



**HAL**  
open science

# Rôle de la polyadénylation intronique dans la réponse des cellules cancéreuses au cisplatine

Alexandre Devaux

► **To cite this version:**

Alexandre Devaux. Rôle de la polyadénylation intronique dans la réponse des cellules cancéreuses au cisplatine. Cancer. Université Paris-Saclay, 2024. Français. NNT : 2024UPASL015 . tel-04583916

**HAL Id: tel-04583916**

**<https://theses.hal.science/tel-04583916>**

Submitted on 22 May 2024

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

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

# Rôle de la polyadénylation intronique dans la réponse des cellules cancéreuses au cisplatine

*Role of intronic polyadenylation in cancer cell  
response to cisplatin*

## Thèse de doctorat de l'université Paris-Saclay

École doctorale n° 582, Cancérologie : Biologie, Médecine, Santé (CBMS)  
Spécialité de doctorat : Sciences du cancer  
Graduate School : Life Science and Health. Référent : Faculté de Médecine

Thèse préparée dans l'unité de recherche Intégrité du Génome, ARN et Cancer  
(Université Paris-Saclay, CNRS), sous la direction de **Martin DUTERTRE**,  
directeur de recherche

Thèse soutenue à Paris-Saclay, le 25 mars 2024, par

**Alexandre DEVAUX**

## Composition du Jury

Membres du jury avec voix délibérative

<b>Serge PLAZA</b> Directeur de recherche, Université Toulouse III Paul Sabatier	Président
<b>Cyril BOURGEOIS</b> Chargé de recherche, Université Claude Bernard Lyon	Rapporteur & Examineur
<b>Stefania MILLEVOI</b> Directrice de recherche, Université Toulouse III Paul Sabatier	Rapporteur & Examinatrice
<b>Valéria NAIM</b> Chargé de recherche, Université Paris Saclay GS Life Sciences and Health	Examinatrice

**Titre :** Rôle de la polyadénylation intronique dans la réponse des cellules cancéreuses au cisplatine.

**Mots clés :** cancer du poumon, cisplatine, épissage alternatif, microprotéines, polyadénylation intronique, U1-AMO

Au cours d'études sur la polyadénylation alternative (APA), des transcrits courts terminant dans un exon final alternatif ont été découverts, on parle de polyadénylation intronique (IPA). L'IPA est régulée par des facteurs de l'épissage (dont U1 snRNP), de polyadénylation, et d'élongation de la transcription (dont CDK12). Les isoformes IPA sont régulées par des agents génotoxiques (induisant des dommages à l'ADN), dont les rayonnements UV et la doxorubicine. Les inhibiteurs de CDK12 augmentent l'IPA dans des gènes de réparation de l'ADN et la sensibilité cellulaire à des génotoxiques.

L'IPA a souvent lieu dans la région codante des gènes, générant des protéines altérées en carboxy terminal. Cependant, des transcrits IPA sont aussi générés dans les premiers introns des gènes, on parle alors de 5'IPA. Des transcrits 5'IPA sont dégradés par l'exosome nucléaire, mais certains sont abondants et ont un faible potentiel codant. Deux d'entre eux, issues des gènes ASCC3 et CDKN1A, ont des fonctions non codantes. Par ailleurs, des études montrent par Ribo-seq et spectrométrie de masse (MS) l'existence, dans des ARNm et des lncRNA, de petits cadres de lecture ouverts (sORF) codant des microprotéines (miP, protéines de moins de 100 aa) qui peuvent être fonctionnelles. Aucune miP n'a été rapportée dans des isoformes 5'IPA.

Le cisplatine (CisPt) est un agent pontant de l'ADN très utilisé dans les cancers du poumon non à petites cellules (NSCLC). Mon équipe a observé par 3'-seq dans des cellules NSCLC que le CisPt augmente l'expression des isoformes IPA par rapport aux ARNm canoniques (ratio IPA:LE) dans de nombreux gènes, et que certaines isoformes IPA sont peu engagées dans les polysomes lourds et sont issues de la région en amont du site d'initiation de la traduction annoté du gène (isoformes 5'UTR-IPA).

Mes objectifs étaient de déterminer le rôle de l'IPA dans la réponse cellulaire au CisPt. Je me suis intéressé premièrement au rôle des isoformes 5'UTR-IPA. Pour deux d'entre elles, issues des gènes PRKAR1B et PHF20, j'ai montré que leur déplétion par siARN augmente la survie de cellules NSCLC au CisPt. Ces deux isoformes sont engagées dans des fractions polysomiques légères. Des analyses de bases de données de Ribo-seq et de MS ont révélé l'existence de sORF dans ces deux isoformes. Par transfection de vecteurs contenant ces isoformes 5'UTR-IPA et en balisant leurs sORF, j'ai observé en ImmunoFluorescence (IF) et Western Blot que l'isoforme 5'UTR-IPA de PRKAR1B code une miP. La délétion de cette isoforme IPA endogène ou la mutation de l'ATG de sa sORF par CRISPR ont donné un phénotype similaire au siARN. Il s'agit de la première isoforme 5'UTR-IPA codant une miP (miP-5'UTR-IPA). Par croisement de nos données de 3'-seq avec des données de Ribo-seq et de MS, nous avons identifié une centaine d'isoformes miP-5'UTR-IPA potentielles induites par le CisPt.

Deuxièmement, je me suis intéressé à la possibilité de sensibiliser des cellules NSCLC au CisPt en ciblant U1 snRNP par un oligonucléotide (U1-AMO) qui induit l'IPA dans de nombreux gènes. Dans plusieurs lignées cellulaires NSCLC, j'ai pu montrer une sensibilisation par U1-AMO au CisPt en termes d'inhibition de croissance cellulaire et d'induction de dommages à l'ADN (foyers  $\gamma$ H2AX). Cette sensibilisation est liée à une réduction (en 3'-seq et RT-qPCR) de l'expression des ARNm canoniques dans des gènes de réparation des pontages de l'ADN (voies de Fanconi et de l'excision de nucléotides). Cependant, U1-AMO prévient les blocages de cycle cellulaire induits par le CisPt, ainsi que les effets du CisPt sur le ratio IPA:LE de nombreux gènes.

Mes travaux montrent l'impact de l'IPA sur la réponse des cellules cancéreuses au CisPt, et révèlent un nouveau paradigme génétique, appelé miP-5'UTR-IPA, dans lequel des gènes produisent par IPA des transcrits courts codant des miP.

**Title :** Role of intronic polyadenylation in cancer cell response to cisplatin.

**Keywords :** lung cancer, cisplatin, alternative splicing, microproteins, intronic polyadenylation, U1-AMO

During studies on alternative polyadenylation (APA), short transcripts ending in an alternative last exon were discovered, known as intronic polyadenylation (IPA). IPA is regulated by splicing factors (including U1 snRNP), polyadenylation factors and transcription elongation factors (including CDK12). IPA isoforms are regulated by genotoxic agents (inducing DNA damage), including UV radiation and doxorubicin. Conversely, CDK12 inhibitors increase both IPA in DNA repair genes and cellular sensitivity to genotoxic agents.

IPA often occurs in the coding region of genes, generating carboxy-terminally altered proteins. However, IPA transcripts are also generated in the first introns of genes, known as 5'IPA. Many 5'IPA transcripts are degraded by the nuclear exosome, but some are abundant and have a low coding potential. Two of these, derived from the ASCC3 and CDKN1A genes, have non-coding functions. In addition, studies using Ribo-seq and mass spectrometry (MS) are showing the existence -in mRNAs and lncRNAs- of small open reading frames (sORF) encoding microproteins (miP, proteins of less than 100 aa) which can be functional. No miP has been reported in 5'IPA isoforms.

Cisplatin (CisPt) is a DNA-crosslinking agent widely used in non-small cell lung cancer (NSCLC). My team observed, by 3'-seq in NSCLC cells, that CisPt increases the expression of IPA isoforms compared to canonical mRNAs (IPA:LE ratio) in many genes, and that some IPA isoforms are poorly associated with heavy polysomes and are derived from the region upstream of the annotated translation initiation site of the gene (5'UTR-IPA isoforms).

My objectives were to determine the role of IPA in cell response to CisPt. I first looked at the role of 5'UTR-IPA isoforms. For two of them, derived from the PRKAR1B and PHF20 genes, I showed that their depletion by siRNA increased the survival of NSCLC cells to CisPt. These two isoforms are associated with light polysomal fractions. Analyses of Ribo-seq and MS databases revealed the existence of sORFs in these two isoforms. By transfecting vectors containing these 5'UTR-IPA isoforms and by tagging their sORFs, I observed by ImmunoFluorescence (IF) and Western Blot that the 5'UTR-IPA isoform of PRKAR1B encodes a miP. Deletion of this IPA isoform or mutation of the sORF ATG endogenously by CRISPR gave a phenotype similar to the siRNAs. This is the first 5'UTR-IPA isoform encoding a miP (miP-5'UTR-IPA). By cross-referencing our 3'-seq data with Ribo-seq and MS data, we identified around a hundred potential miP-5'UTR-IPA isoforms induced by CisPt.

Secondly, I investigated the possibility of sensitizing NSCLC cells to CisPt by targeting U1 snRNP with an antisense oligonucleotide (U1-AMO), that induces IPA in many genes. In several NSCLC cell lines, I showed sensitization to CisPt by U1-AMO in terms of cell growth inhibition and DNA damage induction (yH2AX foci). This sensitization is linked to a reduced expression of the canonical mRNAs of DNA crosslinks repair pathways (Fanconi and nucleotide excision repair), as shown by 3'-seq and RT-qPCR. However, U1-AMO prevented CisPt-induced cell cycle block and the effects of CisPt on the IPA:LE ratio of many genes.

My work shows the impact of IPA on the response of cancer cells to CisPt, and reveals a new genetic paradigm, called miP-5'UTR-IPA, in which genes produce short miP-encoding transcripts by IPA.



## Remerciements

Je tiens à remercier Monsieur Dutertre Martin, Directeur de recherche Inserm à l'Institut Curie, qui m'a encadré tout au long de cette thèse et qui m'a fait partager ses brillantes intuitions. Qu'il soit aussi remercié pour sa gentillesse, sa disponibilité permanente et pour les nombreux encouragements qu'il m'a prodigués.

Cette thèse est le fruit d'une collaboration de plus de trois années avec lui.

Je remercie Monsieur Vagner Stephan, directeur de l'Unité UMR 3348 « Intégrité du génome, ARN et Cancer »

J'adresse tous mes remerciements à Madame Millevoi Stefania, Directrice de recherche à l'Université Toulouse III Paul Sabatier, ainsi qu'à Monsieur Bourgeois, Chargé de recherche à l'Université Claude Bernard à Lyon, de l'honneur qu'ils m'ont fait en acceptant d'être rapporteurs de cette thèse.

J'exprime ma gratitude à Madame NAIM Valéria, Chargé de recherche à l'Université Paris Saclay GS Life Sciences and Health et à Monsieur PLAZA Serge, Directeur de recherche à l'Université Toulouse III Paul Sabatier, qui ont bien voulu être examinateurs.

Je tiens aussi à remercier Madame HENEMAN Amélie, ingénieure d'études CNRS pour m'avoir accompagné pendant 1 an dans mes expériences ainsi qu'à LABBE Céline, ingénieure d'études/bioinformaticienne pour m'avoir aidé sur des outils bioinformatiques et pour avoir réalisé le traitement des séquençages présents dans cette étude.

Je remercie Monsieur FOUILLEUL Quentin, doctorant en 1<sup>er</sup> année, qui sera en charge de poursuivre mes recherches et d'approfondir les questions soulevées par mes travaux.

Je remercie également Madame TRAN QUANG Christine, Chargée de recherche Inserm, Monsieur MESSAOUDI Cédric, ingénieur de recherches, Madame BESSE Laetitia, ingénieure d'études, pour leur apport sur les expériences d'immunofluorescence et Madame LASGI Charlène, Ingénieure d'études Inserm, sur celles de FACS.

Merci aussi à Madame DUCHAMBON Patricia, ingénieure d'étude de l'Inserm, dont la rencontre fortuite et les discussions sur les microprotéines ont permis la mise en place d'un protocole fiable.

Je tiens à remercier Madame UGUEN Patricia, Maître de conférence Université Paris Saclay, qui m'a permis d'effectuer une mission de vacataire pour découvrir le monde de l'enseignement.

Un grand merci aussi à tous les membres de mon équipe et en particulier à Madame FABBRI Lucilla, post-doctorante, Monsieur MARTINEAU Sylvain, ingénieur d'études de l'Institut Curie, Madame DIAN LUISA Ana, doctorante en 4<sup>ème</sup> année dont les discussions, débats et la bonne humeur m'ont aidé à la réalisation de cette thèse.

J'adresse tous mes remerciements aux membres de mon comité de suivi individuel de thèse pour l'encadrement et les conseils prodigués.

Je remercie enfin tous nos collaborateurs de près comme de loin qui ont conduit à l'aboutissement de ces deux articles et de cette thèse.

Mes derniers remerciements vont à toutes les personnes avec qui j'ai partagé mes études et notamment ces années de thèse.

A la mémoire de Marie-Giovanna Bireloze et Rosa Bendicho qui auraient aimé pouvoir lire ces lignes. Je dédie cette thèse à ma famille, qui, peu importe la distance, a toujours su être à mes côtés, me guider et me conseiller. Mes pensées vont également à tous mes amis de Paris, de Toulouse et de Nouvelle-Calédonie pour la joie et le soutien qu'ils m'apportent ainsi qu'à tous les enseignants m'ayant inspiré ma vocation de chercheur.

## Sommaire

<b>I.</b>	<b>ÉPISSAGE ET POLYADENYLATION DES ARN PRE-MESSAGERS .....</b>	<b>1</b>
I.1	ÉPISSAGE .....	1
I.2	MATURATION 3' : CLIVAGE ET POLYADENYLATION .....	3
I.3	ÉPISSAGE ALTERNATIF.....	6
I.4	POLYADENYLATION ALTERNATIVE .....	7
<b>II.</b>	<b>LA POLYADENYLATION INTRONIQUE (IPA) ET LES EXONS TERMINAUX ALTERNATIFS (ALE) ...</b>	<b>10</b>
II.1	DEFINITION ET CATEGORISATION .....	10
II.2	MECANISMES DE REGULATION DE LA POLYADENYLATION INTRONIQUE.....	12
II.2.1	<i>U1 snRNP.....</i>	<i>12</i>
II.2.2	<i>CDK12/CDK13.....</i>	<i>14</i>
II.2.3	<i>Facteurs du complexe de clivage/polyadénylation .....</i>	<i>14</i>
II.2.4	<i>Motifs ADN/ARN.....</i>	<i>15</i>
II.2.5	<i>Taux d'élongation et pauses d'ARN polymérase II.....</i>	<i>16</i>
II.2.6	<i>Marques épigénétiques et épitranscriptomiques.....</i>	<i>18</i>
II.2.7	<i>Exosome et dégradation des ARN.....</i>	<i>19</i>
II.3	REGULATION CELLULAIRE DES ISOFORMES DE POLYADENYLATION INTRONIQUE .....	19
II.3.1	<i>Régulation développementale et physiologique .....</i>	<i>19</i>
II.3.2	<i>Stress et maladies.....</i>	<i>20</i>
II.4	FONCTIONS DE L'IPA ET DES ISOFORMES IPA.....	21
II.4.1	<i>Fonction de l'IPA comme événement de terminaison prémature de la transcription.....</i>	<i>21</i>
II.4.2	<i>Fonctions des isoformes IPA.....</i>	<i>22</i>
<b>III.</b>	<b>PETITES ORF ET MICROPROTEINES .....</b>	<b>26</b>
III.1	METHODES DE DETECTION .....	28
III.1.1	<i>Le Ribo-seq.....</i>	<i>29</i>
III.1.2	<i>Spectrométrie de masse.....</i>	<i>30</i>
III.1.3	<i>Analyse Bioinformatique.....</i>	<i>30</i>
III.1.4	<i>Bases de données.....</i>	<i>31</i>
III.1.5	<i>Pertinence des ORF.....</i>	<i>32</i>
III.2	LOCALISATION DES ALT-ORF DANS LES ARN.....	32
III.3	EVOLUTION ET CONSERVATION .....	35
III.4	MECANISMES D'ACTION DES ALT-ORF ET MICROPROTEINES .....	37
III.4.1	<i>Réservoir évolutif.....</i>	<i>37</i>
III.4.2	<i>Mode d'action des microprotéines et des alt-ORF .....</i>	<i>38</i>
III.4.3	<i>Interactions avec facteurs .....</i>	<i>38</i>
III.4.4	<i>Excrétion/ligands.....</i>	<i>39</i>
III.4.5	<i>Microprotéines membranaires.....</i>	<i>40</i>
III.4.6	<i>Réponses antigéniques et perspectives thérapeutiques .....</i>	<i>40</i>
III.4.7	<i>Fonctions doubles ARN et protéines.....</i>	<i>41</i>

III.5	PREUVES FONCTIONNELLES DES ALT-ORF ET MICROPROTEINES .....	43
III.5.1	Détection de la microprotéine .....	43
III.5.2	Abondance, stabilité et purification .....	44
III.5.3	Implications physiologiques prouvées.....	44
<b>IV.</b>	<b>LES AGENTS GENOTOXIQUES ET LA REPOSE AUX DOMMAGES A L'ADN : FOCUS SUR LE</b>	
<b>CISPLATINE</b>	<b>47</b>	
IV.1	LES AGENTS GENOTOXIQUES .....	47
IV.2	LE CISPLATINE.....	47
IV.3	LA REPOSE AUX DOMMAGES A L'ADN .....	49
IV.3.1	Les voies de réparation.....	50
IV.3.2	Voies de signalisation et activation des points de contrôle du cycle cellulaire .....	59
IV.3.3	Régulation de l'expression des gènes.....	62
IV.3.3.1	Régulation co- et post-transcriptionnelle des ARN.....	63
IV.4	CISPLATINE ET CANCER .....	64
IV.4.1	Toxicité.....	65
IV.4.2	Résistance.....	66
IV.5	LIENS ENTRE IPA ET DDR .....	70
IV.5.1	Régulation des gènes de DDR au niveau de l'IPA .....	70
IV.5.1.1	Régulation par CDK12.....	70
IV.5.1.2	Régulation par U1 snRNA.....	74
IV.5.2	Régulation des isoformes IPA par les agents génotoxiques .....	75
<b>V.</b>	<b>DONNEES PRELIMINAIRES ET OBJECTIFS.....</b>	<b>77</b>
<b>VI.</b>	<b>ARTICLES .....</b>	<b>78</b>
VI.1	LES ISOFORMES DE POLYADENYLATION INTRONIQUE DANS LA PARTIE 5' DES GENES CONSTITUENT UNE SOURCE DE MICROPROTEINES ET SONT IMPLIQUEES DANS LA REPOSE CELLULAIRE AU CISPLATINE .....	78
VI.1.1	Introduction.....	78
VI.1.2	Conclusion .....	79
VI.2	LE BLOCAGE DE U1 SNRNA SENSIBILISE LES CELLULES AU CISPLATINE EN AUGMENTANT LES DOMMAGES A L'ADN	80
VI.2.1	Introduction.....	80
VI.2.2	Conclusion .....	81
<b>VII.</b>	<b>DISCUSSION ET PERSPECTIVES.....</b>	<b>82</b>
VII.1	MECANISMES DE REGULATION DU RATIO D'ISOFORMES IPA:LE EN REPOSE AU CISPLATINE ET U1-AMO ...	82
VII.2	TRADUCTIBILITE DE NOS ISOFORMES MIP-5'UTR IPA .....	84
VII.3	FONCTIONS ET DESTIN DE LA MICROPROTEINE PRKAR1B SORF 2.....	85
VII.4	CONCEPT DES MIP-5'UTR-IPA.....	86
VII.5	FONCTIONS DES ISOFORMES 5'UTR-IPA .....	87
VII.6	SENSIBILISATION PAR U1-AMO AU CISPLATINE : PERSPECTIVES ET LIMITES.....	88
<b>VIII.</b>	<b>BIBLIOGRAPHIE : .....</b>	<b>92</b>

## Listes des figures

Figure 1 : Reconnaissance des Exons/Introns par le spliceosome et cinétique de l'épissage (adaptée de Grange 2016).....	3
Figure 2 : Interaction des facteurs de la machinerie de terminaison 3'aux éléments de séquence du pré-ARNm (Tian et al. 2017).....	5
Figure 3 : Les principaux types d'épissage alternatif (de Dujardin et al. 2016).....	6
Figure 4 : Représentation des ALE et 3'UTR-APA (modifiée de Di Giammartino et al., 2011) ...	8
Figure 5 : Polyadénylation intronique - les différents ARNm générés par IPA dans les gènes (modifiée de Tian et al., 2007) .....	11
Figure 6 : Motifs ARN régulant l'IPA .....	16
Figure 7 : <i>Kinetic coupling model</i> .....	18
Figure 8 : Devenir et fonctions des isoformes IPA.....	24
Figure 9 : Modèle du <i>ribosome scanning</i> et cas particuliers.....	27
Figure 10 : Protocole du ribo-seq (adapté de Hsu et al. 2016) .....	29
Figure 11 : Origines des Alt-ORF chez l'Homme (modifiée de Orr et al. 2020).....	34
Figure 12 : Modes d'actions des microprotéines, alt-ORF et sORF (adapté de Schlesinger et Elsässer 2022).....	42
Figure 13 : Diversité des microprotéines caractérisées chez l'Homme avec leur homologue chez la souris (Hassel et al. 2023).....	46
Figure 14 : Modalités d'entrée et formation d'adduits du cisplatine (adapté de Song et al. 2022 ; Ghosh 2019) .....	49
Figure 15 : Réponses aux dommages à l'ADN (DDR) de cassures doubles brins, dommages simples brin et blocages de fourches de réplication (de Dutertre, Lambert, et al. 2014) .....	50
Figure 16 : Les voies de réparations et leurs dommages associés dus au stress génotoxique (de Hartwig et al. 2020).....	51
Figure 17 : Réparation par excision de nucléotides (NER) (modifiée de Scharer 2013) .....	52
Figure 18 : Interactions des protéines de la voie de réparation Fanconi pour réparer un pontage interbrin en phase S (de Kim et D'Andrea 2012). .....	55
Figure 19 : Cycle cellulaire et points de contrôle du cycle contrôlé par le DDR (du the Cancer Research Product Guide Edition 3, 2015).....	61
Figure 20 : Mécanismes de résistances au cisplatine (de Song et al. 2022) .....	67
Figure 21 : Modèle des miP-5'UTR-IPA.....	87

## Listes des Tableaux

Tableau 1 : Listes des outils bioinformatiques de prédiction d'alt-ORF/sORF (adapté de Leong et al. 2022)..... 31

Tableau 2 : Listes des bases de données de alt-ORF (adapté de Leong et al. 2022) ..... 32

# I. Épissage et polyadénylation des ARN pré-messagers

## I.1 Épissage

Il existe chez l'Homme trois types d'ARN polymérases nommées respectivement ARN polymérase I, II et III pouvant débiter un événement de transcription. Ces trois polymérases possèdent des caractéristiques propres leur permettant de s'associer à des facteurs régulant la transcription à différentes étapes, de l'initiation jusqu'à la terminaison (Calvo et al. 2003; Shatkin et al. 2000; Kwak et al. 2013). Ces polymérases reconnaissent des promoteurs différents sur l'ADN, guidées notamment par des séquences ADN favorisant la liaison de la polymérase (Chalkley et al. 1999; Vannini 2013). Les ARN polymérases synthétisent différents types d'ARN. Il est notamment admis que l'ARN polymérase I synthétise tous les ARN ribosomiques à l'exception de l'ARN 5S alors que l'ARN polymérase III polymérise les ARN de transfert, l'ARN ribosomique 5S et certains snARN. Nous nous focaliserons pour cette partie sur l'ARN polymérase II qui elle synthétise les ARN messagers (ARNm) et de petits ARN tels que les snARN, microARN, siARN et snoARN (Carter et al. 2009).

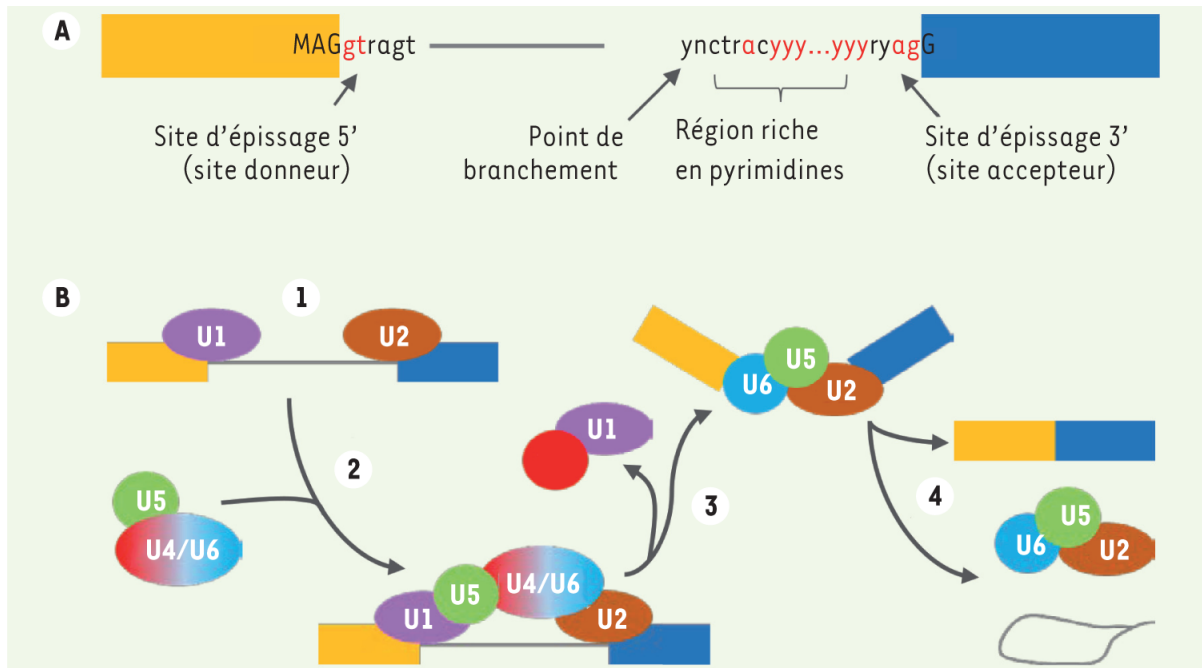
Les ARN pré-messagers subissent plusieurs modifications dues à des facteurs accompagnant l'ARN polymérase II durant la transcription (Shatkin et al. 2000; Calvo et al. 2003) au cours de leurs maturations en ARN messagers. Trois principales modifications sont couplées respectivement à l'initiation, l'élongation et la terminaison de la transcription : le coiffage en 5', l'épissage des introns et la maturation en 3' (Calvo et al. 2003; Shatkin et al. 2000). Il existe d'autres modifications des ARN pré-messagers comme la méthylation (Zaccara et al. 2019). Le coiffage ou *capping* fait intervenir plusieurs enzymes qui vont modifier l'extrémité 5' de l'ARN en ajoutant une guanosine méthylée au triphosphate du premier nucléotide transcrit. L'ARN est ainsi protégé des ribonucléases 5'→3'. La coiffe favorise plus tard l'export de la molécule dans le cytoplasme et la lecture de l'ARN par les ribosomes (Ramanathan et al. 2016).

