèse, nous étions logiquement partis sur ce principe. Cependant, le bruit de fond généré par l'utilisation de billes streptavidine s'est révélé trop élevé, comme le montre cette expérience au pCp (**Figure 39**, piste 1). Un bruit de fond trop élevé risque de rendre l'identification des ARN co-purifiées avec le snoARN d'intérêt difficile. Pour pallier à cette difficulté, nous nous sommes tournés sur un moyen de purification des ARN alternatif. La purification du snoARN lors de l'étape 2 de PIKA se fera via l'utilisation des MagIC Beads. Les MagIC Beads sont des billes produites et développées sur mesure par Zero Elements Biolabs (<https://elementzero.bio/>). Ces billes magnétiques possèdent à leur surface plusieurs

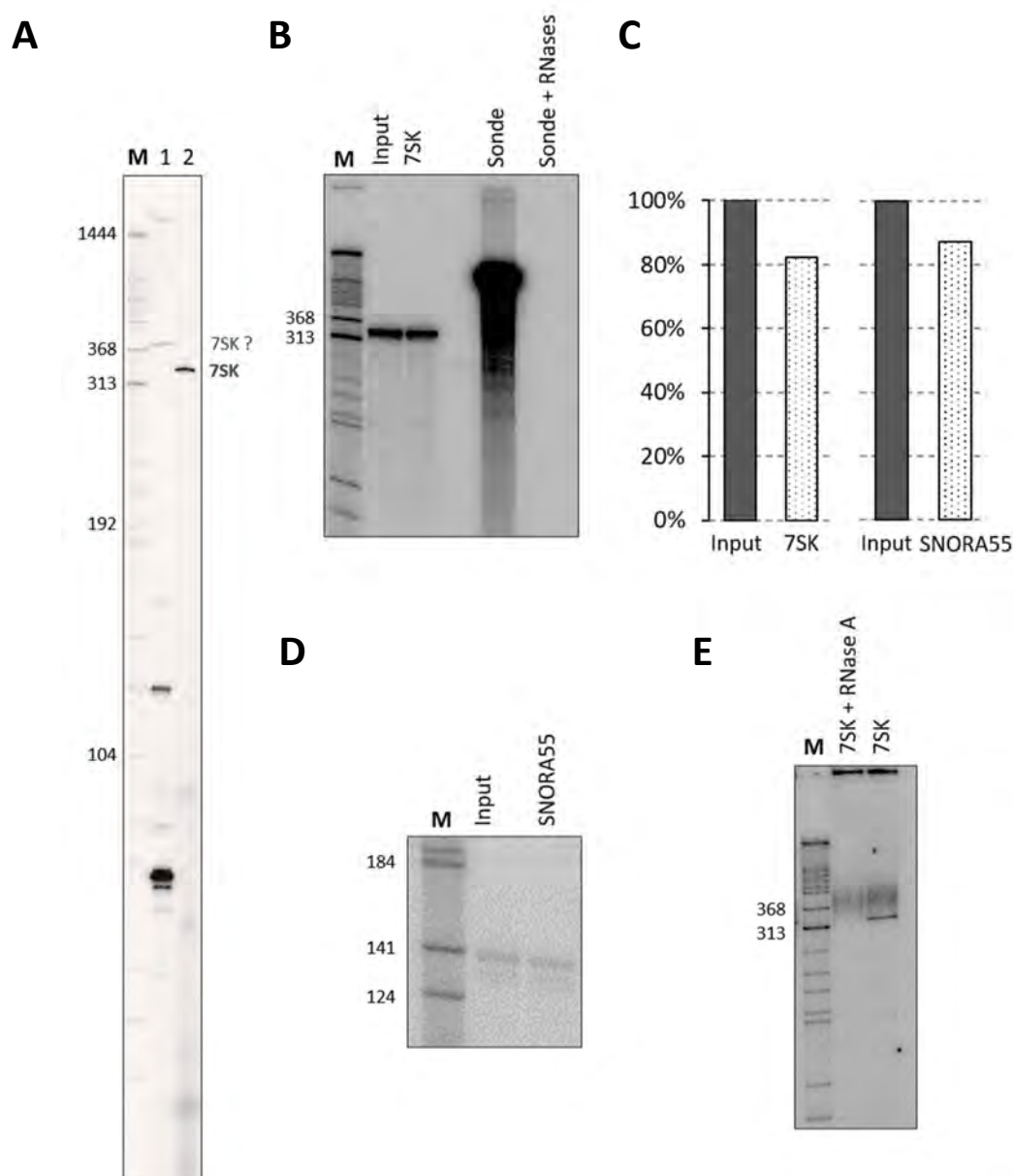


Figure 39 : Mise au point de l'étape de purification d'un snoARN d'intérêt

(A) Radio-marquage au pCp des ARN purifiés par oligonucléotide biotinylé complémentaire à l'ARN 7SK (1) ou par MagIC Beads (2). (B) Mapping aux RNases avec une sonde complémentaire l'ARN 7SK dans des extraits d'ARN totaux et dans un échantillon de 7SK purifié. La sonde seule a été déposée, et sa dégradation par les RNases A/T1 démontrée. (C) Taux de purification des MagIC Beads spécifiques de 7SK, SNORA55 et SCARNA10 par rapport aux extraits d'ARN totaux à partir desquels les purifications ont été effectuées (Input). (D) Mapping aux RNases réalisé avec une sonde complémentaire à SNORA55, dans des extraits d'ARN totaux et dans une fraction purifiée de SNORA55 avec les MagIC Beads. (E) Radio-marquage au pCp d'une fraction purifiée de 7SK par les MagIC Beads et d'une même fraction digérée à la RNase A.

oligonucléotides d'ADN, qui ciblent différentes régions d'un seul et même ARN. Désignées sur mesure, ces billes présentent un coût plus élevé que l'utilisation des billes streptavidine, mais présentent en revanche une efficacité et une spécificité plus élevées (**Figure 39**, piste 2).

Les efficacités des MagIC Beads contre 7SK, SNORA55 et SCARNA10 ont été testées par protection aux RNases A/T1, et leurs spécificités par radio-marquage au pCp. La détection des ARN par protection aux RNases consiste à synthétiser *in vitro* un ARN antisens radiomarké au ^{32}P et complémentaire à un ARN d'intérêt. Suite à l'incubation de cet ARN antisens avec un échantillon d'ARN contenant celui d'intérêt, une digestion avec les nucléases A et T1 est effectuée. Ces nucléases dégradant les ARN simples brins, une partie de la sonde radioactive se retrouve alors protégée par l'hybridation de l'ARN d'intérêt, permettant ainsi sa détection après séparation des ARN protégés sur un gel polyacrylamide. De ce fait, la protection aux RNases possède une sensibilité plus élevée que le Northern Blot. Lors de la détection de 7SK dans un extrait d'ARN totaux, nous observons une forte bande entre 313 et 368 nt, correspondant à la taille de 7SK (**Figure 39B**). Cette même bande apparaît lors d'une purification en amont de 7SK via les MagIC Beads à partir d'un extrait d'ARN totaux. Les MagIC Beads permettent alors de purifier spécifiquement un ARN d'intérêt, avec une efficacité très honorable. En effet, 82.37% des ARN 7SK présents dans l'Input ont été purifiés, en concentration non limitante de billes (**Figure 39C**). De la même manière, les expériences de protection aux RNases ont démontré que les MagIC Beads désignées contre SNORA55 et SCARNA10 sont capables de purifier avec une efficacité très similaires ($\approx 80\%$) des ARN à partir des ARN cellulaires. Nous avons cependant remarqué que malgré l'utilisation de fortes concentrations d'ARN totaux, le signal de SNORA55 reste faible, traduisant une abondance relativement faible au sein des cellules (**Figure 39D**).

Les MagIC Beads nous permettent donc de purifier nos snoARN d'intérêts avec une forte efficacité. En parallèle, nous avons déterminé la quantité maximale d'ARN totaux pour une quantité de billes donnée, afin d'optimiser le rendement (données non montrées). Ainsi, lors de la seconde étape de la méthode PIKA, les billes seront en concentrations non-limitantes pour la quantité d'ARN d'intérêt.

De manière à déterminer le bruit de fond des MagIC Beads, les radiomarquages au pCp des ARN purifiés ont été effectués pour chaque type de billes. Les résultats concernant les MagIC Beads dirigées contre le snARN 7SK peuvent être observés sur la **Figure 39A**. Aucune bande précise n'apparaît en supplément de celle observée vers 330nt, qui correspond à l'ARN 7SK purifié. Quelques bandes diffuses semblent en revanche apparaître avec des tailles

inférieures à 100nt. Cependant, après purification via les MagIC Beads, une tâche diffuse apparaissant entre 350 et 500nt est détectée à chaque expérience de radiomarquage au pCp (**Figure 39E**). Afin de vérifier si ce signal venait d'acides nucléiques, un radiomarquage au pCp d'une purification de 7SK suite à un traitement aux DNases et RNases a été réalisé. Cette expérience a mis en évidence que cette tâche n'était pas le résultat de présence des acides nucléiques, mais une signature des MagIC Beads et d'autres molécules contaminantes étaient à l'origine de leur présence. (**Figure 39E**).

L'utilisation des MagIC Beads permet donc de purifier un ARN d'intérêt avec une forte efficacité et avec une faible co-purification d'autres ARN non spécifiques.

I. 3. 3. Validation intermédiaire de la méthode PIKA par RT-qPCR

Afin de savoir s'il était possible de co-purifier un snoARN avec sa cible via l'utilisation des MagIC Beads et d'un crosslink au psoralène, des expériences de RT-qPCR ont été effectuées en aval des étapes 1 et 2 de la méthode PIKA. Après un crosslink au psoralène *in vivo*, les ARN cellulaires ont été extraits puis le SCARNA10 et 7SK ont été purifiés avec des MagIC Beads anti-SCARNA10 et anti-7SK, respectivement. Les ARN purifiés ont subi une irradiation aux UV d'une longueur d'onde de 254nm durant 5 mins afin d'inverser le crosslink. Les ARN sont rétro-transcrits en ADNc, puis les ADNc de SCARNA10, 7SK et snU5 ont été amplifiés par RT-qPCR (**Figure 40**). Dans les échantillons sans crosslink, la co-purification de snU5, que ce soit avec 7SK ou SCARNA10, est quasiment indétectable. Cependant, dans les échantillons crosslinkés, nous remarquons que l'ARN snU5 est co-purifié lors de la précipitation de SCARNA10 avec succès. Les échantillons issus de la purification par les billes contre SCARNA10 contiennent environ 10 fois plus de snU5 que ceux avec des billes anti-7SK. Cependant, la co-purification d'un snoARN avec sa cible reste, comme attendu, un évènement très rare. En effet, lors de la purification d'une molécule de SCARNA10, il y a environ 0,004% de chance de co-purification d'une molécule de snU5.

De par ces expériences de RT-qPCR, il paraît possible de co-purifier un snoARN avec sa cible par les premières étapes de la méthode que nous avons mise en place.

I. 3. 4. Digestion des duplex ARN – ARN

La co-purification d'un snoARN avec sa cible étant possible avec la méthode PIKA, les deux étapes restantes à optimiser étaient celles de digestion et de ligation. Ces deux étapes sont

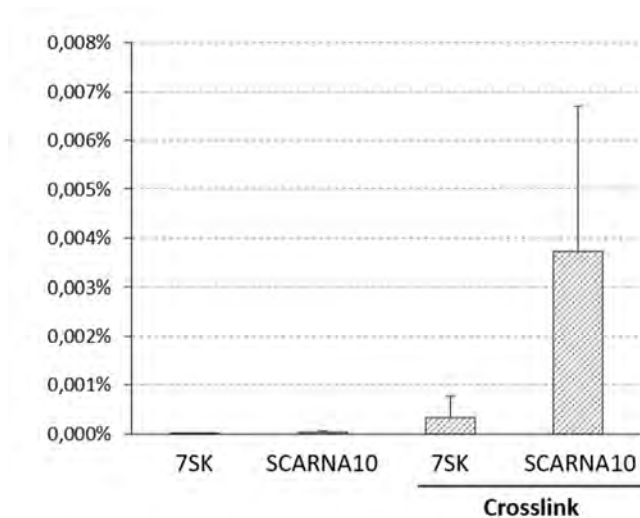


Figure 40 : Détection de la co-purification de snU5 lors de la purification spécifique de SCARNA10 et 7SK.

Quantité de snU5 co-purifiée par rapport aux quantités purifiées de 7SK ou SCARNA10 avec et sans crosslink. Les détections des différents ARN ont été effectuées par RT-qPCR.

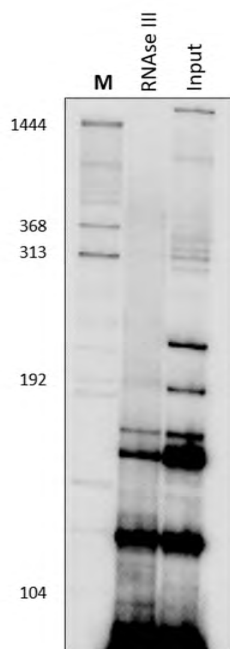


Figure 41 :

Digestion d'un extrait d'ARN totaux par la RNase III à 37°C durant 1h

cruciales pour permettre l'identification des cibles d'un snoARN orphelin par méthodes de séquençage à haut débit.

Pour mettre au point l'étape de digestion, nous nous sommes basés sur les méthodes de crosslink au psoralène PARIS et COMRADES, qui utilisent la RNase III (Lu *et al*, 2016; Ziv *et al*, 2018). L'objectif est de digérer les ARN co-purifiés afin d'obtenir des fragments d'ARN permettant d'indiquer précisément le lieu de l'interaction et de pouvoir les relier via une étape de ligation ultérieure. Pour ce faire, 8µg ARN totaux ont été digérés par 4U de RNase III durant 1h à 37°C. Les fragments d'ARN ainsi obtenus ont été visualisés par radiomarquage au pCp, couplé à une migration sur gel polyacrylamide dénaturant (**Figure 41**). De nombreuses bandes ayant une taille supérieure à 170nt lorsque les ARN n'ont pas été digérés se retrouvent absentes une fois la digestion a été effectuée. Une digestion partielle des ARN par la RNase III est alors observée, donnant lieu à des fragments d'ARN convenant à la préparation ultérieure d'ADNc et au séquençage NGS (Next-Generation Sequencing), validant ainsi cette étape de digestion.

Lors de la méthode PIKA, les ARN à digérer sont issus d'une purification spécifique par des oligonucléotides complémentaires. Les quantités d'ARN à fragmenter sont alors beaucoup plus faibles que dans un échantillon d'ARN cellulaire provenant du même nombre de cellules. En ce sens, des tests de digestion de l'ARN 7SK via les MagIC Beads ont été effectués après purification spécifique de cet ARN (data non montrées). L'ARN 7SK a été digéré en fragments de 20 – 140nt qui ont été révélés sur gel ayant une haute concentration en acrylamide 12%

D'après ces différentes expériences de digestion, la RNase III semble être une nucléase adaptée pour digérer les ARN purifiés en fragments de tailles optimales pour une future étape de ligation de proximité.

Pour mettre au point la ligation de proximité, nous avons décidé d'utiliser la T4 RNA Ligase I, à l'instar des méthodes PARIS, SPLASH et COMRADES (Aw *et al*, 2016; Lu *et al*, 2016; Ziv *et al*, 2018). Cependant, nous n'avons pas pu effectuer de tests préliminaires au laboratoire pour cette étape. En effet, la qualité de la ligation de proximité ne pourra être seulement contrôlé lors de l'analyse des résultats issus du séquençage haut débit après la réalisation de la méthode PIKA pour un ARN donné.

I. 3. 5. Derniers tests de mise au point par séquençage haut-débit

Pour mettre au point l'étape de digestion, ainsi que pour la validation de l'enchaînement des différentes étapes composant la méthode PIKA, différentes expériences ont été réalisées.

La première expérience effectuée a été composée d'une application de la méthode PIKA avec les paramètres précédemment établis avec les ARN SNORA55, SCARNA10 et 7SK. Après crosslink, ces trois ARN ont été spécifiquement co-purifiés, et les ARN ainsi obtenus ont été rassemblés dans un même échantillon. Après fragmentation, ligation, préparation de l'ADNc par transcriptase inverse et séquençage NGS, nous avons espéré être capable de déterminer les interactions de chacun des ARN. Cette décision de rassembler les échantillons a été prise pour pallier aux faibles quantités de matériel obtenues par la purification d'ARN spécifique lors de la méthode PIKA. En effet, le séquençage de petits ARN par la technologie Illumina proposée par GenewizTM nécessite une quantité minimum de 50ng d'ARN par échantillon. Malheureusement, malgré le rassemblement de ces trois expériences parallèles sur trois ARN différents, les échantillons avec et sans crosslink ont difficilement passés l'étape de qualité contrôle, et très peu d'ARN chimériques ont été détectés lors des analyses (data non montrée).

Une seconde expérience a alors été réalisée. De nombreuses expériences de co-purification de SNORA55 et de ces ARN cibles avec ou sans crosslink ont été effectuées en parallèle. L'ensemble de ces expériences ont été assemblées et concentrées par centrifugation sous vide. SNORA55 a pu être aisément être détecté par radiomarquage au pCp dans les échantillons crosslink et non-crosslink (**Figure 42**). Cependant, la quantité de SNORA55 n'excédait pas 60-70ng dans les échantillons. Pour pallier à ce problème, les échantillons ont été enrichis avec des ARNr purifiés d'*E. coli*, qui peuvent être facilement identifiés lors de l'analyse ultérieure des résultats obtenus par séquençage haut-débit. Les échantillons, après passage au contrôle qualité, ont pu être transcrit en ADNc et séquencés. Cependant, lors de l'analyse, trop peu d'ARN SNORA55 ont été détectés pour conclure sur ses ARN cibles.

I. 3. 6. Recherche des ARN interagissant avec 7SK par la méthode PIKA

Afin d'étudier l'étape de ligation tout en ayant une quantité de matériel nécessaire le permettant, la technique PIKA a été appliquée avec l'ARN 7SK. Pour ce faire, 240 millions de cellules HeLa ont été utilisées. La moitié de ces cellules a été crosslinké au psoralène avec les paramètres de la méthode PIKA. Les ARN 7SK ont été purifiés grâce à 600 pmol

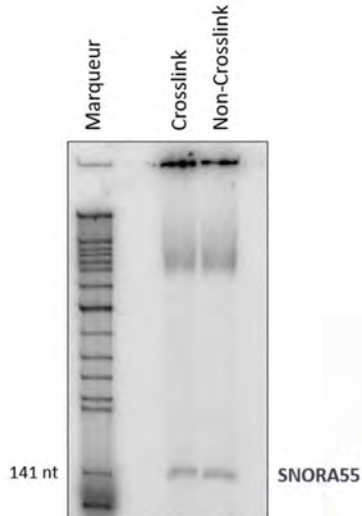


Figure 42 : Détection de SNORA55 avant séquençage haut-débit

Détection au pCp des ARN envoyés au séquençage haut-débit après la réalisation de la technique PIKA. Différentes expériences de co-purifications de SNORA55 et de ses cibles ont été réalisées et rassemblées dans un seul échantillon. Deux conditions ont été effectuées, avec et sans crosslink au psoralène.

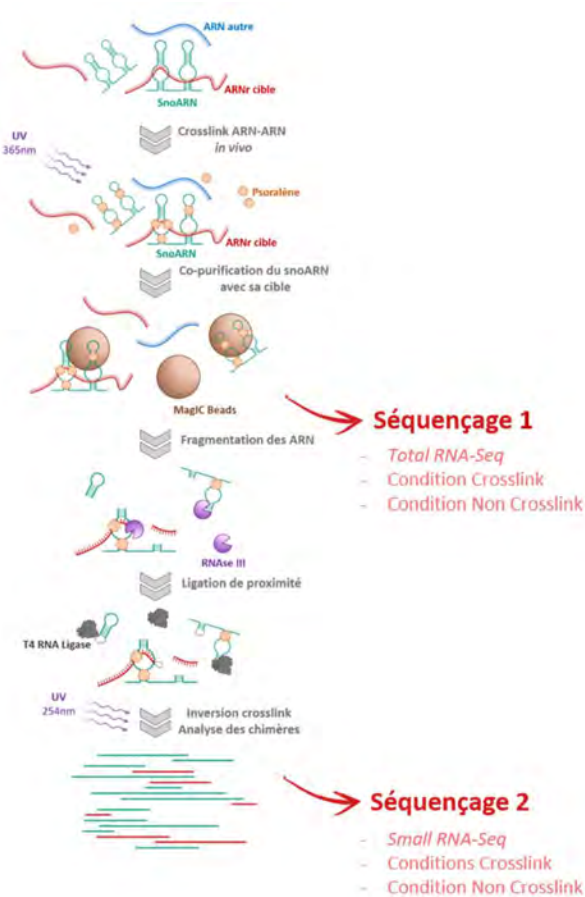


Figure 43 : Spécificités des ARN envoyés au séquençage haut-débit lors de l'utilisation de la méthode PIKA

250ng d'ARN obtenus après purification spécifique de 7SK pour les conditions crosslink et non-crosslink composent les premiers échantillons envoyés au séquençage. 400ng et 361ng d'ARN obtenus après l'utilisation de la méthode PIKA complète pour les conditions crosslink et non-crosslink composent les seconds échantillons envoyés au séquençage.

d'oligonucléotides complémentaires via les MagIC Beads par condition (avec ou sans crosslink). Nous avons obtenu 640ng et 760ng d'ARN pour les conditions contrôle et crosslink, respectivement. 250ng par échantillon ont été utilisés pour un séquençage ARN à haut débit, après inversion du crosslink aux UV (**Figure 43**). Un séquençage lors de cette étape de la méthode PIKA va nous permettre, en comparaison avec un second séquençage ayant lieu à la fin de la technique PIKA, d'étudier précisément l'impact des étapes de digestion et de ligation. Le séquençage est effectué par la technologie Illumina chez GenewizTM avec le programme total RNA-seq.

Les ARN purifiés restant ont été digérés dans un volume de 200µL avec 10 unités de RNase III à 37°C durant 1h. Les fragments d'ARN proches ont été ligués par 50 unités de T4 RNA Ligase I dans un volume de 100µL final pendant une incubation durant la nuit à 16°C. Les ARN ont été concentrés par une purification avec des billes RNA clean XP. Malgré les différentes étapes, très peu de matériel a été perdu : nous avons obtenu 400ng d'ARN dans l'échantillon contrôle et 361ng dans l'échantillon crosslink. Pour finir, ces échantillons d'ARN ont été séquencés par small RNA-seq de la technologie Illumina (GenewizTM).

Concernant l'analyse des échantillons d'ARN pour lesquels les étapes de digestion et de ligation n'ont pas été effectuées, nous avons obtenu 18,3 et 17,8 millions de reads pour l'échantillon avec ou sans crosslink, respectivement. Le logiciel TrimGalore a permis d'enlever les adaptateurs 5' et 3' et la qualité des reads a été vérifiée par le programme FastQC. Les reads ont été alignés à la version hg38 du génome humain grâce au logiciel BWA mem. Afin de recenser combien de fois les reads s'alignent avec quels gènes du génome humain, le logiciel featureCounts a été utilisé. Les analyses exploratoires ont ensuite été réalisées avec le logiciel R au sein de l'environnement de travail Rstudio. 98,05% et 97,72% des reads ont été alignés avec le gène 7SK dans les échantillons crosslink et contrôle, respectivement (**Figure 44A**). Ces scores démontrent la forte spécificité des MagIC Beads et justifient leur utilisation. En dehors de 7SK, les ARN les plus représentés via les reads obtenus sont des ARNr, notamment les 18S et 28S. Ensuite, le nombre de reads obtenus sur un ARN en particulier décroît très vite. La présence de reads obtenus pour ces ARN corrèle avec leur taux d'expression, il y a de grande chance que ces ARN fassent parti des ARN non-spécifiquement capturés par les billes. Pour étudier si certains ARN sont enrichis grâce au crosslink au psoralène, les quantités des reads obtenus dans les échantillons crosslink pour chaque ARN sont analysés en fonction de leur enrichissement par rapport à l'échantillon contrôle (**Figure 44B & C**). Même s'il semble que ce soit le cas pour les ARN CAST, XAB2, RPL7 et HSP90, la

A

ARN	Crosslink	Contrôle
RN7SK	98,05%	97,72%
RNA18S5	0,58%	0,54%
RNA28S5	0,35%	0,19%
PLCG2	0,15%	0,14%
RN7SL2	0,04%	0,04%
RN7SL1	0,02%	0,01%
RNA5-8S5	0,01%	0,01%
EEF2	0,01%	0,02%

B

ARN	Nbre de reads	Coef. d'enrichissement
CAST	83	5,67
XAB2	49	6,69
HSP90	59	5,37
RPL7	58	21,14
SNORD3C	4	4,37
SNORD63	2	1,09
SNORD96A	2	1,10
SNORD3A	80	1,11
SNORA73A	31	1,41
SNORA71B	2	1,09
SNORA64	5	5,47
SNORA6	6	2,19
SNORA1	4	2,19
SNORA5C	2	2,19
SNORA68	12	1,49
SNORA52	4	1,49

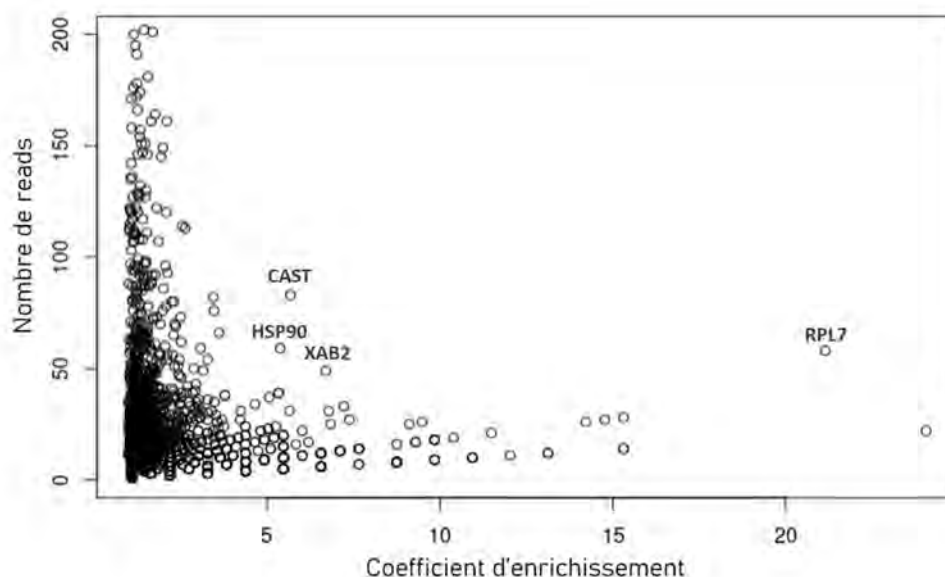
C

Figure 44 : Analyse des reads obtenus par la méthode PIKA avec l'ARN 7SK

Résultats du séquençage des échantillons obtenus par crosslink et purification de l'ARN 7SK. Les étapes de digestion et de ligation n'ont pas été réalisées. **(A)** Quantité de reads obtenu après séquençage des échantillons crosslink et non crosslink. Les reads ont été alignés avec le génome humain hg38. **(B)** Nombre de reads s'alignant avec les ARN CAST, XAB2, RPL7, HSP90 et les ARN H/ACA et C/D obtenus. Les coefficients d'enrichissements des reads entre les conditions crosslink et non-crosslink sont indiqués. **(C)** Représentation schématique du nombre de reads obtenus pour chaque ARN en fonction de leur enrichissement dans la condition crosslink par rapport à la condition contrôle.

quantité de reads retrouvée est très faible (moins d'une centaine sur plus de 18 millions de reads) même si ces ARN ont un enrichissement décent dans l'échantillon crosslink, notamment RPL7, avec un coefficient d'enrichissement dépassant 21. Certains reads s'alignent avec des snoARN H/ACA et C/D, mais possèdent des coefficients d'enrichissement très faibles, environnant 1,5, et très peu de reads sont détectés (moins de 30) (**Figure 44B**). Point important, aucun read ne s'aligne avec le snoRNA70, connu pour guider une pseudouridylation au niveau de l'Uridine 250 de l'ARN 7SK. Cependant, certaines de ces potentielles interactions pourront être infirmées ou confirmées avec l'analyse des ARN chimériques obtenue par small RNA-seq.

Concernant les échantillons séquencés par small RNA-seq, 111 millions et 148 millions de reads ont été obtenus après séquençage haut débit pour les conditions crosslink et contrôle, respectivement. Les ARN, à l'instar du précédent séquençage, ont été séquencés par chaque extrémité, générant des reads forward et reverse. Les adaptateurs ont été retirés avec TrimGalore, et la longueur des reads forward ont été analysées (**Figure 45**). Les résultats obtenus lors de l'analyse des reads reverses sont similaires (data non montrées). La répartition de la longueur des reads entre les conditions crosslink et contrôle sont quasiment identiques. Ces données semblent indiquer que le crosslink a eu peu d'effet sur la longueur des fragments d'ARN séquencés, indiquant la faible proportion des ARN chimériques issus du crosslink. En effet, une longueur de reads plus importante est attendue dans l'échantillon avec crosslink de par la génération d'ARN chimériques.

Les reads Forward et Reverse ont été rassemblés avec le logiciel vsearch, étape nécessaire à l'utilisation du programme ChiRA, permettant d'analyser les reads issus du séquençage d'ARN chimériques (Videm *et al*, 2021). Lors de l'utilisation du programme ChiRA, deux listes de séquences de référence peuvent être spécifiées pour l'alignement. La première séquence que nous avons spécifiée est celle de l'ARN 7SK, et la seconde est l'ensemble du transcriptome humain. De cette manière, ChiRA cherche en premier lieu une complémentarité avec la séquence de 7SK, puis ensuite une complémentarité avec les séquences du transcriptome humain sur les mêmes reads. Donc lors de cette analyse, seuls les reads s'alignant au moins une fois avec 7SK sont analysés. 24914 et 21308 reads ont respectivement été identifiés dans les échantillons crosslink et non crosslink, pour être composé d'une région s'alignant avec 7SK et d'une autre sur une région d'un ARN du transcriptome. Pendant cette analyse, aucune chimère contenant un fragment autre que des fragments de 7SK n'a été trouvé, seulement des ARN chimériques qui comportent deux régions de 7SK ont été recensés. De plus, ces hybrides ont été retrouvés également dans les échantillons dans lequel le

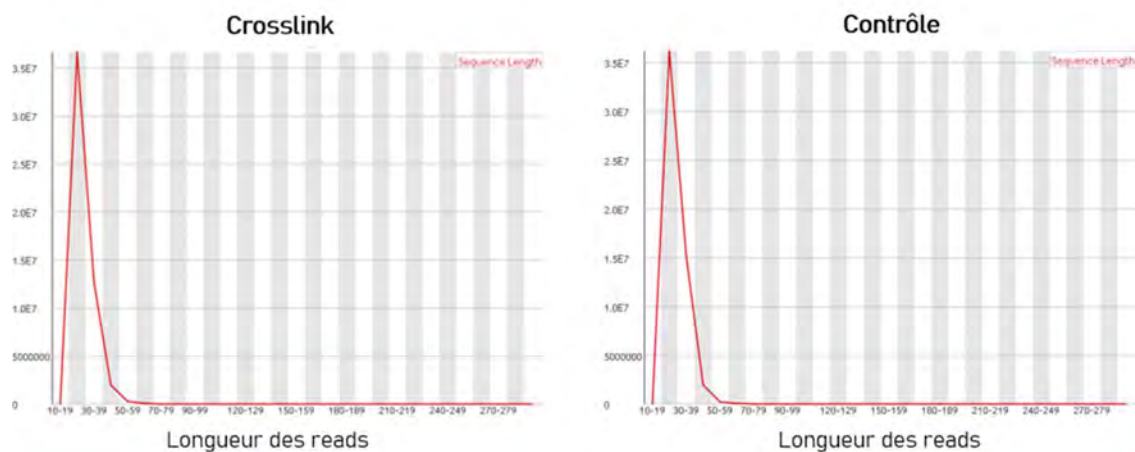


Figure 45 : Répartition des reads en fonction de leur taille

Analyse des données de séquençage issues de l'application de la méthode PIKA avec l'ARN 7SK. La quantité des reads obtenus dans les échantillons avec ou sans crosslink en fonction de leur taille est représentée.

crosslink au psoralène n’a pas été effectué. L’étape de ligation reste alors à optimiser, car elle ne semble pas être spécifique aux fragments crosslinkés. Cependant, ces résultats peuvent servir à vérifier le bon déroulement de l’étape de digestion. Les régions d’alignement des extrémités 5’ des reads qui s’alignent avec l’ARN 7SK ont été analysées (**Figure 46**). Nous pouvons observer un enrichissement d’extrémités de reads s’alignant avec les régions situées vers 10, 60, 210 et 270nt, qui correspondent à des régions double brins de la structure secondaire de 7SK, dans les échantillons crosslink ou contrôle (**Figure 46**). Également, nous observons un léger enrichissement entre 100 et 200 spécifiquement dans les échantillons crosslink. Cette région de 7SK n’est en majorité pas structurée et donc simple brin, mettant en avant une potentielle interaction avec un autre ARN de 7SK.

Cette expérience souligne que l’étape de ligation reste à optimiser afin d’éviter de créer via cette étape des ARN chimériques au sein des échantillons qui ne subissent pas de crosslink au psoralène. En revanche, l’étape de digestion semble fonctionner correctement, de par les régions coupées et la tailles des reads obtenus. Une fois mise au point, la technique PIKA pourrait nous indiquer si une interaction 7SK-7SK existe *in vivo* ou s’il s’agit d’un biais de l’expérience précédemment réalisée.

I. 3. 7. Etude des interactions des ARN C/D avec le transcriptome

En parallèle de l’expérience de crosslink avec l’ARN 7SK, un essai de la méthode PIKA a été effectué en purifiant la protéine cœur des ARN C/D, la Fibrillarine. En effet, l’objectif premier lors du développement de la méthode PIKA est de pouvoir l’appliquer aux snoARN. Nous avons choisi pour la mise au point de PIKA d’effectuer une purification de la protéine Fibrillarine, car (i) les interactions de la totalité des snoARN C/D de la cellule pouvaient être étudiés (ii) cela a permis de générer une importante quantité de matériel, ce qui n’aurait pas été le cas lors de la purification d’un snoARN unique.

Pour cette expérience, 60 millions de cellules HeLa ont été utilisées par condition (crosslink ou non-crosslink) pour effectuer la méthode PIKA. Seule l’étape de purification habituelle via les MagIC Beads a été remplacée, par une immunoprécipitation de la Fibrillarine. Après lyse des cellules par sonication, les extraits ont été incubés avec un anticorps anti-Fibrillarine immobilisé sur des billes agarose durant 1h à 4°C, suivi d’une extraction des ARN des billes par protéinase K. Les étapes suivantes ont été réalisées conformément à la méthode PIKA. 480ng et 720ng d’ARN pour les conditions crosslink et non-crosslink ont été obtenus,

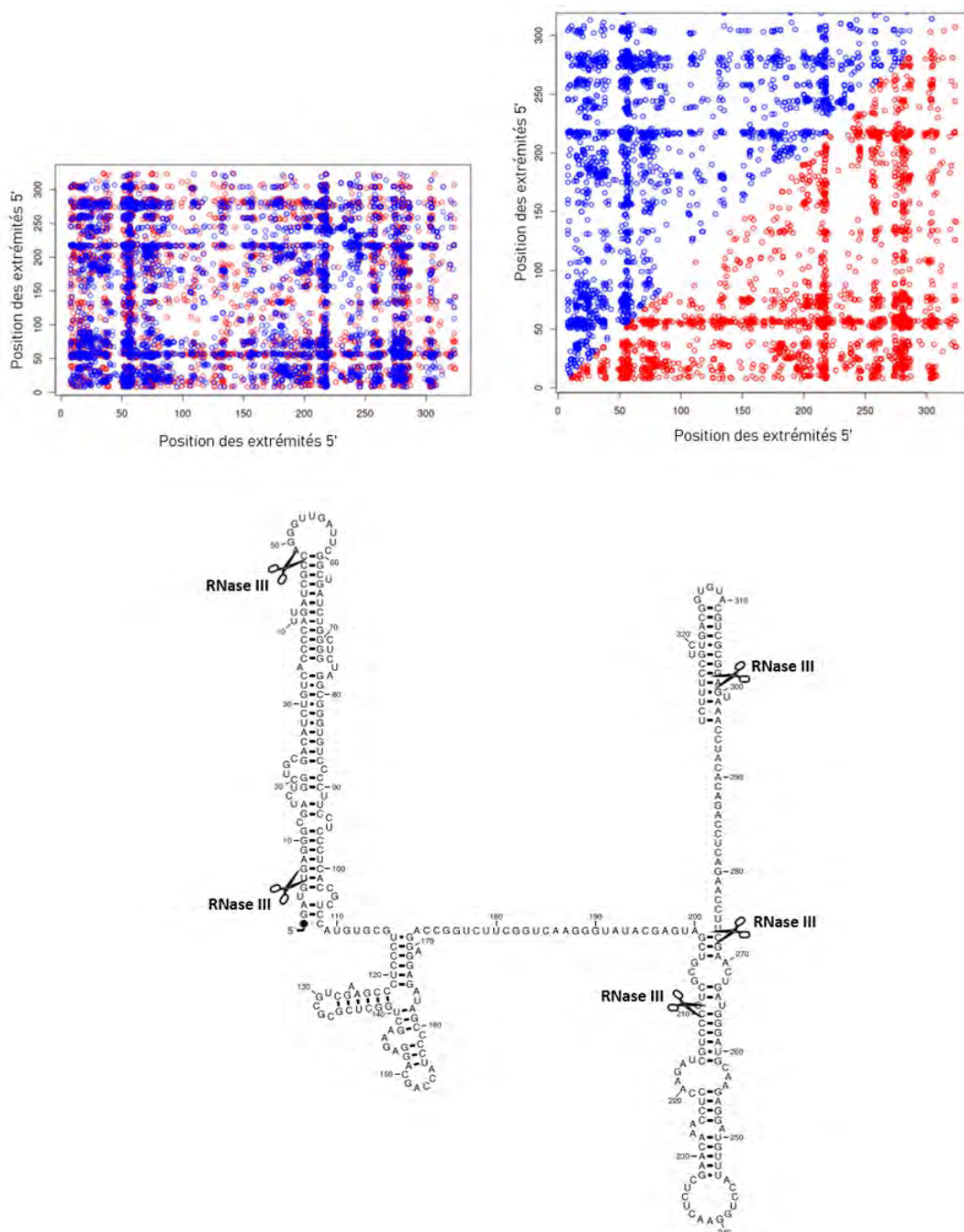


Figure 46 : Analyse des ARN chimériques

Analyse des ARN chimériques obtenus par l'application de la méthode PIKA avec l'ARN 7SK. Les extrémités 5' des régions alignées sur la partie codante du gène 7SK sont représentées. Les points rouges et bleus correspondent respectivement aux reads obtenus avec ou sans crosslink. Les régions putatives digérées par la RNase III sont représentées sur la structure secondaire de l'ARN 7SK.

respectivement. Les ARN ont été envoyés chez GenewizTM pour la préparation de la banque d'ADNc et le séquençage avec la technologie small RNA-seq de Illumina.

74,6 millions et 136,9 millions de reads ont respectivement été obtenus lors du séquençage des échantillons crosslink et contrôle. Les mêmes programmes utilisés lors de l'analyse des reads issus de l'expérience avec 7SK ont été appliqués pour cette analyse. Nous pouvons observer des résultats assez différents concernant la répartition de la longueur des reads (**Figure 47**). L'effet de la ligation dans les échantillons avec crosslink est clairement observé, traduit par la présence d'un pic aux alentours de 180nt, qui n'est pas présent dans l'échantillon contrôle. Lors de cette expérience, l'étape de ligation de proximité semble avoir correctement fonctionné. Après l'utilisation du programme ChiRA, le nombre de reads chimériques n'excède pas 600 000 reads dans l'échantillon contrôle, contre 330 000 dans le crosslink. Lors de l'analyse de chimères obtenues après crosslink, 121 reads ont été identifiés contenant une région appartenant à SCARNA10. Cependant, ces même reads s'alignent au même temps avec les ARNr 28S, 18S et 5S. Également, 7 reads ont été détectés pour l'ARN SNORD 16, qui s'alignent également avec les ARNr 28 et 5S, tandis qu'il est censé guider une méthylation sur l'ARNr 18S (Lestrade & Weber, 2006). Pour d'autres snoARN, la position des alignements correspondait aux interactions attendues entre l'ARN guide et sa cible. Par exemple, 7 sur 13 chimères de SNORD52 correspondaient à sa fonction de guider la méthylation U3904. Sans crosslink, l'analyse des chimères a donné des résultats identiques : la majorité de régions/ARN identifiés pour les ARN guides n'étaient pas la région/ARN attendu avec la connaissance de leur fonction. Récemment, un nombre important d'interactions entre les snoARN avec les ARN ribosomiques a été démontré par CLASH dans la levure (Dunn-Davies *et al*, 2021). Une grande partie de ces interactions ne servent pas aux modifications post-transcriptionnelles, mais peuvent témoigner d'une recherche de complémentarité par les ARN guide, ou même être associées à un rôle régulateur. Nous avons donc attendu des chimères entre les ARN guides et des régions variées des ARN ribosomiques. Mais un grand nombre de chimères a été également recensé sans crosslink préalable, indiquant que ces interactions putatives ne pouvaient pas toutes être le résultat d'interactions transitoires identifiées par crosslink, mais plutôt le résultat d'une ligation non-optimale qui relie des fragments d'ARN non-crosslinkés. L'analyse de cette expérience nous montre que les conditions de la ligation doivent être optimisées afin de réduire drastiquement la possibilité de liguer des fragments non-crosslinkés ensemble.

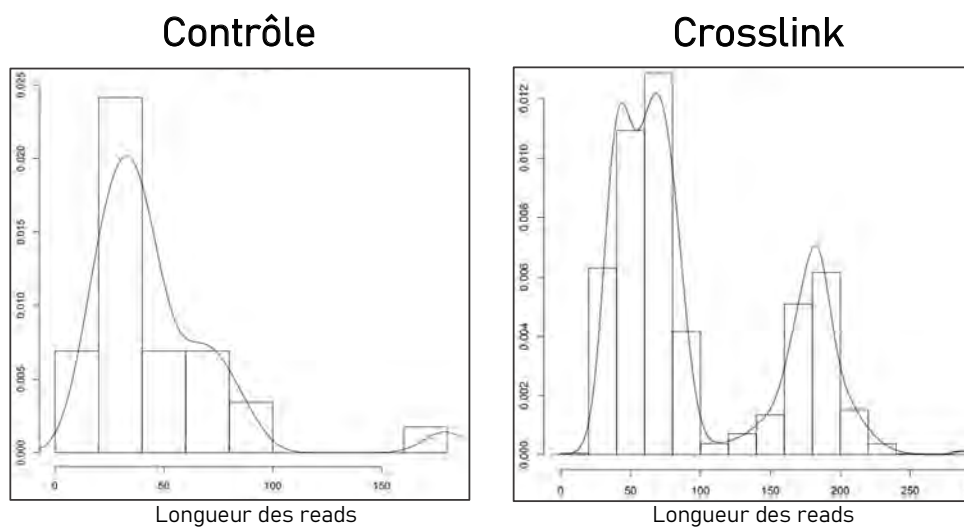


Figure 47 : Répartition des reads en fonction de leur taille

Analyse des données de séquençage issu de l'application de la méthode PIKA avec les ARN C/D cellulaire, obtenues par une immunoprécipitation de la Fibrillarine. La distribution des reads obtenus dans les échantillons avec ou sans crosslink en fonction de leur taille est représentée.

