



## AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact : [ddoc-theses-contact@univ-lorraine.fr](mailto:ddoc-theses-contact@univ-lorraine.fr)

## LIENS

Code de la Propriété Intellectuelle. articles L 122. 4

Code de la Propriété Intellectuelle. articles L 335.2- L 335.10

[http://www.cfcopies.com/V2/leg/leg\\_droi.php](http://www.cfcopies.com/V2/leg/leg_droi.php)

<http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm>

# École Doctorale BioSE (Biologie-Santé-Environnement)

## Thèse

Présentée et soutenue publiquement pour l'obtention du titre de

**DOCTEUR DE L'UNIVERSITE DE LORRAINE**

**Mention : "Sciences de la Vie et de la Santé"**

par **Géraud HECKLER**

---

**Etude de l'effet d'un stress oxydant sur l'activité spliceosomale et  
sur les RNP à boîtes C/D et H/ACA**

---

Le 12 Novembre 2019

### Membres du jury:

|                               |                                   |                                                        |
|-------------------------------|-----------------------------------|--------------------------------------------------------|
| <b>Rapporteurs :</b>          | <b>Mme. Beata JADY</b>            | <b>Chargée de Recherche, CNRS, Toulouse</b>            |
|                               | <b>M. Yves HENRY</b>              | <b>Directeur de Recherche, CNRS, Toulouse</b>          |
| <b>Examineurs :</b>           | <b>M. Bruno CHARPENTIER</b>       | <b>Professeur des Universités, UL, Nancy</b>           |
|                               | <b>Mme. Céline VERHEGGEN</b>      | <b>Directeur de Recherche, CNRS, Montpellier</b>       |
| <b>Membres invités :</b>      | <b>M. Stéphane DALLONGEVILLE</b>  | <b>Ingénieur de Recherche, Institut Pasteur, Paris</b> |
|                               | <b>Mme. Natacha DREUMONT</b>      | <b>Maître de Conférences, UL, Nancy</b>                |
| <b>Directrices de thèse :</b> | <b>Mme. Isabelle BEHM-ANSMANT</b> | <b>Chargée de Recherche, CNRS, Nancy</b>               |
|                               | <b>Mme. Christelle AIGUEPERSE</b> | <b>Ingénieur d'Etudes hors classe, CNRS, Nancy</b>     |

---

UMR 7365 CNRS-UL, Laboratoire IMoPA (Ingénierie Moléculaire et Physiopathologie Articulaire)  
Campus Biologie-Santé, Faculté de Médecine, Bâtiment Biopôle, 9 avenue de la forêt de Haye  
54505 Vandoeuvre-lès-Nancy-France



## REMERCIEMENTS

Je tiens tout d'abord à exprimer ma plus profonde gratitude envers ma directrice de thèse, Isabelle Behm-Ansmant et ma co-directrice de thèse, Christelle Aigueperse. Vous m'avez formé et accompagné durant toutes ces années et vous m'avez soutenu afin que je m'épanouisse dans ce qui me tiens à cœur : la microscopie. Maintenant vous me soutenez de toutes vos forces, par la parole et par les actes pour que je puisse soutenir dans les meilleures conditions avant de me tourner sereinement vers mon nouveau projet. Je vous en suis extrêmement reconnaissant et vous serez à jamais dans mon cœur.

Je remercie très sincèrement Natacha Dreumont et Céline Verheggen d'avoir accepté de faire partie de mon comité de suivi de thèse. Je vous dois de nombreux conseils avisés et un véritable soutien sur le plan scientifique. Je vous remercie également d'avoir accepté d'être membres de mon jury de thèse.

Je remercie Madame Beata Jady et Monsieur Yves Henry pour avoir accepté d'être rapporteurs de mon manuscrit de thèse et d'évaluer mon travail.

Je remercie Monsieur Bruno Charpentier et Monsieur Stéphane Dallongeville pour avoir accepté de juger mon travail et d'être membres de mon jury de thèse.

Je remercie infiniment mon papa et ma maman. Vous êtes ma plus belle source d'inspiration au quotidien et aucun mot n'est assez fort pour exprimer que je vous dois tout, tout simplement.

Je remercie ma sœur Lucile et mon frère Quentin du plus profond de mon cœur. Votre présence me donne chaque jour la force d'avancer.

Je remercie Syrine. Tu m'as accompagné, épaulé et soutenu de toutes tes forces pendant ces nombreuses années. Je ne pourrai jamais l'oublier.

Je pense à toutes ces belles personnes qui m'ont accompagné durant ces années, ces mois, ces jours... Je vous voyais tous les jours ou juste une fois par an, mais cette simple présence quotidienne ou occasionnelle a rendu ma vie plus belle. A vous tous, ma famille, mes amis, mes collègues, à ceux qui ont construit mon passé, qui illuminent mon présent et qui seront à jamais dans mon cœur et mes pensées, je vous dis tous merci ☺





# ABREVIATIONS ET ACRONYMES



## **ABREVIATIONS ET ACRONYMES**

°C : degré Celsius

aa : acide(s) amine(s)

A/mA : ampère/milliampère

ADN : Acide DésoxyriboNucléique

ADNc : ADN complémentaire

ARN : Acide RiboNucléique

ARNm : ARN messenger

ARNr : ARN ribosomique

ARNt : ARN de transfert

ATP : adenosine 5'triphosphate

BSA : "Bovine Serum Albumin" (Albumine de sérum bovin)

CB : "cajal bodies "

C-ter : C-terminal(e)

Da : Dalton

DAVID : "Database for Annotation, Visualization and Integrated Discovery"

DMEM : Dulbecco's modified Eagle's Medium

Dnase : désoxyribonucléase

dNTP : désoxyribonucléotides

DTT : dithiothréitol

EA : épissage alternatif

ESE : "Exonic Splicing Enhancer"

ESS : "Exonic Splicing Silencer"

EST : "Expressed Sequence Tag"

FISH : "Fluorescence In Situ Hybridization" (hybridation de fluorescence in situ)

FITC : Fluorescéine IsoThioCyanate

FRAP : "Fluorescence Recovery After Photobleaching"

GO : "Gene Ontology"

HGPS : "Hutchinson-Gilford Progeria Syndrome"

hnRNP : "heteronuclear RiboNucleic Particle"

ISE : "Intronic Splicing Enhancer(s)"

ISS : "Intronic Splicing Silencer(s)"

kb : kilobase

kDa : kilo Dalton

LB : Luria-Bertani

LNA : "Locked Nucleic Acid" ("acide nucléique verrouillé")

miR, miRNA : microRNA

NLS : "Nuclear Localisation Signal" (signal de localisation nucléaire)

NO : monoxyde d'azote

NPC : "Nuclear Pore Complex" (pores nucléaires)

NRS : "Nuclear Retention Signal", signal de rétention nucléaire

nt : nucléotide(s)

N-ter : N-terminal(e)

PAGE : "Polyacrylamide Gel Electrophoresis"

PCR : "Polymerase Chain Reaction"

pol : polymérase

pré-ARNm : Acide Ribo Nucléique pré-messager

RMN : Résonance Magnétique Nucléaire

RNP : "Ribo Nucleic Particle"

RNS : "Reactive Nitrogen Species"

ROS : "Reactive Oxygen Species"

RT : Rétrotranscription

sat : satellite

scaRNA : scavenger RNA

siRNA : "small interfering RNA", petits ARN interférents

SMN : "Survival Motor Neuron" (protéine ou complexe SMN)

SNP : "Single Nucleotide Polymorphism"

snRNA : "small nuclear RiboNucleic Acid"

snRNP : "small nuclear RiboNucleic Particle"

snoRNA : small nucleolar RNA

SR : "Serin-Arginin Rich"

Tm : "melting Température" (température de fusion)

U snRNA : "Uridine-rich Small Nuclear Ribonucleic Acids"

U snRNP : "Uridine-rich Small Nuclear Ribonucleic Particle"

WB : "Western Blot" ou "Western Blotting"

WT : "Wild Type" (type sauvage)



# LISTE DES FIGURES ET DES TABLEAUX





| <b>N° de Figure</b> | <b>Titre</b>                                                                                                                                                                                                 | <b>Page</b> |
|---------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------|
| <b>INTRODUCTION</b> |                                                                                                                                                                                                              |             |
| <b>Figure I1</b>    | <b>Réactions de trans-estérification se déroulant au cours de l'épissage des pré-ARN</b>                                                                                                                     | <b>1</b>    |
| <b>Figure I2</b>    | <b>Séquences nécessaires pour l'épissage</b>                                                                                                                                                                 | <b>2</b>    |
| <b>Figure I3</b>    | <b>Schéma général de l'assemblage séquentiel du spliceosome au cours de la réaction d'épissage</b>                                                                                                           | <b>4</b>    |
| <b>Figure I4</b>    | <b>Les différents modes d'épissage alternatif des pré-ARN</b>                                                                                                                                                | <b>8</b>    |
| <b>Figure I5</b>    | <b>La régulation de l'épissage alternatif est un processus très complexe impliquant de nombreux niveaux de régulation</b>                                                                                    | <b>10</b>   |
| <b>Figure I6</b>    | <b>Le rôle de la chromatine dans l'épissage alternatif</b>                                                                                                                                                   | <b>12</b>   |
| <b>Figure I7</b>    | <b>Régulation de l'épissage alternatif co-transcriptionnel par les résidus m6A et la protéine hnRNP G</b>                                                                                                    | <b>14</b>   |
| <b>Figure I8</b>    | <b>Pseudouridylation stress-spécifique des UsnRNA de levure</b>                                                                                                                                              | <b>18</b>   |
| <b>Figure I9</b>    | <b>Biogenèse des UsnRNP Sm</b>                                                                                                                                                                               | <b>20</b>   |
| <b>Figure I10</b>   | <b>Structure chimique d'un résidu 2'-O-méthylé (Nm) et d'un résidu pseudouridine (Ψ)</b>                                                                                                                     | <b>22</b>   |
| <b>Figure I11</b>   | <b>Listing des modifications post-transcriptionnelles internes identifiées au sein des UsnRNA humains</b>                                                                                                    | <b>23</b>   |
| <b>Figure I12</b>   | <b>Les modifications post-transcriptionnelles internes des UsnRNA humains sont présentes au sein de segments d'ARN impliqués dans des interactions essentielles au déroulement de la réaction d'épissage</b> | <b>25</b>   |
| <b>Figure I13</b>   | <b>Structure des ARN et des RNP à boîtes H/ACA et C/D</b>                                                                                                                                                    | <b>26</b>   |
| <b>Figure I14</b>   | <b>Structuration de scaRNA en tandem (plusieurs motifs h/aCA ou plusieurs motifs c/D) ou hybride</b>                                                                                                         | <b>28</b>   |
| <b>Figure I15</b>   | <b>Illustration de la diversité des substructures nucléaires sans membranes dans les cellules de mammifères</b>                                                                                              | <b>29</b>   |
| <b>Figure I16</b>   | <b>Organisation fonctionnelle de la protéine coiline</b>                                                                                                                                                     | <b>33</b>   |
| <b>Figure I17</b>   | <b>Rôles proposés de TDP-43 dans la modification post-transcriptionnelle des snARNs U1 et U2 via l'adressage des scaARNs à boîte C/D dans les corps de Cajal</b>                                             | <b>38</b>   |
| <b>Figure I18</b>   | <b>Principales voies de production et de métabolisation des ROS chez les cellules eucaryotes</b>                                                                                                             | <b>41</b>   |
| <b>Figure I19</b>   | <b>Voies de signalisations cellulaires majeures régulées par les ROS</b>                                                                                                                                     | <b>43</b>   |

| N° de Figure | Titre                                                                                                                                          | Page |
|--------------|------------------------------------------------------------------------------------------------------------------------------------------------|------|
| RESULTATS    |                                                                                                                                                |      |
| Figure R1    | Protocole de segmentation automatisé basé sur l'algorithme Watershed                                                                           | 51   |
| Figure R2    | Extraction automatique des marqueurs lors du processus de segmentation                                                                         | 52   |
| Figure R3    | Protocole de segmentation basé sur la combinaison d'une segmentation watershed classique et une segmentation watershed basée sur des marqueurs | 53   |
| Figure R4    | Architecture du protocole <i>Substructure Analyzer</i>                                                                                         | 56   |
| Figure R5    | Analyse cinétique de la cytotoxicité et de l'induction de ROS par l'H <sub>2</sub> O <sub>2</sub> dans des cellules HeLa.                      | 58   |
| Figure R6    | Cinétique de la distribution nucléoplasmique de 53BP1 après un stress oxydant.                                                                 | 59   |
| Figure R7    | Quantification de la protéine Ki67 en immunofluorescence dans des cellules traitées ou non par H <sub>2</sub> O <sub>2</sub> .                 | 61   |
| Figure R8    | Ki-67, un marqueur de phase du cycle cellulaire.                                                                                               | 62   |
| Figure R9    | Analyse du cycle cellulaire après un stress oxydant au cours d'une cinétique de temps                                                          | 64   |
| Figure R10   | Classification non supervisée (clustering) des données d'expression génique au cours d'une cinétique de réponse au stress oxydant              | 65   |
| Figure R11   | Analyse de l'effet du stress oxydant sur l'expression des gènes humains.                                                                       | 67   |
| Figure R12   | Enrichissement des différentes catégories de gènes au sein des gènes différentiellement exprimés en réponse au stress oxydant.                 | 68   |
| Figure R13   | Analyse des voies de signalisation associées aux gènes différentiellement exprimés en réponse au stress oxydant                                | 69   |
| Figure R14   | Analyse de l'effet d'un stress oxydant sur l'épissage des transcrits primaires humains.                                                        | 71   |
| Figure R15   | Nature et fonction des gènes différentiellement épissés en phase précoce de réponse au stress oxydant                                          | 72   |
| Figure R16   | Analyse du recouvrement des régulations de l'expression et de l'épissage alternatif des gènes en phase précoce de réponse au stress oxydant    | 73   |
| Figure R17   | Analyse de la quantité des UsnRNA des spliceosomes majeur et mineur au cours d'une cinétique de réponse au stress oxydant.                     | 78   |
| Figure R18   | Principe de l'enrichissement en UsnRNP actives par RNA-immunoprécipitation (RIP)                                                               | 80   |
| Figure R19   | Proportions relatives de chacune des UsnRNP dans des cellules HeLa soumises à un stress oxydant                                                | 82   |

| <b>N° de Figure</b> | <b>Titre</b>                                                                                                                                                      | <b>Page</b> |
|---------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------|
| Figure R20          | Optimisations de détection des U snARNs du spliceosome mineur et majeur                                                                                           | 85          |
| Figure R21          | Localisation du snARN U1 par RNA-FISH                                                                                                                             | 87          |
| Figure R22          | Localisation du snARN U2 par RNA-FISH                                                                                                                             | 88          |
| Figure R23          | Localisation de la coiline et de SC35 en immunofluorescence lors d'un SO.                                                                                         | 90          |
| Figure R24          | Localisation de Gem2 et Sumo1 lors d'un SO.                                                                                                                       | 90          |
| Figure R25          | Cinétique de la distribution nucléoplasmique de la coilin après un stress oxydant.                                                                                | 92          |
| Figure R26          | Analyse de l'expression de la protéine coiline en réponse au stress oxydant                                                                                       | 94          |
| Figure R27          | Analyse de l'expression de la protéine coiline en réponse au stress oxydant.                                                                                      | 96          |
| Figure R28          | Effet du stress oxydant sur l'exclusion de l'exon 4 de la coiline                                                                                                 | 97          |
| Figure R29          | Composition et structure de l'exon 4 de la coiline                                                                                                                | 98          |
| Figure R30          | Optimisation de la déplétion de la coilin endogène par interférence ARN                                                                                           | 99          |
| Figure R31          | Mise au point de la surexpression de la coiline et de la coiline $\Delta 4$                                                                                       | 100         |
| Figure R32          | Suivi du nombre de foci coiline en fonction du niveau de surexpression de la protéine coiline $\Delta 4$ -EGFP chez des cellules transfectées (tag EGFP en C-ter) | 102         |
| Figure R33          | Caractérisation du niveau d'expression de Coilp1 en réponse au stress                                                                                             | 103         |
| Figure R34          | Effet de différents stress sur la localisation et le niveau d'expression de PA28 $\gamma$                                                                         | 105         |
| Figure R35          | Rôles fonctionnels de WRAP53 au sein des corps de Cajal.                                                                                                          | 107         |
| Figure R36          | Effet du stress oxydant sur la localisation et le niveau d'expression de la protéine WRAP53.                                                                      | 108         |
| Figure R37          | Effet du stress oxydant sur l'expression de TDP-43 et VRK1.                                                                                                       | 110         |
| Figure R38          | Effet du SO sur l'expression et la localisation de la protéine SMN.                                                                                               | 112         |
| Figure R39          | L'expression de certains sno-/scaRNAs est significativement affectée par le SO.                                                                                   | 113         |
| Figure R40          | Analyse de l'expression des protéines composant les RNP à boîtes H/ACA et C/D.                                                                                    | 114         |

| <b>N° de Figure</b> | <b>Titre</b>                                                                                                                                                                | <b>Page</b> |
|---------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------|
| <b>Figure R41</b>   | <b>Visualisation par immunofluorescence de protéines RNP des boîtes H/ACA ou C/D suite à un stress oxydant</b>                                                              | <b>115</b>  |
| <b>Figure R42</b>   | <b>Suivi de la localisation de composants protéiques des U snRNPs spécifiques au cours d'un stress oxydant</b>                                                              | <b>118</b>  |
| <b>Figure R43</b>   | <b>Principe la méthode Ribomethseq</b>                                                                                                                                      | <b>120</b>  |
| <b>Figure R44</b>   | <b>Analyse par la méthode Ribomethseq du contenu en résidus 2'-O-méthylés des UsnRNA de cellules HeLa S3.</b>                                                               | <b>120</b>  |
| <b>Figure R45</b>   | <b>Cartographie des anciennes et des nouvelles méthylations détectées au sein des snARNs U4, U6, U4atac et U6atac</b>                                                       | <b>121</b>  |
| <b>Figure R46</b>   | <b>Extinction de l'expression de la protéine Nop58 par emploi de siRNA</b>                                                                                                  | <b>122</b>  |
| <b>Figure R47</b>   | <b>Analyse de l'effet de l'extinction de Nop58 sur le contenu en résidus 2'-O-méthylés de l'ARNr 28S humain.</b>                                                            | <b>123</b>  |
| <b>Figure R48</b>   | <b>Analyse de l'effet de l'extinction de Nop58 sur le contenu en résidus 2'-O-méthylés des UsnRNA majeurs</b>                                                               | <b>124</b>  |
| <b>Figure R49</b>   | <b>Analyse de l'effet de l'extinction de Nop58 sur le contenu en résidus 2'-O-méthylés des UsnRNA mineurs</b>                                                               | <b>125</b>  |
| <b>Figure R50</b>   | <b>Récapitulatif des effets d'un stress thermique (HS) et d'un stress oxydant (OS) sur le contenu en 2'-O-méthylations des UsnRNAs majeurs et mineurs de cellules HeLa.</b> | <b>127</b>  |

| N° de<br>Tableau     | Titre                                                                                                                                | Page |
|----------------------|--------------------------------------------------------------------------------------------------------------------------------------|------|
| RESULTATS            |                                                                                                                                      |      |
| Tableau R1           | Test ROC (Receiver Operating Characteristic) entre cellules stressées et non stressées au cours de la cinétique de temps             | 60   |
| Tableau R2           | L'expression des protéines SR n'est pas affectée par le SO                                                                           | 74   |
| Tableau R3           | L'expression des protéines hnRNP n'est pas affectée par le SO.                                                                       | 75   |
| Tableau R4           | Le SO affecte significativement l'expression des UsnRNA du spliceosome mineur                                                        | 76   |
| MATERIEL ET METHODES |                                                                                                                                      |      |
| Tableau M1           | Désignations, séquences et usages des oligodésoxynucléotides employés                                                                | 190  |
| Tableau M2           | Désignations, caractéristiques et conditions d'utilisation des anticorps primaires                                                   | 192  |
| Tableau M3.A         | Réaction de qPCR pour la détection de produits spécifiques à partir d'ARN totaux rétrotranscrits (ADNc)                              | 198  |
| Tableau M3.B         | Programme de qPCR sur Step One Plus Real-Time PCR system (Applied Biosystems)                                                        | 198  |
| Tableau M4           | Dilution des ADNc en fonction des ARN amplifié par qPCR et estimation de leur efficacité en fonction de la température d'hybridation | 198  |
| Tableau M5.A         | Réaction de PCR pour la détection de produits spécifiques à partir d'ARN totaux retrotranscrits (ADNc)                               | 199  |
| Tableau M5.B         | Programme de PCR                                                                                                                     | 199  |



# TABLE DES MATIERES





# Table des matières

## INTRODUCTION

|      |                                                                                                              |    |
|------|--------------------------------------------------------------------------------------------------------------|----|
| I.   | L'épissage des pré-ARN chez l'Homme .....                                                                    | 1  |
| 1.   | L'épissage .....                                                                                             | 1  |
| a)   | Définition .....                                                                                             | 1  |
| b)   | Séquences de définition des introns .....                                                                    | 2  |
| c)   | Présentation générale de la machinerie d'épissage : le spliceosome .....                                     | 3  |
| d)   | Les mécanismes de la réaction d'épissage .....                                                               | 4  |
| 2.   | L'épissage alternatif .....                                                                                  | 5  |
| a)   | Définition et occurrence chez l'Homme .....                                                                  | 5  |
| b)   | Les différents types d'épissage alternatif .....                                                             | 6  |
| c)   | Fonctions physiologiques de l'épissage alternatif .....                                                      | 9  |
| d)   | Les mécanismes de régulation de l'épissage alternatif .....                                                  | 9  |
| e)   | Epissage alternatif et stress .....                                                                          | 15 |
| II.  | Biogenèse/Maturation du spliceosome .....                                                                    | 19 |
| 1.   | Biogenèse des U snRNPs .....                                                                                 | 19 |
| 2.   | Les modifications post-transcriptionnelles des UsnRNA .....                                                  | 21 |
| a)   | Contenu en modifications post-transcriptionnelles des UsnRNA humains .....                                   | 21 |
| b)   | Fonction des modifications post-transcriptionnelles des UsnRNA .....                                         | 24 |
| c)   | Formation des modifications post-transcriptionnelles des UsnRNA .....                                        | 26 |
| III. | Les corps de Cajal et les substructures nucléaires .....                                                     | 28 |
| 1.   | Introduction aux substructures nucléaires sans membranes .....                                               | 28 |
| a)   | Propriétés générales et fonctions des corps nucléaires .....                                                 | 28 |
| b)   | Dynamique structurale des corps nucléaires .....                                                             | 30 |
| c)   | Fonctions architecturales des ARNs .....                                                                     | 31 |
| d)   | Importance des domaines protéiques de faible complexité .....                                                | 31 |
| 2.   | Présentation générale des corps de Cajal .....                                                               | 32 |
| 3.   | La coiline, un facteur essentiel assurant la structure et la fonction des CBs .....                          | 33 |
| 4.   | Fonctions des corps de Cajal dans la formation des modifications post-transcriptionnelles des U snARNs ..... | 35 |
| a)   | Adressage des snoRNPs dans les nucléoles .....                                                               | 35 |
| b)   | Rétention des scaRNA au sein des Corps de cajal .....                                                        | 36 |
| 5.   | Effets des stress sur les CBs .....                                                                          | 38 |

|      |                                                                                                                               |    |
|------|-------------------------------------------------------------------------------------------------------------------------------|----|
| IV.  | Le stress oxydant .....                                                                                                       | 40 |
| 1.   | Définition des ROS et fonctions physiologiques.....                                                                           | 40 |
| a)   | Définition des ROS.....                                                                                                       | 40 |
| b)   | Fonctions physiologiques des ROS .....                                                                                        | 41 |
| 2.   | Définition d'un stress oxydant et dommages dans la cellule.....                                                               | 43 |
| a)   | Définition et origines du stress oxydant.....                                                                                 | 43 |
| b)   | Effets/dommage du stress oxydant .....                                                                                        | 44 |
| 3.   | Mécanismes de réponses et d'adaptation au stress oxydant.....                                                                 | 45 |
| 4.   | Stress oxydant et épissage alternatif : état des connaissances .....                                                          | 45 |
|      | OBJECTIFS DE THESE.....                                                                                                       | 47 |
|      | RESULTATS ET DISCUSSION                                                                                                       |    |
| I.   | Développement de protocoles d'analyse d'images .....                                                                          | 50 |
| II.  | Validation des conditions de stress oxydant .....                                                                             | 57 |
| 1.   | Effets du SO sur la viabilité cellulaire et la production de ROS.....                                                         | 57 |
| 2.   | Validation de 53BP1 comme marqueur de stress en immunofluorescence.....                                                       | 59 |
| 3.   | Evaluation de la prolifération cellulaire .....                                                                               | 60 |
| 4.   | Analyse des effets du stress oxydant sur le cycle cellulaire .....                                                            | 63 |
| III. | Effets du stress oxydant sur le transcriptome cellulaire .....                                                                | 64 |
| 1.   | Analyse des effets du SO sur l'expression des gènes humains .....                                                             | 66 |
| 2.   | Analyse des effets du SO sur l'épissage alternatif.....                                                                       | 70 |
| IV.  | Etude des mécanismes cellulaires à l'origine des variations d'épissage observées après un stress oxydant.....                 | 74 |
| 1.   | Le stress oxydant n'affecte ni l'expression, ni l'épissage des gènes codant des protéines SR et hnRNP.....                    | 74 |
| 2.   | Le stress oxydant affecte l'expression des UsnRNA mineurs.....                                                                | 75 |
| 3.   | L'épissage et le niveau d'expression des gènes contenant des introns U12-dépendant ne sont pas affectés en réponse au SO..... | 79 |
| 4.   | Effet du SO sur la quantité de particules UsnRNP actives.....                                                                 | 80 |
| 5.   | Etude des effets du SO sur la localisation des UsnARN.....                                                                    | 83 |
| a)   | Mise au point du RNA-FISH .....                                                                                               | 83 |
| b)   | Effets du SO sur la localisation des snARN U1 et U2 .....                                                                     | 86 |
| 6.   | Effet du SO sur différentes substructures nucléaires.....                                                                     | 89 |
| 7.   | Caractérisation de la redistribution nucléoplasmique de la coiline induite par le SO .....                                    | 91 |
| a)   | Etude des mécanismes à l'origine de la redistribution nucléoplasmique de la coiline.....                                      | 93 |

|     |                                                                                                                                                                            |     |
|-----|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| b)  | Effet du stress oxydant sur l'expression de variants de coiline .....                                                                                                      | 96  |
| 8.  | Etude de l'effet du stress oxydant sur les composants des CB .....                                                                                                         | 104 |
| a)  | Effets du stress oxydant sur la localisation et l'expression de PA28 $\gamma$ .....                                                                                        | 105 |
| b)  | Effet du stress oxydant sur la localisation et l'expression de WRAP53 .....                                                                                                | 106 |
| c)  | Effet du stress oxydant sur la localisation et l'expression de TDP43 .....                                                                                                 | 109 |
| d)  | Effet du stress oxydant sur la localisation et l'expression de la protéine SMN .....                                                                                       | 111 |
| e)  | Effet du SO sur l'expression et la localisation des composants des RNPs à boîte C/D et à boîte H/ACA .....                                                                 | 112 |
| f)  | Effet du SO sur la localisation de composants des UsnRNP .....                                                                                                             | 116 |
| V.  | Effets du SO sur la fonctionnalité des corps de Cajal .....                                                                                                                | 119 |
|     | CONCLUSION .....                                                                                                                                                           | 128 |
|     | ANNEXE 1: Article - Substructure Analyzer: a user-friendly workflow for rapid exploration and accurate analysis of cellular bodies in fluorescence microscopy images ..... | 133 |
|     | ANNEXE 2: Transcriptome Analysis Console (TAC) 4.0.2 .....                                                                                                                 | 162 |
|     | MATERIEL ET METHODES                                                                                                                                                       |     |
| I.  | Matériel .....                                                                                                                                                             | 190 |
| 1.  | Cellules Humaines .....                                                                                                                                                    | 190 |
| 2.  | Souches bactériennes .....                                                                                                                                                 | 190 |
| 3.  | Oligodeoxynucleotides .....                                                                                                                                                | 190 |
| 4.  | Plasmides et constructions génétiques .....                                                                                                                                | 192 |
| 5.  | Anticorps .....                                                                                                                                                            | 192 |
| 6.  | 2'-O-Methylation par RiboMethSeq .....                                                                                                                                     | 193 |
| 7.  | Micropuces jonctions exon-exon .....                                                                                                                                       | 194 |
| 8.  | Tampons et solutions .....                                                                                                                                                 | 194 |
| II. | Méthodes .....                                                                                                                                                             | 195 |
| 1.  | Culture cellulaire et Induction du stress oxydant .....                                                                                                                    | 195 |
| 2.  | Analyse de l'expression des ARN par RT-qPCR (Quantitative Reverse Transcription Polymerase Chain Reaction) .....                                                           | 196 |
| a)  | Extraction des ARN .....                                                                                                                                                   | 196 |
| b)  | Digestion de l'ADN par la DNase RQ1 dans les échantillons d'ARN .....                                                                                                      | 196 |
| c)  | Transcription Reverse (Reverse Transcription ou RT) .....                                                                                                                  | 196 |
| d)  | Détection des produits de RT par PCR quantitative (qPCR) .....                                                                                                             | 197 |
| e)  | Analyse de l'expression du pseudogène coilp1 par PCR en point finale .....                                                                                                 | 199 |
| 3.  | Analyse de l'expression des protéines .....                                                                                                                                | 200 |

|    |                                                                                                                     |     |
|----|---------------------------------------------------------------------------------------------------------------------|-----|
| a) | Extraction des protéines.....                                                                                       | 200 |
| b) | Electrophorèse et Western Blot.....                                                                                 | 201 |
| 4. | Immunoprécipitation .....                                                                                           | 201 |
| a) | Préparation des billes magnétiques .....                                                                            | 201 |
| b) | Préparation des extraits cellulaires et immunosélection.....                                                        | 202 |
| c) | Réduction du signal des chaînes lourdes d'anticorps en Western Blot .....                                           | 202 |
| d) | Détection des U snRNA en gel dénaturant.....                                                                        | 203 |
| 5. | Cartographie et quantification des 2'-O-méthylation des U snRNA et ARNr par le protocole Illumina RiboMethSeq ..... | 203 |
| 6. | Analyse du cycle cellulaire.....                                                                                    | 203 |
| 7. | Analyse <i>in cellulo</i> de la localisation des protéines et des ARNs .....                                        | 204 |
| a) | Analyse <i>in cellulo</i> de la localisation des ARN par Hybridation In Situ de Fluorescence (RNA-FISH)<br>204      |     |
| b) | Analyse <i>in cellulo</i> de la localisation des protéines par Immunofluorescence (IF).....                         | 205 |
| 8. | Transfection des cellules .....                                                                                     | 205 |
| a) | Surexpression de plasmides .....                                                                                    | 205 |
| b) | Extinction de l'expression de la coiline ou de NOP58 .....                                                          | 206 |
| 9. | Analyse transcriptomique par puces à jonctions exon-exon.....                                                       | 206 |
|    | BIBLIOGRAPHIE.....                                                                                                  | 208 |

# INTRODUCTION

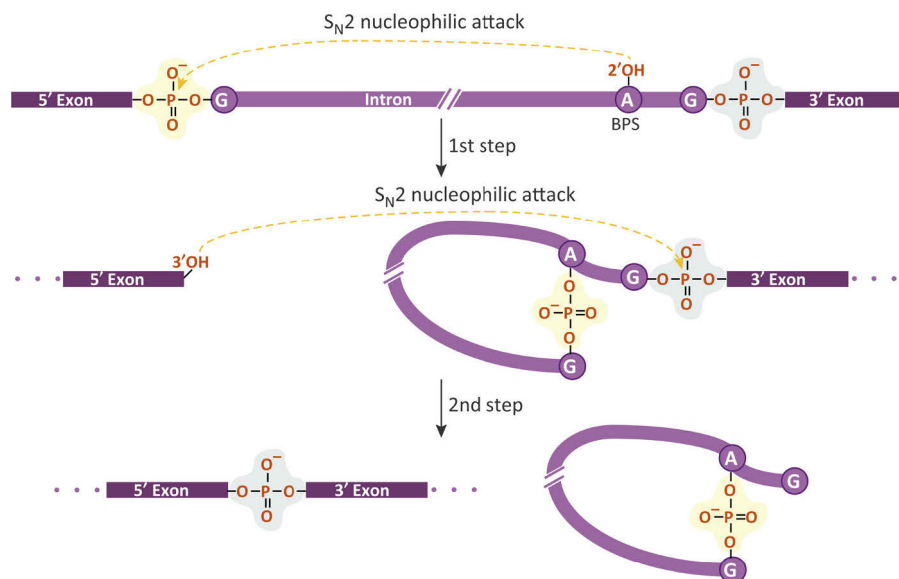


# I. L'épissage des pré-ARN chez l'Homme

## 1. L'épissage

### a) Définition

L'expression des gènes au sein des cellules eucaryotes implique de nombreux processus afin d'obtenir des transcrits ou des protéines fonctionnelles. La plupart des gènes eucaryotes transcrits par l'ARN polymérase II, qu'ils soient codants ou non codants, sont constitués de séquences spécifiques ou « exons » séparées les unes des autres par des séquences nommées « introns ». Le processus d'épissage permet l'excision des introns de l'ARN précurseur (pré-ARN) et la suture des exons (Moore and Sharp, 1993). Ce processus, co- ou post-transcriptionnel, permet d'aboutir à la production d'ARN matures à partir d'un transcrit dit primaire et repose sur deux réactions successives de trans-estérification (Figure I1). La plupart des événements d'épissage se déroulent co-transcriptionnellement.



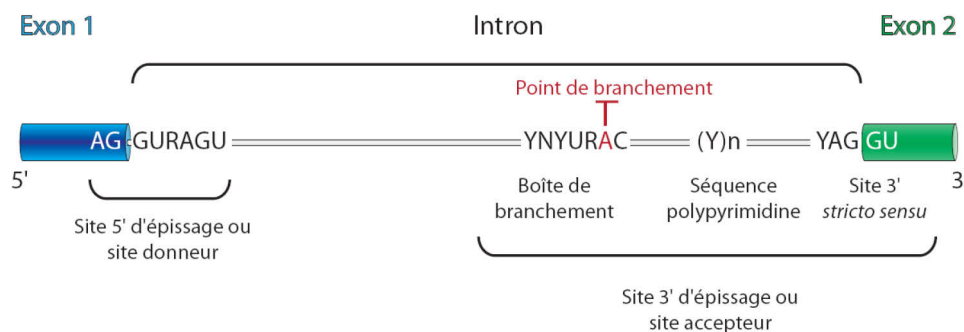
**Figure I1. Réactions de trans-estérification se déroulant au cours de l'épissage des pré-ARN (Papasaïkas et Valcárcel, 2016)**

L'épissage des pré-ARN se déroule en deux étapes de transestérifications successives. Au cours de la première, le groupement 2'-O-hydroxyl de l'adénosine (A) du point de branchement (branch-point site, BPS) réalise une attaque nucléophile sur le groupement phosphate entre l'exon 5' et l'intron, générant une liaison phosphodiester 2'-5' et conduisant à la formation d'un intermédiaire sous forme de lasso. Au cours de la seconde étape, le groupe 3'-OH libre de l'exon 5' attaque le groupement phosphate entre l'intron et l'exon 3', permettant de lier les deux exons et de libérer l'intron en forme de lasso.



### b) Séquences de définition des introns

L'information permettant de définir les séquences introniques au sein de chaque pré-ARN est portée par des séquences conservées appelées: sites 5' et 3' d'épissage (Figure I2). Chez les vertébrés, des séquences consensus ont pu être mises en évidence par comparaison de séquences (Chen and Cheng, 2012; Moore and Sharp, 1993). La séquence consensus au niveau du site 5' donneur d'épissage est AG/GURAGU, où le symbole « / » représente la jonction exon/intron et R, un résidu purine. Le dinucléotide GU de l'extrémité 5' de l'intron est très conservé et joue un rôle primordial dans l'efficacité d'épissage. Le site 3' d'épissage est constitué de 3 éléments directement impliqués dans la réaction d'épissage : i) le site 3' *stricto sensu*, situé à la jonction de l'intron et de l'exon et de consensus YAG/GU (où Y représente un résidu pyrimidine), ii) la boîte de branchement YNYURAC, située 18 à 40 nucléotides en amont du site 3' *stricto sensu* et contenant un résidu adénosine conservé appelé point de branchement, et enfin iii) une séquence polypyrimidine (chez les métazoaires), localisée entre les deux éléments cités précédemment et renfermant de longs enchaînements de pyrimidine (10 à 20 résidus C ou U). Cette séquence permet d'assurer la reconnaissance correcte du site 3' d'épissage.



**Figure I2. Séquences nécessaires pour l'épissage.** Représentation schématique de l'organisation d'un intron situé entre deux exons. Les séquences conservées jouant un rôle important dans la mise en place de la machinerie d'épissage sont indiquées (Moore and Sharp, 1993) (pour revue, Chen and Cheng, 2012; Hoskins and Moore, 2012). Y : Pyrimidine. R : Purine. N : un des 4 résidus A, C, G ou U.

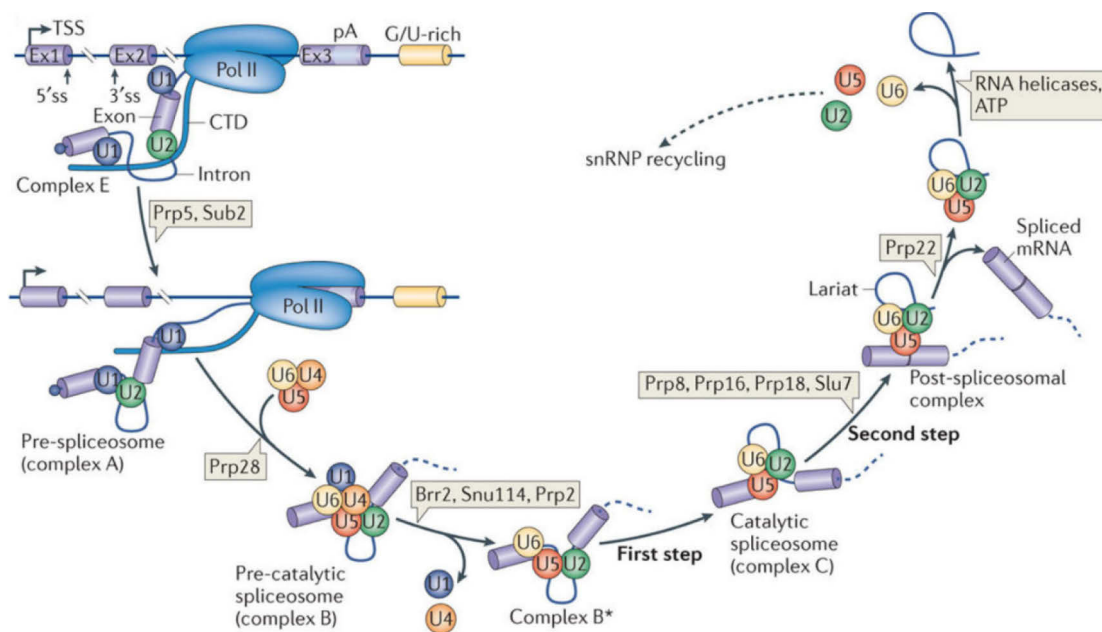
Les séquences consensus des sites d'épissage diffèrent entre deux grands groupes d'introns que sont les introns majeurs (présentés précédemment) et les introns mineurs (moins de 1% des introns). Le site 5' des introns mineurs est caractérisé par une séquence consensus /RUAUCCU, le site 3' *stricto sensu* par YA(C ou G)/N et la boîte de branchement par UCCUUA**A**CU (l'adénosine correspondant au point de branchement est indiqué en gras). Bien que les introns mineurs n'aient qu'une faible représentation dans le génome (882 introns mineurs pour environ 1 million d'introns référencés), ces derniers sont retrouvés au sein de beaucoup de gènes essentiels (réplication et réparation de l'ADN, transcription, maturation de l'ARN et traduction (Will and Lührmann, 2005)).

### c) Présentation générale de la machinerie d'épissage: le spliceosome

La grande diversité des introns, que ce soit en termes de taille ou de séquence implique la nécessité d'intervention d'une machinerie moléculaire élaborée et complexe de plusieurs mégadaltons appelée spliceosome. Les introns majeurs sont pris en charge par le spliceosome U2-dépendant (ou spliceosome majeur), alors que les introns mineurs sont reconnus et épissés par le spliceosome U12-dépendant (ou spliceosome mineur). Le spliceosome est caractérisé par le dynamisme de sa conformation et de sa composition qui lui confèrent une grande précision, ainsi qu'une importante flexibilité. Les deux types de spliceosome sont chacun constitués de 5 particules ribonucléoprotéiques (RNP) appelées UsnRNP (Uridin-rich small nuclear RNP). Alors que la snRNP U5 est commune aux 2 systèmes, le spliceosome U2-dépendant est spécifiquement composé des particules U1, U2, U4 et U6. Ces dernières sont respectivement remplacées par les particules U11, U12, U4atac et U6atac dans le spliceosome U12-dépendant. Chaque UsnRNP mature est composée d'un petit ARN nucléaire riche en Uridine (U snARN) caractérisé par un site Sm autour duquel 7 protéines Sm (B/B', D3, D2, D1, E, F et G) s'assemblent sous forme d'un anneau. Il est à noter que les ARN U6 et U6atac, qui sont transcrits par l'ARN pol III et non l'ARN pol II comme les autres UsnARN, interagissent chacun avec 7 protéines différentes dites LSm (LikeSm) (LSm 2-8). Chaque UsnRNP est également constituée d'un lot de protéines spliceosomales qui lui est propre.

#### d) Les mécanismes de la réaction d'épissage

L'assemblage du spliceosome sur son substrat pré-ARN se déroule *via* le recrutement séquentiel des UsnRNP et de nombreuses autres protéines et s'accompagne d'importants changements dans la composition et la conformation du complexe spliceosomal (Papasaïkas and Valcárcel, 2016; Wahl et al., 2009) (Figure I3).



**Figure I3. Schéma général de l'assemblage séquentiel du spliceosome au cours de la réaction d'épissage (Matera et Wang, 2014)**

L'assemblage du spliceosome débute par la reconnaissance des sites 5' et 3' d'épissages (5' et 3' splice site) par les snRNP U1 et U2, respectivement. Leur recrutement est médié par le domaine C-terminal (C-terminal domain (CTD)) de l'ARN Pol II. L'interaction de la snRNP U1 avec le 5' ss permet de former le complexe précoce E. Puis la snRNP U2 est recrutée pour former le complexe A qui interagit ensuite avec la tri-snRNP U4/U6-U5 (complexe B). Après de nombreux changements de conformation, les snRNP U1 et U4 quittent le complexe, ce qui permet de former un complexe activé B. L'intervention de l'hélicase Prp2 permet de former le complexe catalytiquement actif B\* qui catalyse la première réaction de transestérification conduisant à la formation du complexe C. Puis Prp22 ainsi que 16 autres hélicases interviennent pour permettre à la seconde réaction de se dérouler et conduisant au complexe post-spliceosomal.

Au cours des étapes précoces de l'assemblage du spliceosome, la snRNP U1 reconnaît le site 5' d'épissage, puis la snRNP U2 est recrutée au niveau de la boîte de branchement et permet

d'augmenter l'accessibilité de l'adénosine du point de branchement en la projetant à l'extérieur de l'hélice d'ARN. Par la suite, la tri-snRNP U4/U6.U5 préformée est recrutée sur le pré-ARN. De nombreux changements conformationnels conduisent alors le snARN U6, interagissant initialement avec le snARN U4, à interagir avec le site 5' d'épissage et le snARN U2, entraînant la libération des snRNP U1 et U4 du complexe spliceosomal et la formation d'un complexe catalytiquement activé (Chen and Cheng, 2012; Hoskins and Moore, 2012). Ce dernier est alors capable de catalyser la première réaction de trans-estérification caractérisée par l'attaque nucléophile du 2'-OH de l'adénosine du point de branchement sur le site 5' d'épissage. L'interaction entre les snARN U2 et U6 est essentielle dans la configuration du centre catalytique. Après cette première étape, de nouveaux réarrangements sont nécessaires afin de permettre le repositionnement de l'intermédiaire d'épissage et la poursuite vers la deuxième réaction de trans-estérification entre le 3'OH de l'exon 5' et le 5' phosphate de l'exon 3'. A la fin de la réaction d'épissage, l'intron est libéré sous forme de lasso et les UsnRNP se désengagent du complexe spliceosomal afin d'être recyclées pour de nouvelles réactions d'épissage ou d'être dégradées. Il est à noter que de nombreux ARN non codants sont introniques et dépendent donc de l'épissage et de la maturation du lasso intronique pour devenir fonctionnels. La directionnalité caractérisant la formation des différents complexes spliceosomaux et leurs changements de conformation au cours de l'épissage est conduite par l'activité d'hélicases ATP dépendantes. De plus, la reconnaissance des sites d'épissages par les snARN, le dynamisme des interactions ARN-ARN, ARN-protéine et protéine-protéine, ainsi que le positionnement correct du pré-ARN pour la catalyse sont assurés par un grand nombre de protéines appartenant ou non aux UsnRNP. Ainsi l'analyse protéomique de complexes spliceosomaux humains purifiés suggère que plus de 170 protéines sont associées au spliceosome à certaines étapes de l'épissage (Wahl et al., 2009).

## 2. L'épissage alternatif

### a) Définition et occurrence chez l'Homme

L'analyse de données de séquençage ARN à haut débit (RNA-Seq) du transcriptome indiquent qu'au moins 95% des gènes humains multi-exoniques produisent des transcrits qui sont épissés de façon alternative (Nilsen and Graveley, 2010; Pan et al., 2008; Wang et al., 2008). En effet, contrairement à l'épissage dit constitutif, l'épissage alternatif (EA) est un

processus au cours duquel des sites d'épissage sont différentiellement sélectionnés au sein du pré-ARN, aboutissant à la formation de différentes isoformes d'ARN matures constituées de combinaisons d'exons différentes. L'EA, retrouvé au sein des eucaryotes supérieurs et des virus, participe donc à la diversification du transcriptome et du protéome. Ainsi, les dernières données de GENCODE (v29, Ensembl95) reportent 38 400 gènes d'ARNnc et de pseudogènes pour plus de 123 000 transcrits. Il existe également environ 20 000 gènes codant des protéines pour plus de 83 000 protéines décrites à ce jour.

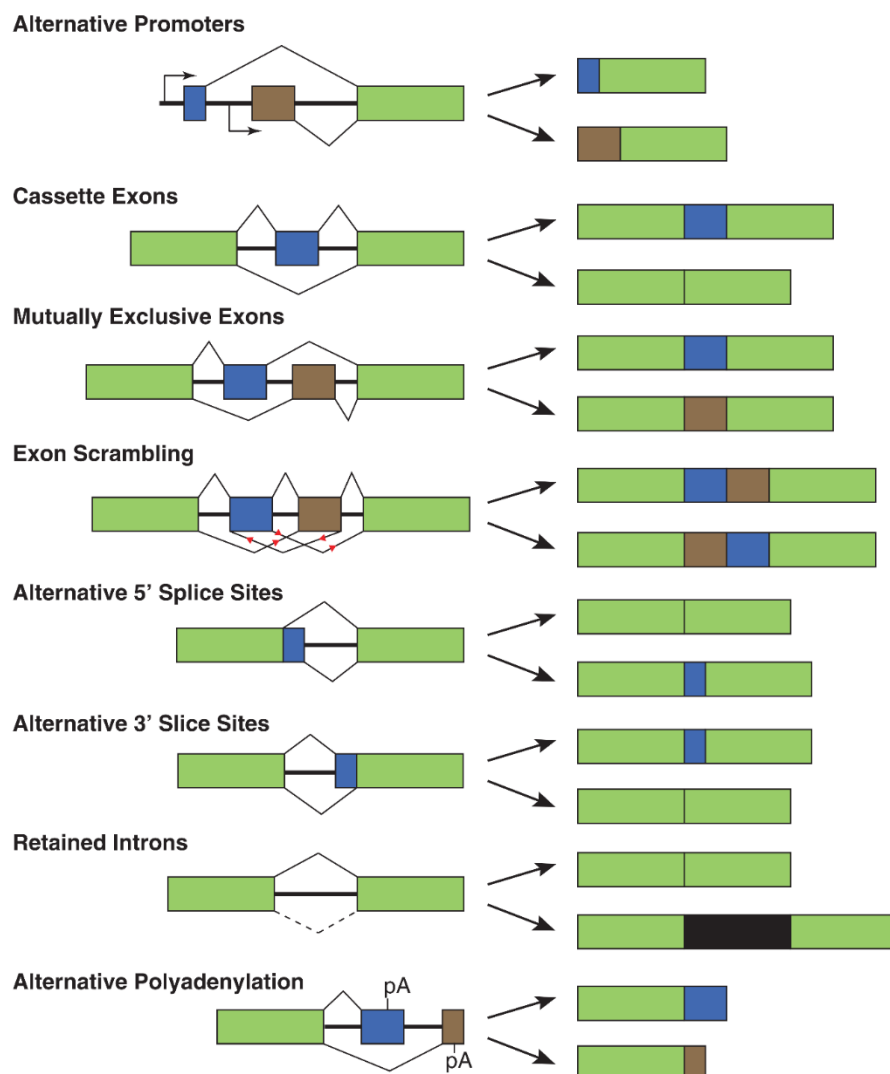
#### b) Les différents types d'épissage alternatif

L'épissage alternatif peut se dérouler via différents mécanismes et aboutir à la formation de différents transcrits matures. Il existe huit grands types d'évènements d'épissage alternatif différents (Figure I4) :

- Exon cassette : l'exon alternatif peut être soit inclus, soit exclus de l'ARN mature.
- Exons mutuellement exclusifs : la reconnaissance par le spliceosome d'un couple de sites 5' et 3' d'épissage au niveau d'un exon donné abolit la reconnaissance d'un autre couple spécifique d'un autre exon. Ainsi, ces exons ne sont jamais présents ensemble au sein du même ARN.
- Les sites 5' ou 3' alternatifs d'épissage : au niveau de certains exons, plusieurs sites 5' ou 3' d'épissage fonctionnels peuvent être présents et entrer ainsi en compétition. L'utilisation de ces sites conduit à l'inclusion ou l'exclusion d'une partie de l'exon et non de l'exon entier.
- La rétention d'intron : constitue le cas le plus rare, au cours duquel les sites 5' et 3' d'épissage ne sont pas reconnus et l'intron associé est alors maintenu au sein de l'ARN mature.
- Promoteurs alternatifs : plusieurs sites d'initiation alternatifs de la transcription peuvent être présents sur la séquence génomique. Ces différents sites d'initiation peuvent moduler soit l'inclusion d'exons en 5' mutuellement exclusifs, soit l'inclusion d'un exon ou d'un groupe d'exons localisés en aval du site d'initiation entraînant la formation d'une isoforme spécifique.

- Sites de polyadénylation alternatifs : différents exons d'un même gène peuvent contenir un site de polyadénylation. Par conséquent, la reconnaissance d'un premier site de polyadénylation peut empêcher l'inclusion d'un ou plusieurs exons en aval de ce site, conduisant à la formation d'isoformes avec des exons terminaux différents. Il peut également exister plusieurs sites de polyadénylation alternatifs au sein d'un même exon, dont l'utilisation peut conduire à la synthèse de différentes isoformes de transcrits matures.
- Exon scrambling : l'ordre des exons dans le transcrit mature n'est pas le même que dans le gène. Près de 90% des transcrits contenant des exons « scrambled » ne sont pas polyadénylés suggérant qu'ils correspondent à des formes circulaires de transcrits matures. Cependant, certains de ces transcrits sont polyadénylés et correspondraient à des formes linéaires d'ARN matures (Caldas et al., 1998).

Chez les eucaryotes, l'analyse des événements d'épissage alternatifs montre que les exons cassettes sont présents dans 49% des transcrits primaires humains. Les exons présentant des sites 5' ou 3' d'épissage alternatifs sont présents respectivement dans 26% et 25% des transcrits (de la Grange et al., 2005). Les sites de polyadénylation alternatifs sont présents dans 20% des pré-ARN (Tian et al., 2007) et les introns sont retenus dans 15% des pré-ARN (de la Grange et al., 2005). Il est à noter que différents types d'événements d'épissage alternatifs sont fréquemment combinés au sein d'un même pré-ARN, augmentant encore les possibilités de production d'isoformes de transcrits matures.



**Figure 14. Les différents modes d'épissage alternatif des pré-ARN.** Les exons constitutifs sont représentés en vert et les exons épissés alternativement sont représentés en bleu et marron. Les flèches coudées représentent des promoteurs alternatifs et la symbole "pA" représente des sites de polyadénylation alternatifs.

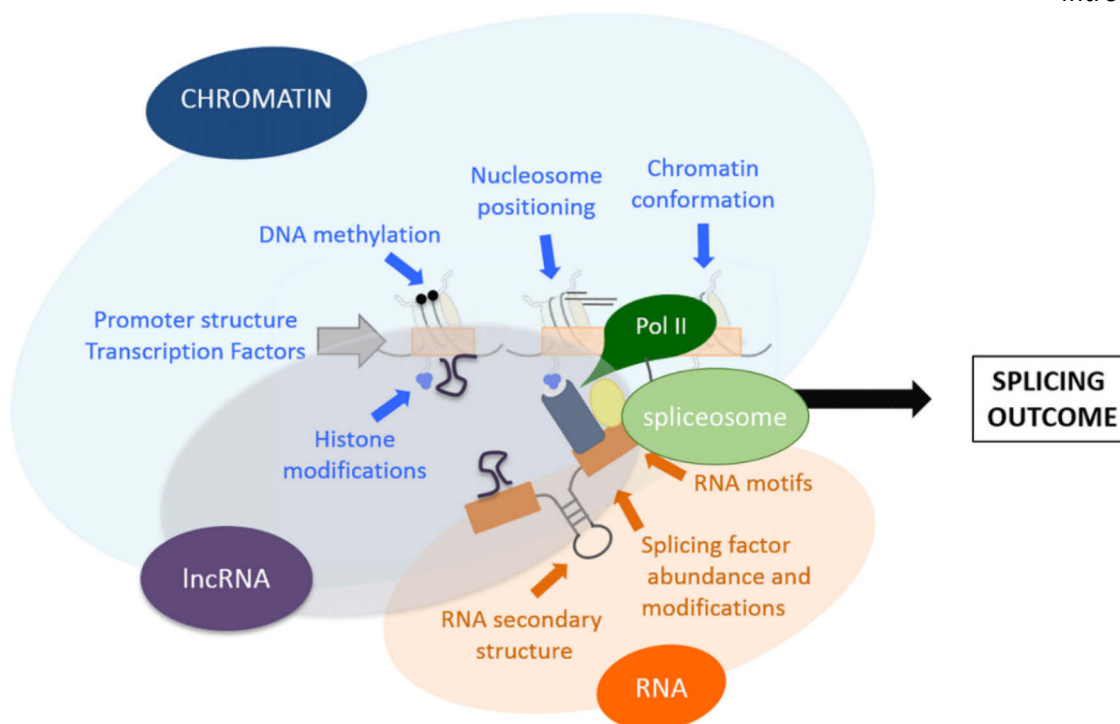
### c) Fonctions physiologiques de l'épissage alternatif

L'épissage alternatif permet de diversifier le génome et constitue un niveau de régulation central de l'expression des gènes. Il joue un rôle essentiel dans pratiquement tous les processus biologiques d'un organisme, y compris la régulation de la division et de la mort cellulaires, la différenciation des tissus dans l'embryon et l'organisme adulte, ainsi que dans la réponse cellulaire à divers facteurs environnementaux (microenvironnement, stress...) (Ramanouskaya and Grinev, 2017). Il joue également un rôle majeur dans la réaction immunitaire en régulant la production des immunoglobulines membranaires vs sécrétées. L'épissage alternatif est tellement répandu que certains considèrent que l'unité de l'information génétique n'est plus le gène, mais l'exon. De ce fait, un gène devrait plutôt être considéré comme une succession d'exons sélectionnés alternativement et permettant la production d'un ensemble de transcrits matures codant autant d'isoformes de protéines ou d'ARN. Ainsi, un gène peut être défini comme étant un message pouvant être interprété par la cellule de différentes manières pour assurer différentes fonctions biologiques selon ses conditions environnementales, son stade de développement et sa spécialisation ([http://www.crcl.fr/Portals/0/Equipes/EquipeAuboeuf/tutorial\\_site\\_internet.pdf](http://www.crcl.fr/Portals/0/Equipes/EquipeAuboeuf/tutorial_site_internet.pdf)). L'importance physiologique de l'épissage alternatif s'illustre notamment au travers du rôle qu'il joue dans le développement de nombreuses pathologies. Ainsi près de 45 % des polymorphismes simples de nucléotides (Single Nucleotide Polymorphisms, SNP) associés à des pathologies sont prédits affecter l'épissage (Krawczak et al., 2007; Lim et al., 2011; Mucaki and Rogan, 2019; Shirley et al., 2018).

### d) Les mécanismes de régulation de l'épissage alternatif

L'EA repose sur l'existence de sites 5' et 3' d'épissage sous-optimaux, c'est à dire présentant une séquence dégénérée par rapport aux séquences consensus. L'efficacité de reconnaissance de ces sites par le spliceosome peut varier considérablement et est sujette à régulation. La régulation de l'EA est un processus très complexe impliquant des mécanismes multiples et dont les règles (« splicing code ») demeurent à établir (Ramanouskaya and Grinev, 2017) (Figure I5).





**Figure I5. La régulation de l'épissage alternatif est un processus très complexe impliquant de nombreux niveaux de régulation** (adapté de Luco and Misteli, 2011)

Le profil d'épissage alternatif d'un pré-ARN est déterminé par tout un ensemble de paramètres incluant i) des éléments de régulation en *cis* recrutant des facteurs de régulation de l'épissage, des éléments de structure secondaire (représentés en orange), ii) les propriétés de la chromatine et de l'ARN polymérase (représentées en bleu), iii) l'expression de longs ARN non codants pouvant affecter l'état de la chromatine, la transcription, mais également le recrutement des facteurs de régulation de l'épissage (violet) et enfin, iv) la disponibilité et la fonctionnalité de la machinerie d'épissage (vert)

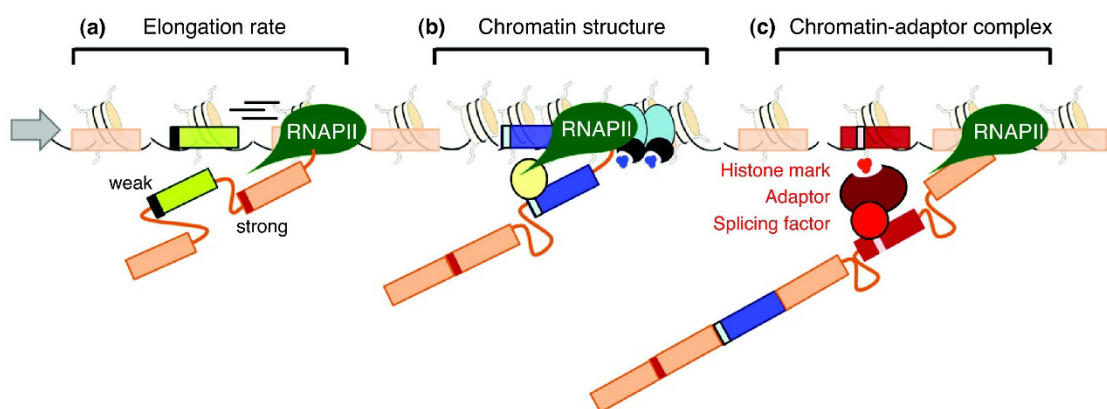
Parmi ces mécanismes, on peut citer par exemple le rôle d'éléments *cis* présents sur la séquence du pré-ARN. Parmi ces éléments, on distingue les séquences dites Exonic/Intronic Splicing Enhancers (E/ISE) et Exonic/intronic splicing silencers (E/ISS). Ces dernières sont reconnues par des facteurs d'épissage qui permettent respectivement de promouvoir ou de réprimer l'utilisation de sites d'épissages spécifiques. Les protéines SR (protéines riches en Sérine et en Arginine) sont des facteurs d'épissage souvent associés à la reconnaissance des ESE et ISE, alors que les protéines hnRNP sont plutôt associées à la reconnaissance des ESS et ISS (Long and Caceres, 2009; Martinez-Contreras et al., 2007). L'inclusion d'un exon alternatif dépend de l'antagonisme entre les différents facteurs d'épissage. La concentration active de ces facteurs peut elle-même être régulée au niveau de leur expression, de leur contenu en modifications post-

traductionnelles ou de leur localisation cellulaire. Par ailleurs, les séquences au sein du pré-ARN qui régulent la sélection de sites d'épissages influencent souvent la formation de structures secondaires. Les protéines régulant l'EA se fixent généralement sur l'ARN simple brin, et la formation de structures particulières (structure tige-boucle, G-quadruplexes, pseudonoeud...) peut masquer ces sites de fixation (Bartys et al., 2019; Hiller et al., 2007; Weldon et al., 2018). Il existe aussi certains exemples de métabolites capables de réguler l'EA et agissant, entre autres, via la modification de ce type de structures ARN (Cheah et al., 2007; Martínez-Montiel et al., 2016; Weldon et al., 2018).

L'épissage étant majoritairement co-transcriptionnel, le choix d'exons alternatifs peut également être régulé par la transcription. Deux modèles, non exclusifs, ont été proposés pour expliquer l'influence de la transcription sur la régulation de l'EA : le modèle de recrutement et le modèle cinétique. Dans le modèle de recrutement, l'ARN polymérase II est proposée interagir directement ou indirectement avec les composants du spliceosome afin de favoriser ou défavoriser la réaction d'épissage. Le domaine carboxy-terminal (CTD) de la grande sous-unité de l'ARN Pol II agirait comme une plateforme de recrutement de facteurs d'épissage (Hsin and Manley, 2012). Dans le modèle cinétique, c'est la vitesse d'élongation de la transcription qui est proposée affecter la cinétique de recrutement du spliceosome. Ainsi, une vitesse d'élongation lente tend à laisser un temps suffisant au spliceosome pour catalyser l'élimination de l'intron et abolit la possibilité de compétition entre sites d'épissage. A l'opposé, une vitesse d'élongation élevée génère la coexistence possible de plusieurs sites d'épissage au sein du transcrit primaire et établit une compétition entre des sites d'épissage de « force » différente et il en résulte de l'épissage alternatif (Figure I6a) (Dujardin et al., 2014; Schor et al., 2013).

La structure de la chromatine peut aussi impacter les décisions d'utilisation de sites d'épissages et de sélection des exons (Braunschweig et al., 2013). Ainsi les protéines histones qui structurent la chromatine sont différenciellement distribuées au niveau des frontières exon-intron, jouant un rôle dans la reconnaissance des exons. Par exemple, les exons présentant des sites d'épissage faibles entourés par de longs introns ont un niveau d'occupation par les nucléosomes plus élevé que les exons avec des sites d'épissages forts ou flanqués par de petits introns (Figure I6b) (Luco and Misteli, 2011). Par ailleurs, les modifications d'histones (H3K36me<sub>3</sub>,

H3 acetyl, H3K4me3, H3K9me<sub>3</sub>...) peuvent communiquer avec la machinerie d'épissage à travers la formation de complexes adaptateurs chromatine-épissage (Figure I6c) (Luco and Misteli, 2011). Ces complexes sont composés d'une protéine adaptatrice de liaison à la chromatine qui reconnaît les marques d'histones et recrute un facteur d'épissage qui va interagir avec le pré-ARN en cours de transcription. Différents complexes adaptateurs chromatine-épissage ont été identifiés à ce jour (H3K36me<sub>3</sub>/MRG15/PTB, H3K9me<sub>3</sub>/HP1/hnRNPs...) et certains impliquent des composants des UsnRNP et notamment de la snRNP U2 (H3K4me<sub>3</sub>/CHD1/U2snRNP ; H3acetyl/Gcn5/U2snRNP) (Gunderson and Johnson, 2009; Luco and Misteli, 2011; Sims et al., 2007).



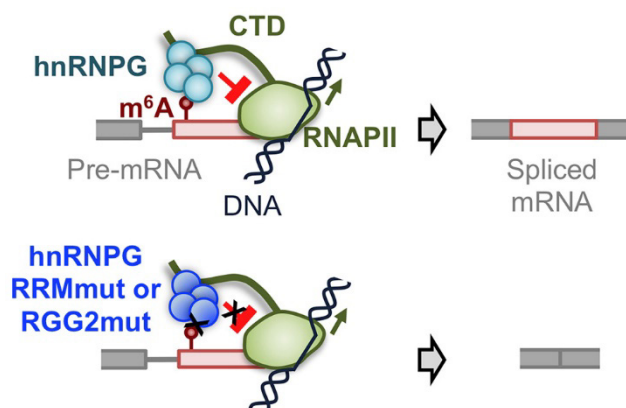
**Figure I6. Le rôle de la chromatine dans l'épissage alternatif.** (a) Le taux d'élongation de l'ARN Pol II affecte le recrutement du spliceosome. Une élongation rapide favorise l'inclusion d'exons avec un site 3' d'épissage fort. (b) Un changement dans la conformation de la chromatine, tel qu'une hétérochromatinisation localisée (ovales bleus et plus grande densité de nucléosomes) ralentit l'ARN Pol II et favorise le recrutement de facteurs d'épissage (ovale jaune) et l'inclusion d'un exon possédant un site 3' d'épissage faible (rectangle bleu). (c) Les modifications d'histones (petits cercles rouges) peuvent recruter directement des facteurs d'épissage via un système adaptateur de chromatine (ovales rouges) constitué d'une protéine se liant à la chromatine qui lit les marques d'histone et module le recrutement du facteur d'épissage sur le pré-ARNm (rectangle rouge) (tiré de Luco and Misteli, 2011)

Les ARNnc, qu'ils soient petits (<200 nt) ou longs (>200 nt), sont aussi des acteurs importants de la régulation de l'épissage alternatif (Romero-Barrios et al., 2018). Les microARN miR-10a/10b, miR-28 et miR-505 ciblent le facteur d'épissage SF2/ASF (SRSF1) et influencent l'épissage des exons régulés par SRSF1 (Meseguer et al., 2011; Verduci et al., 2010). De même,

chez la souris, l'augmentation post-natale de l'expression du microARN miR-23 diminue très fortement le taux des facteurs d'épissage CUGBP et ETR-3-like qui régulent près de la moitié des transitions d'épissage régulées au cours du développement au niveau du cœur (Kalsotra et al., 2010; Kucherenko and Shcherbata, 2018). Certains petits ARN nucléolaires (snoARN), ainsi que des fragments ARN dérivés de ces snoARN (psnoRNA), ont été proposés être exportés du nucléole vers le nucléoplasme et se fixer par complémentarité de séquence au niveau d'éléments régulateurs d'épissage au sein de différents pré-ARN, interférant ainsi avec le recrutement et la fixation de facteurs d'épissage au niveau de ces éléments de séquence (Kishore et al., 2010). Enfin, certains longs ARNnc (lncARN) ont également été associés à la régulation de l'EA. Trois principaux modes de régulation de l'EA par les lncARN ont été mis en évidence : (i) interaction avec des facteurs d'épissage ; (ii) formation de duplex ARN-ARN avec les pré-ARN sujets à EA et (iii) remodelage de la chromatine (Romero-Barrios et al., 2018). Ainsi, le lncARN NEAT1 s'associe avec la protéine SRp40 dont il régule l'état de phosphorylation et la localisation (Cooper et al., 2014; West et al., 2014), tandis que le lncARN MALAT1 régule la phosphorylation de nombreuses protéines SR au sein des speckles nucléaires et du nucléoplasme (Tripathi et al., 2010). L'ARN antisens FGFR2 (FGFR2-AS) induit quant à lui le recrutement en cis du complexe polycomb repressor PRC2 et de la protéine histone H3K36 déméthylase KMD2a. Il inhibe ainsi le recrutement du complexe adaptateur de la chromatine MGR15-PTB et favorise donc l'inclusion de l'exon IIIb (Gonzalez et al., 2015).

Plus récemment, le développement des recherches liées à l'épitranscriptome, a conduit à proposer que la présence de la modification N6-méthyladenine (m6A) pourrait également influencer l'épissage des pré-ARNm. En effet, de nombreux résidus m6A ont été détectés au sein des ARNm eucaryotes. Ces résidus sont formés co-transcriptionnellement par un complexe méthyltransférase (méthyltransférase-like 3 (METTL3) et 14 (METTL14)) qui est enrichi au sein des nuclear speckles et dont la déplétion affecte globalement l'épissage alternatif (Hausmann et al., 2016; Huang et al., 2019; Ke et al., 2017; Knuckles et al., 2017; Liu et al., 2014; Louloui et al., 2018; Ping et al., 2014). La présence des résidus m6A a été montrée pouvoir i) favoriser directement le recrutement de la protéine SRSF2 au niveau d'éléments (Zhao et al., 2014), ii) favoriser le recrutement de la protéine SRSF3 via le m6A reader YTHDC1 (Xiao et al., 2016) et iii)

influencer la processivité de l'ARN Pol II et donc l'épissage alternatif via la protéine hnRNP G, qui d'une part se fixe sur les résidus m6A et d'autre part interagit, grâce à son domaine RGG, avec le domaine CTD de l'ARN Pol II (Zhou et al., 2019) (Figure I7).



**Figure I7. Régulation de l'épissage alternatif co-transcriptionnel par les résidus m6A et la protéine hnRNP G.** (tiré de Zhou et al, 2019)

La formation de résidus m6A au sein du transcrit primaire en cours de transcription permet le recrutement de la protéine hnRNP G, qui via son domaine RGG interagit avec le domaine CTD de l'ARN pol II pour réguler négativement sa processivité et affecter ainsi le taux d'inclusion des exons alternatifs

Enfin, la régulation de la disponibilité des différents composants intervenant au cours de l'épissage est également un moyen efficace de contrôler les événements d'épissage constitutifs et alternatifs. La compartimentalisation des noyaux cellulaires en substructures non délimitées par des membranes et concentrant différents acteurs de l'épissage, telles que les corps de Cajal et les speckles nucléaires, constitue une architecture favorable à la régulation de cette disponibilité (Park et al., 2004). Ainsi, la diminution de l'expression des protéines SmB/B' dans les cellules humaines induit une réduction du niveau des UsnRNP et diminue spécifiquement l'inclusion de certaines d'exons alternatifs, avec peu d'effet sur le niveau d'épissage constitutif (Saltzman et al., 2011). Les processus contrôlant la formation, la dynamique et l'activité des corps de Cajal peuvent donc affecter la régulation de l'épissage alternatif d'un grand nombre de gènes et la coiline, une protéine structurante essentielle des corps de Cajal, constitue un facteur potentiellement déterminant dans la régulation de l'épissage alternatif.

### e) Epissage alternatif et stress

Toutes les cellules sont susceptibles de rencontrer des conditions externes ou internes perturbant leur homéostasie. Ces conditions peuvent correspondre à des variations de température, la présence de toxines, des dérégulations de la balance redox, un statut cancéreux, le vieillissement cellulaire... Nombre de ces conditions modifient ou affectent, en fonction de leur intensité, les fonctions cellulaires. Afin de contrebalancer ces perturbations, les cellules activent des programmes permettant de s'adapter aux nouvelles conditions ou d'activer leur mort programmée, par apoptose ou par nécrose, lorsque les dommages sont trop conséquents. Depuis de nombreuses années, il est connu que des conditions de stress sévères peuvent conduire à une inhibition de divers processus cellulaires, dont l'épissage. Ainsi, dès 1986, il a pu être établi que l'exposition à un stress thermique intense (43-46°C pendant 2h) inhibe la transcription et l'épissage d'une construction rapporteur *Adh* transfectée dans des cellules humaines (Yost and Lindquist, 1986). Quelques années plus tard, il a été montré que la formation de complexes spliceosomaux aberrants (notamment au niveau de la tri-snRNP U4/U6.U5) et la déphosphorylation de la protéine SRSF10 (SRp38), qui se comporte alors comme un répresseur de l'épissage en interagissant avec la protéine U1-70K et en inhibant le recrutement de la snRNP U1 au niveau du site 5' d'épissage, étaient à l'origine de cette inhibition (Shin et al., 2004; Utans et al., 1992). Cependant, ce n'est qu'en 2014, que l'équipe de Chris Burge a formellement démontré une inhibition générale de l'épissage, et plus particulièrement de l'épissage post-transcriptionnel, consécutive à l'exposition de fibroblastes de souris à un stress thermique sévère (Shalgi et al., 2014). Cette équipe a également mis en évidence une modulation globale de l'épissage alternatif lors d'un stress thermique modéré (40-42°C pendant 2h) (Shalgi et al., 2014). D'autres équipes ont observé un effet global sur l'épissage alternatif de différents stress (stress génotoxiques, stress osmotique, traitement par inhibiteur de topoisomérases ; exposition aux UV ou aux radiations ionisantes..) (Anufrieva et al., 2018; Carbonell et al., 2019; Kim Guisbert and Guisbert, 2017). De façon intéressante, il apparaît de plus en plus clairement que les modulations d'épissage alternatif observées en réponse aux stress font partie intégrante des programmes de réponse mis en place par les cellules et jouent un rôle essentiel dans l'adaptation et la survie des cellules à ces stress. Un argument fort en faveur de cette hypothèse est que les modulations

d'épissage alternatif concernant généralement des groupes de gènes différents selon la nature du stress. Par exemple, la privation de nutriments chez la levure altère l'épissage de la majorité des gènes codant les protéines ribosomales et contenant des introns, alors que la présence d'éthanol modifie l'épissage d'un autre ensemble de gènes (Pleiss et al., 2007). De même chez les plantes, des variations spécifiques d'épissage alternatif sont observées selon la nature du stress subit par les plantes avec, entre autres, des modulations très fortes de l'épissage alternatif des gènes de résistance aux pathogènes en réponse à des stress biotiques vs des stress abiotiques (Lenzken et al., 2013; Naro et al., 2015; Shang et al., 2017). Chez l'homme, enfin, des variations spécifiques de l'épissage alternatif des pré-ARN codant les protéines de la DNA Damage Response (DDR) sont spécifiquement observées en réponse à une exposition à des stress génotoxiques (Muñoz et al., 2009; Shkreta and Chabot, 2015). Les données non publiées de notre équipe sur les cellules HeLa et les fibroblastes humains vont également dans ce sens, puisque nous retrouvons très peu d'événements d'épissage alternatif communément régulés en réponse à un choc thermique modéré et à un stress oxydant.

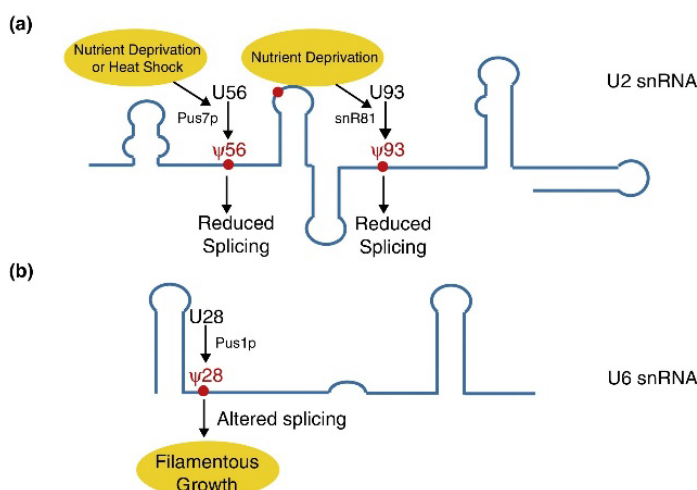
Les mécanismes à l'origine de la modulation de l'épissage alternatif par les stress ont encore été peu étudiés à ce jour. Initialement, l'idée prévalait que les stress sévères inhibaient l'épissage en altérant le fonctionnement de la machinerie d'épissage et que les conditions de stress modérées affectaient préférentiellement l'épissage alternatif via des effets spécifiques sur les régulateurs d'épissage. Cependant, les données obtenues au cours des dernières années ont clairement établi que la régulation de l'épissage alternatif pouvait également résulter d'effets directs des stress sur la machinerie d'épissage et que chacun des niveaux de régulation précédemment décrits peut contribuer à la modulation de l'EA par les stress. En effet, les modifications de l'activité des facteurs d'épissage, tels que les protéines SR, hnRNP, mais également des composants spliceosomaux constituent un des premiers niveaux de régulation de l'épissage alternatif induit par les stress. Ainsi, plusieurs études montrent l'impact des stress sur la localisation subcellulaire de facteurs d'épissage, avec notamment, la séquestration de facteurs protéiques dans le cytoplasme réduisant leur disponibilité pour la réaction nucléaire d'épissage. Par exemple, l'induction d'une ischémie au niveau du cerveau cause l'accumulation

cytoplasmique du facteur tra2- $\beta$ 1, ainsi que de plusieurs protéines SR. La délocalisation de ces facteurs stimule l'inclusion d'un exon alternatif au niveau des transcrits du gène ICH-1, favorisant ainsi la production d'une isoforme anti-apoptotique (Daoud et al., 2002). En réponse aux stress, certains facteurs d'épissage se relocalisent dans des granules de stress, comme la protéine Sam68 se relocalisant dans des granules de stress nucléaires après l'induction d'un choc thermique. D'autres facteurs se relocalisent au niveau de granules de stress cytoplasmique ou au niveau de foyers de réparation de l'ADN (Anderson and Kedersha, 2008; Biamonti and Caceres, 2009; Mirkin et al., 2008). Il est également intéressant de constater que selon la nature et l'intensité du stress, l'épissage d'un même pré-ARN peut être modulé par différents facteurs d'épissage. Par exemple, les protéines SRSF2 et SRSF1 sont impliquées dans la régulation d'un réseau d'évènements d'épissage alternatif au niveau de gènes liés à l'apoptose. SRSF2 régule notamment l'épissage des transcrits caspase 8, caspase 9, Bcl-x et c-flip (Edmond et al., 2011; Merdzhanova et al., 2008) et SRSF1 régule l'épissage d'autres transcrits parmi lesquels figurent également Bcl-x et caspase 9 (Moore et al., 2010).

La régulation de l'activité des facteurs d'épissage induite par les stress se déroule également au niveau post-traductionnel. Ainsi, comme mentionné précédemment, La déphosphorylation de la protéine SRSF10 (ou SRp38) joue par exemple un rôle central dans l'inhibition de l'épissage en réponse au choc thermique sévère en inhibant la fixation de la snRNP U1 au niveau des sites 5' d'épissage (Shin et al., 2004). De même, la modulation de la phosphorylation via p38 de la protéine SPSK1, un composant du spliceosome, est impliquée dans la régulation de l'épissage alternatif des cellules humaines en réponse au stress osmotique (Carbonell et al., 2019). Enfin, les stress peuvent moduler l'épissage alternatif via un effet sur la transcription ou sur l'état de la chromatine. Ainsi, en réponse à un stress métabolique, la protéine Parg1 est proposée réguler l'épissage alternatif via la parylation des histones et l'ouverture de la structure de la chromatine au niveau de certains exons alternatifs (Matveeva et al., 2019). Un dernier niveau de régulation proposé affecter la régulation stress-spécifique de l'épissage alternatif est le contenu en modifications post-transcriptionnelles à la fois des snRNA et des pré-ARN. En effet, chez la levure, l'induction de la formation de résidus pseudourine ( $\Psi$ ) aux positions



56 et 93 du snRNA U2 a été mise en évidence en réponse à une privation nutritionnelle (Wu et al., 2011) (Figure I8). Il est à noter que le résidu pseudouridine en position 56 est également induit suite à un stress thermique (Wu et al., 2011). L'introduction artificielle du résidu  $\Psi$ 93 dans le snRNA U2 de levure diminue l'efficacité de la réaction d'épissage (Wu et al., 2016), d'où l'idée que ces modifications inductibles pourraient permettre de réguler finement l'épissage alternatif afin de s'adapter aux stress (Figure I8). Des profils de pseudouridylation alternatifs des ARNm de mammifères ayant également été observés en réponse à un choc thermique ou un stress oxydant, il est également proposé que la modification inductible des ARNm participe à la régulation de l'épissage alternatif et/ou de la traduction des ARNm (Li et al., 2015; Zhao et al., 2018).



**Figure I8. Pseudouridylation stress-spécifique des UsnRNA de levure (tiré de Zhao et al., 2018)**

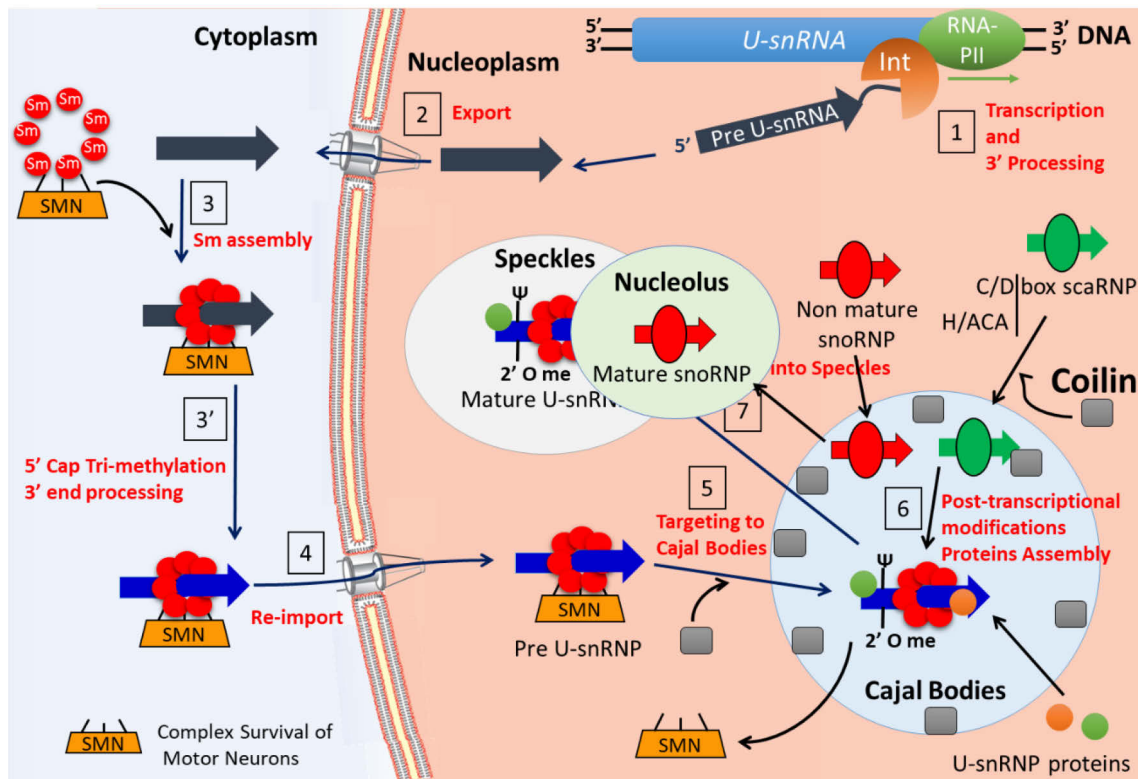
**(a)** La formation du résidu  $\Psi$ 56 au sein du snRNA U2 de levure, par l'ARN pseudouridine-synthase Pus7p, est induite en réponse à une privation nutritionnelle ou un choc thermique. Une privation nutritionnelle induit également la formation du résidu  $\Psi$ 93 au sein du même ARN, mais cette fois-ci via le système des ARN guide (snR81) **(b)** La formation du résidu  $\Psi$ 28 au sein du snRNA U6 de levure, par l'ARN pseudouridine-synthase Pus1p active la croissance filamentaire

## II. Biogenèse/Maturation du spliceosome

### 1. Biogenèse des U snRNPs

Chaque réaction d'épissage dépend de la fonctionnalité de l'ensemble des UsnRNP et de la qualité de l'assemblage des UsnRNP sur le pré-ARN. L'assemblage fidèle du spliceosome est un véritable challenge, impliquant de multiples réarrangements structuraux des ARN et protéines soulignant l'importance de la biogenèse de ces particules qui elle-même implique de multiples étapes de modifications des ARN et d'assemblage des protéines. Tout défaut dans la biogenèse du spliceosome peut avoir de graves conséquences sur l'intégrité et les fonctions cellulaires. Ainsi, à ce jour plus d'une dizaine de pathologies humaines différentes impliquant des défauts dans la biogenèse des UsnRNP ont été décrites : l'atrophie musculaire spinale (SMA, défaut du complexe SMN), la sclérose amyotrophique latérale (ALS, défaut protéine FUS), le syndrome cerebrocostomandibulaire (CCMS, défaut protéine SmB) (Coady and Lorson, 2011; Vazquez-Arango and O'Reilly, 2018; Zhang et al., 2008).

La biogenèse des UsnRNP Sm débute au sein du noyau où les UsnARN Sm sont transcrits par l'ARN Polymérase II puis maturés en 3' par le complexe Integrator (Figure I9). La formation en 5' d'une coiffe 7-méthyl guanosine (m7 G) autour de laquelle s'assemble un complexe d'export permet alors aux snARN Sm d'être exportés dans le cytoplasme. L'export coiffe-dépendant vers le cytoplasme des précurseurs U snARN dépend de CBC, ARS2 et PHAX, et de l'exportine CRM1 recrutée par PHAX (Hallais et al., 2013). Durant la phase cytoplasmique, les protéines Sm sont recrutées et s'assemblent autour du snARN grâce à un réseau complexe de facteurs d'assemblages se regroupant autour de la protéine Arginine Methyltransferase 5 (PRMT5) et du complexe SMN (Survival of Motor Neurons). Le cœur de protéines Sm constitue alors une plateforme de fixation pour le recrutement de la triméthylguanosine synthase 1 (TGS1) qui catalyse l'hyper-méthylation de la coiffe m7G en coiffe 2, 2, 7-triméthylguanosine (coiffe m3G) (Bordonné, 2000; Plessel et al., 1994). Le cœur de protéines Sm et la coiffe m3G constituent alors un signal bipartite pour l'assemblage d'un complexe permettant l'import de la pré-UsnRNP dans le noyau.