L'épissage, deuxième modification, consiste en l'excision d'une partie de l'ARN synthétisé appelée intron et de la ligation des deux parties restantes appelées exons. Ce phénomène se produit en moyenne huit fois par gène (Sakharkar et al. 2004) et l'on peut alors définir des exons et des introns répartis de façon alternée pour un gène donné. L'identification des exons ou introns a été notamment possible historiquement grâce aux travaux sur l'ARNm

maturé de quelques gènes puis plus tard par RNA-seq sur l'ensemble du génome (Jacob et al. 2017). L'étude des facteurs d'épissage formant un complexe appelé spliceosome a aussi permis de mieux comprendre la cinétique de ce processus. L'épissage fait intervenir 5 snRNPs et plus d'une centaine de protéines, nécessitant plusieurs étapes consécutives (Wilkinson et al. 2020). Par l'étude du spliceosome, les critères déterminant si une séquence d'ARN donnée est un exon ou un intron ont pu être identifiés.

Des régions en 5' des introns nommées sites 5' d'épissage ou sites donneurs ont été montrées comme recrutant un facteur essentiel à l'épissage, U1 snRNP. Ce recrutement se fait notamment par la présence d'un motif de liaison sur l'ARN, GU. En 3' de l'intron se trouve le site 3' d'épissage ou site accepteur (typiquement AG), précédé d'une région riche en pyrimidines (U/C) et du point de branchement ; cette région 3' recrute U2snRNP. Une fois la région 3' liée à U2snRNP, une interaction entre U1 snRNP sur la région 5' et U2snRNP sur la région 3' a lieu. La formation de ce complexe favorise ensuite le recrutement de U4snRNP et U6snRNP recrutant à leur tour U5snRNP. Le spliceosome alors complet va dans un premier temps se séparer de U1 snRNP et rapprocher les extrémités de l'intron. Une réaction de transestérification catalysée par U2snRNP et U6snRNP va alors couper l'extrémité 5' de l'intron et la relier sur une adénosine (point de branchement) de la région U/G riche en 3' de l'intron. Cette structure en « lasso » est appelée lariat. Une deuxième réaction de transestérification se produit alors séparant l'extrémité 3' de l'intron de l'extrémité 5' de l'exon 2 et reliant les deux exons ensemble. Le spliceosome se dissocie alors et l'intron sous forme de lariat est dégradé (Figure 1). Ce processus va être répété au cours de la transcription autant de fois que le spliceosome reconnaîtra conjointement un site donneur et un site accepteur. À noter qu'il existe aussi un autre spliceosome dit mineur faisant intervenir d'autres facteurs (à l'exception de U5snRNP, commun aux deux complexes) dont nous ne détaillerons pas le processus, mais dont le fonctionnement est similaire (Bai et al. 2021).





**Figure 1 : Reconnaissance des Exons/Introns par le spliceosome et cinétique de l'épissage (adaptée de Grange 2016)**

**A.** Signaux de l'épissage reconnus par les complexes ribonucléoprotéiques du spliceosome. **B.** La réaction d'épissage Le complexe U1 s'associe au site donneur et U2 se fixe sur le point de branchement. (2) Les complexes U4/U5/U6 interviennent : le complexe U6 se fixe sur le complexe U2 tandis que le complexe U5 va se fixer sur l'extrémité 3' de l'exon (proche du site donneur d'épissage) à proximité du complexe U1. (3) Les complexes U1 et U4 sont libérés, le complexe U5 glisse sur l'intron et le complexe U6 se fixe sur le site donneur qui est coupé. (4) L'extrémité 3' de l'intron est coupée, ce qui libère l'intron (forme d'un « lasso ») et les exons sont liés. Enfin, le complexe se dissocie

La troisième modification de l'ARNm se déroule en fin de transcription. Il s'agit de la maturation en 3'.

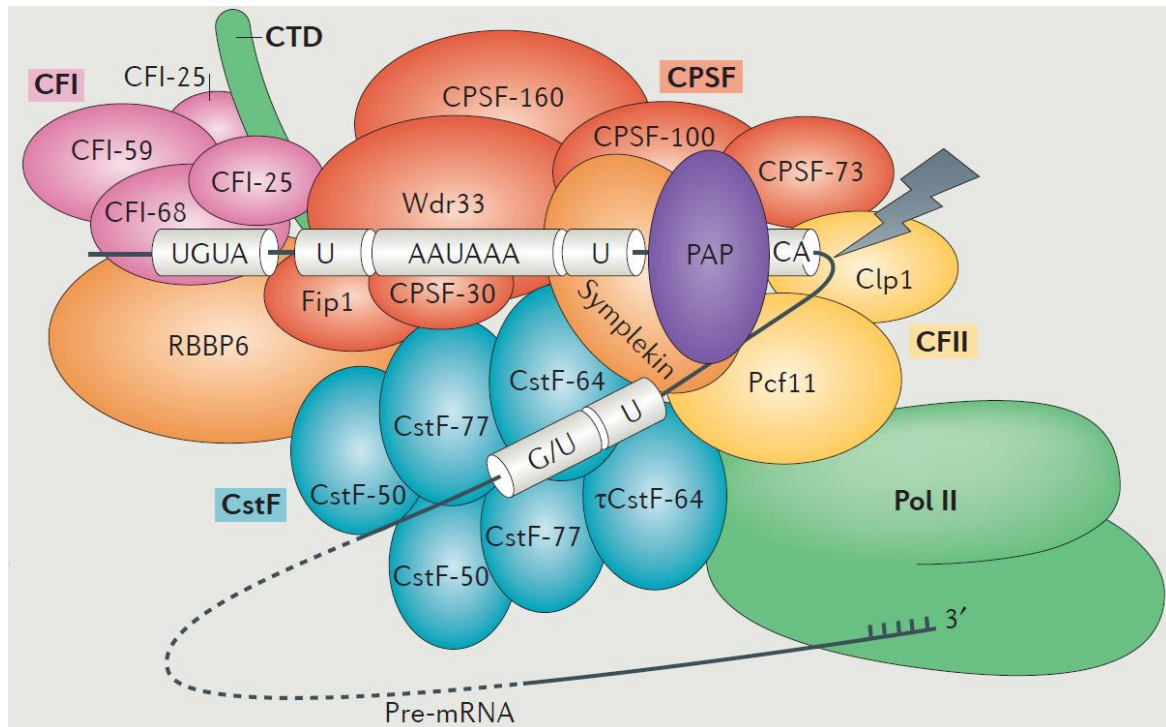
## I.2 Maturation 3' : Clivage et Polyadénylation

La maturation en 3' est divisée en deux étapes couplées : le clivage et la polyadénylation, qui comportent une machinerie d'environ 80 protéines. Quatre régions flanquantes au site de clivage sont importantes pour permettre le recrutement des facteurs de clivage/polyadénylation réalisant ces deux étapes (Figure 2). La principale se nomme le signal de polyadénylation (signal polyA). Elle est constituée d'une séquence conservée (AAUAAA pour 56% des sites ou AUUAAA pour 16% des sites) 15 à 30 nucléotides en amont du site de clivage/polyadénylation (site polyA) et recrute le complexe CPSF. Le complexe CPSF possède une endonucléase, CPSF3, effectuant le clivage et FIP1L1, un facteur recrutant avec CPSF1 la polyA polymérase. En aval du site de clivage (généralement CA) se lie le complexe CstF à une séquence G/U ou U riche appelée le Downstream Element Sequence (DES). Le rapprochement entre le complexe CPSF et CstF connectés par SYMPK est essentiel pour la

sélection du site de clivage. Il existe une variabilité du site de coupure (entre 20-30 nts) malgré un signal polyA précis, mais la fonction n'est à l'heure actuelle pas clairement définie (Tian et al. 2012). La troisième séquence en amont du signal polyA, U(G/A)UA, dite *upstream sequence element* (USE) liant le complexe CFIm (MacDonald et al. 2010) ainsi que la quatrième séquence G riche en aval de DES liant des facteurs inconnus ne participe qu'à améliorer la reconnaissance du site de clivage (Millevoi et al. 2006; Tian et al. 2012). En cas de dégénérescence du signal polyA, il a cependant été montré que les 3 autres séquences devenaient plus importantes pour une coupure correcte (Millevoi et al. 2006; Tian et al. 2012; Di Giammartino et al. 2011). Après clivage, la queue polyA, d'environ 250 nucléotides, est rajoutée par la polyA polymérase guidée par les autres facteurs (Millevoi et al. 2006; Eckmann et al. 2011). Cette longueur est limitée notamment par des facteurs d'export, des *RNA binding protein* (RBP) tel que PABN1 et NAB2 jouant par la suite un rôle dans l'export (Charlesworth et al. 2013; Nourse et al. 2020).

L'étape de clivage/polyadénylation est généralement couplée à la terminaison de la transcription par deux mécanismes principaux : un ralentissement/pause de l'ARN pol II dans des régions GC riches en aval du site polyA et le « torpillage » de l'ARN pol II par Xrn2 dégradant le transcrit en aval du site de clivage à partir de son extrémité 5' (Proudfoot 2016). Mais il existe aussi le cas du clivage cotranscriptionnel « CoTC », permettant une polyadénylation post-transcriptionnelle dans le nucléoplasme après clivage au niveau d'un site CoTC en aval du site polyA (Dye et al. 2001; West et al. 2008; Nojima et al. 2013). Il peut arriver dans des conditions particulières que le dernier signal polyA canonique du gène ne soit pas utilisé aboutissant à des événements de transcription dit « *read-through* ». Ce *read-through* peut aboutir à la génération de transcrit dit de fusion/chimérique avec un autre gène (Grosso et al. 2015; Peck et al. 2019).

À noter que l'événement de clivage au niveau du site polyA (ou du site CoTC) permet le relargage de l'ARNm de la chromatine (Nojima et al. 2013; Dutertre et al. 2021). Le détachement des complexes de clivage et de polyadénylation permet ensuite de continuer le processus d'export (Dutertre et al. 2021).



**Figure 2 : Interaction des facteurs de la machinerie de terminaison 3'aux éléments de séquence du pré-ARNm (Tian et al. 2017)**

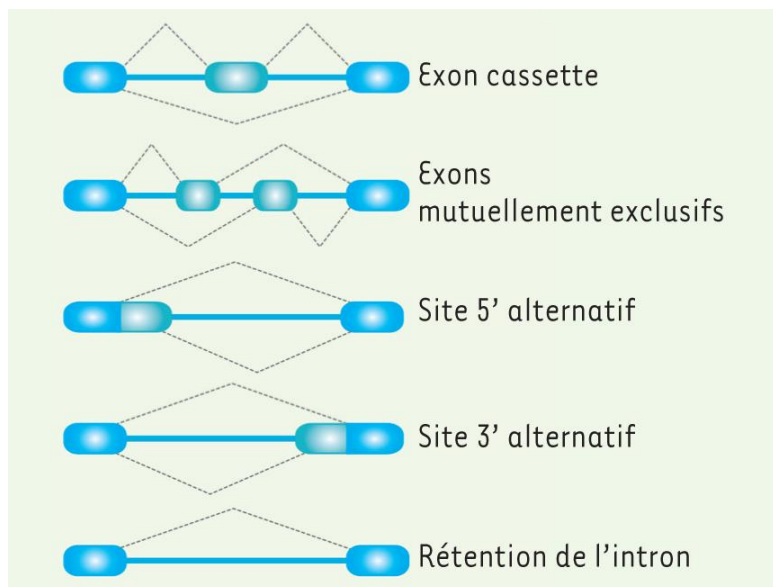
Localisation des cis éléments régulateurs en amont et en aval du site de coupe souvent localisés après un ACA (ici éclair). Les éléments en amont incluent une séquence A[A/U]UAA et d'autres variants dites séquences polyA, des séquences U-riche et UGUA. Les séquences en aval incluent des séquences U et GU riches. La fréquence de ces séquences peut varier suivant la localisation du site APA (moins de UGUA et séquences GU riches dans les APA proximal). 4 complexes protéiques ainsi que des protéines seules lient ces séquences. *Cleavage and polyadenylation specificity factor* (CPSF) se lie à la séquence polyA et aux régions U riches flanquantes. *Le Cleavage Factor I* (CFI) se lie au UGUA et *Cleavage stimulating Factor* (CstF) aux séquences U et GU riche en aval du site de coupe. L'activité endonucléase est portée par le complexe CPSF. D'autres protéines comme l'ARN pol II (queue CTD) ou la symplekin assurent un rôle de support à la machinerie.

La queue polyA protège les ARNm de la dégradation par liaison avec des protéines telles que PABP empêchant la déadénylation (Nourse et al. 2020; Eckmann et al. 2011). La polyadénylation est connue depuis longtemps comme améliorant la traduction par formation de *close loop complex* entre PABP liée à la queue polyA et eIF4G qui fait partie du complexe eIF4F, facteur d'initiation de la traduction liant la coiffe (Eckmann et al. 2011). Elle est aussi impliquée dans la stabilité de l'ARNm, l'export du noyau, et peut aussi être la cible de régulation (Nourse et al. 2020).

Sur un ARN pré-messager, il peut exister plusieurs sites polyA ainsi que plusieurs sites accepteurs ou donneurs d'épissage. On parle alors de polyadénylation alternative et d'épissage alternatif, respectivement.

### I.3 Épissage alternatif

L'épissage alternatif se définit comme l'utilisation facultative de sites d'épissage au cours de la maturation des transcrits d'un gène donné, générant des isoformes d'ARNm avec un contenu exonique différent. En effet, au sein d'un gène, plusieurs sites donneurs (en 5') ou plusieurs sites accepteurs (3') peuvent rentrer en compétition. Cela conduit à des exons alternativement inclus ou exclus, ou à des variations de longueur d'exon. De plus, l'excision d'un intron peut être facultative, on parle de rétention d'intron (Figure 3).



**Figure 3 : Les principaux types d'épissage alternatif (de Dujardin et al. 2016)**

Certains exons peuvent suivant les transcrits être inclus ou exclus (exon cassette). D'autres exons, peuvent s'exclure mutuellement d'un transcrit. Lorsqu'un exon possède un site d'épissage 3' ou 5' faible, ce dernier peut ne pas être utilisé et un site d'épissage plus en amont ou en aval respectivement peut être pris (site 5' et 3' alternatif). Dans certains cas, la totalité d'un intron peut être inclus dans un transcrit ARNm (intron rétention).

Chez l'Homme, l'épissage alternatif concerne 95% des gènes connus avec en moyenne quatre transcrits par gène (Pan et al. 2008; E. T. Wang et al. 2008; Yates et al. 2016). L'épissage alternatif est régulé physiologiquement en fonction du tissu, type cellulaire, stade de développement, des stimuli et stress, et dans diverses maladies telles que le cancer (Marasco et al. 2023; Anczuków et al. 2016; Cherry et al. 2020).

Le taux d'utilisation d'un site d'épissage alternatif (par exemple par rapport à un site compétiteur) dépend en partie du degré de conservation des motifs reconnus par le spliceosome (jonctions Exon/intron [sites 5' et 3' d'épissage], point de branchement, région riche en pyrimidines). De nombreux facteurs d'épissage (protéines régulatrices) se lient à des motifs ARN régulateurs (activateurs ou répresseurs) situés à proximité des sites d'épissage,

dans les exons ou les introns, et modulent le recrutement ou l'activité locale du spliceosome (Marasco et al. 2023). Les dérégulations de l'épissage dans les maladies sont souvent dues à des mutations des sites d'épissage ou de motifs régulateurs, ou à des variations d'expression des facteurs d'épissage (Grange 2016; Marasco et al. 2023; Anczuków et al. 2016; Cherry et al. 2020).

L'épissage intervenant en général de façon cotranscriptionnelle, l'épissage alternatif est aussi régulé par la vitesse de transcription de l'ARN polymérase II, qui peut être modulée (Moehle et al. 2014). Une vitesse lente tend à favoriser l'inclusion d'exons alternatifs, mais peut aussi favoriser l'exclusion de certains exons (Marasco et al. 2023). De plus, des marques d'histone peuvent moduler l'inclusion d'exons alternatifs, soit en modulant la vitesse d'élongation de la transcription, soit en recrutant des facteurs d'épissage (Luco et al. 2010).

Un cas particulier de l'épissage alternatif est la rétention d'un intron normalement excisé dans la séquence finale de l'ARNm mûr. Cette rétention d'intron est favorisée notamment lorsque les motifs ARN des sites donneurs et accepteurs sont très dégénérés ou lorsque les introns sont courts avec un fort taux en GC (Jacob et al. 2017).

L'épissage alternatif a été montré pour de nombreux gènes comme générant des isoformes protéiques différentes modulant l'activité de la protéine, sa stabilité, ses interactions moléculaires et/ou sa localisation subcellulaire (Kelemen et al. 2013). Avec environ 20 000 gènes connus, l'épissage alternatif permettrait une génération estimée de plus de 100 000 protéines différentes (Grange 2016) bien que cette diversité n'a pas encore été montrée à l'échelle du protéome *in vivo*. Cette absence de détection peut s'expliquer par les limitations techniques en protéomique (sensibilité de détection, etc.) et à l'instabilité possible de certaines isoformes protéiques. Par ailleurs, de nombreux variants d'épissage sont instables au niveau ARN du fait de la présence d'un codon stop prématuré, situé en amont du dernier exon du transcrit, conduisant à sa dégradation (*nonsense-mediated RNA decay* [NMD]) (Powers et al. 2020).

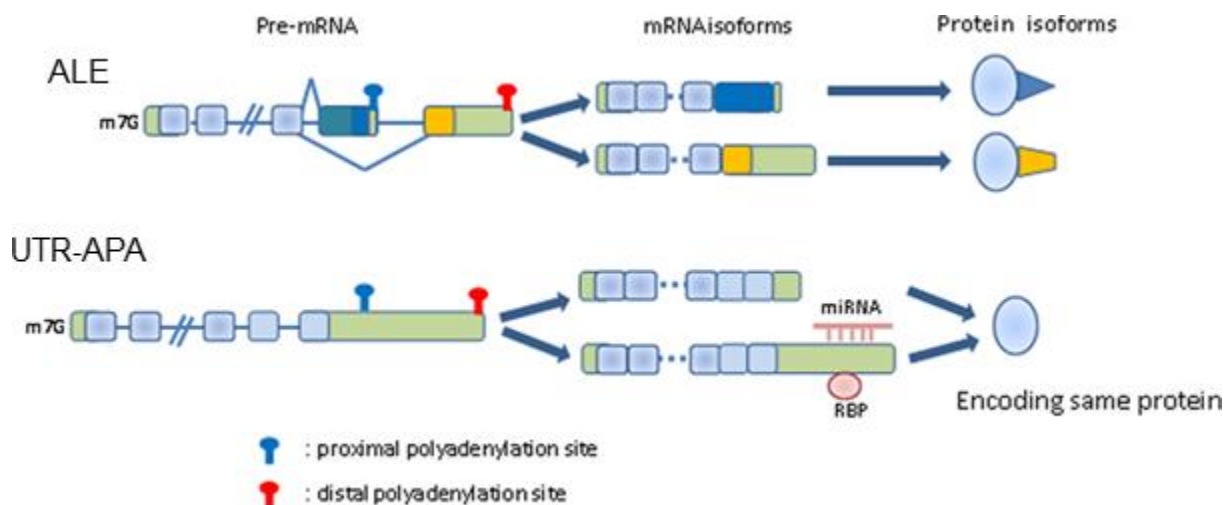
#### I.4 Polyadénylation alternative

De la même façon qu'il existe des sites d'épissage alternatifs dans un ARN pré-messager, il peut y avoir plusieurs sites polyA (Di Giammartino et al. 2011). La sélection d'un site polyA plutôt qu'un autre se nomme polyadénylation alternative (APA). Les différents sites

polyA peuvent être situés soit dans un même exon (sites polyA en tandem, 3'UTR-APA) soit dans des exons différents, appelés exons terminaux alternatifs (*alternative last exons*, ALE, voir Figure 4), on parle dans ce dernier cas de polyadénylation intronique (IPA) (Tian et al. 2012). Chez l'Homme, des isoformes 3'UTR-APA et IPA ont été détectées pour environ 60% et 20% des gènes codants, respectivement. Les isoformes 3'UTR-APA ont été les plus étudiées.

Le choix entre deux sites polyA en tandem a été montré comme pouvant être dû au niveau d'expression des facteurs de polyadénylation (Tian et al. 2017). Par exemple, une faible quantité de la protéine CstF promeut l'usage du site polyA distal alors qu'une faible quantité de CFI favorise les sites proximaux (Tian et al. 2017). Les modifications post-traductionnelles des protéines du complexe polyA peuvent aussi influencer l'APA (Di Giammartino et al. 2011). Par ailleurs, la vitesse d'élongation de transcription influence le choix du site polyA (Devany et al. 2016; Tian et al. 2017). Une élongation ralentie favorise les sites polyA proximaux. Des interactions entre facteurs d'épissage et complexes polyA ont aussi été décrites (Tian et al. 2017) comme médiant une modulation réciproque des deux mécanismes.

Des promoteurs alternatifs (Alfonso-Gonzalez et al. 2023) et des *enhancers* peuvent participer à la régulation de sites polyA alternatifs (Kwon et al. 2022). Des marques épigénétiques pourraient aussi jouer un rôle dans la régulation de l'APA et de la terminaison de la transcription (Di Giammartino et al. 2011). Récemment, l'APA a pu être montrée comme régulée par la méthylation de l'ADN et les boucles chromatinienne (Nanavaty et al. 2020; Terrone et al. 2022). Une forte déplétion de nucléosome a aussi été observée au niveau des sites polyA (Tian et al. 2017).



**Figure 4 : Représentation des ALE et 3'UTR-APA (modifiée de Di Giammartino et al., 2011)**  
 Les ALE donnent des isoformes protéiques avec des régions C-terminal distinctes là où les 3'UTR-APA produisent des ARNm avec des 3'UTR différentes donnant la même protéine. Les 3'UTR plus longues contiennent en général

plus de sites de liaisons aux protéines et aux miARN pouvant déstabiliser et réprimer post-transcriptionnellement les ARNm. ALE, *alternative last exon* ; UTR-APA, *3'UTR-alternative polyadenylation*. Boîte verte : régions non traduites ; Boîte bleu clair : régions codantes partagées ; Boîte bleu foncé et jaune : régions codantes non exclusives ; lignes : introns.

D'un point de vue fonctionnel, les événements d'APA en tandem changent la longueur de la 3'UTR d'un ARNm (d'où son autre appellation de 3'UTR-APA), ce qui altère la présence (et la position) de sites de liaison aux miARN ou à des protéines liant les ARN (*RNA Binding Protein*, RBP). Cela peut aboutir à réguler la stabilité, la traduction ou la localisation des ARNm mais aussi le recrutement d'une protéine interagissant avec la protéine codée, pouvant ainsi affecter sa fonction (par exemple sa localisation) (Di Giammartino et al. 2011; Mayr 2019). En général, une 3'UTR plus longue est plus susceptible de régulation négative par les miARN (Di Giammartino et al. 2011; Mayr 2019), toutefois la taille de la 3'UTR en elle-même pourrait aussi jouer un rôle dans cet effet répressif (Mayr et al. 2009). Contrairement au 3'UTR-APA, l'IPA a souvent lieu au sein de la région codante des gènes, on parle alors de *coding region alternative polyadenylation*, CR-APA (S.-H. Lee et al. 2018). Nous utiliserons les appellations ; isoformes canoniques, isoformes pleine longueur, isoformes longues et isoformes « *last exon* », pour faire référence dans un gène comportant un IPA à l'ARNm transcrit n'utilisant pas l'IPA. Dans la partie suivante, je discuterai du cas de l'IPA qui a été le centre d'intérêt de ma thèse.

## II. La polyadénylation intronique (IPA) et les exons terminaux alternatifs (ALE)

### II.1 Définition et catégorisation

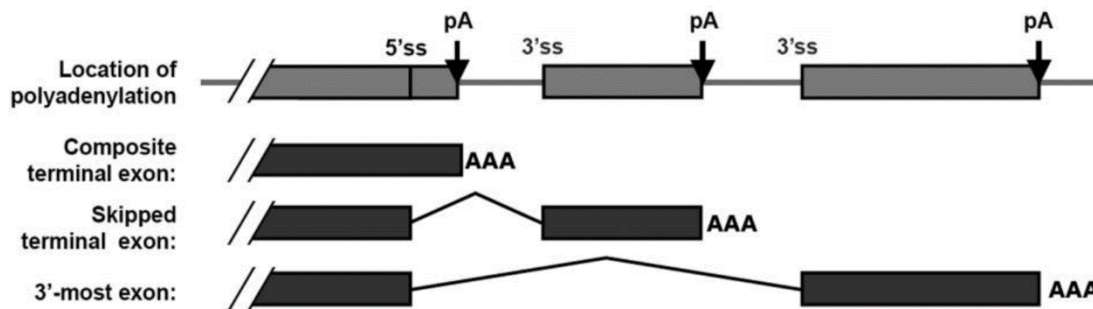
Documentés d'abord pour seulement quelques gènes codant les IgM (F. W. Alt et al. 1980) ou la calcitonine (Amara et al. 1984), les ALE se sont montrés plus étendus à l'échelle du génome. En effet, en 2007 en utilisant des bases de données de cDNA/ESTs, une étude a montré qu'environ 20% des gènes contiennent au moins un événement d'ALE (Tian et al. 2007). Les puces exons et le RNA-seq classique (Illumina) ont permis l'étude de la régulation des ALE (Kaida et al. 2010), mais c'est le développement de méthodes de RNA-seq focalisés sur l'extrémité 3' des transcrits (3'-seq, PAS-seq, 3P-seq, 3'READS) qui a permis de découvrir systématiquement les sites IPA dans différents tissus et lignées cellulaires. En 2013, une étude par 3'-seq chez la souris a trouvé qu'environ 40% des gènes avaient des événements de polyadénylation dans les introns ou les exons internes (Hoque et al. 2013). Une étude sur plus de 46 échantillons de tissus humains sains, lignées primaires et des extraits de myélome multiple (MM) a permis de mettre en évidence 4927 événements d'IPA avec un niveau de confiance élevé (Singh et al. 2018). Prises ensemble, ces études montrent donc l'étendue du processus d'IPA à travers le génome, les tissus et les mammifères. Ces 3'-seq ont permis la publication de bases de données en ligne répertoriant les événements d'IPA fiables à travers les différents tissus, stress, pathologies et espèces étudiées (Gruber et al. 2019; R. Wang et al. 2018).

Des algorithmes ont aussi été développés afin d'identifier plus facilement les événements d'IPA. D'autres essaient d'égaliser le 3'-seq en utilisant des bases de données de RNA-seq déjà disponibles et plus communes pour trouver des IPA. Ils se basent sur les caractéristiques identifiées des introns générant des IPA, mais restent cependant limités avec notamment une difficulté à identifier les IPA *De Novo*. Un nouvel algorithme permet leur identification en se basant sur des changements de nombres de *reads* pour une séquence donnée en RNA-seq témoignant d'une terminaison de transcription prémature (Z. Zhao, Xu, Wei, Wang, et al. 2021).