II. Etude de l'interaction des ARN SNORA55 et l'ARNr 18S

II. 1. Génération de la lignée déplétée pour SNORA55 et étude de l'interaction SNORA55 – ARNr 18S.

Dans l'optique de déterminer les potentiels ARN cibles de SNORA55, une lignée déplétée pour SNORA55 a été générée par la technologie CRISPR/Cas9. Nous avons tenté de retirer une région de 152 paires de base du génome de la lignée cellulaire HAP1, comprenant une partie du gène *SNORA55*. Après transfection des cellules HAP1 avec un plasmide dirigeant l'expression des protéines nécessaires et une ARN guide approprié, nous avons isolé des cellules transfectées afin d'obtenir des lignées monoclonales. Les ADN génomiques des lignées ont été isolés et une partie de la région encodant du gène *SNORA55* a été amplifiée par PCR. L'amplification de la région génomique *SNORA55* s'est révélée impossible pour les clones 1, 2 et 4 (**Figure 48A**) indiquant une délétion génomique. Le clone 4 a été choisi comme lignée KO *SNORA55* pour les expériences subséquentes. L'expression de l'ARN *SNORA55* n'a pas pu être détectée lors d'une protection RNase des ARN cellulaires de la lignée KO *SNORA55* (**Figure 48B**). Ensuite, la délétion attendue de 152nt a été vérifiée par l'amplification d'une grande région génomique comprenant le gène *SNORA55*. Après migration sur gel, nous avons conclu que la coupure induite par le système CRISPR/Cas9 pourrait être plus importante que prévu (**Figure 48C**). La région chromosomique contenant la région déplétée du gène *SNORA55* a été séquencée, et une délétion de 308nt a été observée, soit deux fois plus importante qu'initialement prévue. Cependant, la région perdue ne comprend pas de régions exoniques, et ne semble pas perturber l'expression de l'ARNm hôte PABPC4 (**Figure 48D**).

Grâce à la technologie CRISPR-Cas9, nous avons pu établir une lignée cellulaire déplétée du gène *SNORA55*, un outil génétique nécessaire à l'étude de l'interaction SNORA55 – ARNr 18S.

II. 2. Détection des pseudouridines de l'ARNr 18S par l'étude d'une lignée KO *SNORA55*

Afin d'étudier l'interaction prédite *in silico* de SNORA55 – ARNr 18S, une détection de pseudouridine de l'ARNr 18S dans les lignées cellulaires HAP1 sauvages et KO *SNORA55*, récemment générée, a été effectuée.

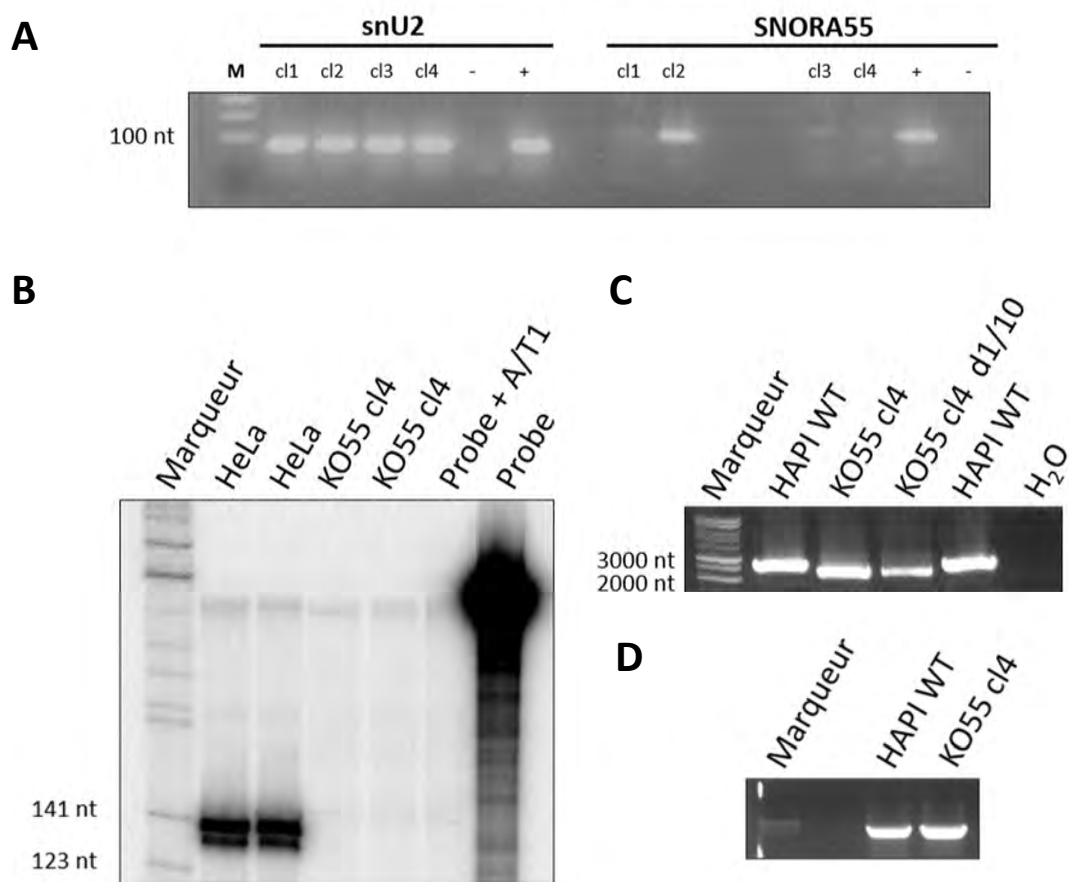


Figure 48 : Génération de la lignée cellulaire KO SNORA55

(A) Criblage des lignées monoclonales issues des cellules transfectées transitoirement avec le plasmide contenant la Cas9 et les ARN guides par amplification génomique du gène *SNORA55*. (B) Détection de l'ARN *SNORA55* dans des cellules sauvages et KO pour *SNORA 55* par digestion aux RNases A/T1 avec une sonde complémentaire à l'ARN *SNORA55*. (C) Observation de la délétion génomique contenant une partie du gène *SNORA55* engendrée par le système CRISPR/Cas9 par PCR (D) Amplification génomique de l'exon 5 à l'exon 8 de l'ADNc du gène *PABPC4* par RT-PCR dans des lignées sauvages et KO pour le gène *SNORA55*.

Après extraction des cellules, les ARN ont été traités au CMCT. Les composés CMC liés au niveau de l'azote N1 des pseudouridines via la réaction avec le CMCT ont été retirés par une hydrolyse alcaline. Une étape de transcriptase inverse avec un oligonucléotide radiomarké a ensuite été réalisée. L'oligonucléotide qui a été utilisée est complémentaire avec une région de l'ARNr 18S située en aval de la pseudouridine Ψ 36 étudiée. Les arrêts de la transcriptase inverse provoqués par la présence de CMC spécifiquement sur les azotes N3 des pseudouridines ayant résistés à l'hydrolyse alcaline ont été analysés (**Figure 49**).

Deux bandes, correspondants aux pseudouridines Ψ 34 et Ψ 36, sont visibles sur les ARN issus de cellules HAPI sauvages traitées au CMC. En revanche, nous observons l'absence de la bande correspondant à la Ψ 36 lorsque le traitement au CMCT a été effectué avec les ARN issus de la lignée cellulaire déplétée de *SNORA55*. La bande correspondant à la pseudouridine à la position 34 est en revanche toujours présente.

Nous en avons conclu que la prédiction *in silico* s'avère exacte, la RNP SNORA55 est nécessaire à l'isomérisation de l'Uridine à la position 36 de l'ARNr 18S.

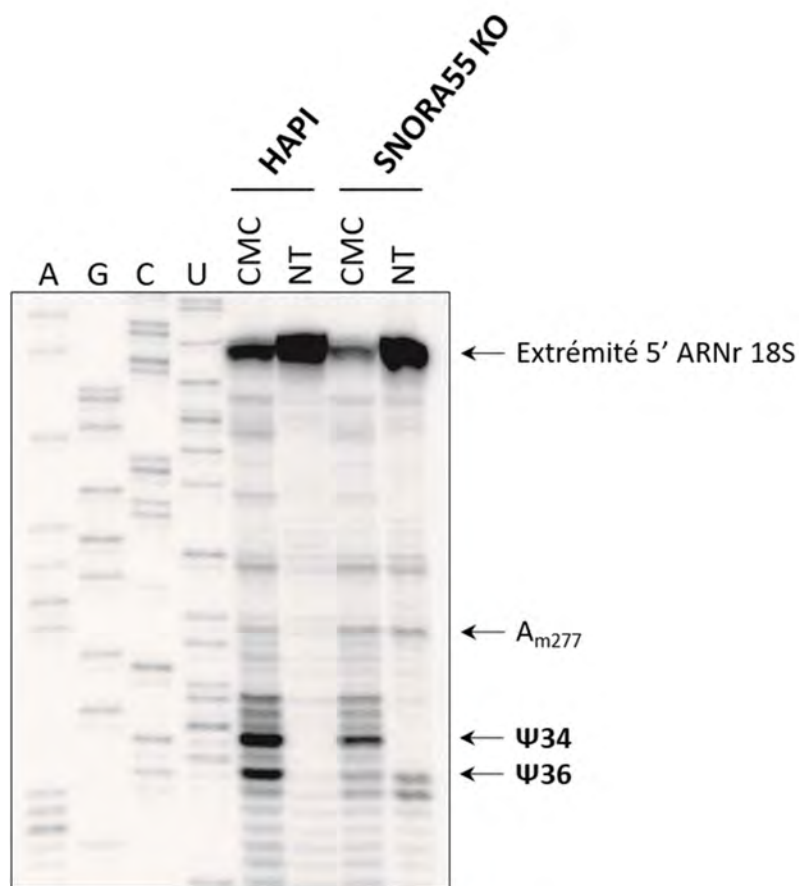


Figure 49 : Détection de la pseudouridine Ψ36 de l'ARNr 18S

Cartographie au CMCT de la région 5' de l'ARNr 18S. Les ARN totaux issus des cellules HAP1 ou HAP1 – SNORA55 KO ont été traités au CMCT suivi d'une transcription inverse avec un oligonucléotide radiomarké complémentaire à une région en aval de la pseudouridine Ψ36 de l'ARNr 18S.

III. Recherche des cibles de SNORD115, un snoARN impliqué dans le syndrome de Prader Willi

III. 1. Le SNORD115, un ARN spécifiquement exprimé dans le cerveau

Vu précédemment, le SNORD115 est un snoARN C/D de 82 nucléotides, impliqué dans le syndrome de Prader-Willi. Cependant, la délétion seule des différentes copies de SNORD115 ne suffit pas à déclencher la pathologie (Runte *et al*, 2005). Des recherches *in silico* ont mis en évidence une complémentarité de 18 nucléotides entre SNORD115 et l'exon V de l'ARNm 5-HT2C (Cavaillé *et al*, 2000). Cependant, une déplétion du SNORD115 dans des modèles murins entraîne seulement de légères différences au niveau de l'épissage de l'ARNm 5-HT2C, et aucun importants défauts physiologiques (Hebras *et al*, 2020).

Malgré une prédiction *in silico* très prometteuse (**Figure 50A**), peu d'effets allant dans le sens de cette prédiction ont été observés *in vivo*. Via une collaboration avec le Dr. Jérôme Cavaillé et son équipe, nous avons mis en place un projet afin de déterminer si l'interaction physique entre l'ARN SNORD115 et l'ARNm 5-HT2C a bien lieu dans les cellules. Pour ce faire, nous utilisons les étapes de la méthode PIKA mises au point, couplés à une détection par RT-qPCR, la cible candidate de SNORD115 étant connue.

SNORD115 est fortement exprimé dans les zones cérébrales, mais est en revanche généralement absent dans les autres types cellulaires. En ce sens, aucune lignées commercialisées, d'origines humaines ou murines, n'expriment SNORD115. Il est en revanche possible de récupérer SNORD115 par un broyage de cerveaux de souris suivis d'une extraction d'ARN, de par sa forte expression dans ces tissus. C'est notamment par cette méthode que le Dr. Cavaillé et son équipe étudient le SNORD115 (Hebras *et al*, 2020). Cependant, cette méthode ne peut pas être couplée à l'utilisation des premières étapes de la technique PIKA. Le crosslink ARN-ARN nécessite d'être réalisé *in vivo*, sur cellules isolées en suspension, et ne peut être fait sur tissus entier. Afin de pallier à ces limites, nous avons mis en place deux protocoles permettant le couplage de la méthode PIKA avec l'étude du SNORD115 (i) Après dissection, les cerveaux murins sont soumis à un broyage mécanique doux, puis une digestion à la trypsine afin de séparer les cellules. Le crosslink est effectué sur les cellules dissociées des tissus. (ii) Des cellules murines cultivées de carcinomes embryonnaires sont différenciées en cellules neuronales, activant ainsi l'expression de SNORD115, sur les quelles est effectué le crosslink au psoralène.

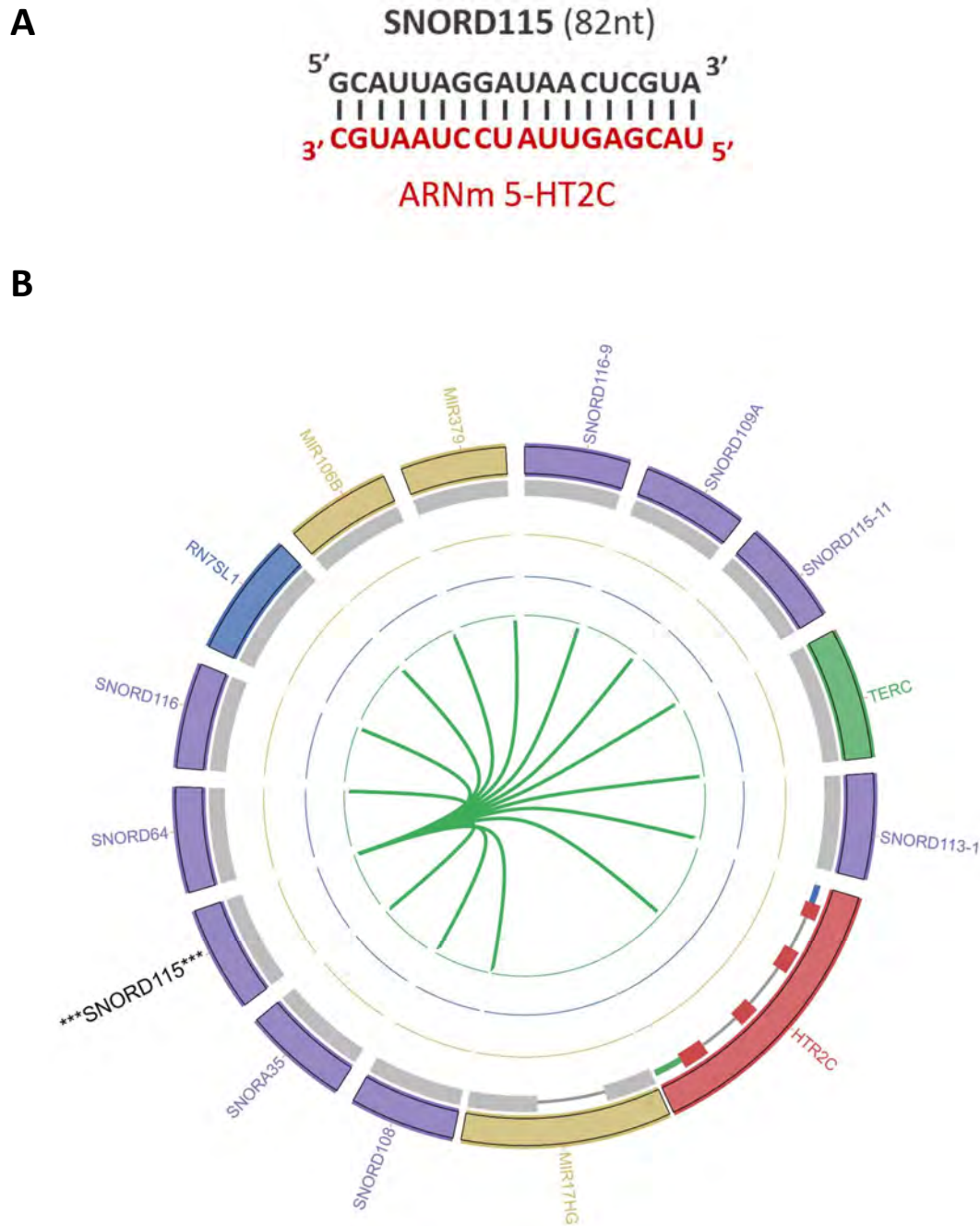


Figure 50 : ARN putatifs interagissant avec SNORD115

(A) Prédiction *in silico* de l'appariement parfait de 18 nucléotides entre l'élément antisens situé en aval de la boîte D de SNORD115 avec l'exon V du pré-ARNm du récepteur à la sérotonine 2C. (B) Cibles potentielles de SNORD115 chez l'homme d'après la base de données RAIN, mise en forme par RISE. Les interactions entre SNORD115 et les ARN candidats sont représentées par les lignes vertes. Les gènes représentés en vert, violet, bleu foncée, rouge et or correspondent à des longs ARN non codants, snoARN, ARN non codant, ARNm et microARN, respectivement.

D'après la base de données RISE, des interactions entre SNORD115 et d'autres ARN sont mises en évidence, notamment avec des snoARN et des micros ARN, en supplément de l'interaction entre SNORD115 et le pré-ARNm 5-HT2C (**Figure 50B**). Toutes ces interactions proviennent de la base de données RISE, aucune interaction pour le SNORD115 n'a été trouvée par l'intermédiaire des méthodes de crosslink au psoralène ou par crosslink avec des protéines appâts.

III. 2. Etude de l'interaction SNORD115 – ARNm 5-HT2C

III. 2. 1. Mise au point de la purification de SNORD115

SNORD115 (82nt) est un snoARN plus court que les snoARN à boîtes H/ACA. Le design des MagIC Beads s'est alors révélé plus complexe, sachant qu'il fallait éviter que les billes ciblent les régions antisens du snoARN, vu que nous voulons co-purifier le SNORD115 avec sa cible. Le nombre de séquences différentes complémentaires présents sur les billes s'est donc retrouvé réduit.

La purification de SNORD115 avec les MagIC Beads à partir des extraits d'ARN totaux issus de cerveaux de souris avec le protocole standard s'est révélée inefficace (**Figure 51A**, piste 5-6). Une diminution de la température d'hybridation de 60°C à 45°C a finalement rendu possible la purification de SNORD115 avec les billes (**Figure 51A**). En revanche, l'efficacité de purification s'est révélée être bien en deçà de celles obtenus lors de la purification de SCARNA10, SNORA55 et 7SK (**Figure 51B**). Nous avons ensuite remarqué que les billes semblaient avoir une meilleure efficacité lors d'une purification en partant de 20µg d'ARN totaux, au lieu de 4µg pour une quantité de billes portant 20pmol d'oligonucléotides complémentaire (**Figure 51C**). Aucune bande non-spécifique n'apparaissait lors de la réalisation d'un radiomarquage au pCp de l'ARN purifié (data non montrée) montrant que la spécificité de la purification reste haute.

Les MagIC Beads permettent de purifier SNORD115, avec une efficacité honorable et une forte spécificité.

III. 2. 2. Obtention de cellules murines exprimant SNORD115

Afin d'être capable de réaliser un crosslink entre SNORD115 et ses cibles, des cellules isolées doivent être obtenues à partir de tissus. Pour obtenir des cellules isolées issues de cerveaux de souris, nous avons couplé une digestion enzymatique avec un broyage mécanique

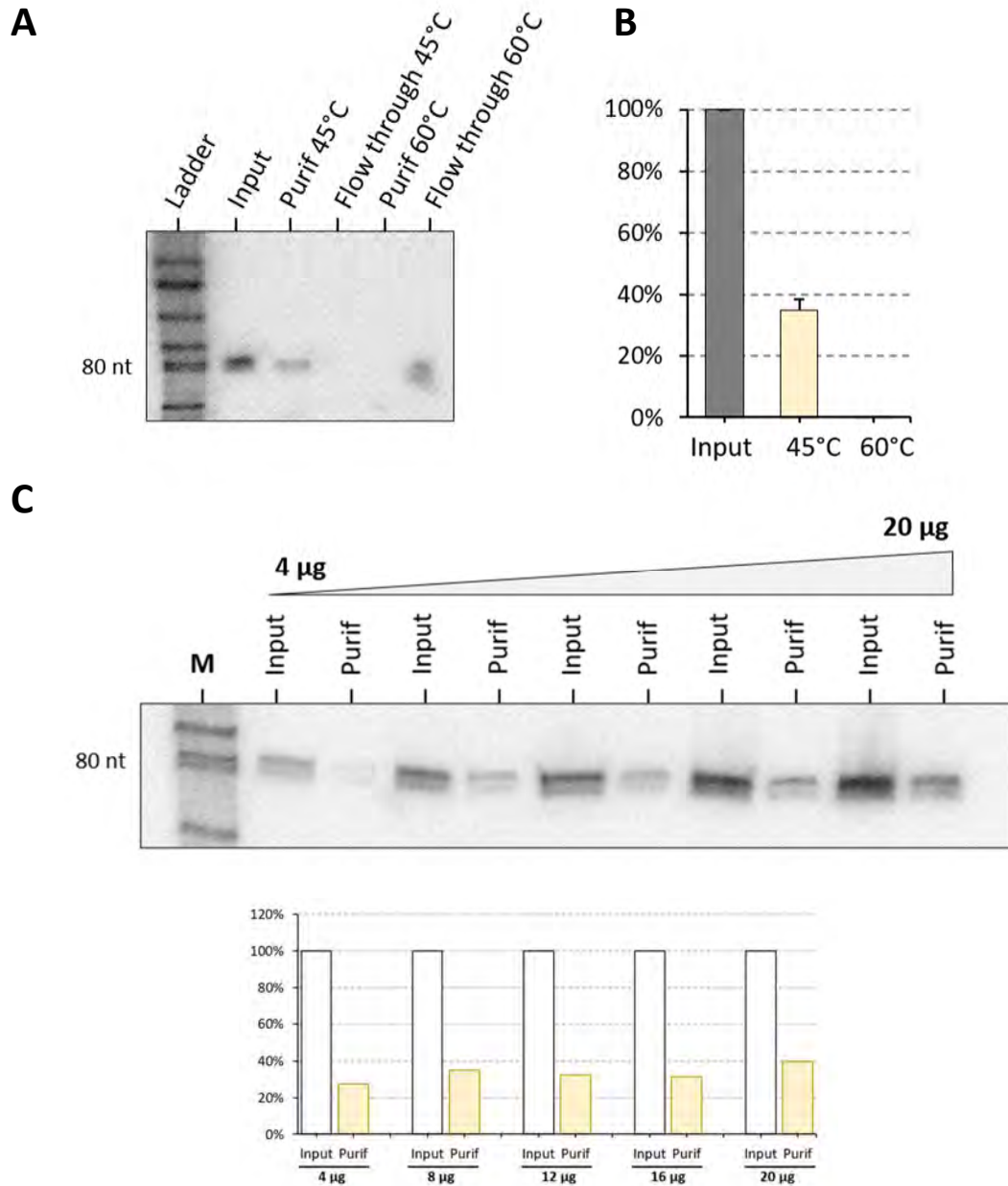


Figure 51 : Mise au point de la purification de SNORD115

(A) Northern Blot montrant la présence de SNORD115 dans des échantillons purifiés d'ARN totaux extraits de cerveaux murins via des MagIC Beads contre SNORD115 avec différentes températures d'hybridation. (B) Efficacité de purification de l'ARN SNORD115 par des MagIC Beads déterminée par rapport à la quantité d'ARN utilisée en Input, selon la température d'hybridation. (C) Efficacité de purification SNORD115 par des MagIC Beads en fonction de la quantité d'ARN totaux utilisés en Input.

doux par aspiration/refoulement. Les cellules dissociées ont ensuite été sélectionnées par filtration. Différentes concentrations enzymatiques et durées de digestion ont été testées, de manière à **(i)** vérifier que nous étions capables d'obtenir des cellules neuronales exprimant SNORD115 dissociées à partir de tissus **(ii)** optimiser les conditions de l'expérience pour maximiser la quantité de cellules isolées. L'expression de SNORD115 dans les cellules dissociées a été vérifiée par Northern Blot (**Figure 52A**). Des « doubles bandes » sont visibles vers 82nt, signature de la présence de SNORD115. La durée d'incubation de la trypsine semble avoir une influence plus forte sur l'isolement des cellules par rapport à la concentration à laquelle la trypsine est utilisée. Il est donc possible, par un broyage mécanique doux couplé à une digestion enzymatique, d'obtenir des cellules cérébrales dissociées exprimant SNORD115, sur lesquelles un crosslink au psoralène peut être effectué. Pour les expériences qui ont suivies, la condition de digestion des échantillons choisie est celle avec de la Trypsine 0,05% - EDTA non diluée durant 30 minutes à 37°C.

Une seconde alternative était de différencier des cellules embryonnaires murines (lignée P19) en cellules neuronales. Pour ce faire, les cellules P19 sont traitées à l'acide rétinoïque (RA), déclenchant ainsi le processus de différenciation. Après cinq jours de traitement, un inhibiteur mitotique est ajouté, pour sélectionner les cellules différenciées. Sept jours après le début du traitement, quelques cellules neuronales sont visibles (**Figure 52B**). Malheureusement, le taux de neurones ne semble pas dépasser les 5% des cellules traitées, et il est très difficile de les faire perdurer plus d'une vingtaine de jours. Lors du protocole pour obtenir des cellules neuronales, les cellules sont en présence de RA durant les 5 premiers jours du processus de différenciation. Afin de déterminer si le taux d'obtention de cellules neuronales pouvait être augmenté, une expérience parallèle a été réalisée, au cours de laquelle le traitement au RA n'a pas été stoppé. Les ARN cellulaires totaux ont été extraits 7, 9 et 12 jours après le début de la différenciation. Les ARN SNORD115 et SNORD116 ont été détectés par Northern Blot avec un oligonucléotide complémentaire radiomarké (**Figure 52C**). L'ARN 7SL a servi de contrôle de charge. Une légère bande peut être observée pour SNORD115 aux jours 7, alors qu'elle est absente, comme attendu, lorsque les cellules n'avaient pas entamé le processus de différenciation. En supplément, l'expression de SNORD115 semble augmenter au cours de la différenciation neuronale. Lors de cette expérience nous avons également voulu observer l'expression de SNORD116, le second snoARN orphelin impliqué dans le syndrome de Prader-Willi pour déterminer si les P19 peuvent être un modèle d'étude intéressant pour étudier cet ARN. Une bande correspondant à SNORD116 est observée au jour 0, et devient de plus en plus

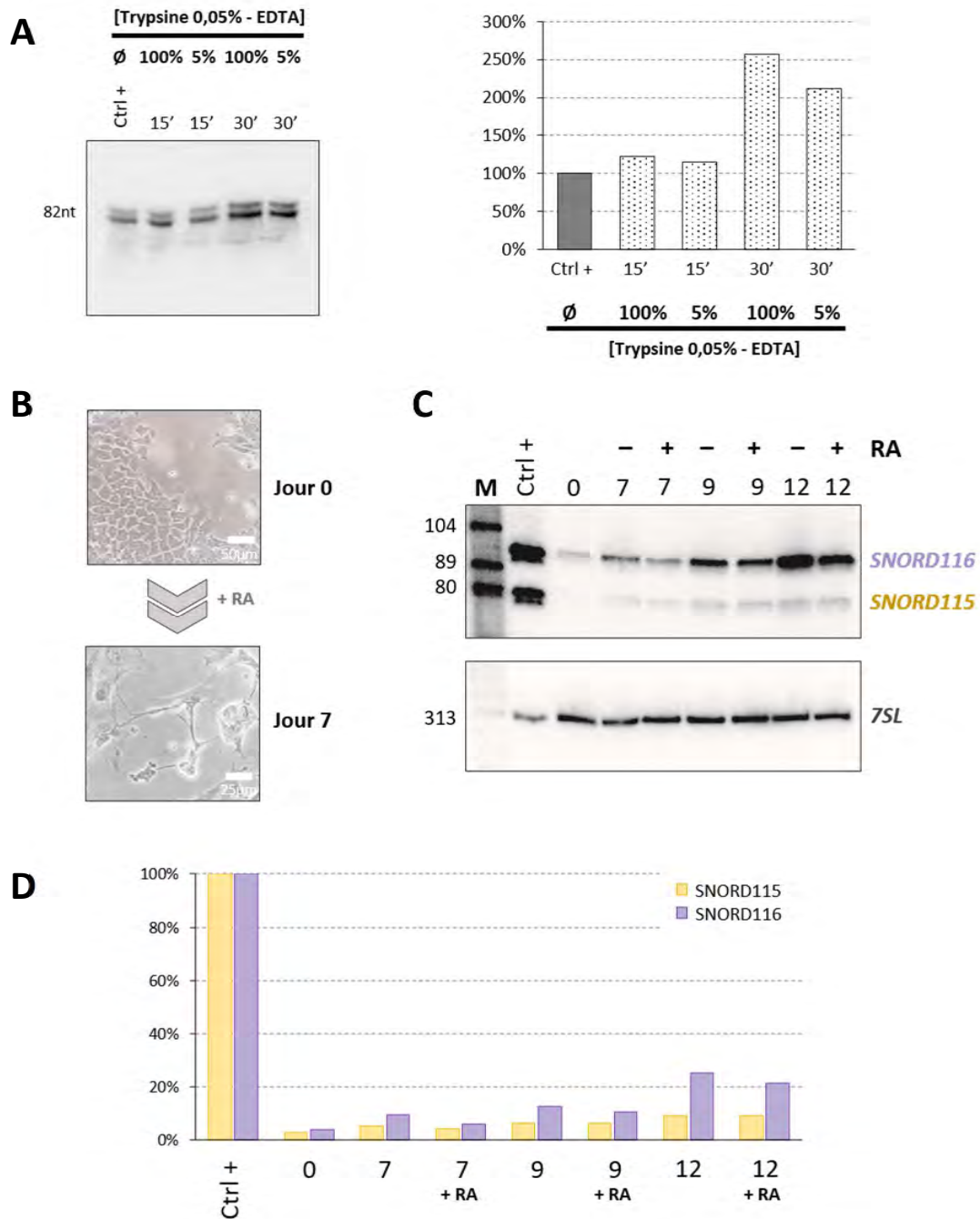


Figure 52 : Méthodes permettant d'obtenir des cellules isolées exprimant SNORD115

(A) Test de dissociation de cellules cérébrales à partir de tissus entiers. Les tissus cérébraux issus de souris ont été incubés durant 15 ou 30 minutes avec différentes concentrations de Trypsine 0,05% - EDTA. La quantité de SNORD115 récupérée est analysée par Northern Blot. Les échantillons sont comparés à des ARN totaux issus de cerveaux de souris obtenus par broyage au Turrax® (broyage mécanique fort). (B) Observations au microscope photonique lors de la différenciation des P19 en cellules neuronales. (C) Détection des ARN SNORD115 et SNORD116 dans les cellules P19 à différents stades de différenciation par Northern blot. (D) Quantification de SNORD115 et SNORD116 dans les cellules P19 au cours de la différenciation neuronale. Les échantillons sont comparés à des ARN totaux issus de cerveaux de souris obtenus par broyage au Turrax®.

intense au cours de la différenciation des cellules embryonnaires en cellules neuronales. Cependant, la durée de traitement des cellules au RA ne semble pas avoir d'effet sur les taux d'expressions des ARN SNORD115 et SNORD116 (**Figure 52D**).

En conclusion, en induisant la différenciation des cellules embryonnaires P19 en cellules neuronales, l'expression de SNORD115 a été activée, puis a augmenté au cours de l'avancée de la différenciation. De plus, ce modèle peut se révéler très pertinent dans le cas de futures études sur SNORD116, de par l'augmentation de son expression au cours de la différenciation neuronale.

Les deux options étant réalisables, nous nous sommes préférentiellement dirigés sur l'isolation des cellules à partir des cerveaux murins dans l'objectif de déterminer l'interaction physique entre SNORD115 et l'ARNm 5-HT2C. L'association entre SNORD115 et sa cible potentielle 5-HT2C est certainement très éphémère. Plus la quantité de matériel utilisé est importante, plus les chances de co-purifier l'ARNm 5-HT2C avec SNORD115 augmente. En ce sens, SNORD115 est fortement exprimé dans les cerveaux murins. L'isolation des cellules des cerveaux de souris permet d'obtenir de fortes quantités de SNORD115, comme en démontre les Northern Blot précédemment observés (**Figure 52C-D**). Lors de la différenciation cellulaire, la quantité de l'ARN SNORD115 obtenue est moins importante, principalement due au fait qu'un taux très faible de cellules se différencient en neurones (**Figure 52B**).

L'ARNm 5-HT2C possède également une expression abondante dans les tissus cérébraux (Hebras *et al*, 2020), contrairement aux différentes lignées cellulaires commercialisées. Toutefois, nous avons détecté l'expression de l'ARNm 5-HT2C dans les neurones différenciés (data non montrées). Lors de l'application de la technique PIKA, le meilleur contrôle de l'expérience est d'utiliser une lignée KO pour le gène étudié. Dr. Cavaillé et son équipe dispose déjà d'une lignée murine dans laquelle les gènes de l'ARN SNORD115 ont été déplétés qui est également en faveur de l'utilisation des cellules provenant de cerveaux murins (Hebras *et al*, 2020). Pour la lignée de P19, la délétion de gènes de l'ARN pour SNORD115 n'a pas encore été accomplie, même s'il était toutefois possible d'utiliser des cellules P19 non différenciées comme cellules n'exprimant pas ce snoARN.

Pour mettre en évidence l'interaction physique entre SNORD115 et l'ARNm 5-HT2C, nous avons donc tenté le crosslink au psoralène sur des cellules dissociées à partir de cerveaux de souris.

III. 2. 3 Etude de l'interaction physique entre SNORD115 et 5-HT2C

Pour démontrer une interaction physique entre SNORD115 et 5-HT2C, des cellules ont été dissociées à partir de cerveaux murins WT et KO pour SNORD115. La moitié des cellules a été incubée avec de l'AMT psoralène, et la totalité a été irradiée aux UV avec une longueur d'onde de 365nm. Après l'extraction des ARN totaux des cellules crosslinkées *in vivo*, les ARN ont été analysés par Northern Blot avec des sondes spécifiques aux ARN SNORD115 et 7SK (**Figure 53A**).

Nous pouvons observer une différence de quantité des ARN 7SK et SNORD115 purifiés entre les différentes souris. Les souris 673 et 674 sont les souris KO SNORD115 et 675 et 676 sont les souris sauvages. Pour chaque échantillon, un demi cerveau de souris a été utilisé, notifié par les lettres G et D, représentant respectivement les hémisphères cérébraux gauches et droites. Lors des étapes de purification, des quantités de billes équivalentes ont été utilisées. Les différences de quantité d'ARN obtenues sont certainement dues aux grandes variations d'efficacité présentes lors du broyage mécanique, de la digestion à la trypsine, et lors de la filtration des cellules. L'ARN 7SK est utilisé pour contrôler l'extraction des différents échantillons. Nous n'avons pas détecté, comme attendu, la présence de SNORD115 chez les souris KO pour le gène *SNORD115*. A contrario, l'ARN SNORD115 est détecté chez les souris WT.

Pour la détection de la co-purification de l'ARNm 5-HT2C lors de la purification de SNORD115, les échantillons 674G, 674D, 675G et 676G, ayant une quantité plus importante en ARN, ont été utilisés. Après purification de SNORD115 grâce aux MagIC Beads, la détection de l'ARNm du récepteur à la sérotonine 2C dans ces échantillons a été analysée par RT-qPCR (**Figure 53B**). La quantité d'ARN 5-HT2C obtenue pour les échantillons a été comparée à des extraits d'ARN totaux de tissus cérébraux de souris par la méthode couramment utilisée par l'équipe du Dr. Cavaillé (Hebras *et al*, 2020). Les ARN SNORD115 et 5-HT2C ne sont pas détectés dans les échantillons d'ARN purifiés à partir des souris KO SNORD115 674G et 674D, comme attendu. L'ARN SNORD115 a bien été enrichi par les MagIC Beads à partir des ARN totaux des cerveaux 675G et 676G, issus de souris sauvages. Cependant, l'ARNm 5-HT2C n'est pas détecté parmi les ARN enrichis.

Nous n'avons pas co-purifié l'ARNm 5-HT2C lors de la purification de l'ARN SNORD115 via un crosslink au psoralène. L'interaction entre l'ARNm 5-HT2C et l'ARN SNORD115, si elle existe, reste difficile à démontrer.

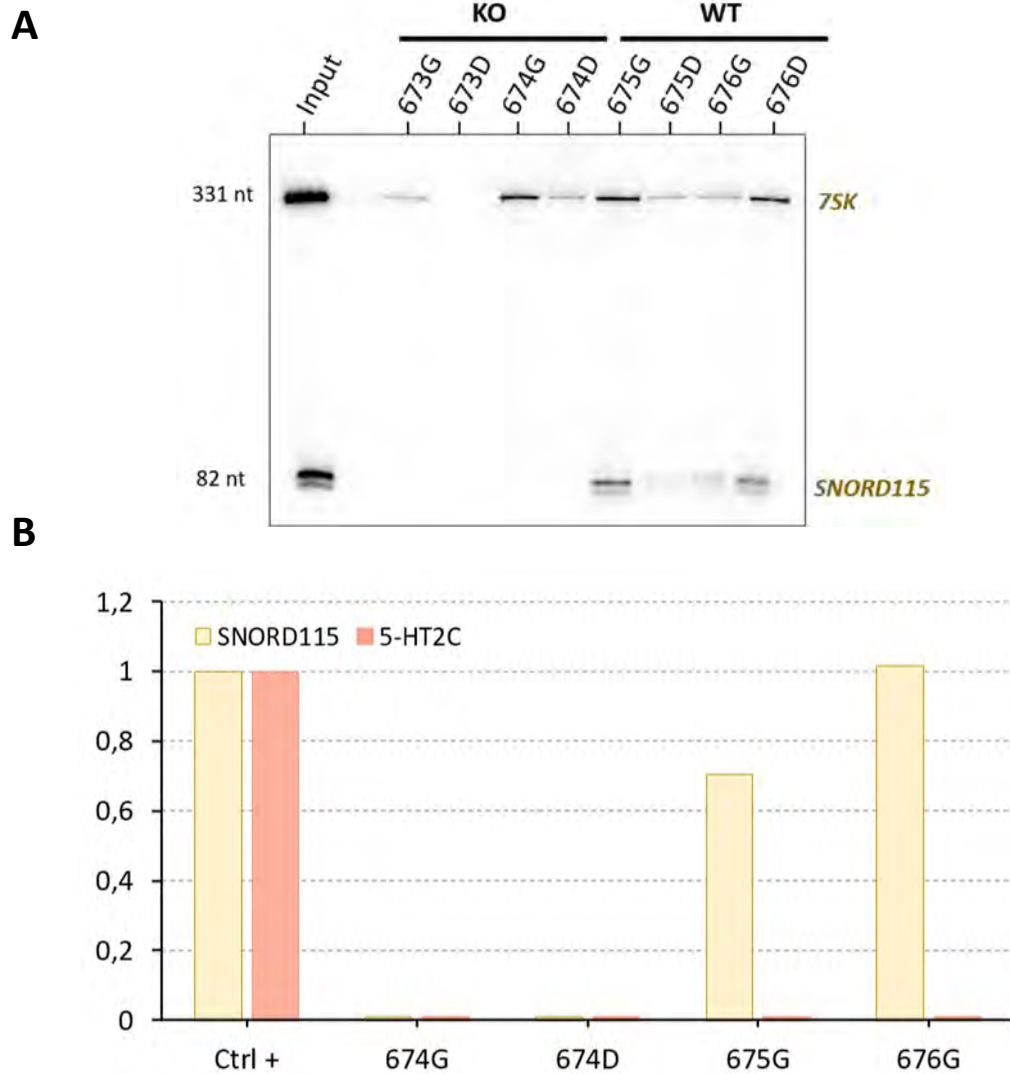


Figure 53 : Co-purification de SNORD115 et 5-HT2C à partir de cerveaux murins

(A) Détection de SNORD115 et 7SK dans les extraits d'ARN totaux des cellules isolées à partir des cerveaux des différentes souris sauvages (675 et 676) ou déplétées du SNORD115 (673 et 674). Les lettres G et D signifient quelles hémisphères cérébrales (Gauche ou Droite) des souris ont été utilisées pour l'expérience. Le contrôle positif sont des ARN totaux issus de cerveaux de souris sauvages obtenus par broyage au Turrax® (broyage mécanique fort). (B) Détection de l'ARN SNORD115 et 5-HT2C par RT-qPCR. RT-qPCR réalisée sur les ARN purifiés issus des différentes souris. Avant extraction, un crosslink à l'AMT psoralène a été effectué *in vivo* sur les cellules isolées des tissus, ensuite l'ARN SNORD115 a été spécifiquement purifié. Le contrôle positif sont des ARN totaux issus de cerveaux murins obtenus par broyage au Turrax® (broyage mécanique fort).

IV. Etude mécanistique et fonctionnelles des snoARN H/ACA

IV. 1. Découverte de nouveaux snoARN H/ACA putatifs

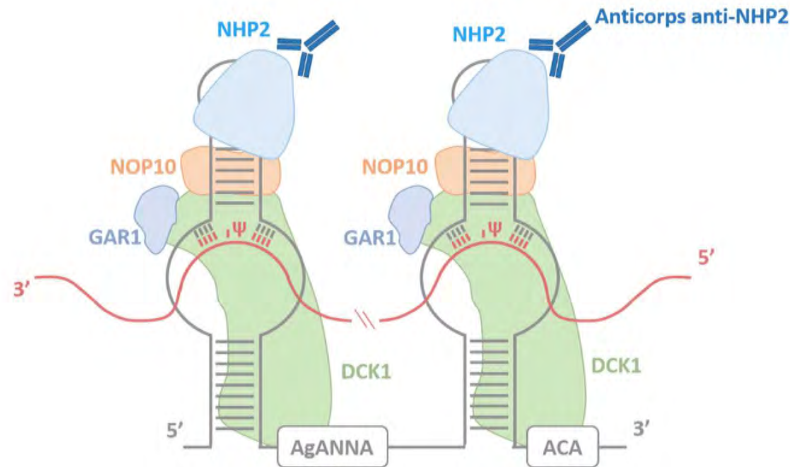
Actuellement, la grande majorité des snoARN localisés dans des régions géniques fortement exprimées ont déjà été découverts et recensés. Ces ARN, comme vu précédemment, ciblent principalement les ARNr et les snARN. L'étude des snoARN situés dans des régions introniques faiblement exprimées pourrait alors aider à la compréhension du rôle des snoARN dans les cellules, ces derniers ciblant potentiellement des ARN cellulaires avec une plus faible abondance, tel que les ARNm. Dans cette optique, un second axe de ma thèse a été de scanner le génome à la recherche de snoARN H/ACA actuellement non répertoriés, puis de vérifier leur accumulation *in vivo*, et de potentiellement découvrir les fonctions et mécanismes qui leur sont associés.