**Figure I9. Biogenèse des UsnRNP Sm (adapté de Kiss, 2014)** Les corps de Cajal jouent un rôle central dans l'assemblage et la maturation finale des UsnRNP. Après leur ré-importation dans le noyau, les UsnRNP immatures sont adressées dans les corps de Cajal où de nombreux facteurs participant à leur maturation sont concentrés, tels que les scaRNP à boîtes C/D et H/ACA qui ciblent les U snARN et induisent la formation de 2' O-méthylation (2' O-CH<sub>3</sub>) et de pseudouridylation (Ψ) respectivement. Certaines protéines qui constituent les UsnRNP sont également recrutées au sein de ces substructures pour être assemblées dans les particules spliceosomales. La formation initiale ou le recyclage du tri-snRNP U4/U6-U5 après une étape d'épissage est facilitée au sein des corps de Cajal grâce à l'augmentation de la concentration effective de chacun de ces constituants. Outre la biogenèse du spliceosome, les corps de Cajal participent aussi à la maturation des scaRNP et des snoRNP avant d'être respectivement concentrées dans ces substructures ou d'être adressées dans le nucléole. Le maintien de la structure des corps de Cajal est assuré par la coiline, une protéine qui participe aussi aux fonctions de ces substructures en assurant notamment le recrutement d'une grande partie de ces constituants tels que les sca/snoRNP (en interagissant avec les sca/snoARN) ou les UsnRNP (via l'interaction avec les protéines Sm et le complexe SMN).

La particule en cours de maturation est ensuite adressée dans des substructures spécifiques du noyau appelées « corps de Cajal » au sein desquelles le complexe SMN se désengage des pré-UsnRNP avant que les étapes finales de maturation ne se déroulent (Figure 19). Parmi ces étapes on peut distinguer des réarrangements structuraux, l'association de protéines spécifiques ou la formation de modifications post-transcriptionnelles au niveau des UsnARN. A la fin de leur maturation, les UsnRNP fonctionnelles sont stockées dans d'autres substructures appelées « Speckles Nucléaires » avant d'être sollicitées pour la réaction d'épissage. Il est à noter que les snRNP LSm U6 et U6atac ont une biogenèse différente puisqu'elles sont transcrites initialement par l'ARN Polymérase III et sont entièrement maturées au sein du noyau où elles transitent par les nucléoles avant d'atteindre les corps de Cajal.

## 2. Les modifications post-transcriptionnelles des UsnRNA

### a) Contenu en modifications post-transcriptionnelles des UsnRNA humains

Les UsnRNAs humains contiennent de nombreuses modifications post-transcriptionnelles (Karijovich and Yu, 2010; Krogh et al., 2016; Massenet and Branlant, 1999; Sloan and Bohnsack, 2018). On distingue généralement les modifications internes et les modifications de la coiffe ou proximales de la coiffe.

#### ➤ Modifications de la coiffe ou proximales de la coiffe

Tous les UsnRNA portent une coiffe à leur extrémité 5', mais la nature et le mécanisme de formation de cette coiffe diffèrent selon qu'il s'agisse d'un UsnRNA Sm ou LSm. Les snRNA U6 et U6atac, transcrits par l'ARN polymérase III (Pol III), portent une coiffe mpppG qui est formée par l'enzyme MEPCE (MEthylPhosphate Capping Enzyme) (Singh and Reddy, 1989). Tandis que tous les autres UsnRNA sont transcrits par l'ARN polymérase II et portent, comme nous venons juste de le voir, une coiffe de type 7-methyl guanosine ( $m^7G$ ) formée par l'action coordonnée de l'enzyme RNGTT (RNA 5' triphosphatase et RNA guanylyltransferase) et du complexe RNMT-RAM (RNA methyltransferase). Cette coiffe  $m^7G$  est ensuite convertie en 2,2,7-triméthylguanosine (TMG ou m3G) par la triméthylguanosine synthase 1 (TGS1). Les UsnRNA Sm, peuvent également être sujets à la 2'-O-méthylations des nucléotides proximaux de la coiffe (cap+1 et cap+2) par les

ARN méthyltransférases CMTR1 et CMTR2, respectivement (Smietanski et al., 2014; Werner et al., 2011).

➤ Modifications internes aux UsnRNA

Les modifications post-transcriptionnelles internes les plus fréquentes au sein des UsnRNA sont les 2'-O-méthylations du ribose (Am, Gm, Cm, Um) et les résidus pseudouridine ( $\Psi$ ) (figure I10). Les 2'-O-méthylations sont formées par transfert d'un groupement méthyle sur le 2'-OH du ribose par une méthyltransférase, tandis que les résidus pseudouridine dérivent des résidus uridine par une réaction d'isomérisation, ne nécessitant ni énergie, ni cofacteur, réalisée par une pseudouridine synthase (Arena et al., 1978; Johnson and Söll, 1970; Kammen et al., 1988; Nurse et al., 1995). La présence de résidus 2'-O-méthylés est proposée stabiliser les hélices d'ARN en favorisant les empilements de bases. Le résidu  $\Psi$  a la particularité de posséder deux groupements imino (NH) dans le cycle de la base, ce qui augmente les capacités d'interaction ARN-ARN et ARN-protéine des segments d'ARN renfermant ce résidu. Par ailleurs, le fait d'avoir une liaison carbone-carbone entre la base et le ribose, impose des contraintes conformationnelles au ribose qui favorisent la participation des segments contenant des résidus  $\Psi$  à la formation de double-hélices.

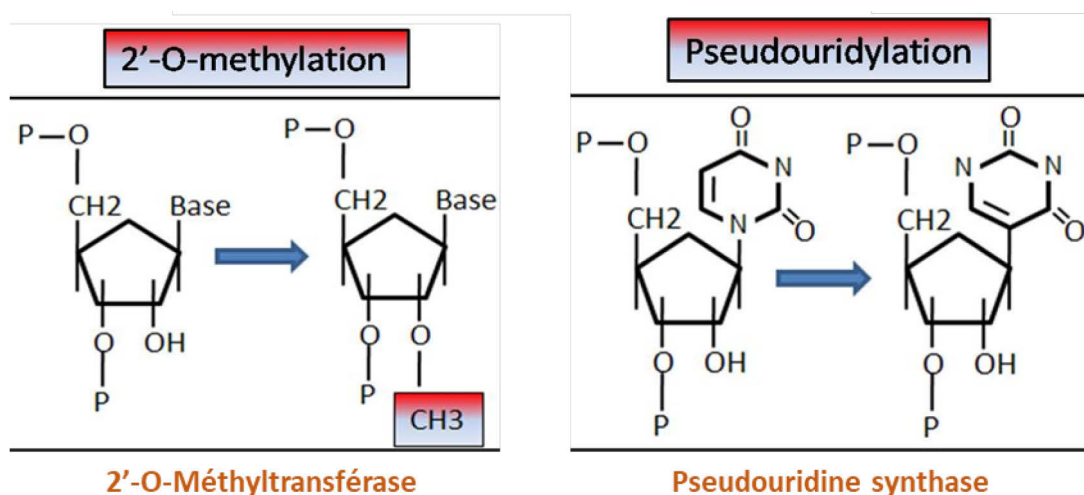


Figure I10. Structure chimique d'un résidu 2'-O-méthylé (Nm) et d'un résidu pseudouridine ( $\Psi$ ).

A l'heure actuelle 28 résidus  $\Psi$  et 24 résidus 2'-O-méthylés ont été recensés au sein des UsnRNA humains. Ces résidus sont répertoriés dans la Figure I11. On note qu'ils sont majoritairement présents au sein des UsnRNA majeurs, puisque seuls 4 résidus  $\Psi$  et 2 résidus 2'-O-méthylés sont présents dans les UsnRNA mineurs. Au sein des UsnRNA majeurs, ce sont les UsnRNA constituant le cœur catalytique du spliceosome, c'est-à-dire les snRNA U2, U5 et U6 qui contiennent le plus de modifications post-transcriptionnelles.

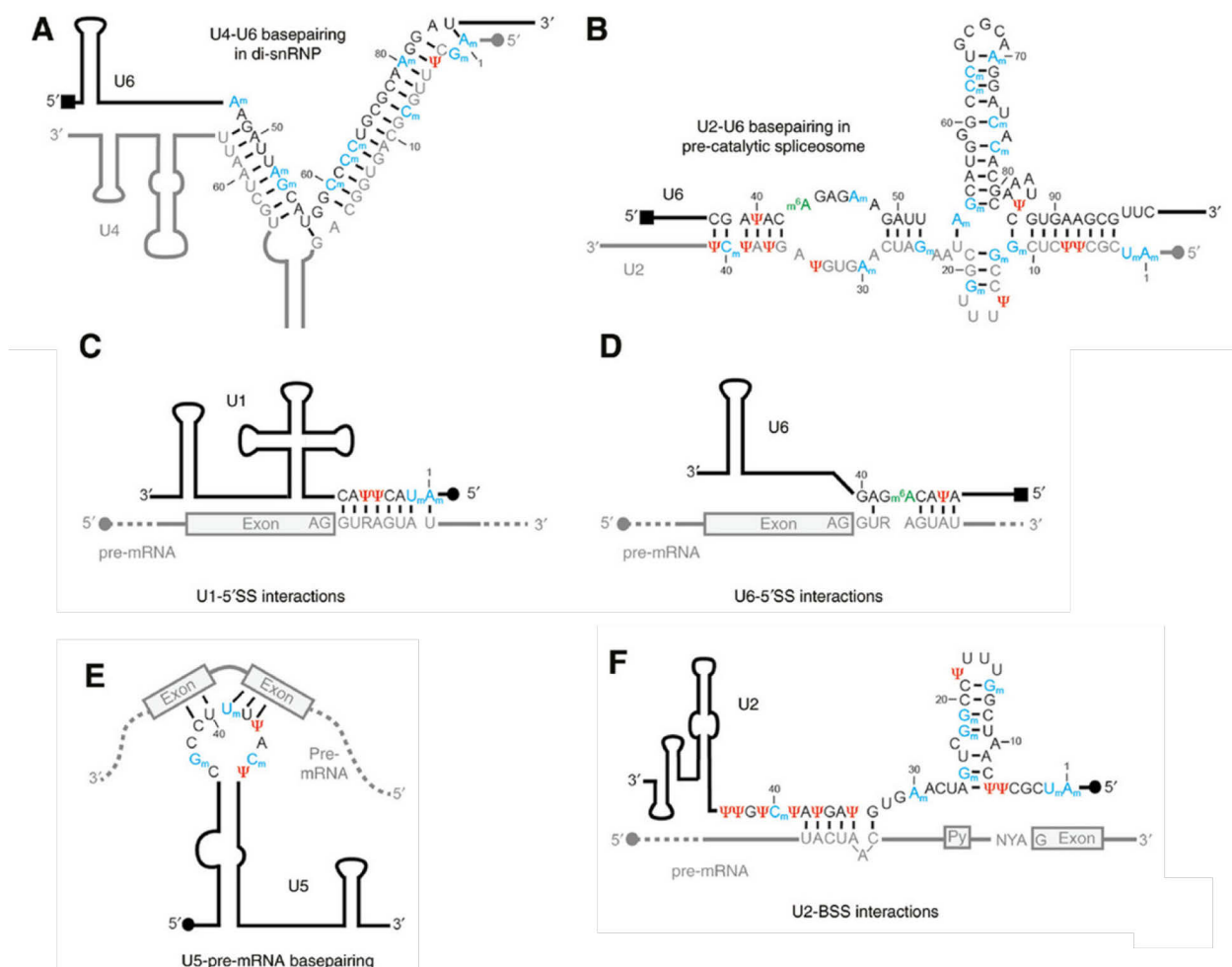
| snRNA  | Mod.                | Enzyme        | Guide RNA   |
|--------|---------------------|---------------|-------------|
| U1     | $\Psi$ 5            | Dyskerin      | SCARNA16    |
|        | $\Psi$ 6            | Dyskerin      | SCARNA18    |
|        | Am70                | Fibrillarin   | SCARNA7     |
| U2     | $\Psi$ 6            | ?             | ?           |
|        | $\Psi$ 7            | Dyskerin      | SCARNA14    |
|        | $\Psi$ 15           | ?             | ?           |
|        | Gm11                | Fibrillarin   | SCARNA2     |
|        | Gm12                | ?             | ?           |
|        | Gm19                | Fibrillarin   | SCARNA9     |
|        | Gm25                | Fibrillarin   | SCARNA2     |
|        | m <sup>6</sup> Am30 | Fibrillarin/? | SCARNA9/-   |
|        | $\Psi$ 34           | Dyskerin      | SCARNA8     |
|        | $\Psi$ 37           | Dyskerin      | SCARNA15    |
|        | $\Psi$ 39           | Dyskerin      | SCARNA4     |
|        | Cm40                | ?             | ?           |
|        | $\Psi$ 41           | Dyskerin      | SCARNA4     |
|        | $\Psi$ 43           | ?             | ?           |
|        | $\Psi$ 44           | Dyskerin      | SCARNA8     |
|        | Um47                | Fibrillarin   | SCARNA28    |
|        | $\Psi$ 54           | Dyskerin      | SCARNA13    |
|        | Cm61                |               | SCARNA2     |
|        | $\Psi$ 88           | ?             | ?           |
|        | $\Psi$ 89           | Dyskerin      | SCARNA1     |
|        | $\Psi$ 91           | ?             | ?           |
| U4     | $\Psi$ 4            | ?             | ?           |
|        | Cm8                 | Fibrillarin   | SCARNA17    |
|        | Am65                | Fibrillarin   | SCARNA5     |
|        | $\Psi$ 72           | ?             | ?           |
| U4     | $\Psi$ 79           | ?             | ?           |
|        | m <sup>6</sup> A100 | ?             | ?           |
| U5     | Gm37                | ?             | ?           |
|        | Um41                | Fibrillarin   | SCARNA5/6   |
|        | $\Psi$ 43           | Dyskerin      | SCARNA11    |
|        | Cm45                | Fibrillarin   | SCARNA10    |
|        | $\Psi$ 46           | Dyskerin      | SCARNA10/12 |
| U6     | $\Psi$ 53           | Dyskerin      | SCARNA13    |
|        | $\Psi$ 31           | Dyskerin      | SNORA7      |
|        | $\Psi$ 40           | Dyskerin      | SCARNA23/3  |
|        | m <sup>6</sup> A43  | METTL16       | -           |
|        | Am47                | Fibrillarin   | SNORD7      |
|        | Am53                | Fibrillarin   | SNORD8/9    |
|        | Gm54                | ?             | ?           |
|        | Cm60                | Fibrillarin   | SNORD67     |
|        | Cm62                | Fibrillarin   | SNORD94     |
|        | Cm63                | ?             | ?           |
|        | Am70                | ?             | ?           |
|        | m <sup>2</sup> G72  | ?             | ?           |
|        | Cm77                | Fibrillarin   | SNORD10     |
|        | $\Psi$ 86           | Dyskerin      | SNORA79     |
| U4atac | $\Psi$ 12           | ?             | ?           |
|        | Gm19                | ?             | ?           |
|        | $\Psi$ 83           | ?             | ?           |
| U6atac | $\Psi$ 19           | Dyskerin      | SCARNA21    |
|        | Gm22                | Fibrillarin   | SCARNA17    |
| U12    | $\Psi$ 28           | Dyskerin      | SCARNA20    |

**Figure I11. Listing des modifications post-transcriptionnelles internes identifiées au sein des UsnRNA humains (adapté de Bohnsack and Sloan, 2018).**

Le nom du UsnRNA, la nature de la modification ainsi que sa position, le système enzymatique et les éventuels ARN guides proposés participer à la formation des modifications sont indiqués.

**b) Fonction des modifications post-transcriptionnelles des UsnRNA**

Si l'on regarde plus en détails la localisation des modifications post-transcriptionnelles au sein des UsnRNA, on s'aperçoit qu'elles sont enrichies au sein de régions impliquées dans des interactions ARN/ARN (Figure I12). Ainsi, près de la moitié des nucléotides impliqués dans l'appariement de base entre les snRNA U4 et U6 au sein de la di-snRNP U4/U6 sont postranscriptionnellement modifiés (Figure I12A). La présence des modifications post-transcriptionnelles dans la région 5' du snRNA U2 est également essentielle pour la formation de la snRNP U2, suggérant un rôle de ces modifications également dans les interactions ARN/protéines (Donmez, 2004; Yu et al., 1998).



**Figure I12.** Les modifications post-transcriptionnelles internes des UsnRNA humains sont présentes au sein de segments d'ARN impliqués dans des interactions essentielles au déroulement de la réaction d'épissage (adapté de Bohnsack and Sloan, 2018). Les résidus  $\Psi$  sont indiqués en rouge, les résidus 2'-O-méthylés en bleu et m6A en vert

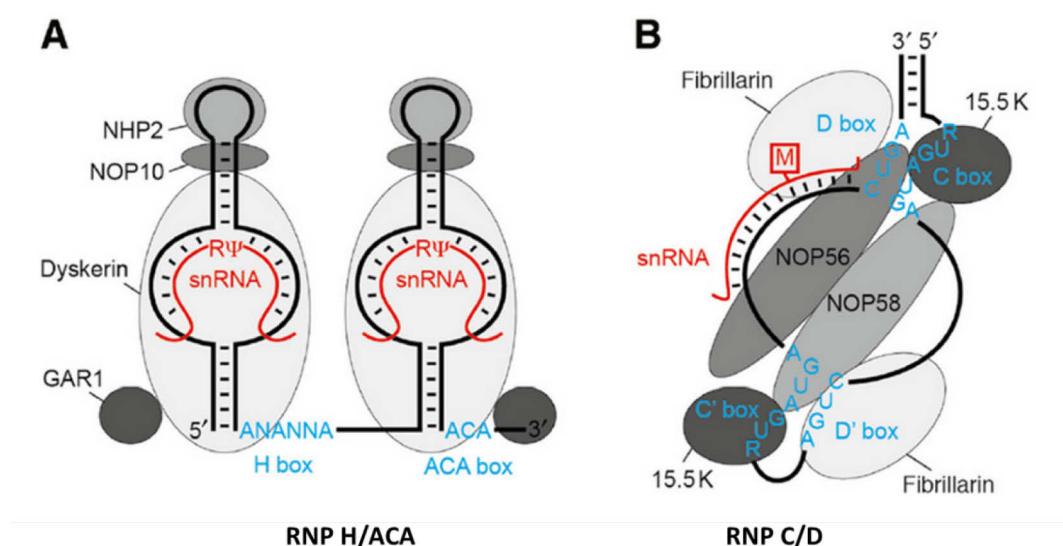
Concernant la fonction des modifications post-transcriptionnelles au sein des UsnRNA, il est également proposé, comme cela a été mentionné précédemment (figure I8), que la formation de certaines modifications au sein des UsnRNA, tels que les résidus  $\Psi$ 56 et  $\Psi$ 93 au sein du snRNA U2 de levure, puisse être inductible et réguler en fonction des conditions physiologiques de croissance l'efficacité et la spécificité de l'épissage et notamment de l'épissage alternatif (Wu et al., 2011, 2016).



## c) Formation des modifications post-transcriptionnelles des UsnRNA

A l'heure actuelle, chez l'homme, la formation de l'ensemble des résidus 2'-O-méthylés et pseudouridylés au sein des UsnRNA est assurée par des complexes ribonucléoprotéiques (RNP pour ribonucleoparticules) contenant une enzyme à activité catalytique (la méthyltransférase fibrillarine et la pseudouridine-synthase dyskérine, respectivement), associée à des partenaires protéiques et à un petit ARN guide qui va assurer la spécificité de la réaction (Sloan and Bohnsack, 2018). Ces RNP peuvent être présentes au sein des nucléoles, où elles assurent la modification des ARN ribosomaux et des snRNA U6 et U6atac, on parle alors de snoRNP (pour small nucleolar RNP) ou bien dans les corps de Cajal où les snRNA Sm sont modifiés, on parle alors de scaRNP (pour small cajal RNP). On distingue les sno-/scaRNP à boîtes C/D, qui catalysent la formation des 2'O méthylations, et les sno-/scaRNP à boîtes H/ACA, qui réalisent la conversion de résidus uridine en pseudouridine (figure I13).

Au sein des RNP à boîtes H/ACA, les ARN H/ACA sont associés à 4 protéines : NHP2, GAR1, NOP10 et l'ARN :  $\Psi$ -synthase Dyskérine (Figure I13) (Kiss et al., 2010; Massenet et al., 2016; Sloan and Bohnsack, 2018). Les ARN à boîtes H/ACA ont une structure particulière : ils se replient en



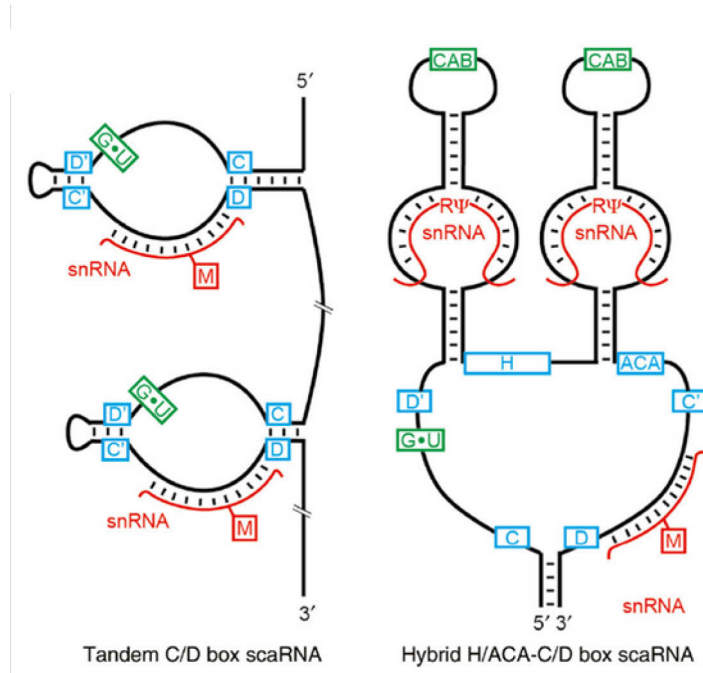
**Figure I13. Structure des ARN et des RNP à boîtes H/ACA (a) et C/D (B) (tiré de Bohnsack and Sloan, 2018)**

L'ARN cible est représenté en rouge, et les nucléotides modifiés sont indiqués par  $\Psi$  pour la pseudouridine et M pour la 2'-O-méthylation. R= purine

deux structures tige-boucles qui sont connectées par une région charnière qui contient une séquence conservée 5'-ANANNA-3', appelée "boîte H" (Figure I13). La seconde tige-boucle est suivie par un segment simple brin contenant une séquence de trois nucléotides 5'-ACA-3' appelée "boîte ACA", qui est localisée exactement trois nucléotides en amont de l'extrémité 3' de l'ARN. L'une ou les deux tige-boucles contiennent des boucles internes de 9 à 13 nucléotides sur chaque brin, qui forment les poches de pseudouridylation. Afin de guider la formation du résidu  $\Psi$ , l'ARN cible s'hybride aux poches de pseudouridylation formant deux hélices intermoléculaires avec l'ARN guide, ce qui permet d'exposer un di-nucléotide 5'-UN-3' non apparié au sommet de la poche. Le résidu U en 5' de ce di-nucléotide est l'uridine ciblée pour la modification.

Les ARN à boîtes C/D, quant à eux, sont associés aux protéines 15.5KD, NOP56, NOP58 et Fibrillarine, qui porte l'activité méthyltransférase (Figure I13) (Liang and Li, 2011; Massenet et al., 2016). Les ARN à boîtes C/D eucaryotes présentent tous une structure contenant deux courtes séquences conservées : la boîte C (« RUGAUGA » où R est une purine) située à l'extrémité 5', et la boîte D (« CUGA ») située à l'extrémité 3' de l'ARN. L'appariement entre les extrémités 5' et 3' de l'ARN permet de former le motif C/D qui a une structure particulière en K-turn (Watkins et al., 2000). Dans la plupart des ARN à boîtes C/D, il existe un second couple de boîtes de type C et D, de séquence moins conservée : le motif C'/D'. Une séquence de 10 à 21 nucléotides appelée "élément anti-sens", située en amont des boîtes D et D', est complémentaire à l'ARN cible (Figure I13). L'appariement de l'ARN guide avec son ARN cible conduit à la formation d'une double-hélice. La réaction de 2'-O-méthylation est catalysée par l'intervention de la 2'-O-méthylase Fibrillarine, qui cible de manière invariante le 5<sup>ème</sup> nucléotide apparié en amont des boîtes D et D'.

Il est à noter que certaines scaRNA peuvent comporter plusieurs motifs H/ACA ou C/D fonctionnels (scaRNA en tandem), mais également à la fois des motifs H/ACA et C/D (scaRNA hybride) chaque motif pouvant guider la formation d'un type de modifications (Figure I14). Les scaRNA ont aussi la particularité de posséder des séquences spécifiques nécessaires à leur rétention au sein des corps de Cajal : la boîte CAB pour les scaRNA H/ACA et un motif riche en G.U pour les scaRNA C/D, éléments qui fixent des partenaires protéiques spécifiques (TCAB1/WRAP53 et TDP43, respectivement) et sur lesquels je reviendrai un peu plus tard dans cette introduction.



**Figure I14. Structuration de scaRNA en tandem (plusieurs motifs h/aCA ou plusieurs motifs c/D) ou hybride (à la fois motifs H/ACA et C/D) (tiré de Bohnsack and Sloan, 2018)**

Les ARN cibles sont représentés en rouge, et les nucléotides modifiés sont indiqués par  $\Psi$  pour la pseudouridine et M pour la 2'-O-méthylation. R= purine

### III. Les corps de Cajal et les substructures nucléaires

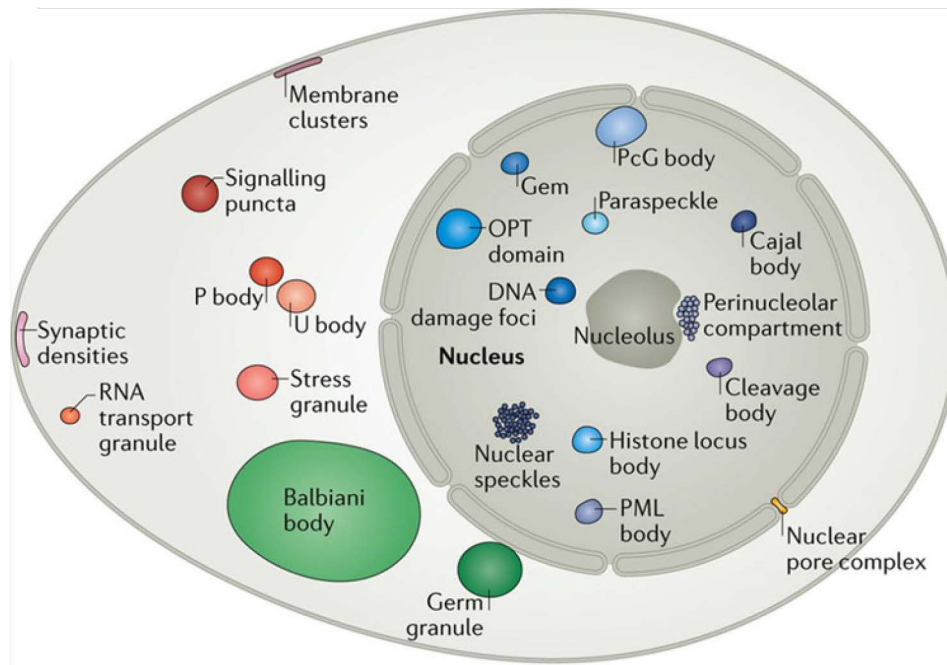
#### 1. Introduction aux substructures nucléaires sans membranes

##### a) Propriétés générales et fonctions des corps nucléaires

L'architecture des cellules de mammifères, et plus particulièrement du noyau de ces cellules, est extrêmement complexe car composée de nombreuses substructures sans membrane, parmi lesquelles les nucléoles, les corps de Cajal, les nuclear speckles, paraspeckles, PML bodies... (figure I15) (Banani et al., 2017).

Ces substructures compartimentent le noyau, permettant d'isoler des processus nucléaires distincts. Le fait de s'affranchir d'une bicouche lipidique permet aussi de se passer de la machinerie impliquée dans la construction et le maintien de la membrane, ainsi que dans le

transport de substances à travers la membrane. Les organelles sans membrane présentent plus de dynamisme dans leur composition et les échanges avec le milieu environnant, et augmente aussi les mécanismes de régulations potentiel. Les substructures nucléaires diffèrent des agrégats protéiques du fait qu'ils échangent de façon dynamique leurs composants avec le nucléoplasme. Les études sur la dynamique physico-chimique et la biogenèse des substructures nucléaires ont mis en évidence la similitude avec les mécanismes de séparation de phase entre deux liquides (Liquid-liquid phase separation ou LLPS) (Courchaine et al., 2016). Ce phénomène se caractérise par la séparation spontanée de différents solutés en phase liquides distinctes au sein d'un solvant correspondant dans ce cas au nucléoplasme.



**Figure I15. Illustration de la diversité des substructures nucléaires sans membranes dans les cellules de mammifères (tiré de Banani et al., 2017).**

De par leurs propriétés physico-chimiques et le rôle des ARN dans leur nucléation, les substructures nucléaires constituent des sites de biogenèse des RNP. Ainsi, les nucléoles se forment sur des loci actifs d'ADN ribosomique, et sont des sites de processing des pré-ARNr et de l'assemblage des sous-unités du ribosome (Boisvert et al., 2007; Falahati et al., 2016; Pederson, 2011). De façon similaire, les corps de Cajal se forment sur les loci actifs des snRNAs et constituent

les plateformes d'assemblage, de contrôle, de maturation et de recyclage des snRNPs (Frey et al., 1999; Machyna et al., 2013; Novotný et al., 2015; Staněk and Neugebauer, 2004). Les speckles nucléaires et paraspeckles concentrent des ARNm interagissant avec des protéines spécifiques, et se forment notamment sur deux longs ARNs non codants que sont MALAT1 et NEAT1 (Mao et al., 2011).

#### **b) Dynamique structurale des corps nucléaires**

Bien qu'elles ne soient pas délimitées par des membranes, les substructures nucléaires se caractérisent par la stabilité de leur structure, permettant de maintenir leur intégrité au cours d'une partie de la vie de la cellule. Au cours de leur existence, ces corps se déplacent au sein du noyau, avec une intensité inversement proportionnelle à leur taille (Dundr et al., 2007; Muratani et al., 2002; Platani et al., 2000). Ainsi, les structures les plus volumineuses, telles que le nucléole et les speckles nucléaires, présentent des mouvements très limités (Kruhlak et al., 2000; Olson and Dundr, 2005), alors que les plus petites, comme les corps de Cajal ou les corps nucléaires PML, se déplacent beaucoup plus au sein du nucléoplasme. De plus, ces derniers sont également caractérisés par une dynamique structurale importante de part des événements de fusion et de fission très fréquents (Chen et al., 2008; Dellaire et al., 2006; Dundr et al., 2007; Platani et al., 2002).

Bien que caractérisés par leur stabilité structurale, les corps nucléaires présentent une dynamique structurale intrinsèque élevée (Handwerger and Gall, 2006; Phair and Misteli, 2000). La biogenèse des substructures nucléaires est extrêmement importante dans la physiologie cellulaire. Par exemple, au cours de la division cellulaire, de nombreuses substructures tels que les CBs se désassemblent lors de l'entrée en mitose impliquant la nécessité de se ré-assembler correctement au sein des cellules filles pour assurer leur fonctionnalité. La formation des substructures nucléaires répond aussi à l'activité cellulaire, impliquant l'importance de leur dynamique au cours de la réponse cellulaire. Par exemple, les lymphocytes qui présentent un taux de production protéiques important et qui nécessitent donc beaucoup d'ARNr ont des nucléoles plus volumineux (Hernandez-Verdun, 2006). L'intégrité et la taille des substructures semblent être dépendantes des interactions protéines-protéines et protéines/ARN. Ces réseaux dynamiques assurent la stabilité de l'architecture et permettent le contrôle de la composition de

ces substructures en fonction des conditions et de l'état de la cellule. Le nombre, la forme et la taille d'une substructure donnée est surtout fonction du type cellulaire. La taille des substructures est contrôlée par la cinétique d'assemblage et de dissociation de ses composants (Dundr and Misteli, 2010).

#### c) Fonctions architecturales des ARNs

L'ARN est un composant majeur structurant les corps nucléaires. Chez les eucaryotes, de nombreuses substructures nucléaires sont structurées par des ARN dit « architecturaux » (arcRNA) (Chujo et al., 2016). Les molécules d'ARN présentent de nombreux avantages permettant d'initier la nucléation et de maintenir la structure des corps nucléaires. Les ARN ont la capacité de positionner de façon précise la nucléation des substructures afin de réguler localement certains événements nucléaires. En effet, les ARNs se concentrent souvent au niveau de leur site de transcription lorsque le niveau de transcription est élevé. De plus, certains transcrits contiennent des répétitions de séquences spécifiques. C'est notamment le cas pour les ARN satIII et hsr (heat shock RNA) omega contenant des répétitions en tandem de 158 et 280 nucléotides respectivement (Garbe et al., 1986; Jolly and Lakhotia, 2006). Les ARN satIII induisent la nucléation des granules de stress nucléaires notamment après l'induction d'un choc thermique. Les ARN hsr omega sont des lncRNA de 10kb structurant les omega speckles, des substructures nucléaires décrites chez la Drosophile. Dans la biogenèse des substructures nucléaires, les ARNs naissant peuvent servir de matrice pour leur nucléation comme c'est le cas pour les nucléoles se formant autour de l'ADNr, pour les histones locus bodies à proximité des clusters de gènes d'histones ou encore pour les granules de stress nucléaires au niveau des ARN sat III en réponse au choc thermique. Les transcrits naissant constituent de véritables plateformes retenant des protéines interagissant avec l'ARN, initiant ainsi la nucléation des corps nucléaires.

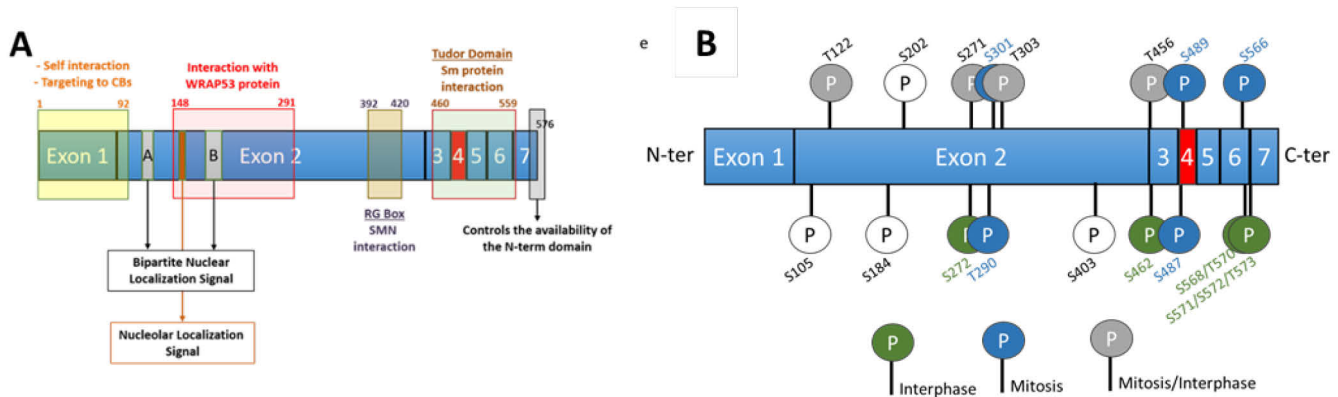
#### d) Importance des domaines protéiques de faible complexité

Les protéines contenant des domaines d'interaction à l'ARN et des domaines à faible complexité (Low Complexity Domains ou LCD) sont également des composants majeurs des substructures nucléaires impliqués dans leur biogenèse et leur dynamique. Toutes les structures contiennent au moins une protéine avec des LCD. Plusieurs études désignent les LCD comme des séquences similaires aux protéines de prion de levure (Alberti et al., 2009; Lancaster et al., 2014).

D'autres études intègrent également les domaines désordonnés parmi les LCD, et suggèrent leur importance dans la formation des systèmes basées sur la séparation de phase comme les granules de stress et les P bodies cytoplasmiques (Decker et al., 2007; Sun et al., 2011).

## 2. Présentation générale des corps de Cajal

Les corps de Cajal (Gall, 2000), sont des substructures nucléaires dynamiques de 0,2 à 1,0  $\mu\text{m}$  de diamètre qui, comme nous venons de le voir, jouent un rôle clef dans la biogenèse des snRNP. Ils sont aussi le lieu de maturation de RNP telles que les snoRNP (small nucleolar RNP), qui sont ensuite adressées dans les nucléoles où elles catalysent la formation de modifications post-transcriptionnelles au sein des ARN ribosomiaux et des UsnRNA U6 et U6atac, et les scaRNP (small cajal bodies RNP), qui sont des acteurs fondamentaux de la maturation des UsnRNP au sein des corps de Cajal (cette fonction a été détaillée dans les paragraphes précédents). La présence de substructures telles que les corps de Cajal permet d'augmenter de façon localisée la concentration effective des facteurs adéquats autour des U snRNP en cours de maturation. Ainsi, il a été établi que la formation de la tri-snRNP U4/U6-U5, que ce soit avant ou après une réaction d'épissage, se déroule environ 10 fois plus rapidement au sein des corps de Cajal que dans le reste du nucléoplasme (Novotný et al., 2011). La nucléation ciblée des corps de Cajal, ainsi que le recrutement des nombreux facteurs impliqués dans la maturation des U snRNPs sont essentiellement assurées par une protéine appelée coiline. En plus de permettre le maintien structural des corps de Cajal, la coiline participe à la régulation du flux des U snRNP au sein de ces substructures, en interagissant notamment avec le complexe SMN et les protéines Sm via ses domaines RG et Tudor respectivement (Toyota et al., 2010) (Figure I.16).



**Figure I16. Organisation fonctionnelle de la protéine coiline.** La partie N-terminale de la protéine lui permet de s'oligomériser pour induire la nucléation des corps de Cajal, et l'extrémité C-terminale contrôle la disponibilité de ce domaine. La protéine contient un signal de localisation nucléaire bipartite ainsi qu'un signal de localisation nucléolaire. Un domaine codé par une partie des exons 1 et 2, interagissant avec la protéine WRAP53 est également décrit. L'interaction de la coiline avec le complexe SMN se fait via un domaine riche en Arginine et en Glycine appelé RG box, alors que l'interaction avec les protéines Sm se fait via le domaine Tudor. La coiline est une phosphoprotéine dont le taux de phosphorylation est régulé via le cycle cellulaire.

### 3. La coiline, un facteur essentiel assurant la structure et la fonction des CBs

La coiline est une phosphoprotéine caractérisée par différents domaines fonctionnels (Figure I16 A) mais aussi par des séquences de faible complexité, à l'image de nombreuses protéines architecturales des substructures nucléaires. L'interactome de la coiline est très vaste et comprend des facteurs protéiques qui seront présentés par la suite mais aussi des ARNs tels que les sno- et les scaARN à boîtes C/D et H/ACA (Machyna et al., 2014). En interagissant avec les gènes U1 et U2, la coiline permettrait de contrôler au niveau de ces loci la formation de corps de Cajal concentrant des facteurs de transcription, et donc de participer à la régulation de leur transcription (Broome and Hebert, 2013). L'état de phosphorylation de la coiline est régulé par le cycle cellulaire et influence l'intégrité des corps de Cajal (Carrero et al., 2011), ainsi que ses fonctions en modifiant sa capacité d'interaction avec ses différents partenaires. Elle est caractérisée par un état hyperphosphorylé pendant la mitose corrélant avec le désassemblage des corps de Cajal (Carmo-Fonseca et al., 1993). A l'inverse, le statut hypophosphorylé de la



coiline est associé à sa relocalisation dans le nucléole ou sa périphérie (Tapia et al., 2010). Certaines observations décrivent aussi l'accumulation nucléolaire de la coiline dans un contexte de stress cellulaire (Trinkle-Mulcahy and Sleeman, 2016), la protéine contenant un site de localisation nucléolaire cryptique (NoLS) (Figure I16). La formation et la taille des corps de Cajal sont principalement liées à l'activité transcriptionnelle et métabolique des cellules, et ces substructures ne sont pas visibles dans tous les types cellulaires. La réduction de l'expression de la coiline affecte le nombre et la taille des corps de Cajal au sein des cellules où ces substructures sont visibles, et la déplétion de la protéine entraîne leur disparition (Strzelecka et al., 2010). Les effets de la réduction du niveau de coiline sur la viabilité cellulaire sont très différents d'un type cellulaire à l'autre. Ainsi sa déplétion dans des embryons de poissons zèbre conduit à une forte diminution de leur viabilité accompagnée d'une diminution du niveau de snRNP et d'ARNm épissés (Strzelecka et al., 2010). A l'inverse, chez les plantes et les drosophiles, aucun effet sur la viabilité ou la fertilité n'a pu être détecté (Collier et al., 2006; Liu et al., 2009). Il existe également des fractions de coiline nucléoplasmique et nucléolaire, non nucléées au sein des corps de Cajal et dont les fonctions demeurent encore mal connues. De plus, même si le domaine C-terminal de la protéine a été étudié par RMN (Résonance Magnétique Nucléaire) (Shanbhag et al., 2010) (voir Figure R29), la structure de la protéine seule ou interagissant avec ses partenaires n'est pas documentée.

La protéine WRAP53 (aussi appelée WDR79 ou TCAB1) qui interagit avec la coiline est un autre composant essentiel des corps de Cajal puisque sa déplétion induit leur démantèlement et empêche leur formation (Mahmoudi et al., 2010). Il est proposé que WRAP53 participe à l'adressage des U snRNP immatures au niveau des corps de Cajal *via* son interaction avec le complexe SMN (voir Figure R36).

#### 4. Fonctions des corps de Cajal dans la formation des modifications post-transcriptionnelles des U snARNs

Comme nous l'avons précédemment décrit, les corps de Cajal augmentent la concentration effective des facteurs de maturations nécessaires à la biogenèse des U snRNPs. Parmi ces facteurs, les scaRNPs jouent un rôle central dans la formation des modifications post-transcriptionnelles des U snARNs. Ainsi, une des fonctions les plus importantes des corps de Cajal dans la maturation du spliceosome est de concentrer efficacement les scaRNPs au niveau de leur substrat. Pourtant, chez les eucaryotes supérieurs, les snoARNs aussi se localisent au sein des CBs avant d'être adressés dans le nucléole (Lange et al., 1998a, 1999, Narayanan et al., 1999a, 1999b; Samarsky et al., 1998; Watkins et al., 2002) et il a été récemment montré que l'intéractome de la coiline chez l'Homme et la souris comprend des centaines de snoARNs (Machyna et al., 2014). De plus, certaines études montrent que les substrats potentiels des sca- et des snoARNs ne se localisent pas toujours au niveau des Corps de Cajal ou des nucléoles. Par exemple, il a été montré que l'ARN mgU6-77, identifié chez la souris et le Xénope, guide des 2'-O-méthylations du U6 et de l'ARNr 28S (Tycowski et al., 1998). De nombreuses prédictions mettent en évidence des scaARNs guidant potentiellement des modifications au sein des U snRNA et des rRNA (Deryusheva and Gall, 2013; Yuan et al., 2003). Ainsi, la fonction des sca- et snoRNP, impliquant notamment la reconnaissance du substrat adéquate, dépend étroitement de leur localisation.

##### a) Adressage des snoRNPs dans les nucléoles

Les CBs jouent un rôle important dans la biogenèse des snoRNPs. L'accumulation du snoARN U3 avec la protéine TGS1 (Trimethylguanosine synthase 1) dans les CBs suggère que l'hyperméthylation de la coiffe se déroule préférentiellement dans ces substructures (Verheggen et al., 2002). Il est également proposé que l'association de protéines cœurs telles que GAR1 et la fibrillarine soit facilitée dans les CBs notamment via leur interaction avec le complexe SMN (Darzacq et al., 2006; Verheggen et al., 2002; Watkins et al., 2004). Après leur passage dans les CBs, les snoRNP sont adressés vers les nucléoles, supposant l'existence de signaux de localisations dont la reconnaissance par des facteurs spécifiques est étroitement régulée au cours de la biogenèse des snoRNP. Des travaux ont montré que les motifs C/D et H/ACA sont des éléments

clés permettant de localiser les snoRNAs dans le nucléole et que les protéines associées à ces motifs jouent probablement un rôle. Les motifs à boîte C/D et H/ACA sont nécessaires et suffisants pour localiser les snoRNAs dans le nucléole chez la levure et les mammifères. Une étude a montré que les 4 protéines cœurs des boîtes C/D constituent collectivement un signal de localisation nucléolaire (Verheggen et al., 2001). Les motifs H et ACA sont des éléments de localisation nucléolaires essentiels, chaque box pouvant supporter individuellement la localisation nucléolaire des snoRNAs à boîte H/ACA (Lange et al., 1999). Au contraire, pour les snoRNAs C/D, les 2 boîtes sont nécessaires, chaque boîte n'étant individuellement pas capable de diriger le snoRNA dans le nucléole (Lange et al., 1998b; Narayanan et al., 1999b).

Il existe 2 isoformes de la TGS1. L'isoforme longue de TGS1 (LF) va dans le cytoplasme et dans les CBs, et est impliquée dans l'hyperméthylation de la coiffe des snRNAs. L'isoforme courte ne va que dans les CBs et modifie la coiffe des snoRNAs. L'extrémité N-terminale de la longue isoforme de TGS1 contient un domaine d'interaction avec le domaine C-terminal de SmB (snRNA). Le domaine C-terminal de TGS1 retrouvé dans les 2 isoformes contient l'activité méthylase ainsi qu'un site de fixation pour NOP56/NOP58 et la dyskérine. Hors ce domaine C-terminal contient un signal de localisation nucléolaire, c'est pourquoi il est proposé que TGS1 en interagissant avec ce domaine, masque ce signal et empêche l'adressage vers le nucléole. Des données *in vitro* montrent que CRM1 interagit avec l'isoforme LF de TGS1 induisant la dissociation de TGS1 du signal de localisation nucléolaire des protéines (Pradet-Balade et al., 2011). CRM1 serait donc un acteur indispensable dans le trafic nucléolaire des snoRNPs.

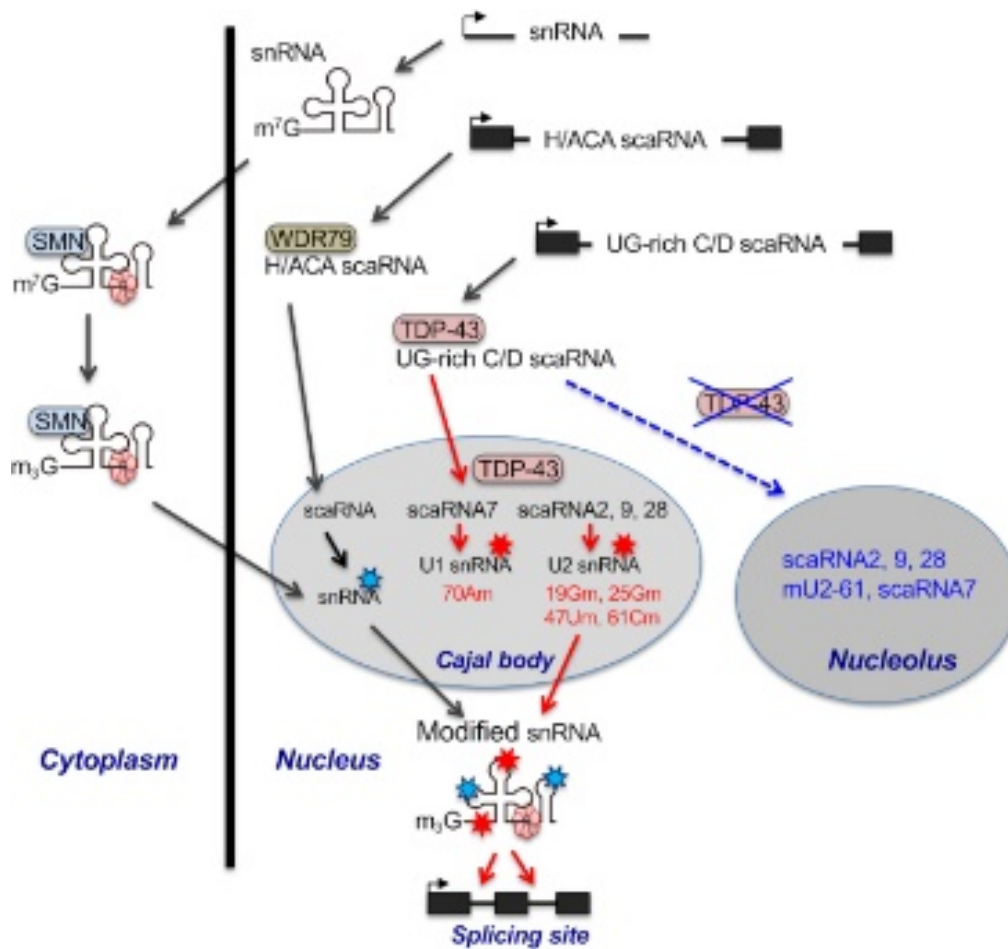
NOPP140 est une protéine surtout présente dans le nucléole mais également dans les CBs. Cette protéine interagit avec les snoRNPs C/D et H/ACA et l'expression de mutant dominant négatifs déplete les snoRNPs des CBs et des nucléoles (Isaac et al., 1998; Yang et al., 2000). Il a donc été proposé que NOPP140 participe au transport des snoRNPs entre les CBs et les nucléoles.

#### b) Rétention des scaRNA au sein des Corps de cajal

De nombreuses expériences de RNA-FISH montrent que les scaRNA ne sont pas observables hors des CBs. Dans ce sens, des expériences de fractionnement cellulaire et de purification des CBs avec le protocole du laboratoire du Pr. Lamond montrent que le scaARN U85

est co-purifié avec la fraction CB, mais difficilement détectable dans les fractions nucléoplasmiques et nucléolaires (Richard et al., 2003).

Les scaARNs contiennent les motifs classiques des boîtes C/D et H/ACA, mais vont également se distinguer par des motifs spécifiques impliqués dans leur localisation au sein des corps de Cajal. Les scaRNA à boîte H/ACA contiennent la séquence consensus ugAG appelée CAB box, localisée sur la boucle terminale des tiges/boucles en 3' et/ou en 5'. Cette séquence est essentielle pour la localisation au sein des CBs (Jády et al., 2004; Richard et al., 2003). Ce motif est impliqué dans l'interaction avec la protéine WRAP53 (TCAB1/WDR79), une protéine aussi indispensable dans la structure des CBs et l'import des U snRNP. Les motifs H/ACA sont aussi importants, suggérant la nécessité d'assembler les protéines cœurs pour interagir avec WRAP53. Le motif CAB est également essentiel pour l'adressage des hybrides C/D et H/ACA au sein des CBs. Les scaRNA à boîtes C/D ne possèdent pas de CAB box, mais WRAP53 est aussi impliquée dans leur adressage au sein des CBs, même si cette protéine reconnaît ces scaRNAs avec les tiges wobble GU/UG environ 20 fois moins que les H/ACA avec une CAB box. Il a récemment été montré que la protéine TDP-43 est l'acteur principal de l'adressage et la rétention des scaRNPs à boîte C/D dans les CBs (Izumikawa et al., 2019) (Figure I17). La protéine TDP-43 (Trans-activating response region DNA-binding protein 43 TARDBP, TDP-43) reconnaît les motifs riches en GU et autorégule son niveau cellulaire en se fixant aux motifs GU de son propre mRNA via un mécanisme de rétrocontrôle négatif (déstabilisation de l'ARN) (Avendaño-Vázquez et al., 2012; Ayala et al., 2011). Parmi les nombreuses fonctions décrites pour TDP-43, la plus caractérisée est la régulation de l'épissage des pré-mRNA en se fixant à des éléments cis- introniques et/ou avec l'action de facteurs d'épissage ou de RNA-binding proteins (Buratti et al., 2001; D'Ambrogio et al., 2009; Rabin et al., 2010; Sephton et al., 2011; Tollervey et al., 2011; Xiao et al., 2011). L'étude d'Izumikawa montre que TDP-43 et WDR79 sont importants pour la localisation des sca2 et sca28 dans les CBs, alors que TDP-43 seulement est nécessaire pour le restant des C/D contenant des motifs riches en GU. Il est à noter que certains scaRNA peuvent former des hybrides avec un ARN H/ACA contenant une CAB box insérée dans la boucle d'une boîte C/D.



**Figure I17. Rôles proposés de TDP-43 dans la modification post-transcriptionnelle des snARNs U1 et U2 via l'adressage des scaARNs à boîte C/D dans les corps de Cajal (Tirée de Izumikawa et al., 2019)**

## 5. Effets des stress sur les CBs

La nucléation et la structure des corps de Cajal dépendent de l'activité métabolique de la cellule et de ses besoins en ARN et RNP matures. Plusieurs conditions de stress perturbant la maturation des ARN et le métabolisme cellulaire altèrent ainsi la structure des corps de Cajal. La privation de nutriments dans des cellules HeLa résulte en une diminution du nombre de CBs, alors que ces dernières présentent des foci plus volumineux et plus intenses en microscopie de fluorescence lorsque les cellules prolifèrent (Andrade et al., 1993). Dans ce cas, le niveau d'expression de la coiline est au moins 2 fois plus élevé comparé à des cellules quiescentes.

L'induction d'un choc thermique sur des oocytes de *Xenopus* (2 à 4h de stress à 31.5-33.5°C au lieu de 18°C) induit la formation de micro-foci contenant de la coiline aussi appelés mini-CBs (Handwerger et al., 2002). Ces derniers n'apparaissent pas pendant le stress, mais plutôt de façon prolongée durant la période de recouvrement entre 6h et 18h après l'induction du stress. De plus, ces mini-CBs présentent la même composition que les CBs de plus grande taille. Les causes de formation de ces micro-foci demeurent encore inconnues. Classiquement, le choc thermique induit une répression générale de la transcription et de la traduction accompagnée de l'activation de l'expression des gènes de réponses au choc thermique (notamment les Heat Shock Proteins ou HSP). Cependant, dans cette étude, la formation des mini-CBs ne dépend pas du changement de l'activité transcriptionnelle, ni de la synthèse de nouvelles protéines ou d'un changement de l'état de phosphorylation de la coiline. Dans les cellules HeLa, bien que la quantité de coiline ne soit pas affectée, la taille et la quantité de CBs diminue significativement lors d'un choc thermique à 39°C alors que la croissance des cellules à des faibles températures (29-32°C) induit la formation de plus gros CBs (Carmo-Fonseca et al., 1993).

Tout comme le choc thermique, l'irradiation aux UV-C et le stress osmotique induit de façon réversible la redistribution de la coiline dans des microfoci. L'effet sur les CBs est maximal 6h après irradiation aux UV-C (Cioce et al., 2006), avec un recouvrement partiel vers 12h. Ce type de stress modifie la composition des complexes contenant la coiline, à l'image de la protéine SMN et du facteur SART3 associé au snRNP U6 qui ne se concentrent plus dans les CBs et se diffusent dans le nucléoplasme. Au contraire, la protéine fibrillarine reste concentrée dans des CBs résiduels. La redistribution nucléoplasmique de la coiline induite par l'irradiation aux UV-C est médiée par une augmentation de l'interaction de la coiline avec PA28 $\gamma$ , un régulateur du protéasome. De plus, un sous-ensemble de cellules irradiées présente également un pool de coiline nucléolaire de façon similaire à ce qui a été observé après inhibition de l'ARN Polymérase I et II (Carmo-Fonseca, 1992; Shav-Tal et al., 2005). La plupart des stress induisant une déstructuration des CBs s'accompagnent aussi d'une inhibition de la transcription. Dans le cadre de l'irradiation aux UV-C, la transcription des snARN est inhibée de façon p53-dépendante (Gridasova and Henry, 2005).

La structure des CBs est également affectée par les infections virales. Dans des cellules HeLa infectées par le virus HSV-1 (Herpes simplex virus type 1), l'expression de la protéine virale ICP40 (viral protein infected cell protein 0) induit la déstabilisation des centromères suivie de la redistribution de la coiline, de SMN et de la fibrillarine au niveau des centromères endommagés (Morency et al., 2007). Dans le contexte des « rosettes » induites par les adénovirus, la coiline et d'autres composants des CBs sont recrutés à la périphérie des centres de réplication virale, où ils sont impliqués dans le processing de transcrits viraux de la phase tardive (James et al., 2010). Selon différentes études, la redistribution de la coiline reflèterait le détournement de la machinerie impliquée dans le processing des transcrits dans les CBs par les virus.

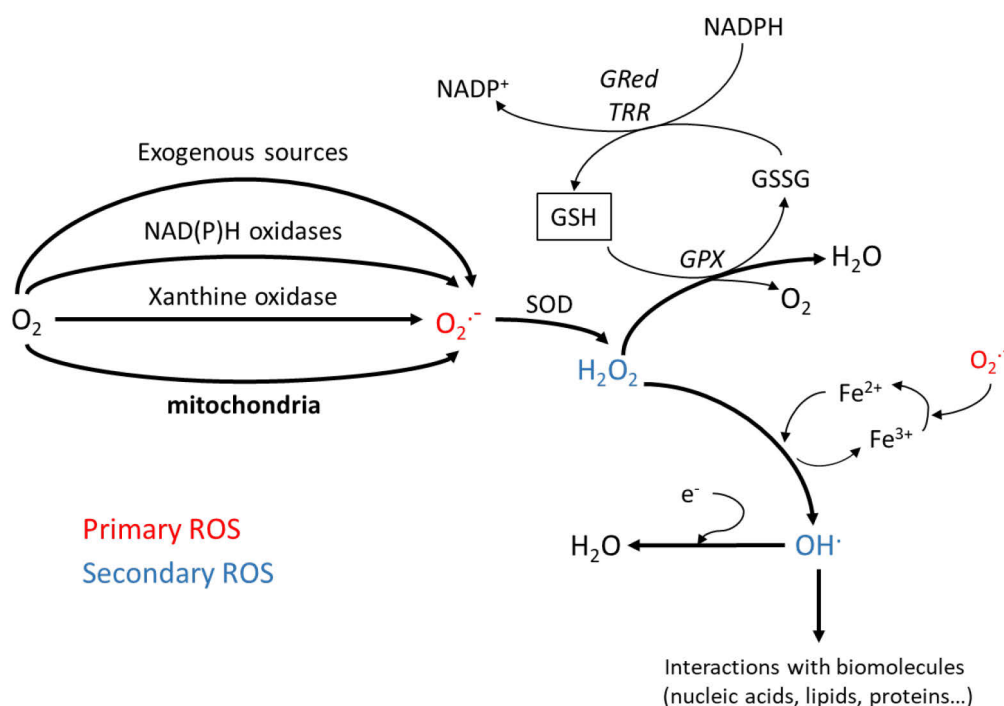
#### IV. Le stress oxydant

##### 1. Définition des ROS et fonctions physiologiques

###### a) Définition des ROS

Toutes les cellules du vivant sont soumises aux variations de facteurs biochimiques et physiques, générant tous types de stress et constituant un véritable défi pour le maintien de l'homéostasie cellulaire. Ainsi, les organismes concernés par le métabolisme aérobie sont exposés à des espèces réactives pouvant être classifiées en 4 grands groupes : les espèces réactives de l'oxygène (ERO ou ROS for Reactive Oxygen Species), de l'azote (ERN), du soufre (ERS) et du chlore (ERC). Les ROS, radicalaires ou non radicalaires sont les produits les plus abondants de ce type de métabolisme, et la réduction partielle de leur atome d'oxygène leur confère une forte capacité oxydante. Leur formation peut être issue de sources exogènes (rayonnements ultra-violet, radiations ionisantes, ozone, polluants, toxines environnementales, agents chimiques...) ou endogènes. Chez les mammifères, la production endogène de ROS se fait majoritairement *via* des systèmes enzymatiques tels que le complexe transmembranaire NADPH oxydase ou Nox (Nicotinamide Adenine Dinucleotide Phosphatonicotinamide oxydase), la xanthine oxydase (XO) et la chaîne mitochondriale de transport d'électrons (Figure I18). L'homéostasie redox cellulaire ou balance redox repose sur un équilibre entre la production des ROS et la diminution de leur niveau intracellulaire grâce à des mécanismes enzymatique de défenses anti-oxydantes (Superoxide dismutase (SOD),

Glutathione peroxidase (GPX), Catalases, Peroxiredoxine (PRX)) et des petites molécules anti-oxydantes (Glutathion, Vitamines E et C).



**Figure I18. Principales voies de production et de métabolisation des ROS chez les cellules eucaryotes**

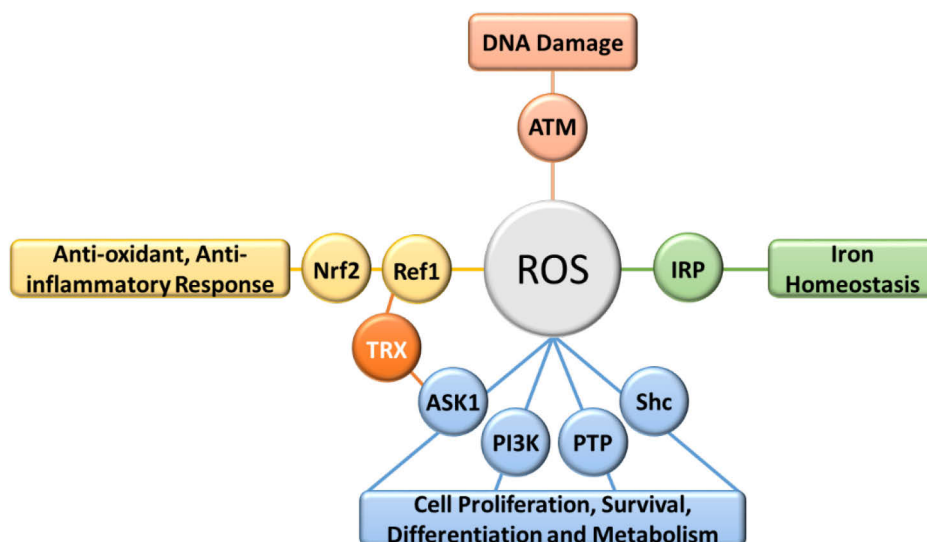
La formation du radical superoxyde ( $O_2^{\bullet-}$ ) peut être induite par des sources endogènes (NADPH oxydases, xanthine oxydase, chaîne respiratoire mitochondriale...) ou exogènes. Le radical est ensuite pris en charge par l'enzyme superoxyde dismutase (SOD) pour être transformé en peroxyde d'hydrogène ( $H_2O_2$ ). Cette molécule est piégée par la glutathione peroxidase (GPx) qui nécessite le glutathion réduit (GSH) en tant que donneur d'électron. Le glutathion oxydé (GSSG) est réduit en GSH par l'enzyme glutathione réductase (Gred) qui utilise le NADPH comme donneur d'électron. Certains métaux de transition ( $Fe^{2+}$ ,  $Cu^+$ , et autres) peuvent induire la formation du radical hydroxyl ( $OH^{\bullet}$ ), via la réaction de Fenton. Ce radical très réactif interagit rapidement avec diverses biomolécules situées à proximité, mais il peut aussi être réduit en  $H_2O$  non délétère pour la cellule.

## b) Fonctions physiologiques des ROS

Lorsque leur concentration est faible ou modérée, les ROS sont impliquées dans de nombreuses fonctions physiologiques activant certaines voies de signalisation cellulaire (Figure I19), si bien que leur disponibilité est étroitement régulée par les défenses antioxydantes. Les ROS



sont les acteurs d'une véritable signalisation redox à l'image des ROS dérivés du système Nox qui activent d'importantes voies de survie cellulaires comme la voie mitogen-activated protein kinase (MAPK) (Patten et al., 2010). Certaines ROS activent des facteurs de transcription tel que Nrf2 (Nuclear factor erythroid 2-related factor 2), un facteur modulant l'expression de plusieurs enzymes anti-oxydantes (SOD, PRX, GPX)(de Vries et al., 2008). L'activation des facteurs de transcriptions par les ROS peut réguler l'expression de plus de 500 gènes incluant des facteurs de croissance, des cytokines inflammatoires, des chemokines, des molécules régulant le cycle cellulaire et des molécules anti-inflammatoires. Le peroxyde d'hydrogène que nous avons utilisé au cours de cette étude pour générer le stress présente aussi des fonctions physiologiques importantes.  $H_2O_2$  est généré en réponse à différents stimuli incluant les cytokines et les facteurs de croissance. Il est aussi impliqué dans la régulation de processus biologiques comme le système immunitaire et le remodelage vasculaire chez les mammifères (Geiszt and Leto, 2004) ou la cloture stomatale et la croissance des racines de plantes (Foreman et al., 2003; Laloi et al., 2004).



**Figure I19. Voies de signalisations cellulaires majeures régulées par les ROS (adapté de Ray et al., 2012)**

Les ROS sont impliqués dans la régulation de plusieurs voies de signalisation via leur interaction avec des molécules clés de la signalisation. Les voies concernées sont liées à de nombreux processus cellulaires tels que la prolifération, le métabolisme, la différenciation et la survie (apoptosis signal-regulated kinase 1 (ASK1), PI3 kinase(PI3K), protein tyrosine phosphatase (PTP), et Src homology 2 domain-containing (Shc)). Ces voies sont également impliquées dans la réponse anti-oxydante et anti-inflammatoire (thioredoxin (TRX), redox-factor 1 (Ref-1), et NFE2-like 2 (Nrf-2)), l'homéostasie du fer (iron regulatory protein (IRP)) et la réponse aux dommages de l'ADN (ataxia-telangiectasia mutated (ATM)).

## 2. Définition d'un stress oxydant et dommages dans la cellule

### a) Définition et origines du stress oxydant

Dans la nature, peu d'organismes vivent dans des conditions physico-chimiques qui leurs sont parfaitement adaptées. En effet, les organismes sont soumis à des contraintes extérieures imposées par l'environnement dans lequel ils évoluent et auxquelles ils doivent s'adapter comme la température, le rayonnement UV et les espèces réactives de l'oxygène. On considère qu'une cellule est soumise à un stress cellulaire lorsqu'elle est soumise à une variation physico-chimique, imposée par son environnement, qui modifie son homéostasie en perturbant son métabolisme. Ainsi, lorsqu'elle est exposée à une accumulation excessive de ROS qui surpasse les défenses anti-oxydantes, la cellule est en état de stress oxydant. Ce déséquilibre de la balance redox peut être

dû à une forte augmentation de la production de ROS et/ou à un dysfonctionnement des systèmes anti-oxydants.

#### **b) Effets/dommage du stress oxydant**

A travers leurs propriétés chimiques, les ROS et plus particulièrement les espèces radicalaires sont très réactives. En effet, elles sont capables de réagir avec un grand nombre de biomolécules et sont impliquées dans la formation de différents dommages cellulaires comme l'oxydation des protéines, la peroxydation des lipides ou l'altération chimique des bases azotées des acides nucléiques. L'oxydation des protéines cause principalement la fragmentation du squelette peptidique, l'altération de leur conformation et la formation de liaisons intermoléculaires anormales (ponts disulfures, ponts dityrosine...) conduisant à la formation d'agrégats protéiques toxiques. Ces derniers sont notamment impliqués dans l'apparition de maladies neurodégénératives. La peroxydation des lipides affecte particulièrement les propriétés physico-chimiques des membranes cellulaires. Il existe de nombreux types de dommages de l'ADN causés par les ROS (Cooke et al., 2003). La plupart d'entre eux conduisent à des mutations où la formation de cassures des brins d'AND. Par exemple, les guanines sont oxydées par le radical  $\text{OH}\cdot$  en 8 oxoguanine (8-oxoG) qui forme une paire de base avec une Adénine plutôt qu'une Cytosine (Shibutani et al., 1991), conduisant à la mutation des G en T. Le radical hydroxyl réagit aussi avec les ions  $\text{H}^+$  de la partie ribose des acides nucléiques, induisant des changements de conformation qui provoquent des cassures simples brins de l'ADN. Les cassures double brins de l'ADN apparaissent alors au cours de la réplication de brins endommagés par les ROS. Les ROS provoquent des dommages à la cellule même lorsqu'elles sont à concentration physiologique. Cependant, dans de telles conditions, ces dommages sont efficacement pris en charges par la cellule et ne s'accumulent pas de façon néfaste pour cette dernière. Outre les dommages directs au niveau des biomolécules, la production non contrôlée de ROS conduit aussi à la dérégulation des processus physiologiques dans lesquels elles sont impliquées. Leur accumulation peut ainsi conduire à la dérégulation de certaines voies de signalisation (apoptose, prolifération et migration cellulaire) et à l'expression anormale de gènes. Ces effets combinés avec l'accumulation de cassures et de mutations de l'ADN, désignent le stress oxydant comme un acteur considérable dans le développement de cancers. De part ces nombreux effets, le stress oxydant est impliqué

dans le vieillissement cellulaire, ainsi que dans le développement d'un grand nombre de pathologies (pathologies cardiovasculaires et neurodégénératives, diabète, cancers...).

### 3. Mécanismes de réponses et d'adaptation au stress oxydant

Au cours de l'évolution, les cellules ont développé des mécanismes de réponses aux stress. Ces mécanismes permettent à la cellule de rétablir un équilibre physico-chimique lui permettant de préserver son intégrité. Afin de s'adapter aux variations de la balance redox, chaque cellule doit mettre en place les réponses adéquates, en fonction de l'intensité et de la durée du stress. Ces réponses à l'échelle cellulaire sont les conséquences d'un réseau complexe de régulations moléculaires qui nécessitent des « Senseurs » pour détecter les variations des conditions redox et l'activation d'« Effecteurs ». Les stress de faibles et moyennes intensités induisent généralement des réponses participant à l'adaptation au stress en régulant par exemple l'expression de gènes impliqués dans le métabolisme des ROS ou dans la réparation des biomolécules endommagées. Par exemple, le protéasome 20S ATP-indépendant participe à la dégradation des protéines oxydées (Jung et al., 2014). Les enzymes OGG1 et OGG2 (8-oxoguanine glycosylase 1 and 2) clivent les bases 8-oxoG interagissant respectivement avec une Cytosine ou une Adénine (Hazra et al., 1998). En contribuant à la diversité du transcriptome et du protéome, l'épissage alternatif constitue une cible stratégique dans la régulation des processus moléculaires qui participent à la réponse au stress oxydant.

### 4. Stress oxydant et épissage alternatif : état des connaissances

Comme pour beaucoup de stress, l'existence de variations isolées d'épissage alternatif en réponse au stress oxydant, a été rapportée depuis de nombreuses années. Mais ce n'est que très récemment que la première étude globale de l'effet d'un stress oxydant sur l'épissage alternatif a été publiée (Lenzken et al., 2017). Cette étude, qui a été publiée au cours de ma thèse, a été réalisée en utilisant des cellules de neuroblastomes humains SH-SY5Y comme modèle cellulaire et le stress oxydant a été effectué en cultivant ces cellules en présence de 2mM de paraquat pendant 48h ou 0,2 mM d'H<sub>2</sub>O<sub>2</sub> pendant 24h. De nombreuses variations d'épissage ont ainsi été mises en évidence avec une prévalence pour des événements de type « site de polyadénylation alternatifs » (APA, Alternative PolyAdenylation) (Lenzken et al., 2017). Il est à noter que les neuroblastomes sont des cellules très sensibles au stress oxydant du fait de l'importance du NO.

Ce type de stress peut par exemple moduler l'inclusion du dernier exon au niveau de l'extrémité 3' des transcrits. La protéine Brahma (BRM), la sous-unité ATPase du complexe de remodelage de la chromatine hSWI/SNF, interagit avec BRCA1/BARD1 qui ubiquitine une sous-unité du facteur CstF impliqué dans le processing des extrémités 3'. Cette ubiquitination permet d'inhiber le clivage du transcrit au niveau du site poly(A) proximal et favorise donc l'inclusion d'un exon distal terminal. Il a récemment été montré qu'au cours d'un stress oxydant, l'expression de BRM est diminuée et l'inhibition du clivage est levée permettant de favoriser l'inclusion de l'exon proximal (Fontana et al., 2017).

Concernant le stress oxydant, les effets sur les facteurs d'épissages ou les composants spliceosomaux ont été très peu documentés et, à ce jour, aucune analyse globale de l'effet d'un stress oxydant sur l'épissage alternatif des cellules humaines n'a été publiée. Enfin, aucune étude ne s'est portée sur l'impact du stress oxydant sur la biogenèse du spliceosome, sur le lien structure-fonction des corps de Cajal ou des Speckles nucléaires, ainsi que sur la disponibilité des sca/snoRNP qui dépend elle-même de l'efficacité d'épissage dans le cas des sca/snoARN introniques. Précédemment, il a été montré que des conditions spécifiques de stress tel que le rayonnement UV peuvent affecter l'intégrité de certains compartiments subcellulaires tels que les cors de cajal (Cioce et al., 2006), mais rien n'a été documenté concernant le stress oxydant. Globalement, les mécanismes moléculaires régissant la régulation de l'épissage alternatif et de la biogenèse du spliceosome lors d'un stress oxydant ne sont pas encore connus et demandent à être étudiés.

# OBJECTIFS DE THESE



## Objectifs de thèse

Lorsque j'ai débuté ma thèse, il apparaissait de plus en plus clairement que si les conditions de stress sévères inhibent le processus d'épissage, des conditions de stress plus modérées, elles, affectent principalement l'épissage alternatif. Les modulations d'épissage alternatif observées dans ce cas font partie intégrante des programmes de réponse mis en place par les cellules et jouent un rôle essentiel dans l'adaptation et la survie des cellules à ces stress. Cependant, les mécanismes moléculaires à l'origine de ces modulations d'épissage alternatif restaient encore largement incompris. Notre équipe avait débuté quelques années plus tôt une étude des mécanismes de régulation de l'épissage alternatif par le stress thermique modéré (Aigueperse et al., en préparation) et venait juste de commencer un projet sur le stress oxydant. Les données préliminaires de l'équipe (données transcriptomiques) sur ces deux stress suggéraient des mécanismes de régulation différents, mais impliquant des effets directs des stress sur la machinerie d'épissage.

En lien avec ces observations, l'objectif de ma thèse était de caractériser les effets du stress oxydant sur l'activité spliceosomale, mais également sur les RNP à boîtes C/D et H/ACA impliquées dans la biogenèse des UsnrNP spliceosomales.

Mon projet de thèse, tel que conçu initialement, se divise en 3 grands axes:

### **1) Etude des effets d'un stress oxydant sur les composants de la machinerie d'épissage et des RNP C/D et H/ACA.**

Ce 1er objectif avait pour but de caractériser les effets du stress oxydant sur les composants ARN et protéines de la machinerie d'épissage et les RNP C/D et H/ACA impliquées dans la modification post-transcriptionnelle des UsnARN. L'effet du stress sur la quantité des composants ARN et protéiques des UsnrNP spliceosomales et des RNP à boîtes C/D et H/ACA, dont des variations ont été détectées par l'étude transcriptomique, devait être évalué par RT-qPCR et Western blot, respectivement à différents temps après induction du stress (phase précoce ou tardive de réponse). La localisation des ARN et protéines devait être analysée par des expériences de RNA-FISH et d'immunofluorescence (IF) selon le protocole établi et publié par notre équipe (Vautrot et al., 2015). Enfin, l'intégrité de sous-structures nucléaires impliquées dans l'épissage, telles que



les corps de Cajal ou les speckles nucléaires devait également être étudiée analysée immunofluorescence.

## **2) Etude des conséquences de ces effets sur la fonctionnalité de la machinerie d'épissage et des RNP C/D et H/ACA**

Mon deuxième objectif avait pour but d'analyser les conséquences des effets observés dans le premier axe sur la fonctionnalité du spliceosome et des RNP C/D et H/ACA. La fonctionnalité des spliceosomes majeurs et mineurs devait être testée par l'emploi de minigènes exprimant une protéine luciférase firefly dont l'expression est directement corrélée à l'épissage d'un intron majeur ou mineur artificiellement inséré au sein de sa phase codante. Les effets du stress sur l'assemblage des spliceosomes majeurs et mineurs devaient être analysés par séparation sur gel natif des complexes spliceosomaux. Par ailleurs, la fonctionnalité des scaRNP à boîtes C/D et H/ACA devait être évaluée par l'analyse du taux des modifications post-transcriptionnelles (résidus pseudourines et 2'-O-méthylations du ribose) au sein des UsnARN. La composition en protéines des RNP dont la localisation ou l'expression de certains composants change au cours de la réponse au stress devait être analysée par des expériences de chromatographie d'affinité couplée à des expériences d'immunoprécipitation.

## **3) Etude du rôle de ces effets dans l'adaptation des cellules humaines au stress oxydant.**

Mon 3ème objectif visait à étudier la contribution de ces effets dans l'adaptation des cellules humaines au stress oxydant. La contribution des facteurs dont des variations du niveau d'expression, de la localisation, ou la fonction auront été observées en réponse au stress oxydant devaient être éteints par ARN interférence ou par utilisation d'oligonucléotides LNA Gapmers. Ces oligonucléotides induisent très spécifiquement et efficacement la dégradation d'ARN nucléaires par formation d'un hybride ARN/ADN et dégradation de l'ARN par l'activité RNase H cellulaire. L'effet de l'extinction de ces facteurs sur les modulations de l'épissage alternatif au cours de la réponse au stress oxydant devait être analysé par RT-PCR et RT-qPCR. Leur effet sur les capacités de réponse de la cellule (viabilité, niveau de ros, cycle cellulaire, etc..) devait également être analysé.

Au cours des 3 années qu’a duré ma thèse, je me suis principalement focalisé sur le premier et le second objectif, je n’ai malheureusement pas eu le temps d’aborder le 3<sup>ème</sup> objectif. Les données relatives à ces 2 premiers axes sont présentées en détails dans la partie « Résultats et Discussion » de ce manuscrit.



# RESULTATS ET DISCUSSION



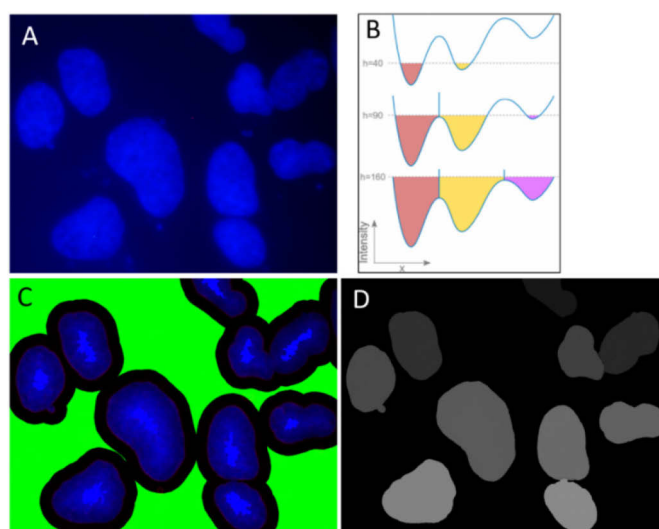
## I. Développement de protocoles d'analyse d'images

Mon projet de thèse a nécessité la réalisation de nombreuses expérimentations en microscopie. Comme nous le verrons par la suite, la caractérisation de l'apparition des microfoci de coiline induit par un stress oxydant a nécessité une analyse fine d'un très grand nombre d'images. L'analyse manuelle des images est très coûteuse en termes de temps et ne permet pas d'accéder avec précision à des paramètres complexes comme la taille des foci. Comme aucun des logiciels gratuits disponibles ne permettaient une analyse complète de mes données, j'ai débuté la mise au point de protocoles informatiques pour automatiser l'analyse de ces images et ainsi optimiser l'extraction des données obtenues par microscopie de fluorescence.

N'ayant reçu aucune formation au cours de mon cursus scolaire sur les aspects de programmation, j'ai commencé par l'apprentissage en autodidacte des principes de bio-imagerie afin d'exploiter pleinement le potentiel des outils mis à disposition dans le logiciel Icy. Ce logiciel, en libre accès, a été développé par une unité spécialisée dans l'analyse d'images biologiques de l'institut Pasteur dirigée par Jean-Christophe Olivo-Marin. Il constitue une plateforme très complète et intuitive pour l'analyse d'images. De plus, il permet aussi l'utilisation de fonctionnalités disponibles sur ImageJ qui est également très utilisé dans le monde de la bio-imagerie. Cependant, il n'existait alors aucun protocole dans Icy adapté à notre problématique.

Lors de la conception de mes protocoles, je me suis heurté à une première problématique couramment rencontrée en bio-imagerie : comment définir de la façon la plus précise possible l'objet à étudier ? Il s'agit de l'étape de segmentation. Par exemple, l'analyse de la localisation de protéines nucléaires nécessite de délimiter précisément chaque noyau de cellule, ce qui peut constituer une tâche assez difficile lorsque plusieurs noyaux sont en contact les uns avec les autres (Figure R1A). Aussi, j'ai développé plusieurs protocoles s'adaptant à différents contextes et permettant de segmenter le plus rapidement et efficacement possible des objets tels que des noyaux de cellules, en deux dimensions. Un de ces protocoles intègre l'algorithme de la Ligne de partage des eaux (surtout connu comme algorithme « Watershed ») (Figure R1B). Dans cet algorithme, chaque image est considérée comme un relief associant le niveau de gris de chaque pixel à une altitude. L'algorithme simule alors une inondation à partir des zones de basses

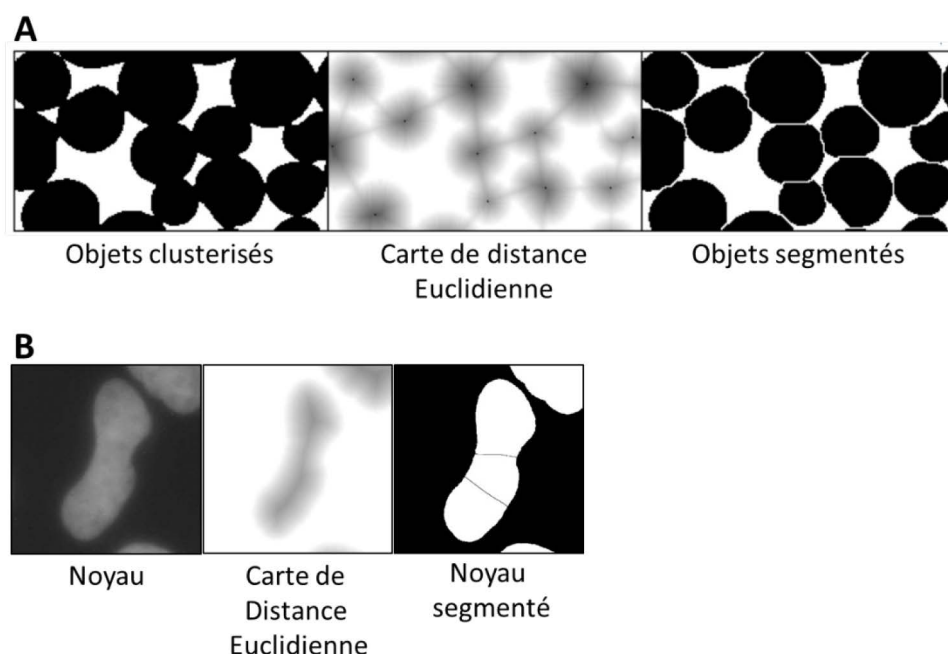
altitudes (zones sombres aussi appelés minimums) et la rencontre de différents bassins permet de définir la ligne de partage des eaux qui constitue la limite de l'objet à définir. Le principe de la segmentation watershed s'est notamment avérée efficace en imagerie pour délimiter individuellement des objets qui se touchent ou même se superposent. Cependant, une des limites de cet algorithme réside dans le fait qu'un seul objet tel qu'un noyau peut être caractérisé par plusieurs minimums à partir desquels l'inondation va être initiée. Cela conduit à une sur-segmentation de l'objet, c'est-à-dire qu'un même objet sera « découpé » en plusieurs objets plus petits et chacun issu des différents minimums. Pour pallier à ce problème de sur-segmentation, une variante de ce processus appelé « marker-based watershed segmentation » simule l'initialisation de l'inondation à partir de marqueurs prédéfinis, le but étant de définir un seul marqueur par objet à segmenter. Le succès de cet algorithme repose surtout sur la détermination optimale de ces marqueurs, étape qui peut être réalisée manuellement pour assurer une bonne qualité de segmentation. L'extraction automatique de tels marqueurs (tâches bleues claires figure R1C) sur des formes non convexes comme nos noyaux de cellules HeLa est relativement complexe, d'autant plus que chaque noyau ne doit en contenir qu'un seul. La figure R1D présente la segmentation finale obtenue.



**Figure R1. Protocole de segmentation automatisé basé sur l'algorithme LPE**

(A) Visualisation par fluorescence de noyaux cellulaires après un marquage au DAPI (B) Le processus de segmentation utilise l'algorithme LPE basé sur le principe de la ligne de partage des eaux (C) Représentation en 2 dimensions de l'inondation (D) Résultat de la segmentation obtenue après l'utilisation de l'algorithme.

Classiquement, l'extraction automatique des marqueurs se fait à partir des minimums obtenus dans la carte de distance Euclidienne des objets à segmenter (Figure R2A). Dans le cas d'objets de forme plutôt convexe, ce minimum est unique pour chaque objet et se trouve en son centre (point noir au niveau de la carte de distance Euclidienne de la Figure R2A), assurant une segmentation efficace. Cependant, lorsque la forme des objets devient plus complexe (Figure R2B) ou que la surface de contact entre 2 objets est trop importante, plusieurs minimums peuvent apparaître pour un même objet qui sera alors sur-segmenté (Figure R2B).

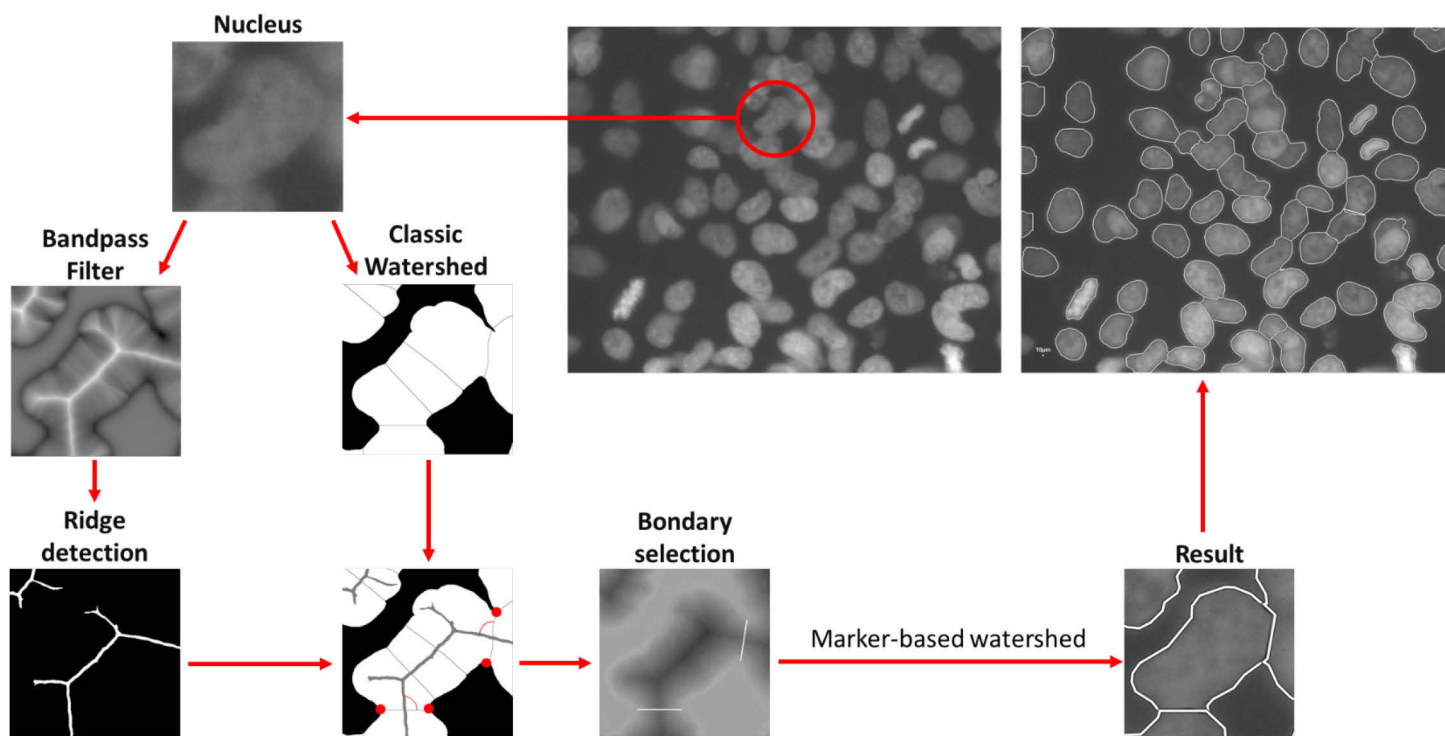


**Figure R2. Extraction automatique des marqueurs** (A) A partir d'une image d'objets clusterisés, l'extraction des marqueurs se fait à partir des minimums obtenus dans la carte de distance Euclidienne pour donner une image d'objets segmentés. (B) Cette extraction montre ces limites de la cas d'objets non convexes qui sont sur-segmentés.