De façon plus générale, les introns associés aux IPA sont plus longs que les introns n'en générant pas et sont enrichis dans les gènes longs (Tian et al. 2007; Singh et al. 2018). Deux



principaux types topologiquement différents d’ALE générés par IPA sont décrits : le *composite terminal exon* correspond à une extension alternative d’exon (allant du site 5’ d’épissage au site polyA) ; le *skipped terminal exon* correspond à un exon entier alternatif (allant d’un site 3’ d’épissage à un site polyA ; Figure 5). Il est à noter que la polyadénylation intronique permet comme observé sur la figure « d’exoniser » une partie de l’intron qui ne se retrouve habituellement pas dans l’ARNm mûré. Nous discuterons plus tard des éventuelles conséquences qui en découlent.



**Figure 5 : Polyadénylation intronique - les différents ARNm générés par IPA dans les gènes (modifiée de Tian et al., 2007)**

pA ; site de polyadénylation, 5'SS ; site 5' d'épissage, 3'SS ; site 3' d'épissage, AAA ; queue polyA

Les facteurs déterminant si un signal polyA dans un intron va aboutir à un *skipped terminal exon* ou un *composite terminal exon* sont nombreux. L'absence de conservation du site 5' d'épissage a par exemple pu être associée au *composite terminal exon* ; alors qu'un signal polyA conservé (AAUAAA) et plus proche de l'extrémité 5' du gène s'est montré corrélé au *skipped terminal exon* (Tian et al. 2007; Singh et al. 2018). En contradiction, une étude a montré que les *skipped terminal exon* se produisaient dans des introns plus longs que les *composite terminal exon* tandis que l'inverse a été conclu par une autre (Tian et al. 2007; Singh et al. 2018). Par ailleurs, une étude des mutations ponctuelles (SNPs) dans 4998 tumeurs de patients a mis en évidence trois caractéristiques orientant la maturation de l'ARN associées aux gènes mutés en IPA plutôt qu'en intron rétention : les introns générant de l'IPA sont plus longs, plus pauvres en GC et contiennent un signal polyA plus fort comparés aux introns générant de l'intron rétention (Z. Zhao, Xu, Wei, Huang, et al. 2021).

Ces caractéristiques permettant à des introns de générer préférentiellement de l'IPA supposent des mécanismes de régulations comme causes ou conséquences de ces caractéristiques observées sur ces introns. L'IPA est en effet régulée par de nombreux acteurs.

## II.2 Mécanismes de régulation de la polyadénylation intronique

Les isoformes issues d'IPA sont régulées par des facteurs à différentes étapes de la genèse et de la maturation de l'ARNm au moment de la transcription, de l'épissage et de l'étape de clivage/polyadénylation.

### II.2.1 U1 snRNP

La U1 snRNP, qui est plus abondante dans les cellules que les autres snRNP du spliceosome, s'est révélée avoir un rôle majeur dans la répression des sites IPA, notamment de nombreux sites « cryptiques » (Berg et al. 2012; Kaida et al. 2010). Le terme « Cryptiques » se rapporte aux sites visibles seulement lorsque l'on induit de façon exogène une répression/inhibition d'un facteur donné. Il a été montré dans ces études dans les lignées cellulaires HeLa que l'inhibition de U1 snRNA, composant de U1 snRNP, entraînait l'induction de transcrits courts issus d'événements d'IPA. Les auteurs ont pu inhiber l'activité de liaison de U1 snRNA au site 5' d'épissage par électroporation d'un morpholino oligomère (ADN chimiquement modifié) complémentaire appelé « U1-AMO » (Berg et al. 2012; Kaida et al. 2010). D'abord par puce exon sur les chromosomes 5, 7 et 16 puis par HIDE-seq, un séquençage comparant les niveaux d'expressions entre deux conditions (pan-génomique), avec une dose très forte de U1-AMO (7,5  $\mu$ M) à un temps de 8 h, les auteurs ont bien sûr observé que U1-AMO réprimait l'épissage des ARNm. Mais aussi qu'il induisait des transcrits finissant à moins de 5 kb du site d'initiation de la transcription annoté du gène et dus à un événement d'IPA. Cet effet semble dose-dépendant, plus on inhibe U1 snRNP plus les événements d'IPA sont précoces et nombreux (à faible dose, U1-AMO induit un raccourcissement des régions 3'UTR des transcrits, appelé *3'UTR shortening*). En inhibant d'autres facteurs d'épissage tels que SF3B1, du complexe U2 snRNP, à l'aide d'un inhibiteur chimique (spliceostatin A) ou par U2-AMO, les auteurs n'ont pu observer le même phénotype vis-à-vis de cette terminaison prématurée, concluant à un effet de U1 snRNP indépendant de sa fonction d'épissage. Par séquençage HIDE-seq, ce serait plus de 6548 gènes présentant des différences de séquences de transcrits à la suite du traitement U1-AMO à 7,5  $\mu$ M (clivage et polyadénylation prématurée (PCPA) et intron rétention inclus). Ces résultats ont été étendus à des cellules murines (3T3) et de drosophile (S2). Par construction d'un minigène NR3C1, les auteurs ont déterminé que U1 snRNP protège les ARNm de l'IPA jusqu'à une distance de 3,5 kb en aval du site 5' d'épissage. L'existence de sites 5' d'épissage cryptiques à l'intérieur des

introns longs permettrait d'étendre la portée de cette protection. Cette hypothèse a été confirmée par expérience de *crosslink* ultraviolets (UV) avec U1 snRNP par RAP-RNA (Engreitz et al. 2014). La surabondance de U1 snRNP par rapport aux autres snRNPs du spliceosome dans les cellules, alors que le complexe s'assemble de façon équimolaire, permettrait cette protection efficace des gènes contre le PCPA (Kaida et al. 2010).

À l'échelle du génome, l'inhibition par U1-AMO touche préférentiellement les gènes longs (39 kb) et affecte peu les gènes de taille faible (6,8 kb) voire favorise leur transcription (Oh et al. 2017). Cette étude indique que les 3590 gènes réprimés au niveau IPA par U1-AMO sont enrichis dans certaines fonctions, telles que la réplication de l'ADN et la réponse aux dommages à l'ADN (Oh et al. 2017). Cette étude ne valide toutefois pas cette observation dont je discuterai plus tard.

Ces effets de U1 snRNP sur l'IPA ont été confirmés dans des cellules endothéliales de veines ombilicales sur 31 isoformes IPA induites dans des gènes de récepteurs à la tyrosine kinase (Vorlová et al. 2011). Par la suite, ces effets ont aussi été validés dans des cellules de myoblaste C2C12, avec environ 100 gènes terminant dans les 2 kb autour du site d'initiation de la transcription (TSS) et 700 avec une CDS tronquée par répression modérée de U1 snRNA avec des siARN (une inhibition importante étant létale) ou U1-AMO (Wencheng Li et al. 2015).

La façon dont U1 snRNP réprime l'IPA est pour le moment mal comprise. *In vitro* dans des extraits de cellules HeLa, la protéine U1-70k (qui fait partie de U1 snRNP) peut interagir avec la poly(A) polymérase  $\alpha$ , responsable majeur de la polyadénylation des ARNm (Gunderson et al. 1998). Cette interaction inhibe alors la polymérase. Elle n'empêche cependant pas le recrutement des facteurs de clivage et polyadénylation au site polyA par gêne stérique et n'est pas impliquée dans l'inhibition de clivage observée. De plus, la position relative du site 5' d'épissage recrutant U1 snRNP par rapport au site de clivage/polyadénylation se révèle aussi importante (Vagner et al. 2000). Dans des cellules HeLa, il a été noté qu'U1 snRNP peut former un complexe avec CFIm, CPSF et CstF, des facteurs de PCPA, sans être associée aux autres éléments du spliceosome. Ce complexe s'associerait avec notamment des facteurs d'élongation de la transcription et d'export des ARNm. Cette étude montre également que U1-AMO modifie la conformation du complexe permettant le recrutement par ce dernier d'autres facteurs activateurs du PCPA expliquant en partie l'effet de U1-AMO sur l'IPA (So et al. 2019). Plus récemment, une étude a montré que U1 snRNP était responsable dans des cellules de mammifères de l'accélération de la transcription de l'ARN polymérase II dans les introns.

L'association de U1 snRNP avec le complexe d'élongation aux sites 5' d'épissage est nécessaire pour stimuler l'élongation et réduire ainsi la fréquence de terminaison prémature et d'arrêt de transcription. U1 snRNP médierait l'accélération dans les introns longs AT riches évitant ainsi les problèmes de transcription (Mimoso et al. 2023).

La déplétion d'autres facteurs du spliceosome tels que SF3B1, du complexe U2 snRNP, et U2AF65, associé à ce complexe, ont été montrés dans des myoblastes comme pouvant favoriser des événements d'IPA dans des introns qui sont normalement efficacement épissés (Wencheng Li et al. 2015). Cet effet de siARN SF3B1 contredit la précédente étude utilisant la spliceostatin A présentée plus tôt. Une autre étude utilisant la spliceostatin A sur des cellules HeLa, observe une induction d'IPA modérée dans 168 introns, doublant le nombre d'IPA en moyenne. Cette étude montre toutefois que la spliceostatin A piège U1 snRNP et limite son renouvellement. Le rôle d'inducteur d'IPA par cet inhibiteur pourrait donc être en partie dû à un effet sur U1 snRNP (Yoshimoto et al. 2021) Grâce à des RNA-seq, il a été aussi montré un lien corrélatif entre induction d'IPA et déplétion des facteurs U2AF1 et U2AF2 dans des fibroblastes chez l'Homme (Z. Zhao, Xu, Wei, Wang, et al. 2021).

## II.2.2 CDK12/CDK13

CDK12 est connu pour réguler l'élongation de la transcription de l'ARN pol II en phosphorylant avec l'aide de la cycline K le domaine C-ter, la plus grande sous-unité de l'ARN pol II (Davidson et al. 2014; Tellier et al. 2020). Une déplétion de CDK12 dans des cellules de neuroblastome et des cellules de cancer du côlon HCT116 par un inhibiteur CDK12 a été montrée comme induisant une inhibition de l'élongation et de la processivité de l'ARN pol II (Krajewska et al. 2019; Chirackal Manavalan et al. 2019). D'abord décrite par Dubbury dans des cellules murines ES, une terminaison prémature de la transcription associée à une IPA a été observée suite aux traitements avec cet inhibiteur. Cette polyadénylation s'est montrée comme ciblant préférentiellement les gènes longs (> 45 kb). Une corrélation entre la longueur du gène et sa répression par l'inhibiteur a notamment été montrée, plus les gènes étaient longs, plus ils étaient susceptibles d'être réprimés (Krajewska et al. 2019). Nous reparlerons de façon détaillée de la régulation de l'IPA par CDK12 en partie IV.

## II.2.3 Facteurs du complexe de clivage/polyadénylation

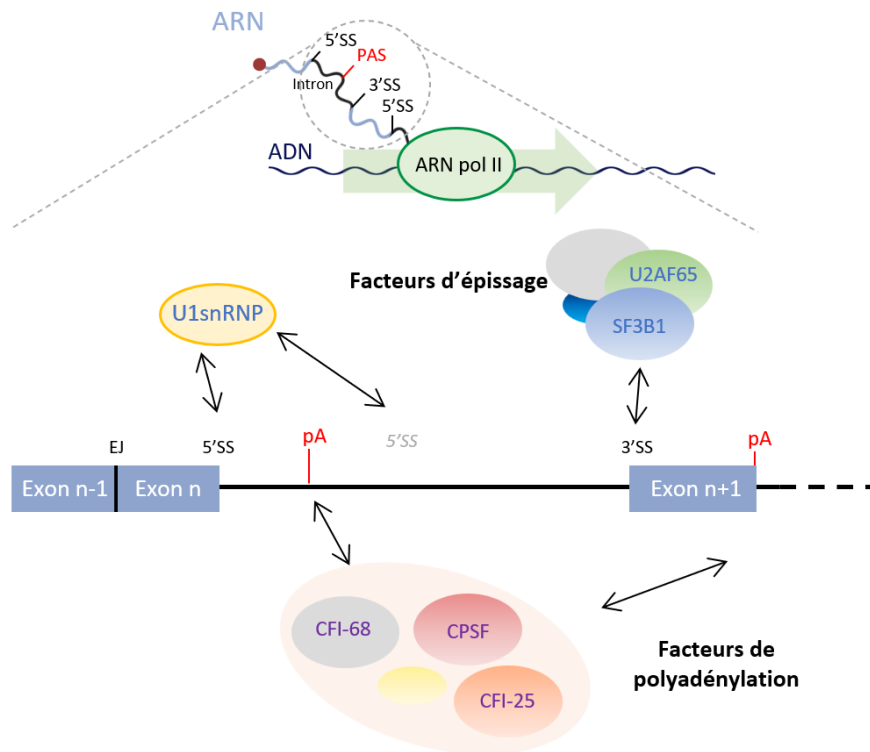
De nombreux facteurs du complexe de clivage/polyadénylation ont été montrés comme impliqués dans la modulation de polyadénylation alternative (Dharmalingam et al.

2021; Xiang et al. 2018; Wencheng Li et al. 2015). Des exemples ont aussi pu être documentés en ce qui concerne l'IPA. Récemment, la diminution du complexe CPSF dans des cellules de kératinocytes a été montrée comme induisant des événements d'IPA notamment dans GRHL3, diminuant ainsi l'expression de l'ARNm pleine longueur (X. Chen et al. 2021). Par ailleurs, une étude a accompli un criblage en siARN de nombreux facteurs du complexe clivage/polyadénylation, suivi d'un séquençage profond dans les cellules C2C12. Ils ont notamment trouvé que la répression de PABPN1, CFI-68, CFI-25 et Pcf11 pouvait induire de l'IPA terminant dans les 2 kb après le TSS dans une centaine de gènes (Wencheng Li et al. 2015; R. Wang et al. 2019). Étonnamment, la répression de PABPN1, protéine liant et protégeant la queue polyA, induirait aussi de l'IPA dans les 2 kb après le TSS lui aussi dans une centaine de gènes. Les chercheurs font alors l'hypothèse que cela pourrait être liée à son rôle dans la dégradation de l'ARN (Wencheng Li et al. 2015). En parallèle de cette étude, la distance entre les sites polyA alternatifs s'est révélée importante dans leur régulation notamment pour CFI-25 et CFI-68. Un autre facteur, CstF-77, pourtant compris dans ce criblage de siARN mais ne donnant pas de résultats remarquables s'est révélé dans une autre étude comme pouvant induire un événement d'IPA au sein même de son gène (W. Luo et al. 2013). Il semblerait que la polyA polymérase non canonique (star-PAP) réprime aussi l'IPA dans des cellules HEK (Weimin Li et al. 2017).

#### II.2.4 Motifs ADN/ARN

Comme nous avons pu le voir, de nombreux facteurs liant l'ARN favorisent les événements d'IPA. De ce fait, la conservation des motifs ARN de liaison à ces facteurs est un élément important pour réguler l'IPA (Z. Zhao, Xu, Wei, Huang, et al. 2021). La présence de motifs de sites 5' d'épissage plus ou moins conservés est par exemple très importante pour le rôle protecteur de U1 snRNP vis-à-vis de l'IPA (Berg et al. 2012; Kaida et al. 2010; Figure 6).

Par ailleurs, des motifs ADN lus par l'ARN pol II peuvent aussi favoriser l'utilisation d'IPA de par leur influence sur la vitesse d'élongation. C'est notamment le cas des îlots CpG, favorisant les pauses de pol II, montrés comme corrélés à l'utilisation des sites polyA en 5' dans les introns (Chiu et al. 2018). Une autre étude corrélative en cellules murines 3T3/NIH et humaines a également montré que les îlots CpG intragéniques associés à H3K4me3 favorisent l'IPA (Amante et al. 2020).



**Figure 6 : Motifs ARN régulant l'IPA**

Lors de la synthèse d'un ARNm par l'ARN pol II, des motifs ARN vont émerger comme des sites d'épissage en 5' ou en 3' ou des signaux polyA. Ces motifs vont être reconnus par différents facteurs appartenant au spliceosome ou contribuant à la polyadénylation. Ces facteurs modulent l'utilisation des signaux polyA. Certains facteurs du spliceosome comme U1 snRNP répriment les signaux polyA situés dans les introns. Les facteurs de polyadénylation peuvent favoriser ou réprimer certains sites polyA. PAS : signal polyA ; 5'SS : site 5' d'épissage ; 3'SS : site 3' d'épissage ; 5'SS en gris : site 5' d'épissage dégénéré ; EJ : Jonction d'exon ; Noms en bleu : Facteurs réprimant l'IPA ; Noms en violet : Facteurs réprimant ou activant l'IPA

## II.2.5 Taux d'élongation et pauses d'ARN polymérase II

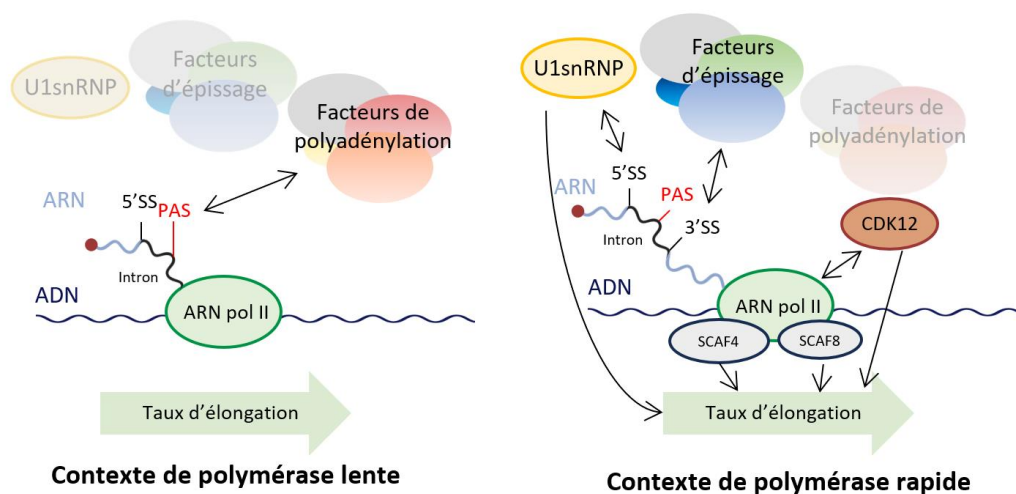
Le taux d'élongation désigne la vitesse à laquelle une ARN polymérase synthétise. En général pour l'ARN pol II, il dépend du contenu en GC et se situe aux alentours de 1000 pb/min (Zamft et al. 2012). Il est à distinguer de deux événements : le *backtracking* de l'ARN pol II qui lui permet de corriger des erreurs de complémentarité de séquences ADN-ARN lors de la synthèse ; et la pause de l'ARN polymérase, qui est un événement de blocage de l'ARN polymérase sur l'ADN. Les causes d'un blocage sont multiples comme les dommages sur l'ADN, des ARN ou ADN polymérase active sur le passage, des nucléosomes ou encore des structures secondaires de l'ARN (Noe Gonzalez et al. 2021). Intrinsèquement, un blocage d'ARN polymérase ou le *backtracking* suivis tous deux d'une reprise de transcription sens peuvent être interprétés comme une diminution du taux d'élongation par des méthodes de séquençage sur l'ARN naissant (GRO-seq ou PRO-seq). Le *backtracking* est d'ailleurs considéré comme mineur dans les études à l'échelle du génome. Pour discerner blocage et perte de vitesse, des méthodes existent comme le ChIP-seq avec des anticorps dirigés contre les sérines

phosphorylées correspondant à l'ARN polymérase à l'arrêt, mais elles sont peu répandues et l'amalgame est souvent fait entre ces deux événements. Une des raisons est que, fonctionnellement, ces deux mécanismes auraient les mêmes conséquences. En effet, le modèle derrière cette idée, le « *kinetic coupling model* », serait qu'une polymérase bloquée ou en perte de vitesse donnerait l'avantage à des facteurs liant l'ARN naissant plus en 5' par rapport à ceux se liant plus en 3' (par exemple pour deux sites 3' d'épissage en compétition) (Naftelberg et al. 2015; Figure 7). Dans le cas de l'IPA, un blocage ou ralentissement dans l'intron en 3' d'un site polyA favoriserait un événement d'IPA.

Une technique appelée CRISPRpas utilisant une deadCas9 se base justement sur de tels blocages de l'ARN pol II pour générer de l'IPA (Shin et al. 2021). En ciblant la région en 3' d'un site polyA par sgARN, il est en effet possible de recruter une protéine Cas9 guidée qui va bloquer la progression de l'ARN pol II et induire une pause. Cette pause va alors promouvoir l'utilisation du site polyA en amont et générer une isoforme IPA au détriment de la forme longue de l'ARN.

Certains facteurs vus précédemment tels que CDK12 ou U1 snRNP tireraient leurs effets sur l'IPA en partie via une augmentation du taux d'élongation (Mimoso et Adelman 2023; Krajewska et al. 2019). Dans une étude sur cellules de sarcome d'Ewing dû à l'oncogène de fusion EWS-FL1 (un facteur de transcription), ce dernier a été montré comme régulant l'expression d'une isoforme IPA de la cycline D1 via une modulation de l'élongation de la transcription (G. Sanchez et al. 2008). D'autres facteurs tels que SCAF4 et SCAF8 ont été montrés comme interagissant avec la queue CTD bi-phosphorylée en ser 2 et ser 5 de l'ARN pol II arrêtée et l'ARN naissant. Cette liaison à l'ARN naissant se fait préférentiellement près des jonctions exons/introns dans les introns. Ce site de liaison corrèle à la localisation générale des sites polyA dans les introns. La liaison de SCAF4 et 8 empêchent les événements de clivage et polyadénylation précoces (Gregersen et al. 2019). De façon inattendue, dans le cadre de la réponse au choc thermique dans les cellules humaines Cugusi et al. ont montré par PRO-seq qu'un accroissement intensif du taux d'élongation augmentait les événements d'IPA, allant à l'encontre du « *kinetic coupling model* » énoncé ci-dessus. Plusieurs explications sont proposées par les auteurs. En premier, une augmentation du mouvement brownien des complexes pourrait favoriser préférentiellement l'IPA. En second, une trop grande vitesse d'élongation ne laisserait pas à U1 snRNP le temps de se lier aux sites 5' d'épissage moins conservés pour protéger les introns ce qui favoriserait l'IPA. De telles hypothèses semblent

cependant très complexes à démontrer dans les faits car il s'agirait de pouvoir prouver pour au moins un groupe de gènes *in vivo* que les facteurs de clivage/polyadénylation en soi interagissent mieux avec l'ARN que U1 snRNP dans un contexte d'ARN polymérase II accélérée.



**Figure 7 : Kinetic coupling model**

Modèle de couplage cinétique. L'ARN pol II peut voir son taux d'élongation varier par interactions avec des facteurs pouvant l'activer tels que U1 snRNP, des facteurs d'élongations tels que SCAF4 et SCAF8, de même que certaines kinases comme CDK12 pouvant la phosphoryler sur son domaine CTD. Cette activation du taux d'élongation favorise des mécanismes tels que l'épissage et réprime l'IPA par un accès plus rapide à la machinerie du spliceosome aux sites 3' d'épissage. En revanche, un ralentissement du taux d'élongation de l'ARN pol II favorise l'induction d'IPA. Les pauses de l'ARN pol II provoquées par des îlots CpG favorisent l'IPA. En grisé, les facteurs désavantagés ; 5'SS : site 5' d'épissage ; 3'SS : site 3' d'épissage ; PAS : signal polyA ; -> : active ; -| : inhibe <-> : interaction/liaison

## II.2.6 Marques épigénétiques et épitranscriptomiques

Très peu d'études décrivent à ce jour des cas d'IPA régulés par des marques épigénétiques ou épitranscriptomiques. Dans des lignées murines ESC, par déplétion siARN de l'exosome Exosc3, une équipe a montré par ChIP-seq une colocalisation entre les nucléosomes stables, les sites d'IPA proche du TSS et les sites de pauses de l'ARN pol II (Chiu et al. 2018). Cette colocalisation entre ces 3 événements est aussi montrée comme supprimée par U1 snRNP. En inhibant PRMT1, une méthyltransférase, par siARN dans des cellules d'adénocarcinome ductale pancréatique, une équipe a montré une augmentation des événements d'APA dont une partie correspondrait à de l'IPA (Giuliani et al. 2021). Une étude citée précédemment a trouvé un lien entre boucle chromatinienne, méthylation de l'ADN et régulation polyadénylation alternative avec un exemple mentionné d'IPA régulée dans le gène *HEATR2* (Nanavaty et al. 2020).



Dans des cellules de lymphocytes et de cancer du poumon, une étude a montré par adaptation d'un protocole de *cross-link immunoprecipitation* par les UV que des milliers de marques m<sup>6</sup>A étaient déposées au niveau des derniers exons de transcrits. Lorsqu'ils réduisent le niveau de méthylation au niveau des sites polyA proximaux par déplétion de méthylases, l'utilisation de ces sites augmente pour certains. Les auteurs ne se focalisent cependant pas sur ce type de régulation pour les IPA (Ke et al. 2015). Une étude *in silico* analysant un RNA-seq de cellules Hela déficientes en YTHDC1, un des seuls facteurs lecteurs de marque m<sup>6</sup>A dans le noyau (Xiao et al. 2016), a pu établir une corrélation entre cette répression et la régulation d'événements d'IPA (Z. Zhao, Xu, Wei, Wang, et al. 2021). Ces résultats restent cependant hypothétiques, car l'équipe développait un algorithme permettant d'utiliser des données de RNA-seq pour trouver des événements *de novo* d'IPA.

## II.2.7 Exosome et dégradation des ARN

Un grand nombre d'isoformes IPA sont détectées et dégradées dans le noyau par l'exosome nucléaire. En effet, une étude déplaçant des facteurs clés de l'exosome par siARN (Exosc3) montre une augmentation des transcrits IPA (Chiu et al. 2018). La déplétion de Mtr4 appartenant au complexe NEXT associé à l'exosome mais aussi à un complexe unique Mtr4/ZFC3H1 stabilise aussi des ARN terminés prématurément souvent par IPA (Ogami et al. 2017).

## II.3 Régulation cellulaire des isoformes de polyadénylation intronique

Une régulation à grande échelle et un rôle des isoformes IPA (ou ALE) ont été décrits à travers différents tissus, stress, processus de différenciation ou de développement, d'activation cellulaire ou encore de maladies.

### II.3.1 Régulation développementale et physiologique

L'IPA est un processus se produisant dans une grande diversité de modèles comme l'Homme, la souris, le rat et le poulet (R. Wang et al. 2018). Comme mentionné précédemment, le premier exemple d'IPA caractérisé concernait le gène de l'immunoglobuline M dans la différenciation des lymphocytes (F. W. Alt et al. 1980) puis étendu à l'échelle du génome et dans d'autres processus de différenciation et de développement des cellules immunitaires (Singh et al. 2018). L'IPA se retrouve dans l'activation neuronale, la spermatogénèse (H. Wang et al. 2006). Pour citer quelques exemples, une étude a en effet

montré dans des lignées murines qu'environ 200 gènes présentés un changement d'expression de l'isoforme IPA par rapport à l'isoforme longue au cours du développement embryonnaire (11 jours vs 14 jours), de la différenciation des cellules 3T3-L1 et C2C12 (Hoque et al. 2013). Dans des lignées de kératinocytes non différenciées, une équipe a montré qu'une IPA dans GRHL3, régulée notamment par CPSF, diminuée l'expression de son gène hôte, un activateur de la différenciation (X. Chen et al. 2021). Un cas d'IPA dans le gène *sgp130* dans les cellules sanguines mononucléaires périphériques humaines a été décrit comme produisant une protéine tronquée jouant probablement un rôle dans la signalisation trans-cellulaire (Sommer et al. 2014).