IV. 1. 1. Elaboration d'une liste de potentiels candidats

Les snoARN H/ACA s'assemblent avec les protéines cœurs Dyskérine, NOP10, GAR1 et NHP2 (**Figure 54A**). Une immunoprécipitation de la protéine NHP2 avec un anticorps spécifique a été effectué dans la lignée cellulaire de lymphoblastes TK6, dans l'optique de copurifier la totalité des complexes ribonucléoprotéiques auxquels la protéine NHP2 est rattachée. La fraction d'acide ribonucléique ainsi obtenue a permis la génération d'une librairie d'ADNc, qui a été séquencée à haut débit. Les reads avec des séquences possédant un triplet ACA ou ATA exactement trois nucléotides en amont de l'extrémité 3' ont été alignés au génome humain. Les séquences alignées aux gènes des snoARN H/ACA déjà identifiées dans la littérature ont été écartées. Les séquences restantes ont été considérées comme provenant de potentiels snoARN H/ACA si dans la séquence génomique une ou plusieurs boîtes H putatives pouvaient être identifiées. Les structures secondaires prédites des ARN candidats ont ensuite été analysées, afin de déterminer si elles étaient conformes au modèle en hairpin-hinge-hairpin-tail (**Figure 54B**).

Également, toutes les séquences des snoARN H/ACA putatifs s'alignant avec les éléments Alu du génome ont été écartés, notifiant qu'ils appartiennent à la classe des ARN AluACA. Les éléments répétitifs Alu constituent plus de 10% du génome humain, mais possèdent une expression relativement faible. Ces éléments peuvent être issus de transcrits par

A



B

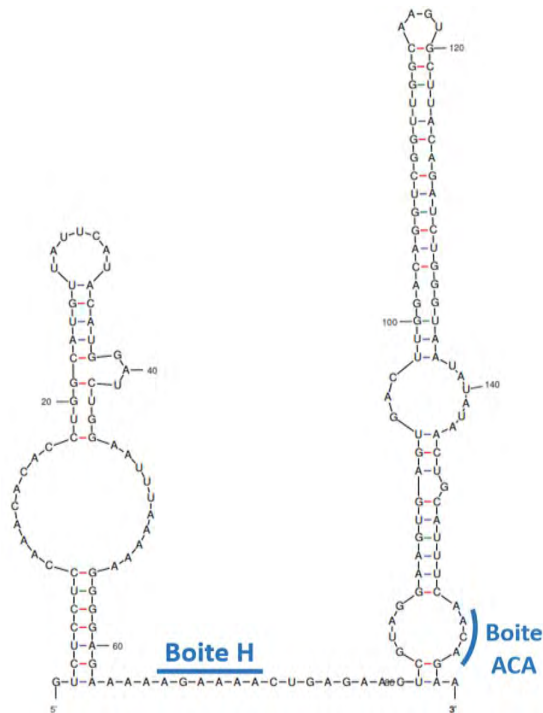


Figure 54 : Etapes permettant l'établissement d'une liste d'ARN H/ACA putatifs

(A) Représentation schématique d'une snoRNP H/ACA. Immunoprécipitation via la protéine cœur NHP2 (B) Structure secondaire prédite du snoARN H/ACA putatif 29, réalisé avec le logiciel mFold.

les polymérase II ou III. Les ARN synthétisés de ces régions peuvent s'accumuler et présentent des caractéristiques analogues aux scaARN, ayant une structure secondaire similaire et possédant les boîtes H, ACA et CAB (Jády *et al*, 2012).

A la suite de ces différentes étapes de sélection, une dizaine de snoARN H/ACA parmi une liste de plus de 600 candidats ont été considérés comme étant de potentiels snoARN H/ACA encore non identifiés (**Figure 55**). Une fois cette liste établie, nous avons testé l'accumulation de ces différents snoARN H/ACA putatifs *in vivo*.

IV. 1. 2. Test d'accumulation des snoARN H/ACA putatifs dans les cellules humaines

Pour tester l'accumulation des snoARN H/ACA putatif dans les cellules, leur présence a été vérifiée par Northern Blot. S'accordant avec le fait que ces snoARN n'ont pas précédemment été répertoriés, leurs expressions n'ont pas pu être détectées via cette méthode (data non montrées).

Pour pallier à cette limite de détection, les gènes correspondants aux différents snoARN H/ACA putatifs, ainsi que les régions d'une cinquantaine de nucléotides les bordants, ont été insérés dans un intron du gène de la β -globine, placé sous le promoteur du Cytomégalo virus (CMV). Les régions situées en amont et en aval des gènes de snoARN peuvent être impliquées dans le repliement de ces derniers. Le vecteur β -globine possède une version tronquée du gène codant pour la globine (Exons 1 à 3). Les snoARN sont insérés dans le second intron, permettant ainsi de simuler les conditions physiologiques de maturation des snoARN.

Les vecteurs obtenus ont été transfectés dans les cellules humaines de la lignée HeLa. Les ARN cellulaires ont ensuite été récupérés et l'accumulation des différents snoARN H/ACA putatifs a été analysée par digestion aux RNases A/T1. Les sondes complémentaires aux snoARN ont été désignées à partir de la région du vecteur comprenant les trois exons de la β -globine ainsi que les régions introniques, contenant le gène du snoARN étudié. Lors de l'étape de digestion aux RNase A et T1, les parties de la sonde correspondant aux trois exons de l'ARNm de la β -globine ont été protégées de la dégradation, au même titre que la région complémentaire au snoARN putatif s'il accumulait suite à sa maturation intronique.

Lors de l'analyse des digestions aux RNases, les exons 1, 2 et 3 de l'ARNm de la globine tronquée ont été détectés, traduisant que des cellules ont bien été transfectées et que le gène hôte a été transcrit. En revanche, aucune bande aux tailles attendues n'ont pas été détectées lors

H/ACA 6, 13^{ème} intron, FAM193A

CTGAAGGGCTGCCTGTACCCAGCCTAGGGTACAGCCCTGGCTGTCACTTAGGGGGCCAAAAGCAATAATGTGATCAG**AGACCAA**
TGGCAACTCAGTAGCTGGCTCGAGGAGGCCTACTAGCCAAAGCTGGG**ACAGAT**

H/ACA 11, 4^{ème} intron, SLC17A9

AGGGGTGTTTCCCTCAGCCGTGCGGGTGGCAGACCCTGAGCGAGCCTACCC**AAAGGA**GTCTGTTTCTCAGTGGCACCAAAAGC
CCTGCCCTGCTGGTCCCTCGGGTGGGAGGTGAGCTAGAGGG**ACAGGT**

H/ACA 29, 1^{er} intron, PTPN11 x2

GTCTCTCCAAACACACCTGGCATGTTATTCATACATGGATCTGGAATTTAAAAAGGGG**AGAAAAAGAAAA**CTGAGAACTCGTAGGA
AGTGAGTGACTTGGACAGGTCGGTTGGCAAGTGCTTACAGATCTGGTAATATATAACTGCATTTC**ACAGAA**

H/ACA 39, 1^{er} intron, DANC

TCCCCACGAGCTTTGCCCGCTCGAGCTCCTCTGCTCCCGCGGCCAGAACGGAGGGGT**AGAGCA**GCCCTCGGCGGCCCGGGGGC
GGGCGCGGTGCCCGTCCCGGGGCTGCGCGAGGC**ACAGGC**

H/ACA 56, 3^{ème} intron, MCOLN2

TGGCCCAACTGCAAACTTATCATTTGTGAAATGAATGGTTGAGTCTCATCGTGGGCCC**AGAGAA**CCTATGGGGACTCTTTATGGAAA
TACCATTTAATACGTATTAATATCTACTGTGCACCCA**ACATTG**

H/ACA 116, 1^{er} intron, SPRY4

CTGCCTCCAAATCAGTGGCGTGATTCTCTCCAGTTTGGGAGAGGCCAGACACACTGGAGGCC**AAAGAA**CTTGAGAGGGGCGGGGA
GCAAGCACTCGTCTGTCTGTAGCTGACCGAGGCACTAACCTCCCTG**ACAAGT**

H/ACA 120, 7^{ème} intron, METTL2

TGTGCCCTTTCTGTAGTGCAAACTGAGGCTGTTTCTGGGCCACCCAGCAGGCAGAGGGGG**AGATCAT**GTCTGCACTGGGCC
ACAGGCCAAGGCAGCGGGTGAAGCTGAAATCCGCCGCTGCTCACTGGCTAGAGCAG**ACTC**

H/ACA 133, 3^{ème} intron, LSP1

TGGGCTGACAGTTCTCCTGGTCAGGGTCTTCTGCCCCCTGAGGGCC**TAACCA**GCCTGGAGCTCTAGGGACCACTGTTGCTGA
GCGACGCTCCCTTGCTCTCT**ACAACC**

H/ACA 186, Overlap 23^{ème} intron/exon, MYO19

CCGGGTGCTGGAGCAGTGTGCCGCTGCATCCAGGGTGGCTGGAGGCGACACCGGCACCG**AGAGCAGG**AGCGGCAGTGGCGGGCC
GTCATGCTCATCCAGGCAGGTAGGTGGCAGGTACCTGTAGCACTGCT**ACAGAC**

H/ACA 265, 8^{ème} intron, BCR x2

AATGGGGAGCTCTGTTGCCCTCCGGCCCCAGAATGGGGGAAAAGTTCAGAGAGACCGAGGCTCCCC**AGAAGA**ACCTCAGGCTGCAG
GTGACAGCAGGGCAAGGTGCTGGTCTGGCCCCAGGACCAAGCTGGCTCCCTGAGAGCACAAGGCCTCATCACTTCTCGAGATGAGG
ACAGGTGCACCGCTGCGGGGAG**ACTG**

Figure 55 : Liste des snoARN H/ACA putatifs

Séquences 5' -> 3' des snoARN H/ACA putatifs. Les régions prédites pour composer les boîtes H et ACA sont en bleus. Pour les ARN 29 et 265, deux séquences ont été prédites pour former un snoARN H/ACA au sein de la même région.

de l'étude de la plupart des snoARN putatifs, signifiant qu'ils ne semblent pas s'accumuler à un niveau détectable au sein des cellules HeLa (**Figure 56A**). En revanche une bande apparaît aux alentours de la taille attendue lors de l'étude de l'ARN H/ACA putatif numéro 120 (**Figure 56B**). Le snoARN H/ACA 120, situé au niveau du septième intron du gène METTL22, semble s'accumuler dans les cellules humaines. Cependant, même via une digestion aux RNases A et T1, aucune bande n'est visible dans les cellules non-transfectées, signifiant un niveau d'expression très faible pour le snoARN endogène.

La maturation du premier intron du gène PTPN11 est prédit pour donner deux snoARN H/ACA putatifs, 29a et 29b, respectivement de 160nt et de 97nt. La séquence de l'ARN putatif 29b est en réalité contenu dans la séquence putative de l'ARN 29a. Une bande très légère en supplément des exons apparaissant vers 104 - 110nt a été observée, puis confirmée avec une sonde plus spécifique via protection aux RNases (**Figure 56C**). Un snoARN issu de la maturation du premier intron du gène PTPN11 semble s'accumuler dans les cellules humaines, avec une taille se rapprochant de celle du snoARN putatif H/ACA 29b.

Parmi 11 snoARN H/ACA putatifs, seuls deux ont révélés s'accumuler *in vivo*, en condition de surexpression. Malgré la sensibilité de la méthode de digestion aux RNases, les snoARN endogènes correspondants n'ont pas pu être détectés.

IV. 1. 3. Le cas du snoARN putatif H/ACA 186

Le snoARN putatif H/ACA 186 possède comme particularité le fait que sa séquence chevauche le 23^{ième} exon et l'intron suivant du gène MYO19. L'exon 23 de du gène *MYO19* est relativement court (130nt). Afin d'éviter l'interférence avec l'épissage tout en testant la capacité de cette région génomique à coder pour un snoARN, le gène du snoARN H/ACA 186 a été inséré sans les séquences le bordant dans le gène de la β -globine tronquée. Cependant, comme les régions adjacentes sont parfois nécessaires à l'accumulation des snoARN, nous avons inséré le même gène avec ces régions le bordant, qui comprenait notamment l'origine du site accepteur d'épissage de l'exon 23. Pour éviter toute interférence, ce site accepteur a été muté (AG → CC).

Des cellules HeLa ont été co-transfectées avec un vecteur d'expression de la Dyskérine tagguée avec un Flag-HA et avec le vecteur contenant le gène de la β -globine tronquée dans lequel a été inséré soit le snoARN H/ACA 186 seul soit le même snoARN mais avec les séquences le bordant et le site accepteur d'épissage muté. Comme contrôle, des cellules HeLa

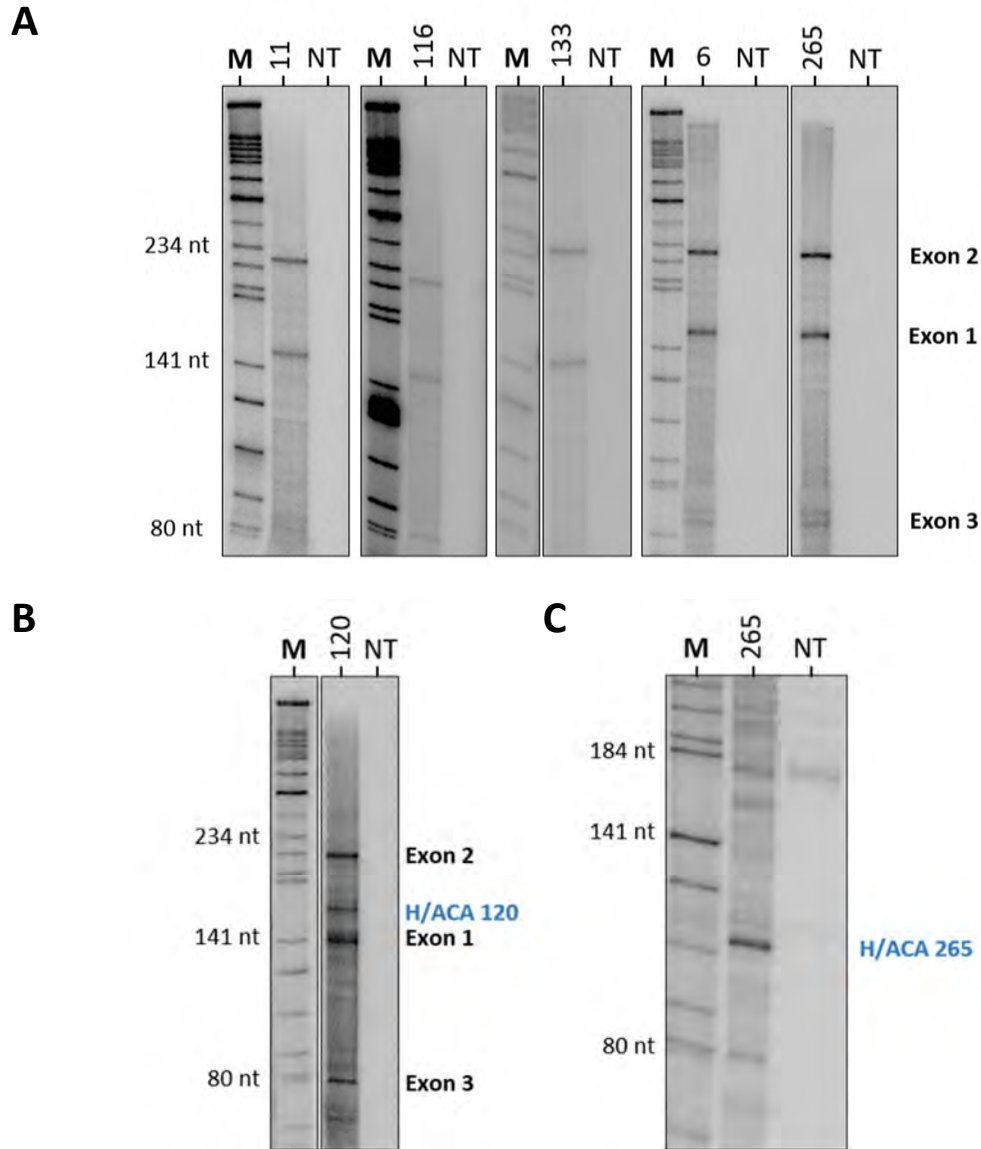


Figure 56 : Test d'accumulation des snoARN H/ACA putatifs

(A) Mapping aux RNases pour tester l'accumulation *in vivo* des snoARN H/ACA putatifs. Les snoARN sont clonés dans le second intron de la β -globine tronquée sous promoteur fort, du vecteur pCGL-1. La sonde antisens est complémentaire aux pré-ARN de β -globine tronquée y compris à son intron portant le snoARN. (B) Mapping aux RNases pour tester l'accumulation *in vivo* du snoARN putatif H/ACA 120. (C) Mapping aux RNases pour tester l'accumulation *in vivo* du snoARN putatif H/ACA 265. La sonde est complémentaire seulement au snoARN 265 et aux régions de 50 nucléotides le bordant.

ont été transfectées seulement avec la Dyskérine tagguée Flag-AH. Les cellules ont ensuite été lysées par sonication et des anticorps reconnaissant le peptide Flag et immobilisés sur billes agarose ont permis une purification spécifique la Fibrillarine tagguée des extraits cellulaires. Les ARN associés à la Dyskérine purifiée ont ensuite été collectés et analysés par digestion aux RNase A et T1 à l'aide de deux sondes complémentaires aux ARN H/ACA 186 et SCARNA10, respectivement (**Figure 57**).

Une bande à la taille attendue est observable vers 330nt, correspondant à la présence de SCARNA10 dans les trois ARN purifiés, signe que les ARN associés à la Dyskérine tagguée ont été correctement purifiés. Une bande diffuse, apparaissant à la taille attendue de 141nt, est observable sur les pistes de gels correspondant à tous les ARN purifiés et l'intensité de cette bande est augmentée quand le snoARN putatif H/ACA 186 est surexprimé.

Le snoARN putatif H/ACA 186 semble s'accumuler dans les cellules humaines, et les régions le bordant ne semble pas être nécessaire à son accumulation. Également, l'enrichissement des snoARN H/ACA associés à la Dyskérine est suffisant pour observer l'accumulation du snoARN H/ACA 186 endogène.

Ces travaux ont permis d'identifier trois nouveaux snoARN H/ACA au sein des cellules humaines, en conditions de surexpression. Seul l'ARN H/ACA 186 endogène a pu être détecté.

IV. 2. Mécanisme de pseudouridylation de deux uridines adjacentes par des snoARN H/ACA humains

IV. 2. 1. Contexte des travaux effectués

Comme vu précédemment, lors de l'interaction entre une snoRNP et sa cible, la poche de pseudouridylation et les protéines cœurs subissent un réarrangement conformationnel de manière à présenter au niveau du site catalytique de la Dyskérine une uridine cible, non appariée aux éléments antisens, permettant ainsi son isomérisation en pseudouridine. Par des mécanismes encore non élucidés, des uridines adjacentes peuvent être isomérisées par une même snoRNP chez le Xénope, la levure et les archées (Schattner *et al*, 2004b; Deryusheva & Gall, 2017; Majumder *et al*, 2020).

Lors de ma thèse, j'ai pu aider l'équipe d'accueil à démontrer que certains snoARN H/ACA humains sont également capable d'effectuer la pseudouridylation de deux uridines adjacentes. Des expériences de mutagénèses nous ont permis de caractériser les mécanismes moléculaires impliqués.

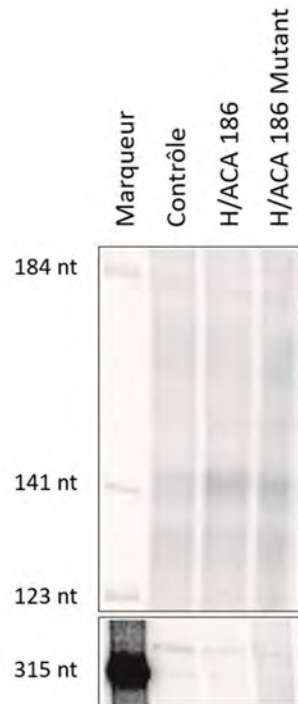


Figure 57 : Test d'accumulation du snoARN H/ACA 186

Des cellules HeLa ont été co-transfectées avec un vecteur exprimant une Dyskérine tagguée FLAG et (i) vecteur pCGL-1 (ii) le vecteur pCGL-1 contenant le gène *H/ACA 186* (iii) le vecteur pCGL-1 contenant le gène *H/ACA 186* avec les régions de 50 nucléotides le bordant et le site d'accepteur d'épissage muté. L'ARN 186 est détecté par digestion aux RNase A/T1 grâce à une sonde complémentaire radiomarquée.

Au cours de cette étude, nous avons mis en évidence que quatre snoARN H/ACA humains, SNORA53, SNORA57, SCARNA8 et SCARNA1 sont capables de guider la synthèse de deux pseudouridines adjacentes, ($\Psi\Psi$ ou $\Psi N\Psi$), au niveau des ARNr 28S, 18S, snU2 et snU2, respectivement. Ces pseudouridylation consécutives sont permises grâce à une réorganisation de l'interaction snoARN-ARN cible, grâce à une remarquable flexibilité des poches de pseudouridylation.

Ma contribution a permis de mettre en évidence que SNORA53 était capable de changer de conformation lors de son interaction avec l'ARNr 28S de manière à effectuer les isomérisations des uridines U3747 et U3749, grâce à la présence d'un « bulge » formé par l'uridine 3752.

De plus, par la génération de mutants de SCARNA1, nous avons démontré que l'efficacité du mécanisme de pseudouridylation dépend des conformations fonctionnelles lors de l'interaction de SCARNA1 avec son ARN cible.

IV. 2. 2. Guide RNA acrobatics: positioning consecutive uridines for pseudo-uridylation by H/ACA

Guide RNA acrobatics: positioning consecutive uridines for pseudouridylation by H/ACA pseudouridylation loops with dual guide capacity

Beáta E. Jály,¹ Amandine Ketele,¹ Dylan Moulis,¹ and Tamás Kiss^{1,2}

¹Molecular, Cellular, and Developmental Biology Department (MCD) UMR 5077, Centre de Biologie Intégrative (CBI), University of Toulouse, Centre National de la Recherche Scientifique, 31062 Toulouse, France; ²Biological Research Centre, 6726 Szeged, Hungary

Site-specific pseudouridylation of human ribosomal and spliceosomal RNAs is directed by H/ACA guide RNAs composed of two hairpins carrying internal pseudouridylation guide loops. The distal “antisense” sequences of the pseudouridylation loop base-pair with the target RNA to position two unpaired target nucleotides 5'-UN-3', including the 5' substrate U, under the base of the distal stem topping the guide loop. Therefore, each pseudouridylation loop is expected to direct synthesis of a single pseudouridine (Ψ) in the target sequence. However, in this study, genetic depletion and restoration and RNA mutational analyses demonstrate that at least four human H/ACA RNAs (SNORA53, SNORA57, SCARNA8, and SCARNA1) carry pseudouridylation loops supporting efficient and specific synthesis of two consecutive pseudouridines ($\Psi\Psi$ or $\Psi N\Psi$) in the 28S ($\Psi 3747/\Psi 3749$), 18S ($\Psi 1045/\Psi 1046$), and U2 ($\Psi 43/\Psi 44$ and $\Psi 89/\Psi 91$) RNAs, respectively. In order to position two substrate Us for pseudouridylation, the dual guide loops form alternative base-pairing interactions with their target RNAs. This remarkable structural flexibility of dual pseudouridylation loops provides an unexpected versatility for RNA-directed pseudouridylation without compromising its efficiency and accuracy. Besides supporting synthesis of at least 6% of human ribosomal and spliceosomal Ψ s, evidence indicates that dual pseudouridylation loops also participate in pseudouridylation of yeast and archaeal rRNAs.

[**Keywords:** pseudouridine; guide RNA acrobatics; box H/ACA RNAs; RNA-guided RNA modification; pseudouridylation; RNA–RNA interaction]

Supplemental material is available for this article.

Received September 28, 2021; revised version accepted November 16, 2021.

Mature cellular RNAs carry >160 different types of post-transcriptionally modified nucleotides (Boccaletto et al. 2018). The most prevalent modified nucleotide is pseudouridine (Ψ), the 5-ribosyl isomer of U, which is present in transfer RNAs (tRNAs), ribosomal RNAs (rRNAs), small nuclear RNAs (snRNAs), and also in polyadenylated mRNAs (Reddy and Busch 1988; Bakin and Ofengand 1998; Carlile et al. 2014; Schwartz et al. 2014; Spenkuch et al. 2014; Li et al. 2015; Taoka et al. 2018). U and Ψ share identical Watson-Crick faces, but the endocyclic N1 group of Ψ can function as an extra proton donor supporting water-coordinated hydrogen bonding that increases the thermodynamic stability of base stacking and the rigidity of the phosphodiester backbone (Arnez and Steitz 1994; Davis 1995; Newby and Greenbaum 2002a,b). Cor-

rect pseudouridylation of spliceosomal snRNAs, rRNAs, and tRNAs is important for efficient and accurate pre-mRNA splicing and protein synthesis (Ge and Yu 2013; Spenkuch et al. 2014; De Zoysa and Yu 2017; Sloan et al. 2017; Bohnsack and Sloan 2018; Adachi et al. 2019). The functional significance of mRNA pseudouridylation is still unclear, but artificially introduced Ψ s can alter stop codon reading, increase mRNA stability and translation efficiency, and suppress RNA immunogenicity (Karikó et al. 2008; Karijolic and Yu 2011; Fernández et al. 2013; Hoernes et al. 2016).

RNA pseudouridylation is achieved by stand-alone and guide RNA-associated Ψ synthases. The stand-alone Ψ synthases recognize local RNA structures and frequently show “region and/or multisite specificity” (Spenkuch et al. 2014). Site-specific synthesis of eukaryotic

Corresponding author: tamas.kiss@univ-tlse3.fr

Article published online ahead of print. Article and publication date are online at <http://www.genesdev.org/cgi/doi/10.1101/gad.349072.121>. Freely available online through the *Genes & Development* Open Access option.

© 2022 Jály et al. This article, published in *Genes & Development*, is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

spliceosomal and eukaryotic and archaeal ribosomal Ψ s located in diverse sequence and structural environments is mediated by H/ACA guide RNA-associated Ψ synthases (Ganot et al. 1997a; Ni et al. 1997; Darzacq et al. 2002). Through base-pairing with the target RNA, the guide RNAs provide high sequence specificity for the associated Ψ synthase, making the pseudouridylation reaction independent from local two-dimensional RNA structures. While archaeal H/ACA RNAs contain one to three hairpin structures, eukaryotic H/ACA RNAs are composed of two hairpins that are connected and tailed with short single-stranded sequences carrying the conserved H (AgAnnA) and ACA (less frequently, AUA) boxes, respectively (Fig. 1A; Balakin et al. 1996; Kiss et al. 1996; Ganot et al. 1997b). The molecular principles of RNA-guided pseudouridylation were established more than two decades

ago (for reviews, see Kiss et al. 2010; Watkins and Bohnsack 2012; Lui and Lowe 2013; Yu and Meier 2014; Czekay and Kothe 2021). The H/ACA hairpins contain internal pseudouridylation guide loops carrying bipartite target recognition (antisense) elements preceding and following the distal stem (DS) that closes the pseudouridylation loop (Ganot et al. 1997a). Base-pairing of the bipartite antisense sequences with the substrate RNA forms an RNA three-way helical junction positioning two unpaired substrate nucleotides (5'-UN-3') into the "pseudouridylation pocket" at the base of the DS (Fig. 1B). Invariantly, the unpaired 5' U that is separated by 13–16 nt from the H or ACA box of the guide RNA is converted into Ψ .

Each H/ACA hairpin associates with four evolutionarily conserved proteins: Nop10, Gar1, Nhp2 (L7Ae in archaea), and the Ψ synthase dyskerin (also called NAP57 in mice and Cbf5 in yeast and archaea) (Fig. 1B). First, the structure of archaeal single-hairpin ACA RNP has been solved (Li and Ye 2006; Liang et al. 2007, 2009; Duan et al. 2009). The ACA RNA hairpin is anchored to the core trimer of Cbf5, Nop10, and L7Ae through bindings of Cbf5 to the ACA box and the proximal stem (PS) and L7Ae to the kink (K)-turn or K-loop in the apical stem-loop. These interactions position the selected target U at the catalytic center of Cbf5. Gar1 associates with the catalytic domain of Cbf5 and controls its activity through promoting accurate substrate placing and releasing (Duan et al. 2009; Wang et al. 2015). Eukaryotic double-hairpin H/ACA RNPs have similar structural organization (Egan and Collins 2010; Li et al. 2011). In contrast to archaeal L7Ae, eukaryotic Nhp2 binds to the distal stem-loop independently from a K-turn motif. While Nhp2 alone has little H/ACA RNA binding affinity, upon incorporation into the dyskerin–Nop10–Nhp2 core complex, it binds to the distal stem-loop with high affinity and specificity (Wang and Meier 2004; Baker et al. 2005; Li et al. 2011). Eukaryotic H/ACA RNPs directing rRNA pseudouridylation accumulate in the nucleolus and are called small nucleolar RNPs (snoRNPs). Guide RNPs mediating snRNA pseudouridylation concentrate in the nucleoplasmic Cajal body (CB) and are termed small CB-specific RNPs (scARNPs) (Darzacq et al. 2002).

Due to their stringent target selection strategy, the H/ACA pseudouridylation guide loops are believed to position only one substrate U for pseudouridylation in their target sequences (Fig. 1B). However, arguing against this notion, the 3'-terminal pseudouridylation loops of *Xenopus* pugU2-43, yeast snR82, and archaeal sR-h45 H/ACA RNAs have been implicated in synthesis of two consecutive spliceosomal and ribosomal Ψ s ($\Psi\Psi$ and $\Psi N\Psi$) through unclear mechanisms (Schattner et al. 2004; Deryusheva and Gall 2013, 2017; Majumder et al. 2020). Here, functional characterization of human H/ACA RNAs reveals that at least four human H/ACA RNAs possess pseudouridylation loops with dual guide capacity. We demonstrate that the 5'-terminal pseudouridylation loop of SNORA53 and the 3'-terminal pseudouridylation loops of SNORA57, SCARNA8, and SCARNA1 direct formation of Ψ 3747/ Ψ 3949 in the 28S rRNA, Ψ 1045/ Ψ 1046 in the 18S rRNA, and Ψ 43/ Ψ 44

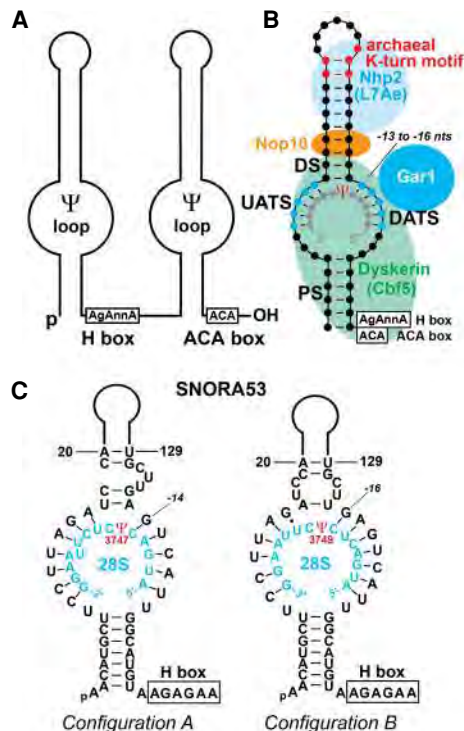


Figure 1. Selection of substrate uridines for pseudouridylation by box H/ACA guide RNAs. (A) Schematic structure of eukaryotic bipartite box H/ACA pseudouridylation guide RNAs. The consensus sequences of the conserved H and ACA box motifs and the internal pseudouridylation guide loops (Ψ -loop) are shown. (B) Schematic consensus structure of the 5' and 3' hairpins of H/ACA pseudouridylation guide RNPs. The antisense sequences (blue dots) and the selected target sequences (gray dots) with the target U (Ψ) positioned for pseudouridylation are shown. The proximal and distal stems (PS and DS) together with the upstream and downstream antisense target stems (UATS and DATS) are indicated. The four H/ACA core proteins (dyskerin, Nop10, Nhp2, and Gar1) are shown. Archaeal ACA RNAs contain K-turns or K-loops. (C) Potential base-pairing interactions of the proposed 5'-terminal pseudouridylation guide loop of human SNORA53 with 28S rRNA positioning U3747 or U3749 for pseudouridylation.

and $\Psi 89/\Psi 91$ in the U2 spliceosomal snRNA, respectively. A detailed mutational analysis of SNORA53, SNORA57, and SCARNA1 confirmed that dynamic alternative guide RNA and substrate RNA base-pairing interactions, termed guide RNA acrobatics, support the sequence specificity and efficacy of the dual pseudouridylation guide loops of human H/ACA RNAs. Since the previously reported dual pseudouridylation guide loops of *Xenopus* pugU2-43, yeast snR82, and archaeal sR-h45 H/ACA RNAs can also form alternative guide-target interactions that can correctly position their target Us for pseudouridylation, we propose that H/ACA guide RNA acrobatics is an evolutionarily conserved strategy of pseudouridylation guide RNAs for accurate and efficient synthesis of two proximate Ψ s.

Results

Human SNORA53 is required for synthesis of both $\Psi 3747$ and $\Psi 3749$ in the 28S rRNA

Human SNORA53, originally named ACA53 (Kiss et al. 2004), is a box H/ACA “orphan” snoRNA, as it has not been assigned for RNA modification or for any other cellular function. Closer inspection of SNORA53 revealed that its 5'-terminal hairpin can be folded into two alternative structures accommodating slightly different putative pseudouridylation loops that, in principle, could direct synthesis of $\Psi 3747$ and $\Psi 3749$ in the 28S rRNA (Fig. 1C). Thus far, no H/ACA snoRNAs have been proposed to govern $\Psi 3747$ and $\Psi 3749$ synthesis (Jorjani et al. 2016). However, in the newly suggested SNORA53–28S rRNA interactions, the alleged pseudouridylation loops, instead of being closed by perfect distal stems, are topped by unusually unstable structures encompassing several bulged and unpaired nucleotides. Moreover, in configuration A assigned for synthesis of $\Psi 3747$, the upstream anti-sense target stem (UATS) is interrupted with the bulged U3742 residue in the 28S rRNA sequence. To test SNORA53 participation in $\Psi 3747$ and/or $\Psi 3749$ synthesis, expression of SNORA53 was knocked out in human HAP1 cells. Mammalian box H/ACA guide RNAs are synthesized within introns of protein-coding genes and are processed from the removed and debranched host introns. The human SNORA53 gene is located within the fifth intron of the *SLC25A3* gene in sense orientation, indicating that SNORA53 is a canonical intronic snoRNA (Fig. 2A). A fragment of the fifth intron of *SLC25A3* encompassing the SNORA53 gene was excised by CRISPR/Cas9 genome editing. The correctness of SNORA53 deletion was verified by PCR amplification and gel electrophoretic analysis of the truncated *SLC25A3* gene fragment (Fig. 2B). RT-PCR detected comparable levels of correctly spliced mature *SLC25A* mRNA in the SNORA53 knockout (KO) and control HAP1 cells, demonstrating that the removal of the SNORA53 gene had no detectable impact on the expression of the *SLC25A3* host gene (Fig. 2C). Finally, RNase A/T1 mappings performed with an internally labeled sequence-specific RNA probe failed to detect SNORA53 expression in SNORA53-KO cells, although accumula-

tion of the U85 control intronic scaRNA remained unaltered (Fig. 2D).

To compare the pseudouridylation states of 28S rRNA at U3747 and U3749 in the presence and absence of SNORA53, cellular RNAs extracted from HAP1 and SNORA53-KO cells were incubated with CMCT [N-cyclohexyl-N'-(2-morpholinoethyl)carbodiimide metho-p-toluenesulfonate], which efficiently and irreversibly reacts with Ψ (Bakin and Ofengand 1998). The covalently bound bulky CMCT molecule blocks reverse transcriptase elongation 1 nt before the modified Ψ (Fig. 2E). Primer extension analysis of CMCT-treated control HAP1 RNAs with a terminally labeled 28S rRNA-specific oligonucleotide primer detected several known Ψ s, including $\Psi 3737$, $\Psi 3741$, $\Psi 3742$, $\Psi 3747$, and $\Psi 3749$ (Fig. 2E, lane 1). In contrast, primer extension mapping of CMCT-modified 28S rRNA from SNORA53-KO cells failed to detect pseudouridylation of U3747 and U3749, although synthesis of $\Psi 3737$, $\Psi 3741$, and $\Psi 3742$ remained unaffected (Fig. 2E, lane 3).

To exclude the formal possibility that disruption of U3747 and U3749 pseudouridylation was a consequence of undesired CRISPR/Cas9 off-target activity, we restored SNORA53 accumulation in the SNORA53-KO cell line by using the pIRESpuro/GL expression plasmid (Fig. 2F; Vitali and Kiss 2019). After selection of a puromycin-resistant monoclonal cell line, accumulation of SNORA53 was confirmed by RNase A/T1 mapping (Fig. 2F, lane 3). Primer extension mapping of 28S rRNA pseudouridylation showed that restoration of SNORA53 accumulation re-established pseudouridylation of U3747 and U3749, demonstrating that SNORA53 is responsible for the synthesis of both $\Psi 3747$ and $\Psi 3749$ in the human 28S rRNA (Fig. 2E, lane 5).

The 5'-terminal pseudouridylation loop of SNORA53 directs synthesis of both $\Psi 3747$ and $\Psi 3749$ in the 28S rRNA

The 3' hairpin of SNORA53 lacks evolutionarily conserved sequences complementary to rRNAs or other stable cellular RNAs (Supplemental Fig. S1A), suggesting that the elements supporting pseudouridylation of 28S-U3747 and 28S-U3749 are located in the 5'-terminal hairpin of SNORA53. To validate this assumption and to learn about the molecular mechanism supporting the predicted double-guide activity of the 5'-terminal pseudouridylation loop of SNORA53, we performed mutational analysis of SNORA53.

The 5' hairpins of SNORA53 RNAs derived from phylogenetically distant species fold into similar two-dimensional structures composed of a long major hairpin and a short internal stem-loop branched out from the main hairpin (Fig. 3A; Supplemental Fig. S1A). However, upon expression of truncated SNORA53 RNAs lacking either the short internal stem-loop (SNORA53d1) or the distal part of the major 5' hairpin (SNORA53d2) in SNORA53-KO cells (Fig. 3A,B), synthesis of both $\Psi 3747$ and $\Psi 3749$ was restored (Fig. 3C, lanes 5,7), indicating that all guide elements supporting U3747 and U3749 pseudouridylation

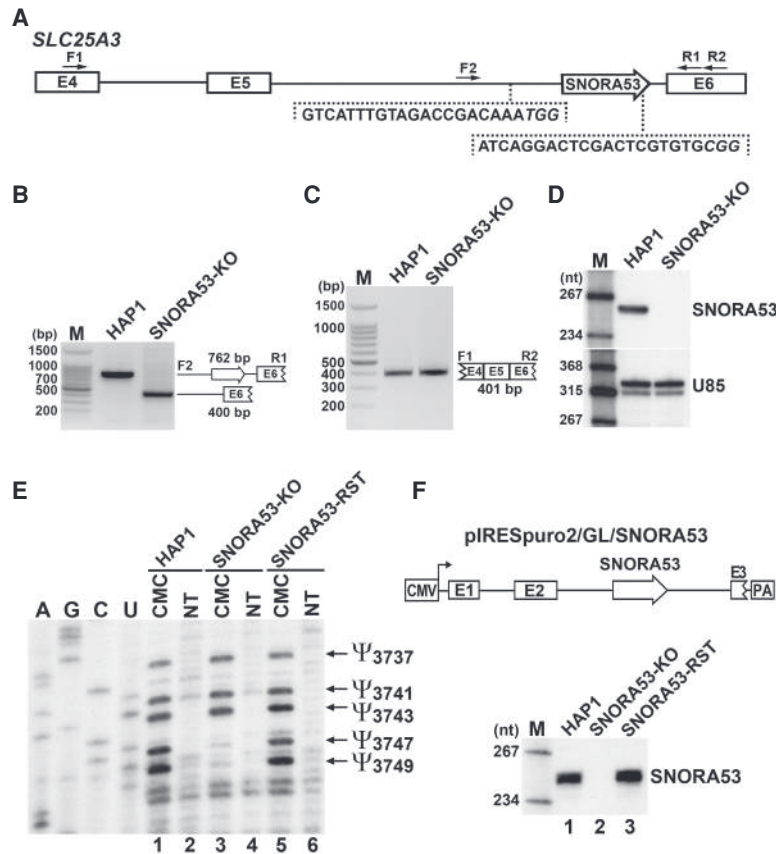


Figure 2. Human SNORA53 is required for synthesis of both Ψ3747 and Ψ3749 in the 28S rRNA. (A) A strategy for CRISPR/Cas9-mediated deletion of human *SNORA53* gene in HAP1 cells. Schematic structure of a fragment of the human *SLC25A3* gene encompassing exons 4–6 (E4–E6) and the intronic *SNORA53* gene (open arrow) is shown. Positions of the forward (F1 and F2) and reverse (R1 and R2) primers used for genomic and RT-PCR analysis are indicated. The genomic sequences targeted by sgRNAs are presented in sense orientation. The PAM sequences are in italics. (B) PCR analysis of genomic DNAs extracted from *SNORA53* knockout (*SNORA53*-KO) and parental HAP1 cells. The amplified DNA fragments were analyzed on 1% agarose gel. (Lane M) DNA size markers in base pairs (bp). Schematic structures and expected lengths of the amplified fragments are shown at the right. (C) RT-PCR analysis of the accumulation of spliced *SLC25A3* mRNA in *SNORA53*-KO and HAP1 cells. Structure of the amplified cDNA fragment is shown. (D) *SNORA53* accumulation. RNAs isolated from HAP1 and *SNORA53*-KO cells were annealed with internally labeled *SNORA53*- or U85-specific antisense RNA probes and digested with a mixture of RNase A and T1. The protected RNA fragments were analyzed on a 6% sequencing gel. (Lane M) Single-stranded DNA size markers in nucleotides (nt). (E) Mapping of 28S rRNA pseudouridylation in HAP1 and *SNORA53*-KO cells before and after restoration of *SNORA53* expression (*SNORA53*-RST). Ψs covalently modified by CMCT were detected by primer extension analysis using a terminally labeled 28S-specific primer and

AMV reverse transcriptase. (Lanes A,G,C,U) Dideoxy sequencing reactions performed on 28S ribosomal DNA with the same 28S-specific primer. (Lanes NT) Primer extension reactions performed on nontreated control RNAs. Ψs corresponding to the detected RT stop signals are indicated at the right. (F) Restoration of *SNORA53* accumulation in *SNORA53*-KO cells. Schematic structure of the pIRESpuro2/GL/*SNORA53* expression construct is shown. The coding region of the human *SNORA53* gene (open arrow) was inserted into the second intron of the truncated β-globin gene (E1–E3) placed under the control of the cytomegalovirus (CMV) promoter and the globin polyadenylation signal (PA). The resulting CMV-globin-*SNORA53* expression unit was inserted into the pIRESpuro2 expression plasmid and transfected into *SNORA53*-KO cells. Puromycin-resistant cell colonies were isolated, and expression of *SNORA53* in *SNORA53*-RST cells was verified by RNase A/T1 mapping.

are confined to the proximal part of the *SNORA53* 5' hairpin. Consistent with its functional importance, the proximal part of the *SNORA53* 5' hairpin contains many phylogenetically conserved nucleotides (Fig. 3A; Supplemental Fig. S1). Significantly, all nucleotides predicted to participate in alternative intra- and intermolecular base-pairing interactions positioning either U3747 or U3749 for pseudouridylation are conserved in the known *SNORA53* sequences (Fig. 1C; Supplemental Fig. S1).