J'ai donc travaillé au développement d'un protocole pour optimiser l'extraction automatique de tels marqueurs dans notre contexte d'étude, afin d'améliorer la segmentation de noyaux aux formes hétérogènes et présentant un niveau de clustering élevé (Figure R3). Ce protocole définit des limites préliminaires après avoir segmenté les noyaux avec un watershed classique. Ce processus aboutit alors à une sur-segmentation (voir Classic Watershed sur la Figure R3). Après avoir appliqué différents filtres sur l'image, des segments définissant la forme des



objets sont extraits (Ridge detection) et superposés aux limites préliminaires. Des points d'inflexions (points rouges sur le résultat de la superposition) sont également extraits pour souligner les zones de contacts pertinentes entre les objets. Le calcul des angles (arc de cercle rouge) formé entre les segments précédemment extraits et le segment formé par 2 points d'inflexions permet d'isoler les limites les plus pertinentes qui séparent 2 vrais objets. Ces limites sont ajoutées sur une image obtenue après application de plusieurs filtres et l'extraction de marqueurs est alors réalisée sur cette nouvelle image. Plusieurs paramètres peuvent être réglés pour s'adapter au type d'image et au type d'objet à segmenter. Je travaille régulièrement à l'amélioration de ce protocole et j'essaie surtout de le rendre applicable au plus grand nombre de problématiques.



**Figure R3. Protocole de segmentation basé sur la combinaison d'une segmentation watershed classique et une segmentation watershed basée sur des marqueurs**

Après plusieurs étapes de « pre-processing » pour améliorer le ratio signal/bruit, différents filtres sont appliqués pour obtenir les segments qui soulignent la forme des objets (Ridge detection). En parallèle, une segmentation watershed classique est réalisée pour obtenir des limites dites préliminaires. La superposition des segments et des limites préliminaires permettent d'obtenir des angles dont les valeurs vont permettre de détecter les véritables limites séparant les objets d'intérêts. Pour affiner le résultat de la segmentation, un marker-based watershed est appliqué sur une image intégrant les limites précédemment extraites.

Ayant réussi à développer un programme de segmentation automatisé applicable à mon contexte d'étude, j'ai ensuite développé un protocole permettant de détecter automatiquement des foci de toutes tailles au sein des noyaux cellulaires et de collecter l'ensemble des données de façon optimale. L'analyse informatique de ces foci a l'avantage d'utiliser les métadonnées de chaque image afin de ne détecter que les éléments ayant une taille et une intensité supérieures à un certain seuil. Ainsi, contrairement à l'analyse manuelle où la valeur des paramètres pris en compte est déterminée arbitrairement d'une cellule à l'autre, le logiciel détecte les foci de façon homogène sur chaque image en tenant compte de ce seuil. Au final, le protocole applique un identifiant à chaque noyau, et reporte dans une feuille de calcul chaque identifiant ainsi que le nombre de foci détectés dans le noyau correspondant. Les foci pris en compte sont également identifiés sur les images pour vérifier qu'il n'y a pas d'erreur ou d'oublis dans la détection.

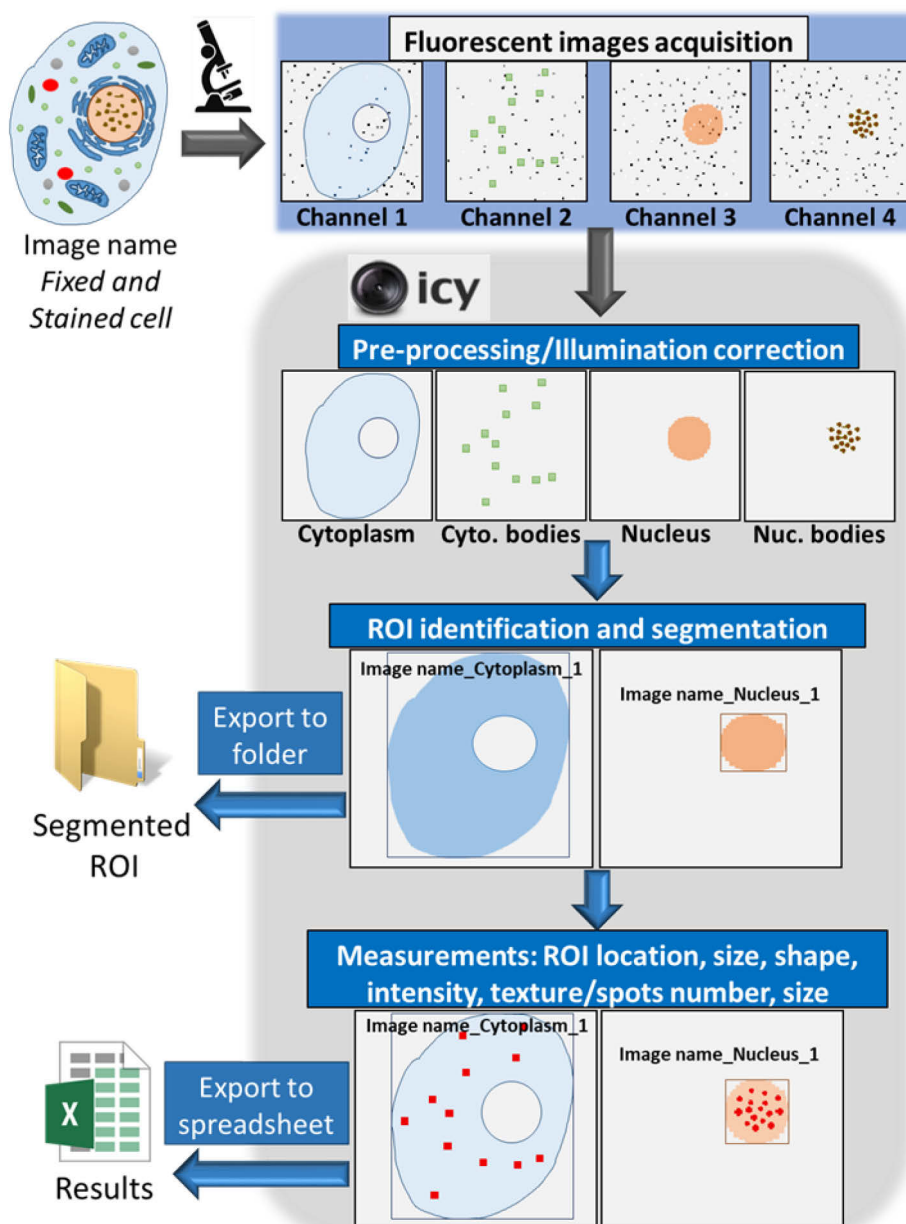
Au cours de ma thèse, de nouvelles problématiques sont apparues et je me suis également formé au langage de programmation Javascript utilisé par de nombreuses plateformes de bio-imagerie (Icy, ImageJ...). Cela m'a permis de diversifier et d'optimiser les tâches d'analyse lors de la conception des protocoles. Ainsi, j'ai conçu et développé un protocole d'analyse complet appelé « *Substructure Analyzer* » et destiné à être mis à disposition sur la plateforme Icy. Notre idée était de développer un protocole pouvant réaliser des tâches complexes et automatiques tout en restant très accessible à la société des biologistes dont les connaissances en bio-imagerie sont limitées. Afin d'informer le plus grand nombre, nous avons soumis ce programme, accompagné de la méthode d'utilisation, pour publication dans la revue JoVE (cf le manuscrit soumis en Annexe)

**Heckler G.**, Aigueperse C., Hettal L., De Chaumont F., Dallongeville S. et Behm-Ansmant I  
*Substructure Analyzer: a user-friendly workflow for rapid exploration and accurate analysis of cellular bodies in fluorescence microscopy images*

Ce workflow, dont l'architecture simplifiée est présentée dans la figure R4, donne accès à des algorithmes simples ou plus complexes qui s'adaptent aux problématiques des utilisateurs. Il est caractérisé par de nombreuses fonctionnalités :

- (i) la superposition d'images issues de canaux différents,
- (ii) la segmentation d'objets complexes et clusterisés,
- (iii) la caractérisation du signal de fluorescence (localisation, intensité, surface, nombre),
- (iv) l'exportation des données dans une feuille de calcul

Par ailleurs, du fait de ces compétences, j'ai été mis à contribution par d'autres équipes de recherche de l'unité et de l'université pour le développement d'un grand nombre de protocoles pouvant s'adapter à des problématiques différentes. J'ai, par exemple, réalisé un protocole de détection de foci capable d'analyser leur distribution dans l'espace par rapport aux structures de la cellule telles que le noyau ou la membrane plasmique. J'ai également développé des protocoles de quantification du signal fluorescent ou non fluorescent.



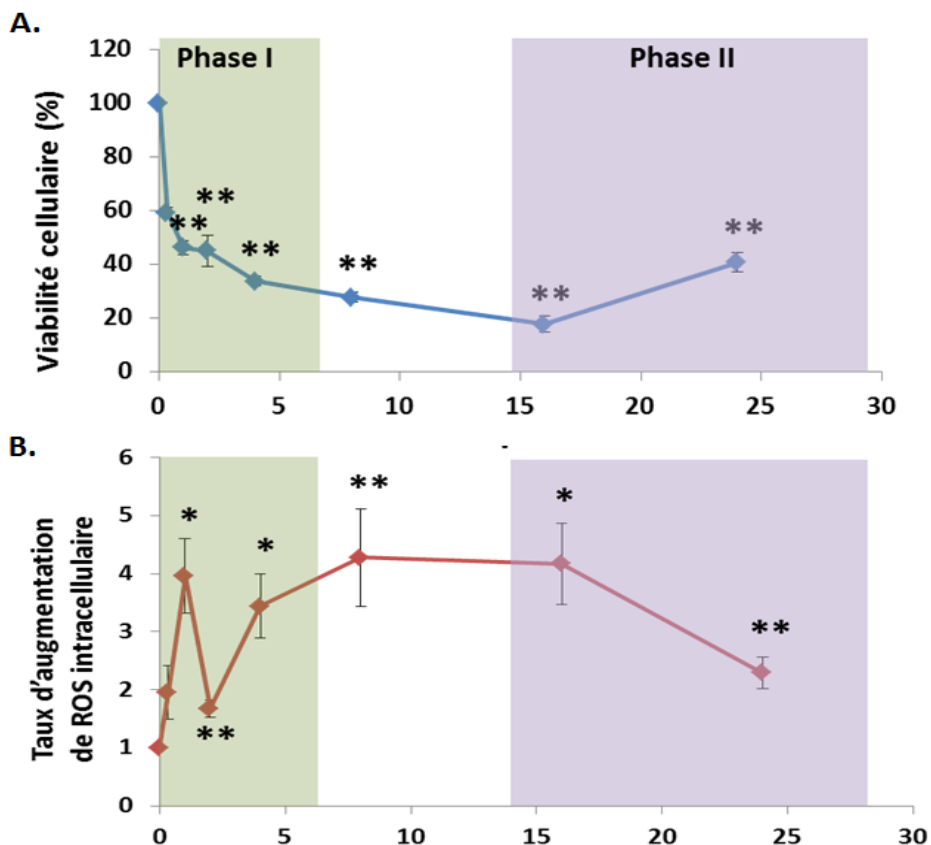
**Figure R4. Architecture du protocole « Substructure Analyzer ».** Le workflow analyse automatiquement des images acquises par microscopie de fluorescence. Premièrement, les images contenant plusieurs canaux sont intégrées dans le workflow et pré-traitées pour améliorer, si nécessaire, le ratio Signal/bruit. Puis, l'étape de segmentation extrait les régions d'intérêts (ROI). Plusieurs méthodes de segmentation sont disponibles en fonction du niveau de clustering et de la nature des objets d'intérêts. Les objets segmentés sont sauvegardés afin de pouvoir être utilisés/ré-utilisés au cours d'autres analyses. Le signal de fluorescence est ensuite détecté et analysé dans les régions d'intérêts et les résultats sont exportés dans des feuilles de calculs de type Excel.

## II. Validation des conditions de stress oxydant

### 1. Effets du SO sur la viabilité cellulaire et la production de ROS

L'induction du stress oxydant est réalisée par ajout de peroxyde d'hydrogène ( $H_2O_2$ ) dans le milieu de culture des cellules à une concentration de 500  $\mu M$ . La durée de stress correspond au temps de culture des cellules en présence d' $H_2O_2$ . Ces conditions de stress avaient été mises au point au cours de la thèse du docteur Jean-Vincent Philippe. Afin de vérifier l'efficacité du stress sur nos cellules HeLa S3, nous avons tout d'abord étudié l'effet de ces conditions sur la viabilité cellulaire grâce au test Vybrant MTT (Invitrogen). Le test MTT est une méthode colorimétrique permettant d'évaluer la viabilité en mesurant l'activité métabolique des cellules : le MTT est un sel de tétrazoium qui est réduit en formazan par la succinate déshydrogénase mitochondriale des cellules vivantes. Les analyses ont été réalisées 20 minutes, 1, 2, 4, 8, 16 ou 24 heures après l'induction du stress et les résultats obtenus sont illustrés dans la Figure R5A.

Ces données montrent une très forte diminution de la viabilité cellulaire au cours des 20 premières minutes avec une perte de près de 40% des cellules. Cette phase est suivie d'une diminution plus modérée, mais progressive jusqu'à 16 heures après l'induction du stress. A 16 heures, seulement 18% des cellules sont viables. Entre 16h et 24h, la viabilité double passant de 18% à environ 40%. Le taux de cellules viables 24h après l'induction du stress est en adéquation avec ce qui a pu être observé dans d'autres études utilisant un traitement similaire sur d'autres lignées cellulaires (Cote et al., 2012; Disher and Skandalis, 2007). Nos conditions de stress impactent donc fortement la viabilité cellulaire au cours des 16 premières heures suivant le début du stress. Cette première phase est alors suivie par une phase de récupération des cellules à partir de 16h. Il est à noter que la viabilité cellulaire ne diminue pas entre les temps 1 et 2 heures, indiquant la possible existence d'une première phase d'adaptation et de récupération au niveau de ces temps.



**Figure R5 : Analyse cinétique de la cytotoxicité et de l'induction de ROS par l'H<sub>2</sub>O<sub>2</sub> dans des cellules HeLa.**

**(A)** Cinétique de viabilité cellulaire suite à un stress oxydant réalisé avec une dose de 500  $\mu$ M H<sub>2</sub>O<sub>2</sub>. La viabilité cellulaire a été analysée par emploi du test MTT 20 min, 1h, 2h, 4h, 8h, 16h et 24h après induction du stress. La viabilité cellulaire est exprimée en % de cellules vivantes par rapport aux cellules contrôles non stressées.

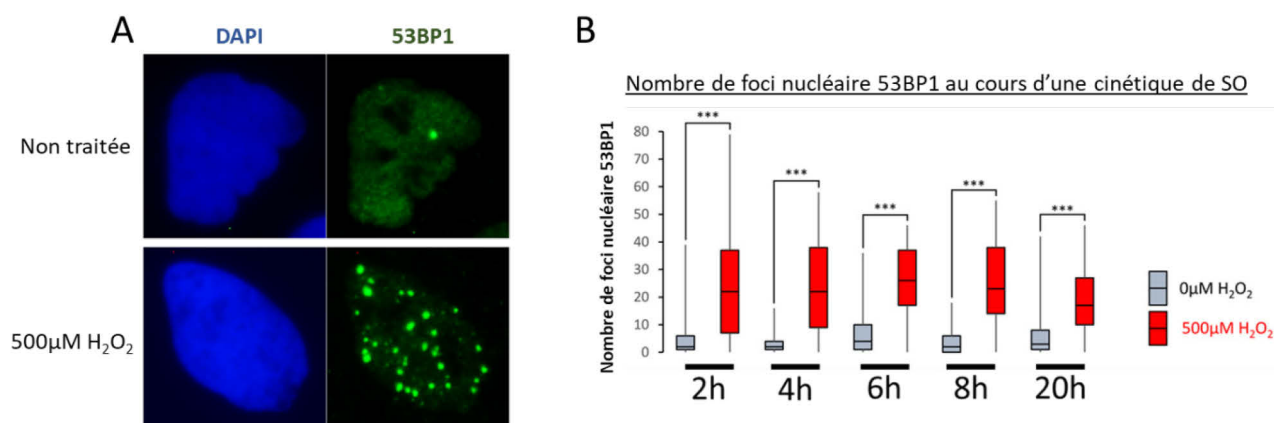
**(B)** Cinétique d'induction des ROS intracellulaires suite à un stress oxydant réalisé avec une dose de 500  $\mu$ M H<sub>2</sub>O<sub>2</sub>. Le dosage des ROS intracellulaires a été réalisé par emploi du H<sub>2</sub>DCFDA 20 min, 1h, 2h, 4h, 8h, 16h et 24h après induction du stress. Le niveau de ROS intracellulaires dans les cellules non traitées est utilisé comme référence pour déterminer le taux d'induction des ROS dans les cellules exposées à l'H<sub>2</sub>O<sub>2</sub>. Chaque barre d'erreur représente la valeur +/- de l'erreur standard (n=3 ; \* P<0,05 ; \*\* P<0,01 par rapport aux cellules non stressées). L'existence de deux phases de réponses au stress oxydant : une phase précoce (1-4 heures) et une phase tardive (16-24 heures) associée à une reprise de l'activité métabolique est indiquée.

En parallèle, j'ai analysé l'effet du stress oxydant sur le taux de ROS intracellulaires (Figure R5B). On constate que la diminution de la viabilité cellulaire est globalement corrélée à une augmentation du taux de ROS intracellulaires et que les phases de récupération des cellules sont associées à une diminution du taux de ROS. La première forte diminution des ROS correspond au temps de ralentissement de la diminution de la viabilité et la phase de récupération (II) des

cellules est associée à la deuxième diminution des ROS. Ces données sont donc en parfait accord avec notre hypothèse d'une réponse en 2 phases.

## 2. Validation de 53BP1 comme marqueur de stress en immunofluorescence

Ces différents tests ne rendant pas compte de l'hétérogénéité de la réponse cellulaire, nous avons poursuivi notre caractérisation du stress avec un marqueur visible en immunofluorescence. Pour cela, nous avons utilisé la propriété du SO de générer un grand nombre de cassures double brin de l'ADN et d'induire la relocalisation de la protéine 53BP1 sur ces cassures (Ward et al., 2003) (Figure R6A). Grâce au protocole « *Substructure Analyzer* » que j'ai développé, j'ai pu quantifier précisément et suivre le nombre de foci nucléaires 53BP1 après différents temps de traitement des cellules HeLa S3 avec  $\text{H}_2\text{O}_2$  (Figure R6B).



**Figure R6. Cinétique de la distribution nucléoplasmique de 53BP1 après un stress oxydant.**

**(A)** Détection des foci 53BP1 par immunofluorescence dans les cellules HeLa S3 non traitées ou traitées avec 500 µM d' $\text{H}_2\text{O}_2$ . Les cellules ont été fixées et co-marquées avec un anticorps anti-53BP1 (canal vert) et un marqueur ADN (DAPI, canal bleu). **(B)** Quantification du nombre de foci 53BP1 basée sur plus de 100 cellules par condition dans 3 expériences indépendantes et avec différents temps de traitement  $\text{H}_2\text{O}_2$  (2h, 4h, 6h, 8h, 20h). L'analyse des données via un T-test montre une augmentation significative du nombre de foci 53BP1 quelle que soit la durée du stress oxydant (\*\*\*:  $p < 0.001$ ).

Ainsi, quelle que soit la durée du traitement H<sub>2</sub>O<sub>2</sub>, on observe des foci 53BP1, reflétant la présence de cassure double brin de d'ADN. Pour chaque temps de la cinétique, les données générées grâce au protocole ont été utilisées pour réaliser des courbes ROC (Receiver Operating Characteristic) afin de déterminer les seuils optimaux de nombre de foci 53BP1 permettant de distinguer les cellules stressées et non stressées. Ces valeurs sont reportées dans le Tableau R1. Les seuils sont accompagnés de différents paramètres (Sensibilité et spécificité du test, nombre de faux positifs/négatifs, nombre de vrais positifs/négatifs).

**Tableau R1 Test ROC (Receiver Operating Characteristic) entre cellules stressées et non stressées au cours de la cinétique de temps**

Seuil : nombre de foci 53BP1 considérés pour le test ROC. VP : Vrai positif ; VN : Vrai négatif ; FP : Faux positif; FN : Faux négatif

| Temps après induction du stress | Seuil | Sensibilité | Spécificité | VP | VN | FP | FN |
|---------------------------------|-------|-------------|-------------|----|----|----|----|
| 2h                              | 16    | 0,579       | 0,893       | 33 | 50 | 6  | 24 |
| 4h                              | 17    | 0,632       | 1           | 36 | 56 | 0  | 21 |
| 6h                              | 17    | 0,754       | 0,929       | 43 | 52 | 4  | 14 |
| 8h                              | 16    | 0,702       | 0,982       | 40 | 55 | 1  | 17 |
| 20h                             | 15    | 0,632       | 0,911       | 36 | 51 | 5  | 21 |

Sur la base de ce tableau, nous avons considéré que les cellules présentant plus de 16 foci nucléaires 53BP1 sont stressées. Ce critère sera utilisé par la suite pour discriminer les cellules stressées des cellules non stressées dans les études d'immunofluorescence.

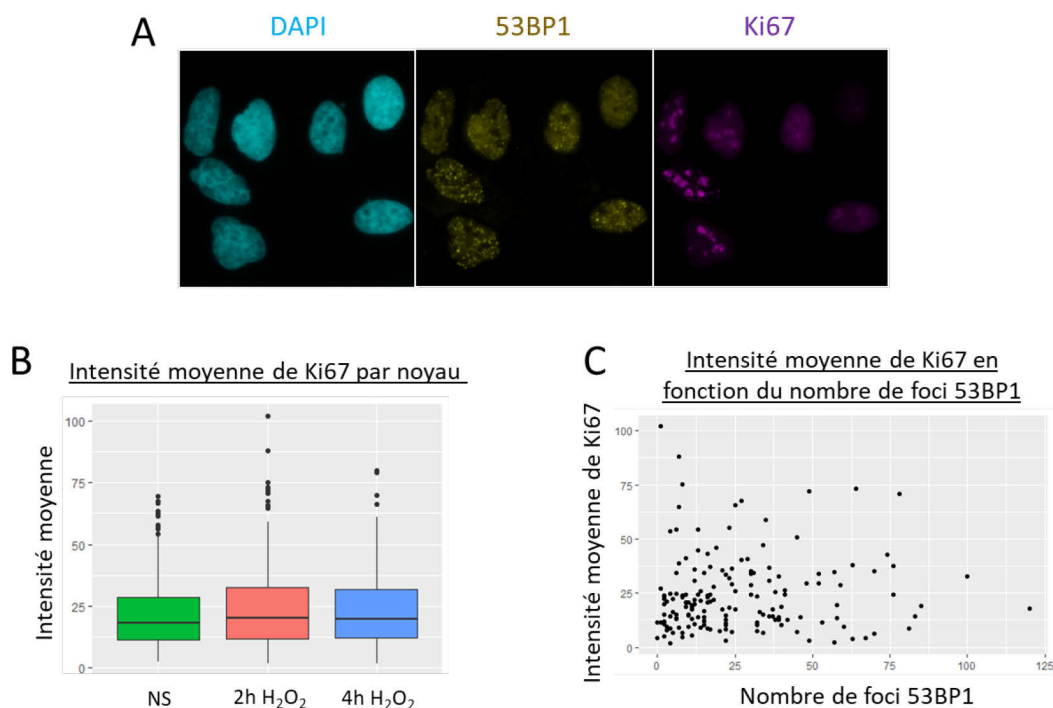
### 3. Evaluation de la prolifération cellulaire

Afin de compléter la caractérisation des effets de nos conditions de SO sur les cellules HeLa S3, j'ai suivi la prolifération cellulaire en immunofluorescence en utilisant la protéine Ki67 comme marqueur, en co-marquage avec 53BP1. Cette dernière, identifiée pour la première fois en tant qu'antigène dans les noyaux de cellules de lymphome de Hodgkin (Gerdes et al., 1983), est fortement exprimée au cours du cycle cellulaire et sous-exprimée dans les cellules quiescentes en phase G0, le signal de Ki67 étant à peine visible à l'œil nu dans les cellules non prolifératives (Figure R7A)(Gerdes et al., 1984). Grâce à mon protocole « *Substructure Analyzer* », j'ai comparé



l'intensit  moyenne du signal de Ki67, qui refl te directement le niveau d'expression de la prot ine, dans les noyaux de cellules non stress es et stress es. Toutes les images ont  t  analys es dans le format 8-bit, signifiant que la valeur de chaque pixel est comprise entre 0 et 255. Pour chaque condition (Non stress e, 2h et 4h apr s induction du stress), environ 150 cellules ont  t  analys es. Les distributions de l'intensit  nucl aire moyenne de Ki67 pour chaque condition ont  t  repr sent es sous forme de boxplot dans la Figure R7B.

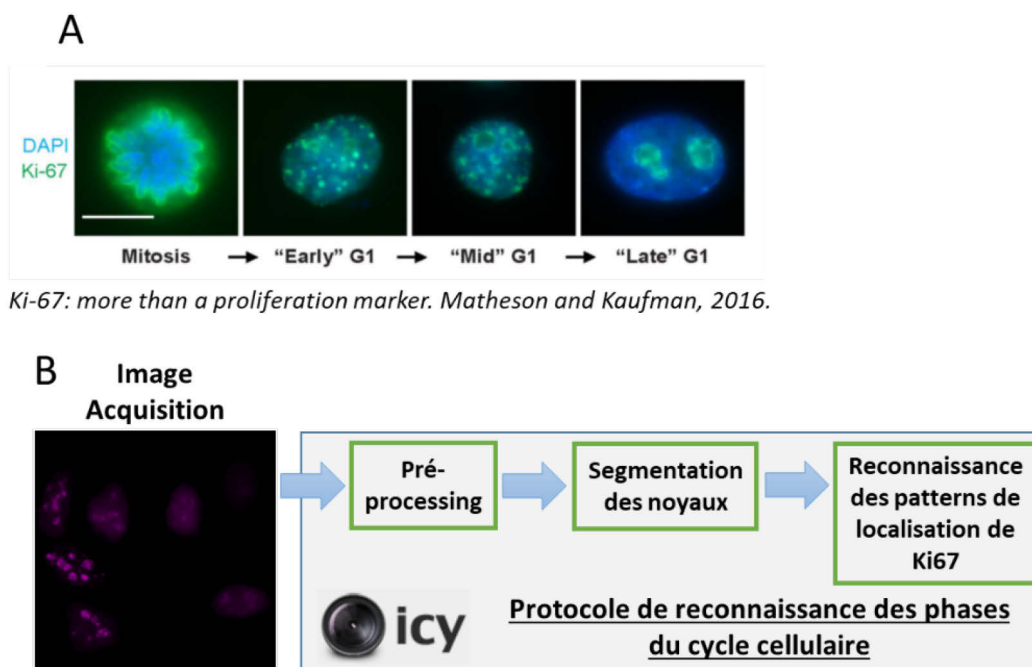
Ces donn es montrent que les cellules des conditions t moins et stress es pr sentent des profils de distribution d'intensit  tr s similaires, sugg rant que le stress n'a pas d'impact significatif sur la proportion de cellules en phase G0. Ces premi res distributions ont  t  r alis es   partir des donn es de toutes les cellules de chacune des conditions (Non stress e, 2h et 4h apr s induction du stress), sans prendre en compte l' tat de stress de la cellule via le nombre de foci 53BP1. Le graphe de la figure R7C, repr sente l'intensit  moyenne de Ki67 en fonction du nombre de foci nucl aires 53BP1 par cellule afin de v rifier si une relation entre ces deux param tres existe. Nos donn es montrent que les cellules stress es (nombre de foci 53BP1 >16) ne pr sentent pas une intensit  de Ki67 diff rentes des cellules non stress es, par cons quent nos conditions de SO ne semblent pas induire un blocage en phase G0.



**Figure R7. Quantification de la protéine Ki67 en immunofluorescence dans des cellules traitées ou non par  $H_2O_2$ .**

**(A)** Détection par immunofluorescence des protéines 53BP1 (jaune) et Ki67 (magenta) dans les noyaux (cyan) de cellules HeLa S3. **(B)** Quantification de l'intensité moyenne de Ki67 dans des cellules non stressées (NS), ou traitées pendant 2h ou 4h avec  $H_2O_2$ . **(C)** Représentation de l'intensité moyenne de Ki67 nucléaire et du nombre de foci 53BP1 par noyau cellulaire.

Des données récentes ont montré que la protéine Ki-67 présente des patterns de distribution nucléaire spécifiques du cycle cellulaire au sein des cellules HeLa S3 (Matheson and Kaufman, 2016) (Figure R8A). J'ai donc entrepris de développer un protocole d'analyse sur le logiciel Icy pour reconnaître ces patterns. Cela permettrait d'analyser des données d'immunofluorescence en fonction de la phase du cycle dans lequel se trouve la cellule (Figure R8B). Ce protocole est actuellement en cours de développement.



**Figure R8. Ki-67, un marqueur de phase du cycle cellulaire.**

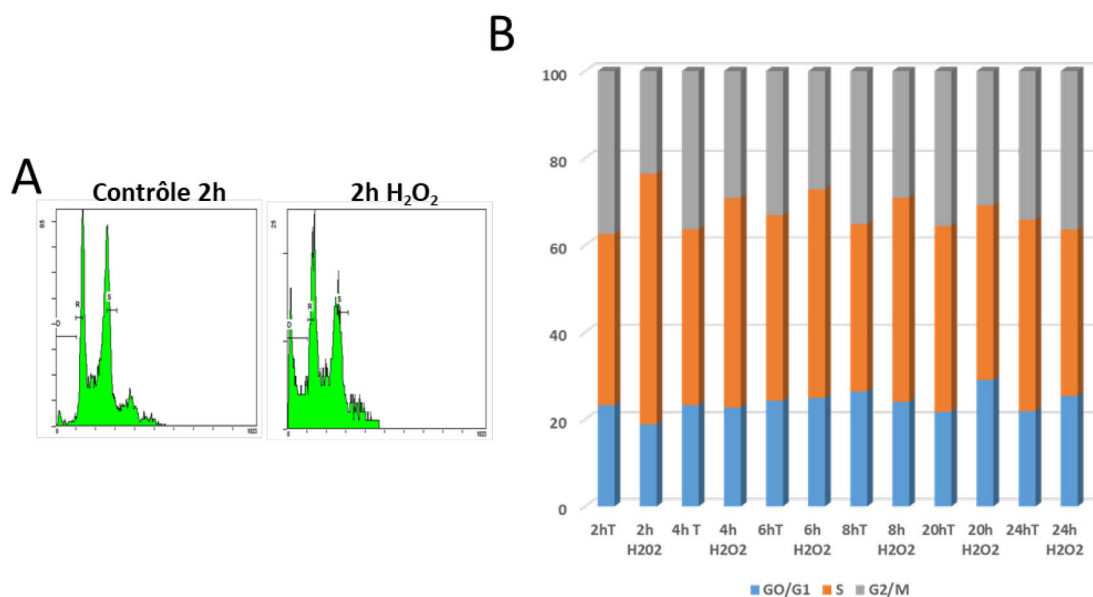
**(A)** Des cellules HeLa S3 sont marquées avec du DAPI pour visualiser le noyau (bleu) et un anticorps anti Ki-67 pour montrer les profils de marquage de Ki-67 en fonction du stade du cycle cellulaire. **(B)** Schéma du projet de protocole pour identifier les profils de distribution de Ki-67 liés au cycle cellulaire.

#### 4. Analyse des effets du stress oxydant sur le cycle cellulaire

Afin de compléter les données sur la prolifération cellulaire, nous avons étudié par cytométrie en flux les effets de nos conditions de stress oxydant sur le cycle cellulaire (Figure R9). L'iodure de propidium (IP) a été utilisé pour marquer les cellules. En effet, l'IP (excitation= 538 nm ; émission= 617 nm) est un fluochrome qui s'intercale dans les acides nucléiques en émettant une fluorescence rouge. En présence de RNase, la quantité d'IP est proportionnelle à la quantité d'ADN. L'IP permettant une analyse stochiométrique de l'ADN, il est bien adapté à l'étude du cycle cellulaire et de la ploïdie. En accord avec nos données obtenues avec le test MTT, nous constatons, après 2h de stress, une augmentation de la population de cellules en phase sub-G1 caractéristique des cellules en phase apoptotique (Figure R9A). L'analyse détaillée des phases du cycle cellulaire montre une augmentation de la proportion de cellules en phase S (de 39% à 59%, Figure R9B), 2h après l'induction du stress oxydant. Cet effet persiste, mais décroît progressivement aux temps 4h, 6h et 8h jusqu'à disparaître complètement à partir de 20h. Cette variation corrèle avec une diminution de la proportion de cellules en phase G2/Mitose alors que les phases G0 et G1 ne semblent pas affectées. Ces résultats suggèrent que le traitement à 500 $\mu$ M de H<sub>2</sub>O<sub>2</sub> induit un blocage des cellules HeLa S3 en phase S. En effet, la formation de dommages à l'ADN et donc l'activation des systèmes de réparation du génome sont souvent à l'origine d'un blocage en phase S. Ces données corréleront avec l'augmentation du nombre de foci 53BP1 et donc de cassures doubles brins de l'ADN induites par ce même stress oxydant.

Pour résumer les données de caractérisation du stress oxydant : l'analyse de la production de ROS intracellulaires et de la viabilité cellulaire supportent un modèle de réponse au stress oxydant en 2 phases, avec une **phase précoce (1h-8h)** au cours de laquelle i) la concentration effective en H<sub>2</sub>O<sub>2</sub> dans le milieu de culture est encore élevée (car le temps de demi-vie estimé de l'H<sub>2</sub>O<sub>2</sub> dans le milieu de culture est d'environ 8h), ii) de nombreuses lésions de l'ADN sont détectées, et iii) le cycle cellulaire est bloqué avant l'entrée en mitose afin de permettre à la cellule de réparer ces lésions. Cette 1<sup>ère</sup> phase est suivie d'une **phase dite tardive (16h-24h)**, associée à i) une faible concentration résiduelle d'H<sub>2</sub>O<sub>2</sub> dans le milieu de culture, ii) à la réparation des cassures double-brin qui restent certes présentes, mais dont le nombre diminue pour

finalement quasi totalement disparaître environ 36h après l'induction du stress (communication personnelle JV Philippe), iii) à la reprise de la mitose et iv) à la reprise de l'activité métabolique.



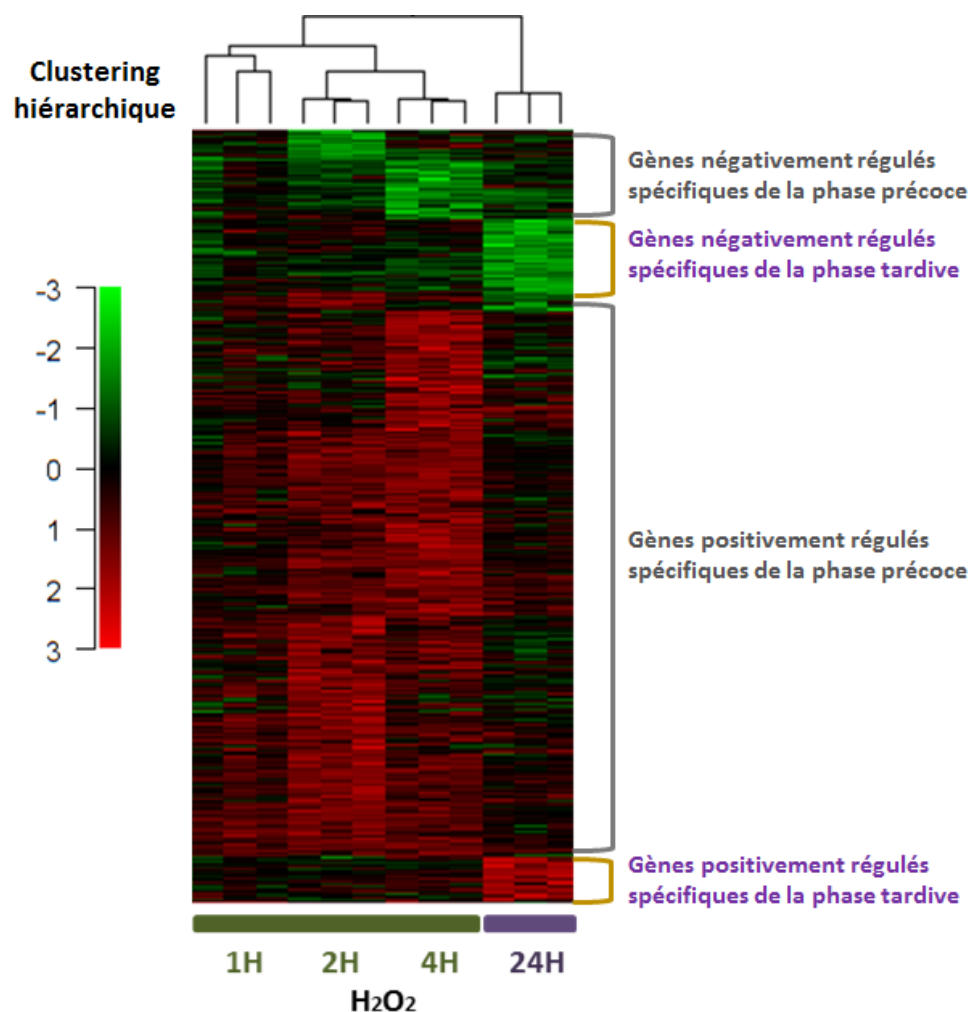
**Figure R9. Analyse du cycle cellulaire après un stress oxydant au cours d'une cinétique de temps.**

Les cellules ont été traitées ou non avec 500  $\mu$ M H<sub>2</sub>O<sub>2</sub> et récoltées aux temps indiqués. L'ADN des cellules a été marqué avec de l'iodure de propidium afin de distinguer les différentes phases du cycle cellulaire en fonction du contenu en ADN des cellules.

### III. Effets du stress oxydant sur le transcriptome cellulaire

Jean-Vincent Philippe, à la fin de sa thèse, avait débuté l'analyse des effets du stress oxydant sur le transcriptome cellulaire en utilisant des puces Affymetrix Exon 1.0 ST. Cette analyse avait été réalisée en collaboration avec l'équipe de Laurent Vallar au Luxembourg Institute of Health (LIH). Les puces Exon 1.0 ST, basées sur la version GRCh37 du génome, contiennent des sondes spécifiques des 306 583 exons identifiés dans la base de données Ensembl à cette époque. Elles sont tout à fait adaptées pour l'analyse des variations d'expression des gènes codants, mais peu adaptées pour l'analyse de l'expression des ARN non codants (car répertoire très incomplet dans cette version du génome) et pour l'analyse de l'épissage alternatif (car absence de sondes couvrant les jonctions). L'analyse des données préliminaires obtenues à partir d'ARN totaux

collectés 1h, 2h, 4h et 24h post-induction du SO (cellules HeLa, 500  $\mu$ M  $H_2O_2$ ,  $n=3$ ) avait néanmoins permis de mettre en évidence de nombreuses variations d'expression et d'épissage et laissait présager un effet global du SO sur l'épissage alternatif. Par ailleurs, ces données transcriptomiques confirmaient l'existence d'une phase précoce et d'une phase tardive de réponse transcriptionnelle au stress oxydant. En effet, la classification non supervisée (clustering hiérarchique) des données révélait une forte similarité des profils d'expression aux temps 1h, 2h et 4h, alors que le profil d'expression au temps 24h était clairement différent (Figure R10).



**Figure R10. Classification non supervisée (clustering) des données d'expression génique au cours d'une cinétique de réponse au stress oxydant.**

Cette analyse a été réalisée à partir des données d'expression génique (intensités normalisées) issues de l'analyse sur puces Affymetrix 1.0 ST des ARN totaux extraits 1h, 2h, 4h et 24h après induction d'un stress oxydant avec 500  $\mu$ M d' $H_2O_2$ . Des groupes de gènes spécifiquement positivement ou négativement régulés lors de ces 2 phases de réponses ont pu être identifiés.

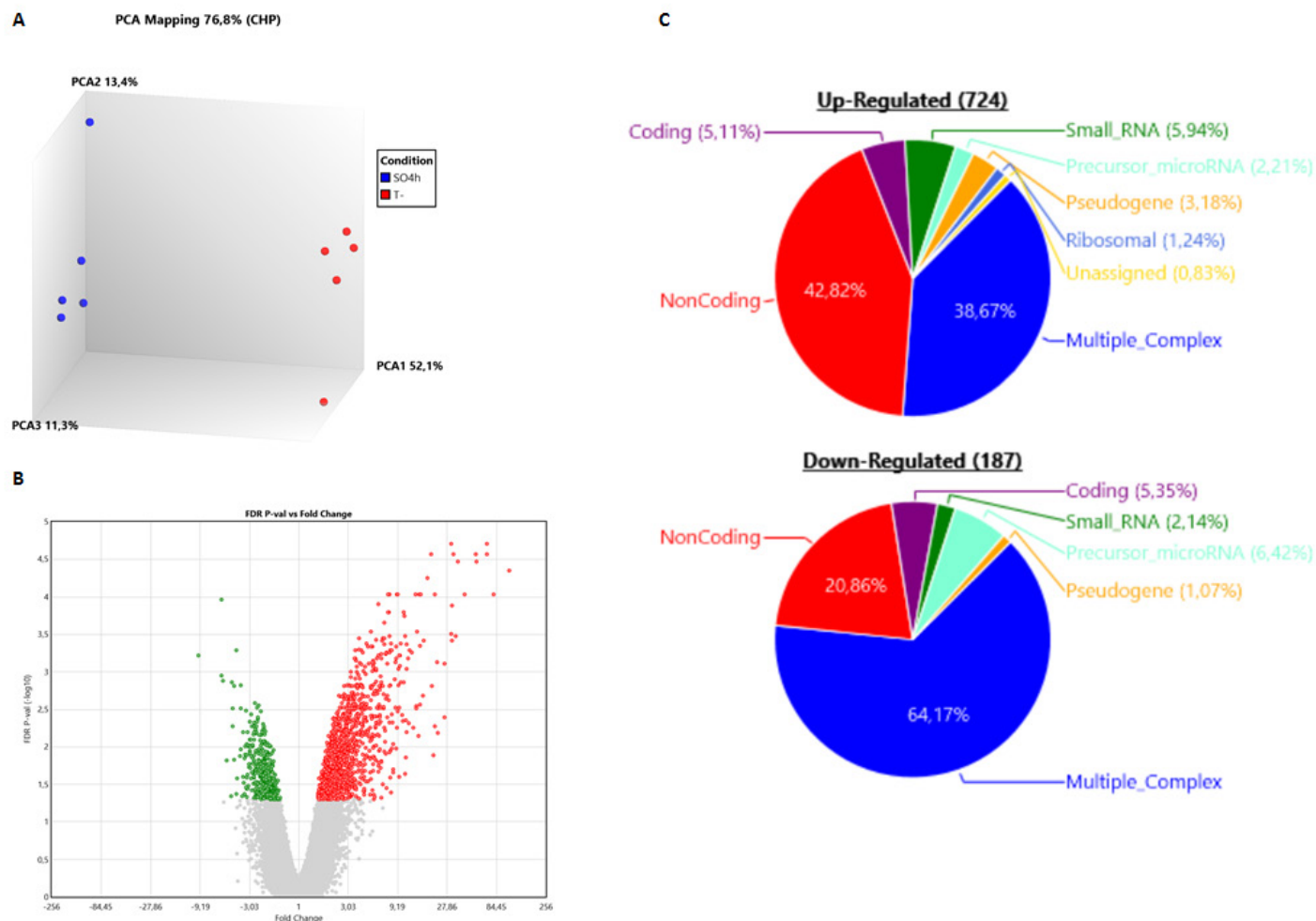
Afin de confirmer ces données et de générer des analyses plus poussées concernant l'épissage alternatif, j'ai entrepris de répéter ces analyses, mais en utilisant la dernière génération de puces Affymetrix, les puces Clariom D. Ces puces, basées sur la version GRCh38 du génome, contiennent des sondes spécifiques pour plus de 948 300 exons, permettent la détection des longs ARN non codants récemment annotés et couvrent, par ailleurs, 484 900 jonctions d'exons. Etant données i) la forte similitude des profils d'expression et d'épissage au sein de la phase précoce de réponse (Figure R10) et ii) la publication par l'équipe de Silvia Barabino de données transcriptomiques générées 24h post-induction d'un stress oxydant à l'H<sub>2</sub>O<sub>2</sub> (Fontana et al., 2017), nous avons fait le choix de répéter nos analyses uniquement pour le temps 4h à une dose de 500 µM d'H<sub>2</sub>O<sub>2</sub> et d'augmenter le nombre de répliques biologiques (5 vs 3) afin d'obtenir des données plus robustes. Notre objectif étant d'étudier la réponse au stress, le temps 4h nous a paru plus judicieux que le temps 24h qui semble correspondre plus à un retour à la normale qu'à la réponse au stress proprement dite.

### 1. Analyse des effets du SO sur l'expression des gènes humains

La préparation des échantillons et l'hybridation sur les puces Clariom D ont été réalisées par Quentin Thuillier (Ingénieur d'étude, CDD) sur la station Affymetrix de notre équipe. L'analyse des données a été effectuée par Isabelle Behm-Ansmant en utilisant la console d'analyse Affymetrix (TAC Transcriptome Analysis Console). L'analyse en principales composantes de variations (Figure R11A) réalisée en utilisant les intensités normalisées d'expression des gènes, révèle 2 groupes bien distincts d'échantillons («T- » vs « SO4h») rendant l'analyse de l'expression et de l'épissage différentiels des gènes possibles.

Concernant l'expression des gènes, cette nouvelle analyse a permis de mettre en évidence l'expression différentielle de 911 gènes (False Discovery Rate ou FDR≤0,05) parmi lesquels 724 sont surexprimés en réponse au SO (Figure R11B). Une analyse détaillée des catégories de gènes dont l'expression est affectée par le stress oxydant a été réalisée (Figure R11C). Ces catégories de gènes (coding, non coding, small RNA, precursor\_miRNA, pseudogene, ribosomal, unassigned et multiple complex) sont celles répertoriées dans la console d'analyse Affymetrix (TAC, Transcriptome Analysis Console). L'appartenance à chacune de ses catégories est mutuellement

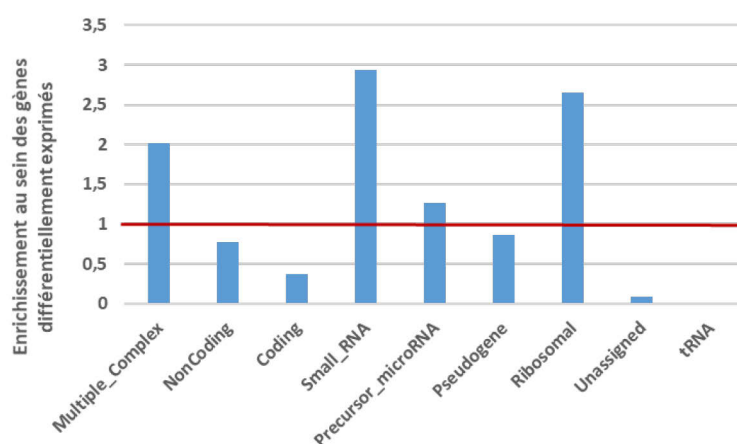
exclusive, except   pour la cat  gorie « multiple complex » qui regroupe justement tous les g  nes appartenant    plusieurs cat  gories    la fois, comme par exemple un g  ne codant exprimant    partir d'un de ses introns un ARN non codant...



**Figure R11. Analyse de l'effet du stress oxydant sur l'expression des g  nes humains.**

**(A)** Analyse en principales composantes de variation. Les intensit  s normalis  es d'expression des g  nes ont   t   utilis  es pour cette analyse. **(B)** Volcano plot pr  sentant les 911 g  nes exprim  s diff  rentiellement (FDR<0,05) entre les cellules expos  es    un stress oxydant durant 4h et les cellules non stress  es contr  les (n=5). L'axe des ordonn  es correspond    la valeur de p-value corrig  e (=FDR, exprim  e en log10) et l'axe des abscisses affiche la valeur de fold change. Les transcrits positivement et n  gativement r  gul  s par le SO sont respectivement repr  sent  s par des points rouges et verts. **(C)** Analyse de la nature des g  nes diff  rentiellement exprim  s en r  ponse au SO. Les cat  gories de g  nes sont celles r  pertori  es dans la console d'analyse Affymetrix (TAC, Transcriptome Analysis Console).

On note que 396 gènes sur les 911 différentiellement exprimés codent des ARN non codants (lncRNA, snoRNA, UsnRNA...) ce qui explique en partie le fait que seuls 31% des gènes différentiellement exprimés dans cette analyse l'étaient également dans l'analyse précédente (réalisée avec les puces affymetrix 1.0 ST ne détectant pas les ARN non-codants) au temps 4h. L'enrichissement de chacune de ces catégories de gènes a été calculé (Figure R12). De façon intéressante, l'enrichissement le plus fort a été trouvé pour la catégorie « small\_RNA », avec notamment une sur-représentation des gènes codant certains UsnRNA et snoRNA (ces résultats seront présentés en détails dans la suite du manuscrit). On note également un fort enrichissement, au sein des gènes différentiellement exprimés, des gènes codant l'ARN ribosomal 5S et de gènes de la catégorie « Multiple\_complex », tandis que les catégories de gènes « coding », « unassigned » et « tRNA » sont clairement sous-représentées (Figure R12).

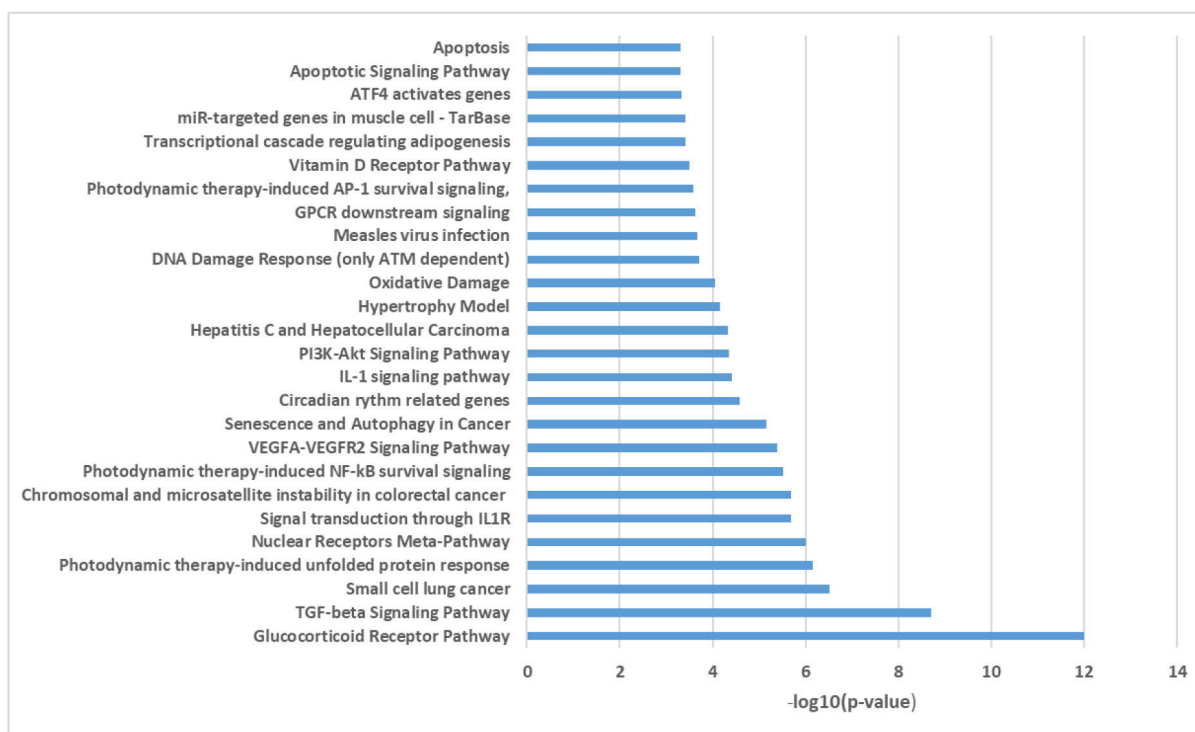


**Figure R12. Enrichissement des différentes catégories de gènes au sein des gènes différentiellement exprimés en réponse au stress oxydant.**

Les catégories de gènes sont celles répertoriées dans la console d'analyse Affymetrix (TAC, Transcriptome Analysis Console). L'enrichissement est calculé en se basant sur le nombre de gènes différentiellement exprimés au sein de nos données et le nombre total de gènes analysés sur les puces Clariom D pour chacune des catégories.

Les fonctions associées aux gènes différentiellement exprimés ont été analysées et les données relatives aux principales voies de signalisation affectées en réponse au stress oxydant sont illustrées dans la Figure R13.





**Figure R13. Analyse des voies de signalisation associées aux gènes différentiellement exprimés en réponse au stress oxydant.** Les 20 pathways les plus significativement associés à l'expression différentielle des gènes en réponse au stress oxydant sont représentés. L'analyse a été réalisée sur la console d'analyse Affymetrix (TAC, Transcriptome Analysis Console).

De façon rassurante, et en lien avec nos données expérimentales, les voies de signalisation « apoptosis », « apoptotic signaling pathways », « DNA damage response » et « oxidative damage » sont parmi les 20 voies les plus affectées par le stress oxydant. De même, l'induction d'une réponse HSP-dépendante (« photodynamic therapy-induced unfolded protein response ») est attendue du fait du rôle central joué par les protéines Heat Shock dans la réponse générale aux stress (Ikwegbue et al., 2018). Par contre, l'effet marqué du stress oxydant sur le « TGF-beta signaling pathway » et le « glucocorticoid receptor pathway » semble, de prime abord, plus inattendu. Cependant, un examen plus approfondi de la littérature a permis de confirmer le lien fonctionnel entre le stress oxydant et ces 2 pathways. Ainsi, un effet répressur de la voie de signalisation du récepteur des glucocorticoïdes sur la réponse antioxydante a été récemment mis en évidence (Alam et al., 2017). Cette répression est liée à l'inhibition, par le récepteur des glucocorticoïdes, de l'activité du facteur de transcription NRF2 (nuclear factor erythroid 2-related

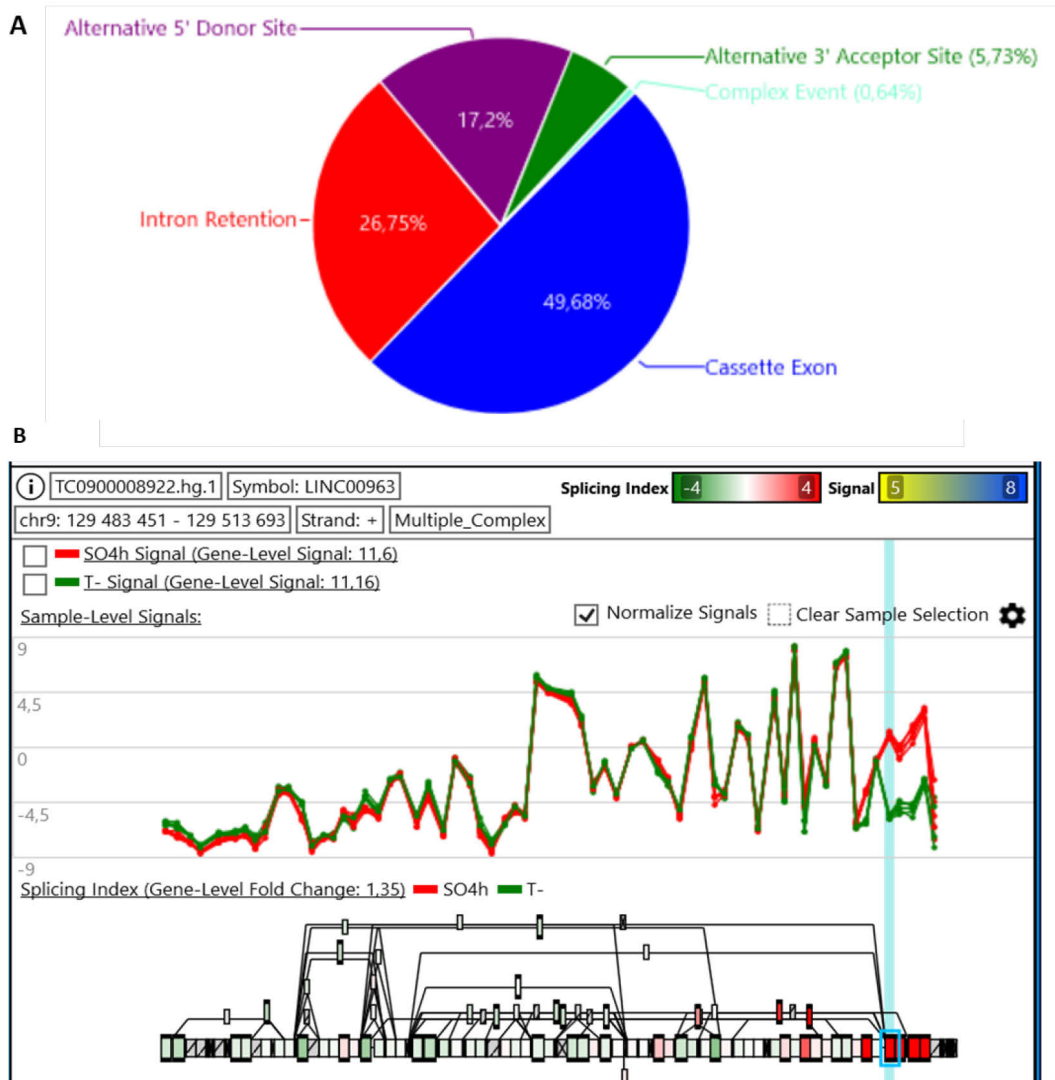
factor 2) qui est responsable de l'activation de la transcription des gènes antioxydants en réponse au stress oxydant (Alam et al., 2017). Quant au TGF $\beta$ , il est décrit réguler l'activité des ROS, d'une part en stimulant leur production (NADH oxydases), et d'autre part en réprimant l'expression d'enzymes antioxydantes (glutaredoxine, catalase, superoxyde dismutase et glutathion peroxydase) (Krstić et al., 2015).

Ces données d'expression génique confirment donc l'état de stress de nos cellules et la mise en place de mécanismes de réponse au stress oxydant. Elles mettent par ailleurs en évidence un effet marqué du stress oxydant sur l'expression de petits ARN, incluant des UsnRNA et des sno-/scaRNA, effet sur lequel nous reviendrons plus en détails dans la suite de ce manuscrit.

## 2. Analyse des effets du SO sur l'épissage alternatif

Notre objectif étant de caractériser les effets du stress oxydant sur l'épissage alternatif, nous nous sommes ensuite intéressés aux données des puces relatives à l'épissage. Notre analyse a mis en évidence 1 654 variations d'épissage significatives (Exon p-value < 0,05) et ainsi confirmé un effet global du stress oxydant sur l'épissage alternatif dès la phase de réponse précoce. La proportion des différents types d'événements d'épissage a été évaluée : les exons cassettes semblent être majoritairement affectés, suivis des rétentions d'introns (événements individuels ne correspondant pas à une inhibition globale de l'épissage), puis des sites 5' donneurs alternatifs (Figure R14A). Un exemple d'événement d'épissage détecté au sein du long ARN non codant LINC00963 est illustré dans la Figure R14B. On notera qu'aucun événement de type « sites de poly-adénylation alternatifs » n'a été détecté, suggérant des modulations d'épissage différentes en phases précoce (4h, notre étude) et tardive (24h, étude de l'équipe de Silvia Barabino) de réponse au stress oxydant. Ces résultats sont cependant à considérer avec précaution, car sur les 1 654 événements d'épissage alternatifs, seuls 349 étaient suffisamment bien annotés dans TAC pour pouvoir effectuer le calcul des proportions ; ces analyses devront donc être répétées après avoir complété l'annotation des exons. Par ailleurs, les neuroblastomes utilisés comme modèle d'étude par l'équipe de Silvia Barabino présentant i) une sensibilité accrue au stress oxydant du fait de leur très forte dépendance au métabolisme du NO et ii) des concentrations cellulaires en facteurs d'épissage différentes de celles des cellules HeLa, il se pourrait que les différences

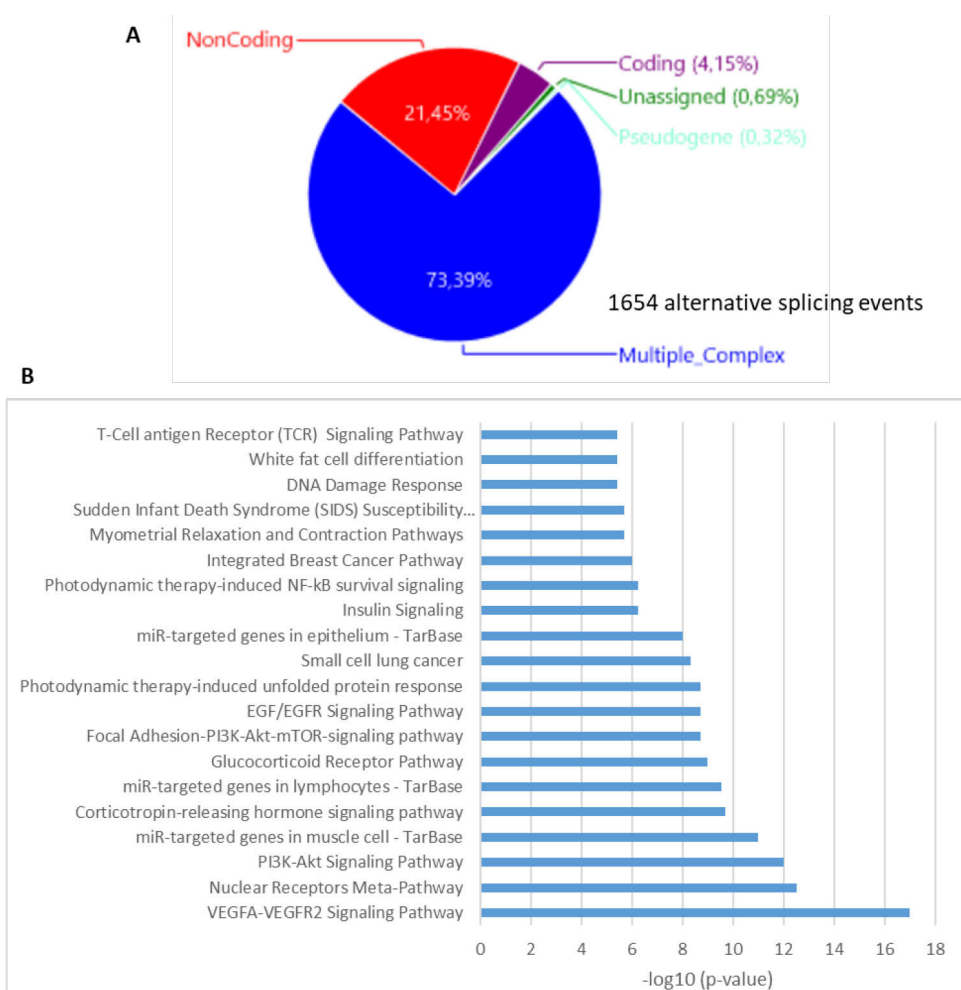
observées entre nos données (4h) et celle de cette équipe (24h) soient imputables aux modèles cellulaires utilisés et non à la cinétique de réponse, même si nos données préliminaires sur puces exon 1.0 ST allaient également dans ce sens.



**Figure R14. Analyse de l'effet d'un stress oxydant sur l'épissage des transcrits primaires humains.**

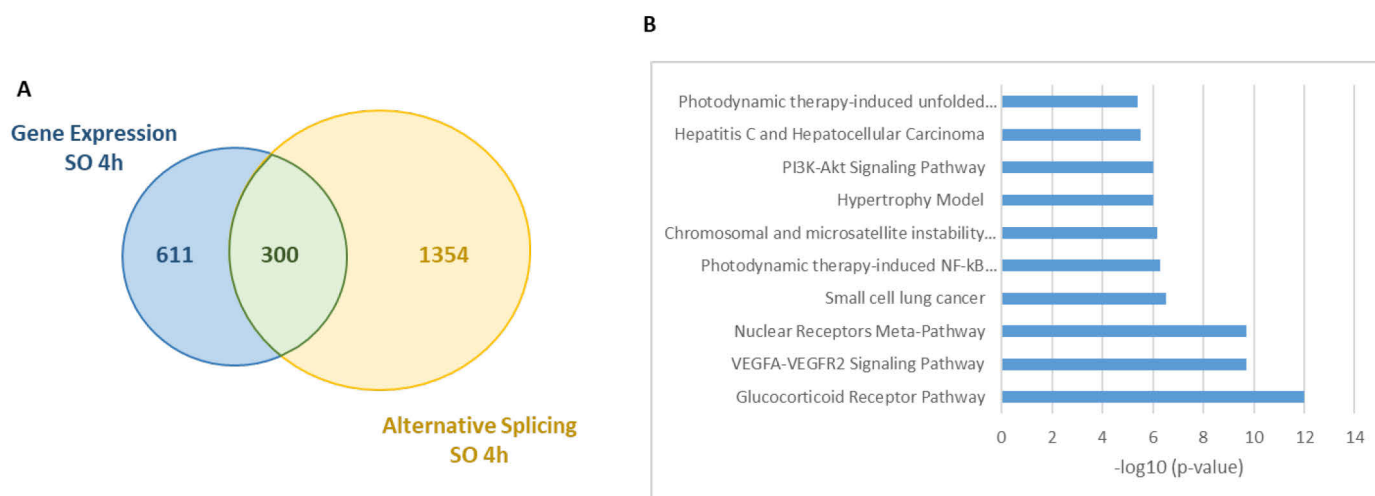
**(A)** Analyse de la nature des événements d'épissage observés en réponse au SO. Les catégories d'événements (Cassette exon, Intron retention, Alternative 5' Donor Site, Alternative 3' Acceptor site et complexe) sont celles répertoriées dans la console d'analyse Affymetrix (TAC, Transcriptome Analysis Console). **(B)** Illustration des variations d'épissage détectées pour le long ARN non codant LINC00963. Les courbes vertes et rouges correspondent à l'intensité moyenne des sondes dans les échantillons contrôles et soumis à un SO, respectivement. On note une claire augmentation de l'inclusion des derniers exons de ce gène suite à une exposition à un SO (illustration produite par capture d'écran à partir de la console d'analyse Affymetrix).

L'analyse détaillée des catégories de gènes dont l'épissage alternatif est affecté par le stress oxydant a révélé une prévalence pour la catégorie « multiple\_complex » (73%), suivie de la catégorie « non coding » (21%) (Figure R15A). De façon intéressante, la voie de signalisation la plus significativement associée aux gènes différentiellement épissés est « VEGFA-VEGFR2 » (Figure R15B). Cette voie, impliquée dans l'angiogenèse, est stimulée suite à un stress oxydant : le facteur de transcription JunD protège les cellules du stress oxydant en réglant négativement l'expression de gènes impliqués dans la production de peroxyde d'hydrogène ( $H_2O_2$ ) et positivement ceux impliqués dans la défense contre ce stress (Gerald et al., 2005).



**Figure R15. Nature et fonction des gènes différentiellement épissés en phase précoce de réponse au stress oxydant.** (A) Analyse des catégories de gènes différentiellement épissés en réponse au SO. Les catégories indiquées (coding, non coding, pseudogene, unassigned et multiple complex) sont celles répertoriées dans la console d'analyse Affymetrix (TAC, Transcriptome Analysis Console). (B) Analyse des voies de signalisation associées aux gènes différentiellement épissés en réponse au stress oxydant. Les 20 pathways les plus significativement associés à l'épissage différentiel des gènes en réponse au stress oxydant sont représentés.

Le croisement de nos listes de gènes différentiellement exprimés et épissés montre que 18% des gènes différentiellement épissés (300 gènes) sont également différentiellement exprimés, tandis que 82% des gènes différentiellement épissés (1 354 gènes) ne présentent pas de variation de leur niveau d'expression (Figure R16A). La modulation de l'épissage alternatif représente donc un niveau de régulation spécifique. La comparaison des voies de signalisation affectées (figures R13, R15B et R16B) montre, quant à elle, que les voies de signalisation les plus affectées par l'expression génique (glucocorticoid receptor pathway) ou l'épissage alternatif (VEGFA-VEGFR2 signaling pathway) sont en réalité régulées à la fois au niveau expression et épissage avec cependant des groupes de gènes au sein de ces voies qui peuvent être communément ou spécifiquement régulés au niveau de leur expression et de leur épissage. Ainsi, au sein de la voie VEGFA-VEGFR2, 11 gènes sont à la fois différentiellement exprimés et épissés, tandis que 15 autres sont uniquement différentiellement épissés.



**Figure R16. Analyse du recouvrement des régulations de l'expression et de l'épissage alternatif des gènes en phase précoce de réponse au stress oxydant. (A) Identification de groupes de gènes régulés spécifiquement à chacun des niveaux (B) Analyse des voies de signalisation associées aux 300 gènes différentiellement exprimés et épissés en réponse au stress oxydant. Les 10 pathways les plus significatifs sont représentés.**

#### IV. Etude des mécanismes cellulaires à l'origine des variations d'épissage observées après un stress oxydant

Ayant confirmé que le stress oxydant module l'épissage alternatif, nous nous sommes ensuite attelés à comprendre les mécanismes par lesquels se fait cette régulation. Disposant des données transcriptomiques, nous avons, dans un premier temps, étudié l'effet du stress oxydant sur l'expression et l'épissage alternatif des gènes codant des facteurs d'épissage ou des composants du spliceosome.

##### 1. Le stress oxydant n'affecte ni l'expression, ni l'épissage des gènes codant des protéines SR et hnRNP

Nous avons commencé cette étude en regardant l'effet du stress oxydant sur l'expression de gènes codant des facteurs de régulation de l'épissage alternatif, telles que les protéines SR (Tableau R2) et hnRNP (Tableau R3). Cependant, aucune variation significative de leur expression ( $FDR \leq 0,05$ ) ou de leur profil d'épissage (exon  $p\text{-value} \leq 0,05$ ) n'a pu être mise en évidence suite à un stress oxydant de 4h.

**Tableau R2. L'expression des protéines SR n'est pas affectée par le SO.**

L'intensité moyenne (en log2) de chaque gène dans la condition stressée (SO4h) ou contrôle (T-) est indiquée dans les 2ème et 3ème colonnes du tableau. Le fold change calculé, ainsi que la p-value et la p-value corrigée (FDR) figurent également dans ce tableau. Sont considérées comme significatives les variations présentant une  $FDR \leq 0,05$ .

| ID<br>count: 15  | SO4h<br>Avg<br>(log2) | T- Avg<br>(log2) | Fold<br>Change | P-val<br>↑ | FDR P-<br>val | Gene Symbol<br>⚡ | Description                               |
|------------------|-----------------------|------------------|----------------|------------|---------------|------------------|-------------------------------------------|
| TC0600012536.... | 6,98                  | 6,56             | 1,33           | 0,0104     | 0,2211        | SRSF12           | serine/arginine-rich splicing factor 12   |
| TC1700011820.... | 12,32                 | 12,08            | 1,18           | 0,1816     | 0,6669        | SRSF2; MIR636    | serine/arginine-rich splicing factor 2... |
| TC0200012297.... | 16,57                 | 16,4             | 1,13           | 0,2170     | 0,7023        | SRSF7            | serine/arginine-rich splicing factor 7    |
| TC1800008705.... | 4,55                  | 4,29             | 1,19           | 0,2282     | 0,7116        | SRSF10P1         | serine/arginine-rich splicing factor 1... |
| TC1100008827.... | 3,79                  | 3,63             | 1,12           | 0,2850     | 0,7544        | SRSF8            | serine/arginine-rich splicing factor 8    |
| TC0100013497.... | 9,89                  | 9,65             | 1,18           | 0,2852     | 0,7547        | SRSF4            | serine/arginine-rich splicing factor 4    |
| TC1700012437.... | 10,3                  | 10,54            | -1,18          | 0,3340     | 0,7882        | SRSF1            | serine/arginine-rich splicing factor 1    |
| TC0100013311.... | 14,14                 | 13,99            | 1,11           | 0,3726     | 0,8106        | SRSF10           | serine/arginine-rich splicing factor 10   |
| TC2100008076.... | 5,46                  | 5,53             | -1,05          | 0,5418     | 0,8877        | SRSF9P1          | serine/arginine-rich splicing factor 9... |
| TC1400007562.... | 13,82                 | 13,89            | -1,05          | 0,6726     | 0,9279        | SRSF5            | serine/arginine-rich splicing factor 5    |
| TC0600007842.... | 14,37                 | 14,28            | 1,06           | 0,6980     | 0,9349        | SRSF3            | serine/arginine-rich splicing factor 3    |
| TC2000007430.... | 11,49                 | 11,44            | 1,03           | 0,8290     | 0,9672        | SRSF6            | serine/arginine-rich splicing factor 6    |
| TC1300009201.... | 3,57                  | 3,55             | 1,02           | 0,8653     | 0,9756        | SRSF1P1          | serine/arginine-rich splicing factor 1... |
| TC1200012163.... | 10,42                 | 10,37            | 1,03           | 0,9577     | 0,9929        | SRSF9            | serine/arginine-rich splicing factor 9    |
| TC0100008692.... | 12,89                 | 12,94            | -1,04          | 0,9998     | 0,9999        | SRSF11           | serine/arginine-rich splicing factor 11   |

**Tableau R3. L'expression des protéines hnRNP n'est pas affectée par le SO.**

L'intensité moyenne (en log2) de chaque gène dans la condition stressée (SO4h) ou contrôle (T-) est indiquée dans les 2ème et 3ème colonnes du tableau. Le fold change calculé, ainsi que la p-value et la p-value corrigée (FDR) figurent également dans ce tableau. Sont considérées comme significatives les variations présentant une FDR<0,05.

| ID<br>count: 18  | SO4h<br>Avg<br>(log2) | T- Avg<br>(log2) | Fold<br>Change | P-val<br>↑ | FDR P-<br>val | Gene Symbol      | Description                                        |
|------------------|-----------------------|------------------|----------------|------------|---------------|------------------|----------------------------------------------------|
| TC0500013423.... | 11,99                 | 11,56            | 1,35           | 0,0076     | 0,1896        | HNRNPH1          | heterogeneous nuclear ribonucleoprotein H1 (H)     |
| TC1000010438.... | 11,06                 | 10,65            | 1,33           | 0,0223     | 0,3127        | HNRNPF           | heterogeneous nuclear ribonucleoprotein F          |
| TC1200007758.... | 15,35                 | 15               | 1,27           | 0,1621     | 0,6457        | HNRNPA1          | heterogeneous nuclear ribonucleoprotein A1         |
| TC0400011168.... | 15,35                 | 15,12            | 1,17           | 0,1834     | 0,6686        | HNRNPD           | heterogeneous nuclear ribonucleoprotein D          |
| TC1000007852.... | 13,12                 | 13,25            | -1,09          | 0,2002     | 0,6866        | HNRNPH3          | heterogeneous nuclear ribonucleoprotein H3 (2H9)   |
| TC0X00011309.... | 9,61                  | 9,7              | -1,06          | 0,2683     | 0,7420        | HNRNPH2          | heterogeneous nuclear ribonucleoprotein H2 (H)     |
| TC1400008625.... | 14,6                  | 14,48            | 1,09           | 0,3515     | 0,7985        | HNRNPC           | heterogeneous nuclear ribonucleoprotein C (C1/C2)  |
| TC0900010588.... | 15,65                 | 15,52            | 1,1            | 0,3811     | 0,8142        | HNRNPK; MIR7-1   | heterogeneous nuclear ribonucleoprotein K; microR  |
| TC0X00011310.... | 9,18                  | 8,96             | 1,17           | 0,3905     | 0,8192        | RPL36A-HNRNP...  | RPL36A-HNRNPH2 readthrough                         |
| TC0100018569.... | 14,22                 | 13,94            | 1,22           | 0,4253     | 0,8362        | HNRNPU           | heterogeneous nuclear ribonucleoprotein U (scaffol |
| TC1900011941.... | 16,15                 | 16,2             | -1,03          | 0,4356     | 0,8415        | HNRNPL           | heterogeneous nuclear ribonucleoprotein L          |
| TC0700010538.... | 14,02                 | 14,05            | -1,02          | 0,4704     | 0,8570        | HNRNPA2B1        | heterogeneous nuclear ribonucleoprotein A2/B1      |
| TC1900006896.... | 13,28                 | 13,21            | 1,05           | 0,5442     | 0,8883        | HNRNPM           | heterogeneous nuclear ribonucleoprotein M          |
| TC0100018424.... | 10,51                 | 10,46            | 1,04           | 0,5768     | 0,8993        | HNRNPR           | heterogeneous nuclear ribonucleoprotein R          |
| TC0500012147.... | 9,61                  | 9,43             | 1,14           | 0,6375     | 0,9192        | HNRNPA0          | heterogeneous nuclear ribonucleoprotein A0         |
| TC0500009652.... | 12,9                  | 12,81            | 1,06           | 0,6414     | 0,9201        | HNRNPAB          | heterogeneous nuclear ribonucleoprotein A/B        |
| TC1100013175.... | 11,43                 | 11,56            | -1,09          | 0,6807     | 0,9301        | HNRNPUL2-BSCL2   | HNRNPUL2-BSCL2 readthrough (NMD candidate)         |
| TC0200010083.... | 13,87                 | 13,82            | 1,03           | 0,7065     | 0,9375        | HNRNPA3; MIR4... | heterogeneous nuclear ribonucleoprotein A3; microR |

## 2. Le stress oxydant affecte l'expression des UsnRNA mineurs

Si nous n'observons pas de variations significatives de l'expression des gènes codant pour les protéines majeures de la régulation de l'épissage alternatif, en revanche, nos données de puces montrent une augmentation significative de l'expression des UsnRNA mineurs U11, U12 et U6atac lors de la phase précoce de réponse au stress oxydant (Tableau R4). Bien que l'expression du snRNA U4atac ait été analysée sur les puces et semble également augmenter en réponse au stress oxydant, ces résultats n'ont pas été pris en considération du fait du niveau d'expression très faible de ce snRNA. Aucune variation significative de l'expression des UsnRNA majeurs (U2, U4, U5 et U6) n'a été mise en évidence dans nos expériences. Concernant le snRNA U1, un des 3 sets de sondes ayant donné des résultats en contradiction avec les 2 autres, il a été exclu de l'analyse lors du contrôle qualité des données et par conséquent nous ne pouvons pas



conclure à partir de nos données de puces quant à une variation de son expression liée à la réponse précoce au stress oxydant.

**Tableau R4. Le SO affecte significativement l'expression des UsnRNA du spliceosome mineur**

L'intensité moyenne (en log2) de chaque gène dans la condition stressée (SO4h) ou contrôle (T-) est indiquée dans les 2ème et 3ème colonnes du tableau. Le fold change calculé, ainsi que la p-value et la p-value corrigée (FDR) figurent également dans ce tableau. Sont considérées comme significatives les variations présentant une  $FDR \leq 0,05$

| ID<br>count: 8   | SO4h<br>Avg<br>(log2) | T- Avg<br>(log2) | Fold<br>Change | P-val<br>↑ | FDR P-<br>val | Gene Symbol | Description                                        |
|------------------|-----------------------|------------------|----------------|------------|---------------|-------------|----------------------------------------------------|
| TC2200007526.... | 9,15                  | 6,28             | 7,34           | 2,44E-06   | 0,0021        | RNU12       | RNA, U12 small nuclear                             |
| TC0100007567.... | 8                     | 5,94             | 4,15           | 5,37E-06   | 0,0031        | RNU11       | RNA, U11 small nuclear                             |
| TC0900011875.... | 8,43                  | 6,87             | 2,95           | 0,0002     | 0,0248        | RNU6ATAC    | RNA, U6atac small nuclear (U12-dependent splicing) |
| TC1200012156.... | 7,73                  | 7,32             | 1,32           | 0,0179     | 0,2838        | RNU4-1      | RNA, U4 small nuclear 1                            |
| TC1700010767.... | 15,59                 | 14,61            | 1,98           | 0,0285     | 0,3458        | RNU2-1      | RNA, U2 small nuclear 1                            |
| TC1500009854.... | 12,71                 | 12,27            | 1,35           | 0,0302     | 0,3542        | RNU6-1      | RNA, U6 small nuclear 1                            |
| TC1500007593.... | 4,5                   | 4,26             | 1,18           | 0,0682     | 0,4859        | RNU5A-1     | RNA, U5A small nuclear 1                           |
| TC1500007595.... | 4,03                  | 3,88             | 1,12           | 0,2064     | 0,6923        | RNU5B-1     | RNA, U5B small nuclear 1                           |

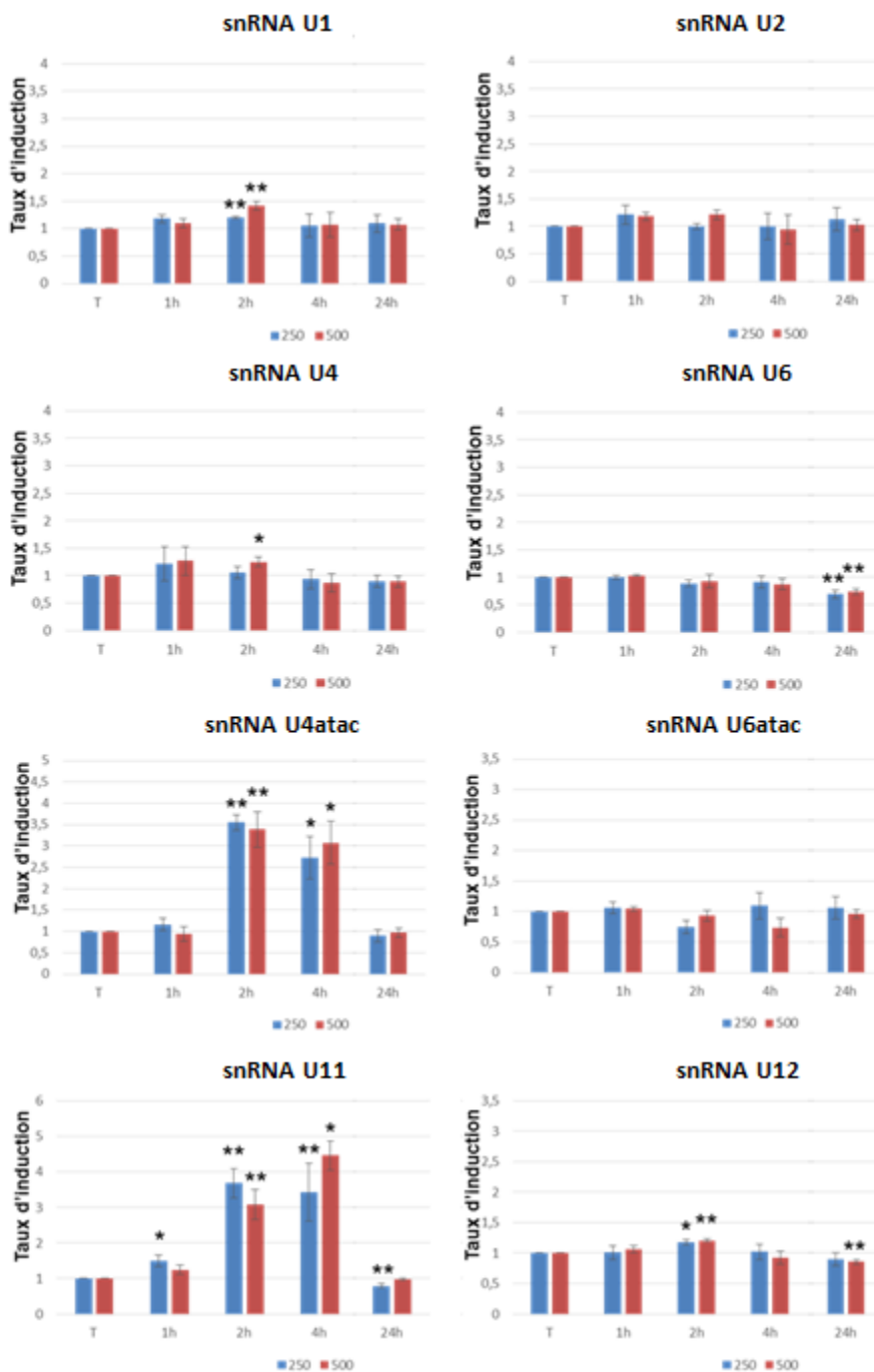
Des expériences de PCR quantitatives (qPCR) ont été réalisées afin de confirmer les variations d'expression observées sur puces pour les UsnRNA. L'efficacité des couples d'oligonucléotides et leur spécificité pour chaque UsnARN des spliceosomes majeur et mineur ont été déterminées. Les différents couples d'oligonucléotides permettant la détection du snARN U5 ne satisfaisant pas les critères de spécificité, nous nous sommes concentrés sur les autres transcrits au cours de cette étude. Les analyses ont été effectuées sur une cinétique de temps d'exposition à l'H<sub>2</sub>O<sub>2</sub> allant de 1h à 24h, avec 2 doses différentes d'H<sub>2</sub>O<sub>2</sub> (250 et 500µM) et les données présentées ont été obtenues en utilisant au minimum trois essais biologiques (avec chacun trois réplicas techniques) pour chaque condition testée. La normalisation des données de qPCR a été réalisée grâce à la détection des transcrits GAPDH, 18S et RplP0 dont l'expression ne varie pas en condition de stress oxydant.