### II.3.2 Stress et maladies

Divers stress régulent les isoformes IPA comme la réponse au choc thermique déjà mentionnée (Cugusi et al. 2022). L'exposition à certaines molécules toxiques à un certain seuil peut aussi moduler l'IPA.

L'anisomycine est un inhibiteur de l'activité peptidyl transférase des ribosomes provoquant un stress ribotoxique. Une étude, en l'utilisant dans les HEK, a observé une augmentation de la taille des transcrits corrélée avec une diminution de l'utilisation de l'IPA et site polyA proximaux et une augmentation de l'utilisation des sites polyA distaux (Hollerer et al. 2016). Le mécanisme n'a cependant pu être résolu.

Des stress induits par des agents occasionnant des dommages directs à l'ADN, appelés agents génotoxiques, ont aussi été documentés comme générant de l'IPA. Je détaillerai cependant ces agents dans la 4ème partie de cette introduction.

De nombreux événements d'IPA sont associés à des maladies (X. Ma et al. 2023), certains sont décrits comme corrélés ou conséquences de la maladie tandis que d'autres peuvent être du moins en partie cause de celle-ci. Lee et al ont par exemple découvert que des isoformes IPA sont spécifiquement exprimées dans des lignées de leucémie chronique lymphocytaire (CLL). Elles sont enrichies dans des gènes suppresseurs de tumeurs et codent des protéines tronquées dont certaines possèdent des propriétés oncogéniques (S.-H. Lee et al. 2018). La même équipe dans des tissus de myélomes multiples a également noté une tendance inverse avec une perte des événements d'IPA dans certains groupes de ces myélomes en comparant avec des tissus sains (Singh et al. 2018). Une étude a aussi montré qu'environ 30% des sites polyA proximaux régulés dans les tumeurs du poumon sont localisés dans les introns (Zingone et al. 2021). En utilisant des bases de données de RNA-seq sur 6 types

de tumeurs différents par rapport à des cellules saines, une étude a montré que plus de 490 événements d'IPA étaient régulés et récurrents dans ces tumeurs (Z. Zhao, Xu, Wei, Wang, et al. 2021). Certains de ces événements sont localisés dans des gènes connus pour être impliqués dans le cancer comme *TSC1*, *SPERD2* ou *CCND2*. En comparant 4998 cancers de patients en RNA-seq et exome, 152 mutations somatiques déclenchant des événements d'IPA ont pu être identifiés (Z. Zhao, Xu, Wei, Huang, et al. 2021). La plupart d'entre elles se produisent là aussi dans des gènes suppresseurs de tumeurs, générant possiblement des protéines tronquées perdant en partie au moins leurs fonctions.

Souvent ces événements d'IPA, comme nous avons pu le voir, sont le fruit d'une mutation/changement d'expression de facteurs les régulant et sont corrélés aux maladies. De plus en plus de papiers tendent à montrer que ces isoformes IPA peuvent être directement impliquées dans des fonctions de la cellule voire dans des pathologies.

## II.4 Fonctions de l'IPA et des isoformes IPA

Deux fonctions distinctes sont apportées par l'IPA : une fonction due à la terminaison prémature de la transcription et une fonction due à l'action de l'isoforme IPA par elle-même. À noter que ces deux fonctions ne sont pas exclusives.

### II.4.1 Fonction de l'IPA comme événement de terminaison prémature de la transcription

Lors de l'expression d'un gène, l'utilisation de site(s) polyA intronique(s) (et/ou la terminaison précoce de la transcription) implique, a priori, une moindre production de transcrit pleine longueur (canonique). Ainsi, une hypothèse serait que l'induction d'IPA dans un gène donné diminuerait l'expression ARNm de sa forme pleine-longueur (canonique). Une variété de modulation des isoformes IPA et longues ont pu être décrites dans la littérature. Déjà en 2012 en 3'-seq, suite aux traitements U1-AMO dans les cellules Hela, Berg et al. pouvaient décrire 3 groupes de gènes : Un groupe avec augmentation de l'isoforme IPA associée à une diminution de l'isoforme longue, un groupe avec augmentation de l'isoforme IPA mais sans changement dans l'expression de l'isoforme longue, et un groupe sans augmentation de l'isoforme IPA mais avec une diminution de l'isoforme longue (Berg et al. 2012). D'autres études ont pu montrer que l'expression d'isoformes IPA participaient à la répression de l'expression de l'isoforme longue du gène correspondant. En effet, dans des kératinocytes, il a été montré que l'expression de l'IPA de *GRHL3* faisait suite à la diminution

du niveau de CPSF pendant la différenciation. Cette augmentation de l'IPA réprime ensuite fortement l'expression de l'isoforme longue codant la protéine fonctionnelle médiant la différenciation (X. Chen et al. 2021). Dans plusieurs lignées de lymphocytes B, il a été observé que quand l'isoforme IPA augmente l'isoforme longue correspondante diminue (Singh et al. 2018). Dans un contexte d'inhibition par U1-AMO dans des cellules HeLa, il a été montré que les gènes longs étaient préférentiellement touchés par des événements d'IPA (Clivage et polyadénylation) empêchant la production de l'isoforme canonique (Oh et al. 2017). Dans un contexte d'inhibition de CDK12 cette fois-ci, il a à aussi été observé qu'une induction d'IPA empêchait la génération d'isoformes canoniques dans les gènes longs (Dubbury et al. 2018).

De façon intéressante, un événement de polyadénylation dans l'intron 3 du gène *CstF-77* dans les cellules Hela a été reporté comme étant favorisé par l'expression de *CstF-77* lui-même. L'isoforme IPA possède un fragment de l'ORF *CstF-77* non détecté avec des anticorps dirigés contre la partie N-ter. Cette interaction serait donc responsable d'un rétrocontrôle négatif. Une augmentation d'expression de *CstF-77* aboutit à la transcription de l'isoforme IPA de *CstF-77* diminuant ainsi l'expression de l'isoforme longue et donc de la protéine (W. Luo et al. 2013).

Un autre type de fonction locale, proposé par Chiu et al., supposerait que les sites IPA proches du TSS sont pour certains des points de contrôle de l'élongation de la transcription lorsqu'ils sont associés avec un nucléosome stable. Il permettrait d'éviter la transcription de transcrits 5'UTR antisens (uaARN).

#### II.4.2 Fonctions des isoformes IPA

Les isoformes IPA peuvent avoir deux types de fonctions : codante ou non codante (voir Figure 8).

De nombreuses isoformes IPA sont connues pour coder des isoformes protéiques tronquées ou avec une extrémité carboxy-terminale (C-ter) alternative. L'un des tout premiers cas reportés était celui touchant le gène de l'immunoglobuline M dans les lymphocytes B. Cet événement génère une protéine tronquée ne possédant plus de domaine transmembranaire permettant ainsi l'excrétion dans le milieu extracellulaire de l'anticorps (Rogers et al. 1980). À la suite de cela, des dizaines d'études font mention de protéines tronquées issues d'IPA (H. Wang et al. 2006; Rogers et al. 1980; Benech et al. 1985; Edwalds-Gilbert et al. 1997; Vorlová et al. 2011; S.-H. Lee et al. 2018). Manquant d'une partie de leurs domaines, cette perte peut

changer leurs fonctions, localisations, interactions par rapport à la protéine complète. Elles peuvent même acquérir des propriétés oncogéniques (G. Sanchez et al. 2008; S.-H. Lee et al. 2018) ou être des dominants négatifs (Vorlová et al. 2011).

À l'échelle du génome, dans une étude récente, mon équipe a observé que la plupart des isoformes IPA détectées par 3'-seq dans le cytosol, sont aussi bien détectées que les isoformes « *last exon* » dans les fractions polysomales lourdes (Chakraborty et al. 2022), suggérant qu'elles sont traduites.

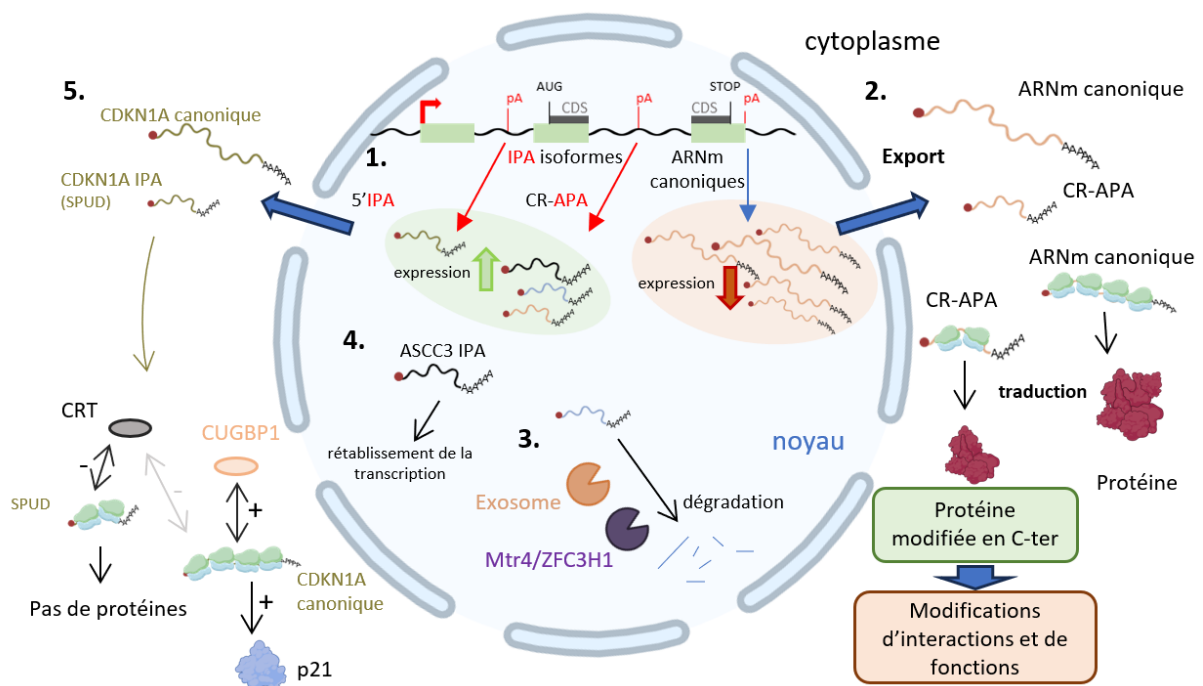
Une étude dans bioRxiv (pour l'instant non relue par les pairs) tend aussi à montrer que des isoformes IPA sont induites dans le cas de cancers puis traduites. Mais ces protéines seraient alors dégradées et les peptides issus de ce processus se retrouveraient présentés sur le complexe CMH à l'extérieur de la cellule. L'étude montre aussi une grande immunogénicité pour certaines séquences peptidiques spécifiquement exprimées par IPA dans le cadre de cancers, cibles de potentielles nouvelles immunothérapies (Ren et al. 2022).

Certaines isoformes IPA ont une fonction non codante. À la suite de l'export dans le cytoplasme des isoformes IPA induites par la déplétion de Mtr4 du complexe NEXT de l'exosome, Ogami et al ont noté que ces isoformes IPA se retrouvent par la suite engagées dans les ribosomes. Une corrélation entre cet événement et une répression globale de la traduction dans la cellule a pu être observée (Ogami et al. 2017). Une équipe a découvert que le gène *ASCC3* en cellules humaines MRC5VA dans le contexte de l'irradiation aux UV-C produisait une isoforme IPA. Bien que se terminant dans la CDS, les auteurs n'ont pas pu détecter de protéine tronquée avec un anticorps liant normalement la partie N-ter d'*ASCC3*. Cette isoforme IPA a été montrée comme ayant un rôle antagoniste avec la protéine *ASCC3* en favorisant le rétablissement de la transcription après dommages à l'ADN alors que la protéine *ASCC3*, elle, l'inhibe. La délétion du codon STOP de l'ORF ne reproduit pas le phénotype apporté par la déplétion de l'isoforme IPA (Williamson et al. 2017). Récemment, Murphy et al ont montré qu'une isoforme IPA, nommée SPUD, était induite en contexte génotoxique (irradiation UV-C) en début de gène *CDKN1A* (le gène codant la protéine p21, qui a un rôle clé dans la réponse aux dommages à l'ADN). Les auteurs montrent que SPUD sous sa forme épissée peut être exportée dans le cytoplasme et se retrouve par la suite dans les polysomes mais ils ne détectent pas de protéine traduite issue de SPUD bien qu'une ORF soit présente. En revanche, ces derniers montrent une interaction de l'ARN SPUD avec les protéines CRT et CUGBP1, respectivement inhibiteur et activateur de la traduction de l'ARNm canonique de *CDKN1A*. Ils

montrent par des tests de *RNA binding immunoprecipitation* (RIP), permettant de détecter des interactions ARN-protéines, sur des cellules HCT116 que SPUD se lie préférentiellement en contexte de dommage à l'ADN CRT alors que l'isoforme longue, elle, lie plutôt CUGBP1. Ce mécanisme permet notamment une activation de la traduction de p21 dans la réponse aux stress génotoxiques.

Enfin, une étude a identifié un ensemble d'isoformes IPA très courtes, terminant près de l'extrémité 5' du gène (5'IPA), qui sont abondantes dans certains tissus ou types cellulaires, mais qui ont un faible potentiel codant prédit d'après l'analyse de leur séquence (Singh et al. 2018). Cependant, ni le statut traductionnel ni la fonction de ces isoformes n'ont été étudiés.

De façon intéressante, depuis plusieurs années, un nombre croissant de publications montrent que des petites ORF au sein d'ARN annotés comme non codants ou au sein de régions ARN annotés comme non traduites (UTR), codent en fait des petites protéines (typiquement, moins de 100 acides aminés) appelées microprotéines. Aucune étude n'a cependant rapporté d'isoformes IPA comme porteuses d'alt-ORF ou comme génératrices de microprotéines.



**Figure 8 : Devenir et fonctions des isoformes IPA**

1. Certains gènes utilisent des sites polyA dans les introns (IPA) sous certaines conditions permettant la génération d'isoformes ARNm IPA différentes de l'ARNm canonique. Parmi ces isoformes IPA, on compte des isoformes 5'IPA terminant dans la 5'UTR ou comprenant la CDS et des isoformes CR-APA comprenant une partie de la CDS qui sont en opposition avec les ARNm canoniques en comprenant la totalité. Dans un gène, lorsqu'un transcrit termine de façon prémature, il génère une isoforme IPA ne devenant pas ainsi un ARNm canonique. Dans certains gènes l'augmentation de l'expression de

l'isoforme IPA (flèche verte) défavorise l'expression de l'ARNm canonique (flèche rouge). **2.** Les ARNm sont généralement exportés dans le cytoplasme afin d'être traduits. Certains CR-APA (en jaune) sont aussi traduits et génèrent une isoforme protéique modifiée en C-ter. Cette modification en C-ter altère les interactions et fonctions de la protéine. **3.** Les isoformes IPA peuvent être la cible de l'exosome ou du complexe Mtr4/ZFC3H1 induisant leur dégradation. **4.** Une CR-APA du gène *ASCC3* (en noir) est connue pour avoir une fonction non codante dans le noyau en permettant le rétablissement de la transcription dans un contexte d'irradiation aux UV. **5.** Une 5'IPA du gène *CDKN1A* nommée SPUD (en vert) est connue pour être notamment exportée dans le cytoplasme et engagée avec les polysomes. Cependant, aucune protéine n'est traduite de cet événement. SPUD se lie préférentiellement à CRT lorsqu'il est en compétition avec CDKN1A dans les polysomes qui lui lie CUGBP1. CRT est un inhibiteur de traduction de CDKN1A tandis que CUGBP1 lui l'active. Ainsi, SPUD favorise la traduction de CDKN1A codant p21.

### III. Petites ORF et microprotéines

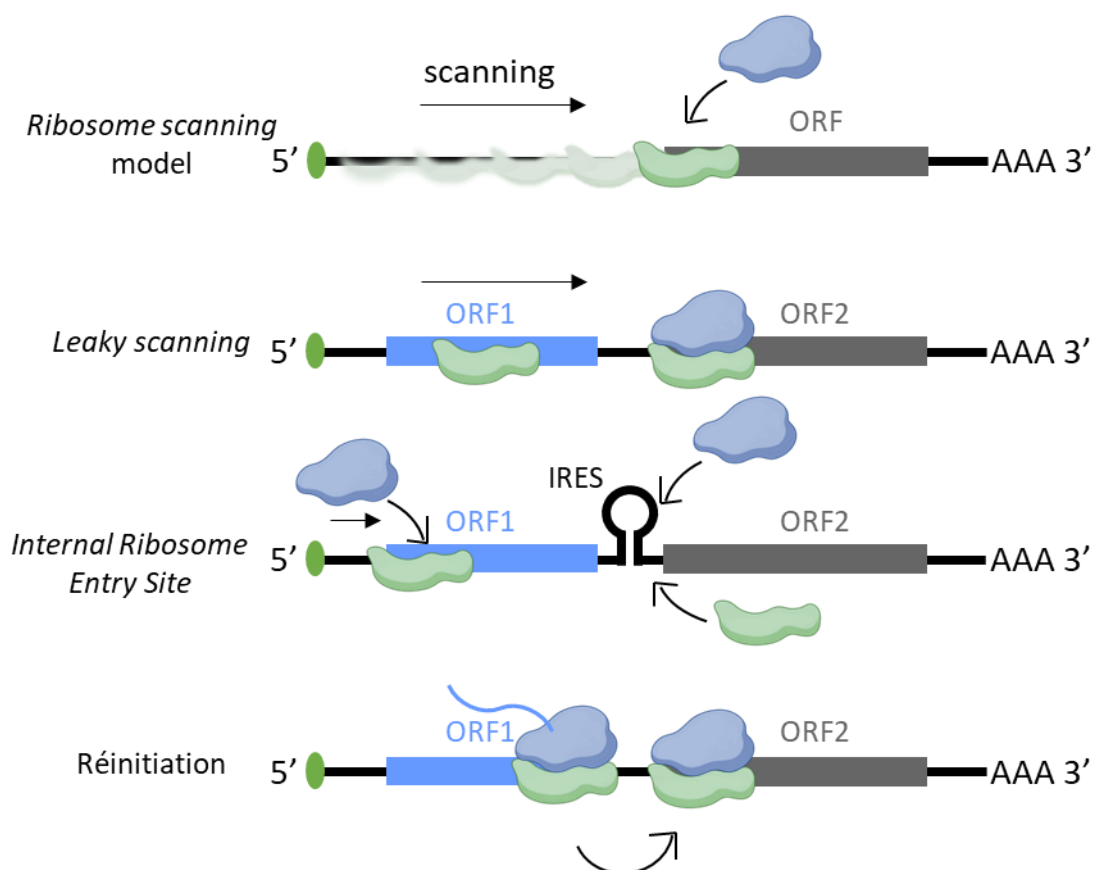
Les microprotéines chez l'Homme se définissent comme des protéines de moins de 100 aa issues de petites ORF (sORF) ayant été exclues des premières annotations de séquences codantes du génome notamment dans le cadre du *Human Genome Project*.

Lancé en 1988 et finalisé en 2003, le *Human Genome Project* avait pour but de séquencer entièrement le génome humain et de définir les séquences codantes du génome. L'identification de ces séquences codantes était basée sur les connaissances et les moyens analytiques de l'époque. De ce fait, ces connaissances et moyens ont notamment orienté l'annotation du génome codant avec 3 conjectures principales pensées comme inhérentes aux eucaryotes (Brunet et al. 2020).

Premièrement, les ARNm ont été supposés comme ne portant qu'une seule séquence codante, plus précisément ils sont dits monocistroniques. Cette idée repose sur le modèle du *ribosome scanning* validé par de nombreuses études (Pestova et al. 1999). Il propose notamment pour la traduction coiffe-dépendante que la sous-unité 40S du ribosome scanne les ARNm en partant de la coiffe en 5' et s'arrête au premier codon d'initiation AUG possédant une séquence de Kozak optimale (la plus conservée)(Mouilleron et al. 2016). Il s'agit de l'étape de reconnaissance du site d'initiation de la traduction (TIS). Cette identification permet ensuite la formation, par recrutement de la sous unité 60S, du ribosome 80S et du recrutement des facteurs d'initiation de la traduction. Après initiation puis élongation, la traduction s'arrête au codon stop en phase et le ribosome se décroche de l'ARNm. Nous savons maintenant que des mécanismes tels que le *leaky scanning*, la réinitiation et les IRES permettent la lecture d'au moins deux ORF sur un même ARNm (Orr et al. 2020; Mouilleron, Delcourt, et Roucou 2016; Figure 9). Le *leaky scanning* correspond à une absence de reconnaissance systématique du TIS lors du *ribosome scanning*. Cette absence de reconnaissance a été montrée comme notamment favorisée lorsque la séquence de Kozak entourant le TIS n'est pas conservée (Mouilleron et al. 2016). Il est notamment connu que seulement 37,4% des ARNm de l'Homme ont une séquence Kozak optimale et que cette dernière n'est pas suffisante pour arrêter tous les ribosomes (Mouilleron et al. 2016). La réinitiation, quant à elle, permet à un ribosome ayant terminé la traduction d'une ORF d'être réengagé en aval du même ARNm (Orr et al. 2020; Brunet et al. 2020). D'abord montré principalement chez les ARNm de virus, l'existence d'*internal ribosome entry site* (IRES) sur des ARNm humains a par la suite été observé bien



qu'il n'y ait pour le moment de consensus établi (Ray et al. 2006; Moulleron et al. 2016). Les IRES sont des structures secondaires d'ARNm permettant le recrutement de la machinerie de traduction de façon indépendante de la coiffe. De ce fait, ce mécanisme outrepassa le scanning effectué par la sous-unité 40S du ribosome (Moulleron et al. 2016). Par ces trois mécanismes, plusieurs ORF distinctes sur un même ARNm peuvent être traduites.



**Figure 9 : Modèle du ribosome scanning et cas particuliers**

*Ribosome scanning model* : La sous-unité 40S (protéine verte) scanne un ARNm coiffé jusqu'à rencontrer un codon d'initiation avec une séquence Kozak. La sous-unité 60s (protéine bleue) est alors recrutée et la traduction débute. *Leaky scanning* : Le scanning de la sous-unité 40S ne s'arrête pas systématiquement au premier codon d'initiation avec une séquence Kozak. Elle reconnaît alors une deuxième ORF située en aval. *Internal Ribosome entry site* : cette structure d'ARN permet le recrutement des sous-unités 40 et 60S du ribosome et permet la traduction d'une ORF de façon coiffe indépendante. *Réinitiation* : À la fin de la traduction d'une ORF sur un ARNm la machinerie de traduction peut être réengagée sur une seconde ORF située en aval de la première.

Deuxièmement, pour des raisons de puissance statistique, un seuil arbitraire de 300 nucléotides (ou 100 acides aminés) avait été fixé pour définir une ORF comme statistiquement significative par rapport à un codon start AUG en phase de lecture de façon aléatoire avec un des trois codons stop possibles (Trotta 2016). Or, il était connu que des protéines telles que MYC pouvaient commencer par un CUG, un des codons d'initiations non canoniques (Wright et al. 2022). Il est en effet rapporté que plusieurs codons d'initiations alternatifs moins

efficaces que AUG existent dans de nombreux modèles tels que CUG (chez l'Homme et la bactérie), GUG (chez l'Homme, la souris et la bactérie), TUG, et plus rarement ACG (chez la levure), UUG (chez la bactérie et la levure), AUU (chez l'Homme, la levure la bactérie) et AUC (chez la bactérie) (Kearse et al. 2017; Cao et al. 2020). Entre autres, la longueur minimale de 100 acides aminés s'est retrouvée comme limitante, car des exemples connus de protéines fonctionnelles comme RPL41 de 25 aa avaient été découvertes là aussi (Wright et al. 2022; Leong et al. 2022).

Troisièmement, les pseudogènes détectés ont été annotés de façon systématique comme non codants. Or, on sait aussi que 10% des pseudogènes seraient altérés dans le cancer avec des changements d'expression ou des mutations ponctuelles (Brunet et al. 2020). Cela suggère d'éventuelles fonctions dans ces pathologies.

À cause de ces 3 hypothèses, comme dit précédemment beaucoup d'ORF ont été exclues des bases de données *Refseq* car ne respectant pas notamment la 1<sup>er</sup> et la 2<sup>ème</sup> hypothèse. Toutes ces ORF se nomment ORF alternatives (alt-ORF), et les alt-ORF exclues par la 2<sup>ème</sup> hypothèse de moins de 300 nts peuvent être nommées *small* ORF (sORF) (Orr et al. 2020). Si une protéine de moins de 100 aa est générée par une alt-ORF ou une sORF, on parle alors de microprotéines.

La première étude faisant mention de microprotéine décrivait en réalité une protéine de 120 acides aminés catégorisée maintenant comme neuropeptide (Grubb et al. 1982). Les études font maintenant la distinction entre les neuropeptides et peptides hormonaux clivés par réaction protéolytique (Makarewich 2020; Bhati et al. 2018) donnant des fragments peptidiques plus petits et les microprotéines issues de la traduction d'une ORF sans modification post-traductionnelle. Bien que de tailles comparables, la biologie de ces deux catégories de peptides est bien différente. Nous ferons donc l'impasse sur les neuropeptides et les peptides hormonaux car n'ayant pas été exclus des annotations *Refseq*.

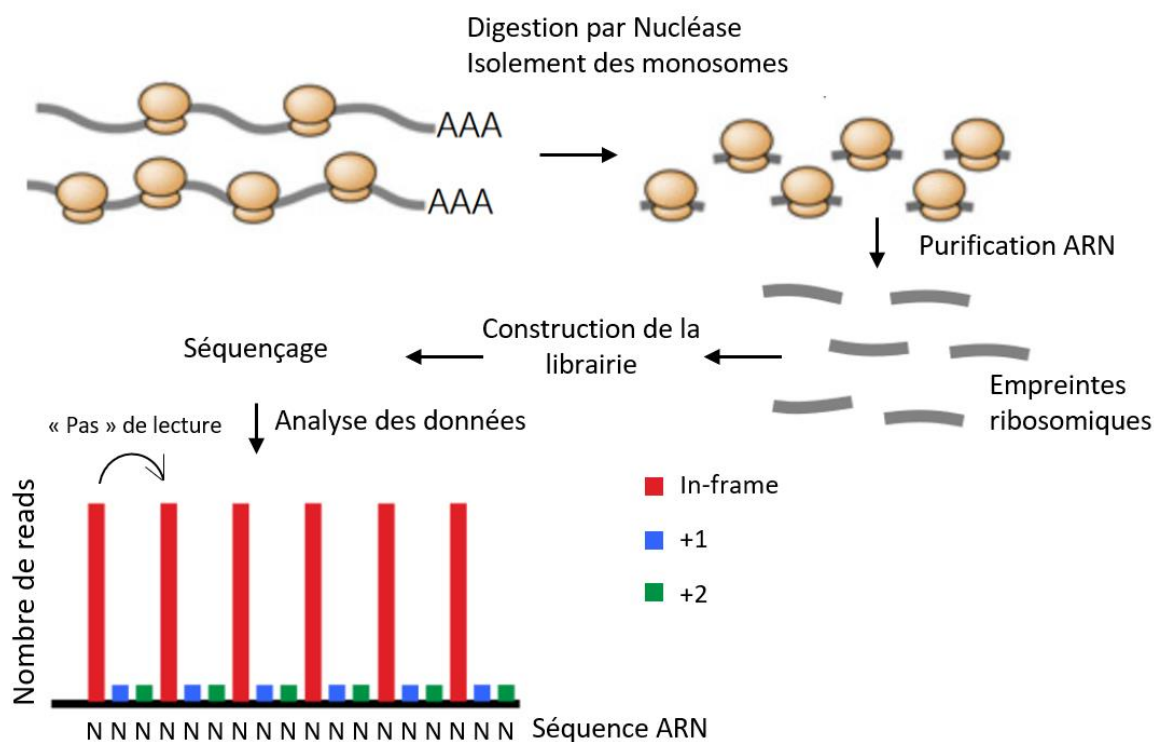
Trois approches indispensables ont permis d'énormes progrès notamment ces 20 dernières années dans l'identification d'une multitude d'alt-ORF codant des microprotéines.