To unambiguously demonstrate that the 5'-terminal pseudouridylation loop region of *SNORA53* is able to direct synthesis of both 28S-Ψ3747 and 28S-Ψ3749, we attempted to separate the two pseudouridylation guide activities of *SNORA53* through construction of mutant *SNORA53* RNAs directing either Ψ3747 or Ψ3749 formation. In the proposed *SNORA53*–28S interactions, in both configuration A and B, the C10–U13 upstream and the G136–U140 downstream antisense sequences are base-paired with the A3753–G3756 and A3742–C3746 regions of 28S rRNA (Fig. 3D, shaded helices). In the center of the guide RNA–substrate RNA three-way helical

junction, alternative base-pairing interactions are predicted to govern the structural rearrangements positioning either U3747 or U3749 for pseudouridylation (Fig. 3D, middle panel). Deletion of the 131-CUU-133 bulged nucleotides (CUU133del) was expected to lock the *SNORA53*–28S rRNA interaction in configuration A and to prevent formation of functional configuration B. Replacement of the 131-CU-132 nucleotides for GA was predicted to stabilize configuration B and to interfere with folding into configuration A. When expressed in *SNORA53*-KO cells, the *SNORA53*-CUU133del snoRNP efficiently synthesized Ψ3747, but failed to convert U3749 into Ψ (Fig. 3C, lane 9). On the contrary, the *SNORA53*-CU132GA snoRNP supported Ψ3749 formation, but failed to synthesize Ψ3747 (Fig. 3C, lane 11). Thus, demonstration that sequence alterations can restrict the pseudouridylation guide ability of *SNORA53* to the synthesis of either Ψ3747 or Ψ3749 in a predictable way validated the proposed dual guide capacity of the 5'-terminal pseudouridylation loop of *SNORA53* and strongly supported the correctness and functional relevance of the

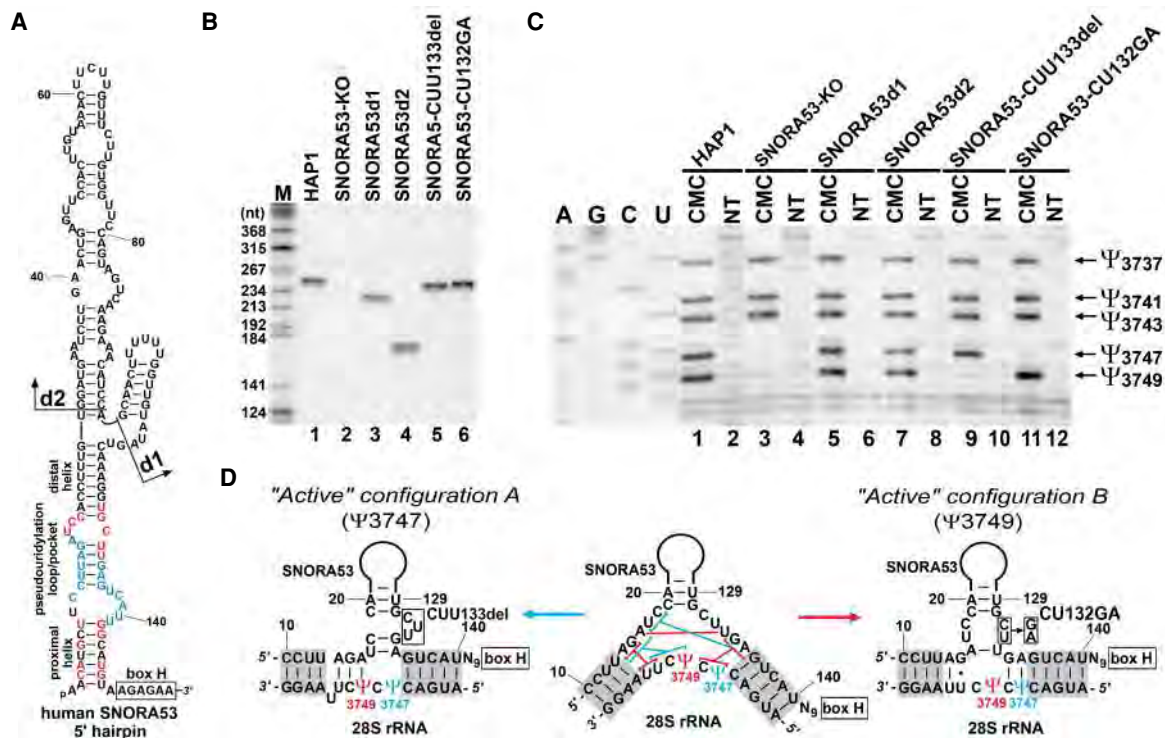


Figure 3. The 5'-terminal pseudouridylation loop of SNORA53 directs pseudouridylation of U3747 and U3749 in the 28S rRNA. (A) A computer-predicted secondary structure of the 5' hairpin of human SNORA53 based on minimum free energy calculation (Zuker 2003). The evolutionarily invariant nucleotides predicted to base-pair with 28S rRNA are in blue. Other conserved nucleotides are in red. Arrows indicate internal deletions d1 and d2. (B) Expression of mutant SNORA53 RNAs in SNORA53-KO cells monitored by RNase A/T1 protection with sequence-specific antisense RNA probes. As controls, accumulation of endogenous SNORA53 was also tested in HAP1 and SNORA53-KO cells. (Lane M) Single-stranded DNA size markers. (C) Primer extension mapping of 28S rRNA pseudouridylation in HAP1 and SNORA53-KO cells either lacking or expressing mutant SNORA53 RNAs as indicated above the lanes. For details, see the legend for Figure 2E. (D) Proposed structural rearrangements of the three-way helical junctions formed by the 5'-terminal pseudouridylation loop of human SNORA53 and 28S rRNA positioning either U3747 or U3749 for pseudouridylation. The proposed alternative base-pairing interactions leading to "active" configuration A or B are indicated by blue and red lines, respectively. The common base-pairing interactions present in both configurations are shaded. The deleted or altered nucleotides in mutant SNORA53-CUU133del and SNORA53-CU132GA RNAs are boxed.

proposed alternative base-pairing interactions of SNORA53 with 28S rRNA positioning U3747 or U3749 for pseudouridylation.

The bulged ribosomal U3752 residue supports the dual guide activity of SNORA53

Normally, the antisense elements of pseudouridylation guide RNAs form perfect double helices with their target sequences (Fig. 1B; Ganot et al. 1997a). However, in configuration A positioning U3747 for pseudouridylation, the upstream antisense target stem (UATS) formed by the upstream antisense element of SNORA53 (10-CCUUAGA-17) and its 28S rRNA target sequence (3749-UCUUAAGG-3756) is interrupted by the bulged U3752 ribosomal residue (Figs. 1C, 3D). Given that the SNORA53 antisense and the 28S target sequences are conserved, we speculated that bulging of 28S-U3752 may contribute to the dual guide capacity of the 5'-terminal pseudouridylation loop of SNORA53. To test this hypothesis, an extra A residue was inserted into SNORA53

between U13 and A14 in order to base-pair with the bulged ribosomal U3452 residue (Fig. 4A). Expression of the mutant SNORA53-U13+A RNA in SNORA53-KO cells restored pseudouridylation of U3747, but failed to support formation of Ψ3749 (Fig. 4A, lane 5).

Next, we tested the *in vivo* pseudouridylation of 28S rRNA sequences lacking U3752. Small fragments of the human 28S rRNA from position G3729 to G3759 either containing (28S-WT) or lacking (28S-U3752del) U3752 were transiently expressed in mouse cells by using the pW(Xb/Xh) mouse ribosomal minigene expression plasmid (Fig. 4B; Hadjiolova et al. 1994). Since the ribosomal minigene is transcribed by RNA polymerase I in the nucleolus, rRNA sequences imbedded within the nascent minigene transcript can undergo snoRNP-mediated modifications (Ganot et al. 1997a). Primer extension analysis after CMCT treatment clearly demonstrated that synthesis of Ψ3749 was largely inhibited or fully abolished in the 28S-U3752del transcript (Fig. 4B, lane 3), but it was unaffected in the control 28S-WT RNA (Fig. 4B, lane 1). Importantly, both 28S-U3752del and 28S-WT RNAs were

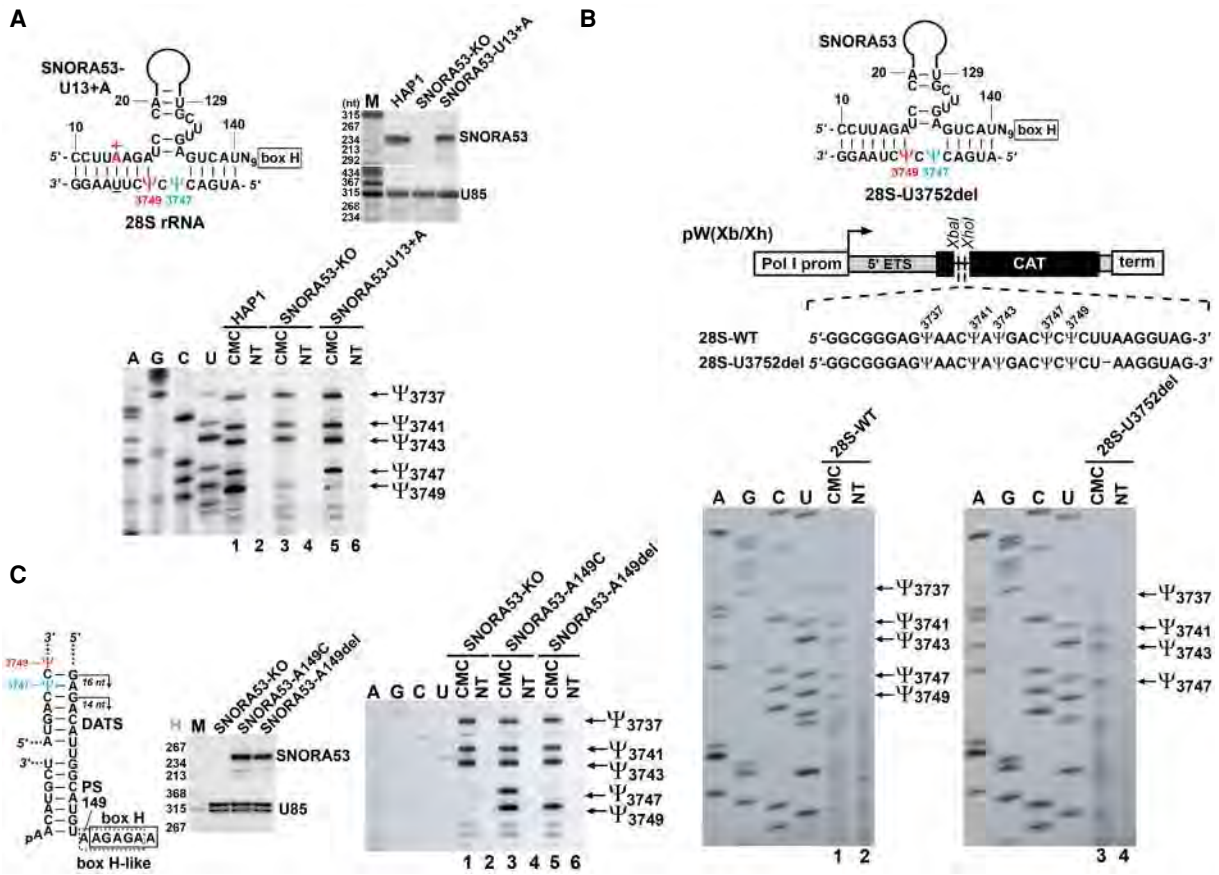


Figure 4. Structural requirements of the dual pseudouridylation guide activity of the 5' hairpin of SNORA53. (A) In vivo pseudouridylation of 28S rRNA with a mutant SNORA53 RNA. (*Top left panel*) A predicted interaction of endogenous 28S rRNA with an ectopically expressed mutant SNORA53 RNA carrying an extra A (+, in red) inserted between U13 and A14. The 28S-U3752 residue is underlined. Please note that 28S-U3751 could also bulge out in configuration A, but the upstream antisense element of SNORA53-U13+A is able to base-pair with both U3751 and U3752. Accumulation of SNORA53-U13+A in SNORA53-KO cells was verified by RNase mapping. As controls, accumulation of endogenous SNORA53 and U85 RNAs was also tested. (*Bottom panel*) Pseudouridylation of 28S rRNA was monitored by primer extension mapping in HAP1 and SNORA53-KO cells lacking or expressing SNORA53-U13+A. For other details, see the legend for Figure 2E. (B) In vivo pseudouridylation of transiently expressed mutant 28S rRNA (28S-U3752del) sequences lacking U3752 in mouse 3T3 cells. (*Top panel*) A predicted interaction of endogenous mouse SNORA53 with 28S-U3752del rRNA and schematic structure of the mouse pW[Xb/Xh] expression construct are shown. The mouse RNA polymerase I promoter (Pol I prom) and terminator (term), the 5'- and 3'-terminal regions of the 5' and 3' external transcribed sequences (ETS; gray boxes), and a fragment of the chloramphenicol acetyltransferase (CAT) gene are indicated (Hadjilova et al. 1994). The nucleotide sequences of wild-type (28S-WT) and mutant (28S-U3752del) 28S rRNA fragments with the known ribosomal Ψs are shown. (*Bottom panel*) Pseudouridylation of the expressed 28S-WT and 28S-U3752del ribosomal minigene transcripts was monitored by primer extension analysis. (C) Correct positioning of the H box relative to the selected substrate Us is essential for the dual guide activity of the 5' hairpin of SNORA53. (*Left panel*) Interaction of the descending strand of the 5' hairpin of SNORA53 with 28S rRNA positioning U3747 and U3749 for pseudouridylation. A putative H-box-like motif is highlighted in the dashed box. The A149 residue was deleted in SNORA53-A149del and replaced for C in SNORA53-A149C. (*Middle panel*) Accumulation of ectopically expressed SNORA53-A149C and SNORA53-A149del RNAs and endogenous U85 scaRNA in SNORA53-KO cells was measured by RNase mapping. (*Right panel*) Primer extension mapping of 28S rRNA pseudouridylation in SNORA53-KO cells lacking or expressing SNORA53-A149C or SNORA53-A149del RNAs.

faithfully pseudouridylated at Ψ 3747 and other known pseudouridylation sites: Ψ 3737, Ψ 3741, and Ψ 3743. Thus, stabilization of the upstream antisense target stem (UATS) of SNORA53 in configuration A, either by removal of the bulged U3752 ribosomal residue or by insertion of a complementary A residue into the upstream antisense sequence of SNORA53, promoted synthesis of Ψ 3747 and prevented formation of Ψ 3749, suggesting that accurate function of the dual guide loops necessitates comparable

thermodynamic stabilities of the alternative antisense target base-pairing interactions.

In most H/ACA guide RNA–substrate RNA interactions, the selected target Us are separated from the H or ACA boxes by 14- to 15-nt-long guide RNA sequences (Supplemental Fig. S2). In the proposed alternative SNORA53–28S rRNA interactions, U3747 is positioned at the optimal 14-nt distance from the H box, but U3749 is separated by 16 nt, which is a less common

configuration (Fig. 4C, left panel). Besides the canonical H box (150-AGAGAA-155), SNORA53 contains another overlapping H-box-like sequence (149-AAGAGA-154) that is separated only by 15 nt from U3749. However, disruption of the potential dyskerin binding capacity of the H-like motif through conversion of A149 into C had no effect on the synthesis of 28S-Ψ3747 and 28S-Ψ3749, suggesting that dyskerin bound to the canonical H box of SNORA53 can convert both U3747 and U3749 into Ψ (Fig. 4C, lane 3). More unexpectedly, deletion of A149 that reduced the distance between the H box of SNORA53-A149del and the U3747 target residue from 14 to 13 nt fully abolished pseudouridylation of U3747 (Fig. 4C, lane 5), contrary to the fact that 13-nt-long spacers separating target Us from the H or ACA boxes are relatively frequent (Supplemental Fig. S2). In addition to the H and ACA motifs, Cbf5/dyskerin also interacts with the minor groove of the proximal stem (PS) and the downstream antisense target stem (DATS) through its PUA (pseudouridine synthase and archaeosine transglycosylase) and catalytic domains, respectively (Fig. 1B). The PS-DATS joint helical stack seems to serve as a molecular ruler measuring the distance between the target Us and H/ACA boxes. In the SNORA53–28S rRNA interaction, the PS and DATS helices are separated by only one unpaired

loop nucleotide (U141) (Fig. 4C, left panel). This may limit the structural plasticity of the PS-DATS helical stack of SNORD53-A149del required for correct positioning of U3747 for pseudouridylation.

Human H/ACA pseudouridylation loops with potential dual guide capacities

To identify additional human H/ACA pseudouridylation loops with potential dual guide capacities, we collected the human spliceosomal and ribosomal Ψs that are located next to each other (ΨΨ) or are separated by 1 nt (ΨNΨ) and listed them together with the corresponding H/ACA hairpins assigned for their site-specific selection (Fig. 5A; Lestrade and Weber 2006; Jorjani et al. 2016; Taoka et al. 2018). This compilation confirmed that most Ψs, even in densely pseudouridylated regions, are selected individually by dedicated pseudouridylation loops. Nevertheless, inspection of the listed H/ACA hairpins revealed that the 3'-terminal pseudouridylation loops of SCARNA14, SCARNA8, and SCARNA1, which had been earlier assigned for pseudouridylation of the U2 snRNA at Ψ7, Ψ44, and Ψ89, respectively (Darzacq et al. 2002; Kiss et al. 2004; Schattner et al. 2006), might also direct synthesis of the neighboring "orphan" Ψ6, Ψ43, and

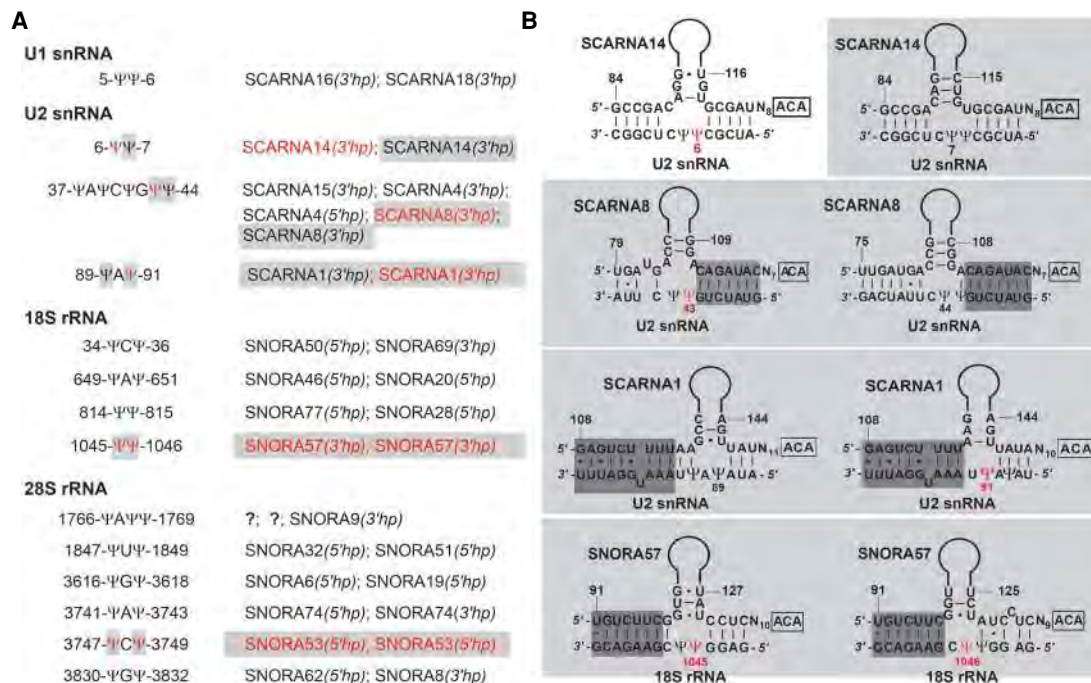


Figure 5. Human pseudouridylation guide loops with possible dual guide capacities. (A) Closely located human spliceosomal and ribosomal Ψs are listed together with the corresponding pseudouridylation guide RNA hairpins (5'hp and 3'hp) connected to their synthesis. Hairpins newly implicated in dual pseudouridylation reactions and their predicted target Ψs are highlighted in red. Participation of the shaded H/ACA hairpins in the synthesis of their predicted target Ψs has been experimentally validated in this study. (B) Potential alternative base-pairing interactions of H/ACA pseudouridylation guide loops with their target RNAs positioning consecutive ribosomal and spliceosomal Us for pseudouridylation. Direction of the synthesis of Ψ7, Ψ44, and Ψ89 in the U2 snRNA by human SCARNA14, SCARNA8, and SCARNA1, respectively, has been proposed earlier (Darzacq et al. 2002; Kiss et al. 2004; Schattner et al. 2006). The newly predicted target Ψs are highlighted in red. The functional interactions shaded in light gray have been experimentally validated in this work. Base-pairing interactions common in the alternative functional configurations are shaded in dark gray. For the interaction of SNORA53 and 28S rRNA, see Figure 3D.

Ψ 91 residues through formation of alternative base-pairing interactions (Fig. 5B). In fact, the 3' hairpin of the *Xenopus* equivalent of human SCARNA8, pugU2-34/44, has already been reported to direct pseudouridylation of the U2 snRNA at both U43 and U44 (Deryusheva and Gall 2013). Finally, re-examination of human orphan H/ACA RNAs revealed that the 3' hairpin of SNORA57 carries a putative pseudouridylation loop that, in principle, could direct synthesis of both Ψ 1045 and Ψ 1046 in the 18S rRNA. Thus, our search identified four novel putative dual pseudouridylation guide loops amenable for experimental testing.

The 3' hairpins of SCARNA1, SCARNA8, and SNORA57 direct synthesis of two adjacent Ψ s in the U2 and 18S RNAs

To examine the *in vivo* pseudouridylation guide competences of SCARNA1, SCARNA8, SCARNA14, and SNORA57, the single-copy genes of these intronic RNAs were deleted with CRISPR–Cas9 genome editing in human HAP1 cells (Supplemental Fig. S3). Disruption of the accumulation of each target H/ACA RNA in the CRISPR-treated KO cells was confirmed by RNase A/T1 mapping (Fig. 6A), and pseudouridylation of the predicted target Us in U2 and 18S RNAs was monitored by primer extension analysis (Fig. 6B). Abrogation of SCARNA1 and SCARNA8 accumulation fully obliterated pseudouridylation of the U2 snRNA at the U89/U91 and U43/U44 residues, respectively. Lacking U2- Ψ 43 formation in the absence of SCARNA8 was somewhat unexpected, because formation of *Xenopus* U2- Ψ 43 had been reported to be a redundant reaction implicating both SCARNA8 and a stand-alone Ψ synthase, Pus1 (Deryusheva and Gall 2017). Disruption of SNORA57 accumulation abolished Ψ 1045 and Ψ 1046 synthesis in the 18S rRNA. Importantly, restoration of SCARNA1, SCARNA8, and SNORA57 accumulations in the corresponding KO cells re-established synthesis of U2- Ψ 89/ Ψ 91, U2- Ψ 43/ Ψ 44, and 18S- Ψ 1045/ Ψ 1046, corroborating the notion that the 3'-terminal pseudouridylation loops of SCARNA1, SCARNA8, and SNORA57 possess dual guide activities. Finally, disruption of SCARNA14 expression demolished formation of Ψ 7 in the human U2 snRNA, but it had no impact on the synthesis of Ψ 6, excluding the proposed dual guide capacity of the 3' hairpin of SCARNA14 and confirming the idea that it directs synthesis of U2- Ψ 7 (Schattner et al. 2006).

Alternative base-pairings with the H/ACA antisense sequences are the major determinants of dual guide activities

We have already demonstrated above that the dual pseudouridylation guide activity of the 5' hairpin of SNORA53 can be confined to the synthesis of one Ψ through structural manipulation of its DS region (Fig. 3). These results seem to support the functional importance of the DS regions of dual pseudouridylation loops. The tested CUU133del and CU132GA mutations of SNORA53 [Fig.

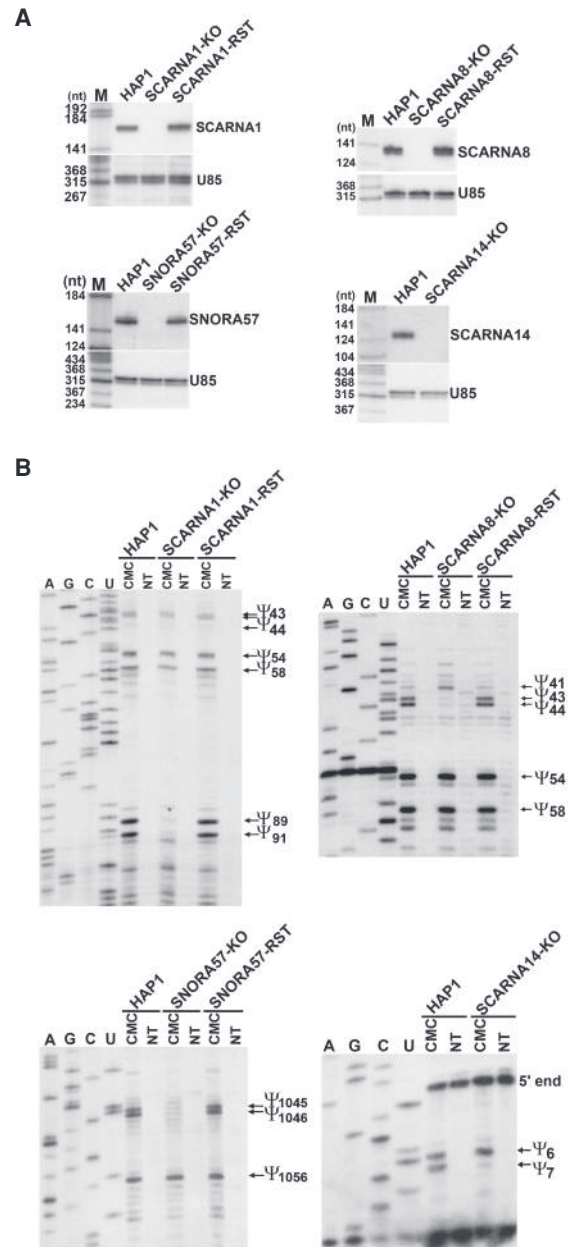


Figure 6. The 3'-terminal pseudouridylation loops of human SCARNA1, SNORA57, and SCARNA8 possess dual guide capacities. (A) Expression of SCARNA1, SNORA57, SCARNA8, and SCARNA14 measured by RNase mappings. After deletion of the SCARNA1, SNORA57, SCARNA8, and SCARNA14 genes in HAP1 cells (Supplemental Fig. S3), each KO cell line was transformed by pIRESpuro/GL plasmids expressing SCARNA1, SCARNA8, or SNORA57 (RST cells). (B) Primer extension mapping of U2 snRNA and 18S rRNA pseudouridylation in the absence and presence of SCARNA1, SCARNA8, SNORA57, and SCARNA14 as indicated above the lanes. For other details, see the legend for Figure 2E.

3D), while promoting formation of one of the two alternative active configurations of SNORA53, prevented formation of the other one. To further understand the

importance of the DS region in the activities of H/ACA dual pseudouridylation guide loops, a series of mutant SNORA57 and SCARNA1 RNAs were expressed in SNORA57-KO or SCARNA1-KO cells, and their pseudouridylation guide capacities were determined (Fig. 7).

Through increasing the stability of the DS regions, the 124-UUC-126 deletion and the U125-to-A conversion in SNORA57 were predicted to promote formation of active configuration A or B and synthesis of 18S-Ψ1045 or 18S-Ψ1046, respectively (Fig. 7A). However, while the 124-UUC-126 deletion was predicted to interfere with formation of active configuration B, the U125A mutation of SNORA57 did not exclude the formal possibility of folding into active configuration A. Upon expression in SNORA57-KO cells (Fig. 7B), the SNORA57-UUC126del RNA proved to be efficient in 18S-Ψ1045 synthesis, but failed to govern 18S-Ψ1046 formation (Fig. 7B, lane 5). In contrast, the SNORA57-U125A RNA retained its dual guide ability to direct synthesis of both Ψ1045 and Ψ1046 (Fig. 7B, lane 7).

Next, we further assessed the contribution of the distal stem (DS) region to the dual pseudouridylation guide activity of the 3' hairpin of SCARNA1. A series of mutant SCARNA1 RNAs carrying nucleotide alterations slightly increasing or reducing the stability of the DS region were expressed in SCARNA1-KO cells (Fig. 7C). Importantly, none of the tested mutations were expected to fully exclude folding into any of the alternative active configurations of the 3' hairpin of SCARNA1. The A144-to-G transition in SCARNA1-A144G was expected to promote DS stability in configuration A directing U2-Ψ89 synthesis and to slightly destabilize the DS region in configuration B governing U2-Ψ91 synthesis. In contrast, the G145-to-U (G145U), the 144-AG-145-to-CU (AG145CU), and the C121-to-G (C121G) base conversions in the DS region of SCARNA1 were predicted to support formation of configuration B and to disfavor folding into configuration A. Insertion of AG dinucleotides between the C122 and A123 residues of SCARNA1 (C122+AG) was predicted to increase the stem stability in both configurations A and B, but leaving six unpaired nucleotides highly destabilizing the proximal part of the DS region in configuration B. Finally, replacement of U147 for a C residue was predicted to disrupt the A117:U146 base-pairing interaction closing the pseudouridylation loop in configuration B. However, when expressed in SCARNA1-KO cells (Fig. 7D, top panel), the mutant SCARNA1 RNAs retained dual guide activity directing efficient synthesis of both Ψ89 and Ψ91 in the U2 snRNA (Fig. 7D, bottom panel), confirming that structural alterations in the distal stem region, if they do not exclude formation of one functional configuration, have no significant impact on the dual pseudouridylation guide capacity of the 3'-terminal guide loop of SCARNA1. These results, together with the observations that minor structural alterations in the upstream antisense target stem (UATS) could completely abolish the dual guide activity of the 5' hairpin of SNORA53 (Fig. 4), suggest that the alternative base-pairing interactions formed between the antisense sequences and the

target RNA are the major determinants of the activity of dual pseudouridylation guide loops.

Discussion

Post-transcriptional pseudouridylation of eukaryotic spliceosomal snRNAs and rRNAs is essential for efficient and correct pre-mRNA splicing and protein synthesis. Site-specific synthesis of ~130 Ψs in human snRNAs and rRNAs is achieved mostly by H/ACA pseudouridylation guide RNPs. In contrast to stand-alone Ψ synthases that recognize local RNA structures and frequently synthesize several closely placed Ψs (Spenkuch et al. 2014), the H/ACA RNPs have been believed to direct synthesis of a single Ψ in their target sequences. However, here we have demonstrated that the pseudouridylation guide loops in the 5'-terminal hairpin of human SNORA53 and in the 3' hairpins of human SNORA57, SCARNA1, and SCARNA8 H/ACA RNAs are able to support synthesis of two consecutive Ψs (ΨΨ or ΨNΨ) through formation of alternative base-pairing interactions with their target RNAs, breaking with the notion that pseudouridylation guide RNPs lack site promiscuity. Apparently, synthesis of two Ψs is not a consequence of defective or inaccurate functions of the dual guide loops of human SNORA53, SNORA57, SCARNA1, and SCARNA8, because their substrate Us are pseudouridylated with practically 100% efficiency (Reddy and Busch 1988; Taoka et al. 2018). Even more telling, all guide RNA nucleotides implicated in inter- and intramolecular base-pairings promoting alternative target RNA interactions are perfectly conserved during evolution (Supplemental Figs. S1, S4). These strong evolutionary constraints clearly demonstrate the functional importance of these nucleotides in correct and efficient dual pseudouridylation reactions.

In principle, synthesis of two Ψs by dual pseudouridylation guide loops could be accomplished through one or two independent target binding events, depending on whether the guide RNP dissociates from or remains associated with its target RNA after the first pseudouridylation reaction. Although we cannot unambiguously distinguish between these two possibilities, we prefer the "one binding–two pseudouridylation" model. Continuous guide RNA–target RNA association could increase the efficiency of the two subsequent pseudouridylation reactions directed by the same guide loop. In this context, it is noteworthy that the proposed alternative guide substrate configurations always share common interacting core domains that could continuously anchor the guide RNP to the target RNA without interfering with the structural rearrangements supporting the two pseudouridylation reactions (Figs. 3D, 5B, shaded helices). In the "two-binding model," after dissociation of the guide RNP following the first pseudouridylation reaction, the dual pseudouridylation loop region of the released guide RNP is expected to undergo structural rearrangements to conform to the second pseudouridylation reaction. Such conformational changes in the guide RNA could

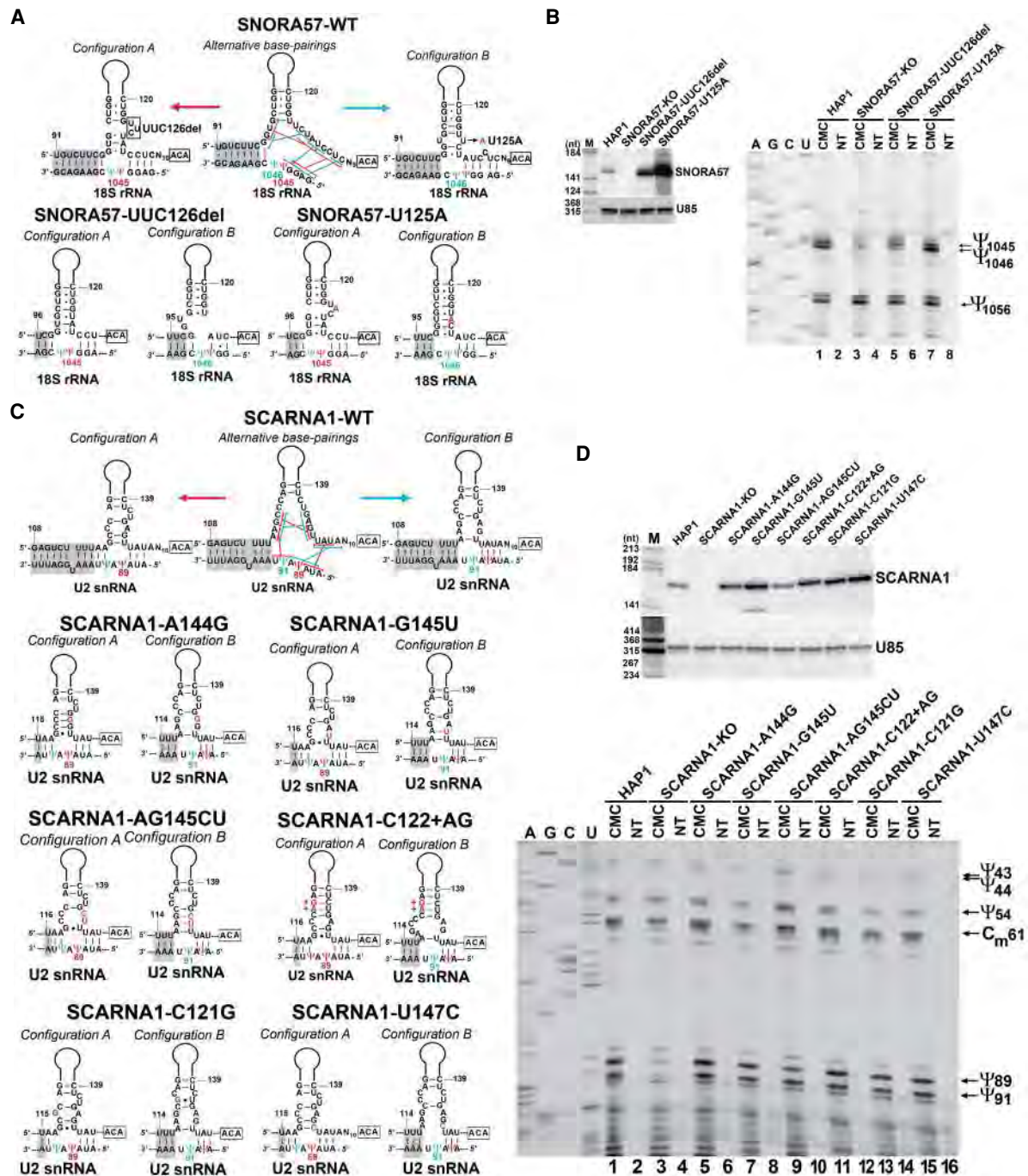


Figure 7. Mutational analyses of SNORA57 and SCARNA1. (A) Predicted base-pairing interactions of wild-type (WT) and mutant SNORA57 RNAs with 28S rRNA in configurations A and B directing synthesis of Ψ1045 and Ψ1046, respectively. (B) Accumulation of mutant SNORA57 RNAs in SNORA57-KO cells and pseudouridylation of endogenous 28S rRNA. (Left panel) Accumulation of SNORA57-UUC126del and SNORA57-U125A RNAs was detected by RNase protection with sequence-specific probes. (Right panel) The pseudouridylation state of 18S rRNA in wild-type HAP1 and SNORA57-KO cells lacking or expressing the indicated mutant SNORA57 RNAs was measured by primer extension. (C) Predicted interactions of wild-type and mutant SCARNA1 RNAs with U2 snRNA directing pseudouridylation of U89 (configuration A) or U91 (configurations B). The nucleotide alterations in the mutant SCARNA1 RNAs are highlighted in red. (D) In vivo pseudouridylation activity of mutant SCARNA1 RNAs. (Top panel) Expression of mutant SCARNA1 RNAs in SCARNA1-KO cells was monitored by RNase protection. The lower band in the SCARNA1-G145U lane likely represents a degradation product of SCARNA1-G145U. (Bottom panel) Primer extension mapping of U2 snRNA pseudouridylation at U89 and U91 in the absence (SCARNA1-KO) or presence of wild-type (HAP1) and mutant SCARNA1 RNAs as indicated above the lanes. For other details, see the legend for Figure 2E.

be governed by the distal stem region through slightly altering and repositioning the single-stranded antisense sequences. However, the dual guide loops of SNORA53, SNORA57, and SCARNA1 are topped by loosely structured and thermodynamically unstable distal stem regions, arguing against a role in shaping their pseudouridylation loops (Figs. 3D, 5B). In line with this conclusion, nucleotide alterations in the distal stem region had no significant impact on the dual pseudouridylation capacity of the 3'-terminal guide loops of SCARNA1 and SNORA57 (Fig. 7). Instead, our data supported the idea that correct remodeling of the target RNA-guide RNA three-way helical junctions is dictated predominantly by formation of alternative antisense target interactions (Fig. 7).

Dissection of the molecular mechanisms driving the appropriate guide RNA-target RNA structural rearrangements supporting the correct and efficient function of dual pseudouridylation guide RNPs remains a challenge for the future. In the secondary structures of H/ACA RNAs defined by minimal free energy calculation, the pseudouridylation loops frequently fold into partially closed structures (see Fig. 3A; Supplemental Fig. S1; data not shown), suggesting that the associated H/ACA proteins contribute to the correct folding of functionally competent pseudouridylation loops. Indeed, the catalytic surface of Cbf5 loosely and nonspecifically associates with the pseudouridylation loop regions of H/ACA RNAs (Li and Ye 2006). In the substrate-bound H/ACA RNP, the conserved arginine-rich thumb region of Cbf5 interacts with the substrate RNA locking the target U at the active site of the RNP (Duan et al. 2009). Gar1 interacts with the catalytic domain of archaeal Cbf5 and controls transitions between the closed and open states of the Cbf5 thumb loop (Duan et al. 2009; Wang et al. 2015). Although it seems that archaeal Gar1 makes no direct RNA contacts, the N- and C-terminal parts of eukaryotic Gar1 contains glycine-arginine-rich RGG domains (Girard et al. 1992). RGG domains are frequent in RNA-binding proteins and they play accessory roles in RNA binding (Thandapani et al. 2013). The RGG domains of yeast Gar1 have been found to facilitate pseudouridylation of the bound substrate RNA and to inhibit its release, raising the possibility that eukaryotic Gar1 may directly interact with target RNAs though its RGG domains (Trucks et al. 2021).

It seems that dual pseudouridylation guide loops are not confined to vertebrate H/ACA RNAs. Genetic depletion of yeast snR82 abolished pseudouridylation of the 27S rRNA at U2350 and also at the nearby U2348 residue (Schattner et al. 2004). While synthesis of Ψ 2350 was predicted to be guided by the 3'-terminal pseudouridylation loop of snR82, snR82-dependent synthesis of Ψ 2348 could not be reasoned, as U2348 was imbedded in the downstream antisense target stem (DATS) of snR82 directing Ψ 2350 formation (Supplemental Fig. S5A, configuration A). However, the 3'-terminal pseudouridylation loop of snR82 can form an alternative interaction with the 27S rRNA to position U2348 for pseudouridylation (configuration B). More recently, the 3' hairpin of the archaeal

Haloferax volcanii sR-h45 H/ACA RNA has been shown to govern pseudouridylation of the 23S rRNA at both Ψ 1940 and Ψ 1942 (Majumder et al. 2020). The investigators proposed that the dual guide activity of the 3' hairpin of sR-h45 was due to its unusual interaction with 23S sequences, placing four unpaired ribosomal nucleotides, 1940-UAUG-1943, including the U1940 and U1942 substrate Us, into the pseudouridylation pocket of the 3'-terminal pseudouridylation loop (Supplemental Fig. S5B, configuration A). In fact, pseudouridylation pockets can accommodate more than two unpaired nucleotides at the base of the distal stem (De Zoysa et al. 2018). The investigators suggested that in this context, Cbf5 bound to the ACA box of sR-h45 was able to convert both the canonically positioned U1940 and the downstream unpaired U1942 into Ψ . Instead, we propose that sR-h45 forms an alternative interaction with 23S rRNA in order to specifically select U1942 for pseudouridylation (configuration B). Importantly, the results of the mutational analysis presented by the investigators are consistent with the existence and functional significance of configuration B supporting Ψ 1942 synthesis. For instance, the presence of U1940 in the 23S rRNA was required for the synthesis of Ψ 1942, supporting the importance of U1940 in formation of DATS in configuration B. In contrast, replacement of the U1942 residue failed to affect U1940 pseudouridylation, as formation of functional configuration A is independent from U1942. Based on these observations, we propose that the 3' hairpins of yeast snR82 and *Haloferax* sR-h45 RNAs carry dual pseudouridylation guide loops directing synthesis of two consecutive Ψ s in the yeast 25S and archaeal 23S rRNAs through formation of alternative base-pairing interactions. Thus, H/ACA guide RNA acrobatics seems to be an evolutionarily widespread strategy for synthesis of two consecutive Ψ s in vertebrate, yeast, and archaeal spliceosomal and ribosomal RNAs.

How many human H/ACA RNAs contain dual pseudouridylation guide loops? Apparently, synthesis of isolated Ψ s dispersed in ribosomal and spliceosomal sequences is directed by specific pseudouridylation guide loops. Although it is unlikely that our search has overlooked many H/ACA RNAs with dual pseudouridylation guide loops (Fig. 5), it is possible that some pseudouridylation loops assigned for synthesis of Ψ s in heavily pseudouridylated sequence environments possess dual guide capacity, as efficient synthesis of some Ψ s might be accomplished by cooperative actions of single and dual pseudouridylation guide loops. Nevertheless, demonstration that four H/ACA RNAs carry dual pseudouridylation loops already shows that synthesis of at least 6% of the human ribosomal and spliceosomal Ψ s is directed by dual guide loops, indicating that formation of closely located Ψ s by dual pseudouridylation loops is a relatively frequent strategy of H/ACA RNPs.