Concernant les UsnARN du spliceosome majeur, une très légère augmentation significative d'environ X1,4 est observée uniquement 2h après l'induction du stress pour les snRNA U1 et U4, alors que les autres transcrits ne semblent pas varier (Figure R17). De façon intéressante, une forte augmentation de l'expression des homologues des snRNA U1 et U4 dans le spliceosome mineur, c'est-à-dire les snRNA U11 et U4atac respectivement, est observée pour



les temps 2h et 4h. Une légère augmentation est aussi observée pour le snARN U12. Ces résultats confirment que le stress oxydant modifie directement le niveau d'expression des UsnARN.

L'effet est surtout marqué pour les composants du spliceosome mineur, suggérant la possibilité de niveaux de régulation au niveau des introns mineurs en réponse au stress. Cette augmentation de la quantité de snARN U1, U4, U11, U4atac et U12 lors de la phase précoce de réponse au stress pourrait refléter le besoin de la cellule en composants spliceosomaux pour assurer l'expression de gènes impliqués dans les défenses anti-oxydantes, la régulation du cycle cellulaire ou la réparation des dommages causés par le stress. Cependant, la surexpression des UsnARN ne corrèle pas nécessairement avec une augmentation de leur activité. Une étude récente a notamment montré qu'une augmentation de l'expression du snRNA U1, induite par la protéine preseniline-1, entraîne une perte de fonction de U1 caractérisée par le clivage prématuré et la polyadénylation de transcrits chez des cellules de neuroblastome Humain (Cheng et al., 2017). En effet, U1 est également connu pour ses fonctions de protection des transcrits naissants contre leur clivage prématuré et leur polyadénylation au niveau de sites de polyadénylation introniques cryptiques (Berg et al., 2012).



**Figure R17. Analyse de la quantit  des UsnRNA des spliceosomes majeur et mineur au cours d'une cin tique de r ponse au stress oxydant.** Les exp riences ont  t  r alis es sur 3 s ries pour chacun des stress oxydant. Les cellules ont  t  pr lev es aux temps 1h, 2h, 4h et 24h suite   l'induction d'un stress oxydant par ajout de 250  M H<sub>2</sub>O<sub>2</sub> (bleu) ou 500  M H<sub>2</sub>O<sub>2</sub> (rouge). Pour chaque UsnRNA test , le taux d'induction de l'ARN a  t  d termin  et un contr le (T), sans ajout H<sub>2</sub>O<sub>2</sub>, a  t  r alis  pour chaque temps. Les barres d'erreur repr sentent la valeur +/- de l'erreur standard et un test de student a  t  r alis  (n=3 ; \*\* : P<0.01 ; \* : P<0.05)

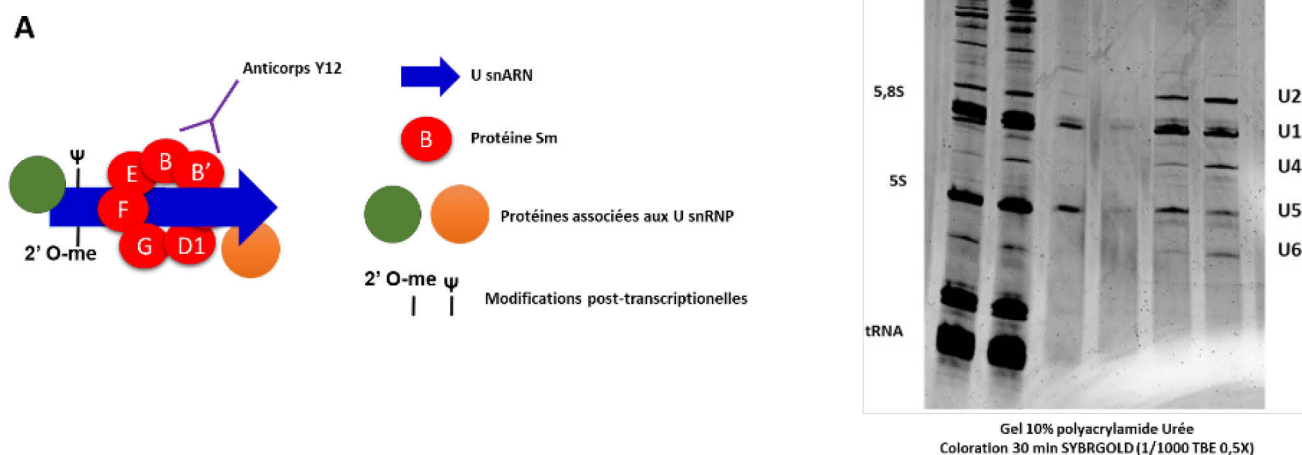
### 3. L'épissage et le niveau d'expression des gènes contenant des introns U12-dépendant ne sont pas affectés en réponse au SO

L'expression des snRNA du spliceosome mineur U4ATAC et U11 étant fortement augmentée lors de la phase précoce de réponse au stress oxydant, se posait la question d'un impact de cette augmentation sur l'épissage des introns U12-dépendants. Ces introns sont-ils plus ou moins efficacement épissés ? Cela affecte-t-il l'expression des gènes les contenant ? Des variations d'événements de rétention d'intron ayant été détectées sur puces, nous avons, dans un premier temps, recherché si ces variations concernaient des introns de type U12 répertoriés dans la U12 intron DataBase (<http://genome.crg.es/cgi-bin/u12db/u12db.cgi>). Cependant nos analyses montrent i) une absence d'enrichissement des introns de type U12 dans les cas de rétention d'introns ; ii) que les introns dont le taux de rétention varie en réponse au SO sont isolés au sein des transcrits primaires et qu'il ne s'agit donc pas d'inhibition globale de l'épissage et iii) le taux de rétention des introns est aussi fréquemment augmenté que diminué en réponse au stress oxydant.

Etant donné que la rétention d'intron conduit généralement à l'introduction de PTC (Premature Termination Codon) et à la dégradation des transcrits matures, nous avons émis l'hypothèse qu'il était peut-être difficile de détecter les événements d'augmentation de rétention des introns du fait qu'il conduisait sûrement à la dégradation des transcrits matures soit par l'exosome nucléaire, soit par la machinerie du NMD. Par conséquent nous avons recherché si dans nos données de puces aux temps 4h, des diminutions d'expression des 726 gènes contenant des introns épissés par le spliceosome mineur étaient observées et enrichies. Nous avons identifié au sein des 911 gènes différentiellement exprimés, 11 gènes contenant des introns de types U12, dont 6 voient leur expression diminuer en réponse au stress oxydant et 5 pour lesquels l'expression augmente en réponse au stress oxydant. Sur la base de ces observations, nous avons conclu à l'absence d'un effet majeur du stress oxydant sur l'épissage des introns de types U12.

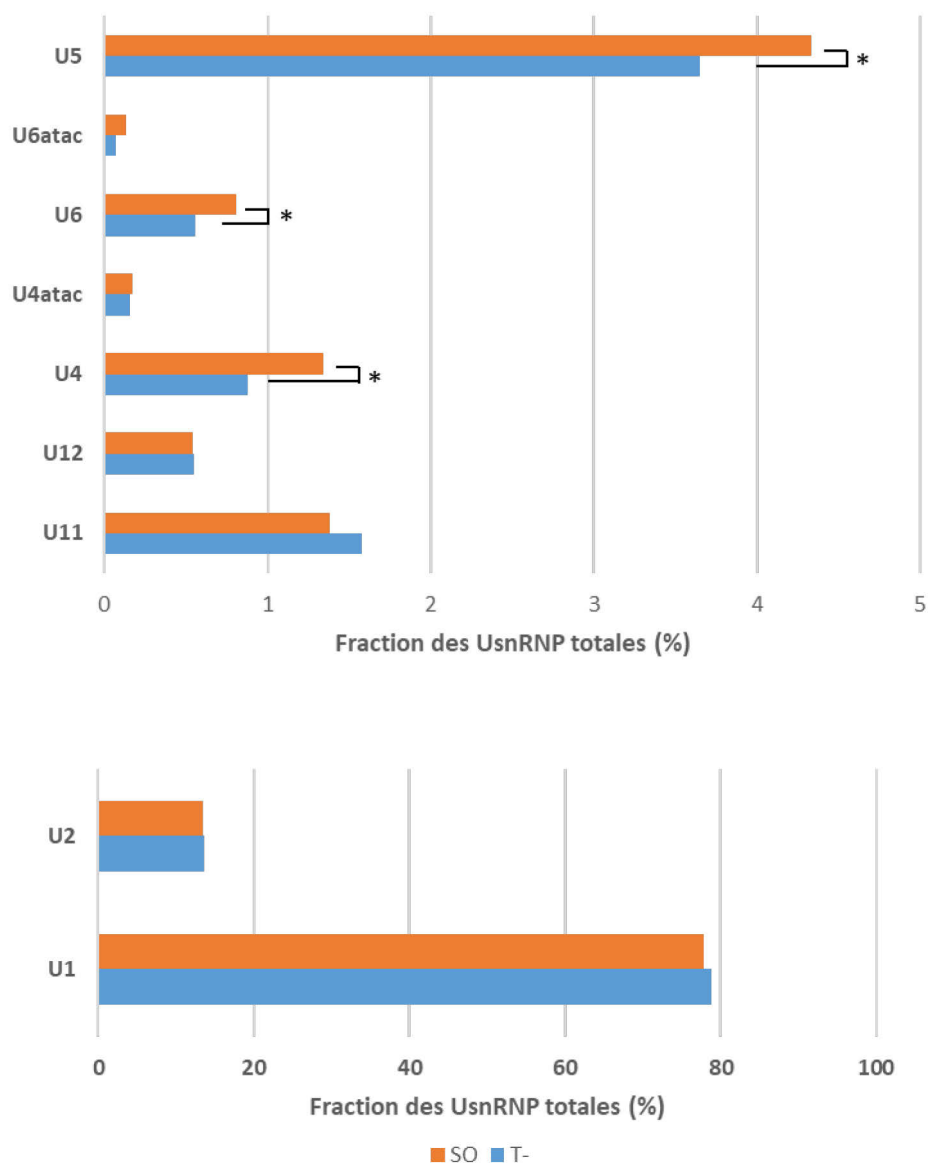
#### 4. Effet du SO sur la quantité de particules UsnRNP actives

Les UsnRNA assurent leur fonction au sein des particules UsnRNP, c'est pourquoi nous avons cherché à savoir si les variations d'expression des UsnRNA observées au temps 4h étaient associées à des variations du taux des UsnRNP correspondantes. Pour cela, nous avons entrepris de purifier les UsnRNP par des expériences d'immunoprécipitation de l'ARN (RNA Immunoprecipitation ou RIP) en utilisant l'anticorps Y12 capable de reconnaître les protéines SmB et SmB, constituants cœur des snRNP de type Sm (Figure R18A). Les UsnRNP Sm-like (U6 et U6atac) sont également enrichies dans les fractions immunoprécipitées du fait de leur présence au sein des di-snRNP (U4-U6 et U4atac-U6atac) et tri-snRNP (U4-U6/U5 et U4atac U6atac/U5). Comme l'équipe s'intéresse aux effets de différents stress sur l'épissage, nous avons réalisé les RIP Y12 dans des conditions de stress thermique (HS= Heat Shock, 2h à 42°C, puis 1h de recouvrement à 37°C) et de stress oxydant (4h post-traitement avec 500  $\mu$ M H<sub>2</sub>O<sub>2</sub>). L'enrichissement des UsnRNAs a été contrôlé par gel dénaturant (Gel Urée, 10% polyacrylamide, coloration SYBRGOLD) (Figure R18B). Des bandes correspondant aux UsnRNA majeurs ont pu être détectées, mais les UsnRNA mineurs, bien que présents sont en trop faible quantité pour être détectés par ce type de coloration (Figure R18B).



**Figure R18. Principe de l'enrichissement en UsnRNP actives par RNA-immunoprécipitation (RIP).** **A.** Principe de la RIP Y12. **B.** Fractionnement sur gel d'acrylamide dénaturant des ARN extraits à partir des fractions collectées lors des expériences de RIP réalisées avec l'anticorps Y12 à partir de lysat cellulaires de cellules HeLa S3 soumises ou non à un choc thermique (HS).

Les expériences de RIP Y12 pour les différentes conditions de stress ont été réalisées en quadruplicats biologiques et les quantités de UsnRNA au sein des fractions immunoprécipitées ont été estimées par RNAseq. Les quantités relatives des UsnRNP (+ /- SO) ont été calculées à partir des données de RNAseq et sont illustrées dans la figure R19. Aucune variation du taux des UsnRNP mineures n'a pu être ainsi mise en évidence. Ces données confirment que l'augmentation du taux des UsnRNA mineurs U11, U12 et U4atac ne reflète pas une augmentation du taux des UsnRNP correspondantes et expliquent l'absence d'effet du stress oxydant sur les événements d'épissage réalisé par le spliceosome mineur. L'augmentation significative de la fraction des UsnRNP U4, U5 et U6 au sein des UsnRNP totales suggèrent par contre un éventuel effet du stress oxydant sur la tri-snRNP U4-U6.U5. La taille et la composition des différentes RNP sera prochainement analysée par gradients de glycérol afin de confirmer ou infirmer ce résultat.



**Figure R19. Proportions relatives de chacune des UsnRNP dans des cellules HeLa soumises à un stress oxydant.** Les UsnRNP ont été immunoprécipitées par emploi de l'anticorps Y12 et les UsnARN ont été extraits à partir des fractions immunoprécipitées, puis quantifiés par RNAseq. Les proportions relatives de chacune des UsnRNP ont été calculées et sont représentées sous forme de graphique. La barre bleue correspond à la condition témoin et la barre orange à la condition de stress oxydant (500 $\mu$ M H<sub>2</sub>O<sub>2</sub>, 4h).

## 5. Etude des effets du SO sur la localisation des UsnARN

Un autre aspect important pour la compréhension du rôle des UsnARN dans les variations d'épissage suite à un stress oxydant est l'étude de leur localisation. En effet, comme il a été décrit au cours de l'introduction, l'assemblage et la maturation des particules spliceosomales se déroule séquentiellement dans des compartiments cellulaires spécifiques. La formation de modifications post-transcriptionnelles au sein des UsnARN, ainsi que l'assemblage et le recyclage de la tri-snRNP U4/U6-U5 sont des étapes essentielles de ce processus de maturation qui se concentre préférentiellement au niveau des corps de Cajal nucléaires. Le stockage de particules matures se déroule, quant à lui, au niveau d'autres substructures appelées speckles nucléaires. Nous avons donc entrepris d'étudier la localisation des UsnARN au cours de la réponse au stress oxydant.

### a) Mise au point du RNA-FISH

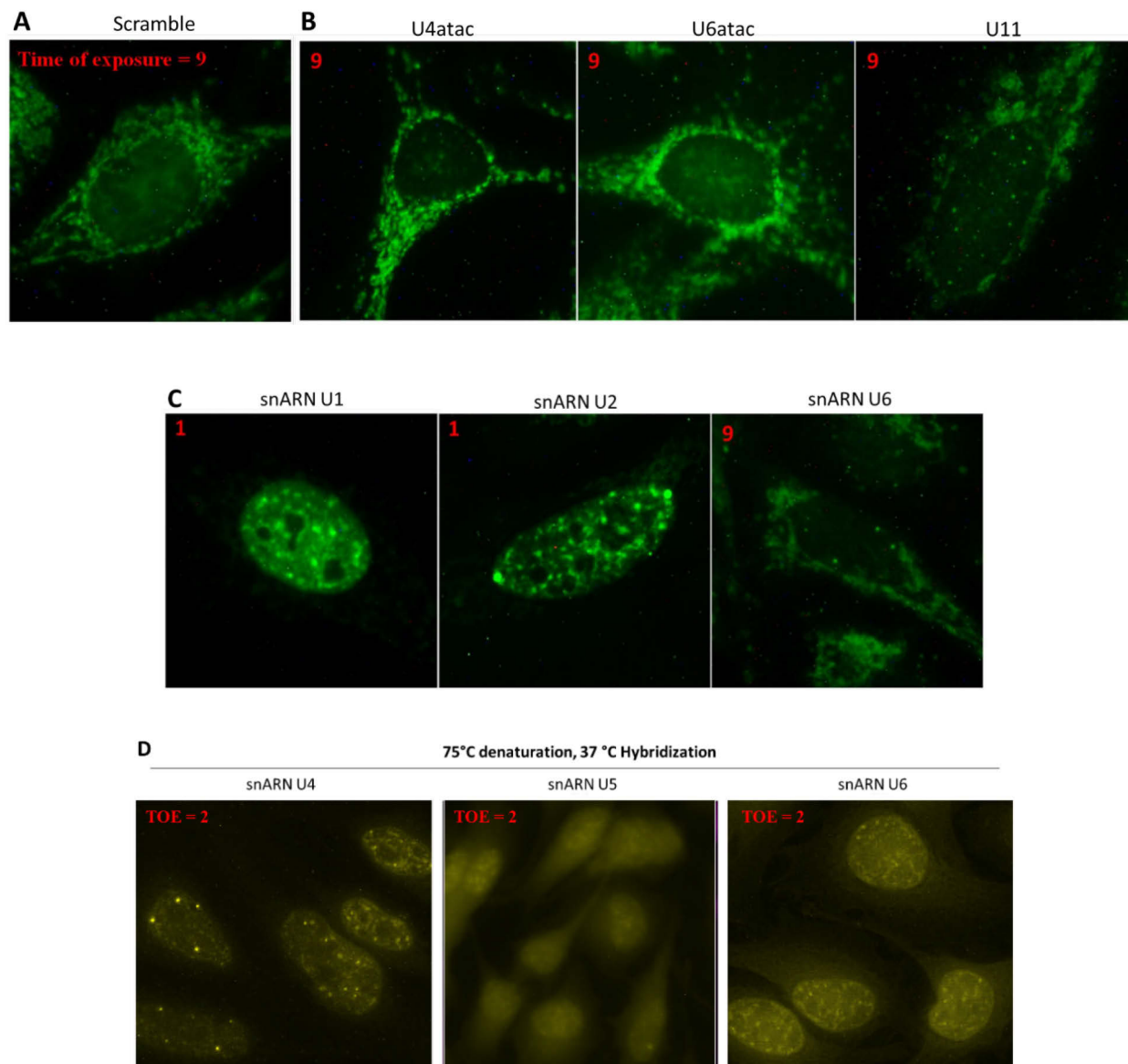
Le RNA-FISH (RNA-Fluorescent In Situ Hybridization) est une méthode puissante permettant de détecter des ARNs spécifiques au niveau de cellules intactes et préalablement fixées métaboliquement. La méthode de RNA-FISH que nous avons employée se base notamment sur l'utilisation de sondes Exiqon dites LNA (Locked Nucleic Acid, voir Matériels et Méthodes) qui sont biotinylées et détectées via un fluorophore streptavidin-Alexa Fluor® 488 et visualisées en microscopie de fluorescence (Vautrot et al., 2015). Comme nous voulions suivre la localisation des UsnARNs du spliceosome majeur et mineur, et que ces derniers sont au moins 100 fois moins abondants que les ARNs du spliceosome majeur, nous avons tout d'abord testé la spécificité du signal de nos sondes en absence de stress oxydant. Pour ce faire, nous disposions d'une sonde commerciale dite « scramble », ne ciblant aucun transcrit au sein des cellules humaines et constituant donc un contrôle pour visualiser le marquage non spécifique (Figure R20). Nous avons utilisé des sondes ciblant spécifiquement les transcrits U4atac, U6atac et U11 du spliceosome mineur (Figure R20B) et U1, U2 et U6 du spliceosome majeur (Figure R20C). Le temps d'exposition (Time Of Exposure ou TOE, en secondes, indiqué en haut à gauche de chaque image) nécessaire pour détecter le signal des snRNA U1 et U2 était 9 fois inférieur à celui du scramble. De plus, la distribution du signal observé pour ces deux sondes diffère clairement de celui du scramble, confirmant que le signal pour U1 et U2 est spécifique. Concernant U6 et les transcrits du

spliceosome mineur, le temps d'exposition et la distribution du signal sont fortement similaires à celui du scramble, ne permettant pas de déterminer si le signal est spécifique. Afin d'amplifier l'intensité du signal, nous avons utilisé un mélange de 2 à 3 sondes différentes pour chaque UsnARN mineur, mais nous n'avons pas constaté d'amélioration. Nous avons également testé des températures d'hybridations des sondes plus élevées allant de 54°C à 60°C afin d'améliorer la spécificité, mais le signal restait très proche du bruit de fond. Nous pouvons noter que les signaux des transcrits U11 et U6 présentent des foci nucléaires non présents au niveau du scramble et qui pourraient correspondre à la fraction de transcrit dans les corps de Cajal.

Ces résultats suggèrent que les sondes Exiqon utilisées ne permettent pas d'amplifier suffisamment le signal des transcrits faiblement exprimés. Nous avons donc poursuivi l'optimisation de la détection de ces UsnRNAs en utilisant des sondes généreusement mises à disposition par le Dr. Verheggen et ayant déjà faites leurs preuves au cours d'expériences précédentes. Avec ces nouvelles sondes, nous avons amélioré la détection du transcrit U6 et complété cette analyse avec les transcrits U4 et U5 (Figure R20D). Ce protocole nous permet de détecter de façon nette la fraction de U4 concentrée dans les CBs, alors qu'elle est beaucoup plus rare, voire absente, pour les transcrits U5 et U6. Des expériences de stress oxydant nous permettront prochainement de déterminer si le stress oxydant modifie la concentration du U4 au sein des CBs.

Concernant le transcrit U1, ces premières données montrent, en accord avec d'autres études précédemment réalisées (Carmo-Fonseca et al., 1991; Matera AG, 1993) qu'en absence de stress le snRNA U1 ne se concentre pas spécifiquement dans les CBs nucléaires et présente un signal nucléoplasmique plus homogène exclu des nucléoles (Figure R20C). Le transcrit U2 présente un signal plus hétérogène, toujours exclu des nucléoles mais enrichi au niveau de foci nucléaires typiques des corps de Cajal (Figure R20C). Compte tenu de la qualité et de la spécificité des signaux obtenus pour les transcrits U1 et U2, nous avons décidé de nous concentrer, dans un 1<sup>er</sup> temps, sur l'étude de leur localisation en condition de stress oxydant.

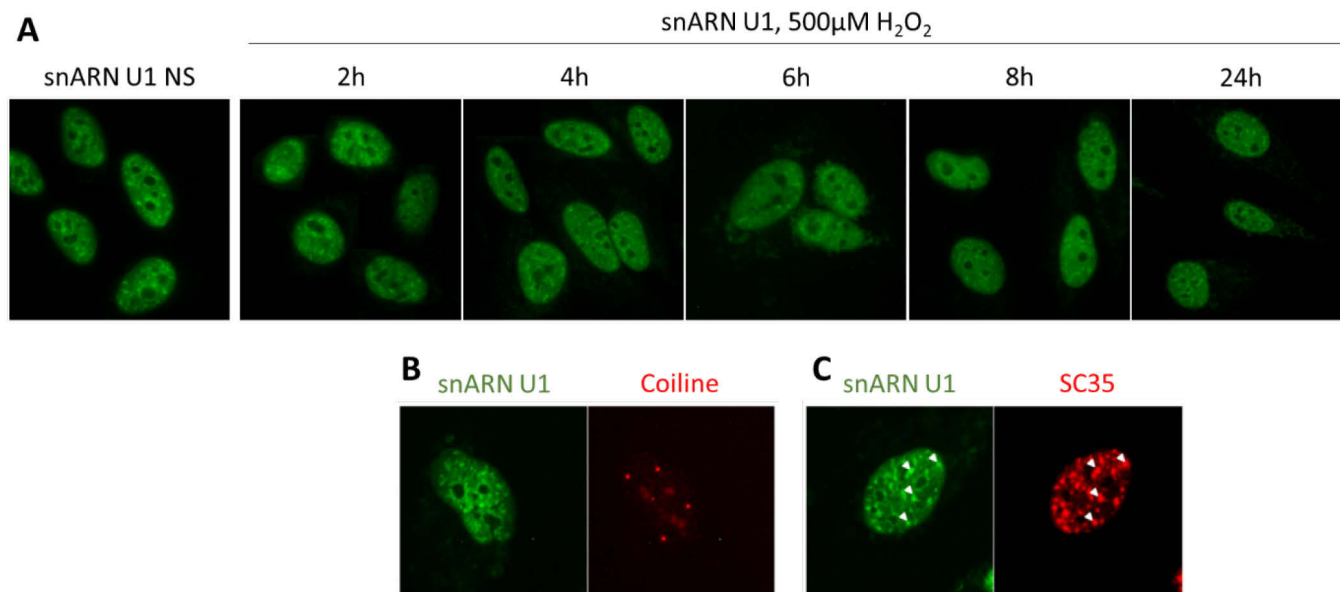




**Figure R20. Optimisations de détection des U snARNs du spliceosome majeur et mineur (B) et du spliceosome majeur (C et D) en RNA-FISH.** Les images **A**, **B**, et **C** sont issues de l'emploi des sondes LNA Exiqon. Les images **D** ont été obtenues grâce à l'utilisation des sondes fournies par le Dr. Céline Verheggen. Afin de déterminer le bruit de fond propre aux sondes Exiqon, nous disposons d'une sonde Scramble ne ciblant aucun transcrit des cellules Humaines (**A**).

### b) Effets du SO sur la localisation des snARN U1 et U2

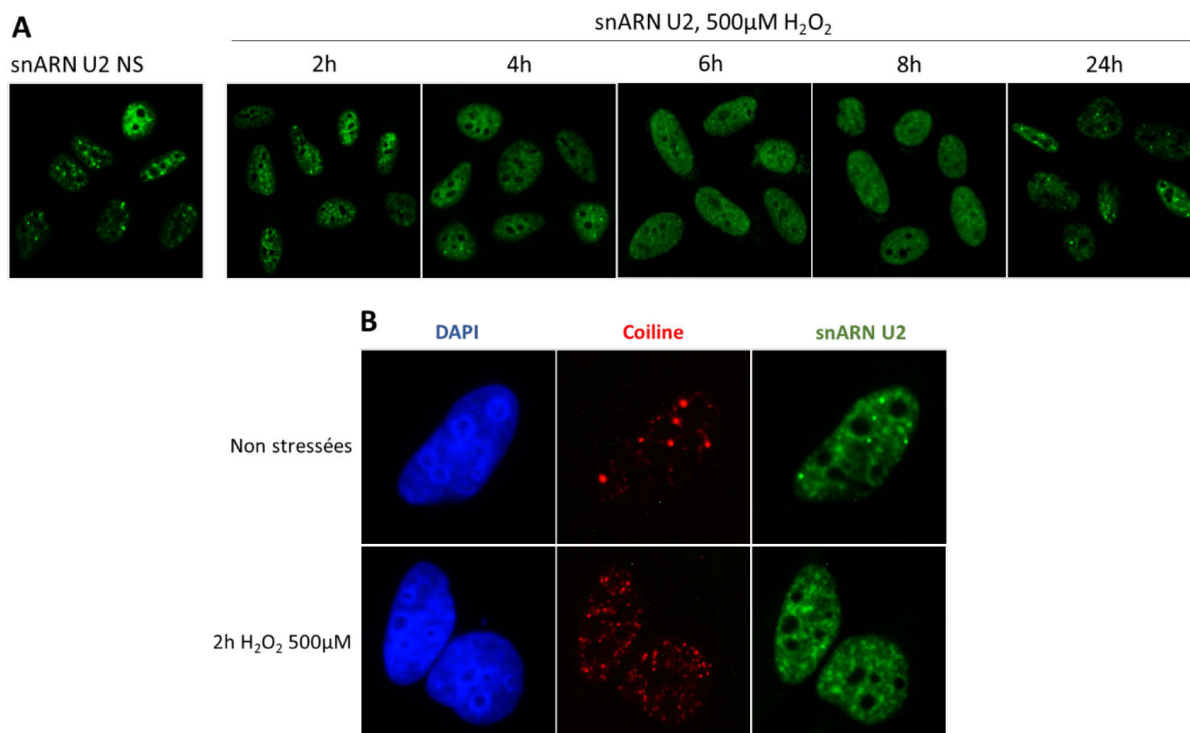
Le snRNA U1, qui est le plus exprimé, présente un signal clair avec une fraction nucléoplasmique hétérogène exclue des nucléoles. Le signal nucléoplasmique de U1 est enrichi dans certains foyers nucléoplasmiques caractérisés par des amas plus clairs. Afin d'identifier la nature de ces amas, nous avons combinés les expériences de RNA-FISH avec des expériences d'immunofluorescence afin de co-marquer U1 avec différents marqueurs de substructures nucléaires. Nous avons ainsi testé des co-marquages avec la protéine coiline, marqueur des corps de Cajal, ou avec la protéine SC35, marquant les speckles nucléaires. Le co-marquage avec la coiline a confirmé l'absence d'enrichissement spécifique du U1 au sein des corps de Cajal (Figure R21B). Cependant, ce co-marquage nous montre aussi que U1 n'est pas exclu de ces substructures, suggérant donc qu'une partie de la maturation de la snRNP U1 puisse se dérouler dans les corps de Cajal, en accord avec ce qui avait été montré chez *Arabidopsis* il y a quelques années (Lorković and Barta, 2008). Le snRNA U1 est un des ARN non codant les plus abondant, présentant un nombre de copies de l'ordre de  $10^6$  par cellule (Guirio and O'Reilly, 2015). Si la fraction de U1 présente dans les speckles nucléaires est majoritaire, il est possible que la forte intensité de signal de cette dernière puisse masquer celui de la fraction de U1 présente dans les corps de Cajal. L'analyse du co-marquage de U1 avec la protéine SC35 montre que la majorité des amas nucléoplasmiques de U1 colocalisent avec le signal de SC35 (Figure R21C), confirmant l'enrichissement du U1 dans les speckles nucléaires déjà observé par d'autres équipes (Carmo-Fonseca et al., 1991; Huang and Spector, 1992; Matera AG, 1993). En réponse au stress, la localisation de U1 ne semble pas être affectée que ce soit dans des temps précoces de réponse au stress (entre 2 heures et 8 heures) ou dans des temps plus tardifs (24 heures).



**Figure R21. Localisation du snARN U1 par RNA-FISH A.** Suivi de la localisation du snARN U1 au cours d'une cinétique de stress oxydant (2 heures à 24 heures après induction du stress avec 500μM d'H<sub>2</sub>O<sub>2</sub>). **B et C.** Caractérisation de la distribution sub-nucléaire de U1 grâce au co-marquage avec les protéines coiline et SC35 marquant respectivement les corps de Cajal (**B**) et les speckles nucléaires (**C**).

Le snARN U2 présente également une localisation nucléoplasmique hétérogène, exclue des nucléoles (Figure R22). Cependant, la fraction présente dans les corps de Cajal est facilement observable via les foci nucléaires très caractéristiques que nous avons confirmé être des corps de Cajal grâce au co-marquage avec la coiline (Figure R22B). De façon intrigante, nous observons une diminution du nombre de ces foci entre 2 heures et 8 heures suivant l'induction du stress (Figure R22A). De plus, les noyaux ne présentant que peu ou plus de foci affichent un signal nucléoplasmique plus homogène. Ces données suggéraient donc une diminution de la concentration de U2 dans les corps de Cajal et une redistribution des transcrits dans le nucléoplasme suite à un stress oxydant. Nous avons donc suivi la localisation du transcrit U2 en co-marquage avec la coiline en réponse au stress (Figure R22B). En analysant les noyaux de cellules stressées pour lesquels U2 n'est plus concentré dans les corps de Cajal, nous avons constaté que la nucléation de ces substructures était elle-même perturbée par le stress. En effet, plutôt que d'observer entre 1 et 8 foci nucléaires d'une taille caractéristique des corps de Cajal

(entre quelques centaines de nanomètres et 1 micromètre), la coiline se distribuait dans des dizaines voire des centaines de micro-foci nucléoplasmiques. Des observations similaires ont été faites en réponse à divers stress tels que les rayonnements UV, les radiations ionisantes ou des infections virales (voir Introduction) mais jamais pour le stress oxydant. Il est à noter que les rayonnements UV, les radiations ionisantes et le stress oxydant ont en commun d'induire des dommages à l'ADN, suggérant une certaine similitude dans la mise en place de certains mécanismes de réponse au stress. Ces premières données suggéraient donc que le stress oxydant pourrait modifier le potentiel de nucléation des corps de Cajal, et que cet effet perturberait la concentration du snARN U2 dans ces substructures. Compte tenu du rôle des corps de Cajal dans la maturation des U snRNP, ces résultats ouvraient alors la réflexion quant à l'impact de la redistribution nucléoplasmique de la coiline sur la fonctionnalité de ces substructures.



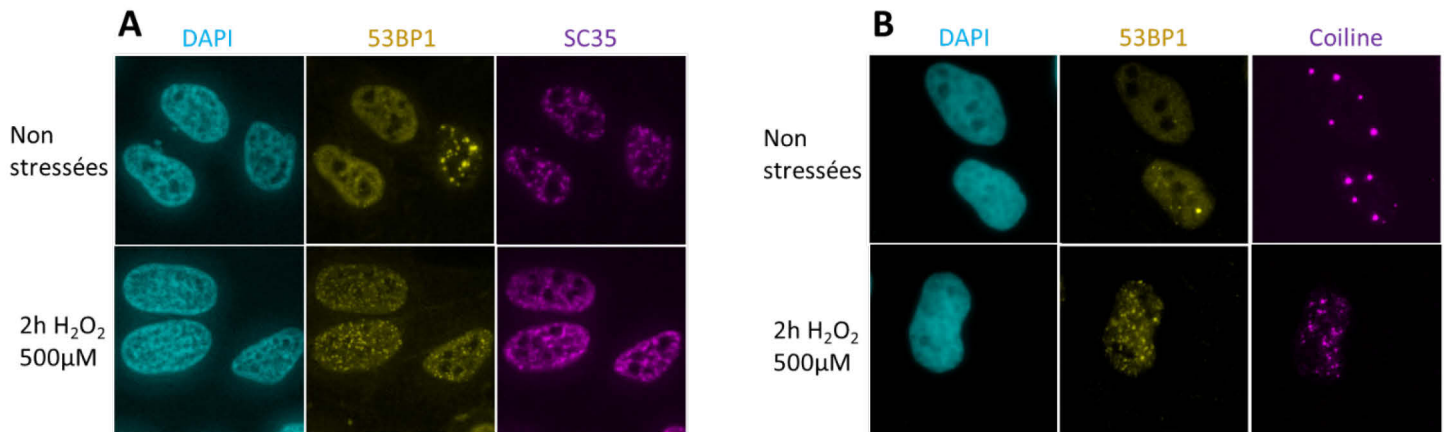
**Figure R22. Localisation du snARN U2 par RNA-FISH A.** Suivi de la localisation du snARN U2 au cours d'une cinétique de stress oxydant (2 heures à 24 heures après induction du stress avec 500μM d'H<sub>2</sub>O<sub>2</sub>). **B.** Suivi de la distribution sub-nucléaire de U2 en réponse au stress oxydant grâce au co-marquage avec la protéine coiline.

## 6. Effet du SO sur différentes substructures nucléaires

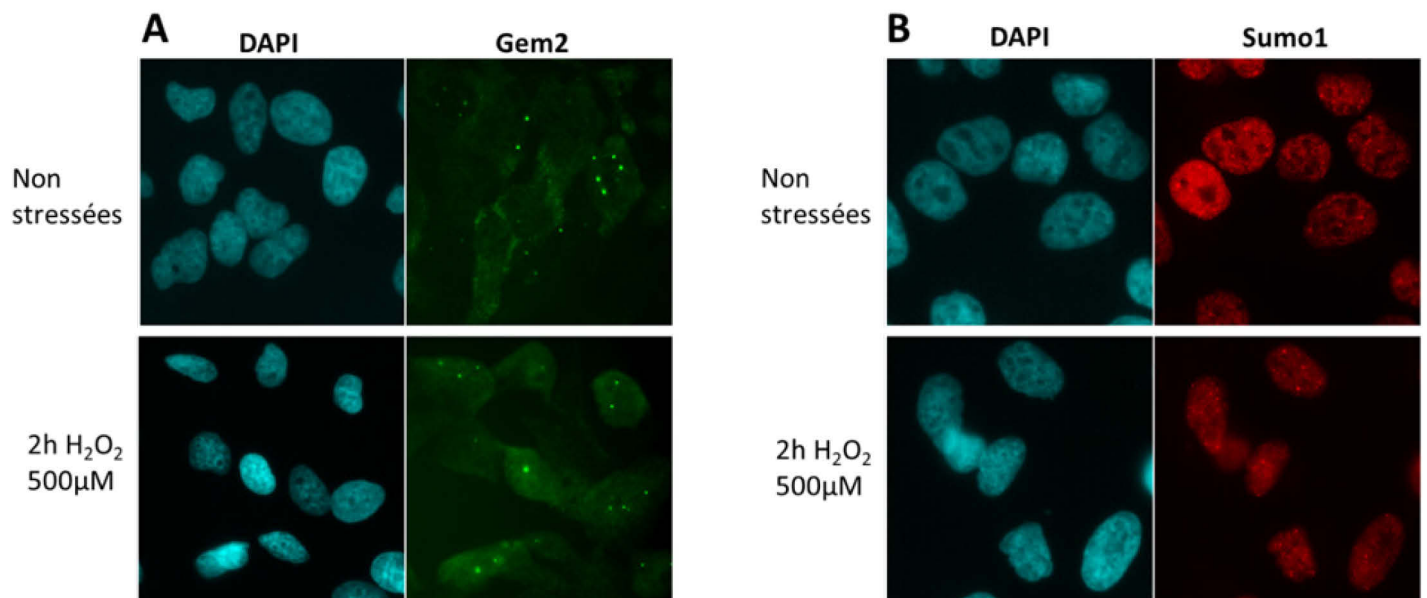
Compte tenu des effets du stress oxydant mis en évidence sur la structure des corps de Cajal, nous avons entrepris d'analyser l'effet de ce type de stress sur d'autres substructures cellulaires. Pour ce faire, nous disposions d'anticorps dirigés contre la protéine SC35 permettant de marquer les speckles nucléaires, mais également les protéines Gem2 et Sumo1 marquant respectivement les gems et les PML bodies (Promyelocytic Leukemia bodies) qui sont tous nucléaires. Comme cité précédemment, les speckles nucléaires permettent notamment de stocker les UsnRNP matures avant leur utilisation pour des réactions d'épissage. Les gems sont des structures qui concentrent le complexe SMN (protéine SMN associées à plusieurs protéines Gemins) après son passage dans les corps de Cajal et avant qu'il soit ré-utilisé ou recyclé dans le cytoplasme. Les gems et les corps de Cajal sont donc deux structures souvent associées les unes aux autres dans le nucléoplasme. Les PML bodies recrutent une grande variété de protéines aux fonctions très différentes. Ces substructures régulent des processus tels que la transcription, l'apoptose, la sénescence, la réponse aux dommages de l'ADN ou la résistance aux micro-organismes (Bernardi and Pandolfi, 2007; Bernardi et al., 2008; Everett, 2006; Salomoni and Pandolfi, 2002). Les gems et les PML bodies forment des foci nucléaires d'une taille généralement comprise entre 0.1 et 1µm.

Les données obtenues montrent que nos conditions de stress oxydant ne modifient pas significativement la forme et le nombre de nuclear speckles (Figure R23A). Comme il avait été observé dans une étude précédente sur les effets des UV-C (Cioce et al., 2006), une partie des cellules stressées présentaient des speckles plus arrondis que dans les cellules témoins. Ces expériences nous ont aussi permis de conclure que contrairement aux corps de Cajal (Figure R23B), le nombre et la structure des gems (Figure R24A) et des PML bodies (Figure R24B) ne sont pas perturbées par nos conditions de stress. De façon intrigante, la régulation des PML bodies par des stress comme le stress oxydant a déjà été décrit dans d'autres études (Villagra et al., 2006; Yamada et al., 2001). De plus, la modification de structure des corps de Cajal ne semble pas affecter les gems qui sont pourtant souvent associés avec ces substructures qui sont toutes deux composées du complexe SMN. Compte tenu de l'importance des corps de Cajal dans la

maturation du spliceosome et des effets observés en réponse au stress oxydant, nous avons choisis de nous concentrer sur l'étude de ces substructures.



**Figure R23. Localisation de la coiline et de SC35 en immunofluorescence lors d'un SO.** Détection de la coiline, SC35 et 53BP1 par immunofluorescence dans les cellules HeLa S3 non traitées ou traitées avec 500 μM d'H<sub>2</sub>O<sub>2</sub>. Les cellules fixées après traitement ont été co-marquées avec des anticorps anti-coilin ou SC35 (canal magenta) et anti-53BP1 (canal jaune).

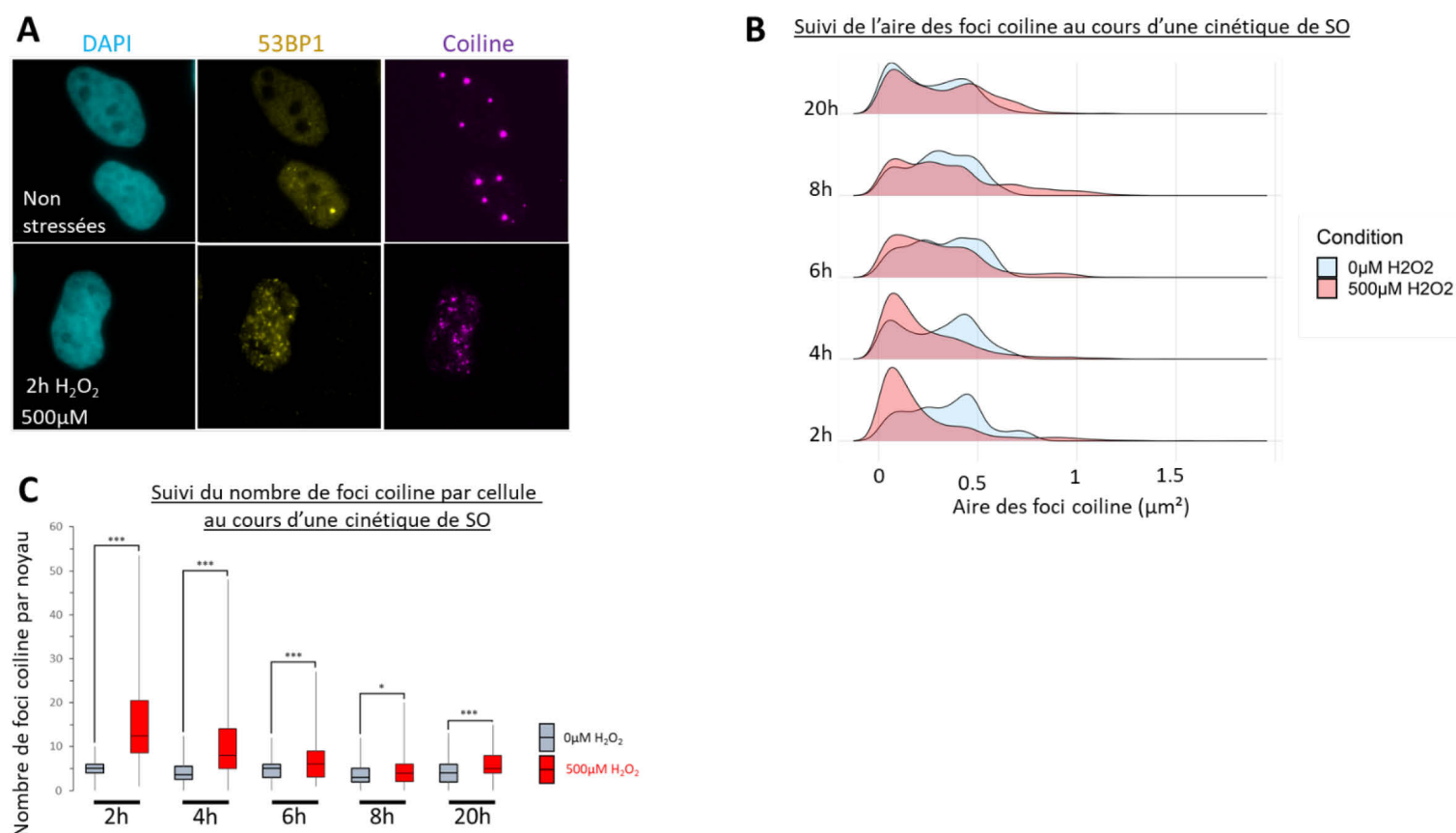


**Figure R24. Localisation de Gem2 et Sumo1 lors d'un SO.** Suivi de la localisation de Gem2 (gems) (A) et de Sumo1 (PML bodies) (B) en immunofluorescence lors d'un stress oxydant. Les cellules fixées 2 heures après l'induction du stress ont été co-marquées avec des anticorps anti-Gem2 ou Sumo1 (canaux vert et rouge respectivement).

## 7. Caractérisation de la redistribution nucléoplasmique de la coiline induite par le SO

Au regard de son rôle central dans la formation des corps de Cajal, la coiline est potentiellement un acteur de la régulation de l'épissage alternatif en réponse au SO. J'ai donc débuté la caractérisation de sa redistribution nucléoplasmique induite par le stress oxydant au cours d'une cinétique de temps. Ainsi, j'ai utilisé le protocole « *Substructure Analyzer* » pour quantifier le nombre et la taille des foci coiline en tenant compte du nombre de foci 53BP1, qui nous permet de discriminer les cellules stressées ou non (cf II les cellules sont stressées si on compte plus de 16 foci 53BP1) (Figure R25A). Cette analyse a été réalisée sur trois séries de SO indépendantes et plus de 2300 cellules. Les données obtenues montrent une augmentation du nombre de foci coiline significative après un stress oxydant, particulièrement pour les temps précoces (2, 4, et 6h) (Figure R25C). L'effet le plus marqué est observé après 2h de SO où le nombre de cellules ayant plus de 10 foci coilin passe de 0% pour les non traitées à 75 % pour les stressées. Cet effet décroît progressivement après 8h de traitement à l'H<sub>2</sub>O<sub>2</sub>. En parallèle de l'augmentation du nombre, nous avons pu mesurer une diminution de la taille des foci coiline en réponse au SO (Figure R25B). Ainsi, après 2h de stress, la proportion de foci ayant une taille inférieure à 0,2µm<sup>2</sup> passe de 26% à 64%. Cet effet diminue également après 8h de traitement.

Comme nous l'avons mentionné dans l'introduction, un grand nombre de facteurs se concentrent dans les CB afin de participer à la biogenèse du spliceosome. Nous avons donc aussi réalisé une analyse de la composition des CB en réponse au stress. Celle-ci sera présentée plus tard dans les résultats afin de présenter dans un seul chapitre toutes les données concernant plus spécifiquement la coiline.



**Figure R25. Cinétique de la distribution nucléoplasmique de la coiline après un stress oxydant.**

L'analyse du nombre et de la taille des foci coiline a été réalisée sur plus de 110 cellules pour chaque point dans 3 expériences indépendantes. **(A)** Détection par immunofluorescence de la coiline (magenta) et de 53BP1 (jaune) dans des cellules HeLa S3 traitées ou non 2h par H<sub>2</sub>O<sub>2</sub> (500µM). **(B)** Analyse de la taille (aire, en µm<sup>2</sup>) des foci coiline dans des cellules traitées ou non au cours d'une cinétique de temps. **(C)** Analyse du nombre de foci coiline dans des cellules traitées ou non au cours d'une cinétique de temps. Un T-test a été réalisé pour voir une différence significative entre les cellules stressées ou non (\*, p<0.05; \*\*, p<0.01; \*\*\*, p<0.001).



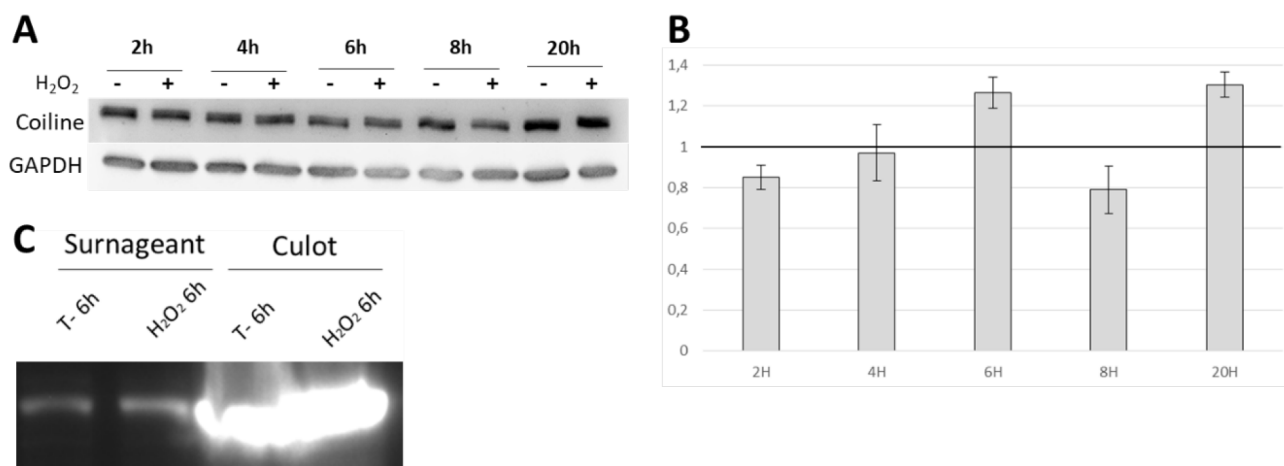
a) Etude des mécanismes à l'origine de la redistribution nucléoplasmique de la coiline

Après avoir mis en évidence un effet du SO sur la distribution de la coiline, nous nous sommes demandés comment cette relocalisation pouvait se mettre en place.

*i. Niveau d'expression de la coiline en réponse au SO*

Le niveau d'expression de la coiline est un facteur déterminant dans la nucléation, la dynamique et la fonction des corps de Cajal. De façon général, la déplétion de la coiline conduit à une disparition des corps de Cajal et une dispersion nucléoplasmique d'un grand nombre de ses composants, comme nous avons pu l'observer pour le transcrit U2. Nous avons donc analysé le niveau d'expression de la protéine coiline par Western Blot. Notre protocole d'extraction de protéines classique consiste à lyser les cellules dans un tampon contenant 1% de détergent Igepal et 150mM de NaCl (détail dans le matériel et méthodes) avec une incubation de 30 min dans la glace, suivie d'une centrifugation. Cependant, au cours de ma thèse, nous avons vu que ce protocole ne permettait de récupérer qu'une faible partie de la coiline, la majorité de la coiline restant dans le culot après l'étape de centrifugation des cellules lysées (Figure R26C). Nous avons donc émis l'hypothèse que l'association de la coiline avec de nombreux complexes et sa forte nucléation dans les cellules HeLa la rendait plus insoluble. Pour optimiser son extraction, les cellules ont directement étaient lysées dans le tampon de lyse nommé Sample Buffer Laemmli et soniquées (afin de casser la chromatine et libérer les protéines). Plusieurs cinétiques de stress oxydant ont été réalisés puis analysées par Western Blot (Figure R26A et R26B). Nos données montrent une très légère diminution de l'expression de la coiline 2 heures et 8 heures après l'induction du stress. Cependant, aucune variation n'est observée à 4 heures et même une légère augmentation au temps 6 heures alors que la redistribution nucléoplasmique de la coiline est encore visible à ces différents temps. Ces résultats ne montrent donc pas de variations d'expression de la coiline cohérentes avec la cinétique de redistribution nucléoplasmique observée en réponse au stress oxydant. Cela suggère que ces changements de structure ne passent probablement pas par une variation de son expression. Cependant, nous pouvons noter que les extraits protéiques sont réalisés avec une population de cellules où 40% des cellules présentent moins de 16 foci 53BP1. De plus, même 2 heures après l'induction du stress, toutes

les cellules ne pr sentent pas un ph notype micro-foci coiline (75%). Ainsi, m me si une grande majorit  des cellules semblent  tre en  tat de stress dans nos conditions, des variations du niveau d'expression de la coiline en r ponse au stress peuvent  tre att nu es du fait que la r ponse cellulaire n'est pas homog ne.



**Figure R26. Analyse de l'expression de la prot ine coiline en r ponse au stress oxydant.** (A et B) Les cellules HeLa S3 ont  t  trait es ou non par H<sub>2</sub>O<sub>2</sub> (500 M) au cours d'une cin tique de temps. **(A)** La quantit  de coiline pr sente dans ces extraits a  t  analys e en WB avec un anticorps anti-coiline. **(B)** La quantit  de coiline est rapport e   la quantit  de GAPDH qui sert de contr le de l'homog n it  des d p ts. La quantit  de coiline dans la condition de stress est rapport e   la condition non stress e. Les quantifications ont  t  r alis es sur 3 exp rimentations diff rentes. **(C)** Apr s un stress de 6h, les cellules ont  t  lys es et les prot ines pr sentes dans le surnageant et le culot ont  t  analys es en WB.

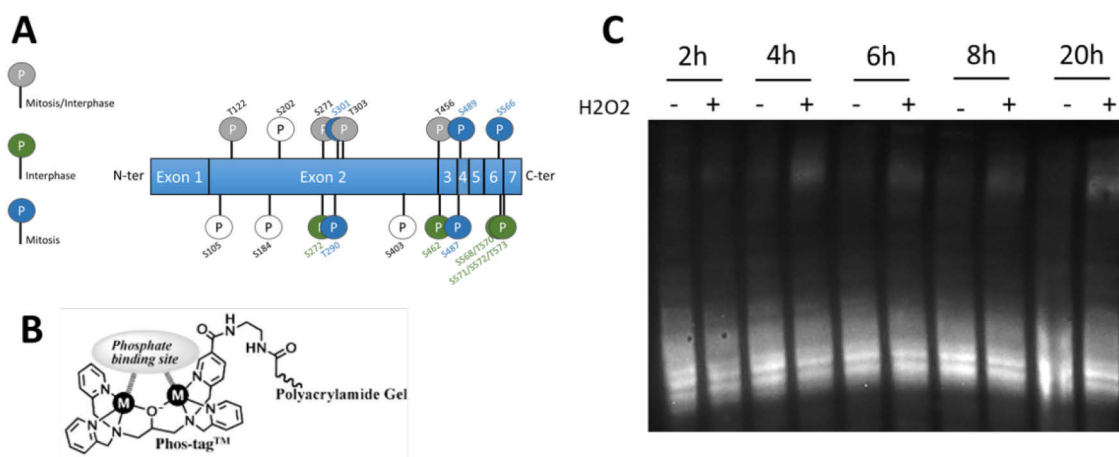
## ii. Effets du SO sur le cycle cellulaire et la phosphorylation de la coiline

Comme mentionn  dans l'introduction, la nucl ation des CBs est sujette   r gulation au cours du cycle cellulaire afin de s'adapter aux besoins de la cellule. En effet, les CBs pr sentent une nucl ation maximale aux cours des phases G1 et S, lorsque l'activit  transcriptionnelle est la plus importante, et sont d mantel s juste avant l'entr e de la cellule en mitose (C. Andrade, 1991; Misteli, 2001). Nous avons fait l'observation, lors de la caract risation de l'effet de nos conditions

de stress sur nos cellules (cf II des Résultats), de l'induction par notre SO de la mort cellulaire par apoptose et d'un effet sur la progression du cycle cellulaire. Nous avons donc émis l'hypothèse que la redistribution de la coiline pourrait être due à cette perturbation du cycle cellulaire induite par le SO. Cependant, notre analyse de la distribution des cellules dans les différentes phases du cycle cellulaire au cours d'une cinétique de stress oxydant (Figure R9) montre un léger arrêt des cellules en phase S au dépend des phases G2/M aux temps précoces (2h à 8h) suivant l'induction du stress. Or la nucléation des CBs est décrite être maximale au cours de la phase S (C. Andrade, 1991; Misteli, 2001). Ces données suggèrent donc que la redistribution nucléoplasmique de la coiline induite par le stress oxydant n'a pas pour origine la perturbation du cycle cellulaire.

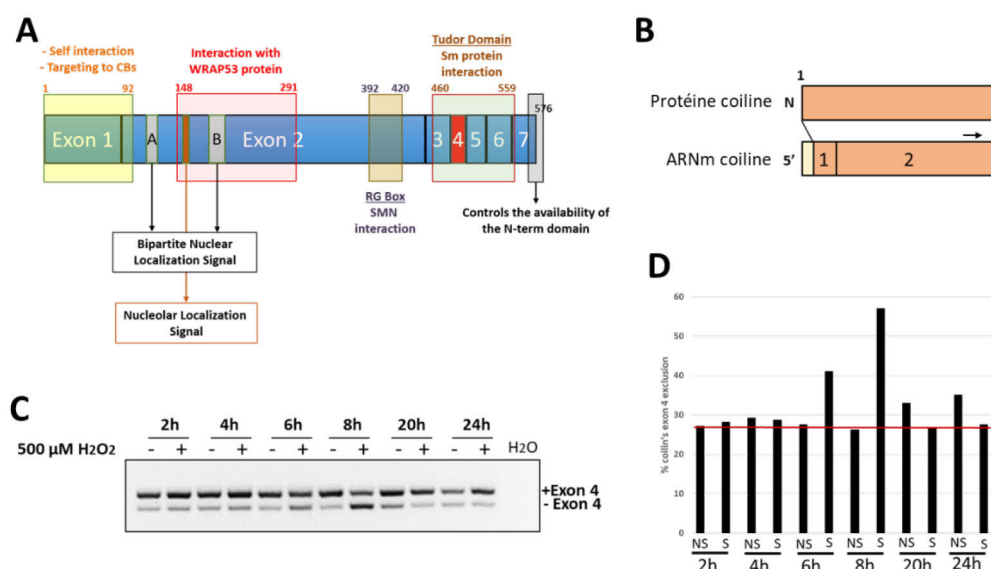
Le potentiel de nucléation de la coiline est également contrôlé par son état de phosphorylation qui est lui-même régulé par d'autres processus tel que le cycle cellulaire (Figure R27A). Il a notamment été proposé que la présence de certains résidus phosphorylés régule l'interaction de la coiline avec ses partenaires SMN et SmB', contrôlant ainsi l'accumulation des UsnRNP en cours de maturation au sein des CBs (Toyota et al., 2010). Afin d'étudier le rôle potentiel de ces modifications post-traductionnelles dans la redistribution nucléoplasmique de la coiline, nous avons récemment débuté l'analyse de son état de phosphorylation. Actuellement, nous optimisons les conditions expérimentales pour réaliser des gels Phos-tag™ (Figure R27B et R27C), ce qui devrait permettre de donner un premier aperçu des différents états de phosphorylation de la coiline après un stress oxydant. En effet, ces gels ont la particularité de contenir la molécule appelée Phos-tag™ qui interagit avec les résidus phosphorylés des protéines.

Les protéines phosphorylées migrent alors moins vite que leurs homologues non phosphorylées. Ces analyses seront complétées par la suite par des expériences de spectrométrie de masse.



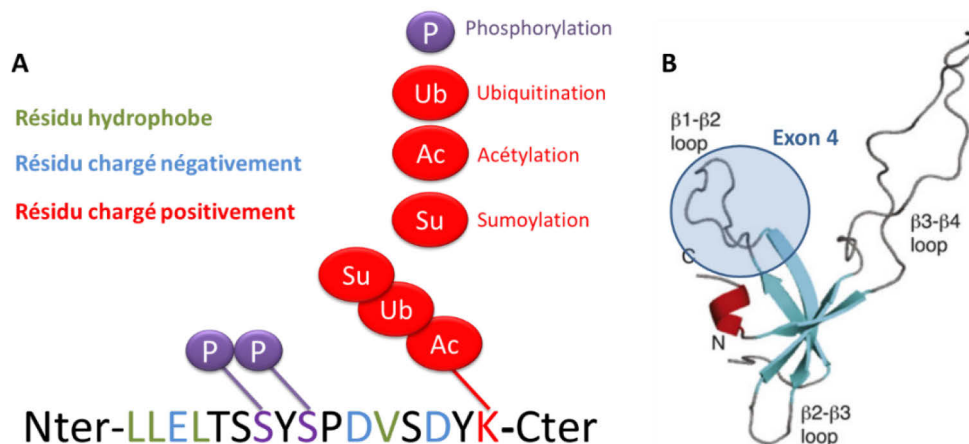
cet exon constitue une partie du domaine Tudor impliqué dans l'interaction avec les protéines Sm et donc dans le recrutement des UsnRNP au sein des corps de Cajal (Figure R28A). Ainsi, nous avons supposé que la perte d'un tel exon pouvait être à l'origine de la perturbation de la structure des corps de Cajal et de la redistribution nucléoplasmique des UsnRNPs que nous avons observée pour le snARN U2, en réponse au SO.

Nous avons donc entrepris de valider par RT-PCR semi-quantitative l'augmentation de l'exclusion de l'exon 4 détectée sur les puces. Cette augmentation a été confirmée grâce à l'utilisation d'un couple d'oligonucléotides couvrant une partie de l'exon 2 et la jonction des exons 5 et 6, afin de ne cibler que les transcrits épissés (Figure R28B, C et D).



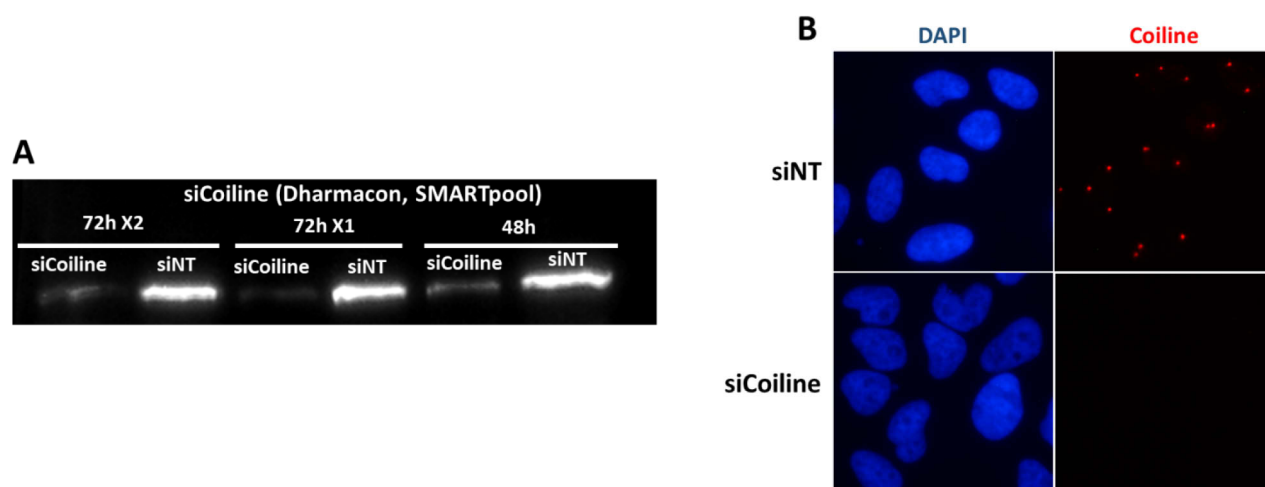
**Figure R28. Effet du stress oxydant sur l'exclusion de l'exon 4 de la coiline.**

**(A)** Schéma de l'organisation structurale de la coiline. **(B)** Positionnement des oligonucléotides sur l'ARNm de la coiline **(C et D)** Analyse par RT-PCR de l'épissage alternatif de l'exon 4 de la coiline, les produits PCR ont été fractionnés sur gel d'agarose et les bandes correspondant aux 2 produits attendus ont été détectées et quantifiées. Sur le graphique, la proportion de l'isoforme sans l'exon 4 est représentée pour chaque condition.



**Figure R29. Composition et structure de l'exon 4 de la coiline** **A.** Propriétés et modifications post-traductionnelles connues des 16 résidus de l'exon 4 de la coiline. **B.** Représentation en ruban de la structure d'une partie de la coiline [460-576] obtenue par RMN (Adapté de Shanbhag et al. 2010). Cette représentation comprend le domaine Tudor de la protéine [460-559] impliqué dans l'interaction avec les protéines Sm et donc dans l'adressage des U snRNPs en cours de maturation dans les corps de Cajal. La partie correspondant à l'exon 4 est mise en évidence par un cercle bleu.

Nous avons alors mis en place différentes stratégies pour déterminer les rôles fonctionnels de la coiline entière et de la coiline déléetée de son exon 4. Nous avons notamment travaillé sur l'optimisation de l'expression exclusive de cette isoforme. Pour cette étude, j'ai produit des constructions plasmidiques permettant l'expression dans les cellules eucaryotes soit de l'isoforme entière, soit de l'isoforme dépourvue de l'exon 4, fusionnée à une étiquette V5 ou GFP (pCDNA-3.1-V5 et pEGFP\_N1). Afin d'étudier exclusivement l'expression des formes exogènes de coiline amenées par le plasmide nous avons fait produire des gènes synthétiques (gBlocks gene Fragments, Integrated DNA Technologies) dans lesquels des mutations silencieuses ont été introduites au niveau des sites de reconnaissance des siARN utilisés au laboratoire pour éteindre l'expression de la coiline endogène. Ainsi l'expression des différentes isoformes peut être étudiée en absence de coiline endogène. Les résultats des optimisations de la déplétion de la coiline endogène sont présentés dans la Figure R30.



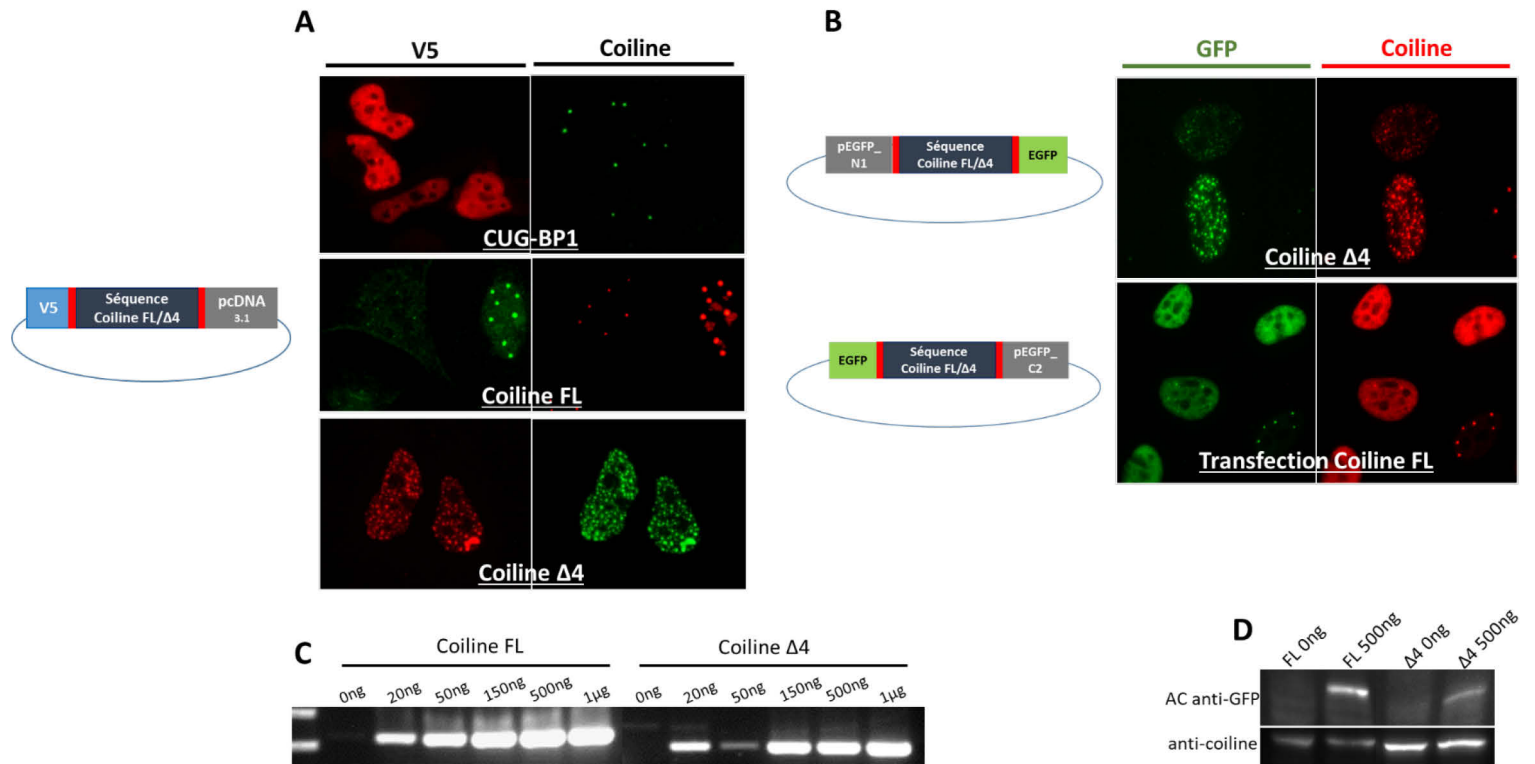
**Figure R30. Optimisation de la déplétion de la coiline endogène par interférence ARN.**

Les cellules ont été transfectées avec un pool de 4 siRNA dirigés contre l'ARNm de la coiline (siCoiline) ou avec des siRNA sans cible (NT) par la technique de phosphate de calcium. Les cellules ont été récupérées après 48 h ou 72h de transfection, et elles ont subi une (X1) ou deux (X2) transfections à 48h d'intervalle. L'efficacité de l'extinction a été contrôlée en WB **(A)** et en IF **(B)**.

Après avoir optimisé le silencing de la coiline endogène, nous avons optimisé les conditions de transfection transitoire des cellules HeLa avec les différentes constructions exprimant les isoformes de coilin que nous avons générées. Nous avons tout d'abord testé le plasmide pcDNA 3.1 codant la protéine entière (coiline Full Length) ou la protéine sans exon 4 ( $\Delta 4$ ), fusionnées à un tag V5 en N-terminal de la protéine (Figure R31A). Nos premières données montrent que l'isoforme  $\Delta 4$  de la coiline n'est pas nucléée dans des corps de Cajal, mais se regroupe plutôt dans de nombreux microfoci nucléaires, alors que la coiline FL transfectée forme des foci de type corps de Cajal dont la taille augmente proportionnellement avec le niveau d'expression de la protéine (Figure R31A). Lorsque nous transfectons les cellules avec une protéine contrôle (V5-CUG-BP 1), la localisation de la coiline endogène n'est pas affectée (Figure

R31A) signifiant que les localisations observées sont bien spécifiques à la coiline. L'exclusion de l'exon 4 semble donc affecter la capacité de la coiline à former des foci de type corps de Cajal.

Nous avons complété cette analyse en sur-exprimant la protéine coilin fusionnée en C-terminal au tag EGFP qui nous a permis de visualiser directement la protéine exogène en microscopie sans utiliser d'anticorps grâce à la fluorescence de la GFP (l'idée étant de pouvoir faire par la suite des analyses en live imaging). Pour chaque transfection, j'ai validé la surexpression de la coiline par PCR semi-quantitative (Figure R31C) et Western Blot (Figure R31D). L'isoforme  $\Delta 4$  ainsi que la protéine FL fusionnées à ce nouveau tag présentaient une localisation similaire à celles observées avec le tag V5 (Figure R31B, en haut). Nous avons également testé de fusionner l'EGFP au niveau de l'extrémité N-terminale de la coiline, mais la capacité de nucléation



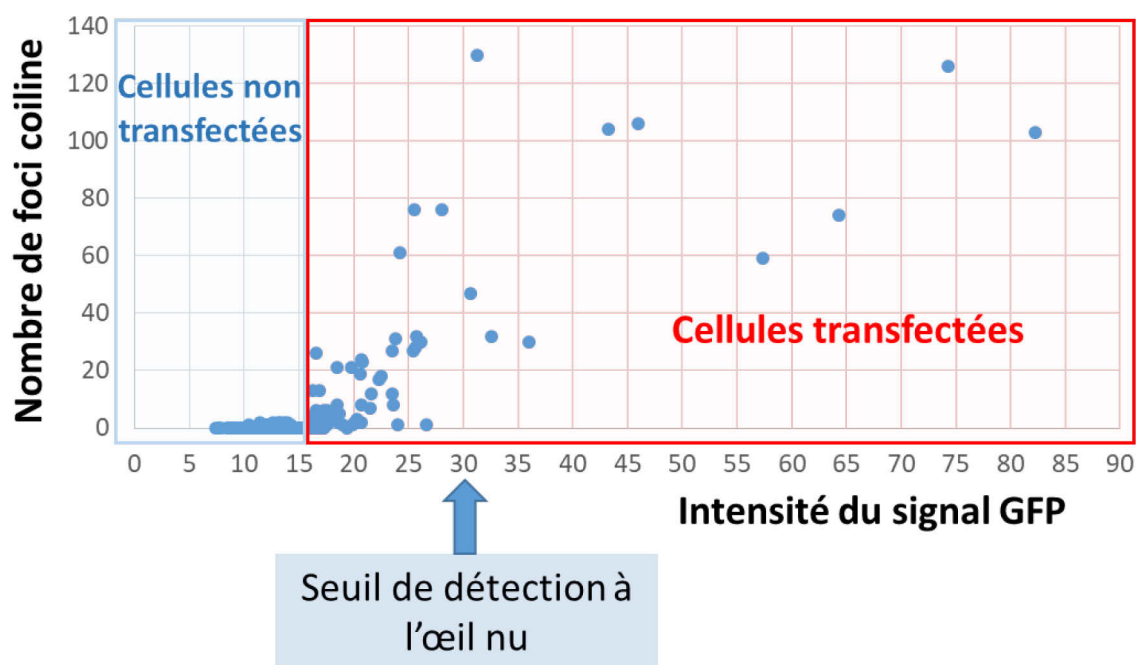
**Figure R31. Mise au point de la surexpression de la coiline et de la coiline  $\Delta 4$**  (A) Localisation en IF de la coiline endogène et des protéines couplées au tag V5, la protéine CUG-BP, la coiline et l'isoforme delta 4 de la coiline. (B) Localisation en IF de la coiline  $\Delta 4$  avec un tag GFP en N-terminal ou en C-terminal ou de la coiline entière (FL) avec un tag GFP en C-terminal. (C) RT-PCR en point final pour suivre le niveau de surexpression des isoformes d'ARNm coilin. (D) Western blot pour suivre le niveau de surexpression des isoformes de coiline couplées à la GFP.



de la protéine exogène dans les corps de Cajal semblait être affectée lorsque la surexpression dépassait un certain seuil (Figure 31B, en bas). En effet, les cellules les plus faiblement transfectées présentent un phénotype corps de Cajal classique alors que la protéine exogène se redistribue de façon homogène dans le nucléoplasme dans les cellules où le signal GFP est plus fort. Ce phénotype a été observé pour les protéines FL et  $\Delta 4$ . L'extrémité N-terminale de la coiline jouant un rôle dans l'oligomérisation de la protéine et son adressage dans les corps de Cajal (Figure R28A), il est possible que la présence d'un tag volumineux comme la GFP puisse perturber les fonctions de ce domaine. Sachant qu'il est connu qu'une grande fraction de coiline est nucléoplasmique et ne se concentre pas dans les corps de Cajal, les résultats obtenus avec le tag EGFP en N-terminal de la protéine n'en demeurent pas moins intéressants.

Suivant le nombre de copies de plasmide qui a été intégré, ce type de transfection transitoire conduit généralement à des populations de cellules exprimant de façon hétérogène la protéine exogène. Ainsi, en immunofluorescence, l'intensité du signal est liée à la quantité de plasmide que la cellule a intégrée : les cellules présentant un signal fort sont plus facilement détectables et analysables par l'utilisateur, bien que les données provenant de ces cellules ne soient pas forcément les plus représentatives compte tenu du fort niveau de surexpression de la protéine exogène. Nous nous sommes donc interrogé sur la pertinence des résultats obtenus pour la coiline  $\Delta 4$ , et avons émis l'hypothèse que le phénotype observé pouvait être causé par un niveau de surexpression trop important. J'ai donc développé un nouveau protocole informatique permettant d'automatiser l'analyse de la localisation de la protéine exogène en prenant en compte son niveau de surexpression. Ce protocole permet de considérer l'hétérogénéité du niveau de surexpression dans l'ensemble des cellules transfectées. L'automatisation de l'analyse était d'autant plus pertinente qu'une partie des cellules transfectées n'est pas visible à l'œil nu lorsque le signal moyen de la GFP est inférieur à 30 environ (valeur qui peut varier selon l'observateur). Plusieurs centaines de cellules non transfectées ont tout d'abord été analysées afin de déterminer le seuil d'intensité du signal GFP permettant de discerner les cellules non transfectées des cellules transfectées. Les foci coiline ont ensuite été quantifiés ainsi que l'intensité du signal GFP pour chaque cellule transfectée avec la coiline  $\Delta 4$  taggée avec l'EGFP en C-terminal (Figure R32). Grâce à cette analyse, j'ai pu mettre en évidence que le nombre de foci

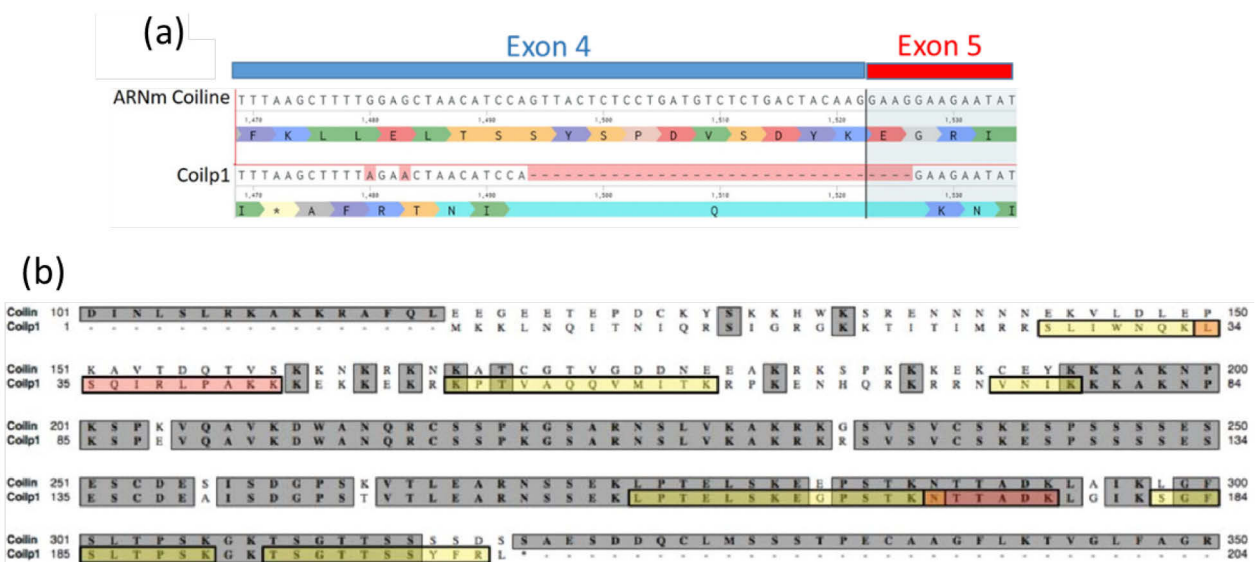
coiline augmente significativement y compris dans les cellules où le niveau de surexpression est faible (entre 15 et 20). La localisation observée pour la coiline  $\Delta 4$  est donc propre à la protéine elle-même et n'est pas un effet dû à un fort niveau de surexpression.



**Figure R32. Suivi du nombre de foci coiline en fonction du niveau de surexpression de la protéine coiline  $\Delta 4$ -EGFP chez des cellules transfectées (tag EGFP en C-ter).** Lorsque l'intensité du signal est inférieure à un certain seuil, il est considéré que les cellules n'ont pas ou peu intégré de plasmide.

A ce stade de notre étude sur les effets du SO sur la coiline, nous avons donc mis en évidence une redistribution nucléoplasmique, sous forme de microfoci, de la coiline, ainsi que la production d'une nouvelle isoforme d'ARNm coiline, déletée de l'exon 4, permettant la production d'un variant de coilin ayant un pattern de localisation très similaire à celui observé en réponse au stress oxydant. Nous n'avons, cependant, pas encore pu prouver la production de la protéine coiline  $\Delta 4$  et toutes nos tentatives pour détecter cette isoforme avaient jusqu'alors échouées, et ce, malgré l'utilisation d'anticorps reconnaissant des peptides conservés entre la

coilin FL et la coilin  $\Delta 4$ . Afin de nous assurer qu'il n'y avait pas de décalage du cadre de lecture lié à l'utilisation, par exemple d'un site cryptique d'épissage à proximité des bornes exoniques de l'exon 4, nous avons décidé de cloner et séquencer le produit de RT-PCR amplifié lors de l'analyse de l'inclusion de l'exon 4. Après alignement des séquences obtenues, nous avons constaté que ce produit ne correspondait pas à l'isoforme  $\Delta 4$ , mais au transcrit d'un pseudogène de la coiline appelé *Coilp1*. Ce transcrit présente une délétion partielle de l'exon 4 et de l'exon 5, conduisant à un produit d'amplification par RT-PCR d'une taille proche de celle attendue pour la délétion complète de l'exon 4 (Figure R33A), d'où la confusion sur gel et la détection sur puces.



**Figure R33. Similarité de séquence nucléotidique et peptidique de Coilp1 et de la coiline.**

**(a)** Alignement des séquences des produits de RT-PCR coilin et coilp1. **(b)** Alignement du polypeptide Coilp1 avec la coiline (Poole et al., 2016). Cet alignement montre une forte conservation des résidus 194 à 315 de la coiline correspondant à une partie du domaine d'interaction avec la protéine WRAP53.

Une étude récente a caractérisé un polypeptide de 28kDa comme étant un produit de l'expression de ce pseudogène coilp1 (Poole et al., 2016) (Figure R33B). Ce peptide de 204 acides aminés présente une très forte similitude avec les acides aminés 194 à 315 de la coiline correspondant à une partie du domaine d'interaction avec la protéine WRAP53. Le peptide Coilp1 est enrichi dans le nucléoplasme et dans le nucléole. Cependant, il n'est pas concentré dans les

CBs, ce qui pourrait être expliqué par l'absence du domaine de « self interaction » (Figure R28A), indispensable pour induire la nucléation des CBs. En revanche, Coilp1 interagit avec les scaRNAs 2 et 9 ainsi qu'avec l'ARN de la télomérase hTR qui sont tous concentrés dans les CBs. Ces données suggèrent que Coilp1 est capable d'interagir avec certains partenaires de la coiline, mais qu'il ne présente pas le même potentiel de nucléation nécessaire à l'assemblage des CBs. Coilp1 contient une partie du domaine d'interaction avec la protéine WRAP53, mais ne possède pas les domaines d'interactions avec SMN et les protéines Sm. Une augmentation de son expression pourrait donc potentiellement perturber l'interaction de la coiline avec ses partenaires ainsi que la nucléation des CBs. Cependant, l'augmentation du transcrit Coilp1 intervient entre 5 heures et 8 heures après l'induction du stress, alors que la redistribution nucléoplasmique de la coiline est observée dès 2 heures. Bien qu'il soit possible que Coilp1 ait un effet sur la dynamique des CBs, nos données suggèrent que le pseudogène n'est pas à l'origine de la redistribution de la coiline en réponse au stress oxydant. Il est à noter que les anticorps utilisés en WB et en IF ne reconnaissent pas d'épitopes de Coilp1 et reflètent donc seulement l'expression et la localisation de la coiline.

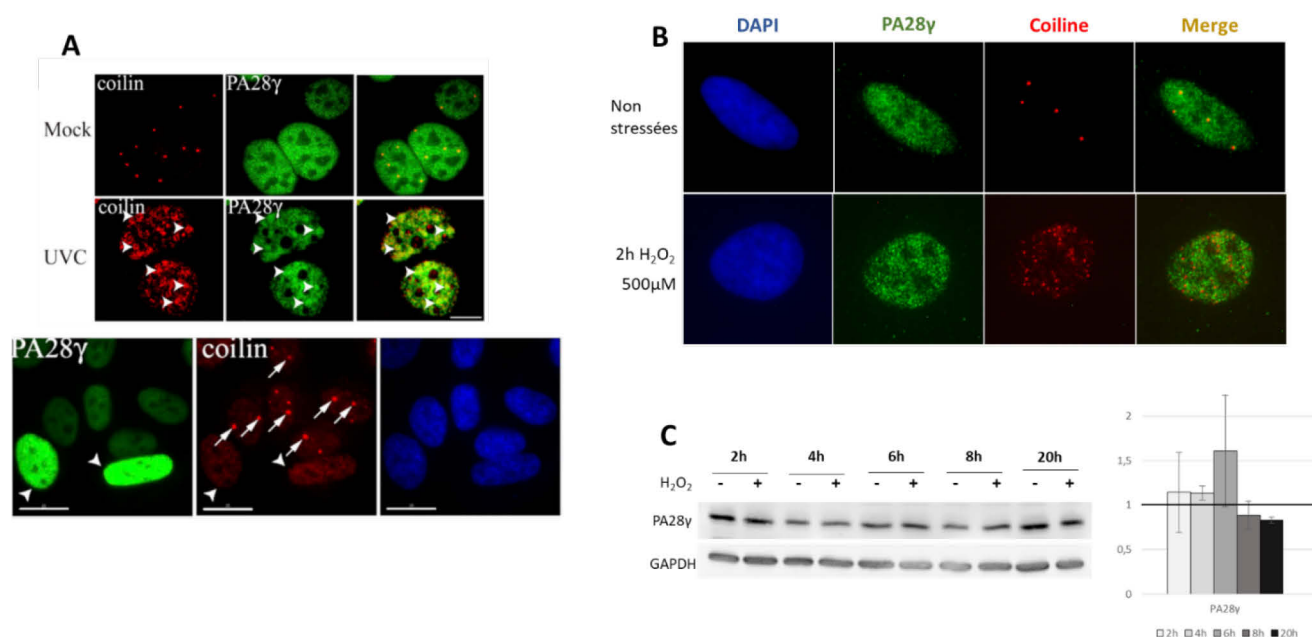
#### 8. Etude de l'effet du stress oxydant sur les composants des CB

Comme nous l'avons vu au début du chapitre précédent, le stress oxydant change le pouvoir de nucléation des CBs induisant une redistribution de la coilin en foci nucleoplasmiques plus nombreux et de plus petite taille. Or les CB ont des localisations précises dans le noyau et peuvent affecter l'expression des gènes en s'associant à des loci spécifiques (Wang et al., 2016). Ce changement dans leur structure pourrait s'accompagner d'un changement dans leur fonction. Cependant nous devons aussi nous demander si ce remodelage est une simple conséquence du SO ou bien s'il s'agit un acteur important dans la réponse au stress.