### III.1 Méthodes de détection

Ces trois approches sont le séquençage sur ribosome (*ribosome profiling* ou Ribo-seq), la spectrométrie de masse (MS) et l'analyse bioinformatique (Brunet et al. 2020).

### III.1.1 Le Ribo-seq

Mis en place par Ingolia et al. en 2009, le Ribo-seq étudie les séquences ARNm interagissant avec les ribosomes et supposément donc traduites (voir Figure 10). Ce séquençage permet de connaître avec une précision nucléotidique les séquences ARNm qui sont lues. Pour se faire, la technique consiste à fixer de façon covalente les ribosomes aux ARNm avec lesquels ils interagissent et à digérer les nucléotides non protégés par les interactions avec les ribosomes fixés. On séquence ensuite les régions ARNm non dégradées. Pour un ARNm donné, une population de fragments de 30 nts correspondant à l’empreinte ribosomique va pouvoir être lue et alignée sur les transcrits ou le génome. Le ribosome lisant des codons de 3 nucléotides à la suite, ce séquençage permet de savoir quel codon était en cours de lecture. En compilant l’ensemble des codons en cours de lecture, on peut déterminer quelle ORF était lue avec son « pas » de lecture associé.



**Figure 10 : Protocole du ribo-seq (adapté de Hsu et al. 2016)**

Les ribosomes sont fixés par réaction chimique et les ARNm sont digérés par des nucléases. Après purification des fragments d’ARN non digérés correspondant aux empreintes des ribosomes, ils sont séquencés. Les fragments ont une taille d’environ 30 nts correspondant à la taille du ribosome. Il est possible de savoir quel codon était en cours de lecture par le ribosome par rapport à la séquence d’ARN protégée. Une analyse bioinformatique permet ensuite un alignement au génome ou transcriptome de la population de fragments d’ARN. En comptabilisant le nombre de *read* pour chaque codon, on peut calculer la fiabilité du « pas » du ribosome traduisant l’ARNm.

Cette technique montre rapidement des résultats étonnants. Plus de 40% des événements d’initiation de la traduction sont localisés dans des ORF non annotées (Brunet et

al. 2020). Le Ribo-seq s'est amélioré avec le temps devenant de plus en plus fiable et sensible (Hsu et al. 2016). La multiplication des études l'utilisant a permis une augmentation du nombre d'alt-ORF documentées.

### III.1.2 Spectrométrie de masse

Globalement, cette technique permet d'identifier par leurs masses et leurs charges des molécules ionisées, projetées à travers un champ magnétique. Elle est notamment utilisée en biologie pour obtenir la séquence en acides aminés d'extraits protéiques. Cette technique a obtenu de nombreuses améliorations et variantes au cours des années permettant de regarder des molécules possédant des groupements, interactions ou partenaires spécifiques. Dans le cas de l'étude des protéines, la MS a le plus souvent recours à des enzymes protéolytiques afin d'analyser chaque fragment issu de la digestion. Pour qu'un fragment soit reconnu par le capteur, il faut qu'il soit en concentration suffisante à travers plusieurs réplicats. Une des difficultés majeures pour l'étude des microprotéines en spectrométrie de masse était que ces dernières de par leur petite taille peuvent être dépourvues de sites de digestion protéolytique. Elles peuvent aussi être perdues là aussi à cause de leur petite taille pendant les différentes étapes de préparation et de purification de l'échantillon avant analyse sur le spectromètre. Récemment, des innovations ont permis d'accroître la sensibilité des spectromètres de masse et de diminuer ce problème de purification (Hassel et al. 2023).

Plus récemment, une nouvelle méthode appelée immunopeptidome a été créée. Elle permet l'analyse en MS de fragments de protéines présentés par le complexe CMH1 en tant qu'antigènes. Cette méthode permet d'identifier des microprotéines plus enclins à être dégradées (Chong et al. 2020).

### III.1.3 Analyse Bioinformatique

De nombreux logiciels ont vu le jour afin de pouvoir distinguer les ORF d'intérêt du simple bruit de fond. Ces derniers tentent d'établir des liens entre niveau de conservation des séquences, composition intrinsèque et pertinence traductionnelle (Tableau 1). D'autres essaient de prévoir la structure, les interactions et les fonctions des microprotéines. Récemment, de nouveaux logiciels utilisant la technologie des réseaux neuronaux ont vu le jour afin de prédire la traductibilité d'une ORF sans forcément essayer de comprendre derrière la logique de cette traductibilité.

**Tableau 1 : Listes des outils bioinformatiques de prédiction d'alt-ORF/sORF (adapté de Leong et al. 2022)**

Prediction tool	References	Website	Description
Coding Non-Coding Identifying Tool (CNIT)	[126]	<a href="http://cnit.noncode.org/CNIT/">http://cnit.noncode.org/CNIT/</a>	Distinguishes between coding and non-coding regions based on intrinsic sequence compositions
Coding Region Identification Tool Invoking Comparative Analysis (CRITICA)	[127]	<a href="http://rdpwww.life.uiuc.edu/">http://rdpwww.life.uiuc.edu/</a>	Analyses nucleotide sequence composition and conservation at the amino acid level
Coding Potential Calculator (CPC)/CPC2	[128, 129]	<a href="http://cpc.cbi.pku.edu.cn">http://cpc.cbi.pku.edu.cn</a> <a href="http://cpc2.gao-lab.org/">http://cpc2.gao-lab.org/</a>	Assess protein-coding potential based on important features (ORF size, coverage, integrity); CPC2 improves run speed and accuracy
Coding Potential Predictor (CPred)	[130]	<a href="http://www.rnabinding.com/CPred/">http://www.rnabinding.com/CPred/</a>	Predicts the coding potential of RNA transcript
CPred-sORF	[131]	<a href="http://www.rnabinding.com/CPred-sORF/">http://www.rnabinding.com/CPred-sORF/</a>	Addition of 2 new features from CCPred i.e., GCcount, mRNN-11 codons and CUG, GUG start codons
MicroPeptide Tool (MiPepid)	[21]	<a href="https://github.com/MindAI/MiPepid">https://github.com/MindAI/MiPepid</a>	Identifies coding sORFs based on existing microproteins subpopulation set
sORF Finder	[52]	<a href="http://evolver.psc.riken.jp/">http://evolver.psc.riken.jp/</a>	Identifies sORF with high coding potential based on nucleotide composition bias and potential functional constraint at the amino acid level
smORFunction	[132]	<a href="https://www.cuilab.cn/smorfunction/home">https://www.cuilab.cn/smorfunction/home</a>	Provides function prediction of sORFs/microproteins
miPFinder	[133]	<a href="https://github.com/DaStraub/miPFinder">https://github.com/DaStraub/miPFinder</a>	Identifies and evaluates microproteins functionality using information on size, domain, protein interactions and evolutionary origin
PhastCons	[134]	<a href="http://compugen.cshl.edu/phast/">http://compugen.cshl.edu/phast/</a>	Based on conservation scoring and identification of conserved elements
PhyloCSF	[135]	<a href="http://compbio.mit.edu/PhyloCSF">http://compbio.mit.edu/PhyloCSF</a>	Determines a conserved protein-coding region based on formal statistical comparison of phylogenetic codon models
uPEPperoni	[136]	<a href="http://u pep-scmb.biosci.uq.edu.au/">http://u pep-scmb.biosci.uq.edu.au/</a>	Specifically for 5'UTR sORFs, based on conservation
AnABLAST	[34]	<a href="http://www.bioinfocabd.upo.es/ab/">http://www.bioinfocabd.upo.es/ab/</a>	Identifies putative protein-coding regions in DNA regardless of ORF length and reading frame shifts
Small Peptide Alignment Discovery Application (SPADA)	[137]	<a href="https://github.com/orionzhou/SPADA">https://github.com/orionzhou/SPADA</a>	Homology-based gene prediction programme
Deep Neural Network for coding potential prediction (DeepCPP)	[138]	<a href="https://github.com/yuuuzhang/DeepCPP">https://github.com/yuuuzhang/DeepCPP</a>	Effective on RNA coding potential prediction, specifically sORF mRNA prediction

### III.1.4 Bases de données

De nombreuses bases de données ont vu le jour (Tableau 2), compilant les Ribo-seq, analyses protéomiques et apportant de nouveaux outils bioinformatiques. Certaines sont spécialisées dans une seule méthode, tandis que d'autres et de plus en plus combinent au moins deux approches *in vivo* à travers divers modèles biologiques. Certaines sont même spécialisées dans des pathologies ou immunologie (X. Luo et al. 2022), tandis que d'autres ont pu être reliées à des fonctions par crible CRISPR *in vivo* (J. Chen et al. 2020).

**Tableau 2 : Listes des bases de données de alt-ORF (adapté de Leong et al. 2022)**

Database	References	Website	Type	Description
sORFs.org	[14]	<a href="http://www.sorfs.org">http://www.sorfs.org</a>	sORF repository	Obtains experimental data from RIBO-seq with conservation analyses and rescanning MS data from PRIDE for updated small peptide validation
SmProt	[109]	<a href="http://bioinfo.ibp.ac.cn/SmProt/">http://bioinfo.ibp.ac.cn/SmProt/</a>	sORF repository	Database on small proteins specifically from lncRNA, obtains data from RIBO-seq, literature mining and MS data, integrates conservation analyses
OpenProt	[111, 112]	<a href="https://www.openprot.org/">https://www.openprot.org/</a>	altORF resource	Contains information on protein isoforms and altORFs with experimental evidence, integrates RIBO-seq, MS, conservation analyses and functional domains
ARA-PEPs	[108]	<a href="http://www.biw.kuleuven.be/CSB/ARA-PEPs">http://www.biw.kuleuven.be/CSB/ARA-PEPs</a>	sORF repository	Repository of putative sORF-encoded peptides specifically in <i>Arabidopsis thaliana</i> , data obtained from in-house Tiling arrays and RNA-seq data
PsORF	[107]	<a href="http://psorf.whu.edu.cn/">http://psorf.whu.edu.cn/</a>	sORF repository	Database of sORF across different plant species, incorporating genomic, transcriptomic, RIBO-Seq and MS data
MetamORF	[110]	<a href="http://metamorf.hb.univ-amu.fr/">http://metamorf.hb.univ-amu.fr/</a>	sORF repository	A repository of unique sORFs in <i>H. sapiens</i> and <i>M. musculus</i> genomes by experimental and computational methods
nORFs.org	[113]	<a href="https://norfs.org/">https://norfs.org/</a>	novel ORF (nORF) repository	Provides aggregated information from databases such as sORFs.org, OpenProt and OpenCB

### III.1.5 Pertinence des ORF

Une des raisons pour lesquelles les études recourent les différentes méthodes pour générer ces bases de données est la fiabilité modérée des alt-ORF potentielles trouvées par ces méthodes. Ainsi, sORFs.org et OpenProt recensent respectivement plus de deux millions et plus de 400000 alt-ORF par analyse de Ribo-Seq, mais seulement quelques dizaines de milliers de microprotéines sont détectées en MS (respectivement 89 459 et 30 096) (Brunet et al. 2020). Il faut alors trouver des critères pour pouvoir être plus restrictif afin de pouvoir éliminer les faux positifs. La nouvelle difficulté n'est donc plus d'identifier de nouvelles ORF comme cela a été le cas durant les deux dernières décennies, mais bien de distinguer les ORF fonctionnelles.

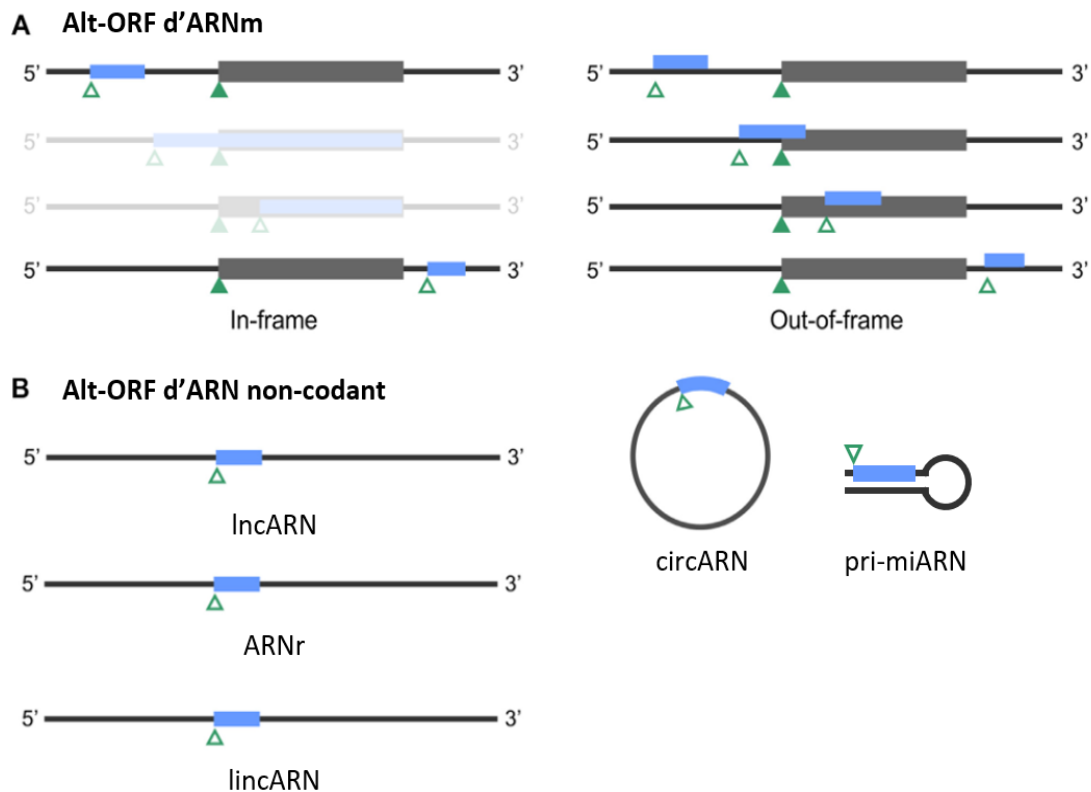
Grâce à ces bases de données, il est possible de dégager des caractéristiques communes aux alt-ORF ainsi que de déterminer les ARN et régions d'ARN dont elles proviennent.

### III.2 Localisation des alt-ORF dans les ARN

Les bases de données obtenues par Ribo-seq, MS et analyses bioinformatiques ont permis de mettre en évidence deux types d'ORF non canoniques, mentionnées précédemment : il s'agit des alt-ORF et de son sous-groupe associé les sORF. Les alt-ORF désignent l'ensemble des ORF présentes sur un ARNm avec une CDS déjà annotée ou sur un

ARN annoté comme non codant (voir Figure 11). Ces alt-ORF d'ARNm sont nommées en fonction de leurs positions relatives à la séquence canonique, *main* ORF en anglais que nous nommerons ici ORF canonique. Il y a donc le groupe des alt-ORF en 5'UTR aussi nommées *upstream* ORF ou uORF, les alt-ORF chevauchantes à la CDS et décalées de son cadre de lecture, dit alt-ORF *out-of-frame*, les alt-ORF prolongeant ou raccourcissant la CDS dans le cadre de lecture ou alt-ORF *in-frame* et les alt-ORF en 3'UTR (Figure 11A). Ce dernier groupe est celui le moins enrichi biologiquement représentant 2,8% des alt-ORF identifiées sur OpenProt et SmProt (Wright et al. 2022). Il a cependant été montré que ce faible nombre d'alt-ORF identifiées dans les 3'UTR pouvait être due à une faible efficacité inhérente au Ribo-seq vis-à-vis de cette région de l'ARNm (Miettinen et al. 2015). Dans cette thèse, nous ne détaillerons pas les régulations apportées par les alt-ORF sur l'ARNm, mais nous utiliserons certaines microprotéines issues de ces alt-ORF comme exemples fonctionnels. Ces alt-ORF sont cependant notamment connues pour pouvoir réguler la traduction et la stabilité/temps de demi-vie de leur ARNm de par leur action direct en Cis (Wright et al. 2022). Nous excluons de nos analyses et raisonnements les alt-ORF *in frame* prolongeant ou raccourcissant la CDS d'une protéine déjà connue grâce à l'utilisation de codon d'initiation non canonique mentionné ci-dessus. Ces alt-ORF diffèrent en effet, des autres alt-ORF de par la taille des protéines qu'elles peuvent produire, ajoutant ou supprimant un domaine en N-ter supplémentaire à une protéine canonique de fonctions et de régulations connues.

D'autres alt-ORF présentes sur des ARN non codants n'avaient pas été trouvées, car d'une taille inférieure à 300 nts. Mentionnées précédemment, on les appelle *small* ORF (sORF ou smORF) et elles codent des microprotéines car d'une taille inférieure à 100 acides aminés (certains auteurs sont moins stricts et disent environ 100 ou moins) (Orr et al. 2020). Les protéines de sORF ont une taille moyenne de 54-57 aa ou 29 aa dans les régions intergéniques (Wright et al. 2022; Cardon et al. 2019; Neville et al. 2021). Les alt-ORF d'ARNm comprises dans les séquences UTR ou chevauchant l'ORF canonique codent des protéines de 39 aa et 27 aa en moyenne respectivement (Neville et al. 2021). Pour comparaison, les protéines Refseq ont une taille moyenne de 344 aa (Cardon et al. 2019) ou 557 aa sur UniProtKB/Swiss-Prot (Neville et al. 2021).



**Figure 11 : Origines des Alt-ORF chez l’Homme (modifiée de Orr et al. 2020)**

A : Les alt-ORF sont localisées sur les ARNm au niveau des UTR (5’ et 3’), chevauchantes (5’ et 3’) ou en décalage de phase (en interne de l’ORF). B : Les alt-ORF ont été reportées comme localisées dans un ensemble d’ARN non codant comme les lncARN, lincARN, ARNr, pri-miARN et circARN. Gris : ORF canonique (CDS annotée) ; Bleu : ORF non canonique ; triangle plein : AUG annoté de l’ORF canonique ; Triangle creux : codon d’initiation de l’ORF non canonique ; Blanchies : les alt-ORF de grande taille

Les alt-ORF d’ARN non codants ont été montrées comme pouvant être localisées dans des lincARN, lncARN (S. Zhang et al. 2020). Il a même été montré chez la Drosophile, puis chez l’Homme que des pri-miARN, ARN précurseurs des miARN, pouvaient contenir des alt-ORF traduites (Montigny et al. 2021; Ormancey et al. 2023). Des alt-ORF ont pu aussi être découvertes dans des ARN circulaires, transcrits alternatifs issus de *back-splicing* (Zheng et al. 2019; M. Zhang et al. 2018; Legnini et al. 2017). Le *back-splicing* est un cas particulier d’épissage liant l’extrémité 3’ d’un exon à l’extrémité 5’ d’un exon situé en amont dans l’ARN primaire. L’ARN circulaire monocaténaire créé peut comporter un ou plusieurs exons du gène dont il est issu ainsi que des introns par rétention d’introns (Ebbesen et al. 2017). Par ailleurs, le génome des mitochondries a été montré comme porteur d’alt-ORF. Le gène de l’ARNr 12S s’est révélé coder une microprotéine de 16 aa appelée MOTS-C (C. Lee et al. 2015). Aucune étude ne rapporte d’isoformes IPA comme vecteurs d’alt-ORF. L’idée est mentionnée de façon hypothétique par Singh et al. lors de l’analyse de ses isoformes IPA courtes, 5’IPA, et dans une revue (Kamieniarz-Gdula et al. 2019) mais aucune analyse de traduction ou de fonction n’a été réalisée. On pourrait en effet anticiper que les sections d’introns exonisées par IPA, séquences



d'ARN uniques et absentes totalement de l'ARNm canonique, puissent être des sources potentielles d'alt-ORF. De même, des uORF pourraient se retrouver isolées de leur ORF canonique associée sur l'ARNm en raison de sites polyA entre, leur permettant d'être traduites.

Avec toutes ces données sur les alt-ORF ainsi que leur localisation, des études se sont intéressées à l'évolution et à la conservation de ces dernières.

### III.3 Evolution et conservation

En s'appuyant sur une étude appliquant des filtres stricts et découvrant 7264 alt-ORF jugées de haute fiabilité (Mudge et al. 2022), une équipe a montré d'importants résultats concernant l'histoire évolutive de ces alt-ORF en regardant notamment leur séquence en acides aminés, leurs structures prédites et leurs interactions (Sandmann et al. 2023). En effet, l'article montre que plus de 90% des alt-ORF de microprotéines humaines ont émergé récemment dont 4101 de régions non codantes. On parle alors de microprotéines *de novo*. Au niveau de la structure des microprotéines, certaines de ces émergences sont très récentes, 1370 structures ont émergé à la séparation Homme-macaque et 222 sont spécifiques à l'Homme. Les auteurs notent aussi un phénomène de pression de sélection pour conserver l'ATG de l'ORF mais pas forcément la structure de la microprotéine. Grâce à la méthode PRISMA, accrochant à une matrice des peptides d'intérêt de 14 acides aminés qu'on incube avec des extraits protéiques avant d'effectuer une MS sur les interactants, ils trouvent plus de 266 interactions avec leurs 45 microprotéines dont 30 ont été détectées en endogène par MS. L'étude complète par ailleurs la précédente étude de Mudge en allant au-delà de la limite de 16 aa définie dans l'étude. Ils détectent en effet, 221 alt-ORF traduites de taille allant de 15 jusqu'à 3 aa dont 38 avec preuve de détection endogène (Sandmann et al. 2023).

Une étude *in silico*, s'appuyant sur un recoupement de bases de données de J. Chen et al. 2020 et Hon et al. 2017 d'alt-ORF fonctionnelles, pour certaines réalisées chez l'Homme, a aussi voulu comprendre l'histoire évolutive de ces ORF. Pour eux, de par leur petite taille, les auteurs supposent qu'il est beaucoup plus probable qu'une alt-ORF émerge de séquences non codantes par le hasard. En comparant avec différents génomes d'autres organismes, l'étude observe que sur les 715 alt-ORF que compte le recoupement, 155 auraient émergé *de novo* dont 2 au moment de la séparation Chimpanzé-hominidés. En comparant les alt-ORF qu'ils qualifient de jeunes avec les plus anciennes, les auteurs observent que les alt-ORF jeunes sont moins GC riches à l'exception des alt-ORF *de novo* chevauchant les ORF canoniques et

bénéficiant ainsi de la conservation du gène codant (Vakirlis et al. 2022). Il est en effet connu que les ORF de gènes canoniques sont plus GC riches (Oliver et al. 1996).

Une réponse à cet article, écrite par Bozorgmehr 2023, nuance cependant les conclusions et observations des auteurs. En effet, Bozorgmehr note que cette étude n'a pas tenu compte de l'impact des divers éléments transposables que le génome humain possède. De plus, certains transcrits porteurs d'alt-ORF considérées par l'auteur comme non codants sont en réalité des pseudogènes. Or, nous en avons déjà parlé, les pseudogènes peuvent être réactivés dans certains cas particuliers (Brunet et al. 2020). Par ailleurs, si un pseudogène était acquis au cours de l'évolution humaine, la détection d'une « alt-ORF » ne serait pas le fruit d'une émergence *de novo*. Il souligne aussi que même dans une étude comme celle de J. Chen des faux positifs existent dans les expérimentations de crible. D'autre part, des auteurs s'interrogeaient si nous avons à disposition les bons outils analytiques pour pouvoir mesurer la conservation d'ORF si petites (Wright et al. 2022). Vakirlis et al. 2022, regardant la séquence en acides aminés et en ADN ont, en effet, au cours de leur étude *in silico*, du modifier de façon plus ou moins arbitraire les seuils de significativité et de conservation attendus pour pouvoir en extraire des alt-ORF. Wright et al. se demandent notamment si ces séquences codantes sont réellement *de novo* et s'il ne faudrait pas plutôt regarder la conservation de conformation comme testé par Sandmann et al. 2023. Il est notamment connu pour les microprotéines fonctionnellement homologues associées à SERCA que bien qu'elles aient des séquences en acides aminés différentes, ces dernières conservent la même conformation (Makarewich 2020).

Vakirlis et al. 2022 prédisent aussi que les alt-ORF *de novo* sont plus spécifiques à des tissus, plus sensibles à l'exosome et ont plus de régions avec hélices moins intrinsèquement désordonnées. Ils trouvent aussi que les alt-ORF *de novo* ont plus de résidus aromatiques, prédisent une plus grande accessibilité au solvant et avec des propriétés d'agrégation plus marquées. Toutes ces caractéristiques potentielles déduites permettent d'établir les modes d'actions possibles des alt-ORF et microprotéines.

## III.4 Mécanismes d'action des alt-ORF et microprotéines

### III.4.1 Réservoir évolutif

Les résultats de l'étude *in silico* de Vakirlis et al. sont à mettre en parallèle avec une étude récente. En effet, des auteurs ont notamment étudié les différences de stabilité des parties C-ter de protéines (Kesner et al. 2023). Ils ont greffé en C-ter d'une protéine rapportrice 12 000 séquences ADN de 90 nts différentes appartenant à divers groupes génomiques (introns, CDS, lncARN, 5'UTR et 3'UTR) et plus de 500 000 de séquences aléatoires de 39 nts à ce même système rapporteur. Ils montrent notamment que l'hydrophobicité moyenne des séquences d'acides aminés en C-ter favorise la dégradation protéosomale des protéines. La queue greffée aurait cependant une influence moindre lorsque cette dernière fait moins de 13 aa. Un lien anti-corrélatif est aussi établi entre dégradation et région intrinsèquement désordonnée et anti-corrélatant aussi avec l'hydrophobicité moyenne. Les régions intrinsèquement désordonnées seraient moins hydrophobes et donc protégées de la dégradation. Ils observent aussi que dans les protéines connues, il y a une diminution du nombre d'acides aminés hydrophobes dans les 30 derniers aa en C-ter. En comparant les groupes annotés, ils trouvent que le groupe terminal CDS et 5'UTR ont le score de dégradation le plus bas et le groupe intron le plus haut.