In summary, guide RNA acrobatics supporting accurate and efficient conversion of nearby substrate Us (UU or UNU) into Ψ is a novel, unexpected, relatively frequent, and, most probably, evolutionarily widespread strategy of box H/ACA pseudouridylation guide RNPs.

Materials and methods

General procedures

Unless stated otherwise, all DNA and oligonucleotide manipulations were performed according to standard laboratory procedures. Oligodeoxynucleotides were purchased from Eurofins MWG Operon. Human HAP1 cells were grown in Iscove's modified Dulbecco's medium (IMDM) supplemented with 10% (v/v) fetal bovine serum, 1 mM sodium pyruvate (Gibco), and 100 µg/mL penicillin and streptomycin (Gibco). Cells were transfected by electroporation with Gene Pulser (Bio-Rad) using exponential decay at settings 270 V and 950 µF.

Expression plasmids

Human HAP1 genomic fragments encompassing the *SNORA53*, *SNORA57*, *SCARNA1*, and *SCARNA57* intronic H/ACA RNA genes were PCR-amplified by using sequence-specific primers encompassing the *Cl*AI and *X*hoI restriction sites (Supplemental Table S1A). The resulting DNA fragments were inserted into the *Cl*AI and *X*hoI sites of the pBST vector and the pCMV/globin expression vector (Darzacq et al. 2002). The H/ACA RNA gene together with its host globin gene was PCR-amplified and inserted into the *E*coRI and *N*otI sites of pIRESpuro2 expression plasmid (Addgene). Mutant guide RNA genes were generated by the megaprimer PCR amplification approach using appropriately designed mutagenic oligonucleotide primers and wild-type sno/scaRNA gene-containing recombinant pBST plasmids as templates. For nucleolar expression of mouse ribosomal minigene transcripts, appropriate oligonucleotides encoding the wild-type and mutant fragments of the human 28S rRNA from G3729 to G3759 were annealed and inserted into the *X*baI and *X*hoI sites of the pW(Xb/Xh) expression plasmid (Hadjilova et al. 1994). The resulting pW-28SWT and pW-28S-U3752del plasmids were transfected into 3T3 mouse cells with JetPrime transfection reagent (Polyplus Transfection SA). Two days after transfection, cellular RNA was extracted. The identities of all expression constructs were verified by sequence analysis.

Excision of HAP1 intronic H/ACA RNA genes by CRISPR/Cas9

To generate synthetic DNA encoding appropriately designed sgRNAs, appropriately designed oligodeoxynucleotides (Supplemental Table S1B) were annealed and inserted into the *B*bsI site of pSpCas9(BB)-2A-Puro (PX459) V2.0 by overnight incubation with T4 DNA ligase (NEB) at room temperature. The resulting pX459/sgRNA expression vectors were transfected into HAP1 cells. Twenty-four hours after transfection, cells were transferred into puromycin-containing (1.5 µg/mL) selective medium and grown for 48 h before establishment of monoclonal stable cell lines by limited dilution under nonselective conditions.

Genomic PCR and RT-PCR

Cellular DNA was purified with GenElute mammalian genomic DNA miniprep kit (Sigma), and the targeted genomic fragments were PCR-amplified with appropriately designed oligonucleotide primers (Supplemental Table S1C) and Phusion DNA polymerase (Thermo Scientific). cDNAs were generated with AMV reverse transcriptase (Promega) using total cellular RNAs as templates and sequence-specific oligonucleotide primers. The PCR-amplified genomic and cDNA fragments were analyzed by electrophoresis on 1% agarose gels.

RNA purification and primer extension mapping of pseudouridylation

The guanidine thiocyanate-phenol-chloroform purification procedure was used for RNA extraction from human HAP1 and HAP1-derived cells as described before (Kiss and Jady 2004). Cellular RNAs were reacted with N-cyclohexyl-N'-(2-morpholinoethyl)carbodiimide metho-p-toluenesulfonate (CMCT) and subjected to primer extension analyses (Bakin and Ofengand 1998; Kiss and Jady 2004). The oligonucleotide primers used for pseudouridylation mapping of 18S and 28S rRNAs, U2 snRNA, and mouse ribosomal minigene transcripts are listed in Supplemental Table S1D.

RNase A/T1 protection analysis

RNase A and T1 protection analysis was performed according to Goodall et al. 1990. For mapping of endogenous and ectopically expressed wild-type and mutant *SNORA53*, *SNORA57*, *SCARNA1*, *SCARNA8*, and *SCARNA14* RNAs, complementary RNA probes were synthesized in vitro by T3 RNA polymerase (Promega), using 30 µCi [α -³²P]CTP and *Cl*AI-linearized pBST-*SNORA53*, pBST-*SNORA57*, pBST-*SCARNA1*, pBST-*SCARNA8*, pBST-*SCARNA14*, and pBST-U85 plasmids as templates. Before hybridization with test RNAs, each probe was purified on a 6% denaturing gel. The protected fragments of the probe RNAs were analyzed on 6% sequencing gels.

Competing interest statement

The authors declare no competing interests.

Acknowledgments

Our research was supported by grants from l'Agence National de la Recherche (ANR-18-CE12-0008-01) and Institut National de la Sante et de la Recherche Medicale Plan Cancer to T.K.

Author contributions: B.E.J., A.K., and D.M. carried out experiments. B.E.J. and T.K. conceived and designed experiments and analyzed results. T.K. wrote the manuscript.

References

- Adachi H, De Zoysa MD, Yu YT. 2019. Post-transcriptional pseudouridylation in mRNA as well as in some major types of non-coding RNAs. *Biochim Biophys Acta Gene Regul Mech* **1862**: 230–239. doi:10.1016/j.bbagrm.2018.11.002
- Arnez JG, Steitz TA. 1994. Crystal structure of unmodified tRNA^{Gln} complexed with glutamyl-tRNA synthetase and ATP suggests a possible role for pseudo-uridines in stabilization of RNA structure. *Biochemistry* **33**: 7560–7567. doi:10.1021/bi00190a008
- Baker DL, Youssef OA, Chastkofsky MI, Dy DA, Terns RM, Terns MP. 2005. RNA-guided RNA modification: functional organization of the archaeal H/ACA RNP. *Genes Dev* **19**: 1238–1248. doi:10.1101/gad.1309605
- Bakin AV, Ofengand J. 1998. Mapping of pseudouridine residues in RNA to nucleotide resolution. *Methods Mol Biol* **77**: 297–309.
- Balakin AG, Smith L, Fournier MJ. 1996. The RNA world of the nucleolus: two major families of small RNAs defined by different box elements with related functions. *Cell* **86**: 823–834. doi:10.1016/S0092-8674(00)80156-7

- Boccaletto P, Machnicka MA, Purta E, Piątkowski P, Bagiński B, Wirecki TK, de Crécy-Lagard V, Ross R, Limbach PA, Kotter A, et al. 2018. MODOMICS: a database of RNA modification pathways. 2017 update. *Nucleic Acids Res* **46**: D303–D307. doi:10.1093/nar/gkx1030
- Bohnsack MT, Sloan KE. 2018. Modifications in small nuclear RNAs and their roles in spliceosome assembly and function. *Biol Chem* **399**: 1265–1276. doi:10.1515/hsz-2018-0205
- Carlile TM, Rojas-Duran MF, Zinshteyn B, Shin H, Bartoli KM, Gilbert WV. 2014. Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. *Nature* **515**: 143–146. doi:10.1038/nature13802
- Czekay DP, Kothe U. 2021. H/ACA small ribonucleoproteins: structural and functional comparison between archaea and eukaryotes. *Front Microbiol* **12**: 654370. doi:10.3389/fmicb.2021.654370
- Darzacq X, Jady BE, Verheggen C, Kiss AM, Bertrand E, Kiss T. 2002. Cajal body-specific small nuclear RNAs: a novel class of 2'-O-methylation and pseudouridylation guide RNAs. *EMBO J* **21**: 2746–2756. doi:10.1093/emboj/21.11.2746
- Davis DR. 1995. Stabilization of RNA stacking by pseudouridine. *Nucleic Acids Res* **23**: 5020–5026. doi:10.1093/nar/23.24.5020
- Deryusheva S, Gall JG. 2013. Novel small cajal-body-specific RNAs identified in *Drosophila*: probing guide RNA function. *RNA* **19**: 1802–1814. doi:10.1261/rna.042028.113
- Deryusheva S, Gall JG. 2017. Dual nature of pseudouridylation in U2 snRNA: Pus1p-dependent and Pus1p-independent activities in yeasts and higher eukaryotes. *RNA* **23**: 1060–1067. doi:10.1261/rna.061226.117
- De Zoysa MD, Yu YT. 2017. Posttranscriptional RNA pseudouridylation. *Enzymes* **41**: 151–167. doi:10.1016/bs.enz.2017.02.001
- De Zoysa MD, Wu G, Katz R, Yu YT. 2018. Guide-substrate base-pairing requirement for box H/ACA RNA-guided RNA pseudouridylation. *RNA* **24**: 1106–1117. doi:10.1261/rna.066837.118
- Duan J, Li L, Lu J, Wang W, Ye K. 2009. Structural mechanism of substrate RNA recruitment in H/ACA RNA-guided pseudouridine synthase. *Mol Cell* **34**: 427–439. doi:10.1016/j.molcel.2009.05.005
- Egan ED, Collins K. 2010. Specificity and stoichiometry of subunit interactions in the human telomerase holoenzyme assembled *in vivo*. *Mol Cell Biol* **30**: 2775–2786. doi:10.1128/MCB.00151-10
- Fernández IS, Ng CL, Kelley AC, Wu G, Yu YT, Ramakrishnan V. 2013. Unusual base pairing during the decoding of a stop codon by the ribosome. *Nature* **500**: 107–110. doi:10.1038/nature12302
- Ganot P, Bortolin ML, Kiss T. 1997a. Site-specific pseudouridine formation in preribosomal RNA is guided by small nucleolar RNAs. *Cell* **89**: 799–809. doi:10.1016/S0092-8674(00)80263-9
- Ganot P, Caizergues-Ferrer M, Kiss T. 1997b. The family of box ACA small nucleolar RNAs is defined by an evolutionarily conserved secondary structure and ubiquitous sequence elements essential for RNA accumulation. *Genes Dev* **11**: 941–956. doi:10.1101/gad.11.7.941
- Ge J, Yu YT. 2013. RNA pseudouridylation: new insights into an old modification. *Trends Biochem Sci* **38**: 210–218. doi:10.1016/j.tibs.2013.01.002
- Girard JP, Lehtonen H, Caizergues-Ferrer M, Amalric F, Tollervey D, Lapeyre B. 1992. GAR1 is an essential small nucleolar RNP protein required for pre-rRNA processing in yeast. *Embo j* **11**: 673–682. doi:10.1002/j.1460-2075.1992.tb05099.x
- Goodall GJ, Wiebauer K, Filipowicz W. 1990. Analysis of pre-mRNA processing in transfected plant protoplasts. *Methods Enzymol* **181**: 148–161. doi:10.1016/0076-6879(90)81117-d
- Hadjiolova KV, Normann A, Cavaille J, Soupene E, Mazan S, Hadjiolov AA, Bachellerie JP. 1994. Processing of truncated mouse or human rRNA transcribed from ribosomal minigenes transfected into mouse cells. *Mol Cell Biol* **14**: 4044–4056.
- Hoernes TP, Clementi N, Faserl K, Glasner H, Breuker K, Lindner H, Hüttenhofer A, Erlacher MD. 2016. Nucleotide modifications within bacterial messenger RNAs regulate their translation and are able to rewire the genetic code. *Nucleic Acids Res* **44**: 852–862. doi:10.1093/nar/gkv1182
- Jorjani H, Kehr S, Jedlinski DJ, Gumienny R, Hertel J, Stadler PF, Zavolan M, Gruber AR. 2016. An updated human snoRNAome. *Nucleic Acids Res* **44**: 5068–5082. doi:10.1093/nar/gkw386
- Karijovich J, Yu YT. 2011. Converting nonsense codons into sense codons by targeted pseudouridylation. *Nature* **474**: 395–398. doi:10.1038/nature10165
- Karikó K, Muramatsu H, Welsh FA, Ludwig J, Kato H, Akira S, Weissman D. 2008. Incorporation of pseudouridine into mRNA yields superior nonimmunogenic vector with increased translational capacity and biological stability. *Mol Ther* **16**: 1833–1840. doi:10.1038/mt.2008.200
- Kiss T, Jady BE. 2004. Functional characterization of 2'-O-methylation and pseudouridylation guide RNAs. *Methods Mol Biol* **265**: 393–408.
- Kiss T, Bortolin ML, Filipowicz W. 1996. Characterization of the intron-encoded U19 RNA, a new mammalian small nucleolar RNA that is not associated with fibrillarin. *Mol Cell Biol* **16**: 1391–1400. doi:10.1128/MCB.16.4.1391
- Kiss AM, Jady BE, Bertrand E, Kiss T. 2004. Human box H/ACA pseudouridylation guide RNA machinery. *Mol Cell Biol* **24**: 5797–5807. doi:10.1128/MCB.24.13.5797-5807.2004
- Kiss T, Fayet-Lebaron E, Jady BE. 2010. Box H/ACA small ribonucleoproteins. *Mol Cell* **37**: 597–606. doi:10.1016/j.molcel.2010.01.032
- Lestrade L, Weber MJ. 2006. snoRNA-LBME-db, a comprehensive database of human H/ACA and C/D box snoRNAs. *Nucleic Acids Res* **34**: D158–D162. doi:10.1093/nar/gkj002
- Li L, Ye K. 2006. Crystal structure of an H/ACA box ribonucleoprotein particle. *Nature* **443**: 302–307. doi:10.1038/nature05151
- Li S, Duan J, Li D, Yang B, Dong M, Ye K. 2011. Reconstitution and structural analysis of the yeast box H/ACA RNA-guided pseudouridine synthase. *Genes Dev* **25**: 2409–2421. doi:10.1101/gad.175299.111
- Li X, Zhu P, Ma S, Song J, Bai J, Sun F, Yi C. 2015. Chemical pull-down reveals dynamic pseudouridylation of the mammalian transcriptome. *Nat Chem Biol* **11**: 592–597. doi:10.1038/nchembio.1836
- Liang B, Xue S, Terns RM, Terns MP, Li H. 2007. Substrate RNA positioning in the archaeal H/ACA ribonucleoprotein complex. *Nat Struct Mol Biol* **14**: 1189–1195. doi:10.1038/nsmb1336
- Liang B, Zhou J, Kahen E, Terns RM, Terns MP, Li H. 2009. Structure of a functional ribonucleoprotein pseudouridine synthase bound to a substrate RNA. *Nat Struct Mol Biol* **16**: 740–746. doi:10.1038/nsmb.1624
- Lui L, Lowe T. 2013. Small nucleolar RNAs and RNA-guided post-transcriptional modification. *Essays Biochem* **54**: 53–77. doi:10.1042/bse0540053
- Majumder M, Mukhopadhyay S, Kharel P, Gupta R. 2020. The presence of the ACA box in archaeal H/ACA guide RNAs

- promotes atypical pseudouridylation. *RNA* **26**: 396–418. doi:10.1261/rna.073734.119
- Newby MJ, Greenbaum NL. 2002a. Investigation of overhauser effects between pseudouridine and water protons in RNA helices. *Proc Natl Acad Sci* **99**: 12697–12702. doi:10.1073/pnas.202477199
- Newby MJ, Greenbaum NL. 2002b. Sculpting of the spliceosomal branch site recognition motif by a conserved pseudouridine. *Nat Struct Biol* **9**: 958–965. doi:10.1038/nsb873
- Ni J, Tien AL, Fournier MJ. 1997. Small nucleolar RNAs direct site-specific synthesis of pseudouridine in ribosomal RNA. *Cell* **89**: 565–573. doi:10.1016/S0092-8674(00)80238-X
- Reddy R, Busch H. 1988. Small nuclear RNAs: RNA sequences, structure and modifications. In *Structure and function of major and minor small nuclear ribonucleoprotein* (ed. Birnstiel ML), pp. 1–37. Springer-Verlag Press, Berlin, Germany.
- Schattner P, Decatur WA, Davis CA, Ares M Jr., Fournier MJ, Lowe TM. 2004. Genome-wide searching for pseudouridylation guide snoRNAs: analysis of the *Saccharomyces cerevisiae* genome. *Nucleic Acids Res* **32**: 4281–4296. doi:10.1093/nar/gkh768
- Schattner P, Barberan-Soler S, Lowe TM. 2006. A computational screen for mammalian pseudouridylation guide H/ACA RNAs. *RNA* **12**: 15–25. doi:10.1261/rna.2210406
- Schwartz S, Bernstein DA, Mumbach MR, Jovanovic M, Herbst RH, León-Ricardo BX, Engreitz JM, Guttman M, Satija R, Lander ES, et al. 2014. Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. *Cell* **159**: 148–162. doi:10.1016/j.cell.2014.08.028
- Sloan KE, Warda AS, Sharma S, Entian KD, Lafontaine DLJ, Bohnsack MT. 2017. Tuning the ribosome: the influence of rRNA modification on eukaryotic ribosome biogenesis and function. *RNA Biol* **14**: 1138–1152. doi:10.1080/15476286.2016.1259781
- Spenkuch F, Motorin Y, Helm M. 2014. Pseudouridine: still mysterious, but never a fake (uridine)!. *RNA Biol* **11**: 1540–1554. doi:10.4161/15476286.2014.992278
- Taoka M, Nobe Y, Yamaki Y, Sato K, Ishikawa H, Izumikawa K, Yamauchi Y, Hirota K, Nakayama H, Takahashi N, et al. 2018. Landscape of the complete RNA chemical modifications in the human 80S ribosome. *Nucleic Acids Res* **46**: 9289–9298. doi:10.1093/nar/gky811
- Thandapani P, O'Connor TR, Bailey TL, Richard S. 2013. Defining the RCG/RG motif. *Mol Cell* **50**: 613–623. doi:10.1016/j.molcel.2013.05.021
- Trucks S, Hanspach G, Hengesbach M. 2021. Eukaryote specific RNA and protein features facilitate assembly and catalysis of H/ACA snoRNPs. *Nucleic Acids Res* **49**: 4629–4642. doi:10.1093/nar/gkab177
- Vitali P, Kiss T. 2019. Cooperative 2'-O-methylation of the wobble cytidine of human elongator tRNA^{Met}(CAT) by a nucleolar and a Cajal body-specific box C/D RNP. *Genes Dev* **33**: 741–746. doi:10.1101/gad.326363.119
- Wang C, Meier UT. 2004. Architecture and assembly of mammalian H/ACA small nucleolar and telomerase ribonucleoproteins. *EMBO J* **23**: 1857–1867. doi:10.1038/sj.emboj.7600181
- Wang P, Yang L, Gao YQ, Zhao XS. 2015. Accurate placement of substrate RNA by Gar1 in H/ACA RNA-guided pseudouridylation. *Nucleic Acids Res* **43**: 7207–7216. doi:10.1093/nar/gkv757
- Watkins NJ, Bohnsack MT. 2012. The box C/D and H/ACA snoRNPs: key players in the modification, processing and the dynamic folding of ribosomal RNA. *Wiley Interdiscip Rev RNA* **3**: 397–414. doi:10.1002/wrna.117
- Yu YT, Meier UT. 2014. RNA-guided isomerization of uridine to pseudouridine–pseudouridylation. *RNA Biol* **11**: 1483–1494. doi:10.4161/15476286.2014.972855
- Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31**: 3406–3415. doi:10.1093/nar/gkg595

DISCUSSION

La méthode PIKA : Le bilan

Le développement récent de méthodes utilisant le crosslink au psoralène, couplées au séquençage NGS, a permis d'étudier les différentes interactions ARN-ARN au niveau du transcriptome. Concernant l'étude des snoARN, il est légitime de se demander si ces nouveaux outils possèdent une sensibilité assez élevée permettant d'identifier les ARN cibles des snoARN orphelins. Plusieurs contraintes entrent en jeu, notamment le fait que (i) les snoARN orphelins ont généralement une faible abondance au sein des cellules, et il y a de forte chance qu'ils ciblent également des ARN peu abondants (ii) L'interaction entre un snoARN et sa cible est éphémère dans le temps (iii) Le crosslink au psoralène peut créer des pontages entre des ARN très proches sans réelles interactions physiques, pouvant générer des faux-positifs. Le point crucial pour le fonctionnement de ces méthodes sur les snoARN est donc d'arriver à faire ressortir une interaction rare dans la cellule entre deux ARN peu abondants parmi des interactions non spécifique. Dans un premier temps, les méthodes LIGR-Seq, SPLASH, PARIS, COMRADES et MARIO ont pu mettre en avant des interactions déjà connues entre des snoARN ayant des abondances relativement élevées et des ARNr ou snARN (Aw *et al*, 2016; Sharma *et al*, 2016; Lu *et al*, 2016; Ziv *et al*, 2018; Nguyen *et al*, 2016). De plus, le LIGR-seq a proposé l'interaction physique entre l'ARN SNORD 83B et les ARNm NOP14, RPS5 et SRSF3 (Sharma *et al*, 2016). Même si les validations expérimentales de ces interactions physiques et/ou fonctionnelles n'ont pas clairement été démontrées, ces différentes méthodes semblent capables de pouvoir identifier des interactions physiques entre les snoARN orphelins et leurs cibles. Cependant, lors de l'analyse des données obtenues par ces cinq méthodes, compilées dans la base de données RISE, les contraintes énumérées plus tôt semblent n'avoir pas été toutes contournées avec succès (Gong *et al*, 2018). En effet, aucune données n'ont été générées pour SNORA55 via ces différentes méthodes. De plus, lors de l'analyse de SCARNA10, un snoARN relativement abondant, l'interaction avec le snARN U5 n'est pas retrouvée. En revanche, la méthode PARIS trouve une interaction SCARNA10 – ARNr 5,8S, et le LIGR-Seq une interaction SCARN10 – SNORA48. Ces interactions, si elles existent, ne sont pas connues comme étant des interactions fonctionnelles associées à des fonctions biologiques. Au sein de la cellule, il est hautement probable que des interactions physiques entre deux molécules d'ARN aient lieu, très éphémères, sans qu'une fonction biologique en découle. Des pontages au psoralène lors de l'étape de crosslink peuvent alors se former entre ces deux molécules, générant des interactions que nous qualifions de non-spécifiques.

Dans l'optique d'améliorer la sensibilité des méthodes précédentes, la méthode PIKA a été mise au point. L'enjeu du développement de cette méthode est de pouvoir déterminer des interactions transcriptomiques associées à une fonction biologiques, parmi d'autres interactions physiques non-spécifiques.

Pour ce faire, l'idée était de se concentrer sur une seule interaction, comprenant un snoARN orphelin. La purification d'un snoARN spécifique a été réalisée par les MagIC Beads. Ces billes, utilisant des oligonucléotides d'ADN complémentaires, se sont révélées avoir une efficacité élevée et une forte spécificité. Également, le fait que des oligonucléotides ciblant des régions différentes soient utilisés sur les billes, cela permet de co-purifier un snoARN avec sa cible sans que cette interaction ne gêne l'accessibilité au snoARN. Lors de l'utilisation de ces billes, après radiomarquage au pCp, une tâche diffuse apparaît généralement aux alentours de 400-500 nucléotides. Une digestion par RNase III ayant montrée que cette tâche n'est pas due à la présence d'acides ribonucléiques, l'hypothèse la plus probable est qu'elle soit générée via l'utilisation de glycogène lors de la précipitation des ARN. De plus, lors de l'utilisation de la méthode PIKA avec l'ARN 7SK, 97-98% des reads obtenus après séquençage correspondent à 7SK dans les échantillons, démontrant la forte spécificité de ces billes.

A l'instar des autres méthodes utilisant le psoralène, la méthode PIKA est composée d'une étape de digestion et de ligation. L'étape de digestion utilise la RNase III, coupant spécifiquement les ARN double brins. Mes travaux de thèse ont montré que cette étape fonctionne, notamment lors de l'analyse de l'expérience au cours de laquelle l'ARN 7SK a été purifié. En effet, de par l'analyse des séquences des extrémités 5' des reads, il y a un enrichissement des sites de coupures au niveau de structures double brin de l'ARN 7SK. La RNase III semble couper de manière à générer des fragments de longueurs équivalent aux miRNA, d'une vingtaine de nucléotides (Ji, 2008). La génération de tel fragments par la RNase III pourrait expliquer la longueur des reads chimériques obtenus lors de l'expérience 7SK, qui avoisinent les 40-50 nucléotides, l'étape de ligation liant deux fragments.

En revanche, les différentes purifications réalisées avec la méthode PIKA soulignent différentes possibilités d'optimisation de l'étape de digestion. En effet, lors de l'expérience de recherche des interactions avec les snoARN C/D, nous avons trouvé un nombre comparable de reads chimériques avec ou sans crosslink. Ces fragments d'ARN n'ont pas été crosslinkés *in vitro*, mais ligés par l'étape de ligation de proximité. Ces informations nous montrent que les conditions de cette étape de ligation n'étaient pas optimales. Afin d'éviter de générer de faux ARN chimériques, cette étape doit être effectuée dans un volume beaucoup plus conséquent en

fonction de la quantité d'ARN purifiée. Afin d'estimer correctement le volume, nous pouvons nous appuyer sur les expériences de purification de l'ARN 7SK, où aucun reads chimériques entre 7SK et un autre ARN du transcriptome n'ont été observés. Dans ce cas-là, le ratio quantité d'ARN/volume de réaction était beaucoup plus important et était probablement suffisant pour réduire considérablement la ligation aléatoire des fragments d'ARN n'étant pas physiquement liés.

Un second problème à contourner lors de cette étape de ligation vient de la nature des clivages engendrés par la RNase III. Cette enzyme coupe dans les régions double brins des ARN, ne possédant pas d'activité hélicase, les fragments générés peuvent rester liés par appariement de base. Lors de l'analyse des nombres de reads chimériques de 7SK ainsi que leurs positions, nous avons émis l'hypothèse que ces fragments, qui n'ont pas été séparés, ont aisément été reliés par la T4 RNA Ligase I. Pour éviter cela, une étape de dénaturation des ARN pourrait être rajoutée entre l'étape de digestion et de ligation, qui permettrait de séparer les fragments qui n'ayant pas été ligués par l'étape de crosslink. Si ces ajustements du protocole de ligation diminuent de manière importante le nombre de reads chimériques sans crosslink par rapport de nombre de reads chimérique après crosslink, nous serons enfin capables d'estimer l'efficacité du crosslink au psoralène. Cette estimation nous permettrait par la suite d'avoir une meilleure estimation du nombre de molécules nécessaires pour pouvoir détecter un événement de crosslink et ainsi mieux estimer le nombre de cellules nécessaires pour les identifier plus facilement.

Une fois les étapes de digestion et de ligation optimisées, l'application de la méthode PIKA permettra de co-purifier un snoARN avec sa ou ses cibles. Cependant, aux vues des résultats obtenus par les précédentes méthodes au psoralène, il semble complexe de dissocier les interactions retrouvées associées à une fonction biologique de celles qui ne semblent pas l'être.

Allant dans ce sens, lors de l'application de la méthode PIKA pour identifier les ARN cibles des snoARN C/D, la quantité d'ARN chimériques comprenant un snoARN était très faible, complexifiant les analyses de ces ARN. Ces données confirment le fait que les interactions entre un snoARN et sa cible sont extrêmement temporaires dans la cellule. Pour pallier à cette contrainte, l'application de la méthode PIKA requiert une quantité de matériel très importante. Via une augmentation du nombre de cellules et de billes utilisées, la méthode PIKA pourrait permettre d'identifier les interactions des ARN guides.

Si malgré les efforts futurs, la méthode PIKA ne peut être optimisée pour l'identification systématique des ARN cibles des snoARN orphelins, la méthode développée durant ma thèse pourra confirmer des interactions putatives entre snoARN orphelins et ARN cibles candidats. En effet, les snoARN interagissant par appariement direct de paires de bases, les méthodes bio-informatiques semblent au final être les mieux placées pour émettre une liste d'ARN candidats, à l'instar d'une étude récente sur l'ARN orphelin SNORD116 (Baldini *et al*, 2021). L'addition d'étapes supplémentaires de RT-qPCR, moins onéreuses que les séquençages NGS permettront alors d'observer la co-purification de l'ARN cible par la purification d'un snoARN orphelin, et inversement, validant ou non l'interaction biologique étudiée. Cette méthode peut être très utile lorsque la modification sur l'ARN cible s'avère être compliquée ou le KO de l'ARN guide impossible.

Interactions de l'ARN 7SK avec le transcriptome

Le rôle du snARN 7SK est de réguler la transcription de la polymérase II par l'inactivation du facteur positif de l'élongation (P-TEFb). A ce jour, il est connu que sa fonction au sein de la machinerie cellulaire ne nécessite aucune interaction avec d'autres ARN cellulaires (Egloff *et al*, 2018). Cependant, l'ARN 7SK, à l'instar de la grande majorité des ARN cellulaires, est porteur de modifications post-transcriptionnelles. En ce sens, une pseudouridine et six méthylations en 2'-O-ribose ont été identifiées sur l'ARN 7SK, et leurs synthèses peuvent nécessiter l'intervention d'un ARN guide (Kishore *et al*, 2013; Zhao *et al*, 2016).

Lors de l'étude de 7SK via l'application de la méthode PIKA, la quasi-totalité des reads issus du séquençage RNA-Seq classique ont été alignés avec l'ARN 7SK. Les interactions entre les snoARN et leurs cibles étant très transitoires, les résultats obtenus n'ont donc pas été surprenants. Cependant, aucun read n'a pu être aligné avec la séquence de SNORA 70, qui est connu pour être impliqué dans la synthèse de la seule pseudouridine identifiée (Ψ 250) et portée par 7SK (Zhao *et al*, 2016). Ce snoARN pourrait alors ne pas être l'ARN guidant cette modification. L'ARN 7SK étant abondant au sein des cellules, une détection des pseudouridines au CMCT, ou l'utilisation de Nanopores dans une lignée KO pour SNORA 70 pourrait répondre à cette interrogation. Concernant les méthylations en 2'-O-ribose, aucun ARN SNORD n'a spécifiquement été enrichi entre les échantillons avec ou sans crosslink. Cependant, le nombre de reads retrouvé pour SNORD3A est, de manière intéressante, supérieur aux autres snoARN identifiés, même si ce fait est potentiellement dû à la forte abondance de SNORD3A dans les cellules. Il y a donc de fortes probabilités que les méthylations 2'-O-ribose portées par 7SK

soient catalysées par des enzymes autonomes. Allant dans ce sens, aucun élément antisens des snoARN C/D connus ne possède de complémentarité avec la région de l'ARN 7SK portant les nucléotides 2'-O-méthylés.

D'un autre côté, la majorité des reads ayant une séquence s'alignant avec le transcriptome, 7SK exclu, obtenus par ce RNA-Seq, s'aligne avec les ARNr. Sachant qu'actuellement aucune observation n'a montré que l'ARN 7SK transite par les nucléoles, il y a peu de chance que la co-purification des ARNr lors de la purification de 7SK soit associée à des interactions ayant une fonction biologique. De plus, lors de l'analyse du Small RNA-Seq, aucune chimère 7SK – ARNr n'a été identifiée. Considérant cela, il y a de fortes chances que l'enrichissement des ARNr dans les échantillons crosslink soient dues au hasard, de par la forte abondance de ces ARNr dans la cellule. Ainsi, l'application de la méthode PIKA sur l'ARN 7SK montre que les fonctions biologiques de cet ARN ne nécessitent probablement aucune interaction avec d'autres ARN cellulaires, en accord avec ce qui est connu aujourd'hui (Egloff *et al*, 2018).

SNORD115, le snoARN orphelin par excellence

Les gènes SNORD115 sont localisés sur le domaine SNURF-SNRPN, dont la perte est associée au syndrome PWS. SNORD115 possède une complémentarité parfaite sur 18 nucléotides avec le cinquième exon de l'ARNm du récepteur à la sérotonine 2C. Le nucléotide prédit pour être méthylé est une adénine, connue pour être désaminée en inosine. L'ajout de cette méthylation inhiberait l'édition de cette adénine, favorisant l'inclusion de l'exon V et ainsi engendrer un récepteur fonctionnel. La perte de cette méthylation tendrait alors à la formation d'un récepteur tronqué, pouvant expliquer les symptômes du syndrome de Prader-Willi.

Malgré tous ces arguments, l'interaction entre l'ARN SNORD115 et l'ARNm 5-HT2C n'a toujours pas été prouvée.

Dans l'optique de prouver cette interaction physique, durant mes travaux de thèse, des essais de co-purification de SNORD115 avec l'ARNm 5-HT2C ont été effectués à partir de tissus cérébraux de souris sauvages ou KO pour SNORD115. Lors de l'analyse des RT-qPCR, l'interaction entre ces deux ARN n'a pas été observée. L'expérience est à réitérer plusieurs fois afin de pouvoir affirmer ce résultat, dans l'idéal où la méthode PIKA est assez sensible pour pouvoir détecter la probablement très faible quantité de SNORD115 co-purifié avec

SNORD115, si l'interaction existe. De plus, un contrôle positif de Crosslink est manquant à l'expérience. En effet, même si l'accomplissement de l'étape de crosslink au psoralène peut être déduite durant l'expérience par une coloration jaunâtre du milieu des échantillons crosslink, une co-purification de SCARNA10-snU5 ou SNORD94-snU6 est beaucoup plus approprié pour montrer l'efficacité du crosslink.

L'ARN SNORD115 faisant 82 nucléotides, les zones pouvant être ciblées par des oligonucléotides antisens se retrouvent être très restreintes. Il n'est pas impossible, que, malgré le cahier des charges, les deux oligonucléotides différents sur les MagIC Beads ciblent une région dans laquelle l'élément antisens pour la purification est présent. Dans ce cas, lors de la purification, seuls les SNORD115 libres, n'étant donc pas en interaction avec leurs cibles, seraient sélectionnés.

Cependant, l'hypothèse allant dans le fait que l'interaction entre SNORD115 et l'ARNm n'existe pas, n'est pas, au contraire, à mettre de côté. En effet, la déplétion de SNORD115 chez la souris n'engendre aucun effet sur **(i)** l'expression de l'ARNm 5-HT2C dans différents tissus cérébraux **(ii)** sur l'épissage alternatif de l'ARNm 5-HT2C **(iii)** de défaut comportemental pouvant être associé au syndrome PWS chez les souris (Hebras *et al*, 2020).

Malgré tous les arguments dans le sens d'une interaction SNORD115 – 5-HT2C, SNORD115 demeure toujours être un snoARN orphelin.

PIKA et la différenciation neuronale, des outils d'études de l'ARN SNORD116

Le gène SNORD116, présent en de nombreuses copies au niveau du domaine génomique SNURF-SNRPN soumis à l'empreinte parentale, localisé sur le locus chromosomique 15q11q13, est associé, à l'instar de SNORD115, au syndrome de Prader-Willi. En effet, SNORD116 fait constamment partie des gènes déplétés lorsque des micro-délétions de la région 15q11q13 sont observés chez les patients atteint du PWS, et des délétions de SNORD116 chez la souris induisent les symptômes de PWS (Polex-Wolf *et al*, 2018; Adhikari *et al*, 2019). Par cela, SNORD116 fait aujourd'hui parti des snoARN orphelins les plus étudiés au sein de la communauté scientifique. Malgré une forte expression dans certains tissus, aucune séquence de complémentarité avec les ARNr et les éléments antisens de SNORD116 n'ont été identifiés. SNORD116 est encore aujourd'hui considéré comme faisant partie de la classe des snoARN orphelins.

Cependant, de récents travaux ont pu établir une liste d'ARN candidats pouvant être la cible du SNORD116. Des programmes bio-informatiques ont pu établir une liste de différents ARN pouvant être la cible de SNORD116. La base de données RAIN prédit principalement des interactions avec d'autres snoARN, dont SNORD115, chez l'humain (Junge *et al*, 2017). La méthode expérimentale MARIO met en avant une interaction avec l'ARN SNORA 44 chez la souris (Nguyen *et al*, 2016). Le programme SnoTARGET prédit une interaction notable avec l'ARNm Ankrd11, un régulateur de la chromatine impliqué dans le développement neuronal (Baldini *et al.*, 2021; Bazeley *et al.*, 2008). Une étude précisément centrée sur SNORD116 a mis en avant une complémentarité des éléments antisens de SNORD116 avec les ARNm Dgkk, Nlgn3 et Rsbn1l, vérifiée par qRT-PCR (Baldini *et al.*, 2021).

Dans l'optique de confirmer ou d'infirmer ces différents ARN candidats, il serait intéressant de prouver l'interaction de SNORD116 avec ces ARN cibles par co-purification *in-vivo*. Dans cette optique, la présence des ARNm pouvant aisément être détecté par qRT-PCR, l'application de la méthode PIKA permettrait d'étudier la co-purification de SNORD116 avec les ARNm Dgkk, Nlgn3 et Rsbn1l. En ce sens, nous avons commandé et mis au point des MagIC Beads purifiant efficacement et spécifiquement SNORD116 au laboratoire (data non montrées). En supplément, la différenciation neuronale de cellules embryonnaires de souris a permis d'augmenter l'expression cellulaire de SNORD116. Etudier SNORD116 dans des cellules neuronales serait un avantage pour pouvoir déterminer ses ARN cibles, la quantité de matériel étant cruciale pour les expérimentations *in vivo* de co-purification. De plus, le Dr. Jérôme Cavaillé et son équipe ont généré une lignée KO pour le gène *SNORD116* dans la lignée cellulaire humaine NT2D1. La lignée NT2D1 sont des cellules pluripotentes de carcinome embryonnaire testiculaire, qu'il est possible de différencier en cellules neuronales via l'ajout d'acide rétinoïque (Goodfellow *et al*, 2011). Une mise en parallèle des données de co-purification de SNORD116 avec ses cibles chez l'humain et la souris pourrait compléter les connaissances sur les interactions de SNORD116 avec le reste du transcriptome.

Une méthode alternative pouvant être proposée est de cartographier les nucléotides méthylés présents sur les ARNm Dgkk, Nlgn3 et Rsbn1l dans des lignées cellulaires sauvages et KO pour le SNORD116. Si une interaction SNORD116 avec ces trois ARNm existe, la différence de profil de méthylations permettrait de démontrer ces interactions, dans le cas où elles soient associées à des fonctions canoniques. Cependant, les ARNm sont des ARN peu abondants dans la cellule, pouvant compromettre la détection des méthylations par nanopores ou par Nm-Seq par exemple (Dai *et al*, 2017; Begik *et al*, 2021).

Une interaction entre SNORA55 et l'ARNr 18S : Oui, mais comment ?

La RNP SNORA55 est composée d'un ARN H/ACA dont les régions nécessaires à ses fonctions ou à sa structure secondaire sont fortement conservées au cours de l'évolution (**Annexe 1**). SNORA55 est classé dans le groupe des snoARN orphelins. Les bases de données snoRNA-LBME-db et RAIN ont prédit que SNORA55 ciblerait l'ARN 18S pour y guider deux pseudouridylation, sur les Uridines aux positions U36 et U681 (Lestrade & Weber, 2006; Junge *et al*, 2017). Les Uridines U36 et U681 seraient reconnues respectivement par les éléments antisens situés en 5' et 3' de l'ARN SNORA55.

Un des axes de mon projet doctoral était d'utiliser la méthode PIKA afin de vérifier les interactions prédites SNORA55 – ARNr18S, mais cela n'a pas pu être accompli dû à des difficultés techniques. En contrepartie, une lignée cellulaire déplétée du gène *SNORA55* a été générée grâce à la technologie CRISPR/Cas9. Cette lignée a permis la réalisation d'une cartographie des pseudouridines de l'ARNr 18S au CMCT. Nous avons pu montrer que la pseudouridine à la position Ψ36 n'était pas présente lors de la cartographie réalisée avec les cellules KO SNORA55, contrairement au contrôle effectué avec des cellules sauvages. La tige-boucle située dans la région 5' de l'ARN SNORA55 n'est désormais plus orpheline, guidant la pseudouridylation de l'Uridine U36 sur l'ARNr 18S. Dans un futur proche, nous planifions, par la même procédure, d'étudier si l'interaction SNORA55 – ARNr18S permet la formation de la pseudouridine située à la position Ψ681 de l'ARNr18S. Cependant, cette prédiction est moins robuste que la précédente. En effet, l'alignement est imparfait et de nombreux mésappariements sont présents. A contrario, l'interaction SNORA55 – ARNr18S pour la formation de la pseudouridine Ψ36 résulte d'un appariement parfait sur 7 et 4 nucléotides, bordant respectivement les deux nucléotides non-appariés présentés au site catalytique de la Dyskérine. Cependant il ne serait pas étonnant que la tige-boucle en 3' de SNORA55 guide une modification sur l'ARNr18S. En effet, certains snoARN sont connus pour guider plusieurs modifications sur un même ARN cellulaire, à l'instar de SCARNA10 (Kiss & Jádý, 2004).

Une étude a fait le lien entre le taux d'expression de SNORA55 et le développement du cancer de la prostate (Crea *et al*, 2016). Il serait intéressant d'étudier l'impact de la perte de la ou des modifications chimiques guidées sur l'ARNr 18S par SNORA55 dans un contexte pathologique. En effet, des mutations de la Dyskérine, la protéine catalytique des snoRNP H/ACA, causent la maladie de dyskératose congénitale liée au chromosome X (Mitchell *et al*,

1999c). Les patients touchés par cette pathologie ont par ailleurs montrés un risque élevé de développement de cancer (Penzo *et al*, 2013).

L'identification de nouveaux snoARN H/ACA : la fin d'une quête ?

Depuis la découverte des pseudouridines en 1951 grâce aux travaux de Cohn (Cohn, 1951), ce « cinquième nucléoside » a été le sujet de nombreuses études. De par leur forte abondance sur les ARNr et les snARN, les rôles associés aux pseudouridines ont été fortement caractérisés dans les contextes de traduction et d'épissage (Karijovich *et al*, 2010). En parallèle, les mécanismes impliqués dans la formation des pseudouridines a des endroits précis du transcriptome sont longtemps restés une énigme. Les isomérisations des uridines en pseudouridines ont premièrement été associées à des pseudouridines synthèses spécifique et autonomes. Il aura fallu attendre 1997 pour qu'un second mécanisme impliqué dans la formation des pseudouridines soit découvert chez les eucaryotes : les snoARN H/ACA (Ganot *et al*, 1997). Depuis, le nombre de snoARN H/ACA identifiés dans les transcriptomes des eucaryotes n'a cessé d'augmenter. L'identification des snoARN H/ACA a notamment été facilité par la présence de boîtes caractéristiques et conservées évolutivement. Des programmes bio-informatiques, tel que ACA Seeker, ont ainsi émergé pour automatiser la recherche de snoARN H/ACA dans les transcriptomes (Yang *et al*, 2006).

Dans cette optique d'enrichir le patrimoine connu des snoARN H/ACA, un axe de mes travaux de doctorat a été impliqué dans la recherche de snoARN H/ACA chez l'Homme. Pour proposer une alternative à la recherche informatique de snoARN H/ACA, une protéine cœur des snoRNP H/ACA, NHP2, a été immunoprécipitée *in vivo*. Après transcriptase inverse et séquençage haut débit, une liste de 600 nouveaux snoARN H/ACA putatifs a été établie. Après analyse des séquences consensus et des structures secondaires prédites de ces ARN, seulement une dizaine de potentiels snoARN, encore non-identifiés par la communauté scientifique, ont été retenus. Les essais de détection *in vivo* de ces snoARN par Northern Blot n'ont pas pu démontrer une accumulation facilement détectable de ces snoARN dans les cellules humaines. L'utilisation de la méthode de digestion au RNase A/T1, plus sensible, couplée à une surexpression des snoARN putatifs dans les cellules ont permis l'identification de trois nouveaux snoARN H/ACA, nommés 29, 120 et 186. En revanche, les snoARN endogènes n'ont pas pu être détectés par digestion RNase A et T1.