Les résultats de localisation du transcrit U2 montraient une diminution maximale des foci nucléaires entre 4 et 8 heures après l'induction du stress. Il semblerait donc que le stress affecte la nucléation des différents composants des corps de Cajal de manière séquentielle. Nous avons donc étudié l'effet du stress oxydant sur la localisation d'autres composants des CB, ainsi que sur leur niveau d'expression.

a) Effets du stress oxydant sur la localisation et l'expression de PA28 $\gamma$ 

Comme mentionné en introduction, le traitement aux UV-C de cellules cancéreuses MCF 7 conduit également à une redistribution nucléoplasmique de la coiline sous forme de microfoci (Cioce et al., 2006) (Figure R34A). Les auteurs ont, tout comme nous, cherché à comprendre les bases moléculaires de cette redistribution et ont mis en évidence une augmentation de l'interaction de la coiline avec le régulateur de l'immunoprotéasome PA28 $\gamma$  C (Cioce et al., 2006). Ils ont également observé que la surexpression de PA28 $\gamma$  induit une redistribution de la coiline (Figure R35A), et de certains de ses partenaires, similaire à celle induite par le traitement aux UV-C et que, à l'opposé, l'extinction de PA28 $\gamma$  atténue la redistribution induite par le traitement aux UV-C (Cioce et al., 2006). Sachant que ce type de rayonnement génère des espèces réactives de l'oxygène (de Jager et al., 2017), nous nous sommes donc interrogés sur le rôle potentiel de PA28 $\gamma$  dans la redistribution nucléoplasmique de la coiline suite à un stress oxydant.



**Figure R34. Effet de différents stress sur la localisation et le niveau d'expression de PA28 $\gamma$ .**

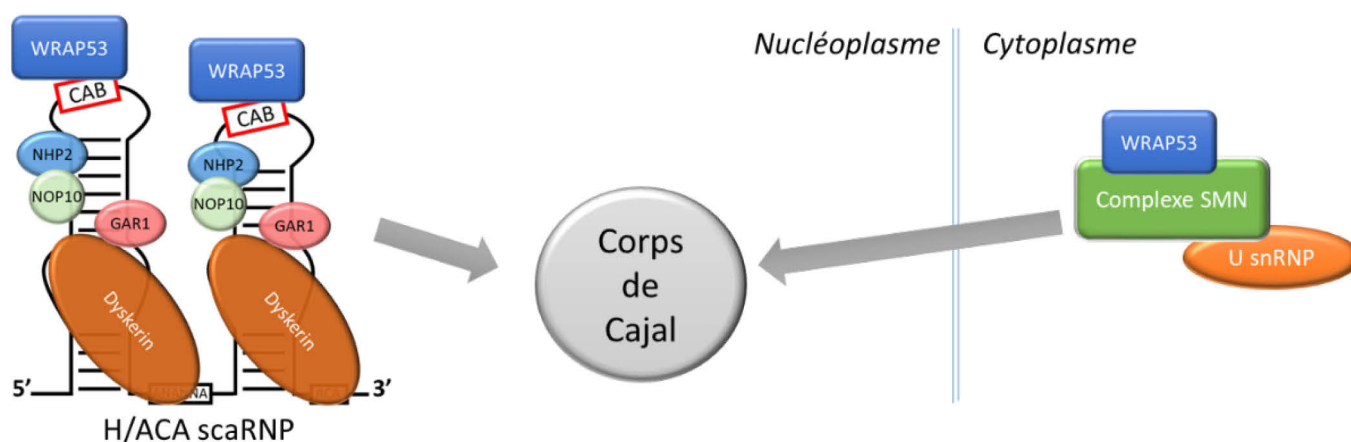
(A) Images issues de Cioce et al 2006 : panneau du haut, PA28 $\gamma$  (vert) et la coiline (rouge) colocalisent partiellement dans des cellules HaCaT traitées aux UV-C-treated cells ; Panneau du bas, des cellules MCF-7 ont été transfectées avec le vecteur d'expression FLAG-PA28 $\gamma$  et marquées avec des anticorps anti-PA28 $\gamma$  (vert) et anti-coilin (rouge). (B) Localisation des protéines coiline (rouge) et PA28 $\gamma$  (vert) dans des cellules traitées avec H<sub>2</sub>O<sub>2</sub> 2h ou non. (C) Les cellules HeLa S3 ont été traitées ou non par H<sub>2</sub>O<sub>2</sub> (500 $\mu$ M) au cours d'une cinétique de temps. La quantité de PA28 $\gamma$  présente dans les extraits protéiques a été analysée en WB avec un anticorps anti-PA28 $\gamma$ . La quantité de PA28 $\gamma$  est rapportée à la quantité de GAPDH qui sert de contrôle de l'homogénéité des dépôts. La quantité de PA28 $\gamma$  dans la condition de stress est rapportée à la condition non stressée. Les quantifications ont été réalisées sur 3 expérimentations différentes.

Nous avons tout d'abord suivi la localisation de PA28 $\gamma$  en IF en conditions de stress oxydant en effectuant un co-marquage avec la coiline (Figure R34B). Bien que le signal de PA28 $\gamma$  ne soit pas exclu des zones où la coiline se redistribue, nos images ne montrent pas d'enrichissement spécifique de PA28 $\gamma$  au niveau des micro-foci de coiline. Ces premières observations doivent être affinées par de la microscopie confocale. Le niveau d'expression de PA28 $\gamma$  au cours d'une cinétique de SO a aussi été analysé par Western Blot (Figure R34C), mais aucune variation significative de son expression n'a été observée. Ces données nous ont donc conduit à invalider l'hypothèse d'un rôle majeur de PA28 $\gamma$  dans la redistribution de la coiline en réponse à un stress oxydant.

#### b) Effet du stress oxydant sur la localisation et l'expression de WRAP53

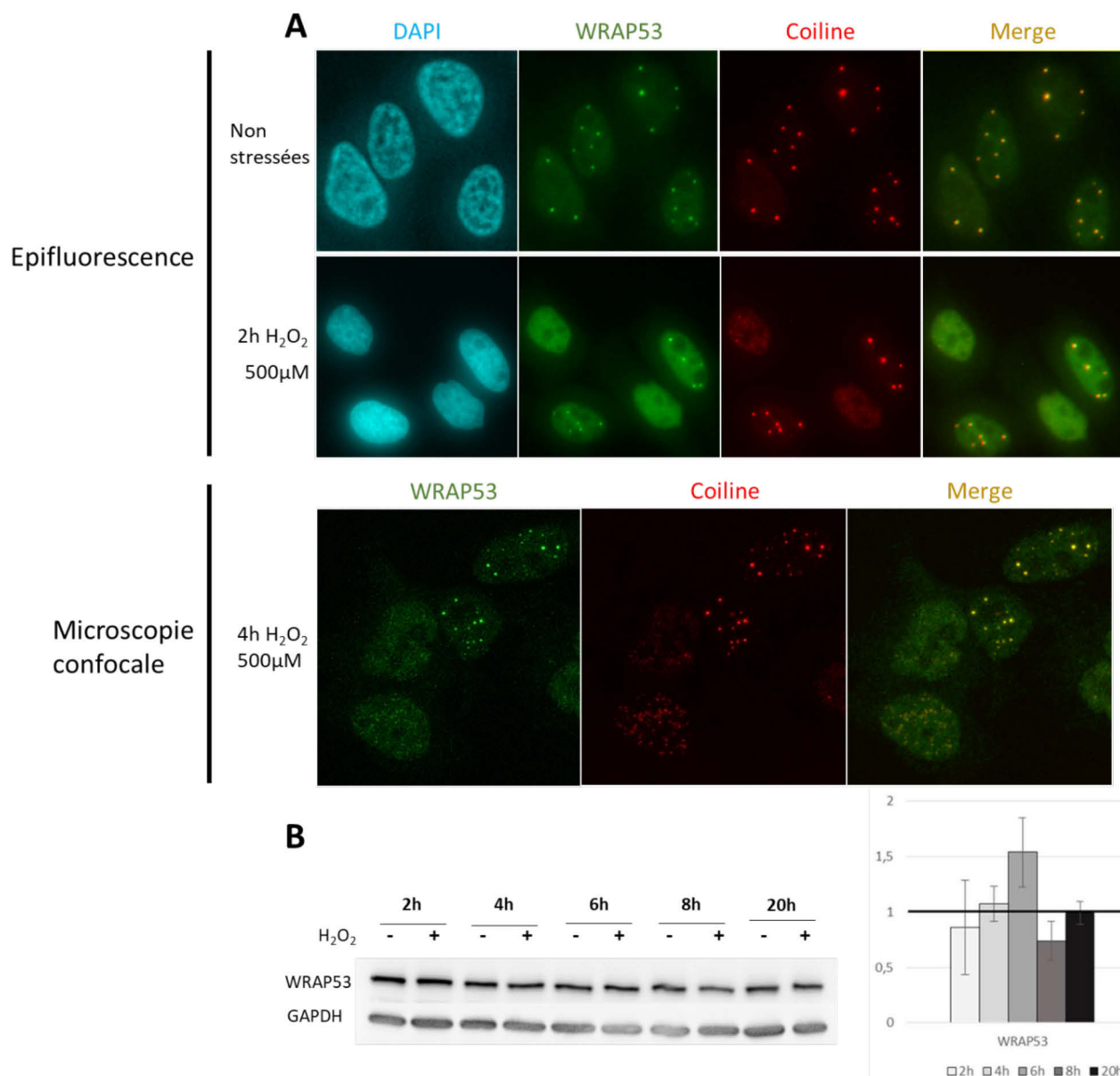
Parmi les composants des CBs impliqués dans la structuration des corps de Cajal mais également dans la maturation du spliceosome, nous nous sommes intéressés à la protéine WRAP53 (= TCAB1 = WDR79) (Mahmoudi et al., 2010). Cette protéine permet de retenir les scaRNP contenant une ou plusieurs boîtes H/ACA au sein des CBs via son interaction avec la CAB-box de leur ARN (Tycowski et al., 2009). WRAP53 serait aussi impliquée dans le ré-import nucléaire et l'adressage des UsnRNP en cours de maturation dans les CBs en interagissant avec le complexe SMN (Figure R35). De plus, il a été récemment montré que WRAP53 est recrutée au niveau des sites de cassures de l'ADN après exposition aux radiations ionisantes dans des cellules cancéreuses (U2OS, H1299) et des fibroblastes humains (Henriksson et al., 2014). Nous avons donc émis l'hypothèse que le recrutement de WRAP53 au niveau des cassures de l'ADN induites par le stress oxydant pourrait être à l'origine de la redistribution de la coiline. Dans un premier temps le niveau d'expression de WRAP53 au cours d'une cinétique de SO a été analysé par western blot, mais aucune variation significative n'a été mise en évidence (Figure R36B). Concernant la localisation de WRAP53, nous avons constaté que, tout au long de la cinétique de SO, WRAP53 est redistribuée dans les mêmes microfoci que la coiline. Il semble donc que la redistribution nucléoplasmique de la coiline ne perturbe pas son interaction avec WRAP53. Cependant, nous n'avons pas pu mettre en évidence un recrutement de WRAP53 dans d'autres foci pouvant correspondre aux sites de cassure de l'ADN (Figure R36A). Nous venons seulement

d'acquérir un anticorps souris dirigé contre la protéine  $\gamma$ H2AX, fonctionnant en immunofluorescence, et qui pourra être utilisé avec l'anticorps anti-WRAP53 lapin pour confirmer l'absence de recrutement de WRAP53 au niveau des cassures double brin lors d'un stress oxydant. Des expériences d'IP sont prévues afin de confirmer que la coiline et WRAP53 continuent d'interagir en conditions de stress oxydant.



**Figure R35. Rôles fonctionnels de WRAP53 au sein des corps de Cajal**

WRAP53 permet de retenir les scaRNP contenant une ou plusieurs boîtes H/ACA au sein des CBs via son interaction avec la CAB-box de leur ARN (Tycowski et al., 2009). WRAP53 serait aussi impliquée dans le ré-import nucléaire et l'adressage des UsnRNP en cours de maturation dans les CBs en interagissant avec le complexe SMN



**Figure R36. Effet du stress oxydant sur la localisation et le niveau d'expression de la protéine WRAP53.**

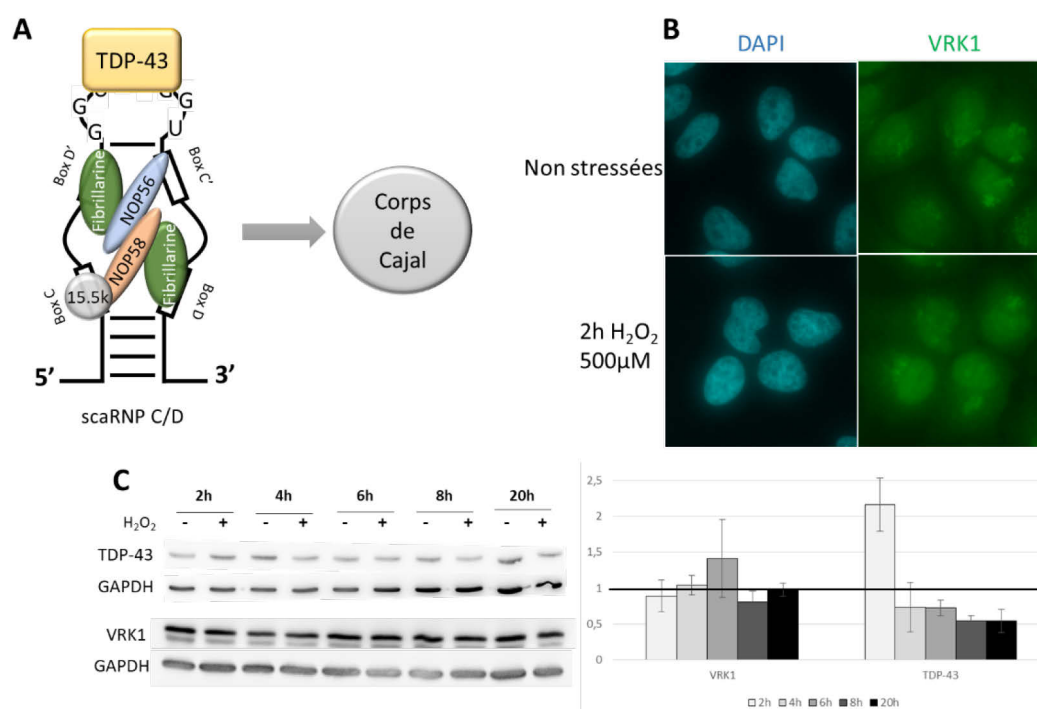
**(A)** Détection par immunofluorescence de la coilin (rouge) et de WRAP53 (vert) dans des cellules HeLa S3 traitées ou non 2h ou 4h par H<sub>2</sub>O<sub>2</sub> (500μM). Les images ont été acquises soit sur un microscope à épifluorescence soit sur un microscope confocale. **(B)** Les cellules HeLa S3 ont été traitées ou non par H<sub>2</sub>O<sub>2</sub> (500μM) au cours d'une cinétique de temps. La quantité de WRAP53 présente dans ces extraits a été analysée en WB avec un anticorps anti-WRAP53. La quantité de WRAP53 est rapportée à la quantité de GAPDH qui sert de contrôle de l'homogénéité des dépôts. La quantité de WRAP53 dans la condition de stress est rapportée à la condition non stressée. Les quantifications ont été réalisées sur 3 expérimentations différentes



## c) Effet du stress oxydant sur la localisation et l'expression de TDP43

L'expression de la protéine TDP-43 (Tat-activating regulatory DNA-binding protein-43) est également nécessaire au maintien de la structure des corps de Cajal (Sawyer et al., 2016). De plus, il a été récemment montré que la protéine TDP-43 permet de retenir des scaRNPs à boîte C/D en interagissant avec des motifs ARN riches en G et en U (Izumikawa et al., 2019) (Figure R37). Cette protéine résidant principalement dans le noyau a été initialement décrite comme interagissant avec des motifs riches en UG retrouvés au sein de milliers de pré-ARNm (Ayala et al., 2008). De façon intéressante, TDP-43 est décrit comme un régulateur de l'épissage en se fixant de façon direct ou indirect au niveau d'éléments *cis* introniques (Buratti et al., 2001; D'Ambrogio et al., 2009; Jeong et al., 2017; Rabin et al., 2010; Sephton et al., 2011; Tollervey et al., 2011). L'étude récente d'Izumikawa suggère la possibilité d'autres mécanismes de régulation de l'épissage par TDP-43 via le contrôle du niveau de 2'-O-méthylations sites-spécifiques au niveau des transcrits U1 et U2. Cette étude suggère également que WRAP53 et TDP-43 sont toutes deux importantes pour la localisation des scaARN à boîte C/D scaRNA2 et scaRNA28 dans les corps de Cajal. Nous nous sommes aussi intéressés à l'effet du stress oxydant sur l'expression et la localisation de cette protéine. L'analyse de l'expression de TDP-43 montre une augmentation 2h après l'induction du stress, puis une diminution (Figure R37C). Cette cinétique n'est pas en accord avec la redistribution nucléoplasmique de la coiline qui est maximale au temps 2 heures et se poursuit aux temps 4h et 6h. Cependant il sera nécessaire d'analyser l'expression de TDP-43 à des temps plus précoces afin de pouvoir conclure quant à une origine TDP-43 dépendante de la redistribution précoce de la coiline. Nous ne pouvons pas non plus exclure qu'il existe un lien entre la redistribution de la coiline et le niveau d'expression de TDP-43 sur les temps plus tardifs. Ces analyses doivent être complétées par le suivi de la localisation de la protéine. Plusieurs optimisations ont été réalisées récemment pour visualiser TDP-43 en microscopie de fluorescence, mais aucun signal spécifique n'a pu être obtenu jusqu'ici indiquant que l'anticorps dont nous disposons n'est pas adapté à cette technique.

Nous avons également complété cette analyse en suivant le niveau d'expression et la localisation de la protéine VRK1 qui protège la coiline de la dégradation et dont la déplétion impact la formation des CBs (Cantarero et al., 2015). Nous n'avons pas constaté de variation significative de son niveau d'expression en réponse au stress (Figure R37C). De plus, sa localisation ne semble pas non plus affectée, même si nous prenons ces résultats avec beaucoup de précaution compte tenu de la qualité médiocre du signal obtenu par épifluorescence (Figure R37B). Nous n'avons pas pu détecter de fractions sub-nucléaires éventuelles de VRK1 en raison du fort bruit de fond apporté par l'ensemble du signal nucléoplasmique. L'analyse des lames en microscopie confocale devrait permettre de gagner non seulement en résolution spatiale mais aussi en qualité du signal.



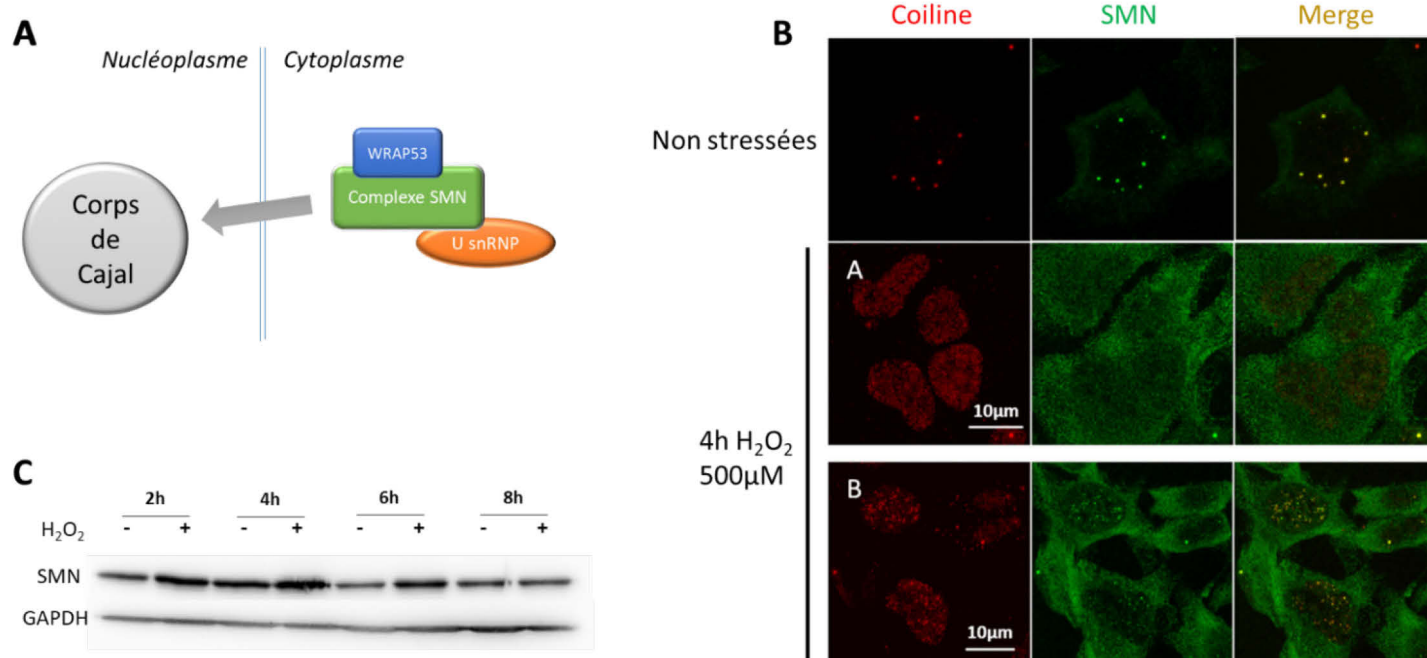
**Figure R37. Effet du stress oxydant sur l'expression de TDP-43 et VRK1.**

**(A)** Schéma d'interaction de la protéine TDP-43 avec un scaRNP C/D. **(B)** Détection par immunofluorescence de la protéine VRK1 dans des cellules HeLa S3 traitées ou non pendant 2h avec H<sub>2</sub>O<sub>2</sub> (500μM). **(C)** Les cellules HeLa S3 ont été traitées ou non par H<sub>2</sub>O<sub>2</sub> (500μM) au cours d'une cinétique de temps. Les quantités de TDP-43 et de VRK1 présentes dans ces extraits ont été analysées en WB. La quantité de chaque protéine est rapportée à la quantité de GAPDH qui sert de contrôle de l'homogénéité des dépôts. L'histogramme présente le ratio de chacune des protéines dans la condition de stress par rapport à la condition non stressée. Les quantifications ont été réalisées sur 3 expérimentations différentes.

d) Effet du stress oxydant sur la localisation et l'expression de la protéine SMN

Le complexe SMN est un composant essentiel des corps de Cajal puisqu'il est impliqué dans le maintien de leur structure et qu'il joue un rôle déterminant dans le recrutement des U snRNPs en cours de maturation en partenariat avec la protéine WRAP53 (Figure R38A). Nous avons donc analysé l'effet du stress oxydant sur le complexe SMN. Les données obtenues en microscopie confocale montrent que la protéine SMN se redistribue également dans les micro-foci coiline en réponse au stress oxydant (Figure R38B). De plus, elle s'accumule davantage dans le cytoplasme et semble presque exclusivement cytoplasmique dans les cellules où la coiline est redistribuée de façon homogène dans le nucléoplasme. L'interaction de la coiline avec SMN devra être analysée par la technique d'immunoprécipitation (IP). Alors que des études montrent qu'un déficit en SMN conduit à un stress oxydant (Dominguez et al., 2017), nous constatons ici une légère augmentation de l'expression de SMN en réponse au stress (Figure R38C).

Pour résumer, nos données montrent donc que les micro-foci coiline sont toujours composés de WRAP53 et de SMN, soulevant la question de leur fonctionnalité. Est-ce que cette multiplication des foci améliore les fonctions des CB ou à l'inverse la réduction de leur taille réduit-elle leur processivité ? Nous avons également observé que le stress oxydant a un effet négatif sur le ré-import de SMN dans le noyau. Notre hypothèse est que la fraction de SMN initialement présente dans les CBs se redistribue dans les micro-foci coiline après l'induction du stress, alors que les complexes SMN cytoplasmiques ne sont plus importés dans le noyau. La fraction cytoplasmique de WRAP53 est difficilement observable, il n'est donc pas possible de dire si la rétention cytoplasmique de SMN concerne aussi WRAP53. Des fractionnements noyau/cytoplasme devront être réalisés afin de confirmer l'accumulation cytoplasmique de SMN en réponse au stress oxydant.



**Figure R38. Effet du SO sur l'expression et la localisation de la protéine SMN.**

**(A)** Schéma du rôle du complexe SMN dans l'import nucléaire des U snRNP et de WRAP53. **(B)** Les cellules HeLa S3 ont été traitées ou non par  $H_2O_2$  (500  $\mu M$ ) au cours d'une cinétique de temps. La quantité de protéine SMN présente dans ces extraits a été analysée en WB. La GAPDH sert de contrôle de l'homogénéité des dépôts. **(C)** Détection par immunofluorescence des protéines SMN (vert) et coiline (rouge) dans des cellules HeLa S3 traitées ou non pendant 4h avec  $H_2O_2$  (500  $\mu M$ ).

e) Effet du SO sur l'expression et la localisation des composants des RNPs à boîte C/D et à boîte H/ACA

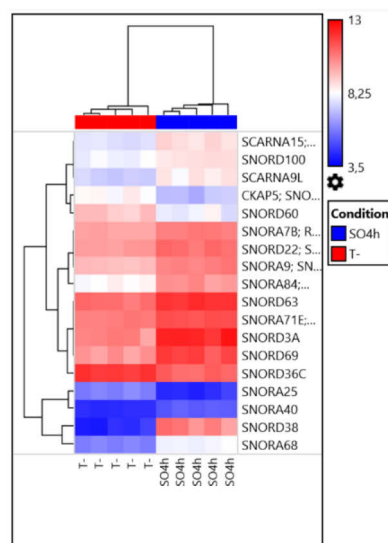
Compte tenu du rôle de TDP-43, WRAP53 et SMN dans l'adressage des scaRNPs et des UsnRNP au sein des corps de Cajal, nous nous sommes interrogés sur l'impact potentiel des effets observés sur la fonctionnalité des corps de Cajal. Nous avons alors entrepris de caractériser les effets du stress oxydant directement sur l'expression et la localisation des composants des scaRNPs. Les snoRNA étant associés aux même protéines que les scaRNA et transitant également par les corps de Cajal, nous avons décidé de les inclure dans nos analyses. Comme mentionné plus tôt dans ce manuscrit, l'analyse des effets du SO sur le transcriptome des cellules HeLa avait révélé un effet marqué du SO sur l'expression d'un panel de snoRNA/scaRNA. Ces variations sont

illustrées dans la figure R39. On note que les sno-/scaRNA dont l'expression est modifiée en réponse au SO sont généralement sur-exprimés en conditions de stress (Figure R39A) et que les variations détectées sont suffisamment fortes pour permettre de séparer les échantillons contrôles des échantillons stressés lors d'une classification non supervisée (clustering hiérarchique) (Figure R39B). Parmi les ARN différentiellement exprimés, on observe une répartition égalitaire entre les RNA à boîtes H/ACA et les RNA à boîtes C/D. La majorité de ces guides sont impliqués dans la modification des ARNr 28S et 18S, quelques-uns sont orphelins et enfin 3 ciblent des UsnRNA : scaRNA15  $\rightarrow$   $\Psi$ 37 dans le snRNA U2, scaRNA9  $\rightarrow$  2'-O-méthylations en position 19 et 30 du snRNA U2 et snoRD67  $\rightarrow$  2'-O-méthylation en position 60 du snRNA U6. Ces résultats laissent présager un possible effet du SO sur les modifications post-transcriptionnelles des UsnRNA et des ARNr

A

| ID<br>count: 18  | SO4h<br>Avg<br>(log2) | T- Avg<br>(log2) | Fold<br>Chan... | P-val    | FDR P-<br>val | Gene Symbol                      |
|------------------|-----------------------|------------------|-----------------|----------|---------------|----------------------------------|
| TC1900009290.... | 10,59                 | 4,28             | 78,98           | 1,37E-08 | 9,27E-05      | SNORD38                          |
| TC0X00006807.... | 8,03                  | 5,83             | 4,58            | 1,44E-07 | 0,0004        | SNORA68                          |
| TC0900010791.... | 10,28                 | 8,36             | 3,79            | 6,90E-06 | 0,0037        | SNORA84; MIR3651                 |
| TC1500008098.... | 8,91                  | 7,7              | 2,31            | 1,22E-05 | 0,0049        | SCARNA15; SNHG21                 |
| TC0700010942.... | 10,62                 | 9,5              | 2,18            | 1,60E-05 | 0,0058        | SNORA9; SNHG15                   |
| TC0800011816.... | 4,26                  | 5,85             | -3,01           | 3,82E-05 | 0,0096        | SNORA25                          |
| TC0500012184.... | 12,07                 | 10,98            | 2,13            | 6,87E-05 | 0,0134        | SNORD63                          |
| TC0500013329.... | 5,24                  | 4,36             | 1,84            | 7,85E-05 | 0,0146        | SNORA40                          |
| TC1700007158.... | 12,35                 | 10,64            | 3,28            | 9,89E-05 | 0,0169        | SNORD3A                          |
| TC2000007318.... | 11,54                 | 10,67            | 1,83            | 0,0002   | 0,0286        | SNORA71E; SNORA60; SNHG11        |
| TC0X00009217.... | 8,77                  | 7,3              | 2,76            | 0,0003   | 0,0299        | SCARNA9L                         |
| TC0300007493.... | 11,83                 | 10,22            | 3,06            | 0,0003   | 0,0302        | SNORD69                          |
| TC0300012391.... | 10,66                 | 9,89             | 1,71            | 0,0003   | 0,0338        | SNORA7B; RPL32P3                 |
| TSUnmapped0....  | 11,08                 | 11,99            | -1,87           | 0,0004   | 0,0355        | SNORD36C                         |
| TC1100010713.... | 7,07                  | 8,36             | -2,44           | 0,0005   | 0,0429        | CKAP5; SNORD67                   |
| TC1600009095.... | 7,86                  | 9,55             | -3,22           | 0,0006   | 0,0469        | SNORD60                          |
| TC0600009490.... | 8,86                  | 7,98             | 1,84            | 0,0006   | 0,0479        | SNORD100                         |
| TC1100011109.... | 10,97                 | 10,12            | 1,81            | 0,0006   | 0,0486        | SNORD22; SNORD29; SNORD31; SN... |

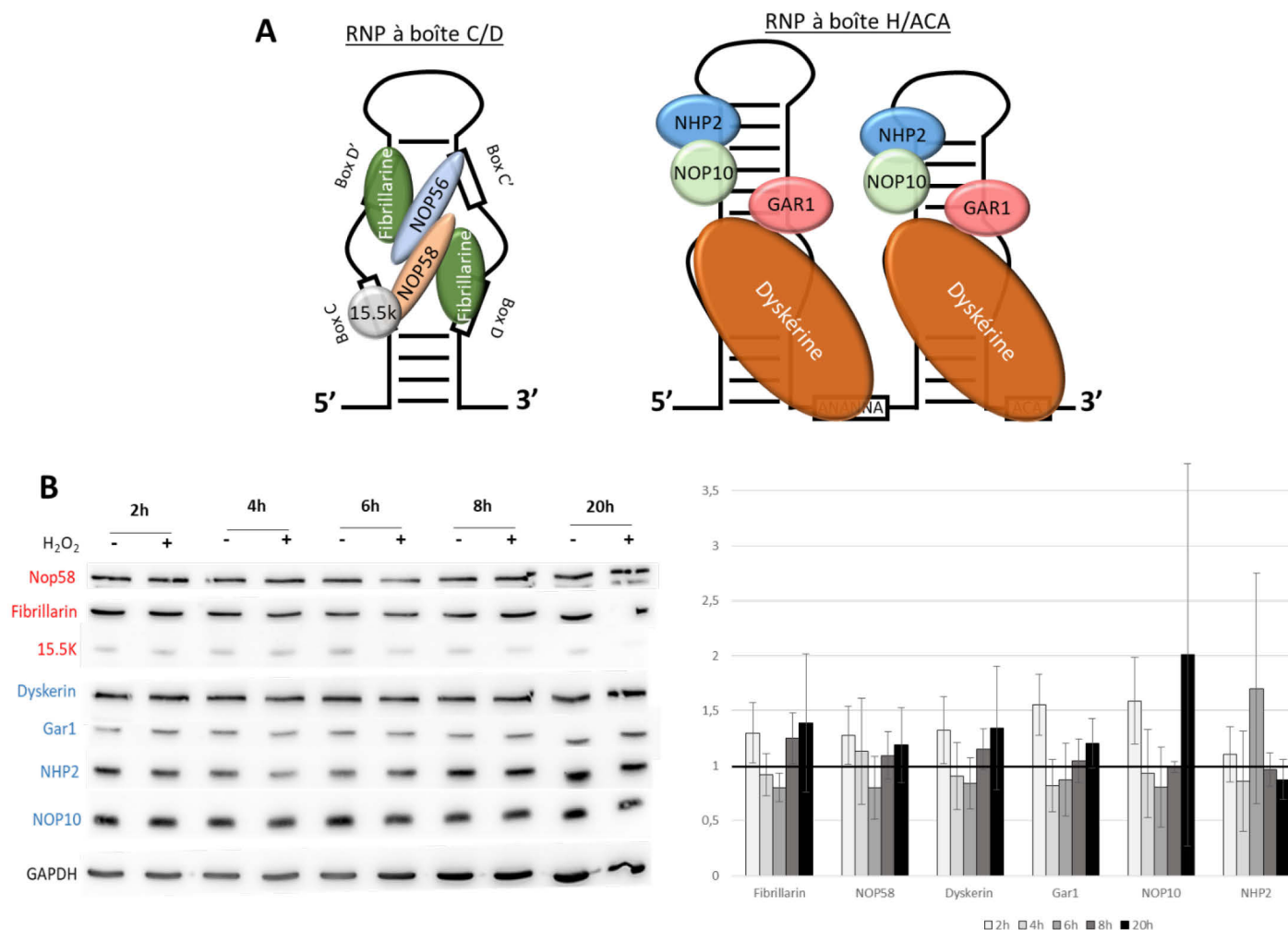
B



**Figure R39. L'expression de certains sno-/scaRNAs est significativement affectée par le SO.**

**A.** Tableau récapitulatif des variations d'expression des sno/scaRNA observée dans nos données transcriptomiques. L'intensité moyenne (en log2) de chaque gène dans la condition stressée (SO4h) ou contrôle (T-) est indiquée dans les 2ème et 3ème colonnes du tableau. Le fold change calculé, ainsi que la p-value et la p-value corrigée (FDR) figurent également dans ce tableau. Sont considérées comme significatives les variations présentant une  $FDR \leq 0,05$ . **B.** Clustering non supervisé des échantillons analysés sur puces Clariom D sur la base des intensités relatives des sno-/scaRNA dont l'expression varie en réponse au SO.

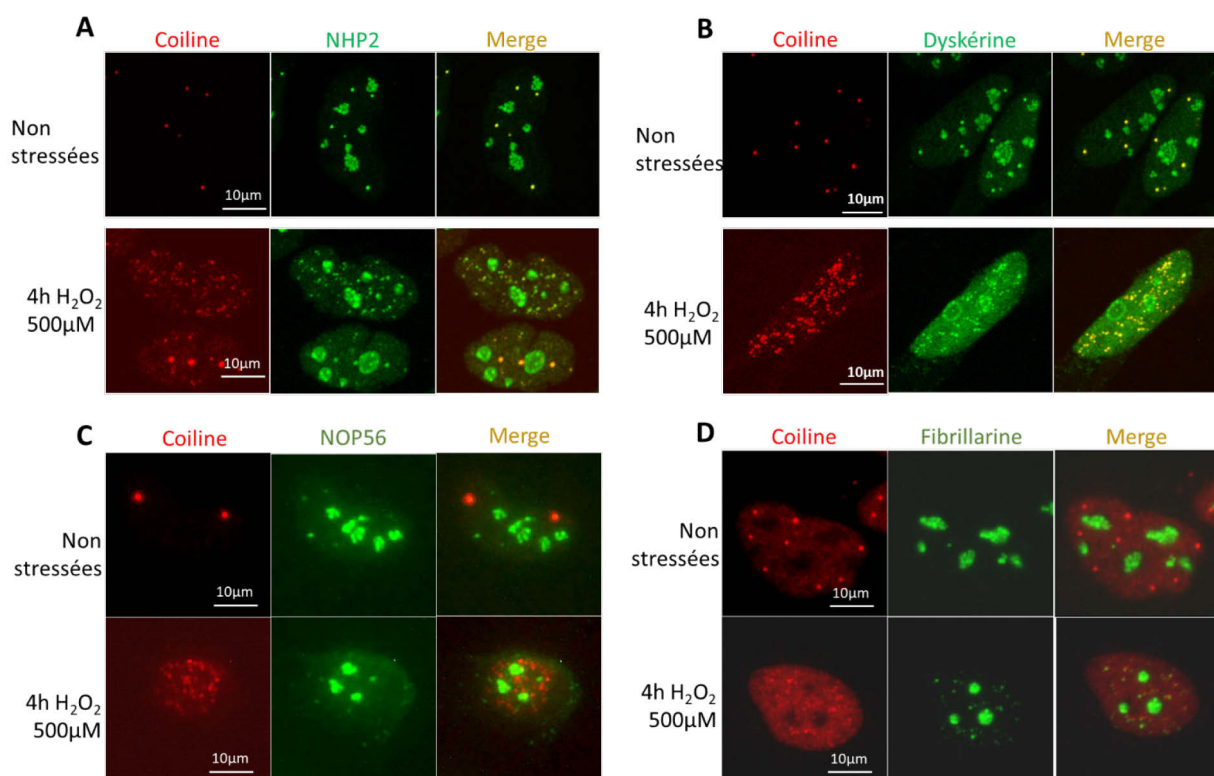
Dans un second temps, l'expression des protéines cœurs des RNP à boîtes C/D (Fibrillarine, Nop56, Nop58 et 15,5K) et H/ACA (Dyskérine, Nhp2, Nop10 et Gar1) a été suivie par western blot au cours d'une cinétique de stress oxydant (Figure R40). Cette analyse n'a révélé aucun effet significatif induit par le stress (Figure R40B), même si les données doivent être complétées par des répliques biologiques pour la protéine 15.5K. Notons aussi que l'analyse de l'expression de Nop56 dont l'anticorps n'a pas fonctionné en Western Blot doit être refaite.



**Figure R40. Analyse de l'expression des protéines composant les RNP à boîtes H/ACA et C/D.**

**(A)** Schéma des RNP à boîte C/D et à boîte H/ACA. **(B)** Analyse de l'expression de protéines des RNP à boîtes H/ACA et C/D dans des cellules traitées ou non par 500µM H<sub>2</sub>O<sub>2</sub> au cours d'une cinétique de temps. La quantité de chaque protéine est rapportée à la quantité de GAPDH qui sert de contrôle de l'homogénéité des dépôts. L'histogramme présente le ratio de chacune des protéines dans la condition de stress par rapport à la condition non stressée. Les quantifications ont été réalisées sur 3 expérimentations différentes.

Nous avons également étudié la localisation de ces différentes protéines cœurs en réponse à un stress oxydant. Les scaARNs à boîte H/ACA sont environ 20 fois plus exprimés que les scaARNs à boîtes C/D, c'est pourquoi il est plus facile de les observer en immunofluorescence. Comme observé pour SMN et WRAP53, le stress oxydant induit dans les temps précoces la redistribution de la dyskérine et de NHP2 dans les micro-foci coiline (Figure R41).



**Figure R41. Visualisation par immunofluorescence de protéines RNP des boîtes H/ACA ou C/D suite à un stress oxydant.** Les cellules sont traitées ou non pendant 4h avec H<sub>2</sub>O<sub>2</sub> (500µM), puis elles sont fixées et co-marquées avec différents anticorps contre les protéines : NHP2 (A) ou Dyskérine (B) ou NOP56 (C) ou Fibrillarine (D) et la coiline (A, B, C, D)

De plus, on observe que le stress oxydant modifie leur localisation dans les nucléoles, eux-mêmes sous-structurés en trois parties selon les fonctions mises en jeu (fibrillar center, dense fibrillar component, granular component). Ainsi, NHP2 présente une distribution nucléolaire plus diffuse suite au stress oxydant suggérant une possible restructuration des compartiments internes comme cela a pu aussi être observé avec d'autres stress (Boulon et al., 2010; Govoni et

al., 1994). La Dyskérine, quant à elle, se concentre plutôt sur la périphérie des nucléoles et présente un signal nucléoplasmique plus important suggérant une diminution de la fraction nucléolaire. Pour la Fibrillarine, alors que dans les cellules non stressées, il est difficile de mettre en évidence une colocalisation avec la coiline, elle semble se relocalisée dans les microfoci coiline suite à un SO. En revanche, nos images obtenues en épifluorescence concernant la localisation nucléolaire de la fibrillarine ne nous permettent pas de conclure et devront être refaites en imagerie confocale. A l'inverse, le marquage de la protéine NOP56 ne montre pas une colocalisation de celle-ci dans les microfoci de coiline après un SO. Ces résultats d'immunofluorescence montrent que les micro-foci coiline concentrent aussi certains des composants protéiques des scaRNP à boîte H/ACA, comme NHP2, dyskérine. En revanche, il est plus difficile de conclure sur les composants des scaRNP à boîte C/D, car on observe d'une part une concentration de la fibrillarine dans les microfoci, et d'autre part une absence de NOP56 au sein des microfoci.

Cette analyse sera complétée par l'analyse des effets du stress oxydant sur la localisation des scaRNAs par RNA-FISH. Des expérimentations pour les scaRNAs 2 et 13 ont d'ores et déjà été réalisées, mais ne nous ont pas permis d'obtenir un signal discernable du bruit de fond. Les sondes LNA utilisées alors étaient couplées à une seule molécule de biotine, ce qui ne permettait pas d'amplifier suffisamment le signal pour compenser le faible niveau d'expression de ces transcrits. Nous avons récemment fait l'acquisition du kit « AmpliScribe™ T7-Flash™ Biotin-RNA Transcription Kit » permettant de produire des sondes ayant l'avantage d'être biotinylées au niveau de chaque résidu uridine, et ainsi de générer des sondes pouvant contenir plus de 20 molécules de biotine. La transcription de ces sondes sera réalisée à partir de plasmides mis à disposition par l'équipe de Tamas Kiss (LBME UMR 5099), codant les scaRNAs à boîte C/D (U90, U91 et U2-47), à boîte H/ACA (ACA26, et U92) ou possédant les deux types de motifs (U85, U87 et U88). En parallèle, nous optimisons les conditions d'IP et de RNA-IP (RIP) afin d'approfondir l'analyse de l'effet du stress oxydant sur la composition des complexes coiline en scaRNP.

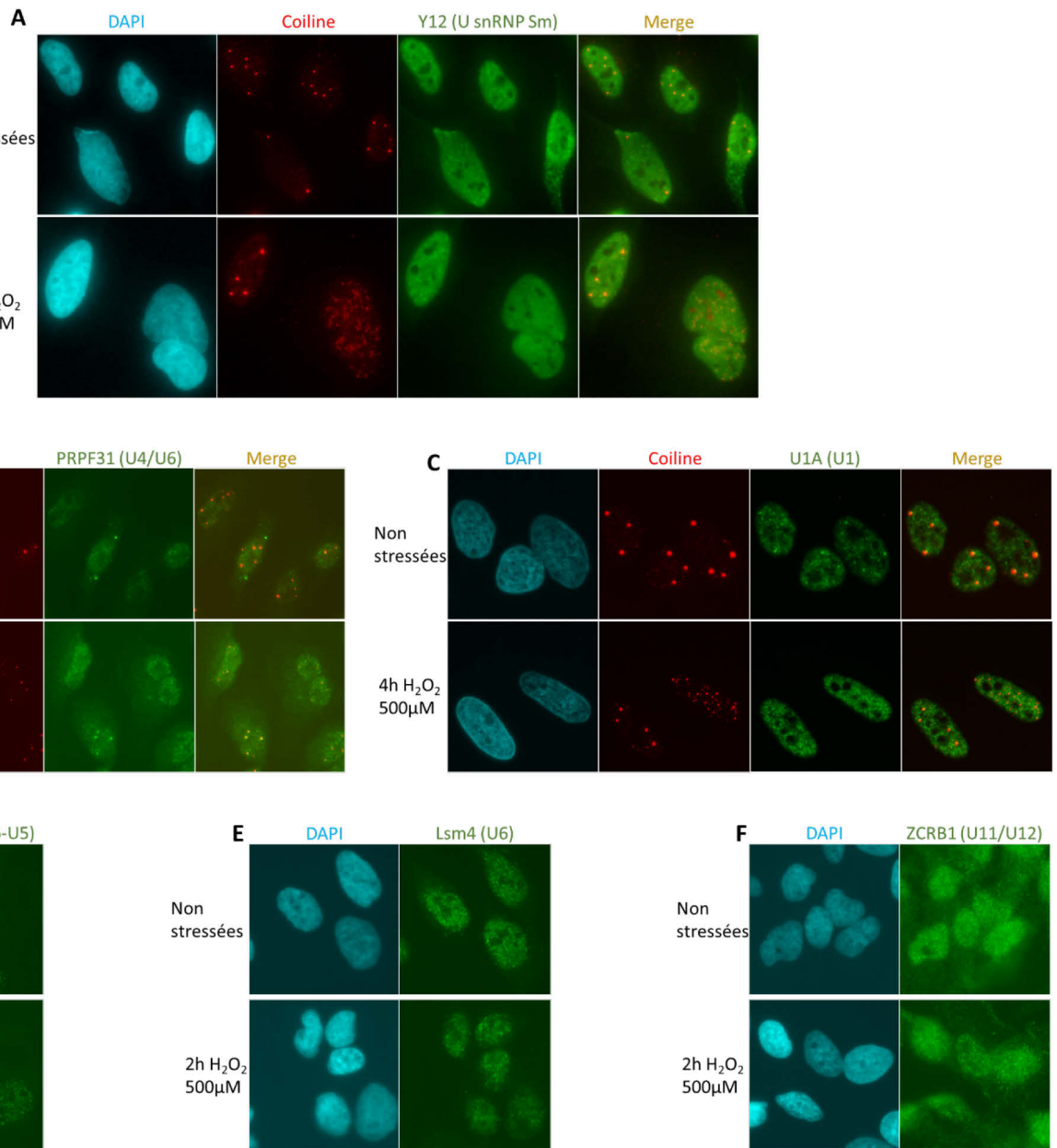
#### f) Effet du SO sur la localisation de composants des UsnRNP

Comme nous l'avons vu le stress oxydant modifie la localisation de nombreuses RNP, nous sommes donc interrogés sur l'aspect fonctionnel de ces modifications et des effets éventuels



sur le spliceosome. Nous nous sommes logiquement intéressés à la localisation de protéines constituant les UsnRNP. Pour cela, nous avons utilisé l'anticorps Y12 dirigé contre les protéines SmB/B' que nous avons utilisé précédemment pour les expériences d'immunoprécipitation. Ainsi, nous espérons visualiser la distribution subnucléaire des UsnRNP matures et notamment la fraction présente dans les corps de Cajal. Or le signal nucléoplasmique de SmB/B' est complètement homogène à l'exception de son exclusion des nucléoles (Figure R42). Un léger signal cytoplasmique est également observé, en accord avec le fait que leur assemblage avec les UsnRNP a lieu dans le cytoplasme. Le stress oxydant ne semble pas perturber la distribution générale de SmB et B', cependant un effet mineur pourrait être masqué par l'intensité du signal nucléoplasmique total.

Nos données de RNA-FISH montraient que la snRNP U1 n'est pas préférentiellement enrichi dans les corps de Cajal, en revanche, la protéine U1A qui lui est associée, présente un signal caractéristique de ces substructures en condition non stressée (Figure R42C). La fraction de U1A présente dans les corps de Cajal disparaît dans la majorité des cellules stressées au profit d'une localisation nucléoplasmique plus homogène, mais toujours exclue des nucléoles. Ces observations sont en faveur d'une redistribution nucléoplasmique de la snRNP U1 induite par le stress oxydant, comme ce que nous avons observé pour le snRNA U2. La protéine PRPF31 qui s'associe à la di-snRNP U4/U6 présente aussi un enrichissement dans des corps de Cajal, mais seulement dans environ 30% des cellules (Figure R42B). De façon générale, le signal nucléoplasmique est complètement homogène et une fraction cytoplasmique est aussi visible. La proportion de cellules présentant un enrichissement de PRPF31 dans les corps de Cajal ne semble pas être modifiée par le stress oxydant. Nous avons également débuté l'analyse de la localisation des protéines Sart1 (tri-snRNP U4/U6-U5), Lsm4 (U6) et ZCRB1 (U11/U12) (Figures R42D, E et F), en absence de co-marquage avec la coiline. Cependant, le signal obtenu pour ces trois protéines ne présentait pas de différence entre les conditions témoins et stressées.



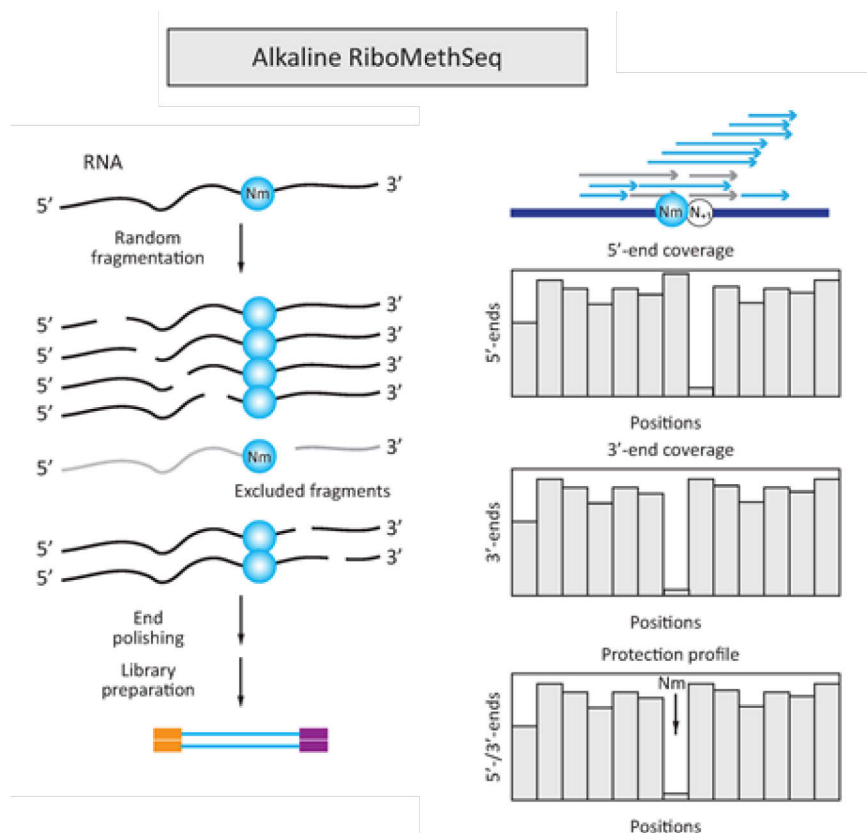
**Figure R42. Suivi de la localisation de composants protéiques des U snRNPs spécifiques au cours d'un stress oxydant.** Ces données ont été obtenues en épifluorescence.

## V. Effets du SO sur la fonctionnalité des corps de Cajal

Compte tenu des variations d'expressions et de localisations observées en conditions de stress sur des composants clés de la maturation du spliceosome, nous nous sommes aussi interrogés sur les effets potentiels du SO sur la fonctionnalité des corps de Cajal. Connaissant l'implication de ces substructures dans la régulation de la cinétique de formation des 2'-O-méthylations et des pseudouridylations au sein des UsnRNAs, nous avons entrepris d'analyser le taux de ces modifications en réponse au stress. Dans un premier temps nous nous sommes intéressés aux 2'-O-méthylations et avons tiré profit de la technique de RiboMethSeq développée par Virginie Marchand et Yuri Motorin au sein de la plateforme de NGS du biopôle (UMS2008 IBSLor) (Marchand et al., 2016). Cette méthode permet de détecter des résidus 2'-O-méthylés par séquençage à haut débit sur la base de la résistance à l'hydrolyse de ces résidus (Figure R43).

Ce protocole nécessitant de travailler à partir d'ARN enrichis, nous avons utilisé les ARN immunoprécipités par l'anticorps Y12 dirigé contre les protéines SmB/B' (RIP). Cette expérimentation a déjà été présentée précédemment dans les résultats (Figure R18).

Les expériences de RIP Y12 pour les différentes conditions de stress utilisées (choc thermique et stress oxydant) ont été réalisées en triplicatas biologiques et les ARN extraits à partir des fractions immunoprécipitées ont été donnés à la plateforme de séquençage pour analyse de leur contenu en 2'-O-méthylations (n=9). Afin de valider l'efficacité de notre approche, nous avons, tout d'abord, vérifié notre capacité à détecter les résidus 2'-O-méthylés déjà connus au sein des ARN non stressés. Il est à noter que le protocole Ribomethseq utilisé ne permet pas de détecter les résidus 2'-O-méthylés présents au niveau des positions 1 ou 2 des ARN (Marchand et al., 2016). Sur les 27 méthylations connues, 25 ont pu être détectées avec notre protocole, soit une efficacité de 92,5% (Figure R44). De plus, 33 nouvelles positions ont été détectées, dont 22 au niveau des UsnRNAs du spliceosome mineur pour lesquels aucune analyse globale de leur contenu en modifications n'avait été réalisée jusqu'à présent (Figure R44).



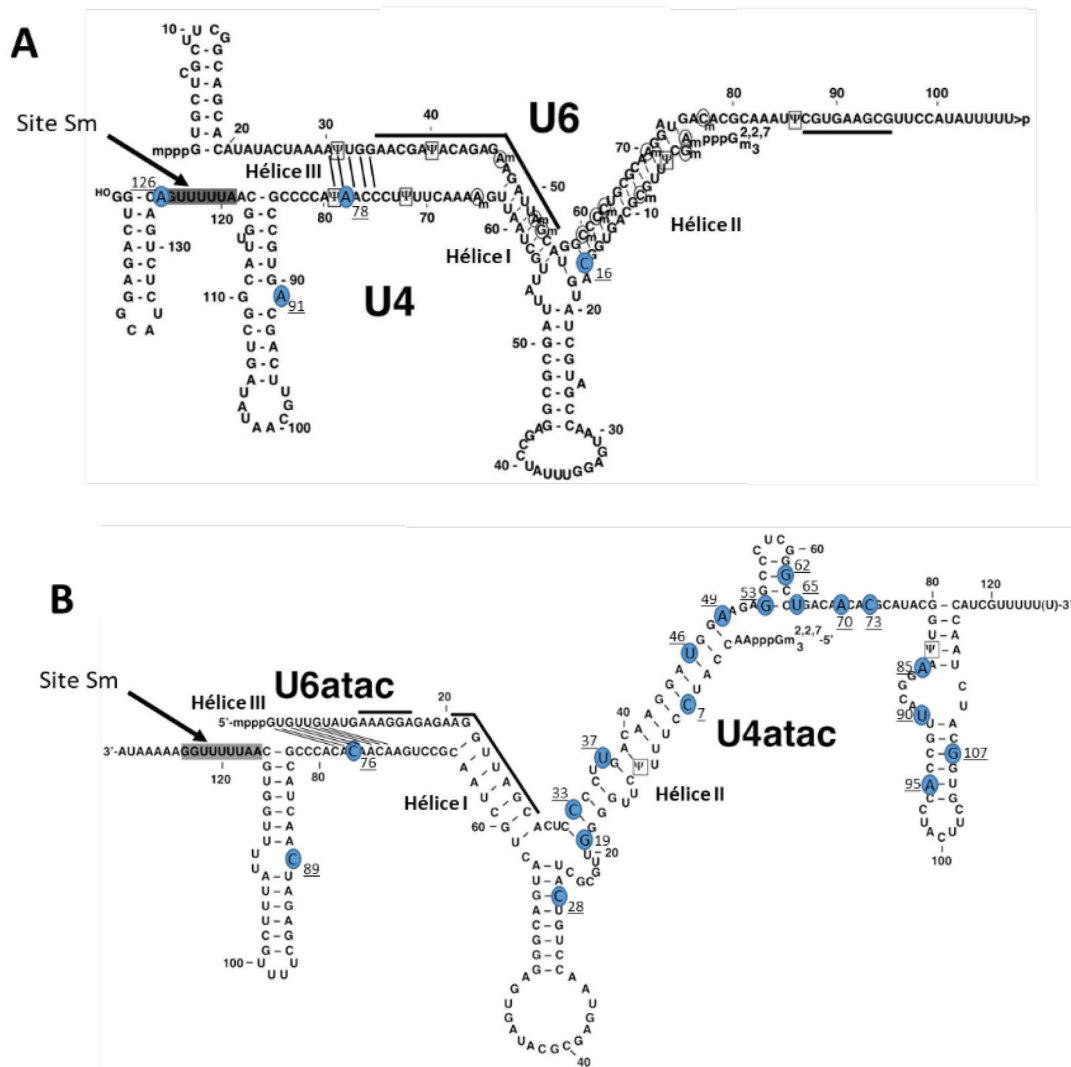
**Figure R43. Principe la méthode Ribomethseq**

La fragmentation alcaline des ARN utilisée dans RiboMethSeq exclue les fragments d'ARN se terminant par un nucléotide 2'-O-méthylé et par conséquent aussi ceux commençant par le nucléotide n+1 d'un résidu 2'-O-méthylé. Ces fragments sont sous-représentés au sein de la librairie de séquençage et lorsque les reads sont alignés sur les séquences références, cela se traduit par une chute caractéristique de la couverture au niveau de la position 2'-O-méthylée (3'end) et de la position N+1 (5'end).

| U snARN | 2'-O-méthylations connues (excluant extrémité 5') | 2'-O-méthylations détectées | 2'-O-méthylations candidats |
|---------|---------------------------------------------------|-----------------------------|-----------------------------|
| U1      | 1(2)                                              | 1                           | 2                           |
| U2      | 8(2)                                              | 7                           | 2                           |
| U4      | 2(2)                                              | 2                           | 4                           |
| U5      | 3(2)                                              | 3                           | 3                           |
| U6      | 9                                                 | 9                           | 0                           |
| U11     | 0                                                 | 0                           | 0                           |
| U12     | 3                                                 | 3                           | 5                           |
| U4atac  | 1                                                 | 0                           | 4                           |
| U6atac  | 0                                                 | 0                           | 13                          |

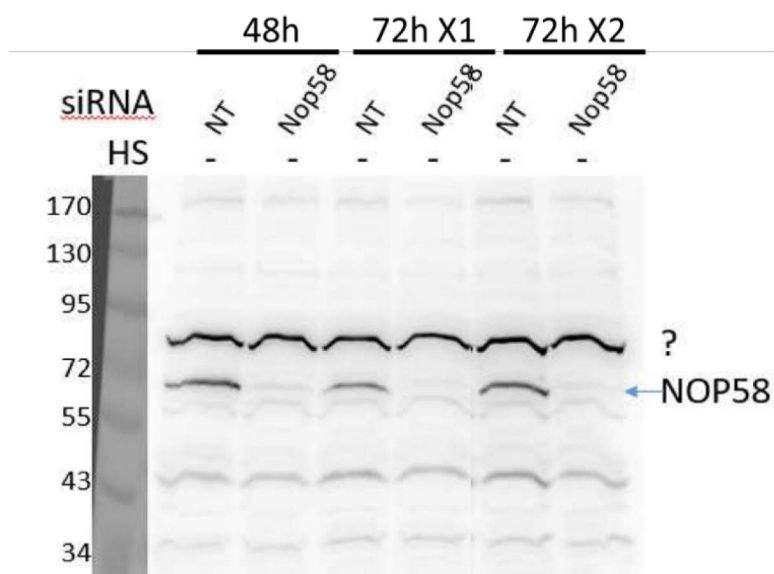
**Figure R44. Analyse par la méthode Ribomethseq du contenu en résidus 2'-O-méthylés des UsnRNA de cellules HeLa S3.**

De façon intéressante, de nombreuses nouvelles méthylations détectées au sein des snRNA U4atac et U6atac sont conservées au sein des snRNA U4 et U6, ce qui est un argument fort en faveur de leur existence (figure R45).



**Figure R45. Cartographie des anciennes et des nouvelles méthylations détectées au sein des snARNs U4, U6, U4atac et U6atac. Les nouvelles positions candidates sont entourées en bleues.**

L'existence des nouvelles modifications détectées au sein des UsnRNA (2'-O-méthylations candidats) sera prochainement validée par des expériences de transcription inverse réalisées en présence de faibles concentrations en dNTP. Par ailleurs, afin de déterminer si la formation de ces nouvelles modifications est réalisée au sein de RNP C/D (sca- ou snoRNP), nous avons mis au point l'extinction d'une protéine cœur de ces RNP, la protéine Nop58 (Figure R46).

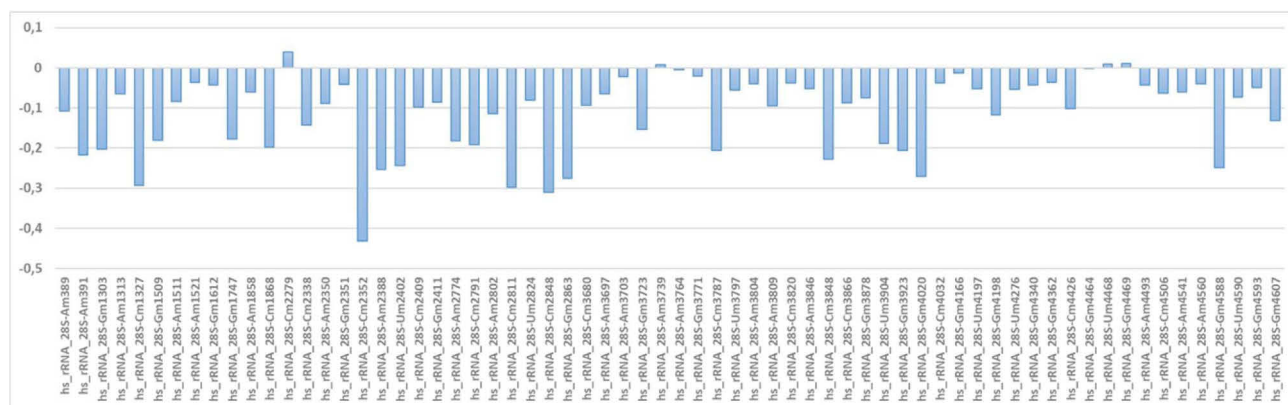


**Figure R46. Extinction de l'expression de la protéine Nop58 par emploi de siRNA.**

L'expression de la protéine Nop58 a été analysée par Western blot 48h et 72h après une transfection unique par le pool de siRNA dirigés contre Nop58, et 72h après une double transfection (transfection à t0 et t48) par ce même pool. La condition 72h post-transfection unique a été conservée pour la suite de nos expériences.

L'efficacité des siRNA Nop58p pour l'invalidation fonctionnelle des RNP C/D a été validée par l'analyse du contenu en résidus 2'-O-méthylés des ARNr 72h post-transfection des cellules HeLa avec les siRNA (Figure R47). Cette analyse a en effet révélé une diminution moyenne d'environ 15% du contenu global en 2'-O-méthylations de l'ARNr 28S, avec certaines diminutions pouvant aller jusqu'à plus de 30%. Le temps de demi-vie des ARNr humains ayant été estimé à  $72\text{h} \pm 4\text{h}$  (Liebhaber et al., 1978), cela signifie que l'extinction de Nop58 permet de diminuer très fortement la modification des ARNr néosynthétisés. Des résultats similaires ont été observés par

l'équipe en utilisant des siRNA dirigés contre la fibrillarine, avec cependant un effet beaucoup plus marqué sur la viabilité cellulaire (Sharma et al., 2017).

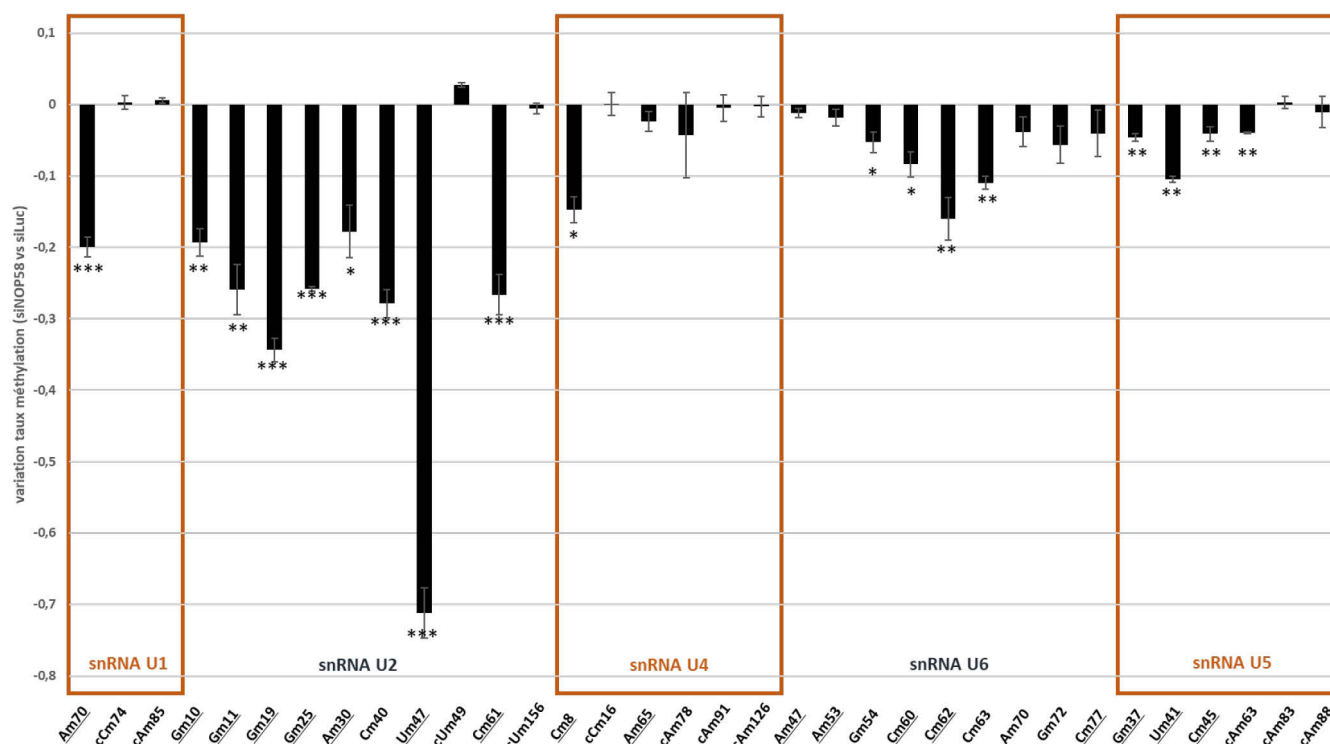


**Figure R47. Analyse de l'effet de l'extinction de Nop58 sur le contenu en résidus 2'-O-méthylés de l'ARNr 28S humain.**

La variation du taux de modification (siNop58 vs siLuc) à chacune des positions 2'-O-méthylée connue de l'ARNr 28S est indiquée.

L'analyse de l'effet de l'extinction de la protéine Nop58 sur la 2'-O- méthylation des UsnRNA majeurs et mineurs est présentée dans les Figures R48 et R49. Concernant les UsnRNA majeurs, une diminution significative du taux de méthylation est observée pour toutes les positions méthylées connues (positions non précédées d'un « c » dans la figure R48) des snRNA U1, U2 et U5. Pour le snRNA U4 sur les 2 méthylations connues seule la position Cm8 est significativement affectée par le siNOP58, la position Am65 présente cependant une tendance à la diminution mais non significative. De même, au niveau du snRNA U6, sur les 9 positions méthylées seules 4 sont significativement affectées par le siNOP58, les 5 autres présentant une faible (Am70, Gm72 et Cm77) voir très faible (Am47 et Am58) tendance à la diminution. Parmi les 11 positions candidates détectées, seule la position cAm63 au sein du snRNA U5 est significativement affectée par le siNOP58, suggérant que cette position a de fortes chances d'exister réellement est d'être modifiée par une RNP à boîtes C/D. De façon intéressante on remarque que certaines positions significativement affectées par le siNOP58, telles que les positions Cm40 du snRNA U2 et Gm54 et Cm63 du snRNA U6, n'ont pas été associées à un ARN C/D guide, alors que d'autres positions non significativement affectées par le siNOP58, telles que les positions Am47, Am53 et Cm77 du

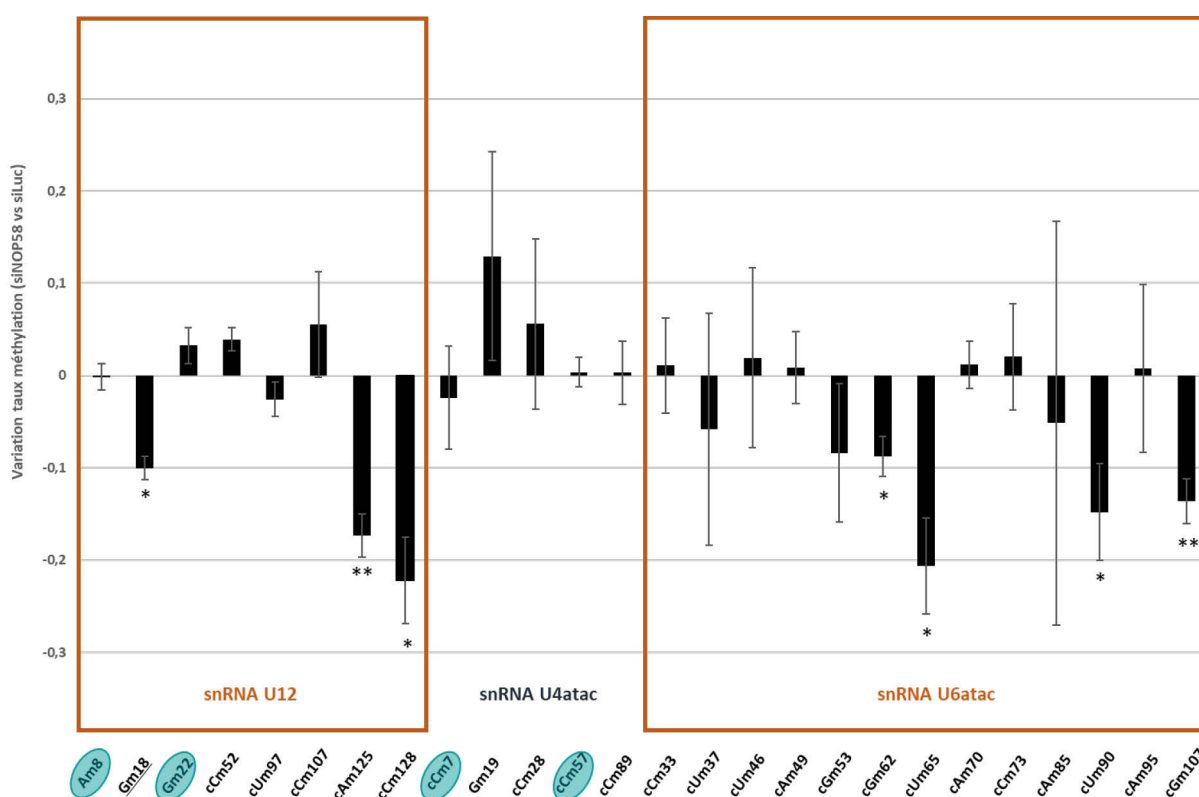
snRNA U6 ont été associées à un ARN guide. Il est cependant important de garder à l'esprit que la plupart des associations « position méthylée/ARN guide » sont basées sur des prédictions informatiques et n'ont pas été vérifiées expérimentalement. Une recherche d'ARN guide a été effectuée (snoscan server, <http://lowelab.ucsc.edu/snoscan/>) pour toutes les positions orphelines incluant la position cAm63 mise en évidence au niveau du snRNA U5 en utilisant les séquences de tous les ARN à boîtes C/D déjà identifiés et ceux prédits dans l'article de Jorjani et coll. en 2016, mais cette recherche ne nous a pas permis de proposer de nouvelles associations « position méthylée/ARN guide ».



**Figure R48. Analyse de l'effet de l'extinction de Nop58 sur le contenu en résidus 2'-O-méthylés des UsnRNA majeurs.** L'effet moyen du siNop58 sur le taux de méthylation de chaque position est représenté. Le nom de la position est indiqué en bas du graphique : il est précédé d'un « c » minuscule dans le cas des positions candidates détectées au cours de notre étude ; le nom est souligné quand, dans la littérature, un ARN guide a pu être associé à la position modifiée. Les données présentées ont été produites à partir de 3 expériences indépendantes et un test de student a été réalisé pour tester la significativité de l'effet observé (\*,  $p < 0.05$ ; \*\*,  $p < 0.01$ ; \*\*\*,  $p < 0.001$ ).



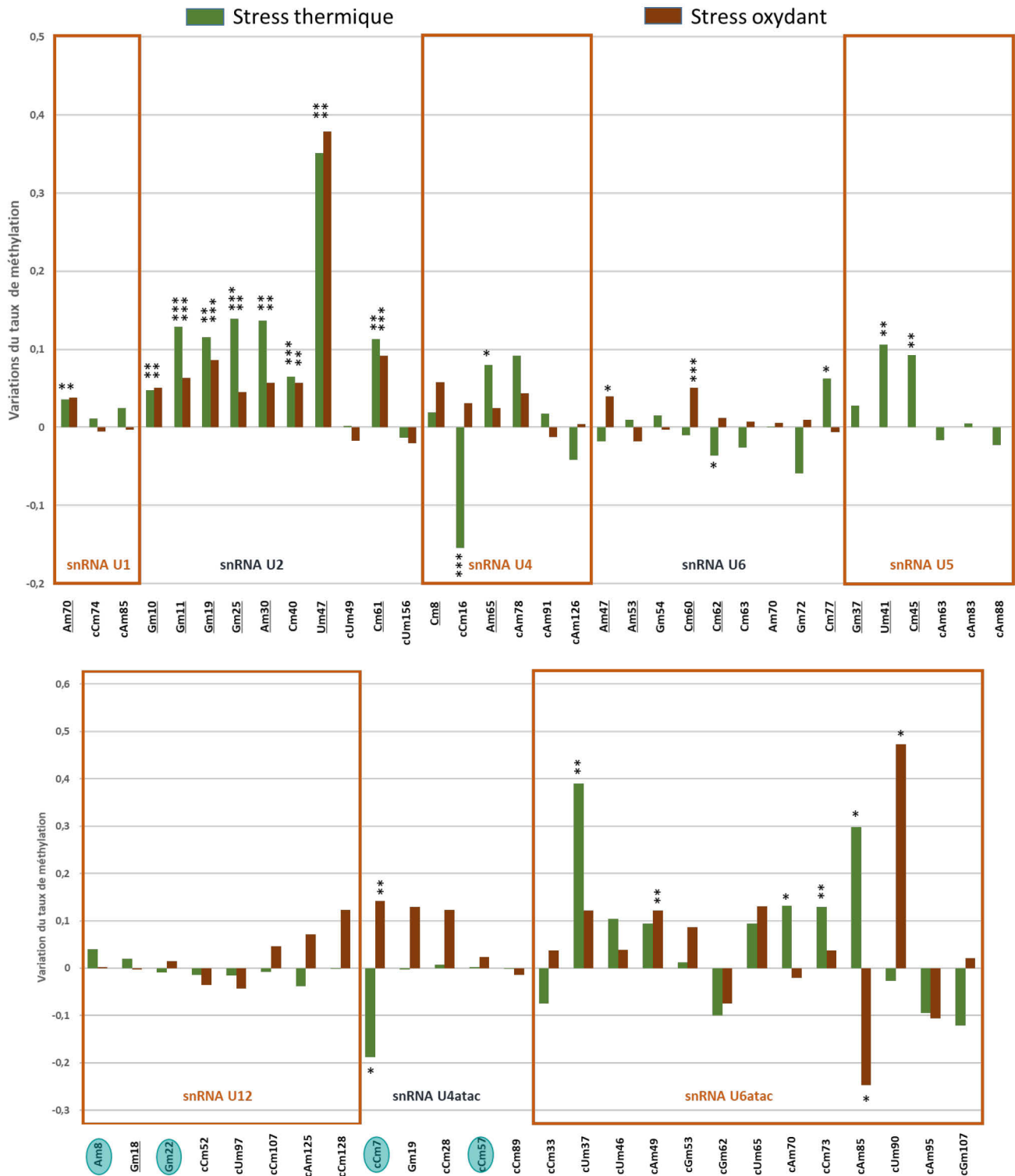
L'analyse de l'effet de l'extinction de la protéine Nop58 sur la 2'-O- méthylation des UsnRNA mineurs, présentée dans la Figure R49 a, quant à elle, révélé que la méthylation de 3 positions au sein du snRNA U12 était significativement affectée par le siNOP58, parmi lesquels 2 positions candidates identifiées lors de nos analyses, cAm125 et cCm128. Les méthylations précédemment publiées en position 8 et 22, tout comme les 3 autres méthylations candidates que nous avons identifiées ne sont pas significativement affectées par le siNOP58. Concernant le snRNA U4atac aucune des positions candidates n'est significativement affectée par le siNOP58 et pour le snRNA U6atac seules 4 positions candidates sur les 13 identifiées sont significativement affectée par le siNOP58.



**Figure R49. Analyse de l'effet de l'extinction de Nop58 sur le contenu en résidus 2'-O-méthylés des UsnRNA mineurs.** L'effet moyen du siNop58 sur le taux de méthylation de chaque position est représenté. Le nom de la position est indiqué en bas du graphique : il est précédé d'un « c » minuscule dans le cas des positions candidates détectées au cours de notre étude ; le nom est souligné quand, dans la littérature, un ARN guide a pu être associé à la position modifiée. Le nom est placé au sein d'une ellipse verte lorsque nous avons pu prédire une association avec un ARN guide C/D Les données présentées ont été produites à partir de 3 expériences indépendantes et un test de student a été réalisé pour tester la significativité de l'effet observé (\*,  $p < 0.05$ ; \*\*,  $p < 0.01$ ).

Etant donné que seule la position Gm18 du snRNA U12 avait pu être associée à un ARN C/D guide (scaRNA21, Jorjani et al., 2016), nous avons recherché si certains des ARN C/D connus pouvaient guider la méthylation aux positions déjà connues ou nouvellement identifiées comme étant potentiellement 2'-O-méthylées dans les UsnRNA mineurs. Nous avons ainsi pu identifier 4 nouvelles associations « position méthylée/ARN guide ». Ces prédictions seront validées expérimentalement en diminuant l'expression des ARN guides par emploi de LNA (Locked Nucleic Acids) gapmers et en analysant l'effet sur les méthylations cibles prédites. Le niveau d'expression des scaRNAs validés comme ciblant les modifications d'intérêts sera analysé par RT-qPCR en réponse aux différents stress.

Ayant analysé le contenu en résidus 2'-O-méthylés des UsnRNA de cellules HeLa en absence de stress, nous avons ensuite entrepris de détecter d'éventuelles variations de ce contenu en réponse au choc thermique (n=3) et au stress oxydant (n=4). Les données obtenues sont résumées dans la Figure R50. La significativité des variations a été contrôlée pour chaque position par un test de Student. Les effets les plus significatifs (test Student  $p \leq 0,05$ , \*\*) sont observés au niveau du snRNA U2 pour lequel toutes les positions méthylées connues sont hyperméthylées en réponse aux 2 stress. On notera notamment la position Um47 qui, bien que déjà répertoriée, est très peu détectée dans les cellules non stressées et qui augmente fortement dans les deux conditions de stress (35% en choc thermique et 38% en stress oxydant). Les 2 positions candidates que nous proposons au sein du snRNA U2 ne varient pas. Les variations significatives (test Student  $p \leq 0,1$ , \*) détectées au sein des snRNA U1, U4, U4atac, U6 et U6atac sont plus hétérogènes et semblent plus spécifiquement affectées par l'un des 2 stress. Ainsi, la position cCm16 du snRNA U4 est significativement hypométhylée en réponse au stress thermique, alors qu'elle est non affectée par le stress oxydant. De même les positions cUm37, cAm70 et cCm73 du snRNA U6atac sont hyperméthylées spécifiquement en réponse au stress thermique. On notera également que des profils opposés de variation selon la nature du stress sont observées pour les positions cCm7 du U4atac et cAm85 du U6atac. Bien que ces données demandent à être validées, elles suggèrent que les variations de méthylation observées résultent de programmes de régulation différents, spécifiques à chacun des stress appliqués.



**Figure R50. Récapitulatif des effets d'un stress thermique (HS) et d'un stress oxydant (OS) sur le contenu en 2'-O-méthylations des UsnRNAs majeurs (en haut) et mineurs (en bas) de cellules HeLa.**

Les nouvelles 2'-O-méthylations détectées lors de notre étude sont indiquées avec un « c » (candidat) devant le nom de la position. Les données présentées ont été produites à partir de 3 et 4 expériences indépendantes de stress thermique et de stress oxydant, respectivement. Un test de student a été réalisé pour tester la significativité de l'effet observé (\*,  $p \leq 0,1$  ; \*\*,  $p \leq 0,05$  ; \*\*\*,  $p \leq 0,01$ ).

# CONCLUSION



Le stress oxydant est un stress très étudié par la communauté scientifique du fait de son incidence dans de nombreuses pathologies. L'originalité de mon projet de thèse réside dans le fait d'appréhender ce stress à travers les dysfonctionnements liés aux régulations fines des ARN. En effet, les ARN se sont révélés être des leviers de régulations pertinents pour la cellule, particulièrement quand elle doit faire face à un stress et modifier très rapidement le panel de protéines et ARN exprimés. Pour ce faire, nous avons réalisé des analyses globales du transcriptome en utilisant la technologie des puces à ADN (puces à jonctions d'exons Clariom D de chez Affymetrix) et le séquençage au débit. Les premières nous ont renseigné sur les variations d'expression des ARN (ARNm, scaRNA...), mais aussi sur la production des isoformes d'ARN en réponse aux stress oxydant. Le séquençage au débit, quant à lui, nous a apporté un éclairage sur la cartographie des modifications des ARN en réponse à différents stress. Ainsi, nous observons pour la 1ère fois des variations du contenu en résidus 2'-O-méthylés des UsnRNA de cellules HeLa en réponse aux stress. De façon intéressante, certaines variations sont communes aux 2 stress, alors que d'autres semblent plus spécifiques d'une condition de stress donnée. Bien que l'impact des variations du contenu en modifications post-transcriptionnelles des UsnRNA sur l'activité du spliceosome reste à établir, ces données suggèrent la possibilité d'une spécialisation du spliceosome en réponse aux stress, via la modulation de l'épitranscriptome des UsnRNA. Cet aspect est d'autant plus intéressant que la découverte des fonctions liées aux modifications de l'épitranscritome est un pan de recherche en plein essor actuellement.

Par ailleurs, nous nous sommes concentrés sur les acteurs directs de l'épissage que représentent le spliceosome, mais aussi tous les régulateurs de cette machinerie qui a une maturation complexe. Nos questionnements se sont tout d'abord portés sur les localisations des différents acteurs. Ainsi, j'ai mis en évidence un effet encore jamais décrit du stress oxydant sur la structure des corps de Cajal. Les analyses en immunofluorescence ont montré qu'un stress oxydant de 2h induit une redistribution nucléoplasmique de la coiline qui joue un rôle central dans la structure et la fonctionnalité des CBs. Cette redistribution conduit soit à la formation d'un grand nombre de foci nucléaires de plus petites tailles, soit à une localisation nucléoplasmique plus homogène qui perdurent jusqu'à 6h après le stress. Les corps de Cajal contiennent des facteurs requis pour l'épissage, la transcription ou la biogenèse du ribosome. Ils peuvent diffuser

de façon dynamique dans l'espace inter-chromatinien et s'associer avec des loci de gènes spécifiques comme les gènes codant les UsnARNs ou les histones (Platani et al., 2002). Certaines données ont aussi démontré que les corps de Cajal régulent la topologie de l'organisation des chromosomes et peuvent induire la colocalisation des gènes de sn/snoARN afin de stimuler leur expression (Wang et al., 2016). Sachant que le stress oxydant induit un grand nombre de dommages à la cellule, nous pouvons imaginer que la redistribution de la coiline ne se fait pas de façon aléatoire, mais que les complexes coiline sont plutôt relocalisés sur des loci particuliers afin d'activer/stimuler l'expression de gènes impliqués dans la réponse au stress.

Il est difficile de déterminer si le stress-oxydant induit directement le désassemblage des foci coiline ou s'il empêche leur réassemblage. La présence de ces micro-foci reflète une perturbation de la nucléation et donc de la composition des complexes au sein de ces substructures. C'est pourquoi nous nous sommes intéressés aux effets du stress oxydant sur les autres composants des corps de Cajal. Ainsi, nous avons observé que cet effet s'accompagne d'une redistribution similaire de WRAP53, une protéine qui est non seulement impliquée dans la nucléation et la structuration, mais aussi dans la rétention des scaRNP à boîte H/ACA. La colocalisation de WRAP53 et la coiline dans les microfoci suggère que les 2 protéines interagissent toujours ensemble et que le recrutement des scaRNP à boîte H/ACA est toujours effectif dans les microfoci. Cette idée est renforcée par l'observation de microfoci dyskérine et NHP2, deux composants cœurs des RNPs à boîte H/ACA, colocalisant avec les microfoci coiline après un stress oxydant. La fraction des scaRNPs à boîte C/D présente dans les corps de Cajal est plus difficile à observer, mais les premières données obtenues avec la fibrillarine et NOP56 semblent également montrer une redistribution nucléoplasmique induite par le stress. Des expériences supplémentaires en microscopie confocale devraient permettre d'établir si les microfoci coiline colocalisent avec ces deux protéines. De plus, le suivi de la localisation de la protéine TDP-43 impliquée dans l'adressage des scaRNPs à boîte C/D dans les corps de Cajal devrait aussi apporter des éléments de réponse dans cette étude des complexes coiline en réponse au stress oxydant. En effet, notre étude sur les niveaux d'accumulation des différents acteurs n'a pas pu mettre en

évidence de variation claire de leur expression après un stress oxydant. La seule exception concerne la protéine TDP-43 qui voit son expression augmenter après 2h de stress puis diminuer. Même si actuellement cette variation ne peut pas être directement reliée au phénotype d'apparition des microfoci, cela n'en reste pas moins une piste intéressante pour comprendre les mécanismes mis en jeu lors de la réponse cellulaire au stress oxydant.