Un biais existerait dans le code génétique entre enrichissement des codons en uracile et traduction d'acides aminés hydrophobes. En comparant la conservation des CDS à travers les espèces, ils ont montré que les CDS estimées comme jeunes étaient AT riches comparées aux CDS anciennes plus GC riches. Pour eux un modèle se dégage dans lequel, les nouvelles ORF émergeant de séquences non codantes AT riches seraient plus enclins à coder des acides aminés hydrophobes ou aromatiques. Ce biais dans le code génétique favoriserait la dégradation de ces microprotéines. Cependant, les auteurs révèlent aussi un mécanisme dans lequel les nouvelles microprotéines plus hydrophobes pourraient être engagées dans les membranes justement grâce à leurs structures hydrophobes en hélice et ainsi échapperaient à la dégradation. Ils trouvent que sur 64 microprotéines *de novo* 74% seraient membranaires et seulement respectivement 14 et 12% seraient cytoplasmiques ou nucléaires (Kesner et al. 2023). Ce double destin antagoniste des microprotéines hydrophobes reste pour le moment mal compris. Les auteurs suggèrent qu'un complexe BAG6 médierait la localisation des microprotéines aux membranes tout en jouant un rôle dans leur dégradation par le

protéasome. Le choix entre les deux voies reste là aussi mal compris et des validations sont nécessaires afin de confirmer le rôle dual de BAG6.

Une étude bioinformatique compilant plus de 80 protéomes du cœur humain, trouve et valide *in vivo* un bon nombre de microprotéines issues de circARN et lncARN comme associées aux mitochondries, plus précisément les membranes des mitochondries. Ces auteurs suggèrent que les mitochondries pourraient être un terrain de jeu évolutif permettant à des microprotéines d'être stabilisées (van Heesch et al. 2019). Nous nous alignerons ici à un point de vue plus global où les membranes de cellules et d'organelles pourraient servir de base à l'émergence de nouveaux gènes fonctionnels. Ces alt-ORF stabilisées par les membranes pourraient par la suite s'enrichir en GC et ainsi échapper à la dégradation protéosomale. De plus ces alt-ORF peuvent par la suite gagner des domaines fonctionnels. Chez les rongeurs, il a par exemple été documenté que l'alt-ORF NOXA ayant émergé *de novo* a, par la suite, doublé de taille avec la duplication du domaine BH3 (Schlesinger et al. 2022).

#### III.4.2 Mode d'action des microprotéines et des alt-ORF

Très peu d'études à ce jour ont identifié la structure précise de microprotéines. De plus, les algorithmes actuels ont des difficultés à pouvoir prédire une structure du fait qu'ils ont été entraînés ou conçus pour résoudre des structures peptidiques beaucoup plus longues (Mohsen et al. 2023).

Il a déjà été décrit ci-dessus qu'une composition riche en acides aminés hydrophobes favorise la dégradation ou la localisation des microprotéines aux membranes cellulaires ou des organelles (Kesner et al. 2023). Des études récentes ont montré que les microprotéines sont plus enrichies que les protéines en régions intrinsèquement désordonnées (Mohsen et al. 2023). Ces microprotéines peuvent alors médier leurs fonctions (détaillées dans la Figure 12) et leurs interactions avec d'autres facteurs par de courts motifs d'interaction linéaires ou *short linear interaction motif* (SLIMs).

#### III.4.3 Interactions avec facteurs

Suggéré pour la première fois par Staudt et Wenkel, les microprotéines de par leur faible taille (54-57 aa en taille moyenne (Cardon et al. 2019; Wright et al. 2022)) auraient la capacité d'interagir de façon précise sur de grands complexes protéiques sans avoir nécessairement de difficulté à accéder au site de liaison. Diverses études, dont plusieurs à

grande échelle, montrent maintenant des interactions de microprotéines avec des protéines ou complexes plus ou moins grands (Cardon et al. 2019; Wright et al. 2022). Ces interactions microprotéines-protéines peuvent perturber une interaction déjà existante et agir comme un inhibiteur/dominant négatif. La microprotéine STORM, homologue au domaine de reconnaissance de SRP19, rentre en compétition avec elle pour se lier sur un site de l'ARN 7S. Cette compétition perturbe aussi l'interaction de l'ARN 7S avec d'autres sous-unités SRP et perturbe ainsi l'assemblage du complexe SRP (Min et al. 2017). D'autres interactions se sont révélées comme des facteurs manquants à des complexes. C'est le cas de ASDURF, microprotéine et sous-unité manquante du complexe chaperon PAQosome (Mohsen et al. 2023). Les microprotéines, de par leurs interactions, ont aussi été montrées comme favorisant la formation de complexes dont elles font partie. En effet, des évidences biochimiques et génétiques suggéreraient que POLR2L est requis pour l'assemblage et l'activation des ARN polymérases eucaryotes (Schlesinger et al. 2022). POLR2L de par sa conformation en 4 hélices joindrait les autres sous-unités ensemble.

Les microprotéines peuvent aussi agir comme régulatrice d'activité de facteurs ou de complexes protéiques. Un exemple bien documenté chez l'Homme est celui de la régulation de l'activité de l'ATPase du réticulum sarco/endoplasmique  $Ca^{2+}$  (SERCA). En effet, deux microprotéines Sarcolipin et phospholamban ont été montrées comme maintenant le complexe SERCA dans une conformation intermédiaire qui ralentit ainsi l'activité enzymatique (Schlesinger et al. 2022). Il est à noter qu'une étude a décrit qu'en plus de son rôle sur SERCA, Phospholamban pouvait aussi s'autoassembler en une structure homopentamérique (Wawrzynow et al. 1992) se fixant à la membrane. Cette structure a une architecture similaire à un canal à ion transmembranaire pouvant être dynamique. La fonction de cette structure n'a cependant pu être élucidée pour le moment (Oxenoid et al. 2005; Maffeo et al. 2009).

#### III.4.4 Excrétion/ligands

Pour l'instant peu étudiées chez l'Homme, deux microprotéines fonctionnelles excrétées ont pu être identifiées. L'humanin est une microprotéine de 24 aa qui a été retrouvée dans le sérum humain. Elle semble protéger la mort cellulaire des neurones dans un contexte de maladie d'Alzheimer avec prévalence génétique. Cette dernière interagirait avec des facteurs intracellulaires et à la membrane jouant donc le rôle d'un ligand. Le mécanisme global d'action n'est cependant pas résolu pour le moment (Mohsen et al. 2023). L'Apelin est une microprotéine de 77 aa pouvant être clivée en des isoformes multiples (Y. Ma et al. 2017;

Tatemoto et al. 1998). Sa ligation acidifie les cellules réceptrices (Tatemoto et al. 1998). Elle a été montrée comme pouvant se lier au récepteur à l'apelin, récepteur couplé aux protéines G. Sa ligation peut s'effectuer par deux sites sur un même récepteur ce qui est une caractéristique unique pour ce type de récepteur (Y. Ma et al. 2017). Entre autres, une étude focalisée sur le traductome dans des tissus de cœur humain suggère que beaucoup de microprotéines sécrétées non caractérisées pourraient exister (van Heesch et al. 2019).

#### III.4.5 Microprotéines membranaires

Les protéines membranaires représentent 1/3 des gènes codants et environ la moitié de la masse en protéines de la cellule. Bien que ces dernières soient très documentées pour leurs fonctions, leurs modalités de régulation, d'assemblage et de localisation restent mal comprises. Certains auteurs suggèrent que les microprotéines permettraient de comprendre ce vide bibliographique (Makarewich 2020). Une étude a notamment caractérisé une microprotéine nommée MOXI enrichie au niveau des mitochondries des muscles. Elle se situe sur la membrane interne de la mitochondrie et a été observée comme interagissant avec plusieurs complexes régulant ainsi le rôle métabolique de la mitochondrie notamment au niveau de l'oxydation des acides gras, la respiration et le potentiel membranaire. La précédente étude fait aussi la supposition, de par les multiples rôles des microprotéines dans les membranes, que ces dernières pourraient participer à une communication interorganelles. Elle s'appuie notamment sur une étude chez l'Homme avec l'exemple de PIGBOS présent à la mitochondrie, mais interagissant avec le réticulum endoplasmique déclenchant ainsi la *unfolded protein response* (Chu et al. 2019).

#### III.4.6 Réponses antigéniques et perspectives thérapeutiques

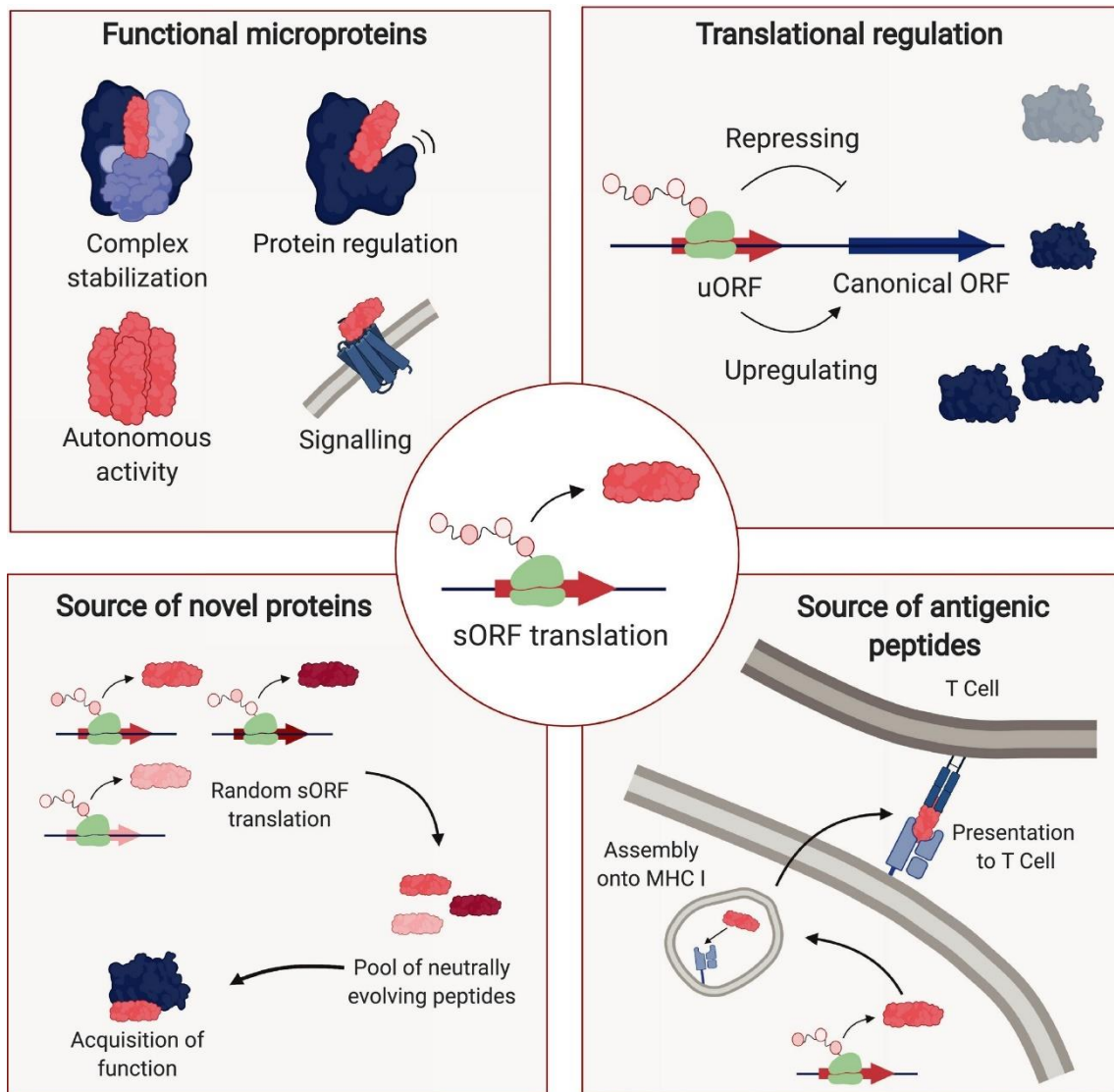
Certaines alt-ORF telles que des uORF ont récemment été montrées comme traduisant des peptides se retrouvant présentés par le complexe CMH1 et pouvant déclencher la réponse immunitaire des lymphocytes T (Wright et al. 2022).

De par leur localisation aux membranes, ces microprotéines pourraient aussi faire l'objet de cible thérapeutique. En effet, plusieurs articles de revue le mentionnent, les microprotéines pourraient être imitées ou ciblées dans le but de déclencher une réponse souhaitée dans le cadre d'une thérapie. Il serait aussi possible de cibler plus facilement des organismes pathogènes arborant des microprotéines spécifiques (Mohsen et al. 2023; Yeasmin et al. 2018; Makarewich 2020).

### III.4.7 Fonctions doubles ARN et protéines

Dans le cas des alt-ORF d'ARNm, il a été montré que ces dernières pouvaient de par leur traduction réguler la traduction de l'ORF canonique. Les uORF peuvent soit favoriser l'ORF canonique soit la réprimer à plus de 80% (Wright et al. 2022). Entre autres, les uORF sont aussi connues pour favoriser en majorité la dégradation de l'ARNm par *Nonsense-mediated decay* (NMD), un mécanisme éliminant les ARNm porteurs de codons STOP prématures (Orr et al. 2020).

Il est à noter qu'un lncARN peut à la fois coder une microprotéine et avoir une fonction non codante. Dans des cellules aortiques endothéliales humaines, il a été montré qu'un lncARN *C15ORF48* servait à la fois à produire un miARN (miR-147b) et une microprotéine (MOCCI), qui exerceraient tous deux une activité anti-inflammatoire en réduisant la production de cytokines (C. Q. E. Lee et al. 2021). Dans d'autres cas, les deux fonctions peuvent être éloignées l'une de l'autre. Chez la souris, le lncARN *Linc-RAM* interagit avec MyoD pour promouvoir la myogenèse alors que la microprotéine produite par sa alt-ORF, elle, régule l'activité du facteur SERCA (Wright et al. 2022).



**Figure 12 : Modes d'actions des microprotéines, alt-ORF et sORF (adapté de Schlesinger et Elsässer 2022)**

De haut en bas et de gauche à droite. Les microprotéines portent un ensemble de fonctions variées. Elles participent en tant que sous-unité à la formation/stabilisation de complexe. Elles régulent l'activité de protéines, peuvent être engagées dans des mécanismes de signalisation intra ou extra cellulaire. Elles peuvent même posséder leur propre activité indépendamment des autres protéines comme Phospholamban. Les alt-ORF, en majorité les uORF du fait de leur proximité avec une autre ORF, agissent en Cis et peuvent moduler la traduction de cette dernière ou la stabilité de l'ARNm. Les alt-ORF sont des réservoirs évolutifs permettant l'émergence de nouvelles protéines pouvant être neutres et non fonctionnelles mais apportant de nouvelles fonctions au cours de l'évolution. Un certain nombre de microprotéines sont la cible des voies de dégradation. Les fragments de ces microprotéines dégradées peuvent être par la suite présentés aux lymphocytes T en tant qu'antigènes et déclencherait des réactions immunitaires. Ces antigènes et les microprotéines plus généralement peuvent être sources de nouvelles stratégies thérapeutiques.



## III.5 Preuves Fonctionnelles des alt-ORF et microprotéines

### III.5.1 Détection de la microprotéine

En utilisant des approches à spectre large (MS, Ribo-seq et analyse bioinformatique), beaucoup d'alt-ORF potentielles sont trouvées (voir partie III.1). Il faut cependant pouvoir valider ces candidats à l'échelle individuelle. De nombreuses études utilisent des vecteurs exogènes (plasmides) pour montrer la détection d'une microprotéine (Meng et al. 2023). Néanmoins, ces résultats sont à prendre avec précaution, car une surexpression de protéine peut perturber le repliement de la protéine en saturant le système de repliement, déréguler le dosage du gène, l'assemblage de complexe et les régulations en aval de ce gène. Une mauvaise localisation de la protéine peut alors avoir lieu (Gibson et al. 2013).

Pour prouver l'existence de la microprotéine au niveau endogène, il faut soit créer un anticorps spécifique à la microprotéine soit mettre une balise (« tag ») sur cette dernière (Hassel et al. 2023). La création d'anticorps est l'approche en théorie optimale pour pouvoir détecter la microprotéine native. Cependant, certaines microprotéines possèdent un très faible potentiel immunogénique et la création de l'anticorps est donc très complexe. De plus, trouver un épitope spécifique unique est moins probable dans le cas d'une microprotéine du fait de sa petite taille (Hassel et al. 2023). Certaines microprotéines possèdent aussi beaucoup de régions intrinsèquement désordonnées là aussi problématiques pour obtenir un bon épitope (Sormanni et al. 2015). Pour le balisage, ce dernier peut être réalisé en N-ter ou en C-ter de la microprotéine avec une influence toutefois possible de ce positionnement. Il est en effet connu que les balises en N-ter peuvent influencer sur la localisation de la protéine tandis que le balisage en C-ter rend moins soluble la protéine (Makarewich et al. 2017). Par ailleurs, la question se pose quant à la stabilité apportée à la microprotéine par le balisage. Kesner et al ont en effet montré que les 30 derniers acides aminés influençaient la stabilité d'une protéine rapportrice. Cependant, il a aussi noté que l'ajout d'une balise d'une taille inférieure à 13 aa n'influçait pas en moyenne la dégradation de sa protéine rapportrice. Dans la littérature, les balises mises en C-ter de microprotéines pour détection en endogène ont une taille pouvant être comprise entre 6 (Histidine-tag) à 22 aa (3x-Flag) pour les plus grands. Certaines balises en C-ter pourraient donc favoriser une certaine stabilisation de la microprotéine à l'étude. Le balisage reste cependant une solution généraliste pratique sans avoir besoin de développer un anticorps pour chaque nouveau candidat.

### III.5.2 Abondance, stabilité et purification

Comme mentionné précédemment, les alt-ORF peuvent être associées à la dégradation de l'ARN par l'exosome et la voie de dégradation NMD. Au niveau de la microprotéine, une étude mentionnée précédemment a montré que les microprotéines pouvaient être dégradées par le protéasome. Les microprotéines sont plus facilement dégradées si elles sont hydrophobes bien que cette hydrophobicité puisse aussi les localiser aux membranes les protégeant de la dégradation (Kesner et al. 2023).

Très peu d'études font état de l'abondance relative des microprotéines dans la cellule. Dans une étude peptidomique travaillant sur 90 microprotéines, les auteurs avaient estimé que chacune des microprotéines à l'étude possédait entre 10 et 1000 molécules dans la cellule. Cette concentration est similaire à certains niveaux de protéines canoniques (Slavoff et al. 2013).

La difficulté de la détection de la microprotéine va donc résider dans la difficulté à pouvoir l'extraire sans perte trop importante en quantité notamment due à sa petite taille (voir partie III.1.2).

### III.5.3 Implications physiologiques prouvées

De façon générale, il semblerait que les microprotéines puissent à l'instar des protéines en fonction de l'ORF être spécifiques à des tissus ou exprimées de façon ubiquitaire. En effet, il a été montré par exemple que la modulation de l'activité du complexe SERCA pouvait s'effectuer par 5 microprotéines différentes, mais de conformations structurales similaires. Parmi ces 5 microprotéines, 3 sont spécifiques aux tissus musculaires tandis que les 2 restantes sont exprimées dans les tissus non musculaires (Anderson et al. 2016)

Les alts-ORF ont été montrées comme impliquées dans la traduction de microprotéines directement engagées dans le complexe CMH 1 jouant ainsi un rôle dans l'immunité et la réponse immunitaire (Schlesinger et al. 2022). Entre autres, pour les alt-ORF, comme mentionné précédemment, ces dernières sont connues pour pouvoir moduler l'expression de l'ORF canonique et touchent ainsi une multitude de fonctions cellulaires.

Les microprotéines ont été montrées comme impliquées dans divers processus tels que la maturation des ARN et des protéines, l'expression des gènes, le métabolisme, la signalisation

et le trafic cellulaire (Schlesinger et al. 2022). Elles sont aussi impliquées dans les mécanismes de réparation dans les voies NHEJ et de la recombinaison homologue (stress génique) (Slavoff et al. 2014; Arnoult et al. 2017; Hung et al. 2018) ainsi que de réponses au stress du réticulum endoplasmique (Chu et al. 2019). Elles auraient aussi des fonctions dans la prolifération et la survie cellulaire (Prensner et al. 2021), la régénération et la fonction musculaire (Matsumoto et al. 2017), la mort cellulaire. De même, des microprotéines ont pu être découvertes dans des processus mitochondriaux. Elles ont aussi été montrées comme impliquées dans l'immunité (Schlesinger et al. 2022) dans différentes pathologies humaines telles que le cancer (Prensner et al. 2021; Hassel et al. 2023), l'obésité et les maladies du cœur (Hassel et al. 2023). Le domaine des microprotéines, bien qu'étant peu documenté et émergent chez l'Homme, semble déjà toucher une grande diversité de processus cellulaires importants. D'autres exemples sont notamment mentionnés ci-dessous.

Microprotein	Function	Human gene	Human AA length	Human UniProt ID	Mouse gene	Mouse AA length	Mouse UniProt ID
ALN	Negative regulator of SERCA	<i>C4orf3</i>	66	Q8WVX3	<i>1810037117Rik</i>	65	Q99M08
APPLE	Regulator of translation initiation through promotion of eIF4A initiation complex assembly	<i>ASH1L-AS1</i>	90	NR	NR	NR	NR
ASAP	Positive regulator of mitochondrial ATP production through interaction with ATP synthase	<i>LINC00467</i>	94	NR	NR	NR	NR
BRAWNIN	Electron transport chain complex III assembly factor	<i>UQCC6</i>	71	Q69YU5	<i>Uqcc6</i>	67	Q8BTC1
CASIMO1	Regulator of cytoskeletal organization, cell migration and proliferation	<i>SMIM22</i>	83	K7EJ46	<i>Smim22</i>	86	V9GXA9
CIP2A-BP	Negative regulator of PI3K/AKT/NFκB pathway through interaction with CIP2A	<i>LINC00665</i>	52	NR	NR	NR	NR
CYREN-1/MRI-1	Regulator of non-homologous end-joining	<i>CYREN</i>	157	Q9BWK5-1	<i>Cyren</i>	157	Q8BHZ5
CYREN-2/MRI-2	Regulator of non-homologous end-joining	<i>CYREN</i>	69	Q9BWK5-4	NR	NR	NR
DWORF	Positive regulator of SERCA2a in heart and slow-twitch skeletal muscle fibers	<i>STRIT1/DWORF</i>	35	P0DN84	<i>Strit1/Dworf</i>	34	P0DN83
ELN	Negative regulator of SERCA3	<i>SMIM6</i>	62	P0DI80	<i>Smim6</i>	56	Q3U0I6
Humanin	Anti-apoptotic activity	<i>MT-RNR2</i>	24	Q8IVG9	<i>Gm20594</i>	37	J3QJY3
MIAC	Negative regulator of actin cytoskeleton through regulation of SEPT2/TGB4 signaling	<i>RP11-469H8.6/AQP5-AS1</i>	51	NR	NR	NR	NR
NR miPEP133	Regulator of p53 transcription and mitochondrial HSPA9 activity	<i>MIR34A</i>	133	NR	<i>Mir34a</i>	NR	NR
MLN	Negative regulator of SERCA1	<i>MRLN</i>	46	P0DMT0	<i>Mrln</i>	46	Q9CV60
MOTS-c	Regulator of insulin sensitivity and glucose metabolism	<i>MT-RNR1</i>	16	A0A0C5B5G6	<i>mt-Rnr1</i>	11	NR
MP31	Negative regulator of lactate-pyruvate conversion through interaction with mitochondrial lactate dehydrogenase	<i>ENSG00000289051</i> 5' UTR of <i>PTEN</i> (uORF)	31	C0HLV8	<i>ENSMUSG00000121574</i> 5' UTR of <i>Pten</i> (uORF)	35	C0HLV9
Mtlbn	Regulator of electron transport chain complex III assembly and function	<i>STMP1/C7orf73</i>	47	E0CX11	<i>Stmp1/1810058I24Rik</i>	47	P0DP99
Mtln	Regulator of fatty acid oxidation, lipid metabolism and respiratory chain activity	<i>MTLN</i>	56	Q8NCU8	<i>Mtln</i>	56	Q8BT35
Myomerge/Myomixer	Regulator of myoblast fusion	<i>MYMX</i>	84	A0A1B0GTQ4	<i>Mymx</i>	84	Q2Q5T5
NEMEP	Regulator of glucose transport and mesoderm development	<i>SMIM43</i>	63	Q4W5P6	<i>Smim43</i>	63	A0A286YD83

NoBody	Regulator of mRNA stability through P-body interaction	NBDY	68	A0A0U1RRE5	Nbdy	68	A0A0N4SUI7
PIGBOS	Regulator of the ER stress response and UPR	PIGBOS1	54	A0A0B4J2F0	Pigbos1	57	A0A5F8MPY0
PLM/FXYD1	Negative regulator of NKA and NCX	FXYD1	92	O00168	Fxyd1	92	Q9Z239
PLN	Negative regulator of SERCA2a	PLN	52	P26678	Pln	52	P61014
pTINCR	Regulator of CDC42 SUMOylation and activation	TINCR	87	A0A1B0GVN0	Tincr	87	A0A1B0GRO3
pTUNAR	Postive regulator of SERCA2 activity during neural differentiation	TUNAR	48	A0A1B0GTB2	Tunar	48	A0A1B0GOX2
SMIM4	Electron transport chain complex III assembly factor	UQCC5	70	Q8WV10	Uqcc5	80	Q8C1Q6
SPAR	Regulator of v-ATPase H+ pump and inhibitor of mTORC1	SPAAR	90	A0A1B0GVQ0	Spaar	75	A0A1B0GSZ0
Stannin/Hemotin	Regulator of endosomal maturation and phagocytic processing in macrophages	SNN	88	O75324	Snn	88	P61807
UQCC3	Electron transport chain complex III assembly factor	UQCC3	93	Q6UW78	Uqcc3	89	Q8K2T4
UQCC4	Electron transport chain complex III assembly factor	UQCC4	132	Q4G010	Uqcc4	136	Q6RUT7

**Figure 13 : Diversité des microprotéines caractérisées chez l'Homme avec leur homologue chez la souris (Hassel et al. 2023)**

De gauche à droite par colonne : nom de la microprotéine ; fonction(s) associée(s) ; gène chez l'Homme ; taille en acide aminé ; UniProtID ; gène ; gène chez la souris ; taille en acide aminé ; UniProtID ; NR : *not reported*, pas de données

Dans cette thèse, nous nous intéresserons aux isoformes IPA comme sources de répression de gènes clés, mais aussi vecteurs d'alt-ORF, sources potentielles de nouvelles microprotéines. Nous nous sommes plus particulièrement concentrés à un contexte biologique particulier, la réponse des cellules cancéreuses du poumon dites non à petites cellules (NSCLC) au cisplatine. Il est intéressant de noter qu'aucune étude ne fait pour le moment mention de microprotéines impliquées dans le stress génotoxique.