Ces résultats démontrent premièrement que la grande majorité des snoARN H/ACA ont déjà été préalablement identifiés par les précédents travaux effectués par la communauté

scientifique, notamment les snoARN H/ACA ayant une abondance relativement élevée dans les cellules. La faible abondance des snoARN H/ACA putatifs 29, 120 et 186 laisse penser qu'ils pourraient cibler des ARN d'abondances beaucoup plus faibles que les ARNr et snARN, à l'instar des snoARN H/ACA actuellement classifiés dans le groupe des snoARN orphelins. Allant dans ce sens, nous savons depuis une décennie que les ARNm sont porteurs de pseudouridylations (Carlile *et al*, 2014). De plus, depuis quelques années, le concept d'hétérogénéité des ribosomes ainsi que leur spécialisation devient de plus en plus évident. Cette spécialisation serait induite par des différences au niveau du contenu protéique des ribosomes ainsi qu'au niveau des modifications chimiques des ARNr (Gay *et al*, 2022). Il n'est donc pas à exclure qu'une faible abondance de certains snoARN H/ACA puisse corrélérer avec la présence de pseudouridines spécifiquement présentes sur une minorité des ARNr.

Réitérer cette expérience d'immunoprécipitation, notamment avec d'autres protéines cœurs composant les snoRNP H/ACA, pourrait être un moyen de compléter le répertoire des snoARN H/ACA connus. De plus, durant mes travaux de thèses, nous nous sommes uniquement concentrés sur les séquences obtenues ayant une structure prédite en "hairpin-hinge-hairpin-tail". Il n'est pas impossible que nous soyons passé à côté de l'identification de snoARN H/ACA possédant des structures secondaires non-canoniques, ayant par exemple une ou trois tiges-boucle, comme étant le cas chez les trypanosomes et les archées (Russell *et al*, 2004; Uliel *et al*, 2004; Bernick *et al*, 2012). Également, nous avons retiré les séquences d'ARN H/ACA putatifs correspondants à des ARN Alu lors de ces travaux. Ces ARN sont issus de régions répétées du génome qui sont très faiblement exprimées. Certains se situent au niveau de régions introniques et peuvent, si transcrits par la polymérase II, s'assembler avec les protéines cœurs des scaRNA H/ACA (Jády *et al*, 2012). Cependant, malgré leur nombre et la présence de poches de pseudouridylations, aucunes fonctions n'ont été associées à ces ARN. Dans l'idée d'identifier des snoARN fonctionnels, nous les avons écartés de nos analyses. Outre cela, il semblerait que, d'après les résultats obtenus, nous arrivons au terme de cette quête d'identification de nouveaux snoARN H/ACA existants dans les cellules humaines.

ARN cibles et fonctions associées aux snoARN H/ACA 29, 120 et 186

L'identification des snoARN H/ACA putatifs 29, 120 et 186 a été permise par l'insertion de leurs gènes respectifs en aval d'un promoteur fort. Seul l'ARN 186 endogène a pu être détecté après enrichissement par une immunoprécipitation d'une des protéines cœurs des complexes H/ACA, la Dyskérine. Dans un premier temps, il serait nécessaire de vérifier

l'accumulation des snoARN 26 et 120 endogène au sein des cellules humaines par les méthodes qui nous ont permis de détecter l'ARN 186. Les immunoprécipitations de la Dyskérine tagguée FLAG pourraient être par ailleurs être réalisées dans des cellules TK6, la même lignée ayant servi à la détection des séquences des ARN 29, 120 et 186.

Les snoARN 120 et 186 se retrouvent être conservés chez les primates, tandis que le snoARN 29 est facilement détectable chez différents vertébrés (**Annexe 2**). De par la conservation de ces snoARN chez d'autres espèces, il est imaginable que ces snoARN soient associés à des fonctions biologiques. En ce sens, dans un second temps, il serait intéressant de déterminer les ARN ciblés par les snoARN H/ACA 29, 120 et 186, permettant de leur associer un rôle au sein de la machinerie cellulaire. Lors de l'appariement physique d'un snoARN avec son ARN cible, l'uridine cible de l'isomérisation se situe dans la majorité des cas à une distance de 14 nucléotides des boîtes H ou ACA (Kiss *et al.*, 2010). Également, la poche de pseudouridylation des snoARN H/ACA est formée par deux séquences de 9 à 13 nucléotides de part et d'autre des brins opposés la formant. Une partie de ces nucléotides forment l'élément antisens. De par ces généralités, les éléments antisens des snoARN H/ACA 29, 120 et 186 peuvent être identifiés et servir d'informations d'entrées lors d'utilisation de programmes tel que snoGPS et RNAsnoop. En effet, ces deux logiciels de prédictions informatiques ont déjà prouvés leur utilité dans la détection des cibles des snoARN H/ACA, notamment chez l'humain (Schattner *et al.*, 2004; Tafer *et al.*, 2010). Dans le cas où des listes d'ARN candidats sont établies par ces logiciels, les interactions pourront être confirmées expérimentalement par des traitements au CMCT couplés à des extensions d'amorces avec l'aide de lignées KO pour les gènes correspondants aux snoARN H/ACA 29, 120 et 186. D'un autre côté, l'utilisation de méthodes *in vivo* de co-purification des snoARN avec leurs cibles, que ce soit par pontages au psoralène ou par protéines appâts, risquent de ne pas avoir une sensibilité assez élevée aux vues de la faible abondance de ces snoARN. En effet, ces différentes méthodes n'ont pas pu établir une liste d'ARN cibles candidats pour l'ARN SNORA55, qui possède une expression relativement plus élevée que les snoARN H/ACA 29, 120 et 186 (Gong *et al.*, 2018). L'enrichissement spécifique d'un de ces snoARN par la méthode PIKA serait long et fastidieux, et la quantité de la cible co-purifiée ne dépasserait probablement pas les ARN composants le bruit de fond généré par le crosslink au psoralène.

Si aucune prédiction informatique ou méthode expérimentale ne pourra établir une liste des ARN cibles putatifs des snoARN H/ACA 29, 120 et 186, ils rejoindront le groupe des

snoARN orphelins, dans l'attente qu'une méthode assez résolutive soit développée avec l'avancement des technologies.

Le snoARN H/ACA 186, un snoARN pas comme les autres

La quasi-totalité des snoARN H/ACA sont issus de la maturation introniques des transcrits de la polymérase II. Le seul snoARN H/ACA humain synthétisé par une voie alternative est celui composant l'ARN de la Télomérase, transcrit par son propre promoteur (Mitchell *et al*, 1999). Le gène de l'ARN snoARN H/ACA 186 possède cette particularité de chevaucher à la fois une région intronique et une région exonique du gène *MYO19*. C'est à ce jour le seul snoARN, classes H/ACA et C/D confondues, possédant cette spécificité.

En premier lieu, une fonction canonique pourrait être associée au snoARN H/ACA 186. Aux vues de sa faible abondance, les hypothèses de guide de pseudouridylation sur des ARNm ou des ARNr spécialisées se posent. Cependant, ces hypothèses restent à prouver. Même si de nombreuses études mettent en avant de nombreux indices d'une modification d'un ARNm par un snoARN, une interaction claire n'a toujours pas été démontrée expérimentalement (Baldini *et al.*, 2021; Hebras *et al.*, 2020; Sharma *et al.*, 2016).

Dans un second temps, l'hypothèse d'une fonction non-canonique associée à l'ARN H/ACA 186 se pose, d'autant plus à la vue de sa position génomique. En effet, l'assemblage des protéines cœurs est effectué en parallèle de la transcription du gène hôte du snoARN par la polymérase II. L'assemblage protéique avec le snoARN pourrait alors bloquer la reconnaissance du site accepteur lors de l'épissage du gène hôte *MYO19*. Ce phénomène pourrait tendre sur le fait que (i) la maturation du snoARN H/ACA 186 et la maturation du pré-ARNm *MYO19* soient deux événements mutuellement exclusifs (ii) la maturation du snoARN186 induise l'épissage alternatif du pré-ARNm *MYO19*, engendrant l'exclusion de l'exon 23, contenant une région du snoARN, de l'ARNm *MYO19* mature. Allant dans ce sens, il a été démontré que la présence de snoARN dans des régions introniques pouvaient affecter l'épissage du pré-ARNm hôte (Hirose & Steitz, 2001). Également, des travaux plus récents ont décrits que l'ARN SNORD86 régissait en *cis* l'épissage du pré-ARNm de son gène hôte (Lykke-Andersen *et al.*, 2018).

Il serait intéressant dans l'avenir d'étudier ces potentiels mécanismes. Pour ce faire, un système artificiel pourrait être élaboré, comprenant la séquence de du gène *MYO19* contenant

ou non de la séquence intronique du snoARN 186 sous promoteur fort. La quantité d'ARNm mature de MYO19 pourra être analysée lors de la maturation parallèle ou non du snoARN.

SnoARN H/ACA 186 est un indice supplémentaire associant les snoRNP à des fonctions cellulaires non-canoniques, diversifiant le rôle des snoARN au sein de la machinerie cellulaire.

MATERIELS ET

METHODES

Procédures générales

La souche bactérienne *E.coli* DH5 α a été utilisée pour amplifier les plasmides d'intérêts. La lignée cellulaire murine P19 a été cultivée dans du milieu DMEM (Dulbecco's Modified Eagle Medium) complémenté avec 10% (v/v) de sérum de veau fœtal. La lignée cellulaire humaine HeLa a été cultivée dans du milieu DMEM complémenté avec 10% (v/v) de sérum de veau fœtal et 100 μ g/mL de pénicilline et streptomycine de chez Gibco™. Les lignées cellulaires humaine HAPI et HAPI KO ont été cultivées dans du milieu IMDM (Iscove's modified Dulbecco's medium) complémenté avec 10% (v/v) de sérum de veau fœtal, 1mM de sodium Pyruvate (Gibco™) et 100 μ g/mL de pénicilline et streptomycine (Gibco™). La lignée cellulaire humaine TK6 a été cultivée dans du milieu RPMI (Roswell Park Memorial Institute) complémenté avec 10% (v/v) de sérum de veau fœtal et 100 μ g/mL de pénicilline et streptomycine de chez Gibco™. Les différentes lignées cellulaires et murines ont été cultivées dans un incubateur à 5% de CO₂ à 37°C.

Toutes les manipulations concernant les oligonucléotides ont été effectuées en accord avec les procédures de laboratoire standard. Le récapitulatif des différents oligonucléotides utilisés sont référencés dans **l'Annexe 3**. La totalité des oligonucléotides utilisées a été achetée auprès d'Eurofins MWG Operon.

Les transfections cellulaires ont été faites par (i) électroporation via l'appareil Gene Pulser (Bio-Rad) en utilisant la décroissance exponentielle à 270V et 950 μ F (ii) via l'ajout du réactif de transcription jetPRIME® de chez Ozyme en appliquant conformément le protocole d'utilisation fourni.

Les procédures animales ont été effectuées par Virginie MARTY, et approuvée par l'Université de Toulouse et par le comité institutionnel du bien-être animal du CNRS (DAP2016061716367988 et DAP2018011214542827). L'animalerie répond aux normes du CNRS. Les souris ont été hébergées dans une pièce à température contrôlée et un cycle lumière-obscurité de 12 heures. Les souris ont un accès à l'eau et à la nourriture *ad libitum*. Lors de l'euthanasie, les tissus ont été prélevés, pesés et (i) immédiatement congelés dans de l'azote liquide et conservés à -80°C (ii) immédiatement digérés à la trypsine et les cellules isolées. Le modèle de souris KO pour le SNORD115 a été généré par l'équipe de Dr. Jérôme Cavaillé (Hebras *et al.*, 2020).

Extraction des ARN cellulaires

Les extractions des ARN à partir des lignées cellulaires humaines, murines, ou isolées à partir de tissus murins ont été effectuées par purification au thiocyanate de guanidine-phénol-chloroforme, méthode décrite dans la publication (Kiss & Jány, 2004). Après précipitation à l'éthanol et un lavage à l'éthanol 70%, les ARN subissent un traitement par RQ DNase (Promega).

Les extractions d'ARN directement à partir de tissus cérébraux de souris congelés ont été effectués par une homogénéisation dans du Tri-Reagent (Molecular Research Center) à l'aide d'un homogénéisateur Ultra-Turrax. Après ajout de chloroforme à l'homogénat (1v5), la phase aqueuse est précipitée à l'isopropanol. Les ARN ont été lavés à l'éthanol 70% et re-suspendus dans de l'H₂O.

Crosslink au psoralène

2.10⁷ cellules humaines ou murines par expérience ont été re-suspendues dans un tampon 150mM NaCl/10mM Tris-HCl avec de l'AMT psoralène (Sigma) à une concentration de 0,375 µg/µL. Les cellules ont été irradiées sur glace durant 20 minutes aux UV d'une longueur d'onde de 365nm.

Purification des ARN d'intérêts

Les ARN d'intérêts ont été purifiés à partir d'extraits cellulaires crosslinkés ou non avec les MagIC Beads de la société ElementZero Biolabs (<https://elementzero.bio/>). Les procédures de purification ont conformément été appliquées aux recommandations du fournisseur en utilisant les tampons fournis, mis à part pour les purifications de SNORD115, où la température d'hybridation a été faite à 45°C. Pour chaque réaction, des billes portant 20pmol d'oligonucléotides antisens ont été utilisées pour purifier les ARN 7SK, SCARNA10, SNORD115 et SNORA55 à partir de respectivement 12µg, 20µg, 20µg, 40µg d'ARN totaux. L'hybridation des ARN sur les billes a été effectuée durant 30 mins à 1200rpm avec des cycles d'agitations de 3sec ON et 30sec OFF. Les paramètres des cycles d'agitations ont été conservés au cours des différents lavages ultérieurs de 10 mins. Les ARN purifiés ont été élués dans 10µL avec du 10mM de Tris-HCl, pH 7.5.

Concernant les immunoprécipitations, 60 millions de cellules HeLA ont été utilisées, lavées dans du PBS froid puis re-suspendues dans 500µL de tampon, composé de 200mM NaCl,

40mM Tris-HCl 7.5, 0,05% NP40. Les cellules ont été lysées par sonication (Bioruptor Plus Sonicator) 5 fois 30s avec 30s de pause. Les extraits cellulaires ont été incubés durant 1h à 4°C en présence de billes d'agarose protéine A (sigma) sur lesquelles ont été fixés des anticorps anti-Fibrillarine ou anti-NHP2. L'élution a été réalisée grâce à la RNase III.

Fragmentation, Ligation, et réversion du Crosslink des ARN

Les ARN purifiés ont été fragmentés par la RNase III (Ambion), utilisée à une concentration de 0,5 U/μg d'ARN durant 1h à 37°C. La réaction enzymatique a été stoppée par un choc thermique à 94°C durant 1 minute. Les ARN ont été extraits avec du phénol/chloroforme/isoamylalcool 25/24/1, puis précipités à l'éthanol. Les ARN précipités ont été re-suspendus dans 32μL d'H₂O puis a été ajouté : 5μL de Tampon de Ligation (T4 RNA Ligase I), 5μL d'ATP à 10mM, 3 μL de Supersasin et 5μL de T4 RNA Ligase. La ligation a été effectuée O/N à 16°C.

Le crosslink ARN-ARN a été retiré par une irradiation aux UV avec une longueur d'onde de 254nm pendant 5 min. La lampe émettrice des UV se situait à environ 10cm des ARN, qui ont été préalablement placés sur glace.

Afin de purifier et concentrer des ARN, des billes RNA Clean XP ont été utilisées. A été ajouté à chaque échantillon : 100μL H₂O, 270μL de billes RNA Clean XP (permettant de capter 150ng d'ARN environ) et 810 μL d'Isopropanol. Après une dizaine d'homogénéisation par pipetage, une incubation à RT durant 5 mins a été effectuée. Les tubes ont ensuite été placés sur un portoir magnétique durant 15mins pour immobiliser les billes. Deux lavages des billes de 1mL avec de l'éthanol à 85% ont été réalisées, puis les billes ont été mises à sécher durant 10 mins. Les ARN ont été élués à RT durant 2 mins dans 20μL d'H₂O.

Méthodes de détection des ARN

Pour les Northern Blot, les ARN ont été séparés en fonction de leur taille sur gels de séquençages à 6% et transférés via un champ électrique sur une membrane en nylon HYbond-N. La détection des ARN a été effectuée par l'ajout d'une sonde oligonucléotidique antisens radiomarquée en 5' au ³²P.

Les protections aux RNases A/T1 ont été effectuées selon (Goodall *et al*, 1990). Pour l'analyse de SNORA55 et 7SK endogènes, les sondes ARN antisens ont été synthétisées *in vitro* par la RNA polymérase T3 (Promega), avec 30μCi de [α-³²P]CTP avec comme template les plasmides

pBST-SNORA55 et pBST-7SK linéarisés par ClaI. Pour l'analyse des snoARN H/ACA putatifs endogènes et en conditions de surexpression, les sondes antisens ont été synthétisées *in vitro* par la RNA polymérase SP6 (Promega), avec 30µCi de [α -³²P]CTP avec comme template les plasmides pCGL-1-snoARN H/ACA linéarisés par Hind III. Les différentes sondes utilisées ont préalablement été purifiées en gel dénaturant 6%, et leur dégradation par les RNases A et T1 a été vérifiée. Les fragments de sonde protégés par les ARN ont été séparés sur gel de séquençage à 6%.

Les radiomarquages des extrémités 3' des ARN au pCp ont été effectués sur 1µg ou moins d'ARN dilués avec de l'H₂O dans un volume final de 4,5µL. Ont ensuite été ajoutés 10µCi de [5'-³²P]pCp, 1 µl du tampon RNA Ligase 10X, 1,5µL de DMSO, 1µL de RNasin (Promega) et 3 unités de T4 RNA Ligase (New England Biolabs). Les réactions ont été incubées O/N à 4°C. Les ARN radiomarqués ont été séparés par migration sur gel de séquençage 6%.

Les signaux radioactifs ont été révélés avec l'utilisation d'un Typhoon Biomolecular Imager (Amersham) et visualisés avec le logiciel Multi Gauge V3.0.

Les analyses par RT-PCR quantitatives en temps réel ont été réalisées sur des ARN purifiés à partir de 4µg d'ARN totaux. Les premiers brins d'ADN ont été synthétisés par la reverse transcriptase GoScript (Promega) et les réactions qPCR ont été effectués sur un thermocycleur C1000Touch™ Thermal Cycler (Bio-Rad) en utilisant le SYBR Green Supermix IQ de BioRad. Pour les analyses, la méthode des Δ Ct a été utilisée et la quantité des ARN purifiée a été rapportée aux Input.

Génération de la lignée SNORA55 KO

Des cellules HAP1 ont été transfectées avec le vecteur CRISPR pX459 puro – ARN guide 1 et 2 pour induire les coupages d'ADN autour de gène *SNORA55*. Après sélection transitoire de 2 jours par puromycine, les cellules ont été isolées dans des plaques 96 puits à raison d'une cellule/puit. Les ARN de différentes lignées monoclonales ont été extraits et la délétion génomique de *SNORA55* a été vérifiée par PCR, l'intégrité de l'ARNm du gène hôte par RT-PCR et l'accumulation de l'ARN SNORA55 par digestion aux RNases. Les oligonucléotides pour générer les ARN guides utilisés sont consignés dans l'**Annexe 3**.

Cartographie de l'ARNr 18S au CMCT

Les ARN cellulaires ont été traités au N-cyclohexyl-N'-(2-morpholinoéthyl)carbodiimide métho-p-toluènesulfonate (CMCT) et soumis à une extension d'amorces (Bakin & Ofengand, 1998; Kiss & Jádý, 2004). Les oligonucléotides utilisés pour la cartographie des pseudouridines de l'ARNr 18S est indiqué dans l'Annexe 3

Différentiation neuronale des cellules murines P19

Pour préparer le processus de différenciation, les cellules murines P19 ont été transférées dans des boîtes de Pétri traditionnelles de 90mm de diamètre afin d'éviter que les cellules n'adhèrent à la surface de la boîte. 0,15 µg/ml d'acide rétinoïque (RA) ont été ajoutés au milieu de culture des cellules, marquant le jour « 0 » du protocole de différenciation des P19 en cellules neuronales. Au jour 2, les cellules n'ont pas adhéré sur le fond des boîtes, et formaient des amas cellulaires, appelés « clumps », flottant dans le milieu. Les clumps ont été prélevés par une aspiration du milieu, et laissés sédimenter par gravité dans un tube Falcon de 15mL. Le surnageant a été écarté et les clumps ont été repris dans du milieu DMEM + 10% SVF + RA et remis en culture dans des boîtes de Pétri de 90mm. Au jour 4, de la même manière que le jour 2, les clumps de chaque boîte ont été isolés dans un tube Falcon 15mL, auquel a été ajouté 4-5mL de Trypsine 0,05% - EDTA (Gibco). Les cellules composant les clumps ont été séparées par une incubation durant 5 mins à 37°C et en agitant légèrement le tube de temps à temps. Après ajout de DMEM (v/v), les cellules ont été centrifugées à 300 x g durant 5 mins. Le surnageant a été écarté et les cellules ont été re-suspendues dans du milieu DMEM + 10% SVF. Les cellules isolées ont été placées (1.10^6 cellules/boîte) dans des boîtes de culture cellulaire 100mm qui ont préalablement été recouverte de PBS + Gélatine 0,1% O/N. Le PBS + Gélatine 0,1% a été retiré 30mins avant l'ajout des cellules. Les cellules isolées ont ensuite été cultivées sur ces boîtes de culture recouvertes de Gélatine. Au jour 5, le milieu des cellules isolées a été remplacé par du milieu DMEM + 10% SVF + cytosine arabinoside (Ara-C) à une concentration finale de 15µg/mL. Au jour 6, le milieu a été remplacé en fonction de si la quantité de cellules mortes étaient trop importante. Jusqu'au jour 12, le milieu des cellules a été renouvelé tous les deux jours environ, et la concentration en Ara-C du milieu a été maintenu.

Dissociation et isolement de cellules neuronales à partir de cerveaux murins

Un demi cerveau murin des lignées WT ou KO SNORD115 a été utilisé pour chaque expérience. Après dissections, effectuées par Virginie MARTY, les demi-cerveaux sont gardés

dans du PBS (Gibco) + 0,5% Glucose. Après un lavage au PBS + 0,5% Glucose, les demi-cerveaux ont été sectionnés en morceaux plus petits à l'aide d'un scalpel. Les morceaux ont été digérés par 0,5 mL de Trypsine 0,5% - EDTA (Gibco), diluée ou non dans 10mL de PBS durant 15 ou 30 minutes. Les délais écoulés, les tissus digérés ont subi un broyage mécanique doux par une dizaine d'aspirations/refoulements à l'aide d'une pipette 5mL puis une seconde dizaine d'aspirations/refoulements à l'aide d'une pipette pasteur fine. Les cellules libérées ont ensuite été filtrées sur un tamis de 70 μ m (BD Falcon), puis centrifugées à 160 x g durant 6 mins. Les cellules ont été re-suspendus dans (i) 600 μ L de tampon d'extraction d'ARN pour la purification des ARN totaux des cellules isolées par la méthode au thiocyanate de guanidine-phénol-chloroforme, décrite dans la publication (Kiss & Jádý, 2004) (ii) 600 μ L de tampon 150mM NaCL/10mM Tris-HCl pour le Crosslink.

Approche Bio-informatiques

Les reads sont issus des séquençages NGS par small RNA-Seq ou RNA-Seq classique effectués par la société GenewizTM. L'analyse NGS a été réalisée dans un environnement Linux à l'aide du terminal. Pour l'analyse des RNA-Seq, le retrait des adaptateurs Illumina standard a été effectué par le programme TrimGalore (0.6.0) en gardant les paramètres par défauts.

```
~path/trim_galore --path_to_cutadapt ~path/cutadapt --paired --retain_unpaired --
illumina --trim-n --core 4 --output_dir ~/directory ~path/forwardreads.fasta
~path/reversereads.fasta
```

La qualité des reads après le retrait des adaptateurs a été contrôlée par le programme FastQC (version 0.11.8). L'alignement des reads au génome humain a été réalisée avec la version annotée hg38, récupérée sur <http://sapac.support.illumina.com>. L'alignement a été effectué avec le programme bwa mem (v. 0.7.17-r1188).

```
bwa mem ~path/genome.fa ~path/ forwardreads.fasta ~path/reversereads.fasta
>alignment.sam
```

Le recensement des reads alignés a été effectué par le programme featureCounts (v 2.0.0).

```
featureCounts -a ~path/genes.gtf -o filename ~path/alignment.sam
```

La suite des analyses exploratoires a été réalisée sur le logiciel R avec l'environnement de travail Rstudio.

Concernant l'analyse des reads obtenus par small RNA-Seq, le retrait des adaptateurs Illumina « small RNA » a été effectué par le programme TrimGalore (0.6.0) en gardant les paramètres par défauts.

```
~path/trim_galore --path_to_cutadapt ~path/cutadapt --paired --retain_unpaired --
adapter AGATCGGAAGAGC --adapter2 GATCGTCGGACT --trim-n --core 4 --
output_dir ~/directory ~path/forwardreads.fasta ~path/reversereads.fasta
```

Les reads Forward et Reverse ont ensuite été assemblés par le programme VSEARCH (v. 2.22.1).

```
vsearch --fastq_mergepairs FILENAME --reverse FILENAME --fastqout FILENAME
```

Les ARN chimères ont ensuite été analysées avec le logiciel ChiRA (v 1.4.3) composé des différentes étapes suivantes :

Collapse merge reads

```
chira_collapse.py -i ~path/filename -o collapsedreads.fasta
```

Alignement of reads :

```
chira_map.py -i collapsedreads.fasta -o ~path -b -f1 path/reference1.fa -f2
path/reference2.fa
```

ChiRA merge:

```
chira_merge.py -b sorted.bed -o ~path -f1 path/reference1.fa -f2 path/reference2.fa
```

Quantify

```
chira_quantify.py -b segments.bed -m merged.bed -o outputdirectory
```

Extract

```
chira_extract.py -l ~path/loci.counts -o outputdirectory -f1 path/reference1.fa -f2
path/reference2.fa
```

ANNEXES

tetraodonSNORA55	--GCACCTCTCAGCAACCCAGTGGACTCTGCT-GAGTGGTCAGATTGGGTATCAGATGGT	57
zebrafishSNORA55	--GCACTCTGC-AAACCCCTGTCTCTGTGT-CAGTGCTGGGCACGGGT-TT-AAAAAGT	54
yellowtailSNORA55	--GCACCTCACAACAACCCCTGTCTGTGT-CAGTGTTTGATACGGGT-CACGCTGGT	56
Salvelinus alpinusSNORA55	--GCACCCAGC-AAACCCCTGTCC--TCCTGT-CTGTGTTGGCCATGGGC-AATGATGGT	53
armadilloSNORA55	GAGCACCTGAATCTTTCCCATTCCTGCTGCT-TTGTGCTGGTGTGGGG--ACAGATGGT	57
lizardSNORA55	ATGCACCTGAATCTTTCCAGGCCCTCGAAAACAAGTGGGGGCTGGGG--ACAAGGGGT	58
turtleSNORA55	GTGCACCTGAATCTTTCCCATTCCTTGCAGTA-CAGTGTTGGGCTGGGG--ACACATGGT	57
xenopusSNORA55	TTGCACTGACTTCTCATCTATCCCTGCTGCT-AAGTGTAGGGCAAG-G--TCACTCAGT	56
opossumSNORA55	CTGCACCTGCATCTTTCCCATTCCTTGCTGCT-TTGTGCTGTCTGGGG--ACAGGTGGT	57
chickenSNORA55	CAGCACCTGAATCTTTCCATTCCTTGCTGTG-CAGTGCAAGTCTGGGG--ACAGTTGGT	57
horseSNORA55	GAGCACCTGAATCTTTCCCATTCCTTGCTGCT-TTGTGCTGGTGTGGGG--ACAGATGGT	57
dogSNORA55	GAGCACCTGAATCTTTCCCATTCCTTGCTGCT-TTGTGCTGGTGTGGGG--ACAGATGGT	57
turkeySNORA55	CAGCACCTGAATCTTTCCATTCCTTGCTGTG-CAGTGCTGGTCTAGGG--ACAGTTGGT	57
humanSNORA55	GAGCACCTGAATCTTTCCCATTCCTTGCTGCC-TCGTGCCGGTGTGGGG--ACAGATGGT	57
brownkiwiSNORA55	CAGCACCTGAATCTTTCCATTCCTTGCTGTG-CAGTGCAAGTCTGGGG--ACAGTTGGT	57
	**** : * : **** . : * *	
tetraodonSNORA55	GCTAGAACG-ACTCACTGTGTATCCAGATGGACCTCCCC-----CCTCTGTGGAGCCTGT	111
zebrafishSNORA55	GCTAAAGTA-CCTCAGAGTGCATCCAGACAAGTCTCCTC-----TTCTATTGGCGCTTGT	108
yellowtailSNORA55	GCCATAGAG-ACTCAATGTAAATCCAGATGGACCTCCTC-----TTCCTATGGAGCCTGT	110
Salvelinus alpinusSNORA55	GCCATAGTG-ACTCAAAGTATATCCAGACGGACCTCCCT-----CTCTATTGGAGCCTGT	107
armadilloSNORA55	GCTACAGTATGAGCAGAGGAAATCCAGACAGGTGTTTTCATTGCTTG-AGGCCTGT	116
lizardSNORA55	GCTAAAGCA-AAGCAGAGGAAATCCAGATGGGCCGCTCTTATCTGCAGTGAGGGCCTGT	117
turtleSNORA55	GCTAACAGCAAAGCAGAGTAAATCCAGACAGGTGTTTTCCTCCCTCATTGAGGGCTTGT	117
xenopusSNORA55	GCTAAAGCA-ATGCAGAGGAAATCCAGATGGAGCATGTT---TTACTTT-GTGTCTGT	110
opossumSNORA55	GCTACAGTACAAGCAGAGGAAATCCAGACAGGTATCTTCTGTTTACTTT-GGGGCTTGT	116
chickenSNORA55	GCTACAGCACA-GCAGCAGAAATCCAGATGGGTACTTTCCACTCGC-TT-GGAACTCAT	114
horseSNORA55	GCTACAGTATGAGCAGAGGAAATCCAGACGGGTGTTTTCATTGTCTT-GGGGCTTGT	116
dogSNORA55	GCTACAGTATGAGCAGAGGAAATCCAGACGGGTGTTTTCATTGTCTT-GGGGCTTGT	116
turkeySNORA55	GCTACAGCA-CAGCAGAGAAATCCAGATGGGTATTTTCCACTT-GCTT-GGAACTCAT	114
humanSNORA55	GCTACAGAAATGAGCAGAGGAAATCCAGACAGGTGTTTTCATTGTCTT-GGGGCTTGT	116
brownkiwiSNORA55	GCTACAGCA-AAGCAGTGGAAATCCAGATGGGTGCTTGCTTTC-ACCT-GGAACTCAT	114
	** * . . * * . . ***** . . *	
tetraodonSNORA55	CTCTACAAC-TTTGAAACATTA-	132
zebrafishSNORA55	CTCTACAAC-TTTGAAACAGC-	130
yellowtailSNORA55	CTCTACAAC-TTTGATACAAAA-	131
Salvelinus alpinusSNORA55	CTCTACAAC-TTTGGCACAAAA-	128
armadilloSNORA55	CTCTACAAC-CTGCCACAGTT-	137
lizardSNORA55	CTCTACAAC-CTGCTACACTG-	140
turtleSNORA55	CTCTACAAC-CTGCAACATTC-	138
xenopusSNORA55	CTCTACAAC-CTGCCAATTC-	131
opossumSNORA55	CTCTACAAC-CTGCCACACTT-	137
chickenSNORA55	CTCTACAAC-CTGCAACATTT-	135
horseSNORA55	CTCTACAAC-CTGCCAATTC-	137
dogSNORA55	CTCTACAAC-CTGCCAATTC-	137
turkeySNORA55	CTCTACAAC-CTGCAACATTT-	135
humanSNORA55	CTCTACAAC-CTGCCAATTT-	137
brownkiwiSNORA55	CTCTACAAC-CTGCAACATTT-	135
	***** . : * * *	

Annexe 1 : Conservation de SNORA55 entre différentes espèces

Les boîtes H et ACA sont en rouge, et les éléments antisens putatifs en bleu

Ref Noms Oligos	Séquence 5' → 3'	Rôle
Putative RNA ACA6	ATA ATC GAT GCT AGG AGG TCA TTG CTT TCT GA	Clonage de l'ARN putatif ACA 6
Putative RNA ACA6	ATA ACG CGT CAA AGT GCT GGG AGC TAC CA	
Putative RNA ACA11	ATA ATC GAT GTG TGC ACA TGT GCA TGT GC	Clonage de l'ARN putatif ACA 11
Putative RNA ACA11	ATA CTC GAG CTA TCT GCT CAG GTT TGG GC	
Putative RNA ACA29A & B	ATA ATC GAT GCT GGA ATT ACA GGC GTG	Clonage de l'ARN putatif ACA 29A & B
Putative RNA ACA29A	ATA CTC GAG GCT CCC TAA AGA ATT AGA AC	Clonage de l'ARN putatif ACA 29A
Putative RNA ACA29B	ATA CTC GAG CAC TGT TCT GTT GAA ATG CAG	Clonage de l'ARN putatif ACA 29B
Putative RNA ACA39	ATA ATC GAT TTC GGT CTC CCG CGG GGA	Clonage de l'ARN putatif ACA 39
Putative RNA ACA39	ATA ACG CGT TAC GGT GCC CAG ACC TGC G	
Putative RNA ACA56	ATA ATC GAT GTT CTC TGA CAT AAG CCA TTG C	Clonage de l'ARN putatif ACA 56
Putative RNA ACA56	ATA CTC GAG AAC AAA CTA CAG TTC CTT CCC	
Putative RNA ACA116	ATA ATC GAT GCT GGA AAG AAG AGT GTA CAG	Clonage de l'ARN putatif ACA 116
Putative RNA ACA116	ATA CTC GAG CAA CTT CTC TGA GCC TTA TAC C	
Putative RNA ACA120	ATA ATC GAT TCG GCC CTC CCA ATA GT	Clonage de l'ARN putatif ACA 120
Putative RNA ACA120	ATA CTC GAG GTG CAG AGT CAT GGT GGA A	
Putative RNA ACA133	ATA ATC GAT CAG GAC CCC CTT GCC TT	Clonage de l'ARN putatif ACA 133
Putative RNA ACA133	ATA CTC GAG AGG CCA GGA CCA GTG GTT	
Putative RNA ACA186	ATA ATC GAT CAA TGG ATG CCT TTT CTC CC	Clonage de l'ARN putatif ACA 186
Putative RNA ACA186	ATA CTC GAG GAC TCG ACG AGC TAT GTT	
Putative RNA ACA265A	ATA ATC GAT GAC ATG AAG CCT CCA CCT CT	Clonage de l'ARN putatif ACA 265A
Putative RNA ACA265A	ATA CTC GAG TGT GAA AAC CCA GTG TCT C	
Putative RNA ACA265B	ATA ATC GAT GCA AAT GGG GAG CTC TG	Clonage de l'ARN putatif ACA 265B
Putative RNA ACA265B	ATA ACG CGT AGT CAG AAG GGT TGG CAG	
SnoRA55	GCC TGA ATC TTT CCC ATT CC	Amplifier cDNA SnoRA55
SnoRA55	AGA CAG GCC CCA AGA CAA AT	Amplifier cDNA SnoRA55
SCARNA 10	GGT GAT TCT TCC AGT TGC TG	(qPCR) Amplifier cDNA U85
SCARNA 10	TGA TAC CCT GAA CAT GCC CT	(qPCR) Amplifier cDNA U85
SnU5	GAT TTC CGT GGA GAG GAA CA	(qPCR) Amplifier cDNA SnU5
SnU5	CCA AAG CAA GGC CTC AAA A	(qPCR) Amplifier cDNA SnU5
SnoRA55	CAC CGT TGC TGC CTC GTG CCG GTG T	KO SNORA55 via CrispR/Cas9
SnoRA55	AAA CAC ACC GGC ACG AGG CAG CAA C	KO SNORA55 via CrispR/Cas9
SnoRA55	CAC CGG GCT TGG GAA CTT CTA CCC	KO SNORA55 via CrispR/Cas9
SnoRA55	AAA CGG GTA GAA GTT CCC AAG CCC	KO SNORA55 via CrispR/Cas9
SnoRA55	ACA CTG TTT CCC AAG CAC	DNAg Snora55 Amplification
SnoRA55	TGA GGT TCC TAC TTA CCC	DNAg Snora55 Amplification
SnoRA55	GAG CTG TAG AGA CAG GCC CCA AGA	(Northern Blot) Snora55 mRNA Detection
SNORA55 region	CAA GCA CTG GAG TCT GGA GA	Check deletion via CRISPR/Cas9 du gène SNORA55
SNORA55 region	ACC ATC CCC ATG TAG TCA GC	Check deletion via CRISPR/Cas9 du gène SNORA55
PABPC4	GTG TCA AGG TGA TGA GAG ATC	Amplifier exon 5 gène PABPC4 (gène hôte SNORA55)
PABPC4	TTA ACT CTG CCT GCC GTT CTA	Amplifier exon 6 gène PABPC4 (gène hôte SNORA55)
PABPC4	GGA TCA ATT ACC AGT GCT AAG	Amplifier exons 7 gène PABPC4 (gène hôte SNORA55)
PABPC4	AGC CCA CAA TGC GTC CAT TCA	Amplifier exons 8 gène PABPC4 (gène hôte SNORA55)
Putative RNA ACA186	ATA ATC GAT GCC GGG TGC TGG	Amplifier seulement la région du gène ACA186
Putative RNA ACA186	ATA CTC GAG TCT GTG AGC AGT GCT	Amplifier seulement la région du gène ACA186
Putative RNA ACA186	ATA ATC GAT AGG AGG TTA CAG AGG CAA	Amplifier seulement la région du gène ACA186
Putative RNA ACA186	ATA ATC GAT CCG GGT GCT GGA GCA GTG	Clonage de l'ARN putatif ACA 186

Annexe 3: Listes des Oligonucléotides utilisés

RÉFÉRENCES

BIBLIOGRAPHIQUES

- Abdalla E, Troy R, Fall S, Elhussin I, Egiebor-Aiwan O & Nganwa D (2020) Racial differences in 5-year relative survival rates of cervical cancer by stage at diagnosis, between African American (black) and white women, living in the state of Alabama, USA. *BMC Cancer* 20: 830
- Abramowski D & Staufenbiel M (1995) Identification of the 5-hydroxytryptamine_{2C} receptor as a 60-kDa N-glycosylated protein in choroid plexus and hippocampus. *J Neurochem* 65: 782–790
- Addepalli B & Limbach PA (2011) Mass Spectrometry-Based Quantification of Pseudouridine in RNA. *J Am Soc Mass Spectrom* 22: 1363–1372
- Adhikari A, Copping NA, Onaga B, Pride MC, Coulson RL, Yang M, Yasui DH, LaSalle JM & Silverman JL (2019) Cognitive deficits in the Snord116 deletion mouse model for Prader-Willi syndrome. *Neurobiol Learn Mem* 165: 106874
- Aittaleb M, Rashid R, Chen Q, Palmer JR, Daniels CJ & Li H (2003) Structure and function of archaeal box C/D sRNP core proteins. *Nat Struct Biol* 10: 256–263
- Alazami AM, Al-Owain M, Alzahrani F, Shuaib T, Al-Shamrani H, Al-Falki YH, Al-Qahtani SM, Alsheddi T, Colak D & Alkuraya FS (2012) Loss of function mutation in LARP7, chaperone of 7SK ncRNA, causes a syndrome of facial dysmorphism, intellectual disability, and primordial dwarfism. *Human Mutation* 33: 1429–1434
- Aris JP & Blobel G (1991) cDNA cloning and sequencing of human fibrillarin, a conserved nucleolar protein recognized by autoimmune antisera. *Proceedings of the National Academy of Sciences* 88: 931–935
- Armanios M & Blackburn EH (2012) The telomere syndromes. *Nat Rev Genet* 13: 693–704
- Aschenbrenner J & Marx A (2016) Direct and site-specific quantification of RNA 2'-O-methylation by PCR with an engineered DNA polymerase. *Nucleic Acids Research* 44: 3495–3502
- Aubert M, O'Donohue M-F, Lebaron S & Gleizes P-E (2018) Pre-Ribosomal RNA Processing in Human Cells: From Mechanisms to Congenital Diseases. *Biomolecules* 8: 123
- Aw JGA, Shen Y, Wilm A, Sun M, Lim XN, Boon K-L, Tapsin S, Chan Y-S, Tan C-P, Sim AYL, *et al* (2016) In Vivo Mapping of Eukaryotic RNA Interactomes Reveals Principles of Higher-Order Organization and Regulation. *Molecular Cell* 62: 603–617
- Ayadi L, Galvanin A, Pichot F, Marchand V & Motorin Y (2019) RNA ribose methylation (2'-O-methylation): Occurrence, biosynthesis and biological functions. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* 1862: 253–269
- Badrock AP, Uggenti C, Wacheul L, Crilly S, Jenkinson EM, Rice GI, Kasher PR, Lafontaine DLJ, Crow YJ & O'Keefe RT (2020) Analysis of U8 snoRNA Variants in Zebrafish Reveals How Bi-allelic Variants Cause Leukoencephalopathy with Calcifications and Cysts. *Am J Hum Genet* 106: 694–706

- Bakin A, Lane BG & Ofengand J (1994) Clustering of pseudouridine residues around the peptidyltransferase center of yeast cytoplasmic and mitochondrial ribosomes. *Biochemistry* 33: 13475–13483
- Bakin A & Ofengand J (1993) Four newly located pseudouridylate residues in Escherichia coli 23S ribosomal RNA are all at the peptidyltransferase center: analysis by the application of a new sequencing technique. *Biochemistry* 32: 9754–9762
- Bakin AV & Ofengand J (1998) Mapping of Pseudouridine Residues in RNA to Nucleotide Resolution. In *Protein Synthesis: Methods and Protocols*, Martin R (ed) pp 297–309. Totowa, NJ: Springer New York
- Balakin AG, Smith L & Fournier MJ (1996) The RNA world of the nucleolus: two major families of small RNAs defined by different box elements with related functions. *Cell* 86: 823–834
- Baldini L, Charpentier B & Labialle S (2021a) Emerging Data on the Diversity of Molecular Mechanisms Involving C/D snoRNAs. *Non-Coding RNA* 7
- Baldini L, Robert A, Charpentier B & Labialle S (2021b) Phylogenetic and Molecular Analyses Identify SNORD116 Targets Involved in the Prader–Willi Syndrome. *Mol Biol Evol* 39: msab348
- Ballarino M, Morlando M, Pagano F, Fatica A & Bozzoni I (2005) The cotranscriptional assembly of snoRNPs controls the biosynthesis of H/ACA snoRNAs in Saccharomyces cerevisiae. *Mol Cell Biol* 25: 5396–5403
- Bao H, Chen X, Liu X, Wu W, Li Q, Xian J, Zhao Y & Chen S (2022) Box C/D snoRNA SNORD89 influences the occurrence and development of endometrial cancer through 2'-O-methylation modification of Bim. *Cell Death Discov* 8: 309
- Barbieri I & Kouzarides T (2020) Role of RNA modifications in cancer. *Nat Rev Cancer* 20: 303–322
- Bazeley PS, Shepelev V, Talebizadeh Z, Butler MG, Fedorova L, Filatov V & Fedorov A (2008) snoTARGET shows that human orphan snoRNA targets locate close to alternative splice junctions. *Gene* 408: 172–179
- Begik O, Lucas MC, Prysycz LP, Ramirez JM, Medina R, Milenkovic I, Cruciani S, Liu H, Vieira HGS, Sas-Chen A, et al (2021) Quantitative profiling of pseudouridylation dynamics in native RNAs with nanopore sequencing. *Nat Biotechnol* 39: 1278–1291
- Berndt H, Harnisch C, Rammelt C, Stöhr N, Zirkel A, Dohm JC, Himmelbauer H, Tavanetz J-P, Hüttelmaier S & Wahle E (2012) Maturation of mammalian H/ACA box snoRNAs: PAPD5-dependent adenylation and PARN-dependent trimming. *RNA* 18: 958–972
- Bernick DL, Dennis PP, Höchsmann M & Lowe TM (2012) Discovery of Pyrobaculum small RNA families with atypical pseudouridine guide RNA features. *RNA* 18: 402–411
- Birkedal U, Christensen-Dalsgaard M, Krogh N, Sabarinathan R, Gorodkin J & Nielsen H (2015) Profiling of Ribose Methylations in RNA by High-Throughput Sequencing. *Angewandte Chemie International Edition* 54: 451–455