Afin d'assurer la maturation du spliceosome, les corps de Cajal concentrent les UsnRNPs au cours de leur biogenèse. Dans notre étude, nous avons observé que la protéine SMN est soit relocalisée dans les microfoci coiline, soit exclue du nucléoplasme. Or, le complexe SMN est impliqué dans l'adressage des UsnRNP en cours de maturation au sein des corps de Cajal. Même si ces données doivent être confirmées par des expériences de fractionnement noyau/cytoplasme, une exclusion de SMN du nucléoplasme pourrait impliquer une perturbation du ré-import des UsnRNP. Cependant, les données collectées en RNA-FISH montrent que l'intensité du signal nucléoplasmique des snRNPs U1 et U2 n'est pas changée par le stress oxydant. Toutefois, le stress modifie la localisation du U2 dans les corps de Cajal, qui se redistribue alors de façon plus homogène dans le nucléoplasme. Les corps de Cajal constituent de véritables plateformes de recrutement afin d'augmenter localement la concentration effective de facteurs spécifiques et de leur substrats. Par exemple, l'assemblage du tri-snRNP U4/U6-U5 se déroulerait 10 fois plus vite dans les corps de Cajal que dans le nucléoplasme environnant, en augmentant la concentration effective de tous les partenaires des UsnRNPs (Novotný et al., 2015). Il a été montré qu'un défaut de l'assemblage de la tri-snRNP pouvait induire la formation de corps de Cajal dans des cellules où ils sont absents, conduisant à l'accumulation dans ces substructures des intermédiaires d'assemblage de snRNPs (Novotný et al., 2015). De manière intéressante, grâce à des expérimentations d'immunoprécipitation des ARN (RIP), nous avons vu une augmentation de la fraction des UsnRNP U4, U5 et U6 au sein des UsnRNP totales en réponse au stress oxydant. Cette augmentation pourrait-elle jouer un rôle dans la redistribution nucléoplasmique de la coilin et de nombreux composants des corps de Cajal observée en réponse au stress oxydant ?

Les corps de Cajal sont des structures très dynamiques dont l'assemblage et le désassemblage est finement régulé par de nombreux processus souvent liés au cycle cellulaire ou



à l'activité métabolique. Même si de plus en plus de mécanismes régulant l'assemblage et le désassemblage des corps de Cajal sont décrits, la compréhension de la dynamique de nucléation et de redistribution nucléoplasmique demeure peu connue. Nos données nous ont permis de conclure que cet effet ne résulte pas d'une perturbation du cycle cellulaire, ni de l'intervention d'un régulateur de l'immunoprotéasome, la protéine PA28 $\gamma$ . Nous avons toutefois constaté que le stress oxydant induit une augmentation de l'expression du pseudogène Coilp1. Cette augmentation n'a été observée que pour les temps 4 à 8 heures, et ne pourrait donc pas expliquer la redistribution précoce de la coiline déjà observée au temps 2 heures. Toutefois, le produit de ce pseudogène est capable d'interagir avec certains constituants des corps de Cajal tels que les scaARNs 2 et 9, l'ARN de la télomérase ou la protéine WRAP53. Nous ne pouvons donc pas exclure que l'augmentation de l'expression de Coilp1 puisse modifier la disponibilité de ces composants et donc perturber la composition des complexes coiline. De plus, il est possible que Coilp1 modifie les complexes coiline en y étant directement intégré. Il paraît donc essentiel à ce stade de notre étude d'analyser la composition des complexes coiline. Nous avons, au cours de la thèse, voulu réaliser des expériences d'immunoprécipitation de la coiline afin d'étudier les partenaires de la coiline en conditions de stress. Mais nous avons été confrontés au problème que, lors de la préparation des extraits protéiques, la coiline est très majoritairement présente au sein de la fraction insoluble. Des optimisations des protocoles de préparation des extraits protéiques et d'immunoprécipitation ont été réalisées et nous allons pouvoir débiter l'étude de la composition des complexes coiline, qui nous l'espérons, sera très informative pour la compréhension de la fonction des CB dans le contexte d'une réponse physiologique de la cellule.

Cette étude a nécessité le traitement d'un grand nombre d'images acquises en microscopie de fluorescence et m'a conduit à me former à l'analyse d'images biologiques. Cette discipline m'a permis de m'initier à la compréhension des propriétés et des caractéristiques d'une image ainsi qu'aux méthodes employées afin de réduire le bruit de fond et améliorer la qualité du signal. Je me suis également initié à différentes techniques de segmentation qui constitue une des étapes les plus complexes dans l'analyse d'images biologiques. J'ai notamment développé un protocole d'extraction automatique de marqueurs pour la segmentation marker-based watershed. Enfin, j'ai développé des compétences en programmation en me formant au langage

Javascript, ainsi qu'à Matlab (offrant également des fonctionnalités de traitement d'images) et à R pour les analyses statistiques des données. Je me suis particulièrement investi dans cette partie du travail de thèse qui m'a fait découvrir un nouvel aspect de la recherche. J'ai d'ailleurs eu l'opportunité de suivre une formation à l'EMBL à Heidelberg sur des techniques avancées de microscopie de fluorescence. Cette dernière m'a notamment initiée à une technique dites de super résolution (PALM ou Photo-Activated Localization Microscopy) constituant une avancée extrêmement prometteuse pour l'avenir de la microscopie photonique. J'ai également été formé à des techniques dites de live-imaging, permettant de suivre la localisation de composants directement *in vivo* ou *in cellulo*. Cette formation m'a aussi donnée l'occasion d'échanger avec des spécialistes de l'analyse d'images biologiques et de renforcer mes connaissances dans le domaine. Vues de toutes les sollicitations que j'ai eues pour aider de nombreux autres scientifiques dans leur analyse d'images, j'ai pris conscience du véritable déficit en outils faciles d'utilisation et gratuits de ce domaine. C'est dans cet esprit que j'ai conçu et développé un programme d'analyse d'images sous Icy « *Substructure Analyzer* » afin de répondre à la forte demande des biologistes. J'espère que la publication de ce workflow dans le journal JoVE permettra une large diffusion et servira la communauté scientifique.

.



# ANNEXES



## ANNEXE 1

### TITLE:

Substructure Analyzer: a user-friendly workflow for rapid exploration and accurate analysis of cellular bodies in fluorescence microscopy images.

### AUTHORS AND AFFILIATIONS:

Géraud Heckler<sup>1</sup>, Christelle Aigueperse<sup>1</sup>, Liza Hettal<sup>1</sup>, Fabrice de Chaumont<sup>2</sup>, Stéphane Dallongeville<sup>2</sup>, Isabelle Behm-Ansmant<sup>1</sup>

<sup>1</sup>Université de Lorraine, CNRS, IMoPA F54000 Nancy France

<sup>2</sup>Institut Pasteur, BioImage Analysis Unit, CNRS UMR 3691, Paris, France.

Corresponding Author:

Isabelle Behm-Ansmant

Isabelle.behm@univ-lorraine.fr

Tel: +33 372746650

Email Addresses of Co-authors:

|                        |                                                                                                        |
|------------------------|--------------------------------------------------------------------------------------------------------|
| Géraud Heckler         | (geraud.heckler@univ-lorraine.fr)                                                                      |
| Christelle Aigueperse  | ( <a href="mailto:christelle.aigueperse@univ-lorraine.fr">christelle.aigueperse@univ-lorraine.fr</a> ) |
| Liza Hettal            | (liza.hettal@univ-lorraine.fr)                                                                         |
| Fabrice de Chaumont    | (fabrice.de-chaumont@pasteur.fr)                                                                       |
| Stéphane Dallongeville | (stephane.dallongeville@pasteur.fr)                                                                    |

### KEYWORDS:

Fluorescence microscopy; Bio-image analysis; Cell segmentation; Foci counting; Cellular bodies; Oxidative stress; Automatization; Translocation analysis; Substructure Analyzer; Icy; Image J

### SUMMARY:

We present a freely available workflow built for rapid exploration and accurate analysis of cellular bodies in specific cell compartments in fluorescence microscopy images. This user-friendly workflow is designed

on the open-source software Icy and also uses ImageJ functionalities. The pipeline is affordable without knowledge in Image Analysis.

#### **ABSTRACT:**

Last decade has been characterized by real breakthrough in fluorescence microscopy techniques illustrated by improvement in spatial resolution but also in live-cell imaging and high-throughput microscopy techniques. This led to a constant increase of the amount and complexity of the microscopy data for a single experiment. Because manual analysis of microscopy data is too much time consuming, pretty subjective, and does not allow quantitative analyses, automation of bio-image analysis is becoming almost unavoidable. We built an informatics workflow named Substructure Analyzer to fully automate signal analysis in bio-images from fluorescent microscopy. This workflow is developed on the user-friendly open source platform Icy (de Chaumont et al., 2012) and is completed by functionalities from another powerful software named ImageJ. It includes the pre-processing of images to improve the Signal to Noise ratio, the individual segmentation of cells (detection of cell boundaries) and the detection/quantification of cell bodies enriched in specific cell compartments. The main advantage of this workflow is to propose complex bio-imaging functionalities to users without image-analysis expertise through a user friendly interface. Moreover, it is highly modular and can be adapted to several problematics from the characterization of nuclear/cytoplasmic translocation, to the comparative analysis of different cell bodies in different cellular sub-structures. We illustrate the functionality of this workflow through the study of the Cajal (coiled) Bodies under oxidative stress (OS) conditions. Our preliminary data from fluorescence microscopy studies show that their integrity in human cells is impacted few hours after induction of an OS. This effect is characterized by a decrease of Coilin nucleation into characteristic nuclear foci, related to a nucleoplasmic redistribution into an increased number of smaller foci. The central role of Coilin in the exchange between CB components and the surrounding nucleoplasm suggests that OS induced redistribution of Coilin could affect the composition and the functionality of Cajal Bodies.

#### **INTRODUCTION:**

Light microscopy and more particularly fluorescence microscopy is a highly powerful and versatile technique commonly used in biological sciences. This tool gives access to the precise localization of various biomolecules like proteins or RNA through their specific fluorescent labelling. The last decade has been characterized by rapid advance in microscopy and imaging technologies. This is particularly evidenced by the 2014 Nobel Prize in Chemistry awarding Eric Betzig, Stefan W. Hell and William E. Moerner for the development of super-resolved fluorescence microscopy which bypasses the diffraction limit of traditional optical microscopy to bring it into the nanodimension. Improvement in techniques like live-imaging or high throughput screening approaches also participates in increasing the amount and the complexity of the data to treat for each experiment. Most of the time, researchers are faced with high heterogeneous populations of cells and want to analyze phenotypes at the single cell level.

Initially, basic analysis like foci counting was performed by eye, which is preferred by some researchers since it provides full visual control over the counting process. However manual analysis of such data is too much time consuming, leads to variability between observers, and cannot give access to more complex features, so that computer assisted approaches are becoming widely used and almost

unavoidable (Meijering et al., 2016). Bio-image informatics methods substantially increase the efficiency of data analysis and are free of the unavoidable operator subjectivity and potential bias of the manual counting analysis. The increased demand in this field and improvement of computer power led to the development of a large number of image-analysis platforms. Some of them are freely available and give access to various tools to perform your own analysis with your own computers. A classification of open access tools has been recently established (Wiesmann et al., 2015) and presents Icy (de Chaumont et al., 2012) as a powerful software combining usability and functionality. Moreover, Icy has the advantage to communicate with ImageJ, another highly functional software.

For users without image-analysis expertise, the main obstacle is to choose the appropriate tool according to the problematic and tune correctly parameters that are often not well understood. Moreover, setup times is often long. Icy software proposes a user-friendly point-and-click interface named “Protocols” to develop your own workflow by combining some plugins found within an exhaustive collection. The flexible modular design and the point-and-click interface make setting up an analysis feasible for non-programmers. Here we present a workflow named “*Substructure Analyzer*”, developed in Icy’s interface, which function is to analyze fluorescent signals in specific cellular compartments and measure different features (mean and sum intensity, foci number and size...). This workflow can be adapted to several problematics like the quantification of signal translocation, the analysis of transfected cells expressing a fluorescent reporter or the analysis of foci from different cellular substructures in individual cells. It allows the simultaneous processing of multiple images and output results are easily readable after an export in a tab-delimited spreadsheet format that can be opened in commonly used programs such as Microsoft Excel or OpenOffice Calc.

The *Substructure Analyzer* pipeline is presented in Figure 1. All the images contained into a specified folder are firstly pre-processed to improve their signal to noise ratio. This step greatly increases the efficiency of the following steps and decreases the running time. Then, the Regions of Interest (ROIs), corresponding to image area where the fluorescent signal should be detected, are identified and segmented. Finally, the fluorescent signal is analyzed and results are exported into a given spreadsheet. Object segmentation (detection of boundaries) is the most challenging step in image analysis and its efficiency determines the accuracy of the resulting cell measurements. The first objects identified in an image (called primary objects) are often nuclei from DNA-stained images (DAPI or Hoechst staining), although primary objects can also be whole cells, beads, speckles, tumors, and so on. In most biological images, cells or nuclei touch each other or overlap, causing the simple fast algorithms to fail. To date no universal algorithm is able to perform perfect segmentation of all objects mostly because their characteristics (size, shape, texture...) modulate the efficiency of segmentation. Cells or cell nuclei segmentation tools commonly distributed with microscopy softwares (for example, MetaMorph Imaging Software by Molecular Devices) are usually based on classical techniques such as correlation matching, thresholding, or morphological operations. Although efficient in basic systems, these overgeneralized methods rapidly present limitations when used in more challenging and specific contexts. Indeed, segmentation can be highly sensitive to experimental parameters such as cell type, cell density and bio-markers, and frequently requires repeated adjustment for a large data set. The *Substructure Analyzer* workflow integrate both simple and more complex algorithms to propose different alternatives adapted to image complexity and user needs. It notably proposes the marker-based watershed algorithm (Meyer and Beucher, 1990) for highly clustered objects. The efficiency of this segmentation method relies on the



selection of individual markers on each object. These markers are most of the time manually chosen to get correct parameters for full segmentation, which is highly time consuming when users are faced to a great number of objects. In our method, we propose an automatic detection of these markers, providing a highly efficient segmentation process. Most of the time, segmentation is the limiting step of image analysis and can change the time of processing depending on the image resolution, the number of objects per image, and the level of object's clustering. Typical pipelines require from a few seconds to 5 minutes per image on a standard desktop computer. Analysis of more complex images can require a more powerful computer and some basic knowledge in image analysis.

The flexibility and functionality of this workflow will be illustrated with various examples in the “Representative results” section. Advantages of this workflow will be notably displayed through the study of nuclear substructures under oxidative stress (OS) conditions. OS corresponds to an imbalance of the redox homeostasis in favor of oxidants and is associated with high levels of reactive oxygen species (ROS). Since ROS act as signaling molecules, changes in their concentration and subcellular localization affect positively or negatively a myriad of pathways and networks that regulate physiological functions, including signal transduction, repair mechanisms, gene expression, cell death and proliferation (D’Aur aux and Toledano, 2007; Schieber and Chandel, 2014). OS is thus directly involved in various pathologies (neurodegenerative and cardiovascular diseases, cancers, diabetes...), but also in cellular aging. Therefore, deciphering the consequences of OS on human cell’s organization and function constitutes a crucial step in the understanding of the roles of OS in the onset and development of human pathologies. It has clearly been established that OS regulates gene expression by modulating transcription through several transcription factors (p53, Nrf2, FOXO3A..) (Davalli et al., 2016), but also by affecting the regulation of several co- and post-transcriptional processes such as alternative splicing (AS) of pre-RNAs (Disher and Skandalis, 2007; Seo et al., 2016; Takeo et al., 2009). AS of primary coding and non-coding transcripts is an essential mechanism which increases the encoding capacity of the genomes by producing transcript isoforms. AS is performed by a huge ribonucleoprotein complex called spliceosome containing almost 300 proteins and 5 U-rich small nuclear RNAs (UsnRNAs) and is tightly controlled in cells (Will and Luhrmann, 2011). Importantly, some steps of the spliceosome maturation occur within membrane-less nuclear compartments named Cajal Bodies. These nuclear substructures are characterized by the dynamic nature of their structure and their composition, which are mainly conducted by multivalent interactions of their RNA and protein components with the coilin protein. Analysis of thousands of cells with the *Substructure Analyzer* workflow allow us to characterize a never described effect of OS on Cajal Bodies. In fact, our data suggest that OS modifies the nucleation of Cajal Bodies, inducing a nucleoplasmic redistribution of the coilin protein into numerous smaller nuclear foci. Such a change of the structure of Cajal Bodies might affect the maturation of the spliceosome and participate to AS modulation by OS.

## PROTOCOL:

User-friendly tutorials are available on Icy’s website <http://icy.bioimageanalysis.org>.

## 1. Download Icy and the protocol on Icy website:

Icy: <http://icy.bioimageanalysis.org/download>

! If you are using a 64 bits OS, be sure to use the 64 bit version of Java. This version will allow you to increase the memory allocated to Icy (Preferences → General → Max memory).

“Substructure Analyzer” protocol: <http://icy.bioimageanalysis.org/protocols?sort=latest>

## 2. Opening the protocol

- 2.1 Open Icy and click on “Tools” in the Ribbon menu.
- 2.2 Click on “Protocols” to open the “Protocols editor” interface.
- 2.3 Click on “Load” and open the protocol “Substructure Analyzer”: Protocol loading can take some seconds. Be sure that opening of the protocol is complete before using it.

## 3. Interacting with the workflow on Icy

- 3.1 The workflow is composed of **13 general blocks** presented in **Figure 2a**. Each block performs a specific task adapted to a given problematic (Image merging, object segmentation, analysis of fluorescent signal...).
- 3.2 Each block works as a pipeline composed of several boxes performing specific subtasks.
- 3.3 Each block/box has a specific rank within the workflow/pipeline as illustrated by a number in the upper right corner. By clicking on this number, the closest possible position to the first is assigned to the selected block/box, while positions of the other blocks/boxes are re-organized. Changing positions of boxes within a block will not affect the position of the block inside the workflow.

! **Respect the right order of the blocks** when preparing the workflow. For example, Spot Detector blocks needs pre-defined ROIs so that Segmentation blocks have to run before Spot Detector blocks.

! Do not modify the position of boxes!

! Never use “.” in image’s name!

- 3.4 Each block/box can be expanded, collapsed or removed by clicking on its upper left corner icon (**Figure 2b**). When expanded, they can also be enlarged or narrowed.
- 3.5 Each pipeline is characterized by a network of boxes connected together via their **input** and **output** (**Figure 2b**). To create a connection, click on an output and maintain your click until the cursor attain an input. **Inputs and outputs have to be compatible to be linked** (for example, a text output will only connect with a text input). **Connection can also be removed by clicking on the output tag.**

## 4. If necessary, merge the channels of interest in the same image

Perform the following steps if several channels are needed for the same run (for example, a channel used to segment nuclei and another to detect spots inside segmented nuclei).

4.1 **If sequences are already merged:** Put all the merged images in the same folder.

4.2 **If color's channels are separated in different sequences:** Use the block "Merge Channels" to generate merged images:

- Sequences to be merged should have **the same name's prefix followed by a distinct separator**. For example, sequences of individual channels from an Image A could be named: ImageA\_**red**, ImageA\_**green** and ImageA\_**blue**.
- In the same folder, create one new folder per channel to merge (for example to merge red, green and blue channels, create 3 folders) and put the corresponding sequences inside.

! For the separator, do not use characters already present in the image's name! For example, if the image's name is "**Phalloïdin Dapi Tubulin (1)**", do not use "**Dapi**" as separator for the dapi image but better use "blue" or "DAPI".

- Click on the number at the upper right hand corner of the "Merge Channels" block to set its position to 1.
- Access to the boxes within the "**Merge Channels**" block.
- For each channel, fill the boxes "**Channel number X**" (boxes 1, 5 or 9), "**Folder channel number X**" (boxes 2, 6 or 10), "**Separator channel number X**" (boxes 3, 7 or 11) and "**Colormap channel nb X**" (boxes 4, 8 and 12) respectively. These boxes are horizontally grouped by four, each line corresponding to the same channel. In each line, a display is also available (boxes 23, 24 or 25) to directly visualize the sequence of the corresponding channel.
- In the box "**Channel number X**", choose which channel to extract (in classical RGB images, 0=Red, 1=Green, 2=Blue). User can easily access to the different channels of an image within the "Inspector" window of Icy, in the "Sequence" tab. The smallest channel's value should be written in the upper line and the highest in the bottom line.
- In the box "**Folder channel number X**", write: \Name of the folder containing images of channel X.
- In the box "**Separator channel number X**", write the separator used for image's name (in the previous example: "\_red", "\_green" and "\_blue").
- In the box "**Colormap channel nb X**", indicate with a number which colormap model to use to visualize the corresponding channel in Icy. All the available colormaps can be accessed within the "Sequence" tab of the "Inspector" window. (Colormap models: 0=Gray, 1=Red, 2=Green, 3=Blue, 4=Magenta, 5=Yellow, 6=Cyan, 7=Ice, 8=Fire, 9=HSV[1.0, 1.0], 10=JET, 11=Glow Under Over, 12=Ametrine, 13=Isolum, 14=Morgenstemning, 15=Parula).

! If you want to **merge 2 channels**, do not fill the four boxes corresponding to the third channel (boxes 9, 10, 11 and 12).

- In the box **“Format of merged images”** (box 28), write the extension to save merged images: .tif, .gif, .jpg, .bmp or .png.
- On the upper left corner of the **“Merge Channels”** Block, click on the link directly to the right of **“folder”**. In the **“Open”** dialog box which appears, **double-click on the folder containing sequences of the first channel that you have defined in boxes “Folder channel number 1” (box 2)**. Then, click on **“Open”**.
- Run the protocol (see part 7).
- Merged images are saved in a **“Merge”** folder, in the same directory as the folders of individual channels. Names of merged images correspond to the common prefix previously used (ImageA in our example above).

## 5. Segmentation of the Regions of Interest

In this protocol, **primary objects** correspond to any cellular substructure with an available fluorescent marker in images (Nuclei, cytoplasm, mitochondria, nucleoli...). Blocks adapted for primary objects segmentation process independently so that several blocks can be used in the same run to compare their efficiency for a particular substructure or to segment different types of substructures.

Blocks adapted for the segmentation of clustered objects are more time consuming. If the clustering level is heterogeneous within the same set of images, it is recommended to process small and highly clustered objects separately in the adapted blocks!

### 5.1 Select the adapted block:

- If objects do not touch each other, or the user does not need to differentiate clustered objects individually, use the block **“Segmentation A: Non clustered objects”**.
- If objects do not touch each other but some of them are in close proximity, use the block **“Segmentation B: Poorly clustered objects”**.
- If objects present a high clustering level and have a convex shape, use the block **“Segmentation C: Clustered objects with convex shapes”**.
- If objects present a high clustering level and have irregular shapes, use the block **“Segmentation D: Clustered objects with irregular shapes”**.
- The block **“Segmentation E: Clustered cytoplasm”** specifically segments touching cytoplasm individually by using segmented nuclei as markers. This block imperatively needs segmented nuclei to process!

5.2 Link the **output0** (File) of the block **“Select Folder”** (block 1) to the **“folder”** input (white arrows in black circles) of the chosen segmentation block.

5.3 Set parameters of the chosen segmentation block:

- **“Segmentation A: Non clustered objects”** and **“Segmentation C: Clustered objects with convex shapes”**:

- In box “**Channel signal**” (box 1), set the channel of the object to segment.
- **Optional:** In box “**Gaussian filter**” (box 2), increase the X and Y sigma values (Z for Z-stacks) if the signal inside objects is heterogeneous (a lot of minima and maxima). Gaussian filter smooth out textures in order to obtain more uniform regions, which can be useful to increase the speed and the efficiency of nuclei segmentation. Avoid high sigma values. The more objects are small, the more sigma value should be low. Defaults values are set to 0.
- In box “**HK-Means**” (box 3), set the “**Intensity classes**” parameter. This step should lead to a **primary detection** of foreground objects in ROIs. By setting 2 intensity classes (minimum value), pixels will be classified in 2 classes: background and foreground. So a 2 value is adapted when contrast between the objects and the background is high. If foreground objects have heterogeneous intensities or if contrast with the background is low, increase the number of classes. Default setting is 2. In the same box, **set the approximative minimum and maximum sizes** (in pixels) of objects to detect. Object’s size can be easily evaluated by drawing manually a ROI around the object of interest. Size of the ROI (Interior in pixels) appears directly on the image when pointing it with the cursor or can be accessed in “ROI statistics” window (open it from the search bar). **Optimal parameters detect each foreground object in a single ROI.** They can be manually defined in Icy (Detection and Tracking → HK-Means).
- **Optional:** In box “**Active Contours**” (box 4), set parameters to optimize the detection of object’s border. Exhaustive documentation of this plugin is available online: [http://icy.bioimageanalysis.org/plugin/Active Contours](http://icy.bioimageanalysis.org/plugin/Active%20Contours). Correct parameters can also be manually defined in Icy (Detection and Tracking → Active Contours).
- **Optional:** In box “**Format of images of segmented objects**” (box 5), set the format for saving segmented objects images (Tiff, Gif, Jpeg, BMP, PNG).
- During the process, a folder will be automatically created to save images of segmented objects. In box “**Text**” (box 6), name this folder (ex: Segmented nuclei). **The folder is created in the folder containing merged images.**
- Run the workflow. See part 7.
  - “**Segmentation B: Poorly clustered objects**”:
- Follow the same steps of 5.3.1 to set parameters of boxes “**Channel signal**”, “**HK-Means**”, “**Active Contours**”, “**Extension to save segmented objects**” and “**Text**” (Boxes ranks are not the same as in 5.3.1!).
- In box “**Call IJ plugin**” (box 4), set the “**rolling**” parameter to control background subtraction. This parameter should be set to at least the size of the largest object that is not part of the background. Decreasing this value will increase background removal, but can also increase removing of foreground signal.

- In box **“Adaptive histogram equalization”** (box 6), contrasts are improved between the foreground objects and the background. Increasing the slope will give more contrasted sequences.
- Run the workflow. See part 7.

- **“Segmentation D: Clustered objects with irregular shapes”:**

This block has to be connected with the block **“Clustered/heterogeneous shapes primary objects segmentation Dialog Box”** to work properly.

This block applies successively **3 different methods of segmentation on each image**: Firstly, **HK-means clustering combined with Active Contours method** are applied. Then, **classical Watershed algorithm** (using the Euclidian distance map) is applied on previously mis-segmented objects. Finally, a **marker-based watershed algorithm** is applied. Only HK-means and marker-based watershed methods need user intervention. For both methods, the same parameters can be applied for all images (“fully automated” version) or be changed for each image (“semi-automated” version). **If the user is not trained to these segmentation methods, the semi-automated processing is highly recommended.**

During the processing of this block, manual intervention is needed. When a segmentation method is finished, user has to manually remove mis-segmented objects before the beginning of the next segmentation method. Successfully segmented objects are saved and not taken into account in the next step(s).

Before the processing:

- Download the ImageJ collection **“MorphoLibJ”** on: <https://github.com/ijpb/MorphoLibJ/releases>. The MorphoLibJ 1.4.0 version is used in this protocol. Place the file **“MorphoLibJ\_-1.4.0.jar”** in the folder **icy→ij→plugins**. If opened, Icy should be restarted. More information about the content of this collection is available on: <https://imagej.net/MorphoLibJ>
- Follow the same steps of 5.3.1 to set parameters of boxes **“Channel signal”**, **“Gaussian filter”**, **“Active Contours”**, **“Extension to save segmented objects”** and **“Text”** (Boxes ranks are not the same as in 5.3.1!).
- Set parameters of the box **“Adaptive histogram equalization”** (see 5.3.2).
- **Optional:** To activate **“Subtract Background”**, write **“yes”** in **“Apply Subtract Background?”** (box 5). Else, write **“no”**. If the plugin is activated, set the rolling parameter (see 5.3.2), in the box **“Subtract Background parameter”** (box 7).
- **Automatization of HK-means:** To apply the same parameters for all images (fully automated processing), set the **“Nb of classes”** (box 11), the **“Minimum size”** (box 12) and the **“Maximum size”** (box 13) (see 5.3.1). These parameters should be set to select a maximum of foreground pixels, and to optimize the individualization of foreground objects. For the semi-automated processing version, no intervention is needed.

- **Automatization of markers extraction:** Markers are extracted by applying an extended-minima transformation on an input image controlled by a “dynamic” parameter. In the marker-based watershed algorithm, flooding from these markers is simulated to perform object segmentation. For successful segmentation of foreground objects, **one single marker per foreground object should be extracted**. For the fully-automated version, expand the box “**Internal Markers extraction**” (box 27) and set the value of the “dynamic” parameter in line 13 of the script. For the semi-automated processing version, no intervention is needed. Setting of the “dynamic” parameter for optimal markers extraction depend mostly on the image’s resolution. Thus, if you are not familiar with this parameter or you do not know which value to apply, use the semi-automated version.
- Run the workflow. See part 7.

During the processing:

- At the beginning of the processing, dialog boxes “**HK-means parameters**” and “**Marker-based watershed**” successively open. To apply the same parameters for all images (fully-automated version) click on “YES”. Otherwise, click on “NO”.
- If you have selected “YES” for the full-automatization of HK-means algorithm, parameters previously set will be applied to all images. If you have selected “NO” an information box will open, asking to “Determine optimal ROIs with HK-Means plugin and close image”. Click on “OK” and manually apply the HK-Means plugin (Detection and Tracking→HK-Means) on the image which automatically opens. “Export ROIs” option should be selected in the HK-Means plugin box. Apply the best parameters to have ROIs containing a **maximum of foreground pixels and to optimize the individualization of foreground objects**. When optimal ROIs are found, directly close the image.
- At the end of the first segmentation method, an information box will open and ask to “Remove unwanted ROIs and close image”. These ROIs correspond to the borders of the segmented objects. Select “OK”, and remove ROIs of mis-segmented objects in the image which automatically opens. A ROI can be easily removed by placing the cursor on its border and using the “Delete” button of the keyboard. Close the image.
- After the second segmentation method, a new information box asks to remove unwanted ROIs. Each ROI corresponds to a limit between two objects. Remove the mis-placed limits like in the previous step and close the image.
- A new box asks to remove unwanted ROIs corresponding to individual objects updated after the removing of ROIs in the previous step. Close the image when ROIs are removed like in the last two steps.

- If you have selected “YES” for the full-automatization of the marker-based watershed algorithm, the previously set parameters will be applied to all images. If you have selected “NO” an information box will open, asking to “Determine and adjust internal markers”. Click on “OK” and within the ImageJ interface of Icy, go to: Plugins→MorphoLibJ→Minima and Maxima→Extended Min & Max. In “Operation”, select “Extended Minima”. Select “Preview” to pre-visualize on the automatically opened image the result of the transformation. Move the dynamic until optimal markers are observed. Markers are groups of pixels with a value of 255 (Not necessary white pixels). Optimal parameters should lead to one marker per object. **Focus on the remaining objects that have not been well segmented with the two previous segmentation methods.** If necessary, additional morphological operations like “Opening” or “Closing” can be applied on the markers to improve them (Plugins→MorphoLibJ→Morphological Filters). When you get the final image of markers, keep it open and close all the other images ending with the image initially used as an input for the “Extended Minima” operation. Click on “No” if an ImageJ box ask to save changes on this image.
- **Optional:** In box “Nb of images with Information Box” (box 14), users can determine how many images with Information Boxes should appear.
  - **“Segmentation E: Clustered cytoplasm”:**  
This block uses previously segmented nuclei as individual markers to initiate cytoplasm segmentation. Be sure that the block of nuclei segmentation has processed before using this block!
- In box “Channel cytoplasm” (box 1), set the channel of the cytoplasmic signal.
- In box “Extension segmented nuclei” (box 2), write the format used to save images of segmented nuclei (tif, jpeg, bmp, png). Default format is “tif”.
- In box “Text” (box3), write: \Name of the folder containing segmented nuclei.
- **Optional:** In box “Format of images of segmented cytoplasms” (box 4), set the format to use for saving segmented objects images (Tiff, Gif, Jpeg, BMP, PNG).
- During the process, a folder will be automatically created to save images of segmented cytoplasms. In box “Text” (box 5), name this folder (ex: Segmented cytoplasms). **The folder is created in the folder containing merged images.**
- Follow the same steps of 5.3.1 to set parameters of boxes “Gaussian filter” and “Active Contours” (Boxes ranks are not the same as in 5.3.1!).
- Run the workflow. See part 7.



## 6. Fluorescent signal detection and analysis

### 6.1 Select the adapted block:

- 6.1.1. In block **“Fluorescence Analysis: 1 Channel Spot Detector”**, detection and analysis of foci is performed in one channel inside one type of segmented object. Example, detection of Coilin foci (red channel) within the nucleus.
- 6.1.2. In block **“Fluorescence Analysis: 2 Channels Spot Detector in the same compartment”**, detection and analysis of foci is performed in two channels inside one type of segmented object. Example: detection of Coilin (red channel) and 53BP1 (green channel) foci within the nucleus.
- 6.1.3. In block **“Fluorescence Analysis: 2 Channels Spot Detector in two compartments”**, detection and analysis of foci is performed in one or two channels specifically inside the nuclei and their corresponding cytoplasm. Example 1: detection of Coilin foci (red channel) both within the nucleus and its corresponding cytoplasm. Example 2: detection of Coilin foci (red channel) within the nucleus and G3BP foci (green channel) within the corresponding cytoplasm.
- 6.1.4. In block **“Global Translocation Analysis”**, percentage of signal from one channel in two cellular compartments (A and B). For example, in a cytoplasm/nucleus translocation assay, calculated percentages of nuclear and cytoplasmic signals for each image are exported in the final “Results” spreadsheet. Formula used to calculate nuclear signal percentage is shown below. This block can be used for any subcellular compartment:

$$\% \text{ nuclear signal} = \frac{\text{Sum of all nuclei signals} \times 100}{(\text{Sum of all nuclei signals} + \text{Sum of all cytoplasmic signals})}$$

- 6.1.5. In block **“Individual Cell Translocation Analysis”**, percentage of signal from one channel in two cellular compartments for each individual cell. This block is especially optimized for nucleus/cytoplasm translocation assay at the single cell level.

! Because the block **“Individual Cell Translocation Analysis”** performs analysis at the single cell level, efficient segmentation of nucleus and cytoplasm is needed!

6.2 Link the **output0** (File) of the block **“Select Folder”** (block 1) to the **“folder”** input (white arrows in black circles) of the chosen block.

### 6.3 Set the parameters of the chosen block:

- 6.3.1. “Fluorescence Analysis: 1 Channel Spot Detector”, “Fluorescence Analysis: 2 Channels Spot Detector in the same compartment” and “Fluorescence Analysis: 2 Channels Spot Detector in two compartment”:  
 ➤ In box “Folder images ROI”, write: \Name of the folder containing images of segmented objects. (For example: \Segmented nuclei).

- In box **“Format of images of segmented objects”** (box 2), write the format used to save images of segmented objects (tif, jpeg, bmp, png). Default format is “tif”.
- In box(es) **“Channel spots signal”**, set the channel where spots have to be detected. In classical RGB images, 0=Red, 1=Green and 2=Blue.
- In box **“Kill Borders?”**, write “Yes” if you want to remove border objects. Otherwise, write “No”. **You need to install the MorphoLibJ collection of ImageJ to use this function!** (See 5.3.3)
- In box(es) **“Wavelet Spot Detector Block”**, set the parameters for the detection of the spots. One detector box is available per channel and should be parameterized according to spot features. User can set the **scale(s)** (referred to spot size), and the **sensitivity of the detection** (a smaller sensitivity decreases the number of detected spots, default value being 100 and minimum value being 0). Exhaustive documentation of this plugin is available online: [http://icy.bioimageanalysis.org/plugin/Spot\\_Detector](http://icy.bioimageanalysis.org/plugin/Spot_Detector). Parameters can also be manually defined in Icy (**Detection and Tracking → Spot Detector**).
- **Optional:** In boxes **“Filter ROI by size”**, filter the segmented objects where spots are detected by setting an interval of size (in pixels). This step could be especially useful to remove under- or over-segmented objects. To manually estimate the size of objects, see 5.3.1. Default parameters do not include filtering of ROIs by size. Block “2 Channels Spot Detector B” contains two boxes: “Filter nuclei by size” (box 19) and “Filter cytoplasm by size” (box 46).
- **Optional:** In boxes **“Filter spots by size”**, filter the detected spots according to their size (in pixels) to remove unwanted artifacts. **Filtered spots are not taken into account for the analysis.** To manually estimate spots size, click on **Detection and Tracking** and open the **Spot detector** plugin. In **Output** options, select **Export to ROI**. Then, all detected spots are exported on your test image as ROI. Features for each spot such as its size appear when pointing the ROI with the cursor. Default parameters do not include filtering of spots by size. The number of fields to enter depends on the number of channels. In each block **“Fluorescence Analysis: 2 Channels Spot Detector”** there are two boxes: “Filter spots A by size” (box 24) and “Filter spots B by size” (box 29) in the block **“Fluorescence Analysis: 2 Channels Spot Detector in the same compartment”**. “Filter nuclear spots by size” (box 23) and “Filter cytoplasmic spots by size” (box 36) in the block **“Fluorescence Analysis: 2 Channels Spot Detector in two compartments”**.
- **Optional:** In boxes **“Filter spots”**, apply an additional filter (contrast, homogeneity, perimeter, roundness...) on the detected spots. Filtered spots are not taken into account for the analysis. Default parameters do not include spot filtering. The number of fields to enter depends on the number of channels. In each block **“Fluorescence Analysis: 2 Channels Spot Detector”** there are two boxes: “Filter spots A” (box 25) and “Filter spots B” (box 30) in the block **“Fluorescence Analysis: 2 Channels Spot Detector in the same compartment”**. “Filter nuclear spots” (box 24) and “Filter cytoplasmic spots” (box 37) in the block **“Fluorescence Analysis: 2 Channels Spot Detector in two compartments”**.

- In boxes “**Name of localized molecule**”, write the name of the molecule localizing into the spots. Number of fields to enter depends on the number of molecules.
- **Optional** : In boxes “**Spot size threshold**”, if wanted, set a threshold for the area (in pixels) of analyzed spots. The number of counted spots below and above this threshold is exported in the final “**Results**” spreadsheet. Number of boxes to be informed depends on the number of channels.
- Run the workflow. See part 7. Data are exported in a spreadsheet “**Results**” saved in the folder containing merged images.

#### 6.3.2. “Global Translocation Analysis” and “Individual Cell Translocation Analysis”:

- In boxes “Folder images ROI” (boxes 1 and 2), write: \Name of the folder containing images of segmented objects. (For example: \Segmented nuclei). In block “Fluorescence Analysis: 2 Channels Spot Detector in the same compartment”, the two types of ROI are identified as ROI A and ROI B. For the block “Fluorescence Analysis: 2 Channels Spot Detector in two compartments”, in “Folder images segmented nuclei” and “Folder images segmented cytoplasms” boxes write the name of folder containing segmented nuclei and cytoplasm respectively.
- In box “**Channel signal**” (box 3), enter the channel of the signal.
- In box “**Format of images of segmented objects**” (box 4), write the format used to save images of segmented objects (tif, jpeg, bmp, png). Default format is “tif”.
- To remove border objects, the “**Kill Borders**” option is also available (see 6.3.1).
- **Optional**: In boxes “**Filter ROI by size**”, filter the segmented objects by setting an interval of size (in pixels). This step could be useful to remove under- or over-segmented objects. To manually estimate object size, see 5.3.1. There are two fields to enter, one per channel. Default parameters do not include ROI filtering by size.
- Run the workflow. See part 7. Data are exported in a spreadsheet “**Results**” saved in the folder containing merged images.

## 7. Run the protocol

- 7.1 **Processing one block in a run**: Remove the connection between the selected block and the block “**Select Folder**”. Place the wanted block at the first rank. On the upper left corner of the wanted block, click on the link directly to the right of “**folder**”. In the “**Open**” dialog box which appears, **double-click on the folder containing the merged images**. Then, click on “**Open**”. Click on “**Run**” to start the workflow. If necessary, the processing can be stopped by clicking on the “**Stop**” button.

- 7.2 **Processing different blocks in a run:** Keep connections of chosen blocks with the block “Select Folder” (block 1). **Be sure that their rank allows the good processing of the workflow!** For example, if a specific block needs segmented objects to process, be sure that the segmentation block will process before. **Before running the workflow, it is also recommended to remove unused blocks and save the new protocol with another name!** Click on “Run” to start the workflow. When the open dialog box appears, **double-click on the folder containing the merged images**. Then, click on “Open”. The workflow will automatically run. If necessary, the processing can be stopped by clicking on “Stop”.
- 7.3 Several signs can appear next to the rank of a block/box, according to its processing states (see **Figure 2b**). If the processing finished successfully, the following message will appear in the lower right hand corner: “The workflow executed successfully (total running time: X seconds)”. If not, the message will be: “The workflow did not execute properly (total running time: X seconds)”. In this case, find the block and the inside box presenting the error sign (**Figure 2b**) to target the error location.
- 7.4 If the workflow executed successfully, all the blocks will be flagged with a green sign (Figure 2b). In this case, a new run cannot be directly started and to process the workflow again, at least one block should be flagged with the sign “ready to process”. To change the state of a block, you can either delete and re-create a link between two boxes inside this block or simply close and re-open the protocol. If an error occurs during the processing, a new run can directly be started. During a new run, all the blocks of the pipeline are processed, even if some of them are flagged with the green sign.

## REPRESENTATIVE RESULTS:

All the described analyses have been performed on a standard laptop (64 bits, quad-core processor at 2.80 GHz with 16 GB RAM) working with the 64-bit version of Java. Random access memory is an important parameter to take into account depending on the amount and the resolution of images to analyze. Using the 32-bit version of Java will limit the memory to about 1300MB, which could be unsuitable for big data analysis whereas the 64-bit version allows to increase the memory allocated to Icy. The time needed for segmentation, one of the most time consuming step in bio-image analysis, for different types of images and different resolutions is reported in **Figure 3** (Note that high resolution can substantially increase the time of segmentation).

The data presented in the next paragraphs demonstrate that the *Substructure Analyzer* workflow can be used to solve most of the common problematics encountered in cellular and molecular biology (cell counting, foci counting and analysis, translocation analysis).

Measuring in numerous cells the precise proportion of molecules concentrated within a given compartment and/or the change of localization across subcellular domains can be a very cumbersome and error prone task, especially when it is performed manually. To this purpose, we firstly tested the ability of the workflow to rapidly and precisely quantify a well described nuclear translocation event: that of the transcription factor NFκB in response to TNFα stimulation. The protocol could be applied for any type of cell's compartments assuming that regions of interest could be precisely defined. Images that are used here were graciously compiled by Ilya Ravkin (<http://www.ravkin.net/SBS/Image-Library.htm>) and are publicly available at the Broad Bioimage Benchmark Collection (Ljosa et al., 2012). This set contains images of MCF7 and A549 cell lines treated with increasing concentrations of TNFα. For this illustration purpose,

we focused our analysis on the MCF7 cell lines. We analyzed more than 40 000 cells and 96 images (48 images for each channel (**Figure 4.a**)). The whole analysis (from import of images to data saving) took 26 minutes. Each image corresponds to a specific TNF $\alpha$  concentration. The nuclear/cytoplasmic proportion of NF $\kappa$ B were homogeneous across all cells for a given image. For this reason, for each image, we summed the values of all nuclear or cytoplasmic pixels to calculate the nuclear and cytoplasmic proportions of NF $\kappa$ B, and individualization of touching nuclei/cytoplasms during their segmentation was not needed. We processed DAPI images in the **“Segmentation C: Clustered objects with convex shapes”** block to segment nuclei and then we used the green channel to delineate the cytoplasms with the **“Segmentation E: Clustered cytoplasm”** block (**Figure 4.b**). We used the **“Global Translocation Analysis”** block to export the sum of nuclear and cytoplasmic pixel values. The obtained data are represented in a dose-response curve (**Figure 4.c**) and illustrate the increase of NF $\kappa$ B nuclear proportion after stimulation with increasing concentrations of TNF $\alpha$ .

The workflow not only analyzes global intensity signal in different sub-cellular compartments, but can also be used to detect foci and extract specific information about their features. In **Figure 5**, we illustrate the detection of P-bodies in individual cells by localizing the enhancer of mRNA decapping 4 (EDC4) protein. We used the **“Segmentation A: Non clustered objects”** block to segment nuclei which present a low level of clustering. Segmented objects are specifically identified (Nucleus\_3, Cytoplasm\_3) and multiple information can be collected into the corresponding worksheet (Nucleus Information and Cytoplasm Information) like the EDC4 mean intensity or the number/size of detected bodies in each ROI. Before being analyzed, detected bodies can be preliminary filtered according to their area, which can be useful to exclude objects below the resolution power of the microscope. In order to have a preliminary overview of the size of conserved objects, an additional area threshold (100 pixels in this example) can be introduced. Areas of all the detected bodies are reported in dedicated worksheets (Nucleus\_Foci size, Cytoplasm\_Foci size). In this example, a total of 17 P-bodies have been detected in the cytoplasm of cell numbered 2 and they all have an area below 100 pixels. In agreement with the description of P-bodies as cytoplasmic granules, no nuclear bodies have been detected.

Cajal Bodies are dynamic structures nucleating and dissolving within the nucleoplasm under the control of specific parameters like the transcription rate or cell cycle progression. Oxidative stress (OS) effect on Cajal Bodies remains unknown. We took advantage of the workflow versatility to study these membrane-less nuclear bodies along a kinetics of oxidative stress (H<sub>2</sub>O<sub>2</sub> 500  $\mu$ M) from 2 to 20 hours in order to visualize potential early and late effects. We visualized Cajal Bodies by localizing their main structural and functional component, the coilin protein. We collected information about their number and size by studying 2300 individual cells from high resolution images (3840X3072) containing 3 channels: DAPI (blue), 53BP1 (green) and coilin (red). Full processing took less than 1h. Because the nuclei presented convex shape and some of them were clustered, we segmented them by using the block **“Segmentation C: Clustered objects with convex shapes”**. As OS is known to induce DNA double strand breaks (DDSBs), we used the p53-binding protein 1 (53BP1), known to be an important effector of DNA damage signaling pathways, as a marker to discriminate between stressed and non-stressed cells. Indeed, in the absence of DNA damages, 53BP1 is homogeneously distributed within the nucleoplasm, whereas following DNA damages it concentrates on DDSBs which form easily distinguishable foci in microscopy. We get the number and the size of both 53BP1 and coilin nuclear foci in each cell by using the block **“Fluorescence Analysis: 2 Channels Spot Detector in the same compartment”**. Firstly, analysis of 53BP1 data helped us to determine for each time point of the kinetics the best threshold to classify cells according to their stress

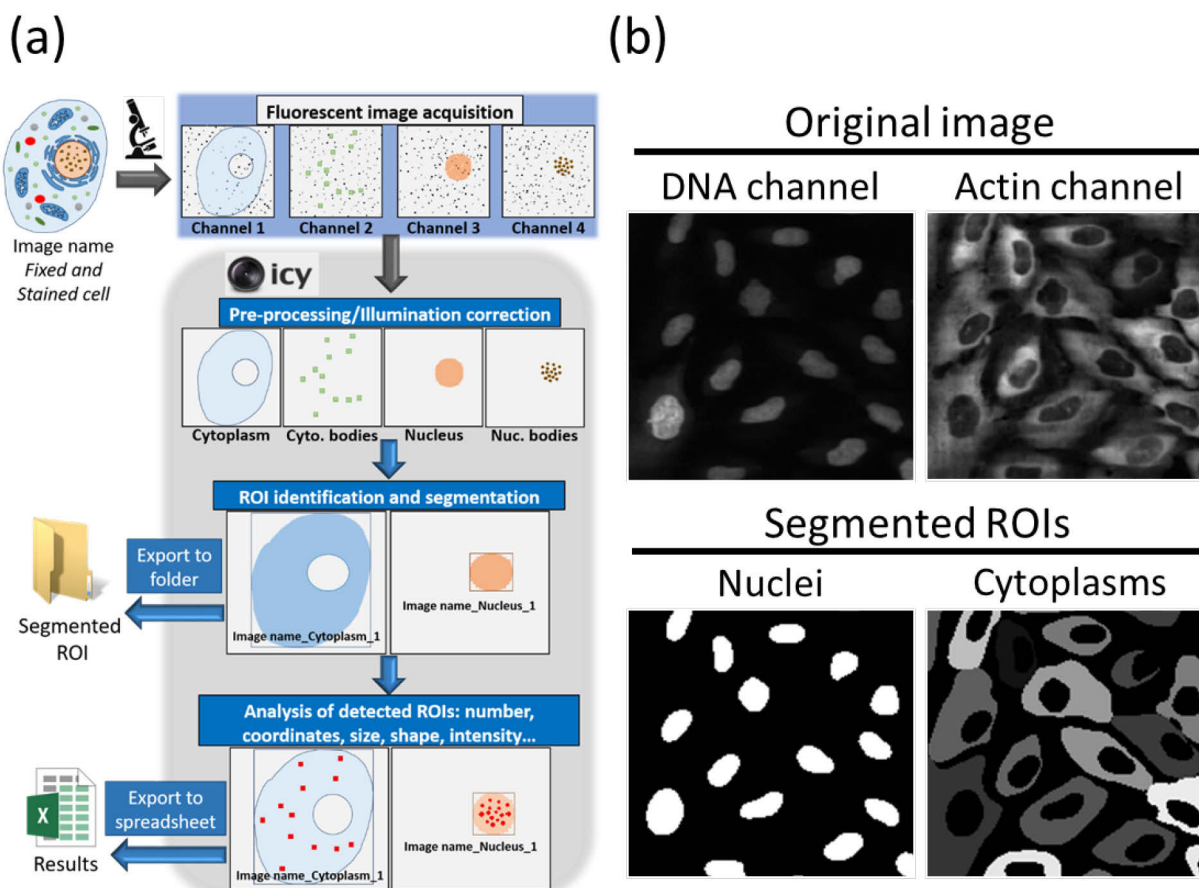
status. Among the whole kinetics, we observed that OS induces a significant increase of the 53BP1 foci number (**Figure 6.b**). At 2h, the median value increases from 2 to 22 foci in stressed cells. We also notice that a small proportion of non-stressed cells have more than 10 foci (14%) which could reflect the cancerous state of HeLa cells and potential defects in their DNA repair mechanisms. When comparing the different median values we observe a significant 11-fold increase at 2, 4 and 8h after the induction of OS. This effect is less important at 6h (6.5-fold) because of an initial high level of DSB in non-stressed cells. Even after 20h, a sustained increase remains (almost 6-fold). These data might reflect the good efficiency of the treatment to induce the OS and show that 53BP1 can be used as a good stress-marker for our early and late observations in microscopy. On another hand, at each time point of the kinetics we always observe a small proportion of stressed-cells with a low level of DDSBs. This suggests that cells are not homogeneously stressed, enforcing the importance of having a marker for our experiments. We performed a ROC curve for each time point of the kinetics to determine the best threshold discriminating efficiently stressed from non-stressed cells. Specificity and sensitivity of threshold-based test have been obtained from the ROC curves. We selected thresholds with optimal specificity instead of sensitivity in order to decrease the number of false positives. Thresholds retained for subsequent analysis are reported in the table of **Figure 6.b**.

Then we analyzed the number and size of coilin nuclear foci according to cell's stress status. We firstly noticed a significant increase of the foci number at 2, 4 and 6h after stress induction (**Figure 6.c**). The most sustained effect being observed at 2h, with a number of cells having more than 10 bodies increasing from 0 to 75% in non-stressed and stressed cells, respectively. Moreover, more than 25% of stressed cells have more than 20 coilin foci. This effect progressively decreases until 8h even if a small proportion of cells with more than 15 bodies remains detectable. At 20h, a small increase of the first and the third quartiles is observed but our data also show that 84% and 80% of cells present less than 8 foci in non-stressed and stressed cells, respectively. These data show that OS has also late effects on parameters controlling the nucleation of Cajal Bodies which might be different from early effects. Interestingly, careful analysis of the features of the coilin nuclear foci revealed that the observed increase of foci number associates with a decrease of their size (**Figure 6.d**). For instance, 2h after OS induction, the proportion of coilin foci with an area below  $0.2 \mu\text{m}^2$  increases from 26% to 64%. This effect decreases until 8h but, unlike Cajal Bodies number, no significant change is observed at 20h. This might reflect that early and late nucleoplasmic redistribution of Cajal Bodies are different events.

All these data strongly suggest that OS changes the nucleation power of Cajal Bodies, inducing a nucleoplasmic redistribution into numerous smaller nuclear foci. Since nucleation of Cajal Bodies is driven by their protein and RNA components, this suggests that OS effect might change the composition of Cajal Bodies. Within Cajal bodies the coilin interacts with several different protein partners such as components of the SMN complex and Sm proteins. These interactions often involved phosphodomains of coilin and one can imagine that modification of the structure of Cajal Bodies could be linked to a changes in the phosphorylation status of coilin. Moreover, recent works demonstrated that Cajal Bodies are not randomly localized within the nucleus, and can affect gene expression through proximal association with specific gene loci (Wang et al., 2016). Considering all these facts, it would not be surprising that changes of their structure would be accompanied with an effect on their functionality. From these results we can also ask if such Cajal Bodies remodeling is a passive consequence of OS or participates in the response by interacting with specific gene loci for example.

To test if a change in coilin expression could alter its localization and change the nucleation of Cajal Bodies, we overexpressed an exogenous GFP-coilin fusion protein. Nuclei were not clustered so we segmented them with the block **“Segmentation A: Non clustered objects”**. Then, to precisely quantify the number and size of coilin foci (red signal, channel 1) according to the level of GFP-coilin (green signal, channel 2) overexpression, we used the block **“Fluorescence Analysis: 2 Channels Spot Detector in the same compartment”**. The level of GFP-coilin was reflected by the mean intensity of the GFP signal in individual nucleus (**Figure 7.a**). (As we analyzed 8-bits images, values of GFP intensity scale range between 0 and 255). In cells with low and medium level of GFP-coilin overexpression (GFP intensity lower than 1), the number of Cajal Bodies per nucleus remains unchanged (**Figure 7.b**). However, in cells expressing higher level of GFP-coilin a slight increase of the number of Cajal Bodies per nucleus has been observed, with some nucleus displaying until 16 foci. We also noticed that the brightness (mean intensity) of Cajal Bodies considerably increases in parallel with the overexpression level leading to intensity saturation. The presence of smaller and fainter foci may be masked by such saturated regions. In overexpressed cells, the size of Cajal Bodies significantly increases (**Figure 7.c**), the median values rising from 0.5 to 1.5  $\mu\text{m}^2$ . Moreover, more than 25% of cells with a high GFP-coilin expression level present foci with an area larger than 2.5  $\mu\text{m}^2$  whereas it never exceeds 0.8  $\mu\text{m}^2$  in non-transfected cells. Coilin overexpression does not change Cajal Bodies number but significantly increases their size. These data might reflect that OS effect on Cajal Bodies structure is induced by a change of their composition rather than by an effect on coilin amount in the cells.

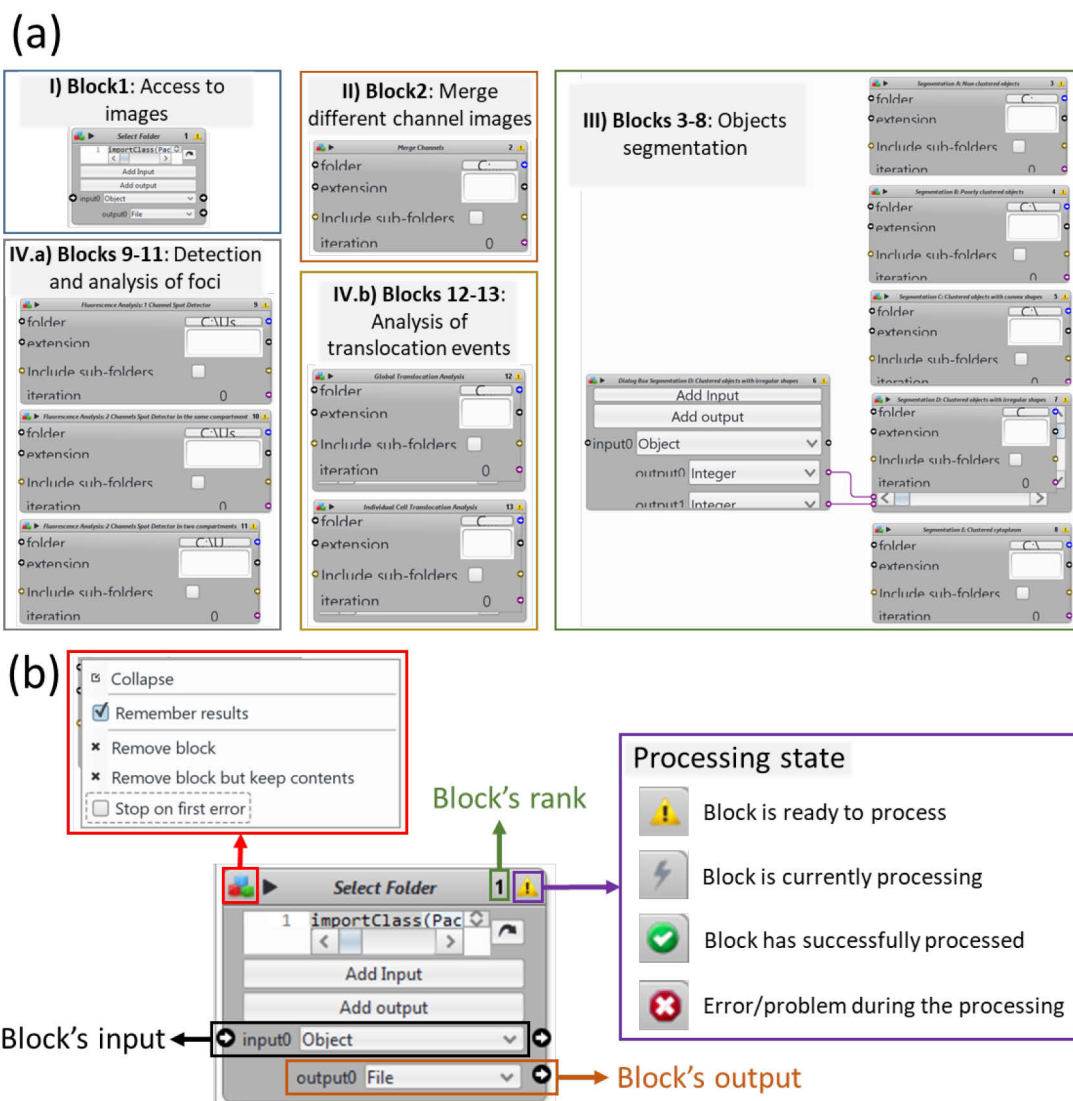
#### FIGURE AND TABLE LEGENDS:



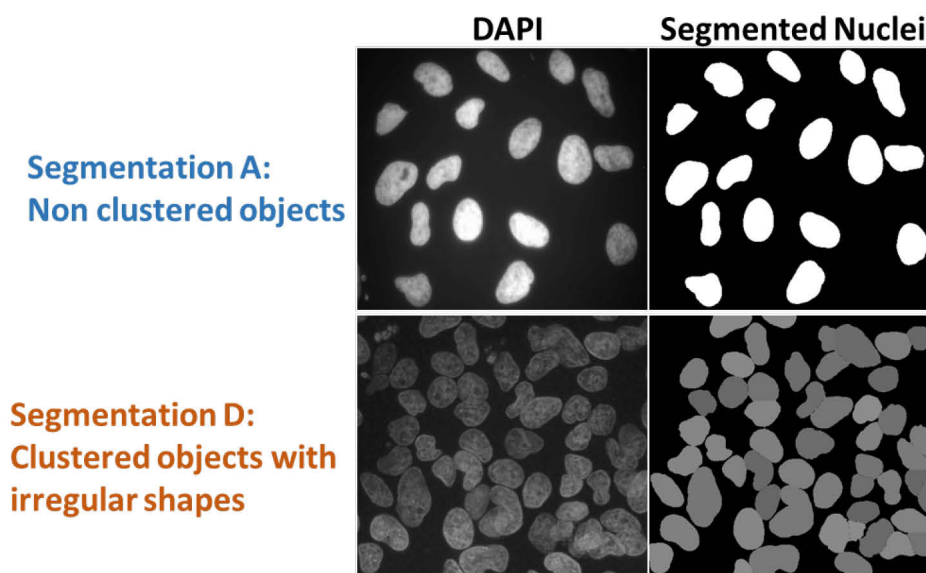
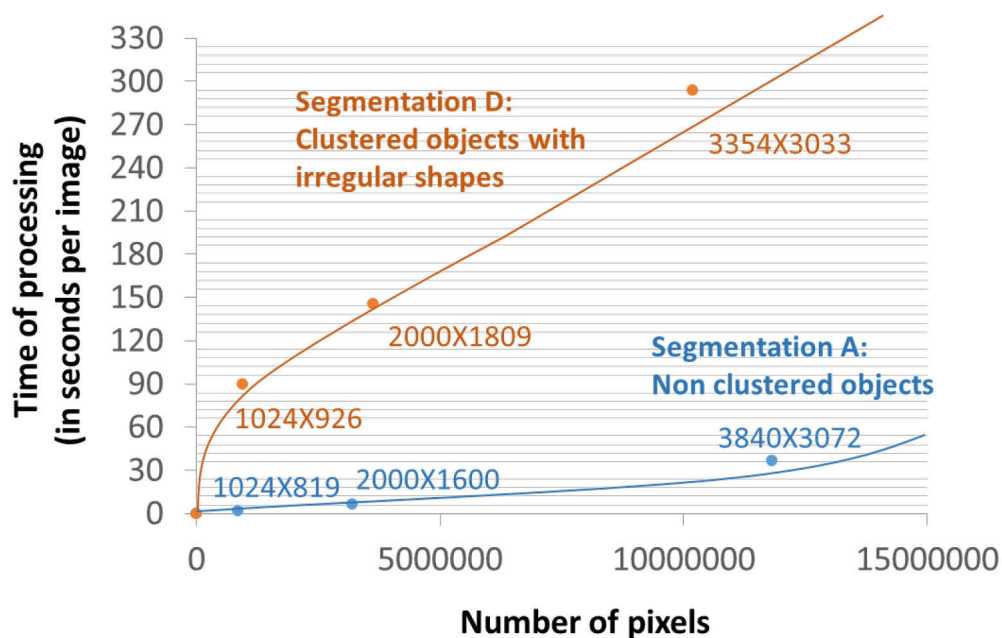
**Figure 1: Substructure Analyzer overview and features**

**(a) Substructure Analyzer pipeline.** The workflow is developed in Icy, an open community platform for bioimage informatics, and performs automatic analysis of multiple fluorescence microscopy images. Firstly, after the selection of the right folder, multi-channel images are automatically loaded within the protocol and pre-processed in order to improve, if necessary, the signal to noise ratio and to remove imaging artifacts. Then, image segmentation isolates regions of interest (ROIs) from the background. Several methods of segmentation are available depending on the level of clustering and the nature of the objects of interest. Segmented objects are saved with a specific descriptor (for example: Image name\_Nucleus\_1) in a specific folder and can be used/re-used for subsequent analysis. Fluorescent signals like spots are then analyzed within the ROIs and multiple features (location, size, shape, intensity, texture, spot number and size) are exported into an automatically created spreadsheet. All the measured features like the number of detected spots are reported to the descriptor of the corresponding ROI. **(b) Example of a correctly segmented image.** Segmentation is the most challenging step in bioimage analysis. Getting successfully segmented objects strongly depends on the quality of the foreground signal but also on object density (presence of touching or overlapping objects). Most of the time, ROIs correspond to cell nucleus or cytoplasm, but can also be any organelle (mitochondria, golgi apparatus...) or cellular substructure (nucleolus...) for which specific labelling is available.



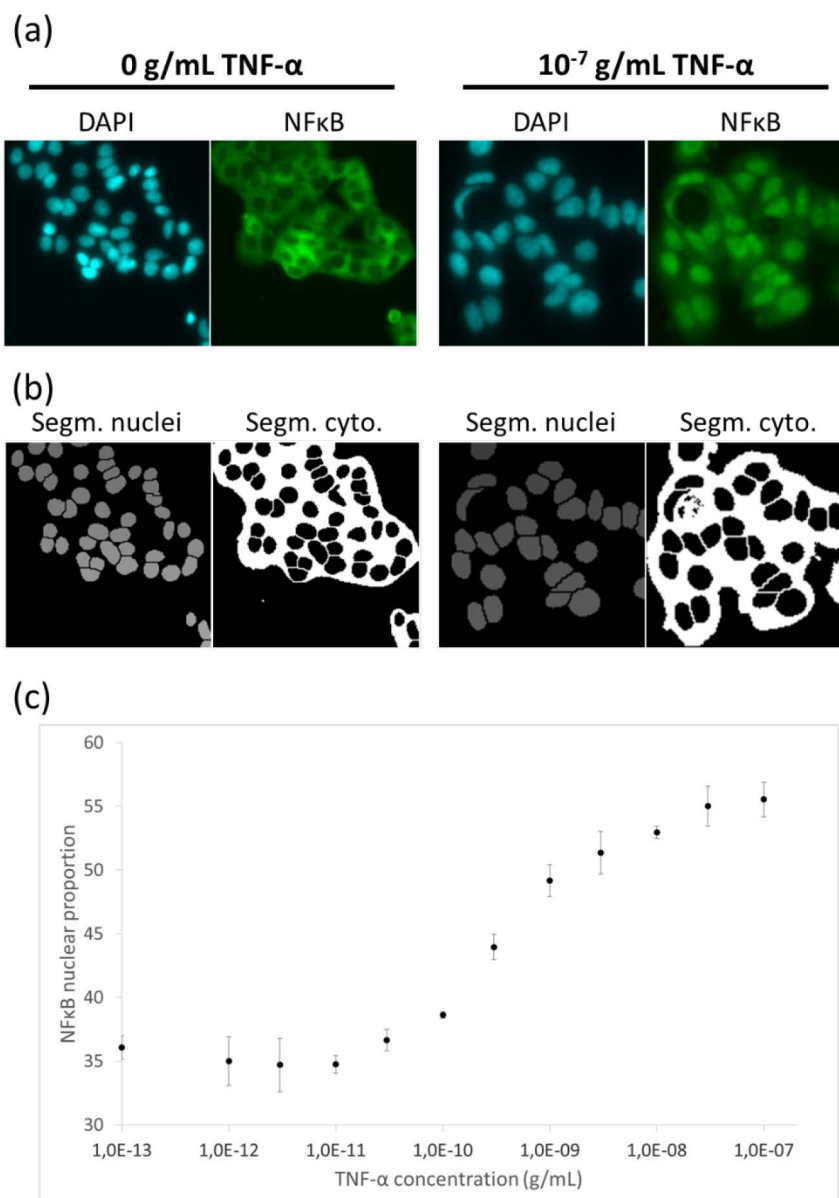


**Figure 2: Graphical interface of the Icy workflow** (a) The workflow is composed of 13 general blocks which can be grouped according to their general function: “Select Folder” (block 1) allow to access to images; “Merge Channels” (block 2) is used to merge different channel images. Blocks 3 to 8 are dedicated to object segmentation, each one being adapted for a specific context: “Segmentation A: Non clustered objects” (block 3), “Segmentation B: Poorly clustered objects” (block 4), “Segmentation C: Clustered objects with convex shapes” (block 5), “Segmentation D: Clustered objects with irregular shapes” (block 7, works in association with block 6), “Segmentation E: Clustered cytoplasm” (block 8). Blocks 9 to 11 can be used to count/analyze foci within subcellular compartments: “Fluorescence Analysis A: 1 Channel Spot Detector” (block 9), “Fluorescence Analysis B: 2 Channels Spot Detector in the same compartment” (block 10) and “Fluorescence Analysis C: 2 Channels Spot Detector in two compartment” (block 11). Blocks 12 and 13 are adapted for the analysis of molecules translocation: “Global Translocation Analysis” (block 12) and “Individual Cell Translocation Analysis” (block 13). (b) Global presentation of a block architecture.



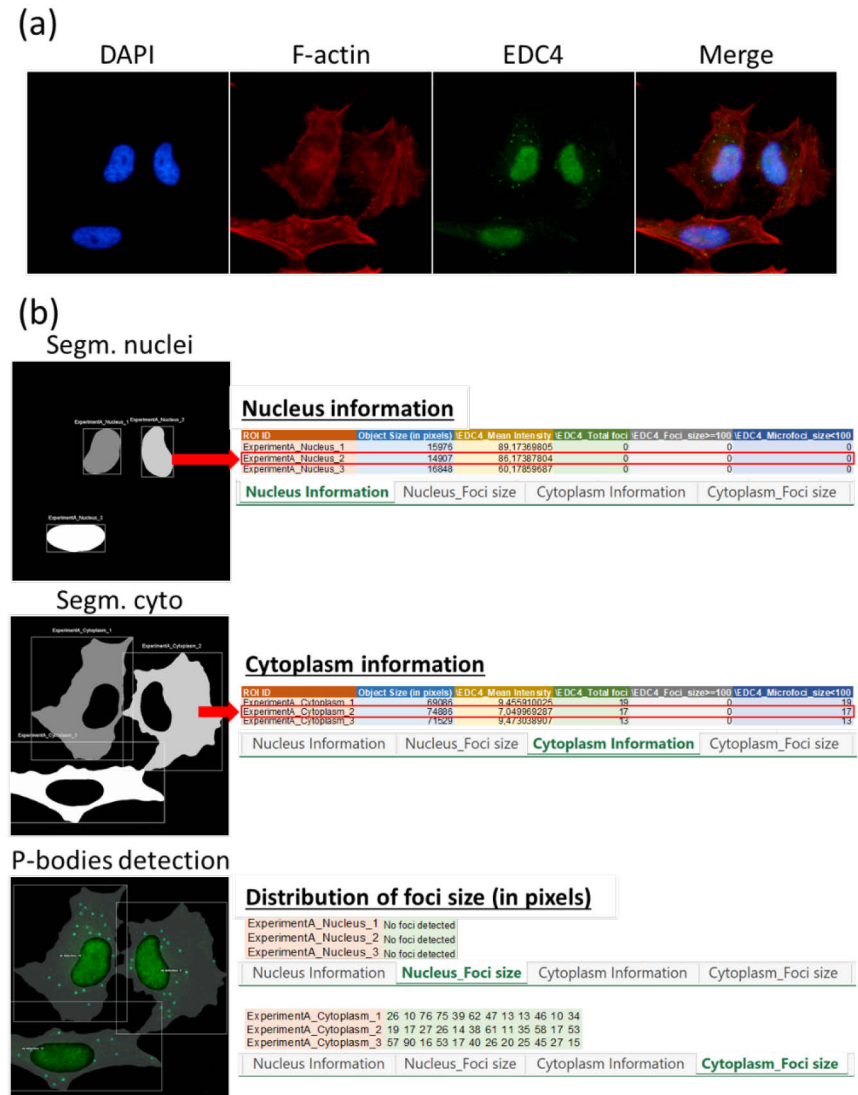
**Figure 3: Time needed for segmentation as a function of image resolution and object clustering level**

**(a)** Graphical illustration of the time needed to perform segmentation depending on image resolution. Two different segmentation blocks have been tested: segmentation A (for non clustered objects) and segmentation D (for clustered objects with irregular shapes). The time of processing (in seconds per image) is plotted as a function of image resolution (number of pixels). **(b)** Example of images analysed with either segmentation A or segmentation D blocks. DAPI staining and segmented nuclei are indicated.



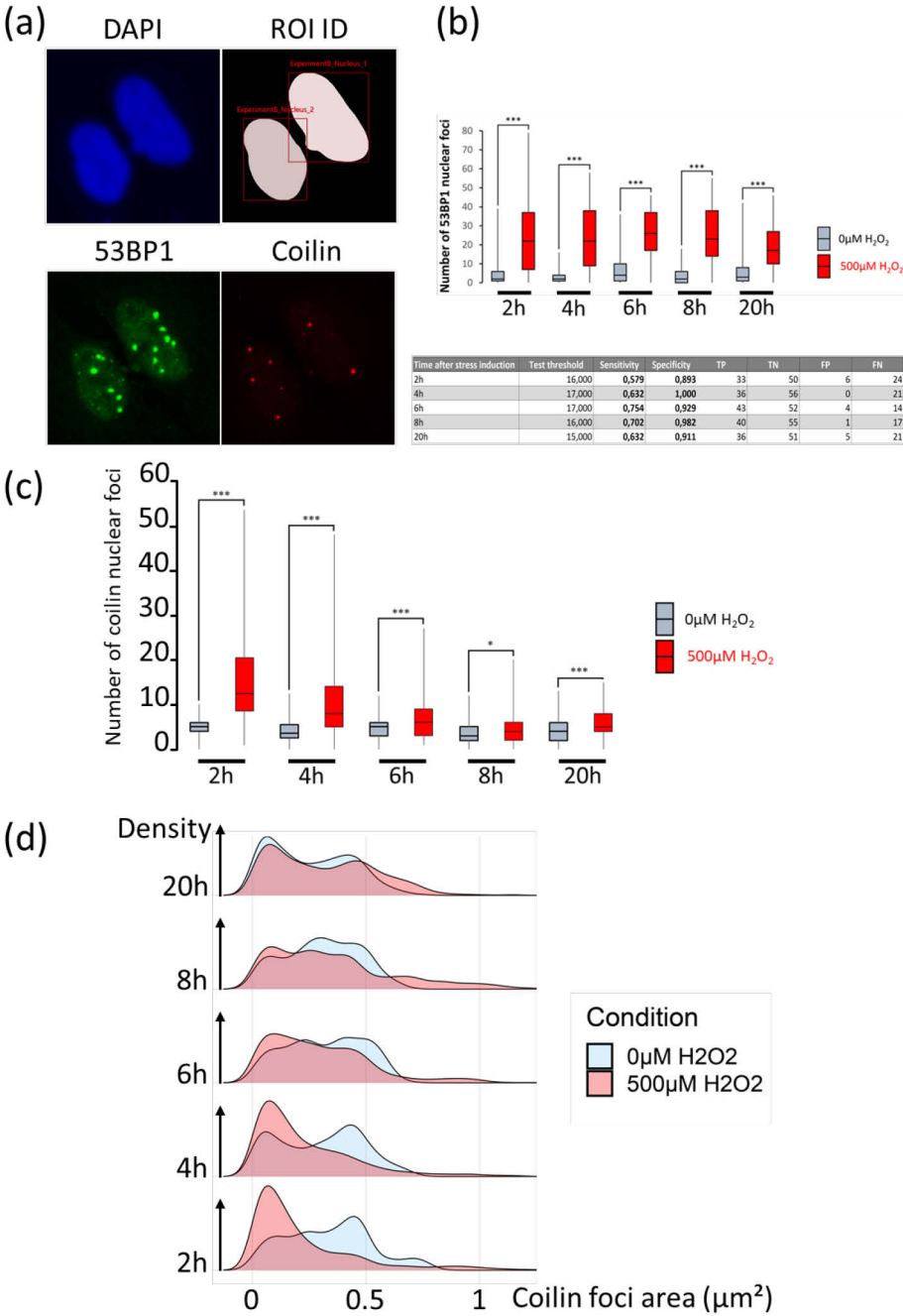
**Figure 4: Workflow analysis of a NFkB nuclear translocation assay in response to TNF $\alpha$**

**(a)** Example of images of human MCF7 cells (human breast adenocarcinoma cell line) treated with TNF $\alpha$  from the BBBC014 image set (Broad Bioimage Benchmark Collection) (Ljosa et al., Nature Methods, 2012). Localization of the transcription factor NFkB has been studied in cells treated with 12 increasing concentrations of TNF $\alpha$  (from 0 to 10<sup>-7</sup>g/ml). For each concentration, 4 replicas were done. Nuclei are localized with DAPI staining (blue channel) and NFkB with FITC staining (green channel). **(b)** Segmented nuclei (grey) and cytoplasm (white) are used to determine the total intensity of NFkB fluorescent signal in each compartment. **(c)** The dose-response curve obtained from our data analysis illustrates the increase of NFkB nuclear proportion as TNF $\alpha$  concentration increases. Image size is 1360 X 1024 pixels and are in 8-bit BMP format. For each concentration, the standard deviation between the 4 replicas is calculated and illustrated as error bars.



**Figure 5: Quantification of EDC4 foci (P-bodies) within the nucleus and the cytoplasm in individual cells**

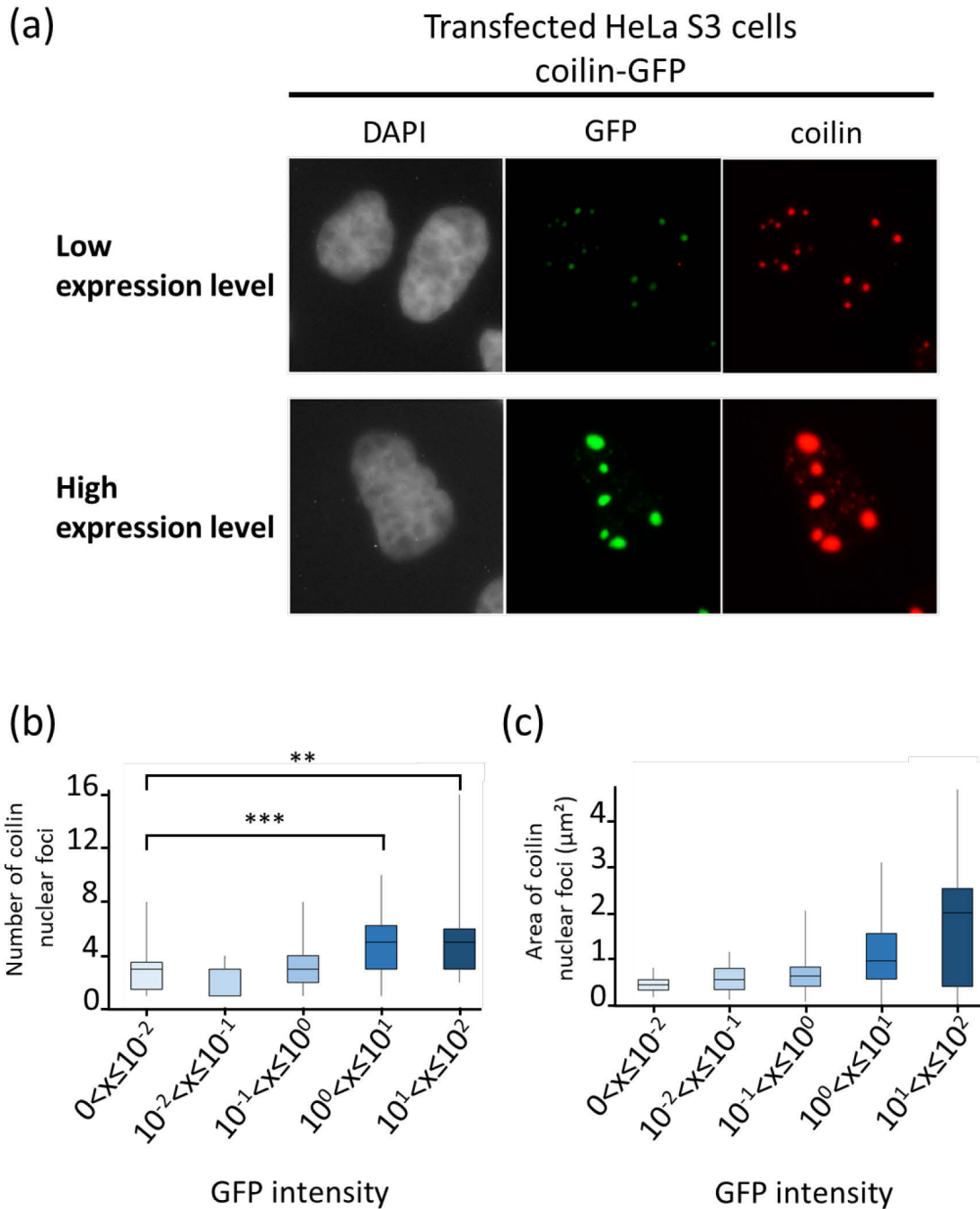
(a) Co-staining with DAPI (blue channel) and Phalloïdin (red channel) allows the segmentation and identification of cell nuclei and F-actin respectively. P-bodies which are detected with antibodies against EDC4 (green channel) are known to be exclusively cytoplasmic. (b) Foci are counted and several features (here their area in pixels) are exported to the *Results* spreadsheet structured by several sheets. Here, two sheets are dedicated to the analysis of nuclei and two to the analysis of cytoplasm. Each row embed the information of a specific ROI. In *Nucleus/Cytoplasm Information* sheets, the characteristics of segmented objects are exported (ROI ID, size and mean intensity) followed by the number of detected foci within each ROI. If necessary, a size threshold (100 pixels in this example) can be added, and the number of foci below and above this threshold is reported in the two last columns. In *Nucleus/Cytoplasm\_Distribution of foci size* sheets, the area (in pixels) of all detected foci is reported in the row of the corresponding ROI. In this example, we highlight the analysis of the cell numbered 2 (Nucleus\_2 and Cytoplasm\_2), where no nuclear EDC4 foci but 17 cytoplasmic ones, all below a size of 100 pixels, have been detected. The size in pixels of each of the 17 cytoplasmic foci is also given.



**Figure 6: Kinetics of coilin and 53BP1 nucleoplasmic distribution after oxidative stress**

The analysis of the number and size of foci was based on more than 110 individual cells for each time point from 3 independent experiments. Analyzed images are in TIF format, with a resolution of 3840X3072 pixels. **(a)** Immunofluorescence detection of coilin and 53BP1 foci in HeLa S3 cells throughout treatment kinetics with 500 µM H<sub>2</sub>O<sub>2</sub>. 53BP1 concentrates on DNA Double Strand Break sites and is used to assess

the efficiency of the treatment on each individual cells. Coilin is enriched in Cajal Bodies. Cells were fixed 2, 4, 6, 8 and 20h after oxidative stress induction and co-stained with anti-coilin (red channel) and anti-53BP1 (green channel) antibodies. **(b)** Quantification of 53BP1 nuclear foci in untreated or H<sub>2</sub>O<sub>2</sub> treated cells after different incubation times. For each time point of the kinetics a ROC (Receiver Operating Characteristic) test has been done to determine the best threshold to discriminate between stressed and unstressed cells. The parameters of each test are reported in the table (Sensitivity, specificity, TP: true positive, TN: true negative, FP: false positive, FN: false negative). From these data, a global threshold of 17 53BP1 foci was determined to discriminate stressed from non-stressed cells. **(c)** Oxidative stress increases the number of coilin nuclear foci. In cells treated with H<sub>2</sub>O<sub>2</sub>, a significant increase is detected and is maximal after 2 and 4h of treatment. T-test analysis demonstrates a significant difference between untreated and H<sub>2</sub>O<sub>2</sub> treated cells (\*, p<0.05; \*\*, p<0.01; \*\*\*, p<0.001). **(d)** The size of coilin nuclear foci (area in  $\mu\text{m}^2$ ) decreases in oxidative stressed cells. From 2h to 4h, the increase of the number of coilin foci coincides with a decrease of their size.



**Figure 7: Study of Cajal Bodies nucleation in cells overexpressing a GFP-coilin fusion protein**

**(a)** HeLa S3 cells were transiently transfected in biological triplicates with 500 ng of plasmid expressing a GFP-Coilin fusion protein using standard JetPEI procedure following the manufacturer's recommendations. After 48h, cells were fixed and stained with antibody against coilin. DAPI (blue channel) permits to segment nuclei. For these analyses more than 150 cells were analyzed. GFP signal reflects

ectopic coilin-GFP protein (green channel) localization whereas coilin signal (red channel) corresponds to both endogenous and exogenous protein localization. For each nucleus, the mean intensity of GFP signal reflects the level of coilin expression. According to this parameter, the features of Coilin foci have been analyzed using the block **“Fluorescence Analysis B: 2 Channels Spot Detector in the same compartment”**. **(b)** Relationship between number of coilin foci and GFP intensity score. Results are shown as box plots, where the central mark is the median, the lower and upper edge of the box are the 25<sup>th</sup> and 75<sup>th</sup> percentiles. T-test analysis demonstrates a significant increase of the number of foci between the lower GFP intensity (<0,01) and GFP intensity >1 (\*,  $p<0.05$ ; \*\*,  $p<0.01$ ; \*\*\*,  $p<0.001$ ). **(c)** Relationship between size of coilin foci and the GFP intensity score.

## DISCUSSION:

Increasing number of free software tools are available for the analysis of fluorescence cell images. Users have to correctly choose the adequate software according to the complexity of their problematic, to their knowledge in image processing and to the time they want to spend in their analysis. Icy, CellProfiler or ImageJ/Fiji are powerful tools combining both usability and functionality. Icy is a stand-alone tool that presents a clear Graphical User Interface (GUI), and notably its “Protocols” point and click interface through which workflows can be easily designed or manipulated. The functionality of this software is enhanced by supporting and utilizing ImageJ plug-ins and a lot of documentation is available thanks to its large community of users. We used this strong combination of usability and functionality of the Icy software to develop the “Substructure Analyzer” workflow. Its main objective is to propose an automated solution to perform full analysis of cell fluorescent signal from multiple images. Analysis encompasses image pre-processing, object segmentation and fluorescent signal analysis.

- Significance of the method with respect to existing/alternative methods

Several protocols are gratefully shared on the Icy website and propose to use Icy functionalities to detect and analyze fluorescent signals like cellular foci. However, some of these protocols only process one image per run, do not export the results in a specific file and/or analyze a single channel. Others do not contain preliminary segmentation to determine Regions of Interest (ROIs), or propose simple algorithm like “HK-means” that are not suited for the segmentation of highly clustered objects. Our workflow automates all the steps of image analysis from image pre-processing to the segmentation of objects and the detection/analysis of fluorescent signals. The protocol also proposes simple or more complex methods for object segmentation and export of the data (names of the segmented objects, foci analysis...) is realized in optimized outputs adapted to the problematic. CellProfiler also proposes powerful pipelines composed by modules that encapsulate functionalities. The available pipelines are well adapted to specific problematics. Icy workflows are network of different boxes each of them performing specific task. Through the “Protocols” interface, boxes composing the network are intuitive and easy to manipulate. “Substructure Analyzer” is adapted to several contexts like the simple merging of channels, the quantification of fluorescent signals distribution between two subcellular compartments, the analysis of foci in one or several channels from one or two cellular compartments in individual cells. Several displays allow to visualize the intermediate results during each run in order to control the processing. Moreover, Icy presents icon bars that group the methods by topic and from which functionalities found in the



workflow can be separately manipulated to manually test specific parameters on some images before using them on a whole image set.

- Critical steps in the protocol

Like in the bio-image analysis field, the most critical step of this protocol is the object segmentation. The protocol proposes the segmentation of primary objects which are most of the time cell nuclei and also contains another block which task is to segment a second type of objects within the cell. Nuclei segmentation can be difficult when the objects are highly clustered so that they touch or overlap together. Different alternative for the segmentation of primary objects are contained in the protocol, according to the shape of the objects and the level of clustering even if to date no universal algorithm success in the full segmentation of highly clustered objects. The segmentation process is mainly limited by the determination of correct parameters that could successfully segment a part of the objects but lead to under- or over-segmentation of another part. The user would have to perform different runs with different parameters to get a final correct segmentation, especially for clustered objects presenting heterogeneous and non-convex shapes.