## IV. Les agents génotoxiques et la réponse aux dommages à l'ADN : focus sur le cisplatine

### IV.1 Les agents génotoxiques

Les agents génotoxiques sont des agents (physiques, chimiques ou métabolites endogènes) qui peuvent entraîner la mort ou la sénescence cellulaire via l'induction de dommages à l'ADN. Ils sont à différencier des agents non génotoxiques carcinogènes s'appuyant sur d'autres types de mécanismes que les dommages à l'ADN ou ne provoquant qu'indirectement des dommages à l'ADN ; ces agents sont souvent dits « promoteurs de tumeur » et possèdent un seuil supposé pour lequel ils n'auraient pas d'effet (Hartwig et al. 2020). Les agents génotoxiques peuvent être des rayonnements ionisants (rayons gamma et X) ou ultraviolets (UV), les sels de platine (Cisplatine etc...) ou encore des agents chimiques tels que la doxorubicine, la camptothécine, le 5-fluorouracile, etc... (Ismailov et al. 2023). De nombreux agents génotoxiques sont utilisés dans le cadre de chimiothérapie ou radiothérapie pour leurs propriétés antiprolifératives et pro-apoptotiques à dose modérée.

### IV.2 Le Cisplatine

Comme mentionné précédemment, les agents génotoxiques provoquent des dommages directs à l'ADN. Ces derniers doivent cependant arriver à l'intérieur des cellules pour pouvoir y accéder. Dans le cas du cisplatine, après administration intraveineuse (Ghosh 2019) ce dernier arrive dans les cellules par diffusion passive ou en utilisant les voies de transport du cuivre CTR1 (H. H. W. Chen et al. 2015). À l'intérieur des cellules, un changement chimique s'opère sur le cisplatine du fait de la faible concentration en ion chlore. Les deux groupements contenus dans le Cisplatine sont substitués par deux molécules d'eau et forment un hydrate de cisplatine chargé positivement (Dasari et al. 2014; Davies et al. 2000).

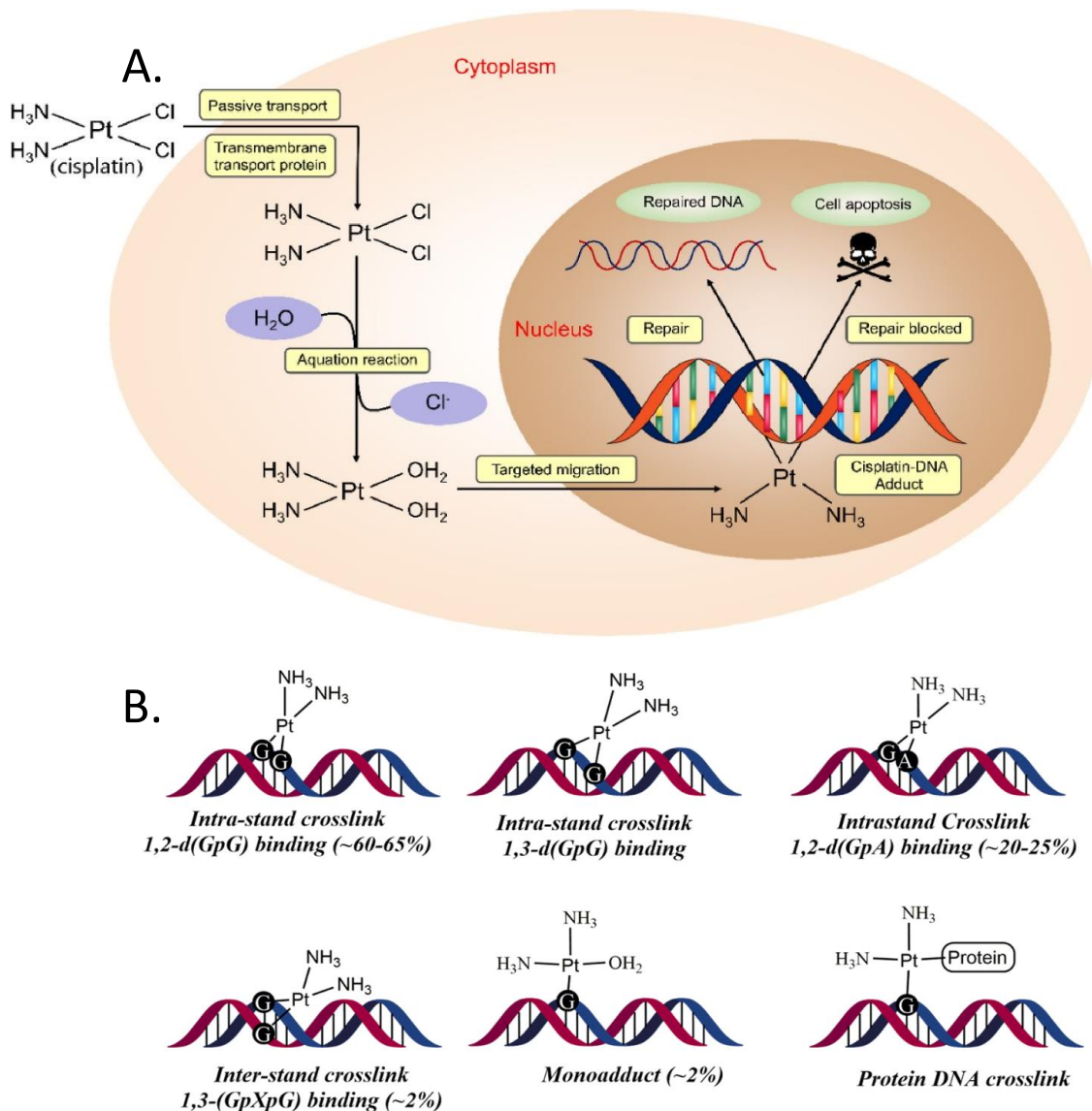
Ces hydrates de cisplatine vont attaquer bon nombre de complexes protéiques, induire de l'apoptose et activer des voies cellulaires variées telles que la voie du calcium, la voie mitochondriale et des récepteurs de signalisation de mort indépendamment de son action sur l'ADN (Florea et al. 2011).

Entre autres, certains agents génotoxiques dont le cisplatine peuvent générer des espèces réactives à l'oxygène (ROS) en fonction du temps et de la concentration (Zhou et al.

2019; Brozovic et al. 2010). Ces ROS très réactifs peuvent à leur tour endommager des molécules/complexes dont l'ADN. Les ROS provoquent notamment des suppressions de bases ou une oxydation des guanines (8-hydroxy-deoxyguanosine ou 8-oxo-dG) provoquant un appariement non conventionnel avec l'adénine (Hartwig et al. 2020). Ils vont être vecteurs de stress cellulaire, déclenchant une réponse antioxydante face à eux (Kanwal et al. 2014).

Les hydrates de cisplatine chargés positivement vont aussi être attirés et réagir principalement avec les acides nucléiques de charge négative (Ghosh 2019)(Figure 14A). Les agents génotoxiques peuvent créer différents types de lésions à l'ADN. Dans le cas du cisplatine, ce dernier va se lier à l'ADN et former des adduits (Todd et al. 2010). Montré *in vitro*, le cisplatine se lie préférentiellement à la position N7 de la guanine mais il peut aussi se lier plus rarement au N7 de l'adénine (Fichtinger-Schepman et al. 1985). Trois topologies d'adduits différents à l'ADN peuvent être formées par le cisplatine. Les pontages intrabrins, formés par les liaisons guanine-cisplatine-guanine ou une guanine-cisplatine-adénine d'un même brin, représentent 90-95% des adduits cisplatine à l'ADN. Les pontages interbrins beaucoup plus minoritaires (~2% des adduits totaux) se forment par liaisons guanine-cisplatine-guanine entre deux brins d'ADN. Les pontages interbrins peuvent dans les conditions physiologiques devenir des pontages intrabrins après passage sous une forme transitoire de mono-adduits (Malinge et al. 1999). Ils peuvent aussi induire des cassures simple brin avec la machinerie de réplication (Noll et al. 2006) ou double brin lors de leur prise en charge par la voie de réparation Fanconi en phase S sur des fourches de réplication bloquées (Long et al. 2011; Räschle et al. 2008; Knipscheer et al. 2009) dont nous reparlerons plus tard. Il est aussi important de noter que la machinerie de réplication rencontrant une cassure simple brin peut générer une cassure double brin (Hanaoka et al. 2016). Le dernier groupe topologique lui aussi minoritaire (~2% des adduits totaux) correspond aux mono-adduits de cisplatine (Figure 14B).

Enfin, les agents génotoxiques (y compris le cisplatine) déforment la conformation native de l'ADN en hélice (Rabik et al. 2007). Globalement, les modifications de l'ADN lorsqu'elles sont détectées permettent le déclenchement de la réponse aux dommages à l'ADN (DDR).



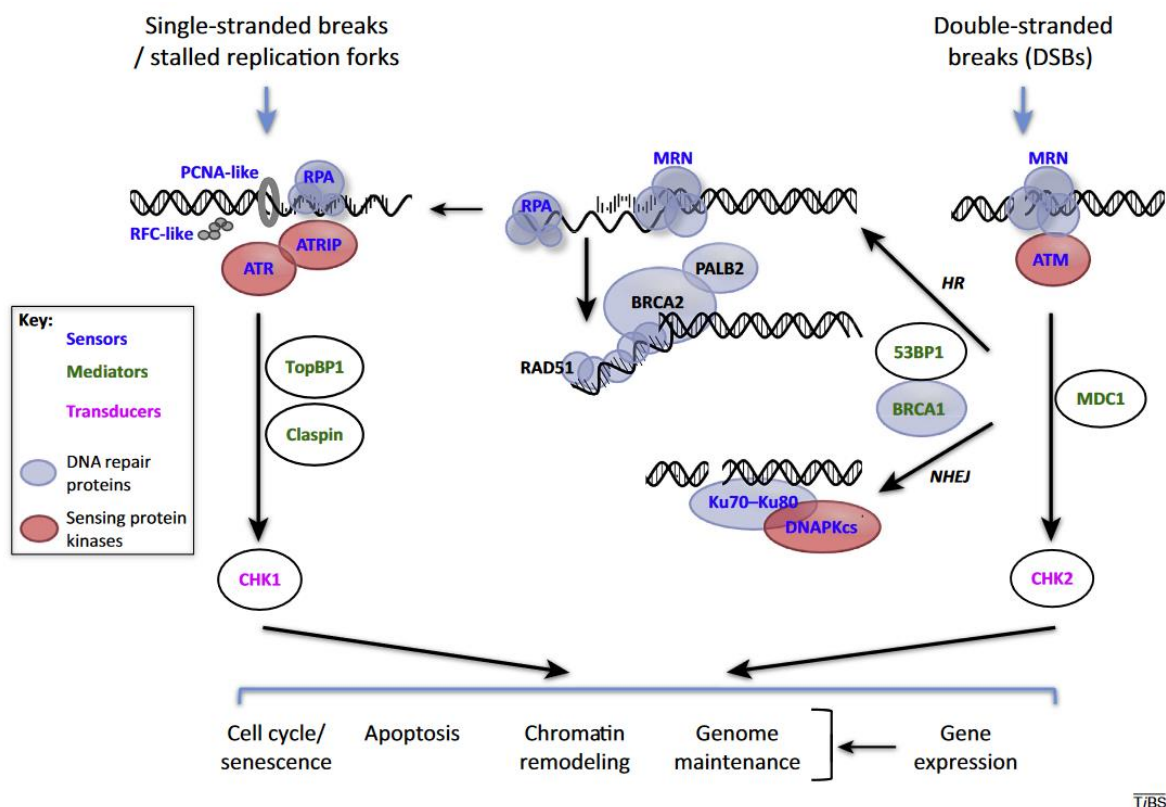
**Figure 14 : Modalités d'entrée et formation d'adduits du cisplatine (adapté de Song et al. 2022 ; Ghosh 2019)**

A : Par diffusion passive ou à l'aide de la pompe à cuivre CTR1, le cisplatine rentre dans la cellule. Due à une concentration faible en ion Chlore dans la cellule, le cisplatine est modifié en hydrate de cisplatine chargé positivement. Ce dernier interagit avec bon nombre de protéines et génère des ROS toxiques pour la cellule. Ces ROS ainsi que les hydrates de cisplatine peuvent interagir avec l'ADN et induire des dommages. Dans le cas des hydrates de cisplatine, ces derniers forment des pontages intra ou interbrins sous formes d'adduits. B : Les différents types d'adduits cisplatine. Le cisplatine forme majoritairement des pontages intrabrin (95%). Ces pontages peuvent être entre N7 de guanines situées l'une à côté de l'autre (1,2-d(GpG)) ou à un nucléotide d'écart (1,3-d(GpG)). Un pontage intrabrin entre N7 de guanine et N7 de l'adénine côte à côte peut aussi se former. De façon plus minoritaire, le cisplatine peut former des monoadduits à l'ADN liés ou non avec une protéine au niveau des guanines (2%). Ils forment aussi des pontages interbrins entre deux guanines. Ces pontages interbrins peuvent se transformer en pontages intrabrin ou induire des cassures simple ou double brin par la suite.

### IV.3 La réponse aux dommages à l'ADN

À la suite de dommages à l'ADN pouvant provenir de plusieurs sources, ici les ROS et les agents génotoxiques, une détection (« *sensing* ») de ces dommages est tout d'abord

essentielle. Elle s'effectue par la reconnaissance de plusieurs types de lésions de l'ADN (délétions, modifications de bases, coupures simple ou double brin) par un ensemble de facteurs appelés senseurs. Cette identification va déclencher une cascade de signalisations transduit à travers d'autres facteurs (transducteurs, médiateurs). Cette cascade permet la modulation d'effecteurs de divers processus, notamment de la réparation, du cycle cellulaire, de voies de survie et de mort cellulaire, ainsi que de régulation de l'expression des gènes. L'ensemble des interactions de ce réseau de facteurs et des régulations de gènes qui en découlent forment la réponse aux dommages à l'ADN (Figure 15).



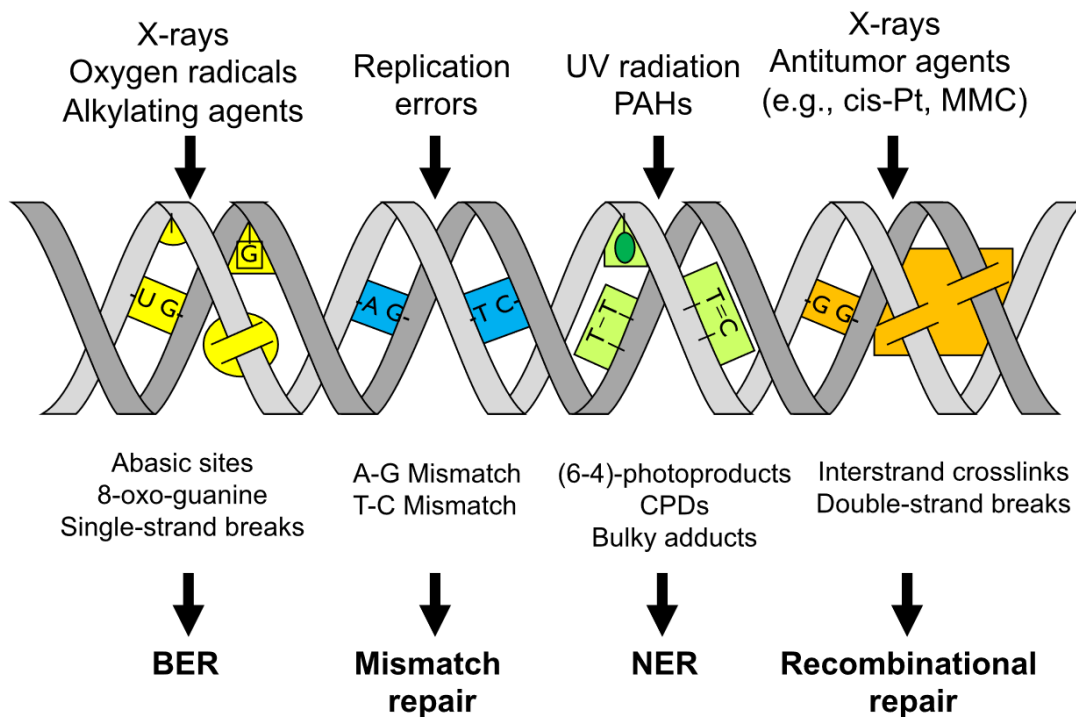
**Figure 15 : Réponses aux dommages à l'ADN (DDR) de cassures doubles brins, dommages simples brin et blocages de fourches de réplication (de Dutertre, Lambert, et al. 2014)**

Les senseurs (en bleu) reconnaissent les dommages à l'ADN et peuvent reconnaître les cassures doubles brins ou les cassures simples brins générées pendant la réparation de dommages simples brins. Parmi ces senseurs, on compte des protéines kinases (facteurs en rouges) comme ATR, ATRIP, ATM et DNA-PKcs phosphorylant localement plusieurs cibles amplifiant le signal. Une aide sera effectuée à l'aide de médiateurs (en vert) permettant l'activation des transducteurs (en violet). Ces transducteurs activent par la suite énormément d'autres facteurs impliqués dans l'arrêt du cycle cellulaire, la réparation de l'ADN, le remodelage de la chromatine, l'expression des gènes et la mort cellulaire. Les facteurs en bleu correspondent aux protéines impliquées dans la réparation de l'ADN.

#### IV.3.1 Les voies de réparation

Selon le type de dommage à l'ADN et selon la phase du cycle cellulaire, différentes voies de réparation et de tolérance des lésions d'ADN peuvent être engagées (Figure 16).





**Figure 16 : Les voies de réparations et leurs dommages associés dus au stress génotoxique (de Hartwig et al. 2020)**

Les rayons X, agents alkylants et les ROS peuvent provoquer des pertes de bases, des cassures simple brin (rayons X uniquement), des oxydations de guanine provoquant un mésappariement avec l'adénine. Ces dommages seront réparés par le *Base Excision Repair* (BER). Les erreurs de réplication provoquent des mésappariements A-G, T-C qui peuvent être réparés par le *Mismatch Repair* (MMR). Les UV-C provoquent des pontages entre deux thymines formant des dimères. Le cisplatine peut provoquer des pontages intrabrins entre guanines ou entre guanines et adénines. Ces pontages vont être réparés par le *Nucleotide Excision Repair* (NER). Le cisplatine, les rayons X et la mitomycine C peuvent générer des pontages interbrins ainsi que des cassures double brin. Les pontages interbrins seront réparés par la suite par la voie Fanconi ou le *Transcription-coupled Nucleotide Excision Repair* couplé aux polymérase de translesion (TLS) (voir ci-dessous). Les cassures double brin sont résolues par le *Non-Homologous End Joining* (NHEJ) et la *Homologous Recombination* (HR).

**Réparation par excision de nucléotides (NER ; Figure 17)** – Le NER élimine une grande variété de dommages à l'ADN affectant un seul des deux brins comme ceux induits par les UV-C et les pontages intrabrins du cisplatine par exemple. Le NER peut être divisé en deux voies alternatives divergeant par la façon dont sont reconnus les dommages à l'ADN mais similaires pour le traitement de ces dommages par la suite. L'une est généraliste appelé *Global Genome repair* (GGR) et l'autre se retrouve couplée à la transcription et se nomme le *transcription coupled nucleotide excision repair* (TC-NER) (Hanaoka et al. 2016). Le GGR reconnaît les dommages à l'ADN par XPC qui reconnaît les déformations de l'ADN (Sugasawa et al. 2001). Il est aussi connu que DDB2 peut assister XPC dans la reconnaissance des dommages (Wakasugi et al. 2002; Q.-E. Wang et al. 2004). Pour le TC-NER, la reconnaissance des dommages se fait à la suite de l'arrêt de l'ARN pol II face aux dommages (Fousteri et al. 2008). Cette dernière est accompagnée par le complexe TFIIH (comportant les sous-unités XPD et XPB) (Drapkin et al.

1994). TFIIH dans le cas du GGR est recruté grâce à sa sous-unité XPB par XPC (Bernardes de Jesus et al. 2008). Les deux voies de réparation se rejoignent donc à cette étape. La sous-unité XPD de TFIIH modulée par les activités ATPase et hélicase portées par XPB (Oksenysh et al. 2009) participe à une deuxième reconnaissance du site de dommage à l'ADN en le liant et le scannant (Hanaoka et al. 2016). XPA, dont le rôle est pour le moment mal compris, est recruté par XPC (Nishi et al. 2013). Des études indiquent néanmoins qu'elle aiderait l'activité d'hélicase de XPB, ainsi qu'à la reconnaissance du site endommagé (Coin et al. 2008) et stabiliserait le complexe (Hanaoka et al. 2016). TFIIH recrute ensuite XPG stabilisant lui aussi le complexe (Zotter et al. 2006) et enlevant XPC (Riedl et al. 2003). XPA recrute ensuite le complexe XPF-ERCC1 (Riedl et al. 2003). Une première coupure simple brin est alors entreprise par le complexe XPF-ERCC1 puis une seconde par XPG (Tapias et al. 2004; Staresinic et al. 2009) permettant l'excision d'un ADN simple brin de 30 nts environ comportant le site endommagé (Cleaver et al. 1991). Les polymérase répliquatives ( $\delta$  et  $\epsilon$ ) avec PCNA et la ligase I peuvent ensuite remplir le trou laissé par le NER et complétant l'ADN simple brin en double brin (Araújo et al. 2000; Shivji et al. 1995). Il a aussi été montré que la polymérase TLS  $\kappa$  peut aussi intervenir *in vivo* à cette étape (Ogi et al. 2006). En outre, le TC-NER a aussi été montré comme pouvant réparer les pontages interbrins du cisplatine à l'aide de polymérase de translesion (Rev1 ou Pol  $\zeta$ ) (Enoiu et al. 2012).

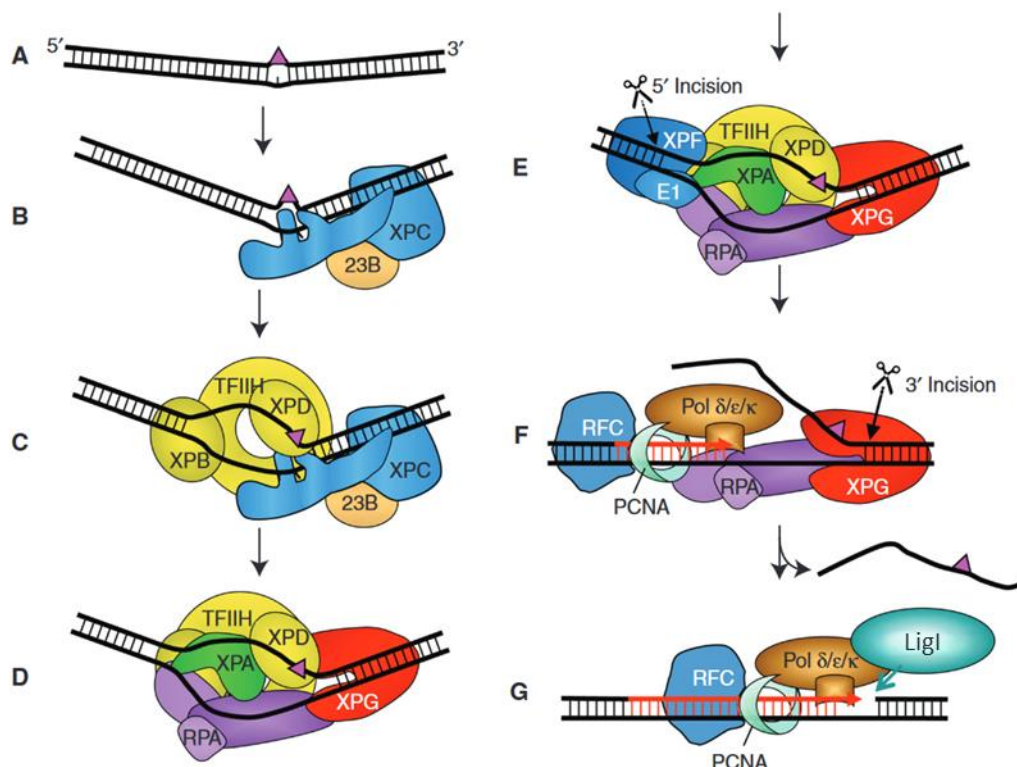


Figure 17 : Réparation par excision de nucléotides (NER) (modifiée de Schärer 2013)

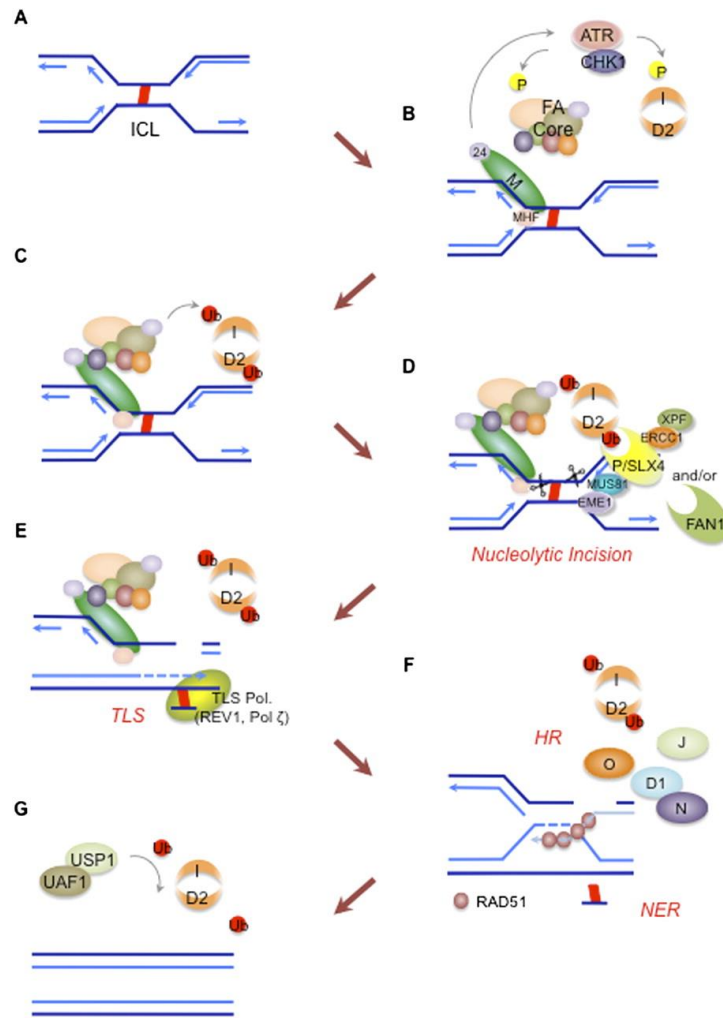
A : Pontage intrabrin. B : Recrutement de XPC-RAD23B reconnaissant le dommage et se liant au brin opposé au pontage. Dans le cas du TC-NER (non montré ici) la reconnaissance s'effectue par le blocage de l'ARN pol II recrutant ensuite CSA et CSB. C : TFIIH interagit avec XPC et ouvre l'ADN avec sa sous-unité XPB permettant à XPD une deuxième reconnaissance du site de dommage. D : Le blocage de XPD au site de dommage permet la formation du complexe de pré-incision en recrutant XPA, RPA et XPG. E : Recrutement par XPA de XPF-ERCC1 permettant une première incision par ce facteur en 5'. F : Recrutement des facteurs de réparation de synthèse Pol  $\delta/\epsilon/\kappa$  et de leurs facteurs associés. La deuxième incision a lieu à la suite. G : Terminaison de la synthèse à l'aide de la ligase I.