- Bleichert F, Gagnon KT, Brown BA, Maxwell ES, Leschziner AE, Unger VM & Baserga SJ (2009) A Dimeric Structure for Archaeal Box C/D Small Ribonucleoproteins. *Science* 325: 1384–1387
- Boccaletto P, Machnicka MA, Purta E, Piątkowski P, Bagiński B, Wirecki TK, de Crécy-Lagard V, Ross R, Limbach PA, Kotter A, *et al* (2018) MODOMICS: a database of RNA modification pathways. 2017 update. *Nucleic Acids Res* 46: D303–D307
- Bohnsack MT & Sloan KE (2018) Modifications in small nuclear RNAs and their roles in spliceosome assembly and function. *Biological Chemistry* 399: 1265–1276
- Boo SH & Kim YK (2020) The emerging role of RNA modifications in the regulation of mRNA stability. *Exp Mol Med* 52: 400–408
- Bortolin-Cavaillé M-L, Quillien A, Thalalla Gamage S, Thomas JM, Sas-Chen A, Sharma S, Plisson-Chastang C, Vandel L, Blader P, Lafontaine DLJ, *et al* (2022) Probing small ribosomal subunit RNA helix 45 acetylation across eukaryotic evolution. *Nucleic Acids Research* 50: 6284–6299
- Boulon S, Marmier-Gourrier N, Pradet-Balade B, Wurth L, Verheggen C, Jádý BE, Rothé B, Pescia C, Robert M-C, Kiss T, *et al* (2008a) The Hsp90 chaperone controls the biogenesis of L7Ae RNPs through conserved machinery. *J Cell Biol* 180: 579–595
- Boulon S, Marmier-Gourrier N, Pradet-Balade B, Wurth L, Verheggen C, Jádý BE, Rothé B, Pescia C, Robert M-C, Kiss T, *et al* (2008b) The Hsp90 chaperone controls the biogenesis of L7Ae RNPs through conserved machinery. *J Cell Biol* 180: 579–595
- Boulon S, Verheggen C, Jádý BE, Girard C, Pescia C, Paul C, Ospina JK, Kiss T, Matera AG, Bordonné R, *et al* (2004) PHAX and CRM1 are required sequentially to transport U3 snoRNA to nucleoli. *Mol Cell* 16: 777–787
- Bragantini B, Tiotiu D, Rothé B, Saliou J-M, Marty H, Cianférani S, Charpentier B, Quinternet M & Manival X (2016) Functional and Structural Insights of the Zinc-Finger HIT protein family members Involved in Box C/D snoRNP Biogenesis. *J Mol Biol* 428: 2488–2506
- Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA & Jemal A (2018) Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians* 68: 394–424
- Burke MF, Logan MK & Hebert MD (2018) Identification of additional regulatory RNPs that impact rRNA and U6 snRNA methylation. *Biol Open* 7: bio036095
- Byerly JH, Port ER & Irie HY (2020) PRKCQ inhibition enhances chemosensitivity of triple-negative breast cancer by regulating Bim. *Breast Cancer Res* 22: 72
- Calvet JP & Pederson T (1979) Photochemical cross-linking of secondary structure in HeLa cell heterogeneous nuclear RNA in situ 1. *Nucleic Acids Res* 6: 1993–2001
- de Carcer G & Medina FJ (1999) Simultaneous localization of transcription and early processing markers allows dissection of functional domains in the plant cell nucleolus. *J Struct Biol* 128: 139–151

- Carlile TM, Rojas-Duran MF & Gilbert WV (2015) Pseudo-Seq: Genome-Wide Detection of Pseudouridine Modifications in RNA. *Methods Enzymol* 560: 219–245
- Carlile TM, Rojas-Duran MF, Zinshteyn B, Shin H, Bartoli KM & Gilbert WV (2014) Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. *Nature* 515: 143–146
- Cassidy SB, Schwartz S, Miller JL & Driscoll DJ (2012) Prader-Willi syndrome. *Genetics in Medicine* 14: 10–26
- Cavaillé J (2017) Box C/D small nucleolar RNA genes and the Prader-Willi syndrome: a complex interplay. *WIREs RNA* 8: e1417
- Cavaillé J, Buiting K, Kieffmann M, Lalande M, Brannan CI, Horsthemke B, Bachellerie J-P, Brosius J & Hüttenhofer A (2000) Identification of brain-specific and imprinted small nucleolar RNA genes exhibiting an unusual genomic organization. *Proc Natl Acad Sci U S A* 97: 14311–14316
- Cavaillé J, Nicoloso M & Bachellerie JP (1996) Targeted ribose methylation of RNA in vivo directed by tailored antisense RNA guides. *Nature* 383: 732–735
- Cavaillé J, Seitz H, Paulsen M, Ferguson-Smith AC & Bachellerie J-P (2002) Identification of tandemly-repeated C/D snoRNA genes at the imprinted human 14q32 domain reminiscent of those at the Prader-Willi/Angelman syndrome region. *Human Molecular Genetics* 11: 1527–1538
- Charette M & Gray MW (2000) Pseudouridine in RNA: what, where, how, and why. *IUBMB Life* 49: 341–351
- Charpentier B, Muller S & Branlant C (2005) Reconstitution of archaeal H/ACA small ribonucleoprotein complexes active in pseudouridylation. *Nucleic Acids Res* 33: 3133–3144
- Chen J-L, Blasco MA & Greider CW (2000) Secondary Structure of Vertebrate Telomerase RNA. *Cell* 100: 503–514
- Choi J, Indrisiunaite G, DeMirci H, Jeong K-W, Wang J, Petrov A, Prabhakar A, Rechavi G, Dominissini D, He C, *et al* (2018) 2'-O-methylation in mRNA disrupts tRNA decoding during translation elongation. *Nat Struct Mol Biol* 25: 208–216
- Chung MS, Langouët M, Chamberlain SJ & Carmichael GG (2020) Prader-Willi syndrome: reflections on seminal studies and future therapies. *Open Biol* 10: 200195
- Clamer M, Höfler L, Mikhailova E, Viero G & Bayley H (2014) Detection of 3'-End RNA Uridylation with a Protein Nanopore. *ACS Nano* 8: 1364–1374
- Clouet-d'Orval B, Gaspin C & Mougin A (2005) Two different mechanisms for tRNA ribose methylation in Archaea: a short survey. *Biochimie* 87: 889–895
- Cohn WE (1951) Some results of the applications of ion-exchange chromatography to nucleic acid chemistry. *J Cell Physiol Suppl* 38: 21–40

- Crea F, Quagliata L, Michael A, Liu HH, Frumento P, Azad AA, Xue H, Pikor L, Watahiki A, Morant R, *et al* (2016) Integrated analysis of the prostate cancer small-nucleolar transcriptome reveals SNORA55 as a driver of prostate cancer progression. *Mol Oncol* 10: 693–703
- Dai Q, Luan G, Deng L, Lei T, Kang H, Song X, Zhang Y, Xiao Z-X & Li Q (2014) Primordial Dwarfism Gene Maintains Lin28 Expression to Safeguard Embryonic Stem Cells from Premature Differentiation. *Cell Reports* 7: 735–746
- Dai Q, Moshitch-Moshkovitz S, Han D, Kol N, Amariglio N, Rechavi G, Dominissini D & He C (2017) Nm-seq maps 2'-O-methylation sites in human mRNA with base precision. *Nat Methods* 14: 695–698
- Darzacq X, Jády BE, Verheggen C, Kiss AM, Bertrand E & Kiss T (2002) Cajal body-specific small nuclear RNAs: a novel class of 2'-O-methylation and pseudouridylation guide RNAs. *EMBO J* 21: 2746–2756
- Darzacq X, Kittur N, Roy S, Shav-Tal Y, Singer RH & Meier UT (2006) Stepwise RNP assembly at the site of H/ACA RNA transcription in human cells. *J Cell Biol* 173: 207–218
- Davis FF & Allen FW (1957) Ribonucleic acids from yeast which contain a fifth nucleotide. *J Biol Chem* 227: 907–915
- De Deurwaerdère P, Lagièrre M, Bosc M & Navailles S (2013) Multiple controls exerted by 5-HT_{2C} receptors upon basal ganglia function: from physiology to pathophysiology. *Exp Brain Res* 230: 477–511
- Deamer D, Akeson M & Branton D (2016) Three decades of nanopore sequencing. *Nat Biotechnol* 34: 518–524
- DEL CAMPO M, RECINOS C, YANEZ G, POMERANTZ SC, GUYMON R, CRAIN PF, MCCLOSKEY JA & OFENGAND J (2005) Number, position, and significance of the pseudouridines in the large subunit ribosomal RNA of *Haloarcula marismortui* and *Deinococcus radiodurans*. *RNA* 11: 210–219
- Dennis PP, Omer A & Lowe T (2001) A guided tour: small RNA function in Archaea. *Molecular Microbiology* 40: 509–519
- Deryusheva S & Gall JG (2013) Novel small Cajal-body-specific RNAs identified in *Drosophila*: probing guide RNA function. *RNA* 19: 1802–1814
- Deryusheva S & Gall JG (2017) Dual nature of pseudouridylation in U2 snRNA: Pus1p-dependent and Pus1p-independent activities in yeasts and higher eukaryotes. *RNA* 23: 1060–1067
- Dez C, Henras A, Faucon B, Lafontaine D, Caizergues-Ferrer M & Henry Y (2001a) Stable expression in yeast of the mature form of human telomerase RNA depends on its association with the box H/ACA small nucleolar RNP proteins Cbf5p, Nhp2p and Nop10p. *Nucleic Acids Res* 29: 598–603

- Dez C, Henras A, Faucon B, Lafontaine DLJ, Caizergues-Ferrer M & Henry Y (2001b) Stable expression in yeast of the mature form of human telomerase RNA depends on its association with the box H/ACA small nucleolar RNP proteins Cbf5p, Nhp2p and Nop10p. *Nucleic Acids Research* 29: 598–603
- Dez C, Noaillac-Depeyre J, Caizergues-Ferrer M & Henry Y (2002) Naf1p, an essential nucleoplasmic factor specifically required for accumulation of box H/ACA small nucleolar RNPs. *Mol Cell Biol* 22: 7053–7065
- Didychuk AL, Butcher SE & Brow DA (2018a) The life of U6 small nuclear RNA, from cradle to grave. *RNA* 24: 437–460
- Didychuk AL, Butcher SE & Brow DA (2018b) The life of U6 small nuclear RNA, from cradle to grave. *RNA* 24: 437–460
- Dimitrova DG, Teyssset L & Carré C (2019) RNA 2'-O-Methylation (Nm) Modification in Human Diseases. *Genes* 10: 117
- Ding Y, Tang Y, Kwok CK, Zhang Y, Bevilacqua PC & Assmann SM (2014) In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature* 505: 696–700
- Dong JT (2001) Chromosomal deletions and tumor suppressor genes in prostate cancer. *Cancer Metastasis Rev* 20: 173–193
- Dong X-Y, Guo P, Boyd J, Sun X, Li Q, Zhou W & Dong J-T (2009) Implication of snoRNA U50 in human breast cancer. *Journal of genetics and genomics = Yi chuan xue bao* 36: 447
- Dong X-Y, Rodriguez C, Guo P, Sun X, Talbot JT, Zhou W, Petros J, Li Q, Vessella RL, Kibel AS, *et al* (2008) SnoRNA U50 is a candidate tumor suppressor gene at 6q14.3 with a mutation associated with clinically significant prostate cancer. *Hum Mol Genet* 17: 1031–1042
- Dong Z-W, Shao P, Diao L-T, Zhou H, Yu C-H & Qu L-H (2012) RTL-P: a sensitive approach for detecting sites of 2'-O-methylation in RNA molecules. *Nucleic Acids Research* 40: e157
- Duan J, Li L, Lu J, Wang W & Ye K (2009) Structural mechanism of substrate RNA recruitment in H/ACA RNA-guided pseudouridine synthase. *Mol Cell* 34: 427–439
- Duc NM, Kim HR & Chung KY (2015) Structural mechanism of G protein activation by G protein-coupled receptor. *Eur J Pharmacol* 763: 214–222
- Duechler M, Leszczyńska G, Sochacka E & Nawrot B (2016) Nucleoside modifications in the regulation of gene expression: focus on tRNA. *Cell Mol Life Sci* 73: 3075–3095
- Duker AL, Ballif BC, Bawle EV, Person RE, Mahadevan S, Alliman S, Thompson R, Traylor R, Bejjani BA, Shaffer LG, *et al* (2010) Paternally inherited microdeletion at 15q11.2 confirms a significant role for the SNORD116 C/D box snoRNA cluster in Prader-Willi syndrome. *Eur J Hum Genet* 18: 1196–1201

- Dunbar DA & Baserga SJ (1998) The U14 snoRNA is required for 2'-O-methylation of the pre-18S rRNA in *Xenopus* oocytes. *RNA* 4: 195–204
- Dunn-Davies H, Dudnakova T, Langhendries J-L, Watkins N, Lafontaine DLJ & Tollervey D (2021) Systematic mapping of small nucleolar RNA targets in human cells. 2021.07.22.451324 doi:10.1101/2021.07.22.451324 [PREPRINT]
- Egloff S, Studniarek C & Kiss T (2018) 7SK small nuclear RNA, a multifunctional transcriptional regulatory RNA with gene-specific features. *Transcription* 9: 95–101
- Elliott BA, Ho H-T, Ranganathan SV, Vangaveti S, Ilkayeva O, Abou Assi H, Choi AK, Agris PF & Holley CL (2019) Modification of messenger RNA by 2'-O-methylation regulates gene expression in vivo. *Nat Commun* 10: 3401
- Emmereichs G, Herdewijn P & Rozenski J (2005) Pseudouridine detection improvement by derivatization with methyl vinyl sulfone and capillary HPLC-mass spectrometry. *J Chromatogr B Analyt Technol Biomed Life Sci* 825: 233–238
- Ender C, Krek A, Friedländer MR, Beitzinger M, Weinmann L, Chen W, Pfeffer S, Rajewsky N & Meister G (2008) A Human snoRNA with MicroRNA-Like Functions. *Molecular Cell* 32: 519–528
- Enright CA, Maxwell ES, Eliceiri GL & Sollner-Webb B (1996) 5'ETS rRNA processing facilitated by four small RNAs: U14, E3, U17, and U3. *RNA* 2: 1094–1099
- Ewens CA, Su M, Zhao L, Nano N, Houry WA & Southworth DR (2016) Architecture and Nucleotide-Dependent Conformational Changes of the Rvb1-Rvb2 AAA+ Complex Revealed by Cryoelectron Microscopy. *Structure* 24: 657–666
- Falaleeva M, Pages A, Matuszek Z, Hidmi S, Agranat-Tamir L, Korotkov K, Nevo Y, Eyraş E, Sperling R & Stamm S (2016) Dual function of C/D box small nucleolar RNAs in rRNA modification and alternative pre-mRNA splicing. *Proc Natl Acad Sci U S A* 113: E1625–E1634
- Falaleeva M & Stamm S (2013) Processing of snoRNAs as a new source of regulatory non-coding RNAs: snoRNA fragments form a new class of functional RNAs. *Bioessays* 35: 46–54
- Falaleeva M, Surface J, Shen M, de la Grange P & Stamm S (2015) SNORD116 and SNORD115 change expression of multiple genes and modify each other's activity. *Gene* 572: 266–273
- Falaleeva M, Welden JR, Duncan MJ & Stamm S (2017) C/D-box snoRNAs form methylating and non-methylating ribonucleoprotein complexes: Old dogs show new tricks. *BioEssays* 39: 1600264
- Farley KI & Baserga SJ (2016) Probing the mechanisms underlying human diseases in making ribosomes. *Biochem Soc Trans* 44: 1035–1044
- Feigon J (2015) Back to the future of RNA structure. *RNA* 21: 611–612

- Fleming AM & Burrows CJ (2022) Nanopore sequencing for N1-methylpseudouridine in RNA reveals sequence-dependent discrimination of the modified nucleotide triphosphate during transcription. 2022.06.03.494690 doi:10.1101/2022.06.03.494690 [PREPRINT]
- Freyhult E, Edvardsson S, Tamas I, Moulton V & Poole AM (2008) Fisher: a program for the detection of H/ACA snoRNAs using MFE secondary structure prediction and comparative genomics – assessment and update. *BMC Res Notes* 1: 49
- Fu D & Collins K (2003) Distinct Biogenesis Pathways for Human Telomerase RNA and H/ACA Small Nucleolar RNAs. *Molecular Cell* 11: 1361–1372
- Gagnon KT, Zhang X, Qu G, Biswas S, Suryadi J, Brown BA & Maxwell ES (2010) Signature amino acids enable the archaeal L7Ae box C/D RNP core protein to recognize and bind the K-loop RNA motif. *RNA* 16: 79–90
- Gall JG, Bellini M, Wu Z & Murphy C (1999) Assembly of the Nuclear Transcription and Processing Machinery: Cajal Bodies (Coiled Bodies) and Transcriptosomes. *MBoC* 10: 4385–4402
- Ganot P, Bortolin M-L & Kiss T (1997) Site-Specific Pseudouridine Formation in Preribosomal RNA Is Guided by Small Nucleolar RNAs. *Cell* 89: 799–809
- Ganot P, Jády BE, Bortolin M-L, Darzacq X & Kiss T (1999) Nucleolar Factors Direct the 2'-O-Ribose Methylation and Pseudouridylation of U6 Spliceosomal RNA. *Molecular and Cellular Biology* 19: 6906–6917
- Garalde DR, Snell EA, Jachimowicz D, Sipos B, Lloyd JH, Bruce M, Pantic N, Admassu T, James P, Warland A, *et al* (2018) Highly parallel direct RNA sequencing on an array of nanopores. *Nat Methods* 15: 201–206
- Garus A & Autexier C (2021) Dyskerin: an essential pseudouridine synthase with multifaceted roles in ribosome biogenesis, splicing, and telomere maintenance. *RNA* 27: 1441–1458
- Gaspin C, Cavaillé J, Erauso G & Bachellerie JP (2000) Archaeal homologs of eukaryotic methylation guide small nucleolar RNAs: lessons from the *Pyrococcus* genomes. *J Mol Biol* 297: 895–906
- Gautier T, Bergès T, Tollervey D & Hurt E (1997) Nucleolar KKE/D repeat proteins Nop56p and Nop58p interact with Nop1p and are required for ribosome biogenesis. *Mol Cell Biol* 17: 7088–7098
- Gay DM, Lund AH & Jansson MD (2022) Translational control through ribosome heterogeneity and functional specialization. *Trends Biochem Sci* 47: 66–81
- Ge J & Yu Y-T (2013) RNA pseudouridylation: new insights into an old modification. *Trends Biochem Sci* 38: 210–218
- Ghanim GE, Fountain AJ, van Roon A-MM, Rangan R, Das R, Collins K & Nguyen THD (2021) Structure of human telomerase holoenzyme with bound telomeric DNA. *Nature* 593: 449–453

- Girard JP, Lehtonen H, Caizergues-Ferrer M, Amalric F, Tollervey D & Lapeyre B (1992) GAR1 is an essential small nucleolar RNP protein required for pre-rRNA processing in yeast. *EMBO J* 11: 673–682
- Godin KS, Walbott H, Leulliot N, van Tilbeurgh H & Varani G (2009) The box H/ACA snoRNP assembly factor Shq1p is a chaperone protein homologous to Hsp90 cochaperones that binds to the Cbf5p enzyme. *J Mol Biol* 390: 231–244
- Gong J, Shao D, Xu K, Lu Z, Lu ZJ, Yang YT & Zhang QC (2018) RISE: a database of RNA interactome from sequencing experiments. *Nucleic Acids Res* 46: D194–D201
- Goodall GJ, Wiebauer K & Filipowicz W (1990) Analysis of pre-mRNA processing in transfected plant protoplasts. *Methods Enzymol* 181: 148–161
- Goodarzi H, Liu X, Nguyen HCB, Zhang S, Fish L & Tavazoie SF (2015) Endogenous tRNA-Derived Fragments Suppress Breast Cancer Progression via YBX1 Displacement. *Cell* 161: 790–802
- Goodfellow CE, Graham SE, Dragunow M & Glass M (2011) Characterization of NTERA2/D1 cells as a model system for the investigation of cannabinoid function in human neurons and astrocytes. *J Neurosci Res* 89: 1685–1697
- Granneman S, Kudla G, Petfalski E & Tollervey D (2009) Identification of protein binding sites on U3 snoRNA and pre-rRNA by UV cross-linking and high-throughput analysis of cDNAs. *Proc Natl Acad Sci U S A* 106: 9613–9618
- Greider CW & Blackburn EH (1989) A telomeric sequence in the RNA of Tetrahymena telomerase required for telomere repeat synthesis. *Nature* 337: 331–337
- Grozdanov PN, Roy S, Kittur N & Meier UT (2009) SHQ1 is required prior to NAF1 for assembly of H/ACA small nucleolar and telomerase RNPs. *RNA* 15: 1188–1197
- Gupte R, Liu Z & Kraus WL (2017) PARPs and ADP-ribosylation: recent advances linking molecular functions to biological outcomes. *Genes Dev* 31: 101–126
- Håkansson KEJ, Goossens EAC, Trompet S, van Ingen E, de Vries MR, van der Kwast RVCT, Ripa RS, Kastrup J, Hohensinner PJ, Kaun C, *et al* (2019) Genetic associations and regulation of expression indicate an independent role for 14q32 snoRNAs in human cardiovascular disease. *Cardiovascular Research* 115: 1519–1532
- Hamma T & Ferré-D'Amaré AR (2004) Structure of Protein L7Ae Bound to a K-Turn Derived from an Archaeal Box H/ACA sRNA at 1.8 Å Resolution. *Structure* 12: 893–903
- Hamma T & Ferré-D'Amaré AR (2006) Pseudouridine synthases. *Chem Biol* 13: 1125–1135
- Han C, Sun L-Y, Luo X-Q, Pan Q, Sun Y-M, Zeng Z-C, Chen T-Q, Huang W, Fang K, Wang W-T, *et al* (2022) Chromatin-associated orphan snoRNA regulates DNA damage-mediated differentiation via a non-canonical complex. *Cell Reports* 38: 110421
- Hasler D, Meduri R, Bāk M, Lehmann G, Heizinger L, Wang X, Li Z-T, Sement FM, Bruckmann A, Dock-Bregeon A-C, *et al* (2020) The Alazami Syndrome-Associated

- Protein LARP7 Guides U6 Small Nuclear RNA Modification and Contributes to Splicing Robustness. *Molecular Cell* 77: 1014–1031.e13
- He X, Chen X, Zhang X, Duan X, Pan T, Hu Q, Zhang Y, Zhong F, Liu J, Zhang H, *et al* (2015) An Lnc RNA (GAS5)/SnoRNA-derived piRNA induces activation of TRAIL gene by site-specifically recruiting MLL/COMPASS-like complexes. *Nucleic Acids Res* 43: 3712–3725
- Hebras J, Marty V, Personnaz J, Mercier P, Krogh N, Nielsen H, Aguirrebengoa M, Seitz H, Pradere J-P, Guiard BP, *et al* (2020) Reassessment of the involvement of Snord115 in the serotonin 2c receptor pathway in a genetically relevant mouse model. *Elife* 9: e60862
- Heiss NS, Knight SW, Vulliamy TJ, Klauck SM, Wiemann S, Mason PJ, Poustka A & Dokal I (1998) X-linked dyskeratosis congenita is caused by mutations in a highly conserved gene with putative nucleolar functions. *Nat Genet* 19: 32–38
- Helwak A, Kudla G, Dudnakova T & Tollervey D (2013) Mapping the Human miRNA Interactome by CLASH Reveals Frequent Noncanonical Binding. *Cell* 153: 654–665
- Henras A, Henry Y, Bousquet-Antonelli C, Noaillac-Depeyre J, Gélugne JP & Caizergues-Ferrer M (1998) Nhp2p and Nop10p are essential for the function of H/ACA snoRNPs. *EMBO J* 17: 7078–7090
- Henras AK, Plisson-Chastang C, Humbert O, Romeo Y & Henry Y (2017) Synthesis, Function, and Heterogeneity of snoRNA-Guided Posttranscriptional Nucleoside Modifications in Eukaryotic Ribosomal RNAs. *Enzymes* 41: 169–213
- Hirose T, Ideue T, Nagai M, Hagiwara M, Shu M-D & Steitz JA (2006) A Spliceosomal Intron Binding Protein, IBP160, Links Position-Dependent Assembly of Intron-Encoded Box C/D snoRNP to Pre-mRNA Splicing. *Molecular Cell* 23: 673–684
- Hirose T & Steitz JA (2001) Position within the host intron is critical for efficient processing of box C/D snoRNAs in mammalian cells. *Proc Natl Acad Sci U S A* 98: 12914–12919
- Ho NW & Gilham PT (1971) Reaction of pseudouridine and inosine with N-cyclohexyl-N'-beta-(4-methylmorpholinium)ethylcarbodiimide. *Biochemistry* 10: 3651–3657
- Hoang C & Ferré-D'Amaré AR (2001) Cocystal structure of a tRNA Psi55 pseudouridine synthase: nucleotide flipping by an RNA-modifying enzyme. *Cell* 107: 929–939
- Hoernes TP, Heimdörfer D, Köstner D, Faserl K, Nußbaumer F, Plangger R, Kreutz C, Lindner H & Erlacher MD (2019) Eukaryotic Translation Elongation is Modulated by Single Natural Nucleotide Derivatives in the Coding Sequences of mRNAs. *Genes (Basel)* 10: 84
- Höfler S & Carlomagno T (2020) Structural and functional roles of 2'-O-ribose methylations and their enzymatic machinery across multiple classes of RNAs. *Current Opinion in Structural Biology* 65: 42–50
- Huang C, Shi J, Guo Y, Huang W, Huang S, Ming S, Wu X, Zhang R, Ding J, Zhao W, *et al* (2017) A snoRNA modulates mRNA 3' end processing and regulates the expression of a subset of mRNAs. *Nucleic Acids Res* 45: 8647–8660

- Huang L, Pookanjanatavip M, Gu X & Santi DV (1998) A Conserved Aspartate of tRNA Pseudouridine Synthase Is Essential for Activity and a Probable Nucleophilic Catalyst. *Biochemistry* 37: 344–351
- Huang S (2002) Building an efficient factory: where is pre-rRNA synthesized in the nucleolus? *J Cell Biol* 157: 739–741
- Hüttenhofer A, Kiefmann M, Meier-Ewert S, O'Brien J, Lehrach H, Bachellerie J-P & Brosius J (2001) RNomics: an experimental approach that identifies 201 candidates for novel, small, non-messenger RNAs in mouse. *The EMBO Journal* 20: 2943–2953
- Hutzing R, Feederle R, Mrazek J, Schiefermeier N, Balwierz PJ, Zavolan M, Polacek N, Delecluse H-J & Hüttenhofer A (2009) Expression and processing of a small nucleolar RNA from the Epstein-Barr virus genome. *PLoS Pathog* 5: e1000547
- Incarnato D, Anselmi F, Morandi E, Neri F, Maldotti M, Rapelli S, Parlato C, Basile G & Oliviero S (2017) High-throughput single-base resolution mapping of RNA 2'-O-methylated residues. *Nucleic Acids Res* 45: 1433–1441
- van Ingen E, van den Homberg DAL, van der Bent ML, Mei H, Papac-Milicevic N, Kremer V, Boon RA, Quax PHA, Wojta J & Nossent AY (2022) C/D box snoRNA SNORD113-6/AF357425 plays a dual role in integrin signalling and arterial fibroblast function via pre-mRNA processing and 2'-O-ribose methylation. *Hum Mol Genet* 31: 1051–1066
- Isaac C, Yang Y & Meier UT (1998) Nopp140 Functions as a Molecular Link Between the Nucleolus and the Coiled Bodies. *The Journal of Cell Biology* 142: 319
- Ismael H, Altmeyer S & Stahl H (2016) Regulation of the U3-, U8-, and U13snoRNA Expression by the DEAD Box Proteins Ddx5/Ddx17 with Consequences for Cell Proliferation and Survival. *Noncoding RNA* 2: 11
- Ito S, Horikawa S, Suzuki T, Kawauchi H, Tanaka Y, Suzuki T & Suzuki T (2014) Human NAT10 Is an ATP-dependent RNA Acetyltransferase Responsible for N4-Acetylcytidine Formation in 18 S Ribosomal RNA (rRNA). *J Biol Chem* 289: 35724–35730
- Ivanov P, Emara MM, Villen J, Gygi SP & Anderson P (2011) Angiogenin-induced tRNA fragments inhibit translation initiation. *Mol Cell* 43: 613–623
- Jády BE, Bertrand E & Kiss T (2004) Human telomerase RNA and box H/ACA scaRNAs share a common Cajal body-specific localization signal. *Journal of Cell Biology* 164: 647–652
- Jády BE, Darzacq X, Tucker KE, Matera AG, Bertrand E & Kiss T (2003) Modification of Sm small nuclear RNAs occurs in the nucleoplasmic Cajal body following import from the cytoplasm. *The EMBO Journal* 22: 1878–1888
- Jády BE, Ketele A & Kiss T (2012) Human intron-encoded Alu RNAs are processed and packaged into Wdr79-associated nucleoplasmic box H/ACA RNPs. *Genes Dev* 26: 1897–1910

- Jády BE & Kiss T (2001) A small nucleolar guide RNA functions both in 2'-O-ribose methylation and pseudouridylation of the U5 spliceosomal RNA. *EMBO J* 20: 541–551
- Jenkinson EM, Rodero MP, Kasher PR, Ugenti C, Oojageer A, Goosey LC, Rose Y, Kershaw CJ, Urquhart JE, Williams SG, *et al* (2016) Mutations in SNORD118 cause the cerebral microangiopathy leukoencephalopathy with calcifications and cysts. *Nat Genet* 48: 1185–1192
- Ji X (2008) The Mechanism of RNase III Action: How Dicer Dices. In *RNA Interference*, Paddison PJ & Vogt PK (eds) pp 99–116. Berlin, Heidelberg: Springer
- Jin H, Loria JP & Moore PB (2007) Solution structure of an rRNA substrate bound to the pseudouridylation pocket of a box H/ACA snoRNA. *Mol Cell* 26: 205–215
- Jorjani H, Kehr S, Jedlinski DJ, Gumienny R, Hertel J, Stadler PF, Zavolan M & Gruber AR (2016) An updated human snoRNAome. *Nucleic Acids Res* 44: 5068–5082
- Joshi AR, Kulkarni K & Shah AU (2021) Labrune Syndrome: A Very Rare Association of Leukoencephalopathy, Cerebral Calcifications, and Cysts. *Indian J Radiol Imaging* 31: 772–775
- Junge A, Refsgaard JC, Garde C, Pan X, Santos A, Alkan F, Anthon C, von Mering C, Workman CT, Jensen LJ, *et al* (2017) RAIN: RNA–protein Association and Interaction Networks. *Database (Oxford)* 2017: baw167
- Karijolic J, Kantartzis A & Yu Y-T (2010) RNA modifications: a mechanism that modulates gene expression. *Methods Mol Biol* 629: 1–19
- Kehr S, Bartschat S, Stadler PF & Tafer H (2011) PLEXY: efficient target prediction for box C/D snoRNAs. *Bioinformatics* 27: 279–280
- Ketele A, Kiss T & Jády BE (2016) Human intron-encoded AluACA RNAs and telomerase RNA share a common element promoting RNA accumulation. *RNA Biol* 13: 1274–1285
- Khoddami V, Yerra A, Mosbrugger TL, Fleming AM, Burrows CJ & Cairns BR (2019a) Transcriptome-wide profiling of multiple RNA modifications simultaneously at single-base resolution. *Proceedings of the National Academy of Science* 116: 6784–6789
- Khoddami V, Yerra A, Mosbrugger TL, Fleming AM, Burrows CJ & Cairns BR (2019b) Transcriptome-wide profiling of multiple RNA modifications simultaneously at single-base resolution. *Proc Natl Acad Sci U S A* 116: 6784–6789
- Kim D-S, Camacho CV, Nagari A, Malladi VS, Challa S & Kraus WL (2019) Activation of PARP-1 by snoRNAs Controls Ribosome Biogenesis and Cell Growth via the RNA Helicase DDX21. *Mol Cell* 75: 1270-1285.e14
- Kim SH, MacFarlane S, Kalinina NO, Rakitina DV, Ryabov EV, Gillespie T, Haupt S, Brown JWS & Taliany M (2007) Interaction of a plant virus-encoded protein with the major nucleolar protein fibrillarin is required for systemic virus infection. *Proceedings of the National Academy of Sciences* 104: 11115–11120

- Kim SH, Spensley M, Choi SK, Calixto CPG, Pendle AF, Koroleva O, Shaw PJ & Brown JWS (2010) Plant U13 orthologues and orphan snoRNAs identified by RNomics of RNA from *Arabidopsis nucleoli*. *Nucleic Acids Res* 38: 3054–3067
- Kirpekar F, Hansen LH, Rasmussen A, Poehlsgaard J & Vester B (2005) The Archaeon *Haloarcula marismortui* has Few Modifications in the Central Parts of its 23S Ribosomal RNA. *Journal of Molecular Biology* 348: 563–573
- Kishore S, Gruber AR, Jedlinski DJ, Syed AP, Jorjani H & Zavolan M (2013) Insights into snoRNA biogenesis and processing from PAR-CLIP of snoRNA core proteins and small RNA sequencing. *Genome Biol* 14: R45
- Kishore S & Stamm S (2006) The snoRNA HBII-52 Regulates Alternative Splicing of the Serotonin Receptor 2C. *Science* 311: 230–232
- Kiss AM, Jády BE, Bertrand E & Kiss T (2004) Human box H/ACA pseudouridylation guide RNA machinery. *Mol Cell Biol* 24: 5797–5807
- Kiss T (2002) Small Nucleolar RNAs: An Abundant Group of Noncoding RNAs with Diverse Cellular Functions. *Cell* 109: 145–148
- Kiss T (2004) Biogenesis of small nuclear RNPs. *J Cell Sci* 117: 5949–5951
- Kiss T, Fayet-Lebaron E & Jády BE (2010) Box H/ACA small ribonucleoproteins. *Mol Cell* 37: 597–606
- Kiss T & Jády BE (2004) Functional characterization of 2'-O-methylation and pseudouridylation guide RNAs. *Methods Mol Biol* 265: 393–408
- Kiss-László Z, Henry Y, Bachellerie JP, Caizergues-Ferrer M & Kiss T (1996) Site-specific ribose methylation of preribosomal RNA: a novel function for small nucleolar RNAs. *Cell* 85: 1077–1088
- Kiss-László Z, Henry Y & Kiss T (1998) Sequence and structural elements of methylation guide snoRNAs essential for site-specific ribose methylation of pre-rRNA. *EMBO J* 17: 797–807
- Klein D j., Schmeing T m., Moore P b. & Steitz T a. (2001) The kink-turn: a new RNA secondary structure motif. *The EMBO Journal* 20: 4214–4221
- Koberna K, Malínský J, Pliss A, Mašata M, Večeřová J, Fialová M, Bednár J & Raška I (2002) Ribosomal genes in focus. *J Cell Biol* 157: 743–748
- Koonin EV (1996) Pseudouridine synthases: four families of enzymes containing a putative uridine-binding motif also conserved in dUTPases and dCTP deaminases. *Nucleic Acids Res* 24: 2411–2415
- Kowalak JA, Pomerantz SC, Crain PF & McCloskey JA (1993) A novel method for the determination of post-transcriptional modification in RNA by mass spectrometry. *Nucleic Acids Res* 21: 4577–4585

- Krogh N, Jansson MD, Häfner SJ, Tehler D, Birkedal U, Christensen-Dalsgaard M, Lund AH & Nielsen H (2016) Profiling of 2'-O-Me in human rRNA reveals a subset of fractionally modified positions and provides evidence for ribosome heterogeneity. *Nucleic Acids Res* 44: 7884–7895
- Krogh N & Nielsen H (2019) Sequencing-based methods for detection and quantitation of ribose methylations in RNA. *Methods* 156: 5–15
- Kudla G, Granneman S, Hahn D, Beggs JD & Tollervey D (2011) Cross-linking, ligation, and sequencing of hybrids reveals RNA–RNA interactions in yeast. *Proceedings of the National Academy of Sciences* 108: 10010–10015
- Kufel J, Allmang C, Chanfreau G, Petfalski E, Lafontaine DLJ & Tollervey D (2000) Precursors to the U3 Small Nucleolar RNA Lack Small Nucleolar RNP Proteins but Are Stabilized by La Binding. *Mol Cell Biol* 20: 5415–5424
- Labrune P, Lacroix C, Goutières F, de Laveaucoupet J, Chevalier P, Zerah M, Husson B & Landrieu P (1996) Extensive brain calcifications, leukodystrophy, and formation of parenchymal cysts: a new progressive disorder due to diffuse cerebral microangiopathy. *Neurology* 46: 1297–1301
- Lafontaine DLJ (2015) Noncoding RNAs in eukaryotic ribosome biogenesis and function. *Nat Struct Mol Biol* 22: 11–19
- Lafontaine DLJ, Riback JA, Bascetin R & Brangwynne CP (2021) The nucleolus as a multiphase liquid condensate. *Nat Rev Mol Cell Biol* 22: 165–182
- Langhendries J-L, Nicolas E, Doumont G, Goldman S & Lafontaine DLJ (2016) The human box C/D snoRNAs U3 and U8 are required for pre-rRNA processing and tumorigenesis. *Oncotarget* 7: 59519–59534
- Lechertier T, Grob A, Hernandez-Verdun D & Roussel P (2009) Fibrillarin and Nop56 interact before being co-assembled in box C/D snoRNPs. *Experimental Cell Research* 315: 928–942
- Lestrade L & Weber MJ (2006) snoRNA-LBME-db, a comprehensive database of human H/ACA and C/D box snoRNAs. *Nucleic Acids Res* 34: D158-162
- Leulliot N, Godin KS, Hoareau-Aveilla C, Quevillon-Cheruel S, Varani G, Henry Y & Van Tilbeurgh H (2007) The box H/ACA RNP assembly factor Naf1p contains a domain homologous to Gar1p mediating its interaction with Cbf5p. *J Mol Biol* 371: 1338–1353
- Levitskii SA, Mukhar'iamova Ks, Veiko VP & Zatsepina OV (2004) [Identification of signal sequences determining the specific nucleolar localization of fibrillarin in HeLa cells. *Mol Biol (Mosk)* 38: 483–492
- Li HD, Zagorski J & Fournier MJ (1990) Depletion of U14 small nuclear RNA (snR128) disrupts production of 18S rRNA in *Saccharomyces cerevisiae*. *Mol Cell Biol* 10: 1145–1152
- Li L & Ye K (2006) Crystal structure of an H/ACA box ribonucleoprotein particle. *Nature* 443: 302–307

- Li S, Duan J, Li D, Yang B, Dong M & Ye K (2011a) Reconstitution and structural analysis of the yeast box H/ACA RNA-guided pseudouridine synthase. *Genes Dev* 25: 2409–2421
- Li S, Duan J, Li D, Yang B, Dong M & Ye K (2011b) Reconstitution and structural analysis of the yeast box H/ACA RNA-guided pseudouridine synthase. *Genes Dev* 25: 2409–2421
- Li X, Xiong X & Yi C (2017) Epitranscriptome sequencing technologies: decoding RNA modifications. *Nat Methods* 14: 23–31
- Li X, Zhu P, Ma S, Song J, Bai J, Sun F & Yi C (2015) Chemical pulldown reveals dynamic pseudouridylation of the mammalian transcriptome. *Nat Chem Biol* 11: 592–597
- Liang B, Zhou J, Kahen E, Terns RM, Terns MP & Li H (2009) Structure of a functional ribonucleoprotein pseudouridine synthase bound to a substrate RNA. *Nature structural & molecular biology* 16: 740
- Liao J, Yu L, Mei Y, Guarnera M, Shen J, Li R, Liu Z & Jiang F (2010) Small nucleolar RNA signatures as biomarkers for non-small-cell lung cancer. *Molecular Cancer* 9: 198
- Lin J, Lai S, Jia R, Xu A, Zhang L, Lu J & Ye K (2011) Structural basis for site-specific ribose methylation by box C/D RNA protein complexes. *Nature* 469: 559–563
- Liu Y, Dou M, Song X, Dong Y, Liu S, Liu H, Tao J, Li W, Yin X & Xu W (2019) The emerging role of the piRNA/piwi complex in cancer. *Mol Cancer* 18: 123
- Livingston JH, Mayer J, Jenkinson E, Kasher P, Stivaros S, Berger A, Cordelli DM, Ferreira P, Jefferson R, Kutschke G, *et al* (2014) Leukoencephalopathy with calcifications and cysts: a purely neurological disorder distinct from coats plus. *Neuropediatrics* 45: 175–182
- Lovejoy AF, Riordan DP & Brown PO (2014a) Transcriptome-wide mapping of pseudouridines: pseudouridine synthases modify specific mRNAs in *S. cerevisiae*. *PLoS One* 9: e110799
- Lovejoy AF, Riordan DP & Brown PO (2014b) Transcriptome-Wide Mapping of Pseudouridines: Pseudouridine Synthases Modify Specific mRNAs in *S. cerevisiae*. *PLoS One* 9: e110799
- Lu B, Chen X, Liu X, Chen J, Qin H, Chen S & Zhao Y (2022) C/D box small nucleolar RNA SNORD104 promotes endometrial cancer by regulating the 2'-O-methylation of PARP1. *J Transl Med* 20: 618
- Lu Z, Zhang QC, Lee B, Flynn RA, Smith MA, Robinson JT, Davidovich C, Gooding AR, Goodrich KJ, Mattick JS, *et al* (2016) RNA duplex map in living cells reveals higher order transcriptome structure. *Cell* 165: 1267–1279
- Lubas M, Andersen PR, Schein A, Dziembowski A, Kudla G & Jensen TH (2015) The human nuclear exosome targeting complex is loaded onto newly synthesized RNA to direct early ribonucleolysis. *Cell Rep* 10: 178–192
- Lukowiak AA, Granneman S, Mattox SA, Speckmann WA, Jones K, Pluk H, van Venrooij WJ, Terns RM & Terns MP (2000) Interaction of the U3-55k protein with U3 snoRNA is