- Limitations of the method

This protocol is mostly limited by the segmentation step which could be highly time consuming and need powerful computer configuration for the most complex cases. More specifically, the more the images are complex (clustered objects to segment, high number of pixels in the images...) and/or numerous, the more the user will require Random-access memory (RAM). Globally, for any problematic, user will need to run the protocol on a 64 bits Java Runtime Environment (freely available on Java website) and to use a 64 bits operating system (OS) to increase the available memory used by Icy (Memory is limited to 1300MB for 32 bits JRE which could). Because of the script part, the protocol does not work correctly on Mac computer. Moreover, the number of parameters increase with the complexity of the problematic and will require more knowledge in image analysis. Concerning the functionality criteria, this protocol has not yet been adapted for the analysis of stacked images (time stacks, z stacks) even if Icy efficiently performs analysis on this type of data.

- Future applications or directions of the method

Future updates will be realized notably to improve the segmentation of complex clustered objects. Additional blocks will be developed to adapt the segmentation to clustered cellular structures with irregular non convex shapes. We will also adapt the workflow for the analysis of time and z-stacks (for live imaging and 3D data).

## **ACKNOWLEDGMENTS:** ([Instructions](#))

G.H. was supported by a graduate fellowship from the Ministère Délégué à la Recherche et aux Technologies. L.H. was supported by a graduate fellowship from the Institut de Cancérologie de Lorraine

(ICL). This work was funded by CNRS and Université de Lorraine (UMR 7365). We thank Quentin Thuillier for his help during the video processing.

# DISCLOSURES: ([Instructions](#))

The authors have nothing to disclose.

# REFERENCES: ([Instructions](#))

1. de Chaumont, F. *et al.* Icy: an open bioimage informatics platform for extended reproducible research. *Nature Methods*. **9** (7), 690–696, doi: 10.1038/nmeth.2075 (2012).
2. Meijering, E., Carpenter, A.E., Peng, H., Hamprecht, F.A., Olivo-Marin, J.-C. Imagining the future of bioimage analysis. *Nature Biotechnology*. **34** (12), 1250–1255, doi: 10.1038/nbt.3722 (2016).
3. Wiesmann, V., Franz, D., Held, C., Münzenmayer, C., Palmisano, R., Wittenberg, T. Review of free software tools for image analysis of fluorescence cell micrographs. *Journal of Microscopy*. **257** (1), 39–53, doi: 10.1111/jmi.12184 (2015).
4. Meyer, F., Beucher, S. Morphological segmentation. *Journal of Visual Communication and Image Representation*. **1** (1), 21–46, doi: 10.1016/1047-3203(90)90014-M (1990).
5. Schieber, M., Chandel, N.S. ROS Function in Redox Signaling and Oxidative Stress. *Current biology : CB*. **24** (10), R453–R462, doi: 10.1016/j.cub.2014.03.034 (2014).
6. D’Autr aux, B., Toledano, M.B. ROS as signalling molecules: mechanisms that generate specificity in ROS homeostasis. *Nature Reviews. Molecular Cell Biology*. **8** (10), 813–824, doi: 10.1038/nrm2256 (2007).
7. Davalli, P., Mitic, T., Caporali, A., Lauriola, A., D’Arca, D. ROS, Cell Senescence, and Novel Molecular Mechanisms in Aging and Age-Related Diseases. *Oxidative Medicine and Cellular Longevity*. **2016**, doi: 10.1155/2016/3565127 (2016).
8. Disher, K., Skandalis, A. Evidence of the modulation of mRNA splicing fidelity in humans by oxidative stress and p53. *Genome*. **50** (10), 946–953, doi: 10.1139/g07-074 (2007).
9. Takeo, K. *et al.* Oxidative stress-induced alternative splicing of transformer 2  (SFRS10) and CD44 pre-mRNAs in gastric epithelial cells. *American Journal of Physiology - Cell Physiology*. **297** (2), C330–C338, doi: 10.1152/ajpcell.00009.2009 (2009).
10. Seo, J., Singh, N.N., Ottesen, E.W., Sivanesan, S., Shishimorova, M., Singh, R.N. Oxidative Stress Triggers Body-Wide Skipping of Multiple Exons of the Spinal Muscular Atrophy Gene. *PLOS ONE*. **11** (4), e0154390, doi: 10.1371/journal.pone.0154390 (2016).
11. Will, C.L., Luhrmann, R. Spliceosome Structure and Function. *Cold Spring Harbor Perspectives in Biology*. **3** (7), a003707–a003707, doi: 10.1101/cshperspect.a003707 (2011).
12. Ljosa, V., Sokolnicki, K.L., Carpenter, A.E. Annotated high-throughput microscopy image sets for validation. *Nature Methods*. **9** (7), 637–637, doi: 10.1038/nmeth.2083 (2012).
13. Wang, Q. *et al.* Cajal bodies are linked to genome conformation. *Nature Communications*. **7**, doi: 10.1038/ncomms10966 (2016).



# Transcriptome Analysis Console (TAC) 4.0.2

## USER GUIDE

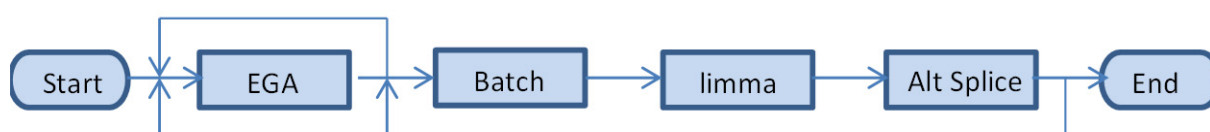
Publication Number 703150

Revision 8



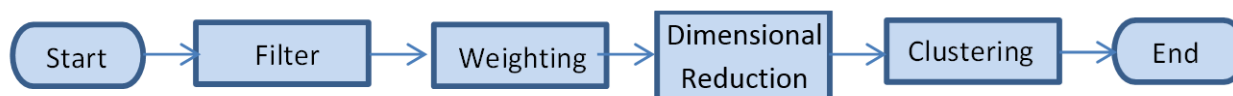
# Algorithms

This appendix describes the algorithms provided by TAC 4.0.2. Most are implemented with a call to an R or Bioconductor package. While the default analysis flow involves limma and alternative splicing analysis for exon-level microarrays, several additional analysis modules enable additional ways to examine the inter-sample relationships and perform batch correction.



The preceding flow chart illustrates the overall algorithmic workflow in TAC 4.0.2 following the pre-processing of the CEL files to CHP files. The first optional step is the exploratory grouping analysis (EGA), which can be iterated. Next, batch effects module and limma differential expression analysis are performed. Finally, in the case of microarrays that include exon-level data, the alternative splicing analysis is run. After examining the results, the user can then go back and iterate the previous steps.

## Exploratory Grouping Analysis (EGA) module



The Exploratory Grouping Analysis (EGA) module provides a variety of tools that enable the user to examine the relationships between the samples included in the experiment. Clearly, the experimental design suggests certain hypothesized relationships among the samples, but assay conditions can create outliers and batch effects. Furthermore, clinical samples can offer surprising differences and sub-clusters. The knowledge of the sample relationships can guide subsequent analysis; for example, outliers can be removed, batch effects corrected, and unexpected sub-clusters can be separated for analytical inspection. EGA supports a workflow with the following stages: gene filtering, gene weighting, dimensional reduction, and clustering. The results are presented as a 3-dimensional visualization and a set of cluster assignments. The Euclidean distance function is used throughout.

The gene filtering has several available controls. First, the user can identify a list of genes for use in the analysis. This approach is especially helpful if attention is focused on a certain pathway or genes of special interest. Second, the user can specify the number of maximally variant genes to be considered. Either way, the maximum



number of genes the EGA module can process is 20,000. The genes with the lowest variance tend to contribute noise to the computation that tends to damp the effect of the more strongly varying genes, so the maximum was chosen as a compromise that removes the noise-generating genes in the low variance tail.

Gene weighting involves the assignment of weight proportional to the square of the variance to each gene. The advantage of using such a weighting is that it focuses on the most variant genes without having to identify them directly or to specify their number.

Dimensional reduction involves projecting objects in a high-dimensional space into a reduced dimensional space that is easier for human perception. In this case, the intensities associated with the many genes in each sample represent the high-dimensional space. As the number of genes is typically thousands, humans could only look at a few genetic dimensions at a time. Dimensional reduction is a mechanism that merges the high-dimensional components (in this case, the gene intensities) into a smaller number of aggregate components (in this case three) that represent the highest variation present in the samples. EGA includes two algorithms for dimensional reduction: Principal component analysis (PCA) and t-distributed stochastic neighbor embedding (tSNE).

PCA computes a set of linearly uncorrelated variables each of which is an orthogonal linear combination of the gene components. These variables are sorted so as to have decreasing variance. Retaining the top few components yields a reduced dimensional view of the samples that represents most of the variance among the samples. PCA is simple to use and requires no control parameters. TAC 4.0.2 calls the R `prcomp` function to compute the PCA.

The tSNE algorithm uses a stochastic non-linear dimensional reduction approach that offers more flexibility than PCA at the cost of some additional complexity. The perplexity parameter can roughly be considered to adjust the effective number of nearest neighbors and can range from 1 to nearly the number of samples / 3. Values of perplexity near the ends of the range can generate deceptive results, so values in the central portion of the range generally lead to the best results. The results for several perplexity values should be compared to find the preferred clustering. The distances between samples in the reduced coordinates do not have meaning, so the scatter plot should be viewed more qualitatively. The seed for the random number generator is fixed so as to generate consistent results for a given data set and perplexity value. TAC 4.0.2 calls the `Rtsne` function in the R package `Rtsne`.

Starting from the dimensionally reduced coordinates, clustering assigns each sample to a group or cluster. EGA offers three algorithms for clustering: k-means, affinity propagation clustering (APC), and density-based spatial clustering of applications with noise (DBSCAN). The distance matrix, which provides the distance between any two samples, is also returned.

The well-known k-means clustering algorithm requires the input of the number of clusters. K-means randomly assigns each sample to one of the clusters, finds the center for each cluster, and then reassigns samples to the nearest cluster center. New cluster centers are determined, and the process iterates to convergence. Depending on the random assignment, the cluster assignment can converge to different results. The k-means algorithm includes a parameter specifying the number of random starts. The final result of these multiple random starts is the result that has the minimum sum of samples to their respective centers. The random number generator's seed is fixed so as to produce consistent results. TAC 4.0.2 calls the R `kmeans` function to compute the k-means clustering.

APC derives from the idea of passing messages between the data points. Unlike k-means, APC determines the appropriate number of clusters. The APC message passing



is used to identify exemplars, which are data points that represent the nuclei of clusters, and to associate other data points to each of the clusters. The affinity parameter ranges from 0 to 1 and controls the degree of aggregation. Smaller values yield fewer clusters. TAC 4.0.2 calls apcluster function in the R apcluster package.

DBSCAN is a density-based method of clustering that uses the number of other samples in the neighborhood of each sample to determine the sample's cluster label. The number of clusters is determined by the algorithm, and the clusters can have an arbitrary shape. There is a control parameter that sets the size of the neighborhood based on the percentile of all the distances between the samples. DBSCAN is different from k-means and affinity propagation clustering in that it may not be able to assign a group to all samples in which case some samples are unclustered. TAC 4.0.2 calls the dbscan function in the R dbscan package. The minPts parameter is fixed at 4.

## EGA parameters

**Figure 213** EGA parameter options

**Exploratory Grouping Analysis**

Variance Filter Number: 20000

Weighted by Variances: False

Dimension Reduction: tSNE

tSNE Perplexity: 5

Clustering: Affinity

Affinity: [0,1] (0 for fewer cluster, 1 for more clusters): 0.25

k-means: number of clusters: 4

k-means: number of random starts: 5

dbscan percentile of intersample distances to choose as epsilon neighborhood: 0.15

### EGA parameter option definitions

**Weighted by Variances** - This mode enhances the clustering of the most strongly varying probesets by transforming the expression levels as follows: subtract the mean expression level for the probeset and multiply the result by the variance of the probeset.

**Dimensional Reduction** - A procedure that converts a high dimensional space, in this case the expression of the thousands of probesets in each sample, to a lower dimensional space by creating weighted combinations of the probeset expression levels. TAC4 supports the linear PCA and the nonlinear tSNE algorithms.

**tSNE Perplexity** - Control parameter for tSNE that roughly represents the effective number of near neighbors.

**tSNE Perplexity Mode** - When mode is set to auto, a reasonable value for perplexity parameter is generated for the user. In manual mode, the user inputs the perplexity value for greater control.

**Similarity** - Should not appear in the interface.

**Clustering** - Identifies the algorithm used to label the samples. The choices include affinity propagation clustering, k-means, and DBSCAN.



**Affinity** (Only active when clustering chooses affinity propagation) - A control parameter for the affinity propagation algorithm. The number of clusters is determined by this algorithm, and this parameter controls the relative propensity to generate a larger number of smaller clusters with a range of values [0,1]. Smaller values of affinity have fewer clusters and larger values have more clusters.

**k-means number of clusters (only active when clustering chooses k-means)** - The number of clusters.

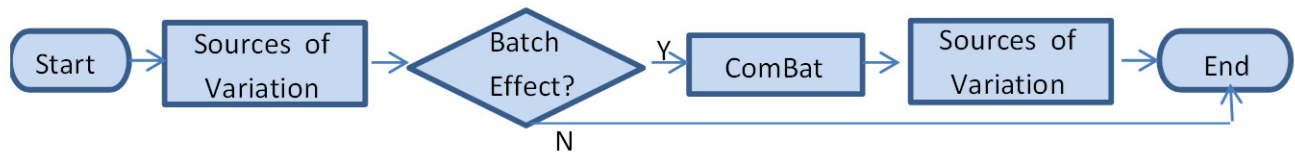
**k-means: number of random starts (only active when clustering chooses k-means)** - K-means starts with a random assignment of the samples to one of the clusters. It then iteratively reassigns clusters until convergence. Different random starting cluster assignments can yield different final assignments. This option will run k-means the indicated number of times and merge the results.

**DBSCAN** - Size of epsilon neighborhood - DBSCAN is a density-based clustering algorithm, which does not require specification of the number of clusters and the clusters may have arbitrary shapes. The epsilon neighborhood controls how the samples are added to a cluster. Smaller values lead to larger number of clusters and larger values fewer clusters.





## Batch Effects module



The TAC 4.0.2 batch effects module offers two functions: source of variation analysis and batch effect removal. This analysis is computed at the gene-level and the exon-level, so in the following, component refers to gene or PSR/JUC as appropriate. First, the variance is computed for the components, and the 1000 components with the highest variance are retained. Next, for each attribute in the sample attribute table (including columns of type "None"), percentage of variance explained is determined as follows:

$$TSS = \text{total sum of squares} = \sum_{\text{all samples}} (e_i - \langle e_i \rangle)^2$$

$$RSS = \text{residual sum of squares} = \sum_{\{\text{all attr levels}\}} \sum_{\{\text{samples in level}\}} (e_i - \langle e_i \rangle)^2$$

$$\text{percentage of variance explained} = \left\langle \frac{TSS - RSS}{TSS} \right\rangle$$

For each retained component, compute the total sum of squares (TSS) as the sum over samples of squares of the expression level for a sample minus the average expression level for the component (shown using angle brackets). The residual sum of squares (RSS) represents a similar sum of squares but the average expression level subtracted is the average over samples in the same attribute level. Only attribute levels with at least two samples contribute to this sum. Finally, the percentage of variance explained is the average across all the retained components of the ratio of the difference of the TSS and the RSS to the TSS.

This computation looks at each attribute independently, so the sum of the percentages is unconstrained. Generally, one expects the percentage of variance explained to be larger for the experimental attributes under investigation, while other secondary attributes such as date or operator should explain relatively smaller percentages of the variance. Of course, attributes that are unique for each sample will trivially explain 100% of the variance. Secondary attributes that explain relatively high levels of variance are candidate batch effects (assuming that they do not confound the experimental attributes of interest). The percentage of variance explained is computed both before and after the ComBat\_Affy\* correction so as to provide an indication as to the success of the batch effect removal.

TAC 4.0.2 batch effect removal operates on known batch effects. The batch effect is discovered either through inspection using the EGA or from the PCA plot and sources



of variation analysis generated during standard analysis. The user indicates the existence of a batch effect by marking an attribute column in the sample attribute table as a batch effect. TAC 4.0.2 provides two separate mechanisms for batch effect removal that are used at different stages of the analysis. `Combat_Affy` batch effect removal generates the transformed expression levels, group averages, group standard deviations, fold changes, and splice indexes. Limma handles the batch effect during the differential expression computation so as to correctly account for the ANOVA degrees of freedom. Limma is discussed in the next section.

`Combat_Affy` models the batches using their mean values and standard deviation per gene. An empirical Bayesian correction uses pooled information across genes so as to improve the model parameter estimates, which is especially useful for the cases when the number of samples in the comparison groups and batches is small. Finally, the modeled batch effect is removed from the expression signals to yield the corrected result. TAC 4.0.2 calls a C++ accelerated port of the SVA `Combat_Affy` code. The call uses the parametric empirical prior option.

Warning. Users should use some caution when setting up the batch effect. Batch effect attribute levels should not be correlated with the attribute levels representing the biological information of interest so as to retain the ability to remove the batch effects. During experimental design, one should attempt to randomize the biological sample types across potential batch effect attributes.

\* `ComBat_Affy`, `ComBat_Affy_Gene`, and `ComBat_Affy_Exon` are modified versions of the `ComBat` function from the Surrogate Variable Analysis package in Bioconductor [11 and 12].

The modified version includes the following changes:

- The original R code was translated to C++.
- Three versions have been created:
  - `ComBat_Affy` loads all of the microarray data into memory for the processing (like the original code).
  - `ComBat_Affy_Gene` buffers the microarray data in files for gene-level analysis so that the processing can run in a smaller memory footprint.
  - `ComBat_Affy_Exon` buffers the microarray data in files for exon-level analysis so that the processing can run in a smaller memory footprint.
- The batch and design indicators are directly loaded through the interface rather than making the batch and design factors and indicators internally.
- If the variance of a gene or exon is  $\leq 1e-8$ , the standardized mean is returned.
- If the hyperparameters cannot be computed, switch to `mean_only` mode.

## Limma Differential Expression analysis

TAC 4.0.2 calls the limma differential expression portion of the Bioconductor package. Limma provides an extremely flexible framework to describe the experimental structure and desired differential comparisons through the use of the design and contrast matrices. Most TAC users, however, may not have experience with this statistical machinery, so TAC 4.0.2 implements a simpler interface that automatically populates the design and contrast matrices based on the sample attributes table. Moreover, there is a natural language interface called the comparison wizard that enables the user to customize the desired comparisons.

Consider the following sample attribute table (Figure 214) that has been artificially generated to show its attribute column types.

**Figure 214** Sample Attribute Table - Example

| File Name (5)                        | Condition #1<br>Type: Comparison | Condition #2<br>Type: Repeated Measure | Condition #3<br>Type: Batch Effect | Condition #4<br>Type: Random Factor (Advanced) | Condition #5<br>Type: Real Covariate (Advanced) |
|--------------------------------------|----------------------------------|----------------------------------------|------------------------------------|------------------------------------------------|-------------------------------------------------|
| 1_0HTA-2_0.sst-rna-combined-dabg.chp | 0                                | 0                                      | 0                                  | 0                                              | 0.13                                            |
| 2_0HTA-2_0.sst-rna-combined-dabg.chp | 1                                | 1                                      | 0                                  | 0                                              | 0.51                                            |
| 3_0HTA-2_0.sst-rna-combined-dabg.chp | 0                                | 2                                      | 1                                  | 0                                              | 0.76                                            |
| 4_0HTA-2_0.sst-rna-combined-dabg.chp | 1                                | 0                                      | 1                                  | 0                                              | 0.28                                            |
| 5_0HTA-2_0.sst-rna-combined-dabg.chp | 0                                | 1                                      | 1                                  | 1                                              | 0.46                                            |
| 6_0HTA-2_0.sst-rna-combined-dabg.chp | 1                                | 2                                      | 1                                  | 1                                              | 0.69                                            |

Each attribute column in the sample attribute table can be set to one of six types: comparison, repeated measure, batch effect, random factor, real covariate, and none. The normalization group attribute type is separate and does not directly affect differential expression analysis. It provides a mechanism to preprocess the CEL files in subgroups rather than as a single group; the subgrouping allows for separate normalization when the samples differ so much that normalizing to a common distribution is not appropriate.

The comparison type specifies fixed attribute levels among which the differential expression can be computed. Moreover, interactions between attributes can be computed. At least one attribute column must be designated as a comparison type. The comparison attributes levels represent the biological factors of interest that the experiment was designed to study. All other attribute types represent various nuisance effects that the analysis will remove so as to enhance the effect of the comparison type attributes.

The repeated measure type indicates that the attribute marks experimental units (e.g. patients) that have multiple conditions associated with them. Differential expression between the conditions on the experimental units is the point of interest; the repeated measure type indicates that the specific effects of the experimental units should be removed.

The batch effect type indicates that the attribute levels represent fixed substrata in the data that represent nuisance factors in the differential expression analysis. Examples of

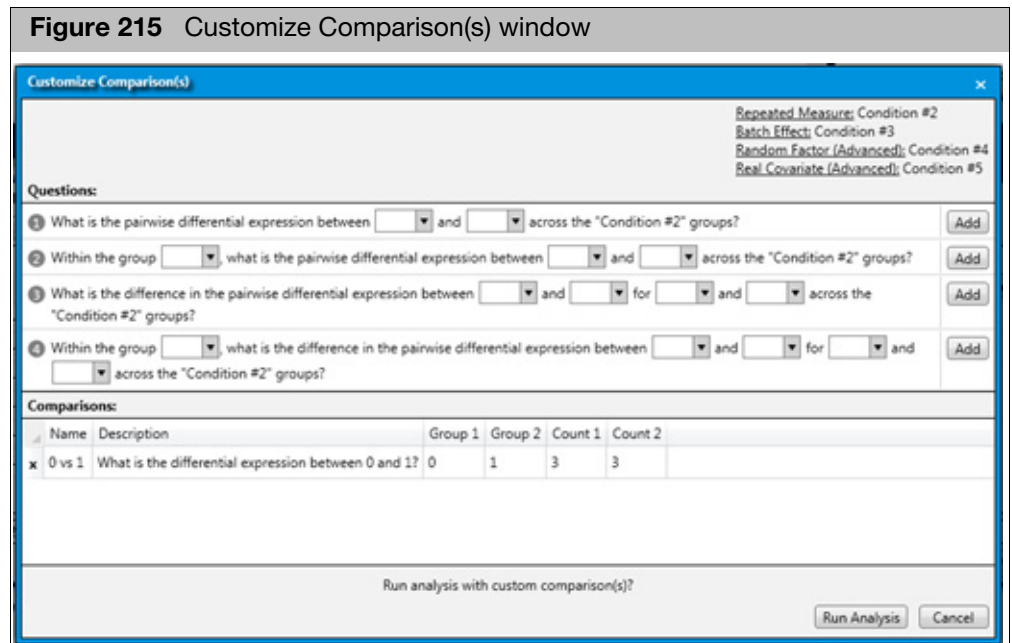
batch effects include assay date, scanner, operator, and so forth. Inclusion of an attribute as a batch effect has the effect of removing it. As in the case of `Combat_Affy`, batch effect attribute levels should not be well-correlated with the comparison attribute levels.

The random factor type represents random factors as supported by limma; that is, the attribute is a special kind of batch effect and the variance components of the random factor are not available. The application of random effects is somewhat subtle and generally a tool for more advanced statistical analysis. A common usage in limma involves comparisons both within and between experimental subjects in which the subject becomes the random factor. Furthermore, the correlation computation for a random factor is expensive, so the use of random factors can significantly extend analysis time.

The real covariate represents a mechanism to incorporate a continuously varying attribute into the analysis. Examples include age, tumor size, degree of contamination, and so forth. Limma will remove the effect of this real covariate on the differential expression results.

The fixed attribute types (comparison, repeated measure, and batch effects) are merged into a compound type. If each of the levels of the resulting type have a single sample, the variance cannot be computed, so all of the resulting comparison p-values for the contrasts will be 1.

The contrast matrix is automatically populated all reasonable comparisons and interactions based on the comparison type attributes. For simpler experiments, this default contrast set should be adequate. For more complex experimental designs, however, it can be helpful to use the comparison wizard to customize the analysis. Generally, this customization can reduce the number of contrasts to simplify the results, as well as reduce the computation time.



In the Custom Comparison(s) window (Figure 215), the default contrasts initially appear in the lower half. The upper half enables the creation of contrasts. The first two questions create contrasts that query differential expression between two groups; the second question specializes to the question of the differential expression between two



groups, both of which reside in a subgroup. The third and fourth questions query interactions, where the fourth question specializes to a subgroup analogously to question 2.

TAC 4.0.2 calls the `lmFit`, `contrasts.fit`, `duplicateCorrelation`, `ebayes`, and `topTable` functions in the `limma` Bioconductor package. The user can select whether to do a simple ANOVA analysis or to add the eBayes empirical parameter improvement (default).

## Anova

Analysis of variance or ANOVA analysis fits a linear model to each probeset independently of the others.

## eBayes

The eBayes analysis corrects the variance of the ANOVA analysis with an empirical Bayes approach that uses the information from all the probesets to yield an improved estimate for the variance. The eBayes correction is especially important when the number of samples being analyzed is small.



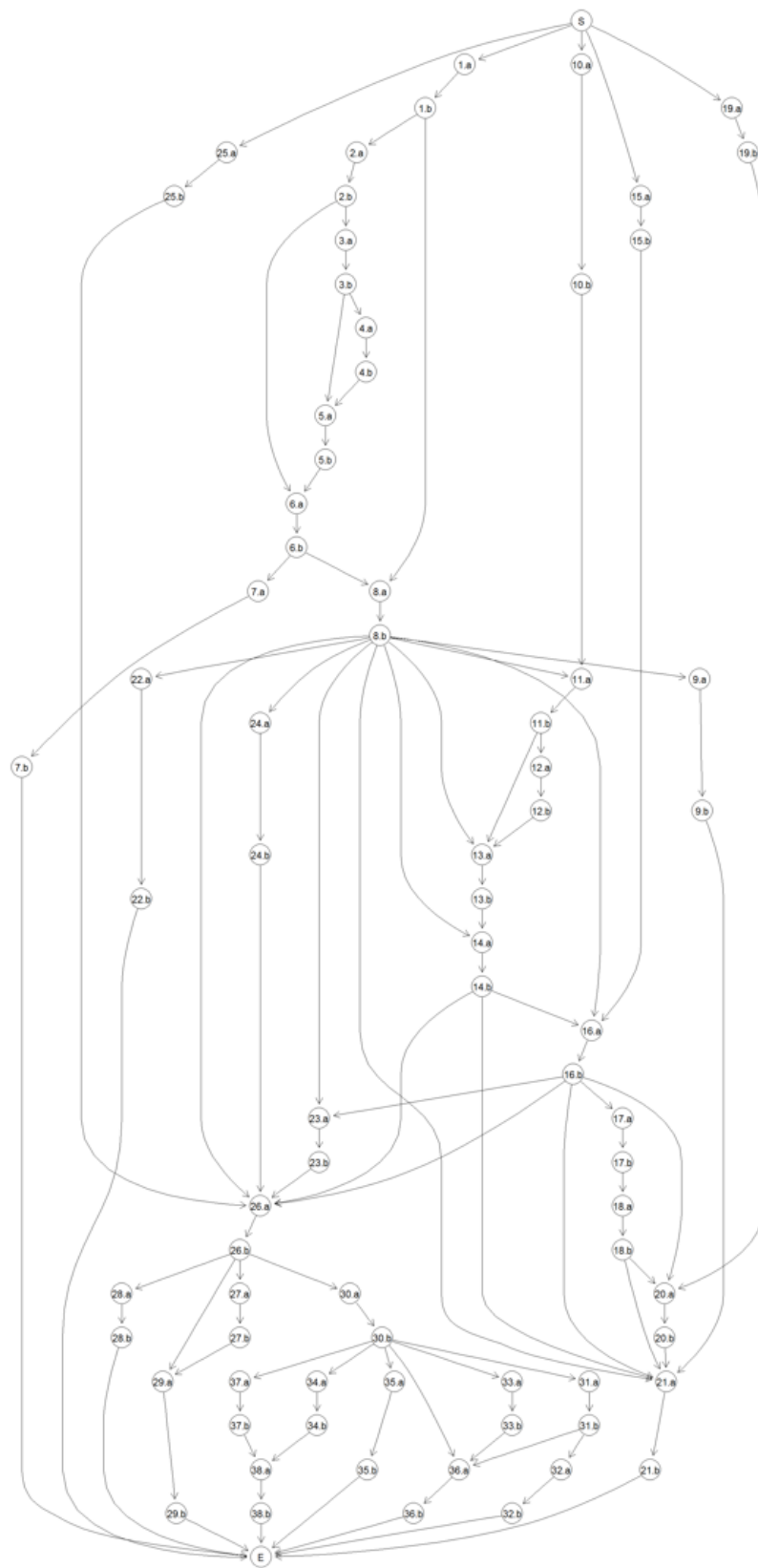
## Alternative Splicing Analysis

### Event Pointer Alternative Splicing Analysis

In Affymetrix microarray designs that enable the inspection of genes for potential alternative splicing events, the genes are composed of probe selection regions (PSRs) that represent exons and junctions (JUCs) that straddle the inter-exon connections. Event Pointer is a statistical algorithm for computing the probability of differential alternative splicing events in genes based on the summary expression signals of the PSRs and JUCs. Event Pointer starts with the set of all known isoforms for a gene from which it creates the splicing graph. Alternative splicing events are extracted from the splicing graph in the form of PSR/JUC signal sets for a reference and a set of associated paths. An extended limma analysis of these signals yields a probability that the differential alternative splicing event occurs between two conditions. The differential alternative splicing events are examined for the same contrasts that are examined for the ordinary limma analysis.

Splicing graph for the NCALD gene appears on the left. At the top, the start node is marked "S". At the bottom, the end node is marked "E". Paths from the start to the end represent isoforms. Nodes or combination of nodes represent different alternative splicing events. A flow through the splicing graph starting at "S" and ending at "E", and at each intermediate node, the flow randomly splits. An event is defined such that the sum of the flows in a set of paths is equal to the flow in some reference. For example, the flows through  $(3b-5a) + (3b-4a-4b-5a) = (2b-3a-3b, 5a-5b-6a)$ . The event defined here is a cassette event. This type of analysis is done for every gene generating a list of potential events. The extended limma analysis computes the probability of these events.

TAC 4.0.2 calls a customized version of Event Pointer with a few changes. First, the number of paths can have an arbitrary number, and not just two. The extended limma analysis is generalized to allow for the arbitrary number of paths. Second, the signal quality of JUC probes is somewhat lower than PSR probes due to the constrained positioning of a junction probe, so JUC probes are down weighted in the analysis. Third, the p-value returned is the average of the one associated with the isoform fold change and the relative concentration change, relative to what is described in the paper [9] [on page 218](#).



The advantages of Event Pointer are that it is a fully statistical analysis that uses information from both the PSRs and the JUCs. It has the full splicing graph information, so it recognizes the standard alternative splicing events (i.e. cassette exon, mutually exclusive exons, alternate 5' donor site, alternate 3' acceptor site, intron retention, alternative first exon, and alternative last exon) as well as more complex events that do not match these categories. The disadvantages of Event Pointer include that it can miss certain events due to the complexity of the splicing graphs in some genes.

## Splicing Index (SI) algorithm

Splicing Index algorithm is a way to measure of how much exon specific expression differs between two conditions after excluding gene level influences. The algorithm first normalizes the exon and junction expression values by the level of gene expression and creates a ratio of normalized signal estimates from one condition relative to another. **Note:** The Splicing Index algorithm in TAC 4.0.2 is different than in previous versions of TAC. TAC 4.0.2 (by default) runs the Event Pointer and Event Estimation simultaneously. For details, see "[Event Pointer Algorithm](#)" on page 210.

## Performing an Alternative Splicing Analysis

In the **Preference** window you can customize algorithm parameters by typing values in the text boxes. ([Figure 216](#)) **Note:** Alternative Splicing Analysis is only available for certain arrays.

**Figure 216** Configurable Parameters 1-5

▲ Is Expressed Criteria

A Probeset (Gene/Exon) is considered expressed if  % samples have DABG values below DABG Threshold.

DABG Threshold:

The Splicing Index algorithm compares normalized signal estimates from one condition to another. See the equation below: ([Figure 217](#))

**Figure 217** Splicing Index Algorithm

$$\text{Splicing index} = \frac{\left\{ \begin{array}{l} \text{Exon 1 Condition 1 Intensity} \\ \text{Gene 1 Condition 1 Intensity} \end{array} \right\}}{\left\{ \begin{array}{l} \text{Exon 1 Condition 2 Intensity} \\ \text{Gene 1 Condition 2 Intensity} \end{array} \right\}}$$

In order for Transcriptome Analysis Console to perform Splicing Index, two key criteria must be met. They are as follows:

1. **Criteria 1:** A Transcript Cluster gene must be expressed in both conditions, therefore for each condition, you must determine whether a gene is expressed or not.
  - i. A gene is expressed in a sample if its gene level DABG < 0.05 (configurable parameter)
  - ii. A condition is expressed if ≥ 50% (configurable parameter) samples are expressed





2. **Criteria 2:** A PSR or Junction can only be analyzed by Splicing Index if it expresses in at least one condition.
  - To decide whether a PSR or junction is expressed in a condition, you need (as an example) to check the DABG p-values (from all samples in that condition) to see whether  $\geq 50\%$  of samples have DABG  $p < 0.05$ . For more information, go to: [thermofisher.com](http://thermofisher.com)

**Note:** PSR/JUC FDR p-value calculation is only based on PSR/JUC that passed the two criteria above.

**Figure 218** Criteria passing examples

|                            | cond1<br>sample1 | cond<br>1<br>sample2 | cond<br>1<br>sample3 | cond<br>1<br>sample4 | cond1<br>summary -<br>whether $\geq 50\%$<br>samples<br>have that<br>PSR/JUC<br>as expressed | cond<br>2<br>sample1 | cond<br>2<br>sample2 | cond2<br>sample3 | cond2<br>sample4 | cond2<br>summary -<br>whether $\geq 50\%$<br>samples<br>have that<br>PSR/JUC<br>as expressed | A PSR<br>or<br>Junction<br>can<br>only<br>be<br>analyzed<br>by<br>Splicing<br>Index<br>if it<br>expresses<br>in<br>at least<br>one<br>condition |
|----------------------------|------------------|----------------------|----------------------|----------------------|----------------------------------------------------------------------------------------------|----------------------|----------------------|------------------|------------------|----------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------|
| PSR/JUC<br>DABG<br>p-value | 0.28713<br>92    | 0.378<br>386         | 0.272<br>059         | 0.39<br>361<br>4     | NO                                                                                           | 0.02<br>062<br>8     | 0.020<br>067         | 0.021189         | 0.021041         | YES                                                                                          | YES                                                                                                                                             |
| PSR/JUC<br>DABG<br>p-value | 0.04889<br>926   | 0.031<br>045         | 0.048<br>693         | 0.05<br>589<br>3     | YES                                                                                          | 0.06<br>610<br>4     | 0.078<br>026         | 0.08502          | 0.081922         | no                                                                                           | YES                                                                                                                                             |
| PSR/JUC<br>DABG<br>p-value | 0.00129<br>8487  | 0.001<br>298         | 0.001<br>771         | 0.00<br>165<br>3     | YES                                                                                          | 0.00<br>017<br>7     | 0.000<br>207         | 0.000148         | 0.000177         | YES                                                                                          | YES                                                                                                                                             |
| PSR/JUC<br>DABG<br>p-value | 0.15584<br>61    | 0.156<br>082         | 0.157<br>794         | 0.13<br>427<br>4     | no                                                                                           | 0.01<br>555<br>2     | 0.017<br>323         | 0.021277         | 0.016437         | YES                                                                                          | YES                                                                                                                                             |
| PSR/JUC<br>DABG<br>p-value | 0.13660<br>51    | 0.092<br>723         | 0.099<br>304         | 0.16<br>844<br>7     | no                                                                                           | 0.02<br>115<br>9     | 0.021<br>661         | 0.01992          | 0.020894         | YES                                                                                          | YES                                                                                                                                             |
| PSR/JUC<br>DABG<br>p-value | 0.15348<br>52    | 0.126<br>306         | 0.120<br>964         | 0.17<br>697<br>6     | no                                                                                           | 0.01<br>752<br>9     | 0.018<br>887         | 0.020776         | 0.023225         | YES                                                                                          | YES                                                                                                                                             |
| PSR/JUC<br>DABG            | 0.06194<br>299   | 0.096<br>116         | 0.065<br>838         | 0.08<br>856          | no                                                                                           | 0.03<br>458          | 0.040<br>754         | 0.034882         | 0.037154         | YES                                                                                          | YES                                                                                                                                             |



**Note:** For each PSR/JUC, the normalized signal is calculated by "PSR/JUC signal"/"gene signal". For example, if the PSR signal is 5.46, gene signal is 9.09, the normalized signal will be -3.63 (this is log scale).

**Figure 219** Normalized Signal calculation example

| normalized signals             | cond1 sample1 | cond1 sample2 | cond1 sample3 | cond1 sample4 | cond2 sample1 | cond2 sample2 | cond2 sample3 | cond2 sample4 |
|--------------------------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
| PSR1                           | -3.63         | -4.1          | -2.83         | -3.3          | -0.97         | -0.94         | -1.25         | -0.89         |
| biweight avg of cond1 for PSR1 | -3.47         |               |               |               |               |               |               |               |
| biweight avg of cond2 for PSR1 | -0.94         |               |               |               |               |               |               |               |
| Splicing index of PSR1         | -2.53         |               |               |               |               |               |               |               |

## Splicing Event estimation score

**IMPORTANT!** The Splicing Event Estimation method is used as a classification tool and not intended for a de novo event discovery.

### Splicing Event Estimation

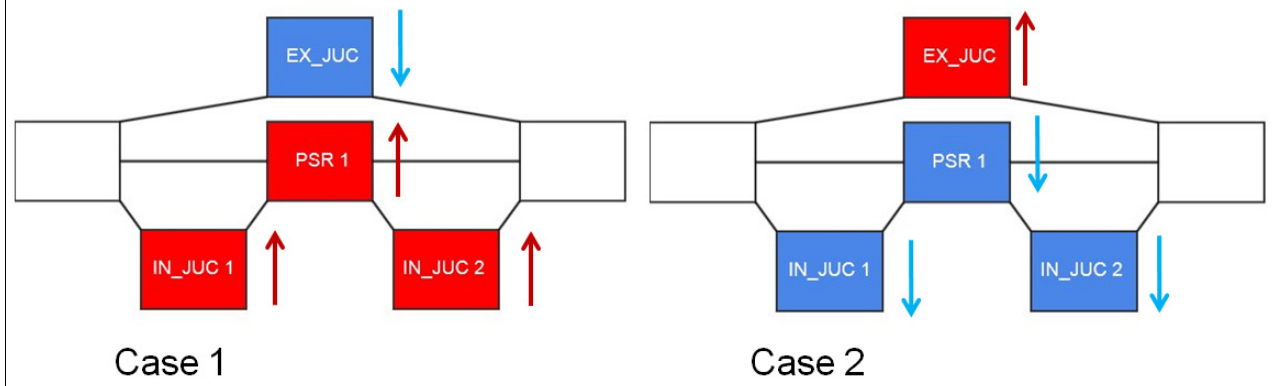
The Splicing Event Estimation is generated by detecting changes between two conditions using SI and p-values. It investigates each mode and finds specific patterns that meet each mode's definition.

The Splicing Event Estimations are based on five following classic modes of alternative splicing:

- **Cassette Exon (CE)** - From a Primary Transcript, one exon is spliced out along with its partnering introns.

For CE, the Splicing Event Estimation is based on PSR and its two inclusion junction patterns moving in the same direction, while its exclusion junction moves in the opposite direction, as shown in [Figure 220](#). The PSR direction influences the direction of the junction.

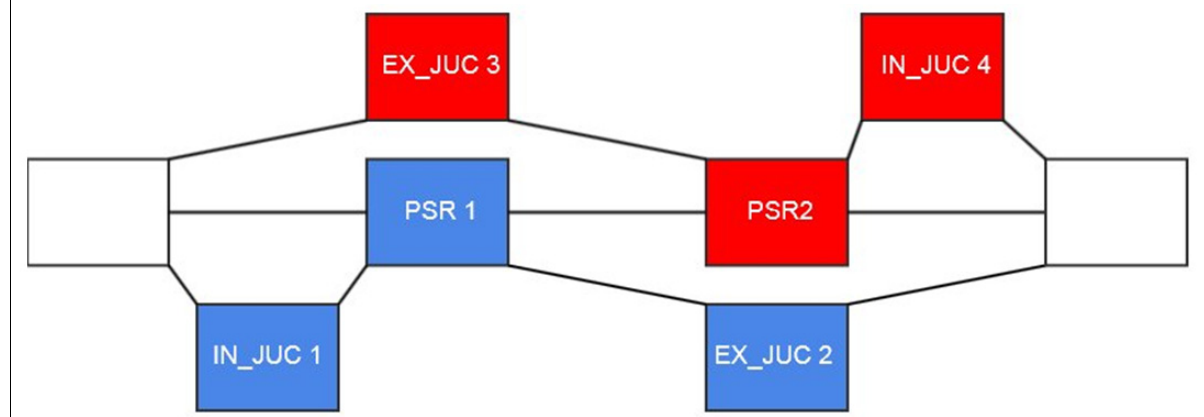
**Figure 220** Cassette Exon (CE) pattern



- **Mutually Exclusive Exons (MXE)** - 2 exons (back to back) that are excluded from the mature mRNA transcript.

For MXE, the Splicing Event Estimate is based on PSR and its inclusion and exclusion junctions moving in the opposite direction of another set, as shown in [Figure 221](#). The PSR direction influences the direction of the junction.

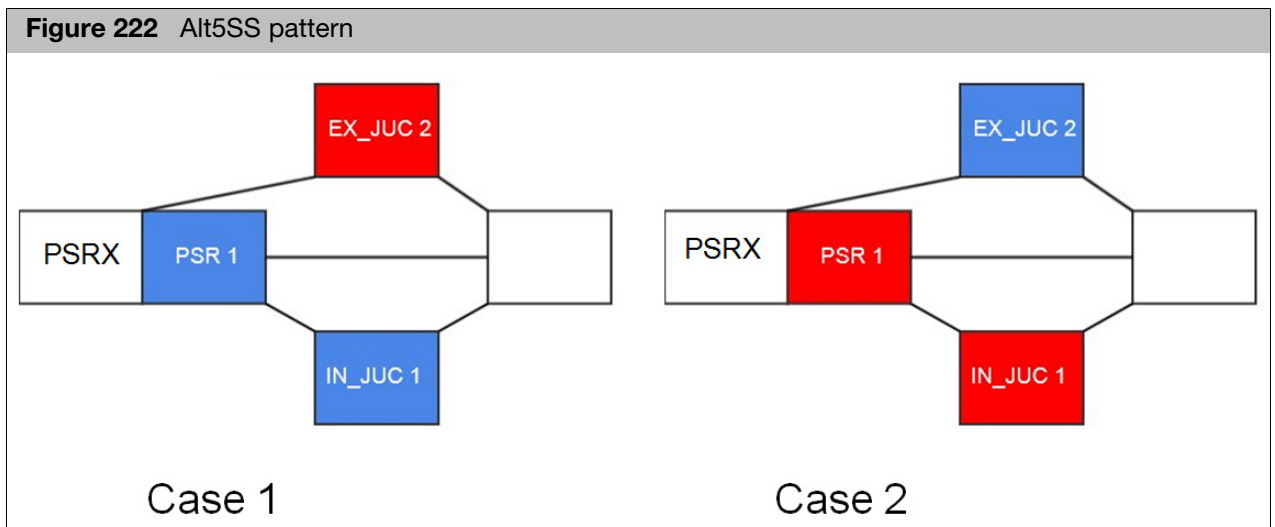
**Figure 221** Mutually Exclusive Exons (MXE) pattern



- **Alternative 5' Sites (Alt 5SS)** - 2 or more splices are recognized at the 3' end of the exon. At this point, an alternative 5' splice junction (donor site) is used. This changes the 3' boundary of the upstream exon.

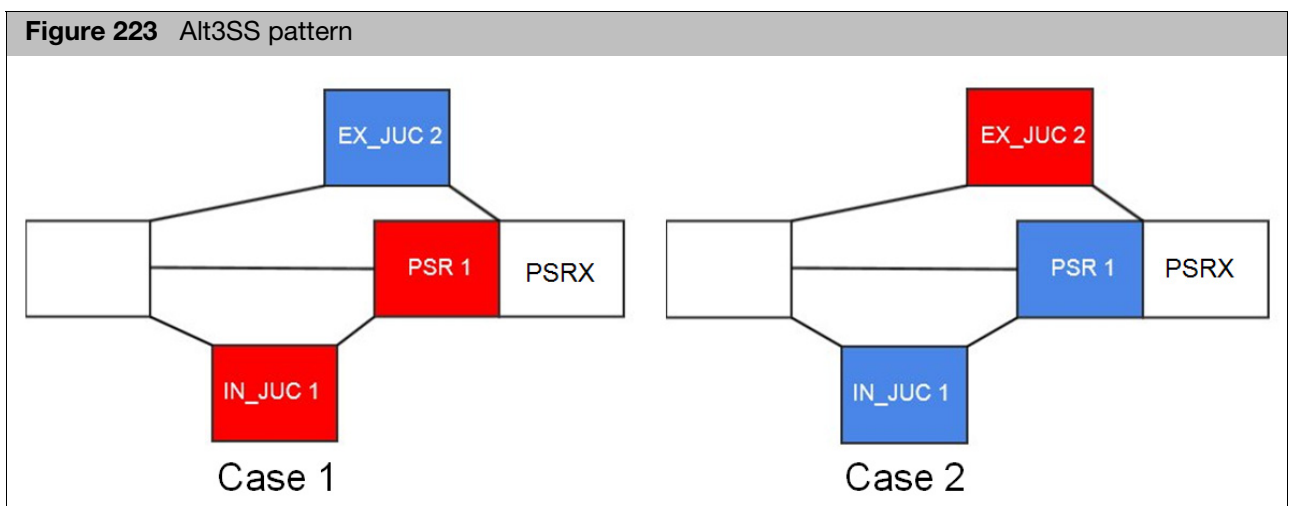
For Alt 5SS, the Splicing Event Estimate is based on PSRs and their inclusion junctions changing in the same direction, while their exclusion junctions move in

the opposite direction, as shown in Figure 222. PSRX and PSR1 are different, but are genomic neighbors. The PSR direction influences the direction of the junction; PSR1's direction decides PSRX's direction.



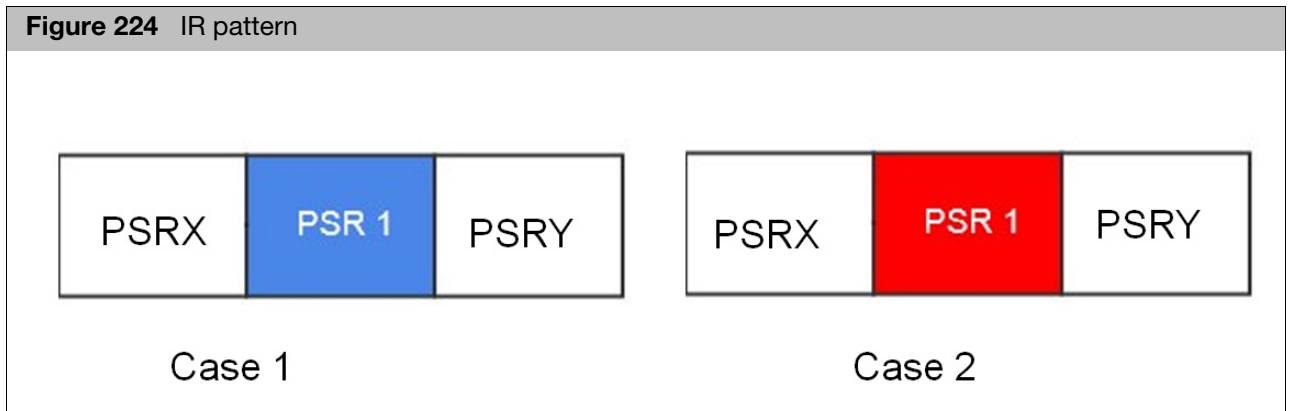
- **Alternative 3' Sites (Alt 3SS)** - 2 or more splices are recognized at the 5' end of the exon. At this point, an alternative 3' splice junction (acceptor site) is used. This changes the 5' boundary of the downstream exon.

For Alt 3SS, the Splicing Event Estimate is based on PSRs and their inclusion junctions changing in the same direction, while their exclusion junctions move in the opposite direction, as shown in Figure 223. PSRX and PSR1 are different, but are genomic neighbors. The PSR direction influences the direction of the junction; PSR1's direction decides PSRX's direction.



- **Intron Retention (IR)** - A sequence is spliced out as an intron or is retained within the mature mRNA transcript.

For IR, the Splicing Event Estimate is based on the pattern of the PSR versus its genomic-related PSRX and PSRY neighbors, as shown in [Figure 224](#).



### Event Estimation Score

- The Event Estimation Score is between 0 and 1 (1 is the highest possible score).
- The Event Estimation Score is based on how well your data fits into pre-defined splicing patterns. PSRs and their related junctions all contribute to an event score.
- PSRs and Junctions can contribute positively or negatively to the event score depending on their specific pattern directions, as described earlier.
- Only PSRs are assigned Event Estimation Score
- PSRs with more than 1 pattern, receive an event estimation score for each of the possible modes of alternative splicing, but the final estimate of the PSR is based on the highest event score.

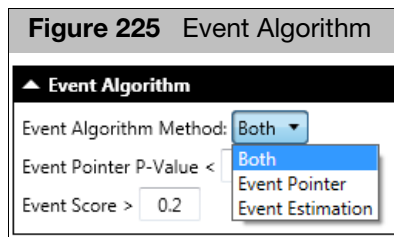
**IMPORTANT!** Real-time PCR is routinely used to verify microarray results due to its high sensitivity and wide dynamic range.

It is recommended to verify the splicing events by using the USB First-Strand cDNA Synthesis Kit [PN 75780] followed by real-time PCR with USB VeriQuest Probe or SYBR® Green qPCR master mixes [PN#75650/75600].

Detailed information can be found in the Data Sheet. Concordance of Affymetrix GeneChip® Human Transcriptome Array 3.0 and real-time PCR results using USB® VeriQuest® qPCR master mixes.

## Event Pointer Algorithm

The Event Pointer is a new option for the Splicing Index (SI) algorithm and is run when Event Algorithm Method is set to **Both** or **Event Pointer**, as shown in [Figure 225](#). For more information on customizing algorithms, see "[Customizing an Installed Array's Algorithm Settings](#)" on page 18.



## Event score

TAC 4.0.2 includes two algorithms for identifying alternative splicing events: the TAC 2.0 algorithm and the new Event Pointer in TAC 4.0.2. The new Event Score leverages both previous TAC 2.0 event estimation score and Event Pointer p-value and sorts the most likely alternative splicing events to the top. The Event Score is computed as follows:

$$\text{EventScore} = (\text{MIN}(1, \text{Log10}(\text{Event Pointer\_P})/\text{Log10}(1\text{E-}20)) + \text{Event Estimation Score}) * 0.5$$

Then adjust the event score slightly by SI.

$$\text{EventScore} = \text{EventScore} - 0.0001 + 0.000005 * \text{Log2}(|\text{SplicingIndex}|)$$

If EventScore is greater than 1, then it will be capped at 1.

Events given high scores by both algorithms tend to be the best events scored by human inspection. The results from each constituent algorithm are available through examination of the hidden columns.



## MAS 5.0, RMA and PLIER

The following are brief descriptions of the MAS 5.0, RMA, and PLIER algorithms. References and links to various publications, describe the MAS 5.0, RMA, and PLIER algorithms in detail. See "[Comparison of algorithms](#)" on page 214 for a side-by-side comparison of the assumptions, advantages and disadvantages of using each individual algorithm.

### MAS 5.0

The MAS 5.0 algorithm uses the Tukey's biweight estimator to provide a robust mean Signal value and the Wilcoxon's rank test to calculate a significance or p-value and Detection call for each probe set. Background estimation is provided by a weighted average of the lowest 2% of the feature intensities. Mismatch probes are utilized to adjust the perfect match (PM) intensity. Linear scaling of the feature level intensity values, using the trimmed mean, is the default to make the means equal for all arrays being analyzed.

Tukey's Bi-weight average is a method to determine a robust average unaffected by outliers.

- **Step 1:** The median is determined to define the center of the data.
- **Step 2:** The distance of each data point from the median is determined. This distance is then used to determine how much each value should contribute to the average. For example, outliers that are far away from the median should contribute less to the average.
- All end result values represented as a Bi-weight average, are shown in a log2 scale (if the CHP files are summarized using RMA)
- For more information, go to: [thermofisher.com](http://thermofisher.com)

The MAS 5.0 algorithm (also known as the Statistical Algorithm) analyzes each array independently. As a result, individual probe-specific affinities cannot be considered and the ability to detect small changes between experiment and control samples is reduced in comparison to either RMA or PLIER. The primary use of the MAS 5.0 algorithm is to obtain a quick report regarding the performance of the arrays and to identify any obvious problems before submitting the final set of arrays to one of the multi-chip analysis methods (RMA, PLIER).

For a more detailed description of the MAS 5.0 algorithm, see the Statistical Algorithms Reference Guide at [thermofisher.com](http://thermofisher.com).

The TAC software displays column headings for individual parameters and outputs in the Sample QC View for each of the 3' Expression algorithms. See the following table for column header definitions.



## MAS 5.0 Column Headers and Definitions

| Column Header      | Definition                                                          |
|--------------------|---------------------------------------------------------------------|
| <b>#P</b>          | The number of probe sets present.                                   |
| <b>%P</b>          | The percent of probe sets present.                                  |
| <b>Signal(P)</b>   | The average signal for the probe sets defined as present.           |
| <b>#M</b>          | The number of probe sets whose detection is marginal.               |
| <b>%M</b>          | The percent of probe sets whose detection call is marginal.         |
| <b>Signal(M)</b>   | The average signal for probe sets whose detection call is marginal. |
| <b>#A</b>          | The number of probe sets absent.                                    |
| <b>%A</b>          | The percent of probe sets called present that are absent.           |
| <b>Signal(A)</b>   | The average signal for the probe sets defined as absent.            |
| <b>Signal(All)</b> | The average signal for all probe sets on the array.                 |

## RMA

The Robust Multi-chip Analysis (RMA) algorithm fits a robust linear model at the probe level to minimize the effect of probe-specific affinity differences. This approach:

- Increases sensitivity to small changes between experiment and control samples.
- Minimizes variance across the dynamic range, but does compress calculated fold change values.

RMA consists of three steps:

1. Background adjustment
2. Quantile normalization
3. Summarization

This is a multi-chip analysis approach. Therefore, all arrays intended for comparison should be included together in the summarization step.

For a more detailed description of the RMA algorithm, see the publication, *Exploration, Normalization, and Summaries of High Density Oligonucleotide Array Probe Level Data, Biostatistics, April 2003; Vol. 4;Number 2: 249-264.*

The TAC software displays column headings for individual parameters and outputs in the Sample QC View. See the following table for column header definitions.





## RMA and PLIER Column Headers and Definitions

Metrics are dependent on the analysis algorithm, array type, and the level; therefore, not all of the metrics are always present in the Full Report.

| Column Header               | Definition                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
|-----------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Spike-probeID-signal</b> | The signal for the probe sets that correspond to the labeling and hybridization spike controls.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
| <b>pm_mean</b>              | The mean signal for all of the PM probes on the array.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
| <b>mm_mean</b>              | The mean signal for all of the MM probes on the array.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
| <b>bgrd_mean</b>            | The average signal for the probes used to calculate the background.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
| <b>pos_vs_neg_auc</b>       | <p>The area under the curve (AUC) for a receiver operator curve (ROC) comparing the intron controls to the exon controls by applying a threshold to the probe set summary. The ROC curve is generated by evaluating how well the probe set summary separates the positive controls from the negative controls (e.g., exon from intron). The assumption (which is only valid in part) is that the negative controls are a measure of false positives and the positive controls are a measure of true positives. An AUC of 1 reflects perfect separation whereas as an AUC value of 0.5 would reflect no separation.</p> <p><b>Note:</b> The AUC of the ROC curve is equivalent to a rank sum statistic used to test for differences in the center of two distributions.</p> |

### SST-RMA

The SST-RMA method is to reduce background and to increase fold change. The SST-RMA method includes GC4 (GC Correction Version 4) background reduction and SST (Signal Space Transformation) intensity normalization to the expression data processing workflow. The GC4 background reduction is achieved through GC count leveling. SST Intensity normalization is an adjustment of probe intensity levels. For detailed information related to GC4 background reduction and SST intensity normalization, please refer to white paper available at [thermofisher.com](http://thermofisher.com).

For column header definitions, refer to the table above ("[RMA and PLIER Column Headers and Definitions](#)" on page 213).

### PLIER

The Probe Logarithmic Intensity Error Estimation (PLIER) algorithm method produces a signal by accounting for experimentally observed patterns in probe behavior and handling error at the appropriately low and high signal values. Similar to RMA, the PLIER algorithm also utilizes data across all arrays submitted for analysis to minimize the effect of probe-specific affinity differences. Unlike RMA, which uses a simple global background correction approach, the standard PLIER configuration uses a probe-specific background providing a higher degree of accuracy at the cost of increased signal variance. Exon Arrays use a more advanced and efficient background correction method utilizing surrogate mismatched probes. (See the technical note, GeneChip® Exon Array Design at [thermofisher.com](http://thermofisher.com)).

Signal variance can be addressed following the PLIER analysis by applying a variance-stabilizing data transformation. A simplistic but effective approach is to add a value of 16 to each and every signal value (PLIER + 16 algorithm). This is a multi-chip analysis approach. Therefore, all arrays intended for comparison should be included together in the summarization step.

For a more detailed description of the PLIER algorithm, see Guide to Probe Logarithmic Intensity Error (PLIER) Estimation at [thermofisher.com](http://thermofisher.com).

For column header definitions, refer to the table above.



## Comparison of algorithms

The table below is a comparison of the MAS 5.0, RMA, and PLIER algorithms.

| Algorithm      | Advantages                                                                                                                                                                                                                                                                                                                                                                                                                                                                        | Disadvantages                                                                                                                                                                                                                                                                                                                                                                 | Assumptions                                                                                                                                                                                                                         |
|----------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>MAS 5.0</b> | <p>Single-array algorithm is independent of other data in the data set:</p> <ul style="list-style-type: none"> <li>■ Less computationally intensive than RMA or PLIER</li> <li>■ Conservative</li> <li>■ Smooth down-weighting of outliers</li> <li>■ Positive output values</li> <li>■ Minimal bias</li> </ul>                                                                                                                                                                   | <p>Not as sensitive as either RMA or PLIER to small changes in target abundance:</p> <ul style="list-style-type: none"> <li>■ Limited ability to adjust for probe-specific affinity differences</li> <li>■ Unstable variance at low end</li> <li>■ Lower precision in signal calculation</li> </ul>                                                                           | <ul style="list-style-type: none"> <li>■ Single-array analysis</li> <li>■ Multiplicative error</li> <li>■ Signal is adjusted by the mismatch probes (PM-MM)</li> <li>■ Background imputed to handle negative differences</li> </ul> |
| <b>RMA</b>     | <p>Minimizes the variance seen across the arrays</p> <ul style="list-style-type: none"> <li>■ Higher reproducibility of signal over single array analyses</li> <li>■ Good differential change detection</li> <li>■ Variance stable on log scale</li> </ul>                                                                                                                                                                                                                        | <p>Not as sensitive as PLIER in the ability to detect small fold changes</p> <ul style="list-style-type: none"> <li>■ In cases where feature intensities disagree, may have more than one solution (mitigated by median polish)</li> <li>■ Positive bias contributed to signal values</li> <li>■ Compresses fold changes for low intensity probe sets</li> </ul>              | <ul style="list-style-type: none"> <li>■ Multiple-array analysis</li> <li>■ Multiplicative error</li> <li>■ PM (perfect match) only</li> <li>■ Single background is used to adjust each intensity</li> </ul>                        |
| <b>PLIER</b>   | <p>Ability to detect small fold changes:</p> <ul style="list-style-type: none"> <li>■ Higher reproducibility of signal (lower coefficient of variation) without loss of accuracy</li> <li>■ Higher sensitivity to changes in abundance for targets near background</li> <li>■ Dynamic weighting of the most informative probes in an experiment to determine signal</li> <li>■ High degree of accuracy for signal and fold change calculations</li> <li>■ Lack of bias</li> </ul> | <p>More variance in individual signals than seen with RMA</p> <ul style="list-style-type: none"> <li>■ Computationally intensive</li> <li>■ In cases where feature intensities disagree, may have more than one solution (mitigated by median polish)</li> <li>■ Performance relative to amount of model data provided</li> <li>■ Variance not stable on log scale</li> </ul> | <ul style="list-style-type: none"> <li>■ Multiple-array analysis</li> <li>■ Mixed error model PM-MM, PM only, etc.</li> <li>■ Multiple background options</li> <li>■ Smoothly handles intensities below background</li> </ul>       |



## Benjamini-Hochberg Step-Up FDR-controlling Procedure

FDR control is a statistical method used in multiple hypothesis testing to correct for multiple comparisons.

- FDR is controlled at certain alpha level (default 0.05 in all result tables). This means that the expected proportion of rejections that are in error is less than alpha. Alpha level can be customized in the result tables by changing the default filtering criteria in the FDR p-value column (0.05) to something else.
- Benjamini, Yoav; Hochberg, Yosef (1995). "Controlling the false discovery rate: a practical and powerful approach to multiple testing". *Journal of the Royal Statistical Society, Series B (Methodological)* 57 (1): 289-300. MR 1325392.

## Fold change

- Fold change is a number describing how much the signal changes from an initial condition group to a final condition group.
- Fold changes are represented in linear space.

## Hierarchical clustering

- Clustering is performed on both probe sets and CHP signals.
- Distance metric used between objects is the Euclidean distance.
- Distances between clusters of objects are computed using the complete linkage method (maximum distance between a pair of objects in the two clusters).
- Results are displayed in a heat map and dendrogram.



## Probe set suffixes

A number of standard metrics are created for particular groups of probe sets on the arrays. The table below describes the suffix and descriptions for the following groups of probe sets:

- **all\_probeset** (all probe sets within a group)
- **bac\_spike** (bacterial spikes and hybridization controls)
- **polya\_spike** (labeling controls)
- **neg\_control** (negative control probes)
- **pos\_control** (positive control probes)

| Suffix for Output          | Definition                                                                                                                                                                                                                                                                                                                                                 |
|----------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>_probesets</b>          | From the QC Report, the number of probe sets actually analyzed.                                                                                                                                                                                                                                                                                            |
| <b>_mean</b>               | The mean signal value for all the probe sets.                                                                                                                                                                                                                                                                                                              |
| <b>_stdev</b>              | The standard deviation for all of the probe sets analyzed.                                                                                                                                                                                                                                                                                                 |
| <b>_mad_residual_mean</b>  | The mean absolute deviation of the residual for a chip versus all chips in the data set.                                                                                                                                                                                                                                                                   |
| <b>_mad_residual_stdev</b> | The standard deviation of the residual for a chip versus all chips in the data set.                                                                                                                                                                                                                                                                        |
| <b>_rle_mean</b>           | The mean absolute relative log expression (RLE) – This metric is generated by taking the probe set summary for a given chip and calculating the difference in log base 2 from the median value of that probeset over all the chips. The mean is then computed from the absolute RLE for all the probe sets for a given CEL file.                           |
| <b>_rle_stdev</b>          | The standard deviation of the relative log expression (RLE) – This metric is generated by taking the probe set summary for a given chip and calculating the difference in log base 2 from the median value of that probeset over all the chips. The standard deviation is then computed from the absolute RLE for all the probe sets for a given CEL file. |
| <b>_percent_called</b>     | The percent of probe sets from the exon analysis called present (DABG <= 0.01).                                                                                                                                                                                                                                                                            |



## Chromosome naming scheme

- In addition to the “regular” chromosomes, the hg19 browser contains nine haplotype chromosomes and 59 unplaced contigs. If an unplaced contig is localized to a chromosome, the contig name is appended to the regular chromosome name, as in chr1\_gl000191\_random. If the chromosome is unknown, the contig is represented with the name “chrUn” followed by the contig identifier, as in chrUn\_gl000211. Note that the chrUn contigs are no longer placed in a single, artificial chromosome as they have been in previous UCSC assemblies. See the sequences page for a complete list of hg19 chromosome names.
- The 9 haplotype chromosomes are as follows:

| Name                 | Accession  | UCSC chr Name   |
|----------------------|------------|-----------------|
| HSCHR6_MHC_APD_CTG1  | GL000250.1 | chr6_apd_hap1   |
| HSCHR6_MHC_COX_CTG1  | GL000251.1 | chr6_cox_hap2   |
| HSCHR6_MHC_DBB_CTG1  | GL000252.1 | chr6_dbb_hap3   |
| HSCHR6_MHC_MANN_CTG1 | GL000253.1 | chr6_mann_hap4  |
| HSCHR6_MHC_MCF_CTG1  | GL000254.1 | chr6_mcf_hap5   |
| HSCHR6_MHC_QBL_CTG1  | GL000255.1 | chr6_qbl_hap6   |
| HSCHR6_MHC_SSTO_CTG1 | GL000256.1 | chr6_ssto_hap7  |
| HSCHR4_1_CTG9        | GL000257.1 | chr4_ctg9_hap1  |
| HSCHR17_1_CTG5       | GL000258.1 | chr17_ctg5_hap1 |



## References

- [1] Jesse H. Krijthe (2015). Rtsne: T-distributed stochastic neighbor embedding using a Barnes-Hut implementation, URL: <https://github.com/jkrijthe/Rtsne>.
- [2] Michael Hahsler (2016). dbscan: Density based clustering of applications with noise (DBSCAN) and Related Algorithms. R package version 0.9-8. <https://CRAN.R-project.org/package=dbscan>
- [3] Ulrich Bodenhofer, Andreas Kothmeier, and Sepp Hochreiter (2011) APCluster: an R package for affinity propagation clustering Bioinformatics 27:2463-2464.
- [4] Brendan J. Frey and Delbert Dueck (2007). Clustering by passing messages between data points. Science 315:972-977.
- [5] Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Research 43(7), e47.
- [6] Phipson, B, Lee, S, Majewski, IJ, Alexander, WS, and Smyth, GK (2016). Robust hyperparameter estimation protects against hypervariable genes and improves power to detect differential expression. Annals of Applied Statistics 10(2), 946-963.
- [7] Smyth, G. K., Michaud, J., and Scott, H. (2005). The use of within-array replicate spots for assessing differential expression in microarray experiments. Bioinformatics 21, 2067- 2075.
- [8] Limma User's Guide, <https://www.bioconductor.org/packages/devel/bioc/vignettes/limma/inst/doc/usersguide.pdf>.
- [9] Juan P. Romero et al, "Event Pointer: an effective identification of alternative splicing events using junction arrays", BMC Genomics 2016 17:467.
- [10] Juan Pablo Romero and Angel Rubio (2015). Event Pointer: analysis of alternative splicing using microarray data. R package version 1.0.
- [11] W.E. Johnson, C. Li, and A. Rabinovic. Adjusting batch effects in microarray data using empirical bayes methods. Biostatistics, 8(1):118-127, 2007.
- [12] Jeffrey T. Leek, W. Evan Johnson, Hilary S. Parker, Elana J. Fertig, Andrew E. Jaffe and John D. Storey (2016). sva: Surrogate Variable Analysis. R package version 3.20.0.).



# MATERIEL ET METHODES





## I. Matériel

### 1. Cellules Humaines

Les cellules HeLa S3 dérivent de la lignée cellulaire parente HeLa qui est une lignée cellulaire immortelle issue de cellules d'adénocarcinome cervical d'une patiente nommée Henrietta Lacks morte en 1951.

### 2. Souches bactériennes

Souche E. coli DH5α. Elle est utilisée pour la production et la purification de plasmides.

### 3. Oligodeoxynucleotides

Les oligodeoxynucléotides utilisés au cours de la thèse sont décrits dans la Table M1

**Table M1. Désignations, séquences et usages des oligodésoxynucléotides employés**

| Numéro Identifiant | Séquence 5' – 3'          | Description                                          |
|--------------------|---------------------------|------------------------------------------------------|
| 7392               | GATACCATGATCACGAAGGTGGTT  | Primer forward U1 Humain et Souris qRT-PCR           |
| 7393               | CACAAATTATGCAGTCGAGTTTCC  | Primer reverse U1 Humain et Souris qRT-PCR           |
| 7394               | GTGCGGAATCGACATCAAGAG     | Primer forward U11 Humain et Souris qRT-PCR          |
| 7395               | CGCCGGGACCAACGAT          | Primer reverse snARN U11 Humain et Souris qRT-PCR    |
| 7399               | CGGCCATACCACCTGAAC        | Primer forward ARN 5S Humain et Souris qRT-PCR       |
| 7400               | GCGGTCTCCCATCCAAGTAC      | Primer reverse ARN 5S Humain et Souris qRT-PCR       |
| 7483               | GTCACCATGCCTCCCTAGAA      | Primer forward scaARN 8 Humain et Souris qRT-PCR     |
| 7484               | ATCTGTCTGCCCCGTATCTG      | Primer reverse scaARN 8 Humain et Souris qRT-PCR     |
| 7568               | TTTCTTGGGGTTGCGCTACTGT    | Primer forward snARN U4atac Humain et Souris qRT-PCR |
| 7569               | GCACCAAAATAAAGCAAAAGCTCTA | Primer reverse snARN U4atac Humain et Souris qRT-PCR |
| 8401               | CGGCCTTTTGGCTAAGATCAAGTG  | Primer forward snARN U2 Humain et Souris qRT-PCR     |
| 8402               | TCCTCGGATAGAGGACGTATCAGA  | Primer reverse snARN U2 Humain et Souris qRT-PCR     |
| 8536               | AACAGCCTCAAGATCATCAGC     | Primer forward GAPDH Humain et Souris qRT-PCR        |
| 8537               | GGATGATGTTCTGGAGAGCC      | Primer reverse GAPDH Humain et Souris qRT-PCR        |
| 9431               | TGCCCAGTGATGACACCATCCTTG  | Primer forward snoRD23 Humain qRT-PCR                |
| 9432               | CCTCAGCCCCTGGGCATTGGCC    | Primer reverse snoRD23 Humain qRT-PCR                |
| A978               | GGGTGAGTGTGAGTGGAC        | Primer forward scaARN2 Humain qRT-PCR                |
| A979               | ATTTCGGGTCGGCCCG          | Primer reverse scaARN2 Humain qRT-PCR                |
| A980               | TTGCCCGGAGCCTGAGAG        | Primer forward scaARN17 Humain qRT-PCR               |
| A981               | TCCCGTCCAGCCTCACC         | Primer reverse scaARN17 Humain qRT-PCR               |

| Numéro Identifiant    | Séquence 5' – 3'                       | Description                                         |
|-----------------------|----------------------------------------|-----------------------------------------------------|
| B486                  | GCACTGACCTTCTTCCAAGC                   | Primer forward snoRD17 Humain qRT-PCR               |
| B487                  | TTGTGAGATGGGTCTCAGAGG                  | Primer reverse snoRD17 Humain qRT-PCR               |
| B491                  | ATCCAGCGGTTGTCTCAGCTAT                 | Primer forward snoRD74A Humain qRT-PCR              |
| B492                  | TCCACTCTCAGAGGTGCAAA                   | Primer reverse snoRD74A Humain qRT-PCR              |
| B493                  | CCCGTAATGTACGGAGGAAG                   | Primer forward snoRA57 Humain qRT-PCR               |
| B494                  | GGCGTTTGAGGATAGAACCA                   | Primer reverse snoRA57 Humain qRT-PCR               |
| B497                  | CCTGGATGATGATAAGCAAATG                 | Primer forward snoRD44 Humain qRT-PCR               |
| B498                  | GAGCTAATTAAGACCTTCATGTTCA              | Primer reverse snoRD44 Humain qRT-PCR               |
| B631                  | CCGACCCGAAATGAAAAGCT                   | Primer forward scaARN2 Humain qRT-PCR               |
| B632                  | CGGCCTCGTCTATCTGATCA                   | Primer reverse scaARN2 Humain qRT-PCR               |
| B633                  | CTGCTCCATTGACATCAGGC                   | Primer forward ARNnc 250 Humain qRT-PCR             |
| B634                  | TAACGGCTTCTCTGGCTT                     | Primer reverse ARNnc 250 Humain qRT-PCR             |
| B635                  | GCACACCAGACTTGACAGAAA                  | Primer forward scaARN13 Humain qRT-PCR              |
| B636                  | GCTTCTTACTGTTGGCGG                     | Primer reverse scaARN13 Humain qRT-PCR              |
| B637                  | TGAAATCAAGCTGACTCTGC                   | Primer forward scaARN15 Humain qRT-PCR              |
| B638                  | TGCTGTTGGTAGATAAGTAGGT                 | Primer reverse scaARN15 Humain qRT-PCR              |
| B639                  | GGGGTCCCTGTTCTCCTTAG                   | Primer forward scaARN20 Humain qRT-PCR              |
| B640                  | CTGAGGCAGGAAAACAAGGC                   | Primer reverse scaARN20 Humain qRT-PCR              |
| FISH snARN U2         | Biotine/ATACTACACTTGATCTTAGCCA/Biotine | Sonde LNA (Exiqon) snARN U2 RNA-FISH                |
| FISH snARN U1         | Biotine/ACCACAAATTATGCAGTCGAGT/Biotine | Sonde LNA (Exiqon) snARN U11 RNA-FISH               |
| FISH snARN U11_1      | Biotine/TCCGCACGCAGAGCAATCGA/Biotine   | Sonde LNA (Exiqon) snARN U11 RNA-FISH               |
| FISH snARN U11_2_3    | Biotine/TGTGCCACTCACGACAGAAGCC/Biotine | Sonde LNA (Exiqon) snARN U11 RNA-FISH               |
| FISH snARN U11_2_2    | Biotine/AATTATGCTTCCGAAATCTCT/Biotine  | Sonde LNA (Exiqon) snARN U11 RNA-FISH               |
| FISH snARN U4atac_2_3 | Biotine/GGTGTGTTGTTTCAGGCGTTA/Biotine  | Sonde LNA (Exiqon) snARN U4atac RNA-FISH            |
| FISH snARN U4atac_3   | Biotine/ACTATGCGCTCATTGGACA/Biotine    | Sonde LNA (Exiqon) snARN U4atac RNA-FISH            |
| FISH snARN U6atac_3   | Biotine/GGCCTCTTCCATCCTTGTC/Biotine    | Sonde LNA (Exiqon) snARN U6atac RNA-FISH            |
| FISH snARN U6_1_1     | Biotine/ATATGTACCACTGAAGCCAGCA/Biotine | Sonde LNA (Exiqon) snARN U6 RNA-FISH                |
| FISH scaARN2_1_1      | Biotine/TAAGATCAAAGTGAAGCGGGA/Biotine  | Sonde LNA (Exiqon) scaARN2 RNA-FISH                 |
| FISH scaARN2_1_2      | Biotine/TCGTCTATCTGATCAATTCAT/Biotine  | Sonde LNA (Exiqon) scaARN2 RNA-FISH                 |
| FISH scaARN13_1_1     | Biotine/TGGCGGATAGATAATGACATA/Biotine  | Sonde LNA (Exiqon) scaARN13 RNA-FISH                |
| FISH scaARN13_1_2     | Biotine/ACACTCCATCTGCTAAACAAT/Biotine  | Sonde LNA (Exiqon) scaARN13 RNA-FISH                |
| Scramble              | Biotine/GTGTAACACGTCTATACGCCCA/Biotine | Sonde LNA (Exiqon) utilisée comme contrôle RNA-FISH |
| siARN J-019894-09     | CGGAGUGUGCUGCGGGUUU                    | siARN ciblant l'exon 2 de la coiline                |
| siARN J-019894-10     | GUGCAGAGUCAGACGACCA                    | siARN ciblant l'exon 2 de la coiline                |
| siARN J-019894-11     | CCAAUAUAUACCGUGAAA                     | siARN ciblant la 3' UTR de la coiline               |
| siARN J-019894-12     | CCAAGAAUGUGCUGCGUAU                    | siARN ciblant la 3' UTR de la coiline               |

#### 4. Plasmides et constructions génétiques

Les plasmides et leurs dérivés portant des délétions de la coiline (sans exon 3 ou sans exon 4) ont été réalisés pour exprimer et déterminer la localisation des isoformes potentielles de la coiline dans les cellules.

Le plasmide pcDNA3.1-V5-coiline : il a été construit à partir du pcDNA3.1-V5 dans lequel la séquence de la coiline a été clonée en aval de l'étiquette V5.

Les plasmides pEGFP-C1-coiline et pEGFP-N1-coiline : ils ont été construits à partir des plasmides pEGFP-C1 et pEGFP-N1 afin de placer la protéine fluorescente GFP soit en aval, soit en amont de la coiline.

#### 5. Anticorps

Les anticorps utilisés au cours de la thèse sont décrits dans la Table M2.

**Table M2. Désignations, caractéristiques et conditions d'utilisation des anticorps primaires**

| Référence de l'anticorps | Dilution          | Organisme hôte | Fournisseur        |
|--------------------------|-------------------|----------------|--------------------|
| Dyskérine sc-48794       | IF: 1/50 WB:1/250 | Lapin          | Santa Cruz Biotech |
| Fibrillarine 72B9        | IF: 1/200         | Souris         | Jérôme Cavaillé    |
| Fibrillarine 16021-1-AP  | WB: 1/500         | Lapin          | Proteintech        |
| Gar1 sc-26984            | WB: 1/500         | Lapin          | Santa Cruz Biotech |
| Gar1 PTG Lab 11711       | IF: 1/200         | Lapin          | PTG Lab            |
| NHP2 15128-1-AP          | WB: 1/500         | Lapin          | Proteintech        |
| NHP2 PTG Lab 15128       | IF: 1/50          | Lapin          | PTG Lab            |
| NOP10 ab134902           | WB: 1/200         | Lapin          | Abcam              |
| NOP56 18181-1-AP         | WB: 1/500         | Lapin          | Proteintech        |
| NOP56 HPA 049918         | IF:1/50           | Lapin          | Sigma              |
| NOP58 14409-1-AP         | WB: 1/500         | Lapin          | Proteintech        |

| Référence de l'anticorps | Dilution                | Organisme hôte | Fournisseur   |
|--------------------------|-------------------------|----------------|---------------|
| NOP58 ab155969           | WB: 1/1000 IF: 1/200    | Lapin          | Abcam         |
| Coiline ab11822          | IF: 1/800               | Souris         | Abcam         |
| Coiline A303-759A        | IP                      | Lapin          | Bethyl        |
| Coiline A303-760A        | WB: 1/2000 IF: 1/200 IP | Lapin          | Bethyl        |
| SC35 ab11826             | IF: 1/200               | Souris         | Abcam         |
| Gemin2 MA1-57777         | IF: 1/200               | Souris         | G. Dreyfuss   |
| Sumo1                    | IF:1/200                | Lapin          | x             |
| SMN 2B1                  | IF: 1/500 WB:1/1000     | Souris         | G. Dreyfuss   |
| GAPDH ab9485             | WB: 1/5000              | Lapin          | Abcam         |
| 53BP1 PA1-16565          | IF: 1/200               | Lapin          | Thermo Fisher |
| V5 2F11F7                | IF: 1/200               | Souris         | Thermo Fisher |
| GFP ab290                | WB: 1/1000              | Lapin          | Abcam         |
| WRAP53 1-4761-1 AP       | IF: 1/100 WB: 1/500     | Lapin          | Proteintech   |
| TDP-43 ab109535          | WB: 1/1000 IF: 1/250    | Lapin          | Abcam         |
| VRK1 MA1-828             | IF: 1/100 WB: 1/1000    | Souris         | Thermo Fisher |
| PA28γ PA5-21789          | WB: 1/1000 IF:1/200     | Lapin          | Thermo Fisher |
| Y12 ab3138               | WB: 1/500 IP            | Souris         | Abcam         |
| Y12 MA5-13449            | IF:1/100                | Souris         | Invitrogen    |
| U1A                      | IF: 1/100               | Lapin          | Thermo Fisher |
| Sart1                    | IF:1/200                | Lapin          | Thermo Fisher |
| Lsm4                     | IF: 1/200               | Lapin          | X             |
| PRPF31                   | IF: 1/200               | Lapin          | X             |
| ZCRB1                    | IF:1/50                 | Lapin          | X             |

## 6. 2'-O-Methylation par RiboMethSeq

L'identification et la quantification des modifications (2'-O-Methylation) ont été réalisées sur la plateforme Epitranscriptomique et Séquençage de l'UMS2008 (campus Brabois-Santé de l'Université de Lorraine à Nancy) grâce au séquenceur à haut-débit Illumina HiSeq 1000. L'analyse bioinformatique a été réalisée par Y. Motorin grâce aux logiciels « adapter trimming software trimmomatic », « alignment software Bowtie 2.0 » et « R environnement ».

## 7. Micropuces jonctions exon-exon

Les puces utilisées pour l'analyse des profils d'expression des ARNm ainsi que leurs isoformes produites en absence ou suite à un stress oxydant sont des modèles GeneChip™ Cartridge Array Clariom D™ humain (Applied Biosystems™). L'analyse a été réalisée sur scanner GeneChip 3000 avec le logiciel Transcriptome Analysis Console (TCA) (Applied Biosystems™). Les puces contiennent des probesets dirigés contre plus de 948 300 exons et 484 900 jonctions exon-exon. Cela couvre plus de 134 700 gènes et 542 500 transcrits. Les puces Clariom D ont été conçues et annotées à partir de nombreuses sources : Ensembl, VEGA, NONCODE, lncRNAWiki, UCSC Genes, AceView, miTranscriptome, RefSeq, MGC, MGI, RGD, Consensus CDS, RNA Central, circBase, Human Body Map, lincRNAdb, Publication-specific gene.

## 8. Tampons et solutions

**Milieu de culture des cellules HeLa S3** : milieu Dulbecco's Modified Eagle Medium (DMEM) supplémenté avec 10 % de sérum de veau foetal (SVF), 2 mM de Glutamine et 100 U/ml de Penicilline et 100 µg/ml Streptomycine.

**Phosphate Buffered Saline (PBS)**: NaCl 137 mM, KCl 2,7 mM, Na<sub>2</sub>HPO<sub>4</sub>.12H<sub>2</sub>O 8.1 mM, KH<sub>2</sub>PO<sub>4</sub> 1,47 mM. Ajuster le pH à 7,4.

**Tris Buffered Saline (TBS)** : NaCl 150 mM ; Tris 50 mM, pH7.6.

**Solution Saline Sodium Citrate (SSC)** : NaCl 150 mM ; citrate de sodium 15 mM

**Solution hybridation des sondes LNA (RNA-FISH)** : Phosphate de sodium 50 mM pH7,4 ; SSC 2X ; Dextran sulfate 10% (v/v) ; formamide 50% (v/v)

**Solution de blocage (RNA-FISH)** : BSA (bovine serum albumin) 3% (p/v) ; SSC 4X ; Tween 20 0,1%.

**Tampon d'hybridation de la streptavidine (RNA-FISH)** : BSA 1% (p/v) ; SSC 4X ; Tween 20 0,1% (v/v) ; Streptavidine Alexa 488 2 µg/mL.

**Tampon de lyse cellulaire (extraction protéines)** : Tris-HCl 50 mM pH 8, NaCl 150 mM, Igepal 1 % (v/v), EDTA 1 mM, extemporanément Halt™ protease&phosphatase inhibitor cocktail-EDTA free (Thermo scientific) (100X)

**Tampon HNTG** : Hepes 20 mM pH7.9 ; NaCl 150 mM ; Triton X-100 1% ;  $MgCl_2$  1 mM ; EDTA 1 mM ; Glycerol 10% ; extemporanément Halt™ protease&phosphatase inhibitor cocktail-EDTA free (Thermo scientific) (100X)

**Tampon RSB100** : Tris-HCl 10 mM pH7.4 ; NaCl 100mM ;  $MgCl_2$  2,5mM ; extemporanément Igepal 0.1% et Halt™ protease&phosphatase inhibitor cocktail-EDTA free (Thermo scientific) (100X)

**Solution SBL (Sample Buffer Laemmli)** : SDS 2 % (p/v) ; glycérol 10 % (v/v) ;  $\beta$ -mercapto- éthanol 100 mM ; bleu de bromophénol 0,01 % (p/v) ; Tris-HCl 63,65 mM pH 6,8.

**LDS sample buffer (Invitrogen) (pour western blot après Immunoprecipitation)** : Tris HCl 106 mM ; Tris base 141 mM, LDS (Lithium dodecyl sulfate) 2% ; Glycerol 10% ; EDTA 0.51 mM ; Blue G250 0.22 mM ; rouge phénol 0.175 mM ; pH 8.5. Extemporaneément ajouter DTT 0.1 M

**Tampon de migration des protéines** : Tris 25 mM ; Glycine 192 mM ; SDS 0,1%

**Tampon de transfert pour western blot** : Tris 25 mM ; Glycine 192 mM ; Ethanol 20 % (v/v). PBS-T : tampon PBS additionné de 0,1 % de Tween 20

TBS-T : tampon TBS additionné de 0,1% de tween 20

**Solution TBE** : Tris 222,5 mM ; acide borique 89 mM ; EDTA 5 mM

**Bleu formamide** : formamide déionisée additionnée d'EDTA (20 mM), Bleu de bromophénol 0.03%, Bleu de xylène cyanol 0.03%

## II. Méthodes

### 1. Culture cellulaire et Induction du stress oxydant

Les cellules HeLa S3 sont cultivées dans un milieu DMEM supplémenté en sérum de veau à 37°C dans une atmosphère contenant 5 % de  $CO_2$ . Vingt-quatre heures avant l'induction du stress oxydant,  $2 \times 10^6$  cellules sont ensemencées dans des boîtes de culture 10 cm et incubées à 37°C. Les cellules sont traitées avec 500  $\mu M$  de peroxyde d'hydrogène ( $H_2O_2$ ). Avant de récupérer les cellules, le milieu contenant l' $H_2O_2$  est retiré et remplacé par 1mL de milieu sans sérum, celui-ci pouvant générer des ROS additionnels. Les cellules sont ensuite récupérées par grattage et transférées dans un Eppendorf pour extraire soit les ARN, soit les protéines ou pour être analysées en cytométrie en flux.