- mediated by the Box B/C motif of U3 and the WD repeats of U3-55k. *Nucleic Acids Res* 28: 3462–3471
- Lykke-Andersen S, Ardal BK, Hollensen AK, Damgaard CK & Jensen TH (2018a) Box C/D snoRNP Autoregulation by a cis-Acting snoRNA in the NOP56 Pre-mRNA. *Mol Cell* 72: 99-111.e5
- Lykke-Andersen S, Ardal BK, Hollensen AK, Damgaard CK & Jensen TH (2018b) Box C/D snoRNP Autoregulation by a cis-Acting snoRNA in the NOP56 Pre-mRNA. *Molecular Cell* 72: 99-111.e5
- Lykke-Andersen S, Chen Y, Ardal BR, Lilje B, Waage J, Sandelin A & Jensen TH (2014) Human nonsense-mediated RNA decay initiates widely by endonucleolysis and targets snoRNA host genes. *Genes Dev* 28: 2498–2517
- Machado-Pinilla R, Liger D, Leulliot N & Meier UT (2012) Mechanism of the AAA+ ATPases pontin and reptin in the biogenesis of H/ACA RNPs. *RNA* 18: 1833–1845
- Machyna M, Kehr S, Straube K, Kappei D, Buchholz F, Butter F, Ule J, Hertel J, Stadler PF & Neugebauer KM (2014) The coilin interactome identifies hundreds of small noncoding RNAs that traffic through Cajal bodies. *Mol Cell* 56: 389–399
- Maden BEH (2001) Mapping 2'-O-Methyl Groups in Ribosomal RNA. *Methods* 25: 374–382
- Maden BEH, Corbett ME, Heeney PA, Pugh K & Ajuh PM (1995) Classical and novel approaches to the detection and localization of the numerous modified nucleotides in eukaryotic ribosomal RNA. *Biochimie* 77: 22–29
- Majumder M, Mukhopadhyay S, Kharel P & Gupta R (2020) The presence of the ACA box in archaeal H/ACA guide RNAs promotes atypical pseudouridylation. *RNA* 26: 396–418
- Makker V, Taylor MH, Aghajanian C, Oaknin A, Mier J, Cohn AL, Romeo M, Bratos R, Brose MS, DiSimone C, *et al* (2020) Lenvatinib Plus Pembrolizumab in Patients With Advanced Endometrial Cancer. *J Clin Oncol* 38: 2981–2992
- Mannoor K, Shen J, Liao J, Liu Z & Jiang F (2014) Small nucleolar RNA signatures of lung tumor-initiating cells. *Mol Cancer* 13: 104
- Marchand V, Pichot F, Neybecker P, Ayadi L, Bourguignon-Igel V, Wacheul L, Lafontaine DLJ, Pinzano A, Helm M & Motorin Y (2020) HydraPsiSeq: a method for systematic and quantitative mapping of pseudouridines in RNA. *Nucleic Acids Res* 48: e110
- Marmier-Gourrier N, Cléry A, Senty-Ségault V, Charpentier B, Schlotter F, Leclerc F, Fournier R & Branlant C (2003) A structural, phylogenetic, and functional study of 15.5-kD/Snu13 protein binding on U3 small nucleolar RNA. *RNA* 9: 821–838
- Massenet S, Bertrand E & Verheggen C (2016a) Assembly and trafficking of box C/D and H/ACA snoRNPs. *RNA Biol* 14: 680–692
- Massenet S, Bertrand E & Verheggen C (2016b) Assembly and trafficking of box C/D and H/ACA snoRNPs. *RNA Biol* 14: 680–692

- Maurizy C, Quinternet M, Abel Y, Verheggen C, Santo PE, Bourguet M, C F Paiva A, Bragantini B, Chagot M-E, Robert M-C, *et al* (2018) The RPAP3-Cterminal domain identifies R2TP-like quaternary chaperones. *Nat Commun* 9: 2093
- McBride AE & Silver PA (2001) State of the arg: protein methylation at arginine comes of age. *Cell* 106: 5–8
- McFadden EJ & Baserga SJ U8 variants on the brain: a small nucleolar RNA and human disease. *RNA Biol* 19: 412–418
- Mei Y-P, Liao J-P, Shen J, Yu L, Liu B-L, Liu L, Li R-Y, Ji L, Dorsey SG, Jiang Z-R, *et al* (2012) Small nucleolar RNA 42 acts as an oncogene in lung tumorigenesis. *Oncogene* 31: 2794–2804
- Meier UT (2005) The many facets of H/ACA ribonucleoproteins. *Chromosoma* 114: 1–14
- Meier UT & Blobel G (1990) A nuclear localization signal binding protein in the nucleolus. *J Cell Biol* 111: 2235–2245
- Meier UT & Blobel G (1992) Nopp140 shuttles on tracks between nucleolus and cytoplasm. *Cell* 70: 127–138
- Meier UT & Blobel G (1994) NAP57, a mammalian nucleolar protein with a putative homolog in yeast and bacteria. *J Cell Biol* 127: 1505–1514
- Mengel-Jørgensen J & Kirpekar F (2002) Detection of pseudouridine and other modifications in tRNA by cyanoethylation and MALDI mass spectrometry. *Nucleic Acids Res* 30: e135
- Michot B, Joseph N, Mazan S & Bachellerie JP (1999) Evolutionarily conserved structural features in the ITS2 of mammalian pre-rRNAs and potential interactions with the snoRNA U8 detected by comparative analysis of new mouse sequences. *Nucleic Acids Res* 27: 2271–2282
- Mitchell JR, Cheng J & Collins K (1999a) A box H/ACA small nucleolar RNA-like domain at the human telomerase RNA 3' end. *Mol Cell Biol* 19: 567–576
- Mitchell JR, Cheng J & Collins K (1999b) A Box H/ACA Small Nucleolar RNA-Like Domain at the Human Telomerase RNA 3' End. *Mol Cell Biol* 19: 567–576
- Mitchell JR, Wood E & Collins K (1999c) A telomerase component is defective in the human disease dyskeratosis congenita. *Nature* 402: 551–555
- Moore T, Zhang Y, Fenley MO & Li H (2004) Molecular Basis of Box C/D RNA-Protein Interactions: Cocystal Structure of Archaeal L7Ae and a Box C/D RNA. *Structure* 12: 807–818
- Morris GE (2008) The Cajal body. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* 1783: 2108–2115
- Motorin Y & Helm M (2011) RNA nucleotide methylation. *WIREs RNA* 2: 611–631

- Motorin Y & Marchand V (2018) Detection and Analysis of RNA Ribose 2'-O-Methylations: Challenges and Solutions. *Genes (Basel)* 9: 642
- Myslinski E, Ségault V & Branlant C (1990) An intron in the genes for U3 small nucleolar RNAs of the yeast *Saccharomyces cerevisiae*. *Science* 247: 1213–1216
- Narayanan A, Lukowiak A, Jády BE, Dragon F, Kiss T, Terns RM & Terns MP (1999) Nucleolar localization signals of box H/ACA small nucleolar RNAs. *EMBO J* 18: 5120–5130
- Newton K, Petfalski E, Tollervey D & Cáceres JF (2003) Fibrillarin is essential for early development and required for accumulation of an intron-encoded small nucleolar RNA in the mouse. *Mol Cell Biol* 23: 8519–8527
- Nguyen TC, Cao X, Yu P, Xiao S, Lu J, Biase FH, Sridhar B, Huang N, Zhang K & Zhong S (2016) Mapping RNA–RNA interactome and RNA structure in vivo by MARIO. *Nat Commun* 7: 12023
- Nguyen THD, Tam J, Wu RA, Greber BJ, Toso D, Nogales E & Collins K (2018) Cryo-EM structure of substrate-bound human telomerase holoenzyme. *Nature* 557: 190–195
- Ofengand J, Del Campo M & Kaya Y (2001) Mapping Pseudouridines in RNA Molecules. *Methods* 25: 365–373
- Ojha S, Malla S & Lyons SM (2020) snoRNPs: Functions in Ribosome Biogenesis. *Biomolecules* 10: 783
- Ooi SL, Samarsky DA, Fournier MJ & Boeke JD (1998) Intronic snoRNA biosynthesis in *Saccharomyces cerevisiae* depends on the lariat-debranching enzyme: intron length effects and activity of a precursor snoRNA. *RNA* 4: 1096–1110
- Orphanos V, McGown G, Hey Y, Boyle JM & Santibanez-Koref M (1995) Proximal 6q, a region showing allele loss in primary breast cancer. *Br J Cancer* 71: 290–293
- d'Orval BC, Bortolin M-L, Gaspin C & Bachellerie J-P (2001) Box C/D RNA guides for the ribose methylation of archaeal tRNAs. The tRNA^{Trp} intron guides the formation of two ribose-methylated nucleosides in the mature tRNA^{Trp}. *Nucleic Acids Research* 29: 4518–4529
- Pace M, Falappa M, Freschi A, Balzani E, Berteotti C, Lo Martire V, Kaveh F, Hovig E, Zoccoli G, Amici R, *et al* (2020) Loss of Snord116 impacts lateral hypothalamus, sleep, and food-related behaviors. *JCI Insight* 5: e137495
- Pacilli A, Ceccarelli C, Tréré D & Montanaro L (2013) SnoRNA U50 levels are regulated by cell proliferation and rRNA transcription. *Int J Mol Sci* 14: 14923–14935
- Palacios JM, Pazos A & Hoyer D (2017) A short history of the 5-HT_{2C} receptor: from the choroid plexus to depression, obesity and addiction treatment. *Psychopharmacology* 234: 1395–1418

- Patteson KG, Rodicio LP & Limbach PA (2001) Identification of the mass-silent post-transcriptionally modified nucleoside pseudouridine in RNA by matrix-assisted laser desorption/ionization mass spectrometry. *Nucleic Acids Res* 29: e49
- Paul A, Tiotiu D, Bragantini B, Marty H, Charpentier B, Massenet S & Labialle S (2019) Bcd1p controls RNA loading of the core protein Nop58 during C/D box snoRNP biogenesis. *RNA* 25: 496–506
- Peattie DA (1979) Direct chemical method for sequencing RNA. *Proc Natl Acad Sci U S A* 76: 1760–1764
- Peculis BA (1997) The sequence of the 5' end of the U8 small nucleolar RNA is critical for 5.8S and 28S rRNA maturation. *Mol Cell Biol* 17: 3702–3713
- Pederson T (2011) The Nucleolus. *Cold Spring Harb Perspect Biol* 3: a000638
- Penzo M, Casoli L, Ceccarelli C, Treré D, Ludovini V, Crinò L & Montanaro L (2013) DKC1 gene mutations in human sporadic cancer. *Histol Histopathol* 28: 365–372
- Phipps KR, Charette JM & Baserga SJ (2011a) The small subunit processome in ribosome biogenesis—progress and prospects. *WIREs RNA* 2: 1–21
- Phipps KR, Charette JM & Baserga SJ (2011b) The SSU Processome in Ribosome Biogenesis – Progress and Prospects. *Wiley Interdiscip Rev RNA* 2: 1–21
- Plumbridge JA, Bäumer HG, Ehrenberg M & Rigler R (1980) Characterisation of a new, fully active fluorescent derivative of E. coli tRNA Phe. *Nucleic Acids Res* 8: 827–843
- Polex-Wolf J, Lam BY, Larder R, Tadross J, Rimmington D, Bosch F, Cenzano VJ, Ayuso E, Ma MK, Rainbow K, *et al* (2018) Hypothalamic loss of Snord116 recapitulates the hyperphagia of Prader-Willi syndrome. *J Clin Invest* 128: 960–969
- Ponti D, Troiano M, Bellenchi GC, Battaglia PA & Gigliani F (2008) The HIV Tat protein affects processing of ribosomal RNA precursor. *BMC Cell Biol* 9: 32
- Pradet-Balade B, Girard C, Boulon S, Paul C, Azzag K, Bordonné R, Bertrand E & Verheggen C (2011a) CRM1 controls the composition of nucleoplasmic pre-snoRNA complexes to licence them for nucleolar transport. *EMBO J* 30: 2205–2218
- Pradet-Balade B, Girard C, Boulon S, Paul C, Azzag K, Bordonné R, Bertrand E & Verheggen C (2011b) CRM1 controls the composition of nucleoplasmic pre-snoRNA complexes to licence them for nucleolar transport. *EMBO J* 30: 2205–2218
- Qian X, Xu C, Wu B, Tang H, Zhao P & Qi Z (2020) SNORD126 Promotes Hepatitis C Virus Infection by Upregulating Claudin-1 via Activation of PI3K-AKT Signaling Pathway. *Front Microbiol* 11: 565590
- Qu G, van Nues RW, Watkins NJ & Maxwell ES (2011) The Spatial-Functional Coupling of Box C/D and C'/D' RNPs Is an Evolutionarily Conserved Feature of the Eukaryotic Box C/D snoRNP Nucleotide Modification Complex. *Mol Cell Biol* 31: 365–374

- Raabe CA, Voss R, Kummerfeld D-M, Brosius J, Galiveti CR, Wolters A, Seggewiss J, Huge A, Skryabin BV & Rozhdestvensky TS (2019) Ectopic expression of Snord115 in choroid plexus interferes with editing but not splicing of 5-Ht2c receptor pre-mRNA in mice. *Sci Rep* 9: 4300
- Rakitina DV, Taliansky M, Brown JWS & Kalinina NO (2011a) Two RNA-binding sites in plant fibrillarin provide interactions with various RNA substrates. *Nucleic Acids Research* 39: 8869–8880
- Rakitina DV, Taliansky M, Brown JWS & Kalinina NO (2011b) Two RNA-binding sites in plant fibrillarin provide interactions with various RNA substrates. *Nucleic Acids Research* 39: 8869–8880
- Ramani V, Qiu R & Shendure J (2015) High-throughput determination of RNA structure by proximity ligation. *Nat Biotechnol* 33: 980–984
- Rashid R, Liang B, Baker DL, Youssef OA, He Y, Phipps K, Terns RM, Terns MP & Li H (2006) Crystal Structure of a Cbf5-Nop10-Gar1 Complex and Implications in RNA-Guided Pseudouridylation and Dyskeratosis Congenita. *Molecular Cell* 21: 249–260
- Rebane A, Roomere H & Metspalu A (2002) Locations of several novel 2'-O-methylated nucleotides in human 28S rRNA. *BMC Mol Biol* 3: 1
- Reddy R, Henning D & Busch H (1985) Primary and secondary structure of U8 small nuclear RNA. *J Biol Chem* 260: 10930–10935
- RENALIER M-H, JOSEPH N, GASPIN C, THEBAULT P & MOUGIN A (2005) The Cm56 tRNA modification in archaea is catalyzed either by a specific 2'-O-methylase, or a C/D sRNP. *RNA* 11: 1051–1063
- Richard P, Kiss AM, Darzacq X & Kiss T (2006) Cotranscriptional recognition of human intronic box H/ACA snoRNAs occurs in a splicing-independent manner. *Mol Cell Biol* 26: 2540–2549
- Rodriguez-Corona U, Sobol M, Rodriguez-Zapata LC, Hozak P & Castano E (2015) Fibrillarin from Archaea to human. *Biology of the Cell* 107: 159–174
- Ross R, Cao X, Yu N & Limbach PA (2016) Sequence mapping of transfer RNA chemical modifications by liquid chromatography tandem mass spectrometry. *Methods* 107: 73–78
- Rozenski J, Crain PF & McCloskey JA (1999) The RNA Modification Database: 1999 update. *Nucleic Acids Res* 27: 196–197
- Runte M, Varon R, Horn D, Horsthemke B & Buiting K (2005) Exclusion of the C/D box snoRNA gene cluster HBII-52 from a major role in Prader–Willi syndrome. *Hum Genet* 116: 228–230
- Russell AG, Schnare MN & Gray MW (2004) Pseudouridine-guide RNAs and other Cbf5p-associated RNAs in *Euglena gracilis*. *RNA* 10: 1034–1046

- Sahoo T, del Gaudio D, German JR, Shinawi M, Peters SU, Person RE, Garnica A, Cheung SW & Beaudet AL (2008) Prader-Willi phenotype caused by paternal deficiency for the HBII-85 C/D box small nucleolar RNA cluster. *Nat Genet* 40: 719–721
- Saliminejad K, Khorram Khorshid HR, Soleymani Fard S & Ghaffari SH (2019) An overview of microRNAs: Biology, functions, therapeutics, and analysis methods. *Journal of Cellular Physiology* 234: 5451–5465
- Santos D, Feng M, Kolliopoulou A, Taning CNT, Sun J & Swevers L (2023) What Are the Functional Roles of Piwi Proteins and piRNAs in Insects? *Insects* 14: 187
- Saxena T, Tandon B, Sharma S, Chameettachal S, Ray P, Ray AR & Kulshreshtha R (2013) Combined miRNA and mRNA signature identifies key molecular players and pathways involved in chikungunya virus infection in human cells. *PLoS One* 8: e79886
- Schattner P, Decatur WA, Davis CA, Ares M, Fournier MJ & Lowe TM (2004a) Genome-wide searching for pseudouridylation guide snoRNAs: analysis of the *Saccharomyces cerevisiae* genome. *Nucleic Acids Res* 32: 4281–4296
- Schattner P, Decatur WA, Davis CA, Ares M, Fournier MJ & Lowe TM (2004b) Genome-wide searching for pseudouridylation guide snoRNAs: analysis of the *Saccharomyces cerevisiae* genome. *Nucleic Acids Res* 32: 4281–4296
- Schopf FH, Biebl MM & Buchner J (2017) The HSP90 chaperone machinery. *Nat Rev Mol Cell Biol* 18: 345–360
- Schubert T, Pusch MC, Diermeier S, Benes V, Kremmer E, Imhof A & Längst G (2012) Df31 Protein and snoRNAs Maintain Accessible Higher-Order Structures of Chromatin. *Molecular Cell* 48: 434–444
- Schwartz S, Bernstein DA, Mumbach MR, Jovanovic M, Herbst RH, León-Ricardo BX, Engreitz JM, Guttman M, Satija R, Lander ES, *et al* (2014) Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. *Cell* 159: 148–162
- Sharma E, Sterne-Weiler T, O’Hanlon D & Blencowe BJ (2016) Global Mapping of Human RNA-RNA Interactions. *Molecular Cell* 62: 618–626
- Sharma S, Langhendries J-L, Watzinger P, Kötter P, Entian K-D & Lafontaine DLJ (2015) Yeast Kre33 and human NAT10 are conserved 18S rRNA cytosine acetyltransferases that modify tRNAs assisted by the adaptor Tan1/THUMPD1. *Nucleic Acids Res* 43: 2242–2258
- Sharma S, Yang J, van Nues R, Watzinger P, Kötter P, Lafontaine DLJ, Granneman S & Entian K-D (2017) Specialized box C/D snoRNPs act as antisense guides to target RNA base acetylation. *PLoS Genet* 13: e1006804
- de Smith AJ, Purmann C, Walters RG, Ellis RJ, Holder SE, Van Haelst MM, Brady AF, Fairbrother UL, Dattani M, Keogh JM, *et al* (2009) A deletion of the HBII-85 class of small nucleolar RNAs (snoRNAs) is associated with hyperphagia, obesity and hypogonadism. *Hum Mol Genet* 18: 3257–3265

- Smith AM, Abu-Shumays R, Akeson M & Bernick DL (2015) Capture, Unfolding, and Detection of Individual tRNA Molecules Using a Nanopore Device. *Front Bioeng Biotechnol* 3: 91
- Snaar S, Wiesmeijer K, Jochemsen AG, Tanke HJ & Dirks RW (2000) Mutational analysis of fibrillarin and its mobility in living human cells. *J Cell Biol* 151: 653–662
- Srinivasan S, Torres AG & Ribas de Pouplana L (2021) Inosine in Biology and Disease. *Genes (Basel)* 12: 600
- Srivastava L, Lapik YR, Wang M & Pestov DG (2010) Mammalian DEAD box protein Ddx51 acts in 3' end maturation of 28S rRNA by promoting the release of U8 snoRNA. *Mol Cell Biol* 30: 2947–2956
- Studniarek C, Tellier M, Martin PGP, Murphy S, Kiss T & Egloff S (2021) The 7SK/P-TEFb snRNP controls ultraviolet radiation-induced transcriptional reprogramming. *Cell Rep* 35: 108965
- Sugimoto Y, Vigilante A, Darbo E, Zirra A, Militti C, D'Ambrogio A, Luscombe NM & Ule J (2015) hiCLIP reveals the in vivo atlas of mRNA secondary structures recognized by Staufen 1. *Nature* 519: 491–494
- Sun Q, Zhu X, Qi J, An W, Lan P, Tan D, Chen R, Wang B, Zheng S, Zhang C, *et al* (2017) Molecular architecture of the 90S small subunit pre-ribosome. *eLife* 6: e22086
- Sur S, Steele R, Shi X & Ray RB (2019) miRNA-29b Inhibits Prostate Tumor Growth and Induces Apoptosis by Increasing Bim Expression. *Cells* 8: 1455
- Tafer H, Kehr S, Hertel J, Hofacker IL & Stadler PF (2010a) RNAsnoop: efficient target prediction for H/ACA snoRNAs. *Bioinformatics* 26: 610–616
- Tafer H, Kehr S, Hertel J, Hofacker IL & Stadler PF (2010b) RNAsnoop: efficient target prediction for H/ACA snoRNAs. *Bioinformatics* 26: 610–616
- Taft RJ, Glazov EA, Lassmann T, Hayashizaki Y, Carninci P & Mattick JS (2009) Small RNAs derived from snoRNAs. *RNA* 15: 1233–1240
- Tan Q, Orsso CE, Deehan EC, Triador L, Field CJ, Tun HM, Han JC, Müller TD & Haqq AM (2020) Current and emerging therapies for managing hyperphagia and obesity in Prader-Willi syndrome: A narrative review. *Obesity Reviews* 21: e12992
- Tanaka K, Yu HA, Yang S, Han S, Selcuklu SD, Kim K, Ramani S, Ganesan YT, Moyer A, Sinha S, *et al* (2021) Targeting Aurora B kinase prevents and overcomes resistance to EGFR inhibitors in lung cancer by enhancing BIM- and PUMA-mediated apoptosis. *Cancer Cell* 39: 1245–1261.e6
- Tang J, Gary JD, Clarke S & Herschman HR (1998) PRMT 3, a type I protein arginine N-methyltransferase that differs from PRMT1 in its oligomerization, subcellular localization, substrate specificity, and regulation. *J Biol Chem* 273: 16935–16945

- Taoka M, Nobe Y, Yamaki Y, Sato K, Ishikawa H, Izumikawa K, Yamauchi Y, Hirota K, Nakayama H, Takahashi N, *et al* (2018) Landscape of the complete RNA chemical modifications in the human 80S ribosome. *Nucleic Acids Res* 46: 9289–9298
- Taucher M, Ganisl B & Breuker K (2011) Identification, localization, and relative quantitation of pseudouridine in RNA by tandem mass spectrometry of hydrolysis products. *Int J Mass Spectrom* 304: 91–97
- Tessarz P, Santos-Rosa H, Robson SC, Sylvestersen KB, Nelson CJ, Nielsen ML & Kouzarides T (2014) Glutamine methylation in histone H2A is an RNA-polymerase-I-dedicated modification. *Nature* 505: 564–568
- Thalalla Gamage S, Bortolin-Cavaillé M-L, Link C, Bryson K, Sas-Chen A, Schwartz S, Cavaillé J & Meier JL (2022) Antisense pairing and SNORD13 structure guide RNA cytidine acetylation. *RNA* 28: 1582–1596
- Tollervey D, Lehtonen H, Jansen R, Kern H & Hurt EC (1993) Temperature-sensitive mutations demonstrate roles for yeast fibrillarin in pre-rRNA processing, pre-rRNA methylation, and ribosome assembly. *Cell* 72: 443–457
- Treger M & Westhof E (2001) Statistical analysis of atomic contacts at RNA–protein interfaces. *Journal of Molecular Recognition* 14: 199–214
- TURNER B, MELCHER SE, WILSON TJ, NORMAN DG & LILLEY DMJ (2005) Induced fit of RNA on binding the L7Ae protein to the kink-turn motif. *RNA* 11: 1192–1200
- Tyc K & Steitz JA (1989) U3, U8 and U13 comprise a new class of mammalian snRNPs localized in the cell nucleolus. *EMBO J* 8: 3113–3119
- Tycowski KT, Shu M-D, Kukoyi A & Steitz JA (2009) A conserved WD40 protein binds the Cajal body localization signal of scaRNP particles. *Mol Cell* 34: 47–57
- Tycowski KT, Shu M-D & Steitz JA (1994) Requirement for Intron-Encoded U22 Small Nucleolar RNA in 18S Ribosomal RNA Maturation. *Science* 266: 1558–1561
- Tycowski KT, Shu M-D & Steitz JA (1996) A mammalian gene with introns instead of exons generating stable RNA products. *Nature* 379: 464–466
- Tycowski KT, You ZH, Graham PJ & Steitz JA (1998) Modification of U6 spliceosomal RNA is guided by other small RNAs. *Mol Cell* 2: 629–638
- Uliel S, Liang X, Unger R & Michaeli S (2004) Small nucleolar RNAs that guide modification in trypanosomatids: repertoire, targets, genome organisation, and unique functions. *International Journal for Parasitology* 34: 445–454
- van Nues RW & Watkins NJ (2017) Unusual C'/D' motifs enable box C/D snoRNPs to modify multiple sites in the same rRNA target region. *Nucleic Acids Res* 45: 2016–2028
- Videm P, Kumar A, Zharkov O, Grüning BA & Backofen R (2021) ChiRA: an integrated framework for chimeric read analysis from RNA-RNA interactome and RNA structurome data. *Gigascience* 10: giaa158

- Vidovic I, Nottrott S, Hartmuth K, Lührmann R & Ficner R (2000) Crystal Structure of the Spliceosomal 15.5kD Protein Bound to a U4 snRNA Fragment. *Molecular Cell* 6: 1331–1342
- Vitali P, Basyuk E, Le Meur E, Bertrand E, Muscatelli F, Cavaillé J & Huttenhofer A (2005) ADAR2-mediated editing of RNA substrates in the nucleolus is inhibited by C/D small nucleolar RNAs. *J Cell Biol* 169: 745–753
- Vitali P & Kiss T (2019) Cooperative 2'-O-methylation of the wobble cytidine of human elongator tRNAMet(CAT) by a nucleolar and a Cajal body-specific box C/D RNP. *Genes Dev* 33: 741–746
- Vitali P, Royo H, Marty V, Bortolin-Cavaillé M-L & Cavaillé J (2010) Long nuclear-retained non-coding RNAs and allele-specific higher-order chromatin organization at imprinted snoRNA gene arrays. *J Cell Sci* 123: 70–83
- Vitali P, Royo H, Seitz H, Bachellerie J, Huttenhofer A & Cavaillé J (2003) Identification of 13 novel human modification guide RNAs. *Nucleic Acids Research* 31: 6543–6551
- Wajahat M, Bracken CP & Orang A (2021) Emerging Functions for snoRNAs and snoRNA-Derived Fragments. *Int J Mol Sci* 22: 10193
- Walbott H, Machado-Pinilla R, Liger D, Bland M, Réty S, Grozdanov PN, Godin K, van Tilbeurgh H, Varani G, Meier UT, *et al* (2011) The H/ACA RNP assembly factor SHQ1 functions as an RNA mimic. *Genes Dev* 25: 2398–2408
- Wang C & Meier UT (2004) Architecture and assembly of mammalian H/ACA small nucleolar and telomerase ribonucleoproteins. *EMBO J* 23: 1857–1867
- Wang G, Li J, Yao Y, Liu Y, Xia P, Zhang H, Yin M, Qin Z, Ma W & Yuan Y (2021a) Small nucleolar RNA 42 promotes the growth of hepatocellular carcinoma through the p53 signaling pathway. *Cell Death Discov* 7: 347
- Wang H, Klein MG, Zou H, Lane W, Snell G, Levin I, Li K & Sang B-C (2015) Crystal structure of human stearyl-coenzyme A desaturase in complex with substrate. *Nat Struct Mol Biol* 22: 581–585
- Wang K, Wang S, Zhang Y, Xie L, Song X & Song X (2023) SNORD88C guided 2'-O-methylation of 28S rRNA regulates SCD1 translation to inhibit autophagy and promote growth and metastasis in non-small cell lung cancer. *Cell Death Differ* 30: 341–355
- Wang X, Li Z-T, Yan Y, Lin P, Tang W, Hasler D, Meduri R, Li Y, Hua M-M, Qi H-T, *et al* (2020) LARP7-Mediated U6 snRNA Modification Ensures Splicing Fidelity and Spermatogenesis in Mice. *Molecular Cell* 77: 999-1013.e6
- Wang Y, Zhao Y, Bollas A, Wang Y & Au KF (2021b) Nanopore sequencing technology, bioinformatics and applications. *Nat Biotechnol* 39: 1348–1365
- Watkins NJ & Bohnsack MT (2012) The box C/D and H/ACA snoRNPs: key players in the modification, processing and the dynamic folding of ribosomal RNA. *WIREs RNA* 3: 397–414

- Watkins NJ, Dickmanns A & Lührmann R (2002) Conserved stem II of the box C/D motif is essential for nucleolar localization and is required, along with the 15.5K protein, for the hierarchical assembly of the box C/D snoRNP. *Mol Cell Biol* 22: 8342–8352
- Watkins NJ, Gottschalk A, Neubauer G, Kastner B, Fabrizio P, Mann M & Lührmann R (1998) Cbf5p, a potential pseudouridine synthase, and Nhp2p, a putative RNA-binding protein, are present together with Gar1p in all H BOX/ACA-motif snoRNPs and constitute a common bipartite structure. *RNA* 4: 1549–1568
- Watkins NJ, Ségault V, Charpentier B, Nottrott S, Fabrizio P, Bachi A, Wilm M, Rosbash M, Branlant C & Lührmann R (2000a) A Common Core RNP Structure Shared between the Small Nucleolar Box C/D RNPs and the Spliceosomal U4 snRNP. *Cell* 103: 457–466
- Watkins NJ, Ségault V, Charpentier B, Nottrott S, Fabrizio P, Bachi A, Wilm M, Rosbash M, Branlant C & Lührmann R (2000b) A common core RNP structure shared between the small nucleolar box C/D RNPs and the spliceosomal U4 snRNP. *Cell* 103: 457–466
- Whitehead SE, Jones KW, Zhang X, Cheng X, Terns RM & Terns MP (2002) Determinants of the interaction of the spinal muscular atrophy disease protein SMN with the dimethylarginine-modified box H/ACA small nucleolar ribonucleoprotein GAR1. *J Biol Chem* 277: 48087–48093
- Wrzesinski J, Nurse K, Bakin A, Lane BG & Ofengand J (1995) A dual-specificity pseudouridine synthase: an Escherichia coli synthase purified and cloned on the basis of its specificity for psi 746 in 23S RNA is also specific for psi 32 in tRNA(phe). *RNA* 1: 437–448
- Wu H & Feigon J (2007) H/ACA small nucleolar RNA pseudouridylation pockets bind substrate RNA to form three-way junctions that position the target U for modification. *Proc Natl Acad Sci U S A* 104: 6655–6660
- Wurtmann EJ & Wolin SL (2009) RNA under attack: Cellular handling of RNA damage. *Crit Rev Biochem Mol Biol* 44: 34–49
- Xiao M, Yang C, Schattner P & Yu Y-T (2009) Functionality and substrate specificity of human box H/ACA guide RNAs. *RNA* 15: 176–186
- Xue S, Wang R, Yang F, Terns RM, Terns MP, Zhang X, Maxwell ES & Li H (2010) Structural Basis for Substrate Placement by an Archaeal Box C/D Ribonucleoprotein Particle. *Molecular Cell* 39: 939–949
- Yagoub D, Hart-Smith G, Moecking J, Erce MA & Wilkins MR (2015) Yeast proteins Gar1p, Nop1p, Npl3p, Nsr1p, and Rps2p are natively methylated and are substrates of the arginine methyltransferase Hmt1p. *Proteomics* 15: 3209–3218
- Yang B, Wang J-Q, Tan Y, Yuan R, Chen Z-S & Zou C (2021) RNA methylation and cancer treatment. *Pharmacological Research* 174: 105937
- Yang C & Söll D (1974) Covalent attachment of fluorescent groups to transfer ribonucleic acid. Reactions with 4-bromomethyl-7-methoxy-2-oxo-2H-benzopyran. *Biochemistry* 13: 3615–3621

- Yang J, Sharma S, Watzinger P, Hartmann JD, Kötter P & Entian K-D (2016) Mapping of Complete Set of Ribose and Base Modifications of Yeast rRNA by RP-HPLC and Mung Bean Nuclease Assay. *PLoS One* 11: e0168873
- Yang J-H, Zhang X-C, Huang Z-P, Zhou H, Huang M-B, Zhang S, Chen Y-Q & Qu L-H (2006) snoSeeker: an advanced computational package for screening of guide and orphan snoRNA genes in the human genome. *Nucleic Acids Res* 34: 5112–5123
- Yang PK, Hoareau C, Froment C, Monsarrat B, Henry Y & Chanfreau G (2005) Cotranscriptional recruitment of the pseudouridylsynthetase Cbf5p and of the RNA binding protein Naf1p during H/ACA snoRNP assembly. *Mol Cell Biol* 25: 3295–3304
- Yip WSV, Shigematsu H, Taylor DW & Baserga SJ (2016) Box C/D sRNA stem ends act as stabilizing anchors for box C/D di-sRNPs. *Nucleic Acids Res* 44: 8976–8989
- Yoshida M & Ukita T (1968) Modification of nucleosides and nucleotides. VII. Selective cyanoethylation of inosine and pseudouridine in yeast transfer ribonucleic acid. *Biochim Biophys Acta* 157: 455–465
- Yoshihama M, Nakao A & Kenmochi N (2013) snOPY: a small nucleolar RNA orthological gene database. *BMC Res Notes* 6: 426
- Yu Y-T & Meier UT (2014) RNA-guided isomerization of uridine to pseudouridine—pseudouridylation. *RNA Biology* 11: 1483–1494
- Zaringhalam M & Papavasiliou FN (2016) Pseudouridylation meets next-generation sequencing. *Methods* 107: 63–72
- Zhang M, Li K, Bai J, Velema WA, Yu C, van Damme R, Lee WH, Corpuz ML, Chen J-F & Lu Z (2021) Optimized photochemistry enables efficient analysis of dynamic RNA structuromes and interactomes in genetic and infectious diseases. *Nat Commun* 12: 2344
- Zhang W & Pan T (2022) Pseudouridine RNA modification detection and quantification by RT-PCR. *Methods* 203: 1–4
- Zhao R, Kakihara Y, Gribun A, Huen J, Yang G, Khanna M, Costanzo M, Brost RL, Boone C, Hughes TR, *et al* (2008) Molecular chaperone Hsp90 stabilizes Pih1/Nop17 to maintain R2TP complex activity that regulates snoRNA accumulation. *J Cell Biol* 180: 563–578
- Zhao Y, Karijolic J, Glaunsinger B & Zhou Q (2016) Pseudouridylation of 7SK snRNA promotes 7SK snRNP formation to suppress HIV-1 transcription and escape from latency. *EMBO Rep* 17: 1441–1451
- Zhong F, Zhou N, Wu K, Guo Y, Tan W, Zhang H, Zhang X, Geng G, Pan T, Luo H, *et al* (2015) A SnoRNA-derived piRNA interacts with human interleukin-4 pre-mRNA and induces its decay in nuclear exosomes. *Nucleic Acids Res* 43: 10474–10491
- Zhu Y, Pirnie SP & Carmichael GG (2017) High-throughput and site-specific identification of 2'-O-methylation sites using ribose oxidation sequencing (RibOxi-seq). *RNA* 23: 1303–1314

- Ziv O, Gabryelska MM, Lun ATL, Gebert LFR, Sheu-Gruttadauria J, Meredith LW, Liu Z-Y, Kwok CK, Qin C-F, Macrae IJ, *et al* (2018) COMRADES determines in vivo RNA structures and interactions. *Nat Methods* 15: 785–788
- Ziv O, Price J, Shalamova L, Kamenova T, Goodfellow I, Weber F & Miska EA (2020) The Short- and Long-Range RNA-RNA Interactome of SARS-CoV-2. *Mol Cell* 80: 1067-1077.e5
- Zurita-Lopez CI, Sandberg T, Kelly R & Clarke SG (2012) Human protein arginine methyltransferase 7 (PRMT7) is a type III enzyme forming ω -NG-monomethylated arginine residues. *J Biol Chem* 287: 7859–7870

RÉSUMÉS

Titre : Caractérisation et analyse fonctionnelle des snoARN orphelins par la mise au point d'une méthode étudiant les interactions snoARN-ARN cibles.

Mots-clefs : Petits ARN nucléolaire, Interactions ARN-ARN, Crosslink au psoralène, Modifications de l'ARN guidée par l'ARN, snoARN orphelins, Pseudouridylation

Résumé : Les ARN cellulaires portent de nombreuses modifications post-transcriptionnelles nécessaires à leur repliement, leur stabilité, et aux interactions dans lesquelles ils sont impliqués. Parmi ces modifications chimiques, les pseudouridylation et les méthylations en 2'-O-ribose sont respectivement guidées par les petits ARN nucléolaires (snoARN) des classes H/ACA et C/D. Via ces modifications chimiques, ces ARN, présents dans les nucléoles et les corps de Cajal des cellules, jouent un rôle primordial dans la maturation des ARN ribosomiques (ARNr) et des ARN de l'épissage (ARNsn). En supplément de ces fonctions de guide, les snoARN sont impliqués dans d'autres mécanismes cellulaires, tels que le clivage des pré-ARNr ou le maintien des télomères. Récemment, plusieurs travaux laissent émerger l'idée que les snoARN cibleraient les ARN messagers (ARNm). Allant dans ce sens, plus d'un tiers des snoARN répertoriés n'ont pas encore de fonctions ou de mécanismes associés. Ils sont philosophiquement nommés « snoARN orphelins ». Les snoARN interagissant physiquement, par appariement direct de paires de base, avec les ARN qu'ils ciblent, le principal objectif de ma thèse a été de mettre au point une méthode de co-purification *in vivo* d'un snoARN avec sa ou ses cibles, de manière à pouvoir les identifier. L'identification des ARN ciblés par les snoARN orphelins se trouve être une information cruciale afin de pouvoir associer un rôle à ces snoARN, et ainsi pouvoir comprendre la place qu'ils occupent au sein de la machinerie cellulaire. Dans cette optique, la méthode mise au point durant mon doctorat consiste en premier lieu à un pontage au psoralène des interactions ARN-ARN au sein de ces cellules. La purification ultérieure d'un snoARN d'intérêt permet alors de co-purifier spécifiquement les ARN cibles physiquement liés à ce dernier. Les ARN obtenus sont fragmentés, puis une étape de ligation de proximité permet de relier entre elles les extrémités des fragments d'ARN proches spatialement. Des ARN chimériques sont alors générés, composés (i) d'un fragment appartenant au snoARN orphelin d'intérêt (ii) d'un fragment d'un ARN cible putatif co-purifié avec le snoARN. Ces ARN chimériques servent ensuite de modèle pour la synthèse d'une librairie d'ADN complémentaire (ADNc) qui est séquencée à haut débit. L'analyse des données obtenues permet l'identification des ARN co-purifiés avec le snoARN. Même si différents aspects de cette méthode restent à peaufiner, j'ai notamment pu apporter, au cours de ma thèse, des informations supplémentaires concernant (i) l'interaction entre le snoARN SNORD115,

impliqué dans le syndrome de Prader-Willi, et le putatif ARNm cible 5-HT2C (ii) l'interaction entre SNORA55, un snoARN orphelin fortement conservé au cours de l'évolution, et l'ARNr 18S.

D'autre part, au cours de mon doctorat, j'ai eu l'opportunité d'étudier en profondeur les caractéristiques moléculaires des snoARN H/ACA. Via une immunoprécipitation d'une des protéines cœurs des complexes H/ACA, j'ai pu identifier et tester l'accumulation de nouveaux snoARN H/ACA. Mes résultats ont montré qu'à ce jour, la quasi-totalité des snoARN H/ACA présent dans les cellules humaines ont été découvert par la communauté scientifique. Cependant, j'ai pu identifier un nouveau snoARN H/ACA, nommé H/ACA 186, situé dans une région génomique atypique pour un snoARN. Le gène *H/ACA 186*, chevauchant une région intronique et une région exonique, pourrait être associé à un nouveau mécanisme d'action des snoARN, comme la régulation de son gène hôte. Également, de par la création d'outils génétiques, j'ai pu participer à démontrer que quatre snoARN H/ACA humains étaient capables, par un réarrangement conformationnel, de réaliser l'isomérisation de deux Uridines consécutives sur un même ARN cible. Ces résultats mettent en valeur l'extraordinaire diversité des mécanismes associés aux snoARN H/ACA dans les cellules humaines.

Title: Characterization and functional analysis of orphan snoRNAs by the development of a novel method to co-purify snoRNAs with their targets.

Keywords: Small nucleolar RNA, RNA-RNA interactions, psoralen crosslink, RNA-guided RNA modifications, orphan snoRNAs, Pseudouridylation

Abstract: Cellular RNAs carry various post-transcriptionally modified nucleotides that are essential for their folding, stability and interactions. Two highly represented chemical modifications are the pseudouridylations and the 2'-O-ribose methylations, which are guided respectively by H/ACA and C/D small nucleolar RNAs (snoRNAs). These snoRNAs accumulating in cell nucleoli and Cajal bodies play a crucial role in ribosomal RNA (rRNA) and splicing RNA (snRNA) maturation. In addition to their guiding functions, snoRNAs are involved in other cellular mechanisms such as pre-rRNA cleavage and telomere maintenance. Recent studies suggested that snoRNAs might also target messenger RNAs (mRNAs). Interestingly, more than one-third of known snoRNAs do not have an identified function or an associated mechanism and therefore referred to as "orphan snoRNAs". To address their function, the main objective of my PhD research was to develop an *in vivo* method for co-purification of a snoRNA with its interacting RNAs in order to identify its possible targets. Identifying the RNAs targeted by orphan snoRNAs is crucial for understanding their cellular role. The method I developed involved RNA-RNA crosslinking using psoralen. Subsequent purification of the snoRNA of interest allowed the specific co-purification of physically bound target RNAs. Fragmentation of the purified RNAs was followed by a proximity ligation step that joined spatially close RNA fragments. This process generated chimeric RNAs composed of (i) a fragment from the orphan snoRNA of interest and (ii) a fragment from a co-purified putative target RNA. These chimeric RNAs were used then to create a complementary DNA (cDNA) library, which was sequenced. Analysis of the sequencing data enabled us the identification of RNAs co-purified with the snoRNA. Although some aspects of this method still require further refinement, I was able to provide additional information on (i) the interaction between the snoRNA SNORD115, associated with Prader-Willi syndrome, and the putative target mRNA 5-HT2C and (ii) the interaction between SNORA55, a highly conserved orphan snoRNA, and the 18S rRNA.

Furthermore, during my doctoral studies, I had the opportunity to extensively study the molecular characteristics of H/ACA snoRNAs. Through immunoprecipitation of one of the core proteins of the H/ACA complexes, I was able to identify and examine the accumulation of new H/ACA snoRNAs. My results showed that, to date, almost all H/ACA snoRNAs present in

human cells have been discovered by the scientific community. However, I was able to identify a new H/ACA snoRNA, named H/ACA 186, located in a genomic region that is atypical for a snoRNA. The H/ACA 186 gene, overlapping an intronic and an exonic region, could be associated to a novel mechanism of snoRNAs, namely the regulation of its host gene. Additionally, by developing genetic tools, contributed to the demonstration that four human H/ACA snoRNAs were capable, through conformational rearrangement, of catalyzing the isomerization of two consecutive uridines on the same target RNA. These results highlight the extraordinary diversity of mechanisms associated with H/ACA snoRNAs in human cells.