## 2. Analyse de l'expression des ARN par RT-qPCR (Quantitative Reverse Transcription Polymerase Chain Reaction)

### a) Extraction des ARN

Les cellules sont centrifugées 5 min à 300 rcf et le culot cellulaire est resuspendu dans 1mL de réactif Trizol (Invitrogen). Après l'ajout de 200µL de chloroforme, le mélange est vortexé et centrifugé à 12 000 rcf à 4°C pendant 15min. La phase aqueuse supérieure qui contient les ARNs totaux est récupérée et mélangée à 1µL de RNase free Glycoblue™ (co-précipitant) et 500µL d'isopropanol pour la précipitation. La solution est centrifugée 15min à 12 000 rcf à 4°C et les sels sont éliminés du culot grâce à un lavage avec 1mL d'éthanol 75%. Après une dernière centrifugation de 8min à 12 000 rcf et 4°C, le surnageant est éliminé et le culot d'ARN est dissous dans 10 à 50 µL d'eau Milli-Q. La concentration finale d'ARN est déterminée en mesurant l'Absorbance à 260nm avec le spectrophotomètre Nanodrop 2000. Les potentiels contaminants en protéines et phénol ou sels sont déterminés en calculant respectivement les ratios A260nm/A280nm et A260nm/A230nm. Pour certains protocoles (RiboMethSeq), la qualité des ARN est aussi vérifiée avec le BioAnalyser Agilent 2100.

### b) Digestion de l'ADN par la DNase RQ1 dans les échantillons d'ARN

Cette étape permet d'éliminer les molécules d'ADN qui peuvent potentiellement contaminer les échantillons d'ARN et être ainsi amplifiées au cours de la qPCR. Pour chaque condition, 2 µg d'ARN total sont incubés avec 2 µL de RQ1 (Promega) dans son tampon de réaction dans un volume final de 20 µL 30 min à 37°C. L'activité de la DNase est stoppée en ajoutant 2 µL de RQ1 DNase Stop Solution et en incubant 10 min à 65°C.

### c) Transcription Reverse (Reverse Transcription ou RT)

Après le traitement à la DNase, les ARN totaux sont convertis en ADN complémentaire (ADNc) par un traitement avec la Moloney Murine Leukemia Virus Reverse Transcriptase (M MLV RT) (Promega). Dans chaque tube, des oligonucléotides hexamériques aléatoires 12 µM final) et



des oligodT (6  $\mu$ M final) sont ajoutés dans un volume final de 16  $\mu$ l. L'ensemble du mix est incubé 5 min à 70°C pour défaire les structures secondaires de la matrice et permettre l'hybridation des amorces. Les tubes sont ensuite refroidis dans la glace 5 min pour empêcher la reformation de ces structures secondaires. Pour contrôler l'absence de contamination génomique, chaque échantillon est séparé en deux, l'un sera traité avec la RT (+RT) et l'autre non (-RT). Pour réaliser la rétrotranscription, 200 unités de MML-V, son tampon et 600  $\mu$ M de dNTP sont ajoutés aux ARN (1  $\mu$ g) dans un volume final de 25  $\mu$ l. La réaction se fait à 37°C pendant 1h.

#### d) Détection des produits de RT par PCR quantitative (qPCR)

Les produits de RT sont analysés par qPCR sur l'appareil Step One Plus Real-Time PCR system (Applied Biosystems). Les mix préparés pour chaque condition et le programme de PCR sont présentés dans les Tables M3.A et 3.B, respectivement. En fonction de l'abondance des produits, la dilution de RT utilisée pour la PCR est ajustée. Les différentes dilutions sont présentées dans la Table M4. Le mix de PCR contient la molécule SYBR Green, un intercalant qui fluoresce lorsqu'il s'hybride avec de l'ADN double brin. Pour chaque échantillon, le CT (Cycle Threshold) est déterminé, correspondant au nombre de cycles nécessaires pour détecter la fluorescence des amplicons au-delà d'un certain seuil fixé par l'appareil de qPCR. L'efficacité d'amplification de la PCR a été calculée pour chaque paire de primers en utilisant la méthode de la courbe standard. Ces efficacités sont reportées dans la Table M4. La quantification relative des amplicons a été faite en utilisant la méthode comparative des CT aussi connue sous le nom de méthode  $2^{-\Delta\Delta CT}$  ou en utilisant une courbe standard si les efficacités ne sont pas comprises entre 90 et 110%. Pour ces deux méthodes de quantification, l'expression du gène cible est normalisé sur l'expression des gènes GAPDH, Rplp0, 18S ou 5S dont l'expression ne change pas dans nos conditions de stress.

**Table M3.A. Réaction de qPCR pour la détection de produits spécifiques à partir d'ARN totaux rétrotranscrits (ADNc)**

| Réactifs qPCR                                | Volume (µL) |
|----------------------------------------------|-------------|
| ADNc (dilution table 5)                      | 1           |
| iTaq Universal SYBR Green Supermix (BIO-RAD) | 5           |
| Oligonucléotides 5' (5µM)                    | 1           |
| Oligonucléotides 3' (5µM)                    | 1           |
| H <sub>2</sub> O nuclease-free               | 2           |
| Total                                        | 10          |

**Table M3.B. Programme de qPCR sur Step One Plus Real-Time PCR system (Applied Biosystems)**

| Nombres de cycles | Etapas                | Temperature | Durée  |
|-------------------|-----------------------|-------------|--------|
| 1                 | Dénaturation initiale | 95 °C       | 10 min |
| 45                | Dénaturation          | 95°C        | 10s    |
|                   | Hybridation           | 55 or 58 °C | 10s    |
|                   | Elongation            | 72 °C       | 10s    |
| 1                 | Courbe de fusion      | 95 °C       | 10s    |
|                   |                       | 60 °C       | 5s     |
|                   |                       | 95 °C       | -      |

**Table M4. Dilution des ADNc en fonction des ARN amplifié par qPCR et estimation de leur efficacité en fonction de la température d'hybridation**

| Oligos (à 500nM)     | Dilution | Efficacité à 55°C | Efficacité à 58°C |
|----------------------|----------|-------------------|-------------------|
| snoRD44 (B497/B498)  | 1/9      | 1,66              | 2,13              |
| sca2 (A978/A979)     | ND       | X                 | 2,007             |
| sca8 (7483/7484)     | ND       | 1,57              | 1,69              |
| sca13 (B635/B636)    | ND       | 1,61              | 1,71              |
| sca15 (B637/B638)    | ND       | 1,58              | 2,42              |
| sca17 (A980/A981)    | ND       | 1,96              | 2,08              |
| sca20 (B639/B640)    | ND       | 1,84              | 1,65              |
| snoRA57 (B493/B494)  | ND       | 1,72              | 1,68              |
| snoRA74A (B491/B492) | ND       | 2,04              | 1,7               |
| GAPDH (8536/8537)    | 1/9      | 2,19              | 2,18              |
| 5S (7399/7400)       | 1/81     | 1,96              | 2,01              |
| Rplp0 (8772/8773)    | 1/9      | 1,93              | 1,98              |

|                    |    |      |      |
|--------------------|----|------|------|
| U1 (7392/7393)     | ND | 2,01 | 2,16 |
| U11 (7394/7395)    | ND | X    | 1,82 |
| U12 (7396/7397)    | ND | 1,78 | 1,83 |
| U6 (7566/7567)     | ND | 1,62 | 1,73 |
| U4atac (7568/7569) | ND | 1,59 | 1,64 |
| U2 (8401/8402)     | ND | 2,08 | 2,16 |
| U4 (A051/A052)     | ND | 1,91 | 1,88 |

#### e) Analyse de l'expression du pseudogène coilp1 par PCR en point finale

L'expression relative du transcrit coilp1 a été analysée par PCR en point finale. Le milieu de réaction et le programme de PCR sont décrits respectivement dans les Tables M5.A et 5.B. Le primer forward s'hybride sur l'exon 2 de la coiline alors que le primer reverse s'hybride sur la jonction des exons 5 et 6. Les produits PCR sont fractionnés sur un gel d'agarose 2% et détectés avec du GelRed (Fluoprobes). La quantification est réalisée avec les logiciels d'analyses GeneSnapacquisition et GeneTools.

**Table M5.A. Réaction de PCR pour la détection de produits spécifiques à partir d'ARN totaux retrotranscrits (ADNc)**

| Réactifs PCR                               | Volume (µL) |
|--------------------------------------------|-------------|
| Tampon DreamTaq 10X (Fermentas)            | 5           |
| Mix dNTP (2mM each)                        | 5           |
| Oligonucléotides Forward (20µM)            | 1           |
| Oligonucléotides Reverse (20µM)            | 1           |
| ADNc                                       | 10pg-1µg    |
| DreamTaq DNA Polymerase (5U/µL, Fermentas) | 0.25        |
| H <sub>2</sub> O nuclease-free             | to 50       |

**Table M5.B. Programme de PCR**

| Nombres de cycles | Etape                | Temperature | Durée |
|-------------------|----------------------|-------------|-------|
| 1                 | Denaturation initial | 95°C        | 5min  |
| 30                | Denaturation         | 95°C        | 30sec |
|                   | Hybridation          | 58°C        | 30sec |
|                   | Elongation           | 72°C        | 30sec |
| 1                 | Elongation finale    | 72°C        | 10min |

### 3. Analyse de l'expression des protéines

#### a) Extraction des protéines

Au cours de ma thèse, nos données ont mis en évidence une faible efficacité de l'extraction de la coiline d'un extrait cellulaire « classique » qui restait dans la fraction insoluble. J'ai donc optimisé les conditions d'extraction de cette protéine.

- Extraction avec un tampon de Lyse « classique »

Les cellules sont centrifugées à 300 rcf pendant 5 min et le culot est repris dans 100 µl de tampon de Lyse. Les cellules sont incubées 30 min dans la glace, et vortexées régulièrement. L'extrait protéique est obtenu après une centrifugation de 15 min à 16 000 rcf à 4°C. Le dosage protéique est réalisé par la méthode colorimétrique de Bradford (QuickStart™ Bradford Protein Assay, Bio-rad).

- Extraction avec tampon SBL

Les cellules sont centrifugées à 300 rcf pendant 5 min et le culot est repris directement dans le tampon SBL. Les échantillons sont ensuite soniqués 5 fois 10 secondes et centrifugés au moins 15 minutes à 16 000 rcf et 4°C puis les protéines sont dosées au Qubit.

- Extraction optimale de la coiline

Les cellules sont centrifugées à 300 rcf pendant 5 min et le culot est repris soit dans un tampon HNTG soit dans un tampon RSB100. Les échantillons sont ensuite soniqués 5 fois 10 secondes et centrifugés au moins 15 minutes à 16 000 rcf et 4°C. Afin de vérifier l'efficacité de l'extraction, les culots (partie insoluble) sont repris dans le SBL pour être aussi déposés sur gel. Les protéines dans le surnageant sont dosées par la méthode colorimétrique de Bradford (QuickStart™ Bradford Protein Assay, Bio-rad).

#### b) Electrophorèse et Western Blot

Entre 10 µg et 30 µg de protéines sont fractionnées sur des gels SDS-PAGE de 8 à 12%. Après la migration, elles sont transférées sur membrane de nitrocellulose 0,45 µm dans du tampon de transfert. Les sites de fixation non-spécifiques de protéines sont bloqués grâce à l'incubation pendant 1h à température ambiante et sous agitation dans du tampon PBS-T contenant 5% de lait écrémé (p/v). Après lavage dans du tampon PBS-T, la membrane est incubée avec une dilution d'anticorps primaire diluée dans une solution de PBS, BSA 3%, Azide de Sodium 0.02% pendant 1h à température ambiante et sous agitation. Après 3 lavages de 15 min dans du PBS-T, la membrane est incubée dans une solution d'anticorps secondaire anti IgG lapin ou anti-IgG souris couplé à une peroxydase (Horseradish peroxidase) (1:20 000 dans du PBS-T) pendant 45 min à température ambiante et sous agitation. Après 3 lavages de 15 min dans du PBS-T, les complexes immuns sont détectés grâce aux solutions de Peroxide et Luminol Enhancer (GE Healthcare) et la chemiluminescence est détectée avec le scanner Fusion Solo (Vilber Lourmat). La quantification de l'expression des protéines est calculée grâce au logiciel Fusion en normalisant l'intensité des bandes avec l'intensité obtenue pour la protéine GAPDH dont l'expression ne varie pas dans nos conditions expérimentales.

#### 4. Immunoprécipitation

Les immunoprécipitations ont été mises en place pour l'analyse des variations d'interactions de la coiline avec ces partenaires protéiques en condition de stress oxydant. Cette technique a également été utilisée pour étudier l'impact du stress oxydant sur le taux de UsnRNA impliqués dans la formation de UsnRNPs (RIP contre les protéines Sm B/B').

##### a) Préparation des billes magnétiques

Les billes magnétiques Dynabeads® Protein A ou G (75µL) (life technologies) sont tout d'abord lavées 3 fois dans du tampon HNTG et incubées avec 10 µg d'anticorps lapin anti-coiline (Bethyl A303 759A) ou d'anticorps souris anti-Sm B/B' (Y12) dilué dans 300 µL de tampon HNTG pendant 1 h à 4 °C sous agitation constante. Les billes sont ensuite lavées 3 fois dans le même tampon afin de retirer les anticorps non fixés.

**b) Préparation des extraits cellulaires et immunosélection**

Un extrait cellulaire total est préparé à partir de 3 à 10 boîtes de culture 10 cm confluentes à 80%. Les culots cellulaires sont repris dans 500 µL de tampon de lyse HNTG et incubés pendant 20min dans la glace ou soniqués 5X 10sec à 0.5 output dans la glace (conditions optimisation de l'extraction de la coiline). Les échantillons sont ensuite centrifugés 15 min à 12 000 rcf à 4°C. Le surnageant est récupéré et 5% (25 µL) de cet extrait est conservé (fraction Input). Ensuite, les billes fraîchement préparées sont incubées avec le reste de l'extrait cellulaire pendant au moins 2h à 4°C et sous agitation constante. En parallèle, un contrôle négatif est réalisé en incubant un extrait cellulaire avec des billes sans anticorps. Les protéines non immuno-sélectionnées sont retirées par 3 lavages de 5min avec du tampon HNTG. Si les ARN et les protéines doivent être récupérés, les billes sont séparées en 2 fractions. Pour les protéines, les billes sont incubées dans 25 µL de LDS sample buffer pendant 5min à 96°C afin de récupérer les protéines avant leur fractionnement sur gel SDS-PAGE et l'analyse par Western Blot. Pour les ARN, les billes sont incubées dans 500 µl de Trizol.

**c) Réduction du signal des chaînes lourdes d'anticorps en Western Blot**

Lorsque les complexes immuns ont été révélés avec des anticorps dirigés contre les IgG de Lapin, les Western Blot ont été réalisés en utilisant le kit commercial ReliaBlot® afin de diminuer le signal des chaînes lourdes et légères de l'anticorps lapin anti-coiline. Le lysat d'IP et les anticorps sont réduits avec du DTT ajouté au LDS sample buffer. L'anticorps secondaire couplé à l'HRP utilisé dans ce kit n'est pas capable de reconnaître la forme réduite des chaînes de l'anticorps primaire lapin. La membrane a été bloquée dans une solution de blocage ReliaBLOT® Block solution pendant 1h et sous agitation douce à température ambiante puis incubée pendant 2h dans les mêmes conditions avec l'anticorps primaire dilué dans cette même solution de blocage. Après 3 lavages dans du tampon TBS-T la membrane a été incubée 1h avec l'anticorps secondaire (ReliaBLOT® HRP Conjugate secondary antibody (1:20 000)) dilué dans la solution de blocage ReliaBLOT® Block solution. Enfin, la membrane est lavée 3 fois dans le TBS-T avant l'étape de révélation.

#### d) Détection des U snRNA en gel dénaturant

Après extraction des ARN, 5 µl de chaque échantillon (sur 10µl) sont mélangés avec 2 µl de bleu de formamide. Les ARN sont dénaturés 2 min à 96°C puis immédiatement placés dans la glace. Ils sont ensuite déposés sur un gel acryl-urée 10% dans un tampon TBE 0,5X. Après migration (1h15 à 30 mA), le gel est rincé 2 fois dans du tampon TBE 0,5X, puis incubé dans un tampon TBE 0,5X avec du SYBR Gold pendant 15 min sous agitation. Le gel est ensuite rincé avec le tampon TBE 0,5X et placé dans un film plastique pour être scanné au Typhoon 9400 (laser à 488nm, filtre d'émission 520 BP 40).

#### 5. Cartographie et quantification des 2'-O-méthylation des U snRNA et ARNr par le protocole Illumina RiboMethSeq

Différents échantillons ont été analysés par cette technique. Les modifications des U snRNA ont été recherchées après les immunoprécipitations avec l'anticorps anti-SmB/B' en conditions de stress oxydant ou en conditions témoin. Les modifications des ARNr ont été recherchées dans des échantillons déplétés ou non de NOP58. Cette partie expérimentale a été réalisée par le personnel de la plateforme épitranscriptomique et séquençage de l'UMS2008, elle sera décrite ici brièvement. Après avoir vérifié la qualité des ARN, ils subissent une hydrolyse alcaline, suivie d'une phosphorylation en 5'. Après purification des ARN avec le kit Rneasy MinElute Cleanup kit, une bibliothèque est constituée avec des adaptateurs spécifique en utilisant le kit NEB Multiplex Small RNA Library Prep Set for Illumina. Les produits sont ensuite purifiés avec le kit GeneJET PCR purification. Après un contrôle qualité les bibliothèques sont séquencées. Les données brutes ont été traitées par Yuri Motorin et avec Isabelle Behm-Ansmant nous avons réalisé l'analyse des résultats.

#### 6. Analyse du cycle cellulaire

Après 5min de centrifugation à 300 rcf, le culot cellulaire est lavé dans 400µL de PBS. Pour la précipitation, 900 µL d'éthanol 100% froid est ajouté goutte à goutte dans les 400 µL sous agitation douce pour empêcher la formation d'agrégats, puis les cellules sont placées toute la nuit à -20°C. Après 3 lavages dans du PBS, les cellules sont incubées dans 500 µL d'une solution de

PBS, RNase A (100µg/mL) et d'iodure de Propidium (PI) (50 µg/mL), 30 min à 37°C à l'abri de la lumière. Le PI est un intercalant utilisé pour quantifier le contenu d'ADN dans les expériences d'analyse du cycle cellulaire. Pour chaque échantillon, environ 20 000 cellules sont analysées avec le cytomètre en flux Gallios (Beckman Coulter) et la proportion de cellules présentes dans chaque phase du cycle est estimée en traitant les données avec le logiciel Gallios 1.2.

## 7. Analyse *in cellulo* de la localisation des protéines et des ARNs

### a) Analyse *in cellulo* de la localisation des ARN par Hybridation In Situ de Fluorescence (RNA-FISH)

Les cellules sont initialement cultivées sur les lames de microscope placées dans les boîtes de culture 10 cm. Après le stress oxydant, les lames sont lavées dans du PBS à 37°C et les cellules sont fixées dans une solution de PBS et 4% de Formaldéhyde pendant 10 min. Après un lavage dans du PBS, les lames sont incubées dans une solution de Tris-HCl 0.1 M, pH 7.5 pendant 5 min puis placées dans du PBS, Glycérol 20 % pendant 1 h. Pour perméabiliser les cellules, les lames sont directement plongées dans de l'azote liquide pendant quelques secondes puis séchées à température ambiante avant d'être placées une nouvelle fois dans la solution de PBS Glycérol. Ce cycle de réfrigération et séchage est répété deux fois supplémentaires. Les cellules sont ensuite déshydratées au cours d'incubations séquentielles dans des solutions d'éthanol de plus en plus concentrées : 70%, 90% et 100%. Les sondes oligonucléotides LNA (Locked Nucleic Acid) biotinylées (Table 1) sont diluées à 30 nM et sont dénaturées à 75°C pendant 5 min. Les cellules déshydratées sont alors incubées avec les sondes dénaturées pendant 30 min à 37°C ou 54°C en fonction des sondes. La technologie LNA consiste à ajouter un pont méthylène (-CH<sub>3</sub>-O-) entre le ribose 2' et le carbone 4'. Dans une double hélice d'acide nucléique, les résidus riboses des nucléotides peuvent passer de façon aléatoire de la conformation S à N et inversement, déstabilisant les interactions entre les deux brins. La présence du pont méthylène permet de figer la conformation du ribose dans une conformation favorable à la formation de paires de bases Watson-Crick, stabilisant fortement les interactions entre les oligonucléotides LNA et leur cible, et augmentant aussi la spécificité de leur interaction. Les lames sont ensuite lavées 3 fois dans une solution de SSC 2X, Tween 20 0.1 % pré-chauffée à 65 °C. Après incubation dans la solution



de blocage 30 min à 37 °C, les cellules sont incubées dans le tampon d'hybridation contenant la Streptavidine conjuguée à l'Alexa fluor 488, pendant 30 min à 37 °C dans une chambre protégée de la lumière. Les lames sont lavées 3 fois dans une solution de SSC 4X, Tween 20 0.1 % préchauffée à 45 °C. A partir de cette étape, les lames sont soit directement montées afin d'être analysées, ou lavées dans du PBS afin de poursuivre avec les étapes d'Immunofluorescence.

#### b) Analyse *in cellulo* de la localisation des protéines par Immunofluorescence (IF)

Comme précédemment, les cellules qui ont adhéré sur les lamelles de microscope sont lavées dans du PBS à 37°C après le stress oxydant, puis fixées covalamment pendant 10 min dans la solution de PBS Formaldéhyde 4 %. Après lavage dans du PBS, les cellules sont perméabilisées dans une solution de Triton X-100 0,5%, NH<sub>4</sub>Cl 50 mM, PBS pendant 15 min. Les cellules sont ensuite lavées 3 fois dans du PBS et incubées avec les anticorps primaires dilués dans du PBS, BSA 1% toute la nuit à 4°C. Après 3 lavages au PBS, les cellules sont incubées 45 min dans une chambre à l'abri de la lumière avec l'anticorps secondaire (1:200) couplé à un fluorophore spécifique et dilué dans une solution de PBS, BSA 1%. Après 3 lavages au PBS, les lamelles sont montées avec le milieu de montage Duolink Mounting Medium contenant du 4',6-diamidino-2-phenylindole (DAPI). Les cellules sont ensuite observées par épifluorescence (Nikon Optiphot 2 Fluorescence microscope) ou en microscopie confocale (SP5-X microscope confocale Leica).

### 8. Transfection des cellules

#### a) Surexpression de plasmides

Afin de suivre la localisation de différentes isoformes de coiline, les cellules sont transfectées avec des plasmides exprimant les isoformes de coiline couplées à la protéine fluorescente GFP ou le tag V5. Pour cela, les cellules sontensemencées dans des plaques 6 puits à 300 000 cellules par puits ou dans des boîtes de 10 cm à 2X10<sup>6</sup> cellules par boîte. Le lendemain, les cellules sont transfectées avec des constructions plasmidiques (500 ng pour les 6 puits ou 3µg pour les boîtes 10 cm) en utilisant le kit JetPEI (PolyPlus transfection) ou le kit lipofectamine 2000 (invitrogen), selon les indications des fournisseurs. Après 24h, les cellules sont récoltées pour analyse protéique ou fixées pour analyse en immunofluorescence.

### b) Extinction de l'expression de la coiline ou de NOP58

Pour éteindre l'expression de ces protéines, les cellules sont transfectées avec 25nM de siRNA ciblant l'ARNm coiline ou l'ARNm NOP58 en utilisant la méthode de transfection au phosphate de calcium. Des siRNA sans cible sont utilisés comme contrôle, Non-targeting siRNA pool (NT). Les séquences ARN ciblées sont listées dans la table 1. La veille de la transfection, les cellules HeLa S3 sontensemencées dans des plaques 6 puits à 150 000 cellules par puits. Trois ou quatre heures avant la transfection, le milieu est changé. Les siRNA sont mélangés avec du Chlorure de calcium (125mM final) et du HBS pH7.05 et incubés 20 min à température ambiante avant d'être ajoutés dans les puits. Le lendemain, le milieu est changé pour enlever les précipités. Soixante-douze heures après la transfection, les cellules sont récupérées pour analyse.

## 9. Analyse transcriptomique par puces à jonctions exon-exon

Cette partie expérimentale a été réalisée par Quentin Thuilier, l'ingénieur d'études responsable de la station affymetrix, elle sera décrite ici brièvement. Afin de vérifier la qualité des ARN extraits selon le protocole cité précédemment, ils sont soumis à une analyse sur des puces de BioAnalyser Agilent 2100 (Eukaryote Total RNA Nano 2.6, Agilent Technologie). Les RIN (RNA Integrity Number) obtenus étaient compris entre 9 et 10, reflétant l'absence de dégradation de nos ARN.

La préparation des ARN est réalisée suivant le protocole « GeneChip™ WT Pico Reagent Kit ». Les ARN sont rétrotranscrits de manière à ajouter en étiquette la séquence du promoteur T7 en 3' des ADNc. Après la synthèse du second brin et la dégradation des ARN par la RNase H, l'ARNc anti-sens est synthétisé grâce au promoteur T7. L'ARNc est rétrotranscrit en présence de dUTP en ADNc, puis éliminé par traitement à la RNase H. Après une étape de purification, les ADNc sens sont fragmentés et biotinylés en 3' puis marqués par un fluorophore (Streptavidine-phycoérythrine). Ce sont ces fragments marqués qui sont hybridés sur les puces jonctions exon-exon. Nous avons analysé 5 puces pour la condition contrôle et 5 puces pour les cellules soumises à un stress oxydant de 4h.

Le traitement des données a été réalisé sur la console d'analyse d'Affymetrix (TAC pour Tools Analysis Console) par Isabelle Behm-Ansmant. La version TAC 4.0.2 de la console a été

utilisée ; cette version repose sur l'utilisation du package bioconducteur Limma et les paramètres par défaut ont été utilisés. Brièvement, le signal des puces a été normalisé en utilisant la méthode SST-RMA, l'annotation des probesets est faite sur la base du fichier d'annotation « Clariom\_D\_Human.r1.na36.hg38.a1.transcript.csv » d'Affymetrix. L'expression différentielle est calculée avec le package Limma et, compte tenu du faible nombre d'échantillons, la correction eBayes est apportée aux résultats de l'analyse ANOVA. L'algorithme EvenPointer est utilisé pour l'analyse de l'épissage alternatif. Le fonctionnement détaillé des différents algorithmes utilisés est disponible dans l'annexe 2.

# BIBLIOGRAPHIE



- Alam, M.M., Okazaki, K., Nguyen, L.T.T., Ota, N., Kitamura, H., Murakami, S., Shima, H., Igarashi, K., Sekine, H., and Motohashi, H. (2017). Glucocorticoid receptor signaling represses the antioxidant response by inhibiting histone acetylation mediated by the transcriptional activator NRF2. *J. Biol. Chem.* 292, 7519–7530.
- Alberti, S., Halfmann, R., King, O., Kapila, A., and Lindquist, S. (2009). A systematic survey identifies prions and illuminates sequence features of prionogenic proteins. *Cell* 137, 146–158.
- Anderson, P., and Kedersha, N. (2008). Stress granules: the Tao of RNA triage. *Trends Biochem. Sci.* 33, 141–150.
- Andrade, L.E., Tan, E.M., and Chan, E.K. (1993). Immunocytochemical analysis of the coiled body in the cell cycle and during cell proliferation. *Proc. Natl. Acad. Sci. U. S. A.* 90, 1947–1951.
- Anufrieva, K.S., Shender, V.O., Arapidi, G.P., Pavlyukov, M.S., Shakhparonov, M.I., Shnaider, P.V., Butenko, I.O., Lagarkova, M.A., and Govorun, V.M. (2018). Therapy-induced stress response is associated with downregulation of pre-mRNA splicing in cancer cells. *Genome Med.* 10, 49.
- Avendaño-Vázquez, S.E., Dhir, A., Bembich, S., Buratti, E., Proudfoot, N., and Baralle, F.E. (2012). Autoregulation of TDP-43 mRNA levels involves interplay between transcription, splicing, and alternative polyA site selection. *Genes Dev.* 26, 1679–1684.
- Ayala, Y.M., Zago, P., D'Ambrogio, A., Xu, Y.-F., Petrucelli, L., Buratti, E., and Baralle, F.E. (2008). Structural determinants of the cellular localization and shuttling of TDP-43. *J. Cell Sci.* 121, 3778–3785.
- Ayala, Y.M., De Conti, L., Avendaño-Vázquez, S.E., Dhir, A., Romano, M., D'Ambrogio, A., Tollervey, J., Ule, J., Baralle, M., Buratti, E., et al. (2011). TDP-43 regulates its mRNA levels through a negative feedback loop. *EMBO J.* 30, 277–288.
- Banani, S.F., Lee, H.O., Hyman, A.A., and Rosen, M.K. (2017). Biomolecular condensates: organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol.* 18, 285–298.
- Bartys, N., Kierzek, R., and Lisowiec-Wachnicka, J. (2019). The regulation properties of RNA secondary structure in alternative splicing. *Biochim. Biophys. Acta Gene Regul. Mech.* 194401–194401.
- Berg, M.G., Singh, L.N., Younis, I., Liu, Q., Pinto, A.M., Kaida, D., Zhang, Z., Cho, S., Sherrill-Mix, S., Wan, L., et al. (2012). U1 snRNP determines mRNA length and regulates isoform expression. *Cell* 150, 53–64.
- Bernardi, R., and Pandolfi, P.P. (2007). Structure, dynamics and functions of promyelocytic leukaemia nuclear bodies. *Nat. Rev. Mol. Cell Biol.* 8, 1006–1016.
- Bernardi, R., Papa, A., and Pandolfi, P.P. (2008). Regulation of apoptosis by PML and the PML-NBs. *Oncogene* 27, 6299–6312.

- Biamonti, G., and Caceres, J.F. (2009). Cellular stress and RNA splicing. *Trends Biochem. Sci.* *34*, 146–153.
- Boisvert, F.-M., van Koningsbruggen, S., Navascués, J., and Lamond, A.I. (2007). The multifunctional nucleolus. *Nat. Rev. Mol. Cell Biol.* *8*, 574–585.
- Bordonné, R. (2000). Functional characterization of nuclear localization signals in yeast Sm proteins. *Mol. Cell. Biol.* *20*, 7943–7954.
- Boulon, S., Westman, B.J., Hutten, S., Boisvert, F.-M., and Lamond, A.I. (2010). The Nucleolus under Stress. *Mol. Cell* *40*, 216–227.
- Braunschweig, U., Gueroussov, S., Plocik, A.M., Graveley, B.R., and Blencowe, B.J. (2013). Dynamic Integration of Splicing within Gene Regulatory Pathways. *Cell* *152*, 1252–1269.
- Broome, H.J., and Hebert, M.D. (2013). Coilin Displays Differential Affinity for Specific RNAs In Vivo and Is Linked to Telomerase RNA Biogenesis. *J. Mol. Biol.* *425*, 713–724.
- Buratti, E., Dörk, T., Zuccato, E., Pagani, F., Romano, M., and Baralle, F.E. (2001). Nuclear factor TDP-43 and SR proteins promote in vitro and in vivo CFTR exon 9 skipping. *EMBO J.* *20*, 1774–1784.
- C. Andrade (1991). Human autoantibody to a novel protein of the nuclear coiled body: immunological characterization and cDNA cloning of p80-coilin. *J. Exp. Med.* *173*, 1407–1419.
- Caldas, C., So, C.W., MacGregor, A., Ford, A.M., McDonald, B., Chan, L.C., and Wiedemann, L.M. (1998). Exon scrambling of MLL transcripts occur commonly and mimic partial genomic duplication of the gene. *Gene* *208*, 167–176.
- Cantarero, L., Sanz-García, M., Vinograd-Byk, H., Renbaum, P., Levy-Lahad, E., and Lazo, P.A. (2015). VRK1 regulates Cajal body dynamics and protects coilin from proteasomal degradation in cell cycle. *Sci. Rep.* *5*, 10543.
- Carbonell, C., Ulsamer, A., Vivori, C., Papasaikas, P., Böttcher, R., Joaquin, M., Miñana, B., Tejedor, J.R., de Nadal, E., Valcárcel, J., et al. (2019). Functional Network Analysis Reveals the Relevance of SKIIP in the Regulation of Alternative Splicing by p38 SAPK. *Cell Rep.* *27*, 847–859.e6.
- Carmo-Fonseca (1992). Transcription-dependent colocalization of the U1, U2, U4/U6, and U5 snRNPs in coiled bodies. *J. Cell Biol.* *117*, 1–14.
- Carmo-Fonseca, M., Pepperkok, R., Sproat, B.S., Ansorge, W., Swanson, M.S., and Lamond, A.I. (1991). In vivo detection of snRNP-rich organelles in the nuclei of mammalian cells. *EMBO J.* *10*, 1863–1873.

- Carmo-Fonseca, M., Ferreira, J., and Lamond, A.I. (1993). Assembly of snRNP-containing coiled bodies is regulated in interphase and mitosis--evidence that the coiled body is a kinetic nuclear structure. *J. Cell Biol.* **120**, 841–852.
- Carrero, Z.I., Velma, V., Douglas, H.E., and Hebert, M.D. (2011). Coilin Phosphomutants Disrupt Cajal Body Formation, Reduce Cell Proliferation and Produce a Distinct Coilin Degradation Product. *PLoS ONE* **6**, e25743.
- de Chaumont, F., Dallongeville, S., Chenouard, N., Hervé, N., Pop, S., Provoost, T., Meas-Yedid, V., Pankajakshan, P., Lecomte, T., Le Montagner, Y., et al. (2012). Icy: an open bioimage informatics platform for extended reproducible research. *Nat. Methods* **9**, 690–696.
- Cheah, M.T., Wachter, A., Sudarsan, N., and Breaker, R.R. (2007). Control of alternative RNA splicing and gene expression by eukaryotic riboswitches. *Nature* **447**, nature05769.
- Chen, H.-C., and Cheng, S.-C. (2012). Functional roles of protein splicing factors. *Biosci. Rep.* **32**, 345–359.
- Chen, Y.-C.M., Kappel, C., Beaudouin, J., Eils, R., and Spector, D.L. (2008). Live cell dynamics of promyelocytic leukemia nuclear bodies upon entry into and exit from mitosis. *Mol. Biol. Cell* **19**, 3147–3162.
- Cheng, Z., Shang, Y., Gao, S., and Zhang, T. (2017). Overexpression of U1 snRNA induces decrease of U1 spliceosome function associated with Alzheimer’s disease. *J. Neurogenet.* **31**, 1–7.
- Chujo, T., Yamazaki, T., and Hirose, T. (2016). Architectural RNAs (arcRNAs): A class of long noncoding RNAs that function as the scaffold of nuclear bodies. *Biochim. Biophys. Acta* **1859**, 139–146.
- Cioce, M., Boulon, S., Matera, A.G., and Lamond, A.I. (2006). UV-induced fragmentation of Cajal bodies. *J. Cell Biol.* **175**, 401–413.
- Coady, T.H., and Lorson, C.L. (2011). SMN in spinal muscular atrophy and snRNP biogenesis. *Wiley Interdiscip. Rev. RNA* **2**, 546–564.
- Collier, S., Pendle, A., Boudonck, K., van Rij, T., Dolan, L., and Shaw, P. (2006). A distant coilin homologue is required for the formation of cajal bodies in Arabidopsis. *Mol. Biol. Cell* **17**, 2942–2951.
- Cooke, M.S., Evans, M.D., Dizdaroglu, M., and Lunec, J. (2003). Oxidative DNA damage: mechanisms, mutation, and disease. *FASEB J.* **17**, 1195–1214.
- Cooper, D.R., Carter, G., Li, P., Patel, R., Watson, J.E., and Patel, N.A. (2014). Long Non-Coding RNA NEAT1 Associates with SRp40 to Temporally Regulate PPAR $\gamma$ 2 Splicing during Adipogenesis in 3T3-L1 Cells. *Genes* **5**, 1050–1063.



- Cote, G.J., Zhu, W., Thomas, A., Martin, E., Murad, F., and Sharina, I.G. (2012). Hydrogen peroxide alters splicing of soluble guanylyl cyclase and selectively modulates expression of splicing regulators in human cancer cells. *PLoS One* 7, e41099.
- Courchaine, E.M., Lu, A., and Neugebauer, K.M. (2016). Droplet organelles? *EMBO J.* 35, 1603–1612.
- D'Ambrogio, A., Buratti, E., Stuani, C., Guarnaccia, C., Romano, M., Ayala, Y.M., and Baralle, F.E. (2009). Functional mapping of the interaction between TDP-43 and hnRNP A2 in vivo. *Nucleic Acids Res.* 37, 4116–4126.
- Daoud, R., Mies, G., Smialowska, A., Oláh, L., Hossmann, K.-A., and Stamm, S. (2002). Ischemia Induces a Translocation of the Splicing Factor tra2- $\beta$ 1 and Changes Alternative Splicing Patterns in the Brain. *J. Neurosci.* 22, 5889–5899.
- Darzacq, X., Kittur, N., Roy, S., Shav-Tal, Y., Singer, R.H., and Meier, U.T. (2006). Stepwise RNP assembly at the site of H/ACA RNA transcription in human cells. *J. Cell Biol.* 173, 207–218.
- D'Autréaux, B., and Toledano, M.B. (2007). ROS as signalling molecules: mechanisms that generate specificity in ROS homeostasis. *Nat. Rev. Mol. Cell Biol.* 8, 813–824.
- Davalli, P., Mitic, T., Caporali, A., Lauriola, A., and D'Arca, D. (2016). ROS, Cell Senescence, and Novel Molecular Mechanisms in Aging and Age-Related Diseases. *Oxid. Med. Cell. Longev.* 2016.
- Decker, C.J., Teixeira, D., and Parker, R. (2007). Edc3p and a glutamine/asparagine-rich domain of Lsm4p function in processing body assembly in *Saccharomyces cerevisiae*. *J. Cell Biol.* 179, 437–449.
- Dellaire, G., Eskiw, C.H., Dehghani, H., Ching, R.W., and Bazett-Jones, D.P. (2006). Mitotic accumulations of PML protein contribute to the re-establishment of PML nuclear bodies in G1. *J. Cell Sci.* 119, 1034–1042.
- Deryusheva, S., and Gall, J.G. (2013). Novel small Cajal-body-specific RNAs identified in *Drosophila*: probing guide RNA function. *RNA N. Y. N* 19, 1802–1814.
- Disher, K., and Skandalis, A. (2007). Evidence of the modulation of mRNA splicing fidelity in humans by oxidative stress and p53. *Genome* 50, 946–953.
- Dominguez, C.E., Cunningham, D., and Chandler, D.S. (2017). SMN regulation in SMA and in response to stress: new paradigms and therapeutic possibilities. *Hum. Genet.* 136, 1173–1191.
- Dujardin, G., Lafaille, C., de la Mata, M., Marasco, L.E., Muñoz, M.J., Le Jossic-Corcos, C., Corcos, L., and Kornblihtt, A.R. (2014). How slow RNA polymerase II elongation favors alternative exon skipping. *Mol. Cell* 54, 683–690.

- Dundr, M., and Misteli, T. (2010). Biogenesis of Nuclear Bodies. *Cold Spring Harb. Perspect. Biol.* 2.
- Dundr, M., Ospina, J.K., Sung, M.-H., John, S., Upender, M., Ried, T., Hager, G.L., and Matera, A.G. (2007). Actin-dependent intranuclear repositioning of an active gene locus in vivo. *J. Cell Biol.* 179, 1095–1103.
- Edmond, V., Moysan, E., Khochbin, S., Matthias, P., Brambilla, C., Brambilla, E., Gazzeri, S., and Eymin, B. (2011). Acetylation and phosphorylation of SRSF2 control cell fate decision in response to cisplatin. *EMBO J.* 30, 510–523.
- Everett, R.D. (2006). Interactions between DNA viruses, ND10 and the DNA damage response. *Cell. Microbiol.* 8, 365–374.
- Falahati, H., Pelham-Webb, B., Blythe, S., and Wieschaus, E. (2016). Nucleation by rRNA Dictates the Precision of Nucleolus Assembly. *Curr. Biol. CB* 26, 277–285.
- Fontana, G.A., Rigamonti, A., Lenzken, S.C., Filosa, G., Alvarez, R., Calogero, R., Bianchi, M.E., and Barabino, S.M.L. (2017). Oxidative stress controls the choice of alternative last exons via a Brahma–BRCA1–CstF pathway. *Nucleic Acids Res.* 45, 902–914.
- Foreman, J., Demidchik, V., Bothwell, J.H.F., Mylona, P., Miedema, H., Torres, M.A., Linstead, P., Costa, S., Brownlee, C., Jones, J.D.G., et al. (2003). Reactive oxygen species produced by NADPH oxidase regulate plant cell growth. *Nature* 422, 442–446.
- Frey, M.R., Bailey, A.D., Weiner, A.M., and Matera, A.G. (1999). Association of snRNA genes with coiled bodies is mediated by nascent snRNA transcripts. *Curr. Biol. CB* 9, 126–135.
- Gall, J.G. (2000). Cajal bodies: the first 100 years. *Annu. Rev. Cell Dev. Biol.* 16, 273–300.
- Garbe, J.C., Bendena, W.G., Alfano, M., and Pardue, M.L. (1986). A *Drosophila* heat shock locus with a rapidly diverging sequence but a conserved structure. *J. Biol. Chem.* 261, 16889–16894.
- Geiszt, M., and Leto, T.L. (2004). The Nox family of NAD(P)H oxidases: host defense and beyond. *J. Biol. Chem.* 279, 51715–51718.
- Gerald, D., Yaniv, M., and Mechta-Grigoriou, F. (2005). Stress oxydatif et angiogenèse : rôle majeur d'AP-1. *MS Médecine Sci.* 21, 233–235.
- Gerdes, J., Schwab, U., Lemke, H., and Stein, H. (1983). Production of a mouse monoclonal antibody reactive with a human nuclear antigen associated with cell proliferation. *Int. J. Cancer* 31, 13–20.
- Gerdes, J., Lemke, H., Baisch, H., Wacker, H.H., Schwab, U., and Stein, H. (1984). Cell cycle analysis of a cell proliferation-associated human nuclear antigen defined by the monoclonal antibody Ki-67. *J. Immunol. Baltim. Md* 1950 133, 1710–1715.

- Gonzalez, I., Munita, R., Agirre, E., Dittmer, T.A., Gysling, K., Misteli, T., and Luco, R.F. (2015). A lncRNA regulates alternative splicing via establishment of a splicing-specific chromatin signature. *Nat. Struct. Mol. Biol.*
- Govoni, M., Farabegoli, F., Pession, A., and Novello, F. (1994). Inhibition of Topoisomerase II Activity and Its Effect on Nucleolar Structure and Function. *Exp. Cell Res.* *211*, 36–41.
- de la Grange, P., Dutertre, M., Martin, N., and Auboeuf, D. (2005). FAST DB: a website resource for the study of the expression regulation of human gene products. *Nucleic Acids Res.* *33*, 4276–4284.
- Gridasova, A.A., and Henry, R.W. (2005). The p53 tumor suppressor protein represses human snRNA gene transcription by RNA polymerases II and III independently of sequence-specific DNA binding. *Mol. Cell. Biol.* *25*, 3247–3260.
- Guirio, J., and O'Reilly, D. (2015). Insights into the U1 small nuclear ribonucleoprotein complex superfamily. *Wiley Interdiscip. Rev. RNA* *6*, 79–92.
- Gunderson, F.Q., and Johnson, T.L. (2009). Acetylation by the transcriptional coactivator Gcn5 plays a novel role in co-transcriptional spliceosome assembly. *PLoS Genet.* *5*, e1000682.
- Hallais, M., Pontvianne, F., Andersen, P.R., Clerici, M., Lener, D., Benbahouche, N.E.H., Gostan, T., Vandermoere, F., Robert, M.-C., Cusack, S., et al. (2013). CBC-ARS2 stimulates 3'-end maturation of multiple RNA families and favors cap-proximal processing. *Nat. Struct. Mol. Biol.* *20*, 1358–1366.
- Handwerger, K.E., and Gall, J.G. (2006). Subnuclear organelles: new insights into form and function. *Trends Cell Biol.* *16*, 19–26.
- Handwerger, K.E., Wu, Z., 'an, Murphy, C., and Gall, J.G. (2002). Heat shock induces mini-Cajal bodies in the *Xenopus* germinal vesicle. *J. Cell Sci.* *115*, 2011–2020.
- Hausmann, I.U., Bodi, Z., Sanchez-Moran, E., Mongan, N.P., Archer, N., Fray, R.G., and Soller, M. (2016). m6A potentiates Sxl alternative pre-mRNA splicing for robust *Drosophila* sex determination. *Nature* *540*, 301–304.
- Hazra, T.K., Izumi, T., Maidt, L., Floyd, R.A., and Mitra, S. (1998). The presence of two distinct 8-oxoguanine repair enzymes in human cells: their potential complementary roles in preventing mutation. *Nucleic Acids Res.* *26*, 5116–5122.
- Henriksson, S., Rassoolzadeh, H., Hedström, E., Coucoravas, C., Julner, A., Goldstein, M., Imreh, G., Zhivotovsky, B., Kastan, M.B., Helleday, T., et al. (2014). The scaffold protein WRAP53 $\beta$  orchestrates the ubiquitin response critical for DNA double-strand break repair. *Genes Dev.* *28*, 2726–2738.

- Hernandez-Verdun, D. (2006). The nucleolus: a model for the organization of nuclear functions. *Histochem. Cell Biol.* *126*, 135–148.
- Hiller, M., Zhang, Z., Backofen, R., and Stamm, S. (2007). Pre-mRNA Secondary Structures Influence Exon Recognition. *PLOS Genet.* *3*, e204.
- Hoskins, A.A., and Moore, M.J. (2012). The spliceosome: a flexible, reversible macromolecular machine. *Trends Biochem. Sci.* *37*, 179–188.
- Hsin, J.-P., and Manley, J.L. (2012). The RNA polymerase II CTD coordinates transcription and RNA processing. *Genes Dev.* *26*, 2119–2137.
- Huang, S., and Spector, D.L. (1992). U1 and U2 small nuclear RNAs are present in nuclear speckles. *Proc. Natl. Acad. Sci. U. S. A.* *89*, 305–308.
- Huang, J., Dong, X., Gong, Z., Qin, L.-Y., Yang, S., Zhu, Y.-L., Wang, X., Zhang, D., Zou, T., Yin, P., et al. (2019). Solution structure of the RNA recognition domain of METTL3-METTL14 N6-methyladenosine methyltransferase. *Protein Cell* *10*, 272–284.
- Ikwegbue, P.C., Masamba, P., Oyinloye, B.E., and Kappo, A.P. (2018). Roles of Heat Shock Proteins in Apoptosis, Oxidative Stress, Human Inflammatory Diseases, and Cancer. *Pharmaceuticals* *11*, 2.
- Isaac, C., Yang, Y., and Thomas Meier, U. (1998). Nopp140 Functions as a Molecular Link Between the Nucleolus and the Coiled Bodies. *J. Cell Biol.* *142*, 319–329.
- Izumikawa, K., Nobe, Y., Ishikawa, H., Yamauchi, Y., Taoka, M., Sato, K., Nakayama, H., Simpson, R.J., Isobe, T., and Takahashi, N. (2019). TDP-43 regulates site-specific 2'-O-methylation of U1 and U2 snRNAs via controlling the Cajal body localization of a subset of C/D scaRNAs. *Nucleic Acids Res.* *47*, 2487–2505.
- Jády, B.E., Bertrand, E., and Kiss, T. (2004). Human telomerase RNA and box H/ACA scaRNAs share a common Cajal body-specific localization signal. *J. Cell Biol.* *164*, 647–652.
- de Jager, T.L., Cockrell, A.E., and Du Plessis, S.S. (2017). Ultraviolet Light Induced Generation of Reactive Oxygen Species. *Adv. Exp. Med. Biol.* *996*, 15–23.
- James, N.J., Howell, G.J., Walker, J.H., and Blair, G.E. (2010). The role of Cajal bodies in the expression of late phase adenovirus proteins. *Virology* *399*, 299–311.
- Jeong, Y.H., Ling, J.P., Lin, S.Z., Donde, A.N., Braunstein, K.E., Majounie, E., Traynor, B.J., LaClair, K.D., Lloyd, T.E., and Wong, P.C. (2017). Tdp-43 cryptic exons are highly variable between cell types. *Mol. Neurodegener.* *12*, 13.

- Jolly, C., and Lakhotia, S.C. (2006). Human sat III and *Drosophila* hsr omega transcripts: a common paradigm for regulation of nuclear RNA processing in stressed cells. *Nucleic Acids Res.* *34*, 5508–5514.
- Jung, T., Höhn, A., and Grune, T. (2014). The proteasome and the degradation of oxidized proteins: Part III—Redox regulation of the proteasomal system. *Redox Biol.* *2*, 388–394.
- Kalsotra, A., Wang, K., Li, P.-F., and Cooper, T.A. (2010). MicroRNAs coordinate an alternative splicing network during mouse postnatal heart development. *Genes Dev.* *24*, 653–658.
- Ke, S., Pandya-Jones, A., Saito, Y., Fak, J.J., Vågbo, C.B., Geula, S., Hanna, J.H., Black, D.L., Darnell, J.E., and Darnell, R.B. (2017). m6A mRNA modifications are deposited in nascent pre-mRNA and are not required for splicing but do specify cytoplasmic turnover. *Genes Dev.* *31*, 990–1006.
- Kim Guisbert, K.S., and Guisbert, E. (2017). SF3B1 is a stress-sensitive splicing factor that regulates both HSF1 concentration and activity. *PLoS ONE* *12*.
- Kishore, S., Khanna, A., Zhang, Z., Hui, J., Balwierz, P.J., Stefan, M., Beach, C., Nicholls, R.D., Zavolan, M., and Stamm, S. (2010). The snoRNA MBII-52 (SNORD 115) is processed into smaller RNAs and regulates alternative splicing. *Hum. Mol. Genet.* *19*, 1153–1164.
- Knuckles, P., Carl, S.H., Musheev, M., Niehrs, C., Wenger, A., and Bühler, M. (2017). RNA fate determination through cotranscriptional adenosine methylation and microprocessor binding. *Nat. Struct. Mol. Biol.* *24*, 561–569.
- Krawczak, M., Thomas, N.S.T., Hundrieser, B., Mort, M., Wittig, M., Hampe, J., and Cooper, D.N. (2007). Single base-pair substitutions in exon-intron junctions of human genes: nature, distribution, and consequences for mRNA splicing. *Hum. Mutat.* *28*, 150–158.
- Krstić, J., Trivanović, D., Mojsilović, S., and Santibanez, J.F. (2015). Transforming Growth Factor-Beta and Oxidative Stress Interplay: Implications in Tumorigenesis and Cancer Progression.
- Kruhlak, M.J., Lever, M.A., Fischle, W., Verdin, E., Bazett-Jones, D.P., and Hendzel, M.J. (2000). Reduced mobility of the alternate splicing factor (ASF) through the nucleoplasm and steady state speckle compartments. *J. Cell Biol.* *150*, 41–51.
- Kucherenko, M.M., and Shcherbata, H.R. (2018). miRNA targeting and alternative splicing in the stress response - events hosted by membrane-less compartments. *J. Cell Sci.* *131*.
- Laloi, C., Apel, K., and Danon, A. (2004). Reactive oxygen signalling: the latest news. *Curr. Opin. Plant Biol.* *7*, 323–328.
- Lancaster, A.K., Nutter-Upham, A., Lindquist, S., and King, O.D. (2014). PLAAC: a web and command-line application to identify proteins with prion-like amino acid composition. *Bioinforma. Oxf. Engl.* *30*, 2501–2502.

- Lange, T.S., Borovjagin, A.V., and Gerbi, S.A. (1998a). Nucleolar localization elements in U8 snoRNA differ from sequences required for rRNA processing. *RNA* 4, 789–800.
- Lange, T.S., Borovjagin, A., Maxwell, E.S., and Gerbi, S.A. (1998b). Conserved boxes C and D are essential nucleolar localization elements of U14 and U8 snoRNAs. *EMBO J.* 17, 3176–3187.
- Lange, T.S., Ezrokhi, M., Amaldi, F., and Gerbi, S.A. (1999). Box H and Box ACA Are Nucleolar Localization Elements of U17 Small Nucleolar RNA. *Mol. Biol. Cell* 10, 3877–3890.
- Lenzken, S.C., Loffreda, A., and Barabino, S.M.L. (2013). RNA Splicing: A New Player in the DNA Damage Response. *Int. J. Cell Biol.* 2013.
- Li, X., Zhu, P., Ma, S., Song, J., Bai, J., Sun, F., and Yi, C. (2015). Chemical pulldown reveals dynamic pseudouridylation of the mammalian transcriptome. *Nat. Chem. Biol.* 11, 592–597.
- Liebhauer, S.A., Wolf, S., and Schlessinger, D. (1978). Differences in rRNA metabolism of primary and SV40-transformed human fibroblasts. *Cell* 13, 121–127.
- Lim, K.H., Ferraris, L., Filloux, M.E., Raphael, B.J., and Fairbrother, W.G. (2011). Using positional distribution to identify splicing elements and predict pre-mRNA processing defects in human genes. *Proc. Natl. Acad. Sci.* 108, 11093–11098.
- Liu, J., Yue, Y., Han, D., Wang, X., Fu, Y., Zhang, L., Jia, G., Yu, M., Lu, Z., Deng, X., et al. (2014). A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. *Nat. Chem. Biol.* 10, 93–95.
- Liu, J.-L., Wu, Z., 'an, Nizami, Z., Deryusheva, S., Rajendra, T.K., Beumer, K.J., Gao, H., Matera, A.G., Carroll, D., and Gall, J.G. (2009). Coilin is essential for Cajal body organization in *Drosophila melanogaster*. *Mol. Biol. Cell* 20, 1661–1670.
- Ljosa, V., Sokolnicki, K.L., and Carpenter, A.E. (2012). Annotated high-throughput microscopy image sets for validation. *Nat. Methods* 9, 637–637.
- Long, J.C., and Cáceres, J.F. (2009). The SR protein family of splicing factors: master regulators of gene expression. *Biochem. J.* 417, 15–27.
- Lorković, Z.J., and Barta, A. (2008). Role of Cajal Bodies and Nucleolus in the Maturation of the U1 snRNP in *Arabidopsis*. *PLoS ONE* 3.
- Louloupi, A., Ntini, E., Conrad, T., and Ørom, U.A.V. (2018). Transient N-6-Methyladenosine Transcriptome Sequencing Reveals a Regulatory Role of m6A in Splicing Efficiency. *Cell Rep.* 23, 3429–3437.
- Luco, R.F., and Misteli, T. (2011). More than a splicing code: integrating the role of RNA, chromatin and non-coding RNA in alternative splicing regulation. *Curr. Opin. Genet. Dev.* 21, 366–372.

- Machyna, M., Heyn, P., and Neugebauer, K.M. (2013). Cajal bodies: where form meets function. *Wiley Interdiscip. Rev. RNA* 4, 17–34.
- Machyna, M., Kehr, S., Straube, K., Kappei, D., Buchholz, F., Butter, F., Ule, J., Hertel, J., Stadler, P.F., and Neugebauer, K.M. (2014). The Coilin Interactome Identifies Hundreds of Small Noncoding RNAs that Traffic through Cajal Bodies. *Mol. Cell* 56, 389–399.
- Mahmoudi, S., Henriksson, S., Weibrecht, I., Smith, S., Söderberg, O., Strömblad, S., Wiman, K.G., and Farnebo, M. (2010). WRAP53 Is Essential for Cajal Body Formation and for Targeting the Survival of Motor Neuron Complex to Cajal Bodies. *PLOS Biol* 8, e1000521.
- Mao, Y.S., Zhang, B., and Spector, D.L. (2011). Biogenesis and Function of Nuclear Bodies. *Trends Genet. TIG* 27, 295–306.
- Marchand, V., Blanloeil-Oillo, F., Helm, M., and Motorin, Y. (2016). Illumina-based RiboMethSeq approach for mapping of 2'-O-Me residues in RNA. *Nucleic Acids Res.* 44, e135.
- Martinez-Contreras, R., Cloutier, P., Shkreta, L., Fiset, J.-F., Revil, T., and Chabot, B. (2007). hnRNP proteins and splicing control. *Adv. Exp. Med. Biol.* 623, 123–147.
- Martínez-Montiel, N., Rosas-Murrieta, N.H., Martínez-Montiel, M., Gaspariano-Cholula, M.P., and Martínez-Contreras, R.D. (2016). Microbial and Natural Metabolites That Inhibit Splicing: A Powerful Alternative for Cancer Treatment.
- Matera AG (1993). Nucleoplasmic organization of small nuclear ribonucleoproteins in cultured human cells. *J. Cell Biol.* 121, 715–727.
- Matheson, T.D., and Kaufman, P.D. (2016). The p150N domain of chromatin assembly factor-1 regulates Ki-67 accumulation on the mitotic perichromosomal layer. *Mol. Biol. Cell* 28, 21–29.
- Matveeva, E.A., Al-Tinawi, Q.M.H., Rouchka, E.C., and Fondufe-Mittendorf, Y.N. (2019). Coupling of PARP1-mediated chromatin structural changes to transcriptional RNA polymerase II elongation and cotranscriptional splicing. *Epigenetics Chromatin* 12, 15.
- Meijering, E., Carpenter, A.E., Peng, H., Hamprecht, F.A., and Olivo-Marin, J.-C. (2016). Imagining the future of bioimage analysis. *Nat. Biotechnol.* 34, 1250–1255.
- Merdzhanova, G., Edmond, V., De Seranno, S., Van den Broeck, A., Corcos, L., Brambilla, C., Brambilla, E., Gazzeri, S., and Eymin, B. (2008). E2F1 controls alternative splicing pattern of genes involved in apoptosis through upregulation of the splicing factor SC35. *Cell Death Differ.* 15, 1815–1823.
- Meseguer, S., Mudduluru, G., Escamilla, J.M., Allgayer, H., and Barettino, D. (2011). MicroRNAs-10a and -10b contribute to retinoic acid-induced differentiation of neuroblastoma cells and target the alternative splicing regulatory factor SFRS1 (SF2/ASF). *J. Biol. Chem.* 286, 4150–4164.

- Meyer, F., and Beucher, S. (1990). Morphological segmentation. *J. Vis. Commun. Image Represent.* *1*, 21–46.
- Mirkin, N., Fonseca, D., Mohammed, S., Cevher, M.A., Manley, J.L., and Kleiman, F.E. (2008). The 3' processing factor CstF functions in the DNA repair response. *Nucleic Acids Res.* *36*, 1792–1804.
- Misteli, T. (2001). The concept of self-organization in cellular architecture. *J. Cell Biol.* *155*, 181–186.
- Moore, M.J., and Sharp, P.A. (1993). Evidence for two active sites in the spliceosome provided by stereochemistry of pre-mRNA splicing. *Nature* *365*, 364–368.
- Moore, M.J., Wang, Q., Kennedy, C.J., and Silver, P.A. (2010). An Alternative Splicing Network Links Cell Cycle Control to Apoptosis. *Cell* *142*, 625–636.
- Morency, E., Sabra, M., Catez, F., Texier, P., and Lomonte, P. (2007). A novel cell response triggered by interphase centromere structural instability. *J. Cell Biol.* *177*, 757–768.
- Mucaki, E.J., and Rogan, P.K. (2019). Expression changes confirm predicted single nucleotide variants affecting mRNA splicing. *bioRxiv* 549089.
- Muñoz, M.J., Pérez Santangelo, M.S., Paronetto, M.P., de la Mata, M., Pelisch, F., Boireau, S., Glover-Cutter, K., Ben-Dov, C., Blaustein, M., Lozano, J.J., et al. (2009). DNA damage regulates alternative splicing through inhibition of RNA polymerase II elongation. *Cell* *137*, 708–720.
- Muratani, M., Gerlich, D., Janicki, S.M., Gebhard, M., Eils, R., and Spector, D.L. (2002). Metabolic-energy-dependent movement of PML bodies within the mammalian cell nucleus. *Nat. Cell Biol.* *4*, 106–110.
- Narayanan, A., Lukowiak, A., Jády, B.E., Dragon, F., Kiss, T., Terns, R.M., and Terns, M.P. (1999a). Nucleolar localization signals of box H/ACA small nucleolar RNAs. *EMBO J.* *18*, 5120–5130.
- Narayanan, A., Speckmann, W., Terns, R., and Terns, M.P. (1999b). Role of the Box C/D Motif in Localization of Small Nucleolar RNAs to Coiled Bodies and Nucleoli. *Mol. Biol. Cell* *10*, 2131–2147.
- Naro, C., Bielli, P., Pagliarini, V., and Sette, C. (2015). The interplay between DNA damage response and RNA processing: the unexpected role of splicing factors as gatekeepers of genome stability. *Front. Genet.* *6*.
- Nilsen, T.W., and Graveley, B.R. (2010). Expansion of the eukaryotic proteome by alternative splicing. *Nature* *463*, 457–463.
- Novotný, I., Blažíková, M., Staneček, D., Herman, P., and Malinsky, J. (2011). In vivo kinetics of U4/U6·U5 tri-snRNP formation in Cajal bodies. *Mol. Biol. Cell* *22*, 513–523.
- Novotný, I., Malinová, A., Stejskalová, E., Matějů, D., Klimešová, K., Roithová, A., Švéda, M.,



- Knejzlík, Z., and Staněk, D. (2015). SART3-Dependent Accumulation of Incomplete Spliceosomal snRNPs in Cajal Bodies. *Cell Rep.* *10*, 429–440.
- Olson, M.O.J., and Dundr, M. (2005). The moving parts of the nucleolus. *Histochem. Cell Biol.* *123*, 203–216.
- Pan, Q., Shai, O., Lee, L.J., Frey, B.J., and Blencowe, B.J. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.* *40*, 1413–1415.
- Papasaikas, P., and Valcárcel, J. (2016). The Spliceosome: The Ultimate RNA Chaperone and Sculptor. *Trends Biochem. Sci.* *41*, 33–45.
- Park, J.W., Parisky, K., Celotto, A.M., Reenan, R.A., and Graveley, B.R. (2004). Identification of alternative splicing regulators by RNA interference in *Drosophila*. *Proc. Natl. Acad. Sci. U. S. A.* *101*, 15974–15979.
- Patten, D.A., Germain, M., Kelly, M.A., and Slack, R.S. (2010). Reactive Oxygen Species: Stuck in the Middle of Neurodegeneration. *J. Alzheimers Dis.* *20*, S357–S367.
- Pederson, T. (2011). The nucleolus. *Cold Spring Harb. Perspect. Biol.* *3*.
- Phair, R.D., and Misteli, T. (2000). High mobility of proteins in the mammalian cell nucleus. *Nature* *404*, 604–609.
- Ping, X.-L., Sun, B.-F., Wang, L., Xiao, W., Yang, X., Wang, W.-J., Adhikari, S., Shi, Y., Lv, Y., Chen, Y.-S., et al. (2014). Mammalian WTAP is a regulatory subunit of the RNA N6-methyladenosine methyltransferase. *Cell Res.* *24*, 177–189.
- Platani, M., Goldberg, I., Swedlow, J.R., and Lamond, A.I. (2000). In vivo analysis of Cajal body movement, separation, and joining in live human cells. *J. Cell Biol.* *151*, 1561–1574.
- Platani, M., Goldberg, I., Lamond, A.I., and Swedlow, J.R. (2002). Cajal Body dynamics and association with chromatin are ATP-dependent. *Nat. Cell Biol.* *4*, 502–508.
- Pleiss, J.A., Whitworth, G.B., Bergkessel, M., and Guthrie, C. (2007). Rapid, Transcript-Specific Changes in Splicing in Response to Environmental Stress. *Mol. Cell* *27*, 928–937.
- Plessel, G., Fischer, U., and Lührmann, R. (1994). m3G cap hypermethylation of U1 small nuclear ribonucleoprotein (snRNP) in vitro: evidence that the U1 small nuclear RNA-(guanosine-N2)-methyltransferase is a non-snRNP cytoplasmic protein that requires a binding site on the Sm core domain. *Mol. Cell. Biol.* *14*, 4160–4172.
- Poole, A.R., Enwerem, I.I., Vicino, I.A., Coole, J.B., Smith, S.V., and Hebert, M.D. (2016). Identification of processing elements and interactors implicate SMN, coilin and the pseudogene-encoded coilp1 in telomerase and box C/D scaRNP biogenesis. *RNA Biol.* *13*, 955–972.

- Pradet-Balade, B., Girard, C., Boulon, S., Paul, C., Azzag, K., Bordonné, R., Bertrand, E., and Verheggen, C. (2011). CRM1 controls the composition of nucleoplasmic pre-snoRNA complexes to licence them for nucleolar transport. *EMBO J.* 30, 2205–2218.
- Rabin, S.J., Kim, J.M.H., Baughn, M., Libby, R.T., Kim, Y.J., Fan, Y., Libby, R.T., La Spada, A., Stone, B., and Ravits, J. (2010). Sporadic ALS has compartment-specific aberrant exon splicing and altered cell-matrix adhesion biology. *Hum. Mol. Genet.* 19, 313–328.
- Ramanouskaya, T.V., and Grinev, V.V. (2017). The determinants of alternative RNA splicing in human cells. *Mol. Genet. Genomics MGG* 292, 1175–1195.
- Richard, P., Darzacq, X., Bertrand, E., Jádý, B.E., Verheggen, C., and Kiss, T. (2003). A common sequence motif determines the Cajal body-specific localization of box H/ACA scaRNAs. *EMBO J.* 22, 4283–4293.
- Romero-Barrios, N., Legascue, M.F., Benhamed, M., Ariel, F., and Crespi, M. (2018). Splicing regulation by long noncoding RNAs. *Nucleic Acids Res.* 46, 2169–2184.
- Salomoni, P., and Pandolfi, P.P. (2002). The role of PML in tumor suppression. *Cell* 108, 165–170.
- Saltzman, A.L., Pan, Q., and Blencowe, B.J. (2011). Regulation of alternative splicing by the core spliceosomal machinery. *Genes Dev.* 25, 373–384.
- Samarsky, D.A., Fournier, M.J., Singer, R.H., and Bertrand, E. (1998). The snoRNA box C/D motif directs nucleolar targeting and also couples snoRNA synthesis and localization. *EMBO J.* 17, 3747–3757.
- Sawyer, I.A., Hager, G.L., and Dundr, M. (2016). Specific genomic cues regulate Cajal body assembly. *RNA Biol.* 14, 791–803.
- Schieber, M., and Chandel, N.S. (2014). ROS Function in Redox Signaling and Oxidative Stress. *Curr. Biol. CB* 24, R453–R462.
- Schor, I.E., Fiszbein, A., Petrillo, E., and Kornblihtt, A.R. (2013). Intragenic epigenetic changes modulate NCAM alternative splicing in neuronal differentiation. *EMBO J.* 32, 2264–2274.
- Seo, J., Singh, N.N., Ottesen, E.W., Sivanesan, S., Shishimorova, M., and Singh, R.N. (2016). Oxidative Stress Triggers Body-Wide Skipping of Multiple Exons of the Spinal Muscular Atrophy Gene. *PLOS ONE* 11, e0154390.
- Sephton, C.F., Cenik, C., Kucukural, A., Dammer, E.B., Cenik, B., Han, Y., Dewey, C.M., Roth, F.P., Herz, J., Peng, J., et al. (2011). Identification of neuronal RNA targets of TDP-43-containing ribonucleoprotein complexes. *J. Biol. Chem.* 286, 1204–1215.

- Shalgi, R., Hurt, J.A., Lindquist, S., and Burge, C.B. (2014). Widespread Inhibition of Posttranscriptional Splicing Shapes the Cellular Transcriptome following Heat Shock. *Cell Rep.* 7, 1362–1370.
- Shanbhag, R., Kurabi, A., Kwan, J.J., and Donaldson, L.W. (2010). Solution structure of the carboxy-terminal Tudor domain from human Coilin. *FEBS Lett.* 584, 4351–4356.
- Shang, X., Cao, Y., and Ma, L. (2017). Alternative Splicing in Plant Genes: A Means of Regulating the Environmental Fitness of Plants. *Int. J. Mol. Sci.* 18.
- Sharma, S., Marchand, V., Motorin, Y., and Lafontaine, D.L.J. (2017). Identification of sites of 2'-O-methylation vulnerability in human ribosomal RNAs by systematic mapping. *Sci. Rep.* 7, 1–15.
- Shav-Tal, Y., Blechman, J., Darzacq, X., Montagna, C., Dye, B.T., Patton, J.G., Singer, R.H., and Zipori, D. (2005). Dynamic Sorting of Nuclear Components into Distinct Nucleolar Caps during Transcriptional Inhibition. *Mol. Biol. Cell* 16, 2395–2413.
- Shibutani, S., Takeshita, M., and Grollman, A.P. (1991). Insertion of specific bases during DNA synthesis past the oxidation-damaged base 8-oxodG. *Nature* 349, 431–434.
- Shin, C., Feng, Y., and Manley, J.L. (2004). Dephosphorylated SRp38 acts as a splicing repressor in response to heat shock. *Nature* 427, 553–558.
- Shirley, B.C., Mucaki, E.J., and Rogan, P.K. (2018). Pan-Cancer Repository of Validated Natural and Cryptic mRNA Splicing Mutations. *bioRxiv* 474452.
- Shkreta, L., and Chabot, B. (2015). The RNA Splicing Response to DNA Damage. *Biomolecules* 5, 2935–2977.
- Sims, R.J., Millhouse, S., Chen, C.-F., Lewis, B.A., Erdjument-Bromage, H., Tempst, P., Manley, J.L., and Reinberg, D. (2007). Recognition of Trimethylated Histone H3 Lysine 4 Facilitates the Recruitment of Transcription Post-Initiation Factors and pre-mRNA Splicing. *Mol. Cell* 28, 665–676.
- Staněk, D., and Neugebauer, K.M. (2004). Detection of snRNP assembly intermediates in Cajal bodies by fluorescence resonance energy transfer. *J. Cell Biol.* 166, 1015–1025.
- Strzelecka, M., Trowitzsch, S., Weber, G., Lührmann, R., Oates, A.C., and Neugebauer, K.M. (2010). Coilin-dependent snRNP assembly is essential for zebrafish embryogenesis. *Nat. Struct. Mol. Biol.* 17, 403–409.
- Sun, Z., Diaz, Z., Fang, X., Hart, M.P., Chesi, A., Shorter, J., and Gitler, A.D. (2011). Molecular determinants and genetic modifiers of aggregation and toxicity for the ALS disease protein FUS/TLS. *PLoS Biol.* 9, e1000614.

- Takeo, K., Kawai, T., Nishida, K., Masuda, K., Teshima-Kondo, S., Tanahashi, T., and Rokutan, K. (2009). Oxidative stress-induced alternative splicing of transformer 2 $\beta$  (SFRS10) and CD44 pre-mRNAs in gastric epithelial cells. *Am. J. Physiol. - Cell Physiol.* 297, C330–C338.
- Tapia, O., Bengoechea, R., Berciano, M.T., and Lafarga, M. (2010). Nucleolar targeting of coilin is regulated by its hypomethylation state. *Chromosoma* 119, 527–540.
- Tian, B., Pan, Z., and Lee, J.Y. (2007). Widespread mRNA polyadenylation events in introns indicate dynamic interplay between polyadenylation and splicing. *Genome Res.* 17, 156–165.
- Tollervey, J.R., Curk, T., Rogelj, B., Briesse, M., Cereda, M., Kayikci, M., König, J., Hortobágyi, T., Nishimura, A.L., Župunski, V., et al. (2011). Characterizing the RNA targets and position-dependent splicing regulation by TDP-43. *Nat. Neurosci.* 14, 452–458.
- Toyota, C.G., Davis, M.D., Cosman, A.M., and Hebert, M.D. (2010). Coilin phosphorylation mediates interaction with SMN and SmB'. *Chromosoma* 119, 205–215.
- Trinkle-Mulcahy, L., and Sleeman, J.E. (2016). The Cajal body and the nucleolus: “In a relationship” or “It’s complicated”? *RNA Biol.* 14, 739–751.
- Tripathi, V., Ellis, J.D., Shen, Z., Song, D.Y., Pan, Q., Watt, A.T., Freier, S.M., Bennett, C.F., Sharma, A., Bubulya, P.A., et al. (2010). The Nuclear-Retained Noncoding RNA MALAT1 Regulates Alternative Splicing by Modulating SR Splicing Factor Phosphorylation. *Mol. Cell* 39, 925–938.
- Tycowski, K.T., You, Z.H., Graham, P.J., and Steitz, J.A. (1998). Modification of U6 spliceosomal RNA is guided by other small RNAs. *Mol. Cell* 2, 629–638.
- Tycowski, K.T., Shu, M.-D., Kukoyi, A., and Steitz, J.A. (2009). A Conserved WD40 Protein Binds the Cajal Body Localization Signal of scaRNP Particles. *Mol. Cell* 34, 47–57.
- Utans, U., Behrens, S.E., Lührmann, R., Kole, R., and Krämer, A. (1992). A splicing factor that is inactivated during in vivo heat shock is functionally equivalent to the [U4/U6.U5] triple snRNP-specific proteins. *Genes Dev.* 6, 631–641.
- Vautrot, V., Aigueperse, C., Branlant, C., and Behm-Ansmant, I. (2015). Fluorescence in situ hybridization of small non-coding RNAs. *Methods Mol. Biol. Clifton NJ* 1296, 73–83.
- Vazquez-Arango, P., and O'Reilly, D. (2018). Variant snRNPs: New players within the spliceosome system. *RNA Biol.* 15, 17–25.
- Verduci, L., Simili, M., Rizzo, M., Mercatanti, A., Evangelista, M., Mariani, L., Rainaldi, G., and Pitto, L. (2010). MicroRNA (miRNA)-mediated Interaction between Leukemia/Lymphoma-related Factor (LRF) and Alternative Splicing Factor/Splicing Factor 2 (ASF/SF2) Affects Mouse Embryonic Fibroblast Senescence and Apoptosis. *J. Biol. Chem.* 285, 39551–39563.
- Verheggen, C., Mouaikel, J., Thiry, M., Blanchard, J.-M., Tollervey, D., Bordonné, R., Lafontaine,

- D.L.J., and Bertrand, E. (2001). Box C/D small nucleolar RNA trafficking involves small nucleolar RNP proteins, nucleolar factors and a novel nuclear domain. *EMBO J.* *20*, 5480–5490.
- Verheggen, C., Lafontaine, D.L.J., Samarsky, D., Mouaikel, J., Blanchard, J.-M., Bordonné, R., and Bertrand, E. (2002). Mammalian and yeast U3 snoRNPs are matured in specific and related nuclear compartments. *EMBO J.* *21*, 2736–2745.
- Villagra, N.T., Navascues, J., Casafont, I., Val-Bernal, J.F., Lafarga, M., and Berciano, M.T. (2006). The PML-nuclear inclusion of human supraoptic neurons: a new compartment with SUMO-1- and ubiquitin-proteasome-associated domains. *Neurobiol. Dis.* *21*, 181–193.
- de Vries, H.E., Witte, M., Hondius, D., Rozemuller, A.J.M., Drukarch, B., Hoozemans, J., and van Horssen, J. (2008). Nrf2-induced antioxidant protection: A promising target to counteract ROS-mediated damage in neurodegenerative disease? *Free Radic. Biol. Med.* *45*, 1375–1383.
- Wahl, M.C., Will, C.L., and Lührmann, R. (2009). The Spliceosome: Design Principles of a Dynamic RNP Machine. *Cell* *136*, 701–718.
- Wang, E.T., Sandberg, R., Luo, S., Khrebtkova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P., and Burge, C.B. (2008). Alternative Isoform Regulation in Human Tissue Transcriptomes. *Nature* *456*, 470–476.
- Wang, Q., Sawyer, I.A., Sung, M.-H., Sturgill, D., Shevtsov, S.P., Pegoraro, G., Hakim, O., Baek, S., Hager, G.L., and Dundr, M. (2016). Cajal bodies are linked to genome conformation. *Nat. Commun.* *7*.
- Ward, I.M., Minn, K., van Deursen, J., and Chen, J. (2003). p53 Binding Protein 53BP1 Is Required for DNA Damage Responses and Tumor Suppression in Mice. *Mol. Cell. Biol.* *23*, 2556–2563.
- Watkins, N.J., Dickmanns, A., and 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, N.J., Lemm, I., Ingelfinger, D., Schneider, C., Hossbach, M., Urlaub, H., and Lührmann, R. (2004). Assembly and maturation of the U3 snoRNP in the nucleoplasm in a large dynamic multiprotein complex. *Mol. Cell* *16*, 789–798.
- Weldon, C., Dacanay, J.G., Gokhale, V., Boddupally, P.V.L., Behm-Ansmant, I., Burley, G.A., Branlant, C., Hurley, L.H., Dominguez, C., and Eperon, I.C. (2018). Specific G-quadruplex ligands modulate the alternative splicing of Bcl-X. *Nucleic Acids Res.* *46*, 886–896.
- West, J.A., Davis, C.P., Sunwoo, H., Simon, M.D., Sadreyev, R.I., Wang, P.I., Tolstorukov, M.Y., and Kingston, R.E. (2014). The long noncoding RNAs NEAT1 and MALAT1 bind active chromatin sites. *Mol. Cell* *55*, 791–802.

- Wiesmann, V., Franz, D., Held, C., Münzenmayer, C., Palmisano, R., and Wittenberg, T. (2015). Review of free software tools for image analysis of fluorescence cell micrographs. *J. Microsc.* 257, 39–53.
- Will, C.L., and Lührmann, R. (2005). Splicing of a rare class of introns by the U12-dependent spliceosome. *Biol. Chem.* 386, 713–724.
- Will, C.L., and Lührmann, R. (2011). Spliceosome Structure and Function. *Cold Spring Harb. Perspect. Biol.* 3, a003707–a003707.
- Wu, G., Xiao, M., Yang, C., and Yu, Y.-T. (2011). U2 snRNA is inducibly pseudouridylated at novel sites by Pus7p and snR81 RNP. *EMBO J.* 30, 79–89.
- Wu, G., Adachi, H., Ge, J., Stephenson, D., Query, C.C., and Yu, Y.-T. (2016). Pseudouridines in U2 snRNA stimulate the ATPase activity of Prp5 during spliceosome assembly. *EMBO J.* 35, 654–667.
- Xiao, S., Sanelli, T., Dib, S., Sheps, D., Findlater, J., Bilbao, J., Keith, J., Zinman, L., Rogaeva, E., and Robertson, J. (2011). RNA targets of TDP-43 identified by UV-CLIP are deregulated in ALS. *Mol. Cell. Neurosci.* 47, 167–180.
- Xiao, W., Adhikari, S., Dahal, U., Chen, Y.-S., Hao, Y.-J., Sun, B.-F., Sun, H.-Y., Li, A., Ping, X.-L., Lai, W.-Y., et al. (2016). Nuclear m(6)A Reader YTHDC1 Regulates mRNA Splicing. *Mol. Cell* 61, 507–519.
- Yamada, M., Sato, T., Shimohata, T., Hayashi, S., Igarashi, S., Tsuji, S., and Takahashi, H. (2001). Interaction between Neuronal Intranuclear Inclusions and Promyelocytic Leukemia Protein Nuclear and Coiled Bodies in CAG Repeat Diseases. *Am. J. Pathol.* 159, 1785–1795.
- Yang, Y., Isaac, C., Wang, C., Dragon, F., Pogacic, V., and Meier, U.T. (2000). Conserved composition of mammalian box H/ACA and box C/D small nucleolar ribonucleoprotein particles and their interaction with the common factor Nopp140. *Mol. Biol. Cell* 11, 567–577.
- Yost, H.J., and Lindquist, S. (1986). RNA splicing is interrupted by heat shock and is rescued by heat shock protein synthesis. *Cell* 45, 185–193.
- Yuan, G., Klämbt, C., Bachellerie, J.-P., Brosius, J., and Hüttenhofer, A. (2003). RNomics in *Drosophila melanogaster*: identification of 66 candidates for novel non-messenger RNAs. *Nucleic Acids Res.* 31, 2495–2507.
- Zhang, Z., Lotti, F., Dittmar, K., Younis, I., Wan, L., Kasim, M., and Dreyfuss, G. (2008). SMN deficiency causes tissue-specific perturbations in the repertoire of snRNAs and widespread defects in splicing. *Cell* 133, 585–600.
- Zhao, X., Yang, Y., Sun, B.-F., Shi, Y., Yang, X., Xiao, W., Hao, Y.-J., Ping, X.-L., Chen, Y.-S., Wang, W.-J., et al. (2014). FTO-dependent demethylation of N6-methyladenosine regulates mRNA splicing and is required for adipogenesis. *Cell Res.* 24, 1403–1419.

Zhao, Y., Dunker, W., Yu, Y.-T., and Karijovich, J. (2018). The Role of Noncoding RNA Pseudouridylation in Nuclear Gene Expression Events. *Front. Bioeng. Biotechnol.* 6.

Zhou, K.I., Shi, H., Lyu, R., Wylder, A.C., Matuszek, Ż., Pan, J.N., He, C., Parisien, M., and Pan, T. (2019). Regulation of Co-transcriptional Pre-mRNA Splicing by m6A through the Low-Complexity Protein hnRNP G. *Mol. Cell* 76, 70–81.e9.

## Résumé

Au cours du processus de maturation des pré-ARN, l'épissage alternatif (EA) inclut ou exclut sélectivement des séquences introniques et exoniques spécifiques, aboutissant à partir du même pré-ARN à différentes combinaisons d'exons/introns également appelées isoformes. Des études récentes étendues à l'ensemble du génome estiment que 95% des gènes humains sont épissés de façon alternative. Aussi, il n'est pas surprenant que des stress cellulaires induisent des changements du profil global d'épissage participant à l'établissement de la réponse cellulaire ou au développement de pathologies. Notre analyse globale de transcriptomique réalisée avec des puces Affimetrix a révélé des variations d'épissages significatives en réponse à un stress oxydant ( $\text{H}_2\text{O}_2$  500 $\mu\text{M}$ ). C'est pourquoi nous nous sommes intéressés aux effets de ce stress sur la machinerie complexe appelée spliceosome, responsable de la catalyse dynamique de l'épissage. Le spliceosome fonctionnel est composé de différentes ribonucléoprotéines (RNPs), caractérisées par de petits ARN nucléaires riches en Uridine (U snRNA, U-rich small nuclear RNA), chacun associé à des protéines spécifiques et formant ainsi les UsnRNPs. Pour être fonctionnelle, chaque U snRNP est maturée suivant différentes étapes comprenant la modification chimique de leur composant ARN. Ces modifications dites post-transcriptionnelles sont induites par d'autres RNPs qui sont particulièrement concentrées dans les corps de Cajal nucléaires et sont ainsi appelés scaRNPs (small Cajal Bodies RNPs). Nos résultats obtenus par des techniques de microscopie de fluorescence (Immunofluorescence et Hybridation in situ de l'ARN par Fluorescence) montrent que le stress oxydant modifie la structure des corps de Cajal en induisant la redistribution nucléoplasmique de son composant structural et fonctionnel majeur, la protéine coiline. Cet effet se caractérise par la formation d'une dizaine à plusieurs centaines de microfoci coiline, corrélant avec une diminution du nombre de foci nucléaires caractéristiques des corps de Cajal. De façon intéressante, nous avons pu mettre en évidence que cette redistribution de la coiline s'accompagne d'un changement de la localisation d'autres composants des corps de Cajal impliqués dans : (i) la formation de modifications post-transcriptionnelles au sein des U snARNs (dyskérine, fibrillarine...), (ii) le recrutement des scaRNPs (WRAP53...), (iii) composition des UsnRNPs (ARN U2, protéine U1A). Le stress oxydant semble aussi exclure la protéine SMN du noyau alors qu'il joue normalement un rôle majeur dans le ré-import nucléaire des U snRNP en cours de maturation. Cette étude a généré un grand nombre de données en microscopie de fluorescence, ce qui m'a conduit à développer différents protocoles d'analyses d'images biologiques afin d'optimiser l'extraction et le traitement de ces données. En parallèle, des expériences d'immunoprécipitation de l'ARN (RIP) couplées à du séquençage haut débit ont révélé (i) une augmentation significative de la fraction des UsnRNP U4, U5 et U6 suggérant un éventuel effet du stress oxydant sur la tri-snRNP U4-U6.U5, (ii) des variations du taux de 2'-O-méthylation induites par le stress oxydant directement au niveau des UsnRNP matures.

**Mots clés :** stress oxydant ; épissage ; U snRNP ; corps de Cajal ; modifications post-transcriptionnelles ; microscopie fluorescente ; analyse d'images biologiques

## Abstract

During pre-RNA maturation process, alternative splicing (AS) selectively includes or excludes specific intronic or exonic sequences. This process leads from a given pre-RNA, to different combinations of exons/introns also called isoforms. Some recent studies estimate that 95% of human genes are alternatively spliced. Then, it is not surprising that stress-induced modifications of splicing profiles participate to the establishment of the cell response to stress or to the development of pathologies. Our global transcriptomic analysis realized with Affymetrix chips has revealed significant splicing variations in response to an oxidative stress (500 $\mu\text{M}$   $\text{H}_2\text{O}_2$ ). That is why we were interested in such stress effects on the complex machinery, named spliceosome, responsible for the dynamic catalysis of splicing. The functional spliceosome is composed by different ribonucleoproteins (RNPs), each being composed of a Uridine-rich small nuclear RNA (U snRNA) associated with specific proteins to form the mature UsnRNPs. To be functional, each UsnRNP should be matured following different steps including the chemical modification of its RNA component. These post-transcriptional modifications are induced by another type of RNPs which are particularly enriched within nuclear Cajal bodies and are named scaRNPs (for small Cajal Bodies RNPs). Our results obtained by fluorescence microscopy techniques (Immunofluorescence and RNA-Fluorescent In Situ Hybridization) show that oxidative stress changes Cajal bodies structure by inducing the nucleoplasmic redistribution of its main structural and functional component, the coilin protein. This effect is characterized by the formation of tens to hundreds of coilin containing microfoci, correlating with a decrease of the number of bigger Cajal bodies specific nuclear foci. Interestingly, we highlighted that coilin redistribution is accompanied with changes of the localization of others Cajal Body components: (i) factors involved in the formation of post-transcriptional modifications in UsnRNAs (dyskerin, fibrillarin...), (ii) factors involved in scaRNPs retention (WRAP53...), (iii) UsnRNPs components (U2 RNA, U1A protein). Oxidative stress also seems to induce the exclusion of SMN protein from the nucleus, knowing that it plays a central role in UsnRNP nuclear re-import during their maturation. This study has generated a large amount of fluorescence microscopy data, so that I have developed different bio-image analysis protocols in order to optimize the extraction and the processing of these data. In parallel, RNA-Immunoprecipitation (RIP) experiments coupled with high-throughput sequencing have revealed (i) a significant increase of U4, U5 and U6 snRNP fraction suggesting a potential effect of oxidative stress on the U4/U6-U5 tri-snRNP, (ii) oxidative stress-induced variations of the 2'-O-methylation level directly in mature U snRNPs.

**Mots clés :** oxidative stress ; splicing ; U snRNA ; Cajal bodies ; post-transcriptional modifications ; fluorescence microscopy ; bio-image analysis