**Réparation par excision de bases (BER)** – Le BER répare principalement les dommages causés par les ROS et les agents alkylants altérant les bases de l'ADN. Il agit sur les sites abasiques et les guanines modifiées (8-oxo-dG). C'est une voie à haute-fidélité couvrant un petit nombre d'altérations de l'ADN. Le BER se divise en deux voies appelées short-patch (SP) et long-patch dépendant de PCNA (LP) réparant respectivement 1 et 2 à 8 nucléotides. Le BER implique 4 enzymes qui sont la DNA N-glycosylase aidée par XRCC1 reconnaissant et excisant les bases endommagées, l'AP endonucléase excisant le brin, l'ADN polymérase  $\beta$  synthétisant les nucléotides manquants et la DNA ligase (I ou III) reliant le 3'OH du dernier nucléotide synthétisé au 5' du brin d'ADN (Robertson et al. 2009; Sattler et al. 2003).

**Réparation de mésappariement (MMR)** – Le MMR est responsable de la réparation des mésappariements causés principalement par des erreurs de l'ADN polymérase pendant la réplication (Hartwig et al. 2020). La reconnaissance de ces mésappariements se fait grâce à différents hétérodimères MSH (Hanaoka et al. 2016). Ces divers hétérodimères permettent d'accroître le type de mésappariement et de déformation de l'ADN reconnue (Fishel et al. 1997). L'hétérodimère MSH2-MSH6 reconnaît les mésappariements uniques entre deux bases ainsi que les indels uniques (Mazurek et al. 2009; Marsischky et al. 1999) tandis que l'hétérodimère MSH2-MSH3 reconnaît les indels plus larges mais reconnaît moins facilement les mésappariements de bases uniques (Harrington et al. 2007; Wilson et al. 1999). D'autres hétérodimères (MSH4-MSH5) reconnaissent les jonctions Holliday présentes notamment dans les événements de recombinaison (Snowden et al. 2004). La reconnaissance de ces mésappariements par les hétérodimères MSH peut être assistée du complexe PCNA/Beta clamp (Lau et al. 2003). Deux types de réactions d'excision peuvent alors avoir lieu principalement en fonction de quel brin synthétisé est touché par un mésappariement (*lagging strand* ou *leading strand*). L'excision en 5' a lieu en général dans le *lagging strand* au niveau des fragments d'Okazaki. La reconnaissance des mésappariements dans ce brin entraîne le recrutement de EXO1 une 5'->3' exonucléase enlevant la région du mésappariement. EXO1 est inhibé par RPA et s'arrête lors de rencontre avec un ADN simple

brin protégé par RPA (Hanaoka et al. 2016). La synthèse (Pol  $\delta$ ) et ligation (ligase I) de la région tronquée est assurée par la machinerie de réplication en phase S (Yanbin Zhang et al. 2005). Pour l'excision en 3', se produisant sur le *leading strand*, cette dernière nécessite le recrutement du complexe MLH/PMS (MLH1 et PMS2 principalement) porteur de l'activité endonucléase (Yanbin Zhang et al. 2005). Cette excision est actuellement encore mal comprise. On sait que, l'activation de l'activité endonucléase s'effectue notamment via une interaction avec PCNA. Un recrutement de EXO1 dans le cas de l'excision 3' est nécessaire aussi pour enlever les fragments d'ADN clivés par l'activité endonucléase de MLH/PMS (Goellner et al. 2014). La synthèse et ligation sont ici de nouveau assurées par la machinerie de réplication de phase S (Hanaoka et al. 2016).

**Réparation par la voie Fanconi** (Figure 18)– Cette voie comptant plus de 16 gènes est engagée dans la réparation des pontages interbrins notamment lors de blocages de fourches de réplication (phase S). Le blocage de fourche active alors FANCM/FAAP24 qui recrute à son tour le complexe centrale FA (FANCA, B, C, E, F, G, L et M) pouvant être associé avec d'autres protéines appelées protéines associées à FA (FAAPs) (Kim et al. 2012). FANCM va alors activer ATR (Kim et al. 2012). Ce complexe, et plus particulièrement FANCL, participe à la mono-ubiquitination de l'hétérodimère FANCD2-FANCI phosphorylé préalablement par ATR activé. Cette modification active l'hétérodimère FANCD2-FANCI permettant la liaison du complexe au site de dommage et servant de support pour recruter des nucléases (Kim et al. 2012). Il recrute notamment soit FAN1, soit SNM1A, soit SLX4 associé avec ERCC1-XPF, MUS81-EME1 ou SLX1 (Kim et al. 2012; Kottemann et al. 2013). Dans toutes les configurations, cela aboutit à la conversion du pontage interbrin en cassure double brin pour un brin et la présence d'un adduit sur l'autre. Des mécanismes de polymérase de translésion (REV1, pol  $\zeta$ ) permettent alors (Kim et al. 2012) de passer outre l'adduit. La HR activée préférentiellement répare ensuite la cassure double brin. Le NER peut enlever les adduits restants et finit de synthétiser des trous restants (Kottemann et al. 2013; Kim et al. 2012).



**Figure 18 : Interactions des protéines de la voie de réparation Fanconi pour réparer un pontage interbrin en phase S (de Kim et D'Andrea 2012).**

A : 2 fourches de réplifications convergent vers le pontage. B : Le complexe FANCM-FAAP24-MHF1 reconnaît le blocage de fourche de réplification et recrute les autres facteurs du Fanconi au pontage. FANCM initie l'activation de la voie ATR-Chk1 qui phosphoryle en retour des facteurs de Fanconi C : Mono-ubiquitinylation du complexe FANCD2-FANCI permettant son activation. D : FANCD2 ubiquitinylé joue le rôle de plateforme pour de multiples nucléases telles que SNM1A, FAN1 ou encore SLX4/FANCP associé à XPF-ERCC1 ou MUS81-EME1 ou SLX1. E : Après clivage, le complexe se décroche laissant un des brins en cassures double brins et le brin complémentaire avec l'adduit. Les polymérases TLS (Rev1 ou  $\zeta$ ) synthétisent au travers de l'adduit. La cassure double brin est réparée principalement par le HR via un recrutement de RAD51 par certains facteurs Fanconi. Le NER enlève l'adduit simple brin. G : FANCD2-FANCI est inactivé par USP1-UAF1 finalisant la réparation.

**Non-Homologous End-Joining (NHEJ)** – Le NHEJ est une des deux voies de réparation impliqué dans la résolution des cassures double brin. Il a lieu dans toutes les phases du cycle cellulaire, là où la voie alternative, sans perte d'information génétique, de réparation des cassures double brin, la recombinaison homologue, a lieu durant la réplication et en G2 (Lieber 2010). Les cassures double brin font partie des types de cassures les plus léthales pour une cellule, aboutissant à la perte de bras de chromosome ou à de sévères réarrangements chromosomiques (Durante et al. 2013). La reconnaissance des cassures double brin se fait par l'hétérodimère Ku70/80 se liant aux deux extrémités créées par la cassure (Walker et al. 2001).

Ku70/80 recrute ensuite la DNA-PKcs (Gottlieb et al. 1993) qui s'active par autophosphorylation après fixation à l'ADN (Chan et al. 1996; Weterings et al. 2003) permettant le recrutement des facteurs suivants (Ding et al. 2003). Si les extrémités sont franches avec un phosphate en 5' et un groupement OH en 3', la ligase IV en complexe avec XRCC4 lie les deux extrémités (Critchlow et al. 1997; Grawunder et al. 1997). Si les extrémités ne sont pas compatibles, des nucléotides ou des groupements phosphates peuvent être ajoutés ou enlevés pour rendre ces extrémités compatibles (Lieber 2010). Il existe une voie alternative au NHEJ, appelée *Alternative End-Joining* ou *Microhomology Mediated End Joining* (MMEJ), indépendante des facteurs NHEJ, se basant sur des microhomologies allant jusqu'à 10 nts. Cette voie est plus propice aux réarrangements chromosomiques (Deriano et al. 2013).

**Recombinaison homologue (HR)** – Deuxième voie réparant les cassures doubles brins, la recombinaison homologue a lieu durant la réplication, en phase G2 (Lieber 2010) et la méiose (Hanaoka et al. 2016) dont nous ferons l'impasse dans cette introduction. Permettant de ne pas générer d'erreurs, ni perdre d'informations génétiques, la HR nécessite cependant la disponibilité du brin homologue non endommagé présent sur le chromosome sœur (Krejci et al. 2012; Renkawitz et al. 2014). Mal réalisée la HR peut toutefois générer des réarrangements chromosomiques créant de l'instabilité génétique pour la cellule (Hoang et al. 2010; Putnam et al. 2009). Une digestion des brins 3' des deux extrémités de la cassure va avoir lieu d'abord par le complexe MRN régulé par Sae2 (Lengsfeld et al. 2007), puis par la Dna2 et le complexe STR ou Exo1 seul étendant cette digestion (Mimitou et al. 2008; Zhu et al. 2008). Le brin d'ADN rendu simple brin permet la fixation de RAD51 (Krejci et al. 2012). La recombinaison homologue se divise alors en 3 sous-voies de réparation en fonction de la topologie ou du locus de la cassure double brin. Elle est à distinguer du *single-strand annealing* (SSA) pouvant alors, indépendamment de la fixation de RAD51 (Mimitou et al. 2008), réparer la cassure double brin si elle se produit dans une région de séquence répétée par homologie avec ces séquences. Le SSA génère en effet une perte d'information génétique par raboutage de séquences répétées (Hanaoka et al. 2016). RAD51 est capable d'aller chercher la séquence complémentaire correspondant à l'ADN rendu simple brin au niveau du site de cassure, formant ainsi une D-loop (Krejci et al. 2012; Renkawitz et al. 2014). Les trois voies qui suivent alors sont le *Break-induced replication* (BIR), le *synthesis-dependant strand annealing* (SDSA) et le *double-strand break repair* (DSBR). Le BIR est notamment utilisé lorsqu'un bout de bras chromosomique est perdu (Malkova et al. 2013). Le SDSA correspond lui à un enlèvement de la structure D-loop créée lorsque la synthèse du brin complémentaire est terminée. Cette

disparition de la structure D-loop se fait notamment par des hélicases (Pâques et al. 1999; McMahon et al. 2007). Le DSBR lui correspond à une stabilisation de la structure D-loop par l'invasion de la deuxième extrémité simple brin du chromosome endommagé sur le chromosome sœur. Dans le cas du BIR et du SDSA on note la formation d'une seule jonction d'Holliday alors que le DSBR lui en compte deux (Llorente et al. 2008). La synthèse des brins complémentaire dans tous les cas a été montrée comme assurée par PCNA, Dpb11 et les polymérase répliquatives ( $\delta$  et  $\epsilon$ ) (Germann et al. 2011; Maloisel et al. 2008). Le choix entre SDSA et DSBR a été montré comme influencé par des hélicases telles que Mph1, Srs2, Sgs1 et Irc20 promouvant le SDSA en enlevant notamment le brin envahissant (Ira et al. 2003; Prakash et al. 2009; Miura et al. 2012; Burkovics et al. 2013). Le DSBR lui est favorisé par Rad52 promouvant l'invasion du second brin (Bi et al. 2004).

**Synthèse de translésion** – Dans l'idéal, tous les dommages à l'ADN sont réparés avant le passage en phase S de la cellule. Cependant, de par leur nombre ou leur complexité (pontages interbrins) certains dommages persistent jusqu'en phase S. Ces dommages peuvent alors bloquer les ADN polymérase haute-fidélité en charge de la répliquion (pol  $\delta$  et pol  $\epsilon$ ) avec leur complexe associé. On parle alors de blocage de fourche de répliquion. Ces dommages peuvent être réparés avec (NHEJ) ou sans perte d'information génétique (HR) aidé par la voie de réparation Fanconi (voir ci-dessus). Dans le cas de la HR, on parle alors de tolérance aux dommages à l'ADN, car elle permet la copie de l'ADN même si des dommages sont encore présents dessus. Il existe toutefois une autre voie de tolérance aux dommages à l'ADN, appelée synthèse d'ADN en translésion (TLS), utilisant le brin endommagé comme support de répliquion. Cette synthèse en translésion s'effectue notamment par le biais d'ADN polymérase spécifiques. Il est à noter que ces polymérase peuvent être employées ponctuellement notamment dans la voie Fanconi et NER décrites plus tôt. Il existe chez l'Homme trois familles de polymérase avec des capacités de TLS, la famille-Y (POLH, POLI, POLK et REV1) (W. Yang 2014), la famille-B (pol  $\alpha$ , pol  $\delta$ , pol  $\epsilon$  et pol  $\zeta$ ) (Kazlauskas et al. 2020) et la famille-A (pol  $\theta$  and pol  $\nu$ ) (Arana et al. 2008; Takata et al. 2006). La famille-Y est engagée exclusivement pour du TLS (Hanaoka et al. 2016) alors que la famille B compte des polymérase répliquatives, mais dotées d'activité TLS (Kazlauskas et al. 2020). Du fait de leur nombre conséquent, nous nous concentrerons ici sur les polymérase participant à la réparation des dommages induits par le cisplatine ou les ROS générés par celui-ci.

Pour la famille-Y, ces polymérase sont moins fidèles que les polymérase répliquatives pour synthétiser l'ADN non endommagé, car dépourvues pour la plupart d'activité de

relecture. Elles possèdent aussi une processivité réduite, mais arrivent à synthétiser l'ADN complémentaire en translésion de façon fidèle (W. Yang 2014). Pour la POLH, cette dernière a été montrée comme réparant efficacement et de façon fidèle les pontages intrabrin (1,2-d(GpG)) générés par le cisplatine (Y. Zhao et al. 2012; A. Alt et al. 2007). À l'inverse, elle génère des erreurs en passant à travers les 8-oxoG générées par des ROS (McCulloch et al. 2009). Elles possèdent une faible processivité, synthétisant ainsi de faibles portions d'ADN (Pavlov et al. 2001). La POLI quant à elle passe moins bien au travers des lésions traitées par la POLH comme les 8-oxoG (Yanbin Zhang et al. 2001) mais passe à travers d'autres types de lésions, dont nous ne traiterons pas ici (Hanaoka et al. 2016). La POLK est capable de passer à travers beaucoup de lésions dont 8-oxoG générées par les ROS incorporant des A en complément et plus rarement des C (Y. Zhang et al. 2000). Elle fait des synthèses plus longues que POLH et POLI de 25nts en pouvant insérer ou supprimer des nucléotides (Ohashi et al. 2000). REV1 insère préférentiellement des dCTPs plutôt que les autres nucléotides (Nelson et al. 1996) réparant ainsi les sites apuriques ou apyrimidiques, et passant à travers les mutations de guanines dont les 8-oxoG induites par ROS mais de façon inefficace (Nelson et al. 1996).

Pour la famille-B, la pol  $\zeta$  a la capacité de façon inefficace de passer à travers beaucoup de dommages dont les sites apuriques/apyrimidiques et la modification de guanines, 8-oxoG, générés par les ROS (Haracska et al. 2001). Elle pourrait aussi déclencher des cascades de signalisation de par sa mauvaise fidélité (Hanaoka et al. 2016).

Pour la famille-A, il est à noter que cette famille est dotée d'une processivité modérée et ne possède pas, comme la famille-Y, d'activité de relecture exonucléase leur donnant ainsi une faible fidélité (Arana et al. 2008; Takata et al. 2006). Il a été montré que la pol  $\nu$  pouvait passer à travers des pontages interbrins causés par le psoralen sans erreur (Zietlow et al. 2009). Ce passage translésionnel de la pol  $\nu$  dans un pontage interbrin n'a pas été montré cependant pour d'autres génotoxiques.

**Liens entre voies** - Ces différentes voies de réparations ne sont pas complètement séparées, ni dans le temps ni dans les facteurs qu'elles utilisent, et forment en réalité un réseau dense qui interagit. En effet, beaucoup de facteurs d'une voie se retrouvent aussi impliqués dans une autre. C'est notamment le cas comme cité ci-dessus des polymérases TLS avec le NER et Fanconi. Des facteurs de NER interviennent aussi dans la voie Fanconi. XRCC1 a par exemple été montré comme un composant essentiel pour le BER et le NER ainsi que pour la réparation de cassures simple brin (Moser et al. 2007). Entre autres, lorsqu'un système de réparation n'arrive pas à réparer un site endommagé voire occasionne même d'autres dommages ou des



mésappariements de nucléotides (comme avec les polymérase TLS), une autre voie peut alors s'en charger. Certaines voies peuvent néanmoins être en compétition comme la HR et le NHEJ par exemple où il a été montré que le complexe Ku70/80 à la base de la voie NHEJ pouvait être enlevé par le complexe HR chez la levure (Shim et al. 2010; Mimitou et al. 2010). Entre autres, les régulations de ces deux voies sont opposées, car dépendantes de facteurs pouvant agir en tant qu'inhibiteur d'une voie et activateur de l'autre. C'est notamment le cas pour BRCA1 favorisant la HR et réprimant le NHEJ et 53BP1, TP53 et RIF1 à l'inverse (Hanaoka et al. 2016). Cette opposition permet de faire le choix entre ces deux voies de réparation en charge de réparer toutes deux les cassures double brin.

**Cas du cisplatine** - Dans le cas du cisplatine, il a été montré que les pontages intrabrins sont préférentiellement réparés par le NER associé ou non à l'ARN polymérase II (Moggs et al. 1997). Cette réparation a été montrée comme 15 à 20 fois plus rapide pour les pontages intrabrins 1,3 par rapport aux pontages intrabrins 1,2. Elle passe par une coupure minimale de 29 nts autour du site du pontage suivi d'une synthèse de la séquence manquante par polymérase (Moggs et al. 1997). Entre autres, il a aussi été reporté que le MMR pouvait réparer de façon non robuste certains pontages intrabrins (Vaisman et al. 1998). Pour les pontages interbrins, certes moins abondants, ces derniers ont été observés comme plus difficilement réparable par le NER. En effet, il a été montré chez l'Homme que des facteurs du TC-NER couplé à REV1 et à la pol  $\zeta$  pouvaient enlever des pontages interbrins induits par le cisplatine, et cela, en dehors de la phase S (Enoiu et al. 2012). Ces pontages persistent donc en partie jusqu'à la réplication de la cellule et peuvent contribuer au stress répliatif bloquant les fourches de réplication (Long et al. 2011; Räschele et al. 2008; Knipscheer et al. 2009). Ils sont pris alors en charge par la voie Fanconi, puis par préférentiellement la HR ou le NHEJ (voir ci-dessus).

Dans le cadre des dommages des ROS générés par le cisplatine, ces derniers peuvent être pris en charge par des polymérase TLS, le BER, NER et MMR.

#### IV.3.2 Voies de signalisation et activation des points de contrôle du cycle cellulaire

Tout au long de ces voies de réparation, des facteurs vont être responsables de la signalisation des dommages de la cellule. Les facteurs clés médiant cette signalisation appartiennent au groupe des phosphatidylinositol 3' kinase (PI3K)-related kinases (PIKKs) et comprennent une sous unité de la DNA-PKcs, ATM et ATR (Lanz et al. 2019). Ils vont être activés en étant recrutés par des facteurs des voies de réparation. Dans le cas d'ATR, ce dernier est activé grâce à son partenaire ATRIP (Blackford et al. 2017). Dans le NER et la HR, l'excision de

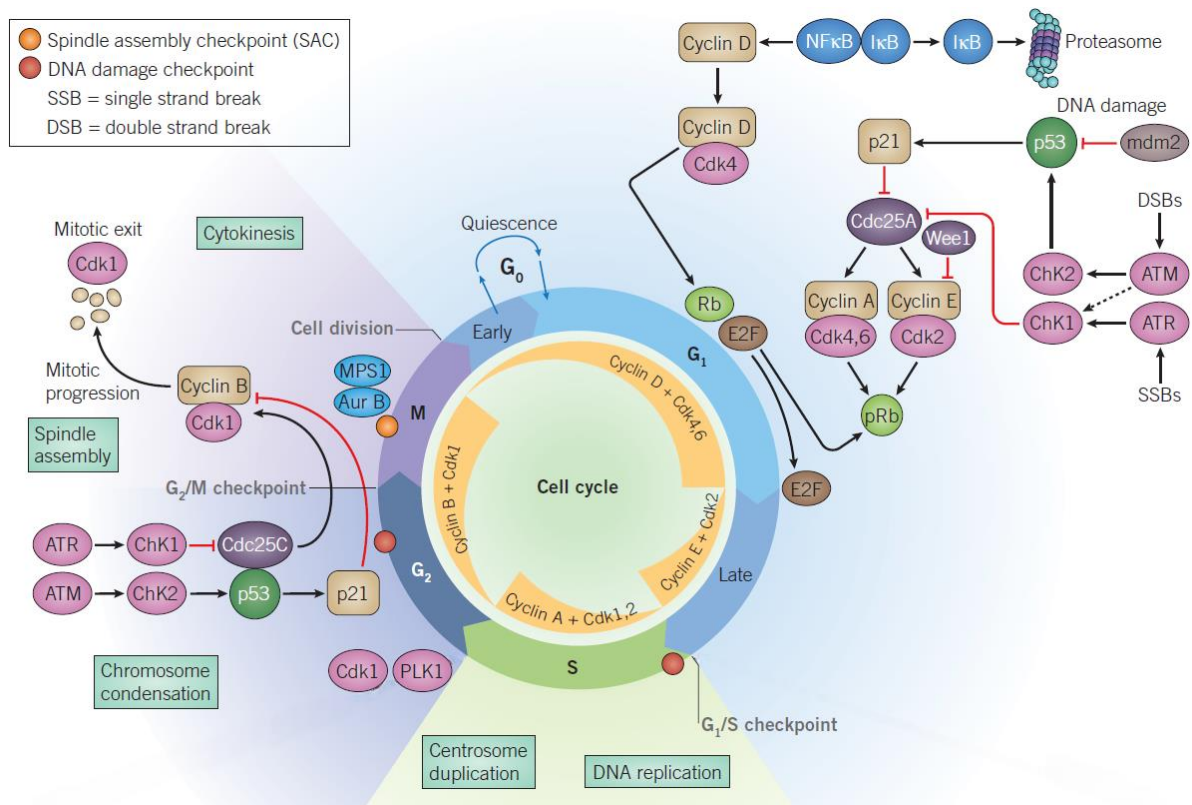
la partie endommagée et la digestion de l'extrémité 3' respectivement laissent la partie complémentaire en simple brin, permettant le recrutement de RPA. Ce recrutement serait notamment reconnu par ATRIP activant donc ATR (Hanaoka et al. 2016). De façon générale, ATR reconnaît ainsi l'ADN simple brin, dont les fourches de réplication à l'arrêt (Lanz et al. 2019) mais aussi une grande variété d'autres dommages. En effet, le recrutement par RPA n'est pas suffisant pour activer totalement ATR et de nombreux coactivateurs ont pu être décrits dans la littérature (Blackford et al. 2017). C'est notamment le cas de FANCM dans la voie Fanconi qui a été montrée comme participant à l'activation d'ATR (voir ci-dessus).

Pour ATM et la DNA-PKcs, ces derniers reconnaîtraient plus spécifiquement les cassures double brin. Pour la DNA-PKcs, cette dernière est recrutée au niveau des cassures par le complexe Ku (voir ci-dessus). Pour ATM, NBS1 sous unité de MRN permettrait là aussi son recrutement aux cassures double brin (Blackford et al. 2017). Entre autres, il a été montré que XPC pouvait aussi participer au recrutement d'ATM (Basu et al. 2010).

Il est aussi à noter que les PIKK peuvent aussi exercer un contrôle sur les voies de réparation en étant par exemple nécessaires à l'activation de certains facteurs de la voie. C'est notamment le cas d'ATR-ATRIP qui participe à la phosphorylation de FANCD2 (Shigechi et al. 2012) dans le complexe FANCD2-FANCI permettant ultérieurement sa mono-ubiquitination l'activant. Un facteur central, P53, va s'activer par phosphorylation à l'aide d'ATR, ATM, Chk2 ou la DNA-PKcs (J. Yang 2003). P53 est à l'interface de nombreuses voies cellulaires comme le cycle cellulaire, l'apoptose, la réponse au stress, la différenciation cellulaire, la prolifération, l'autophagie, la réponse antioxydante et joue donc un rôle extrêmement important dans la réponse aux dommages à l'ADN (Kasthuber et al. 2017). Comme dit précédemment, P53 est aussi déterminant dans le choix s'opérant entre une réparation par NHEJ ou HR. Dans le cas du cisplatine, il a notamment été montré que l'activation de P53 et la signalisation s'effectuaient notamment via une signalisation par ATR et non pas par ATM (Pabla et al. 2008)

Cette signalisation va lors déclencher une cascade de régulation touchant en premier lieu le cycle cellulaire à différentes étapes (Figure 19). Une part importante de la signalisation DDR est l'activation des points de contrôle (« *checkpoints* ») du cycle cellulaire, qui conduisent à un arrêt transitoire du cycle. Ces arrêts transitoires du cycle doivent être distingués de la quiescence, de la différenciation terminale et de la sénescence. La quiescence est un état non prolifératif transitoire (sortie du cycle cellulaire en G0) induit par divers stimuli comme la perte d'adhésion ou l'inhibition par contact, mais elle est réversible lorsque ces stimuli disparaissent.

La différenciation terminale correspond à une sortie permanente du cycle cellulaire sans stress persistant. La sénescence correspond à un état physiologique non prolifératif permanent combiné à un stress persistant comme le raccourcissement des télomères ou l'activation d'oncogènes. Elle agit comme garde-fou pour la cellule (Pack et al. 2019). L'arrêt du cycle cellulaire est un événement ponctuel pouvant conduire à la quiescence, la différenciation terminale, la sénescence, l'apoptose ou le redémarrage du cycle cellulaire. Cet arrêt permet notamment de laisser le temps de réparer les dommages à l'ADN à l'aide des différentes voies de réparation. Il est à noter que l'arrêt de cycle cellulaire se produit tout au long de la vie de la cellule notamment à cause de la réplication, qui est une étape complexe (Pack et al. 2019). Ce stress réplcatif endogène multifactoriel (Zeman et al. 2014) peut alors être accentué grâce aux agents génotoxiques.



**Figure 19 : Cycle cellulaire et points de contrôle du cycle contrôlé par le DDR (du the Cancer Research Product Guide Edition 3, 2015)**

Les dommages à l'ADN peuvent être détectés à tout moment du cycle. Les senseurs et protéines kinases ATM et ATR ainsi que la DNA-PKcs sont recrutés au site de dommage. Les kinases Chk1 et Chk2 sont à l'origine de la cascade de signalisation qui active les points de contrôle en phase G1 et G2. De nombreuses autres kinases et cyclines sont impliquées à divers points de contrôle du cycle.

La voie ATR est essentielle à la cellule, car impliquée dans la réplication de l'ADN, sa réparation et le contrôle du cycle. Chk1 est un des facteurs cible de ATR qui va médier une grande partie de ses fonctions (Hanaoka et al. 2016). ATR peut aussi activer Chk2 de façon plus

minoritaire (Pabla et al. 2008). Comme mentionné précédemment, l'activation d'ATR se fait principalement par la détection d'ADN simple brin lié à RPA résultant soit de dommages à l'ADN soit d'une fourche de réplication à l'arrêt durant la réplication (voir ci-dessus). Dans cette thèse, nous supposons que les arrêts de fourches sont principalement dus à la présence de dommages à l'ADN non réparés bloquant les polymérases répliquatives, bien qu'il existe d'autres causes d'arrêt de fourche dû à des causes endogènes (Zeman et al. 2014). L'activation de Chk1 se fait par polyphosphorylation médiée par ATR (Shimada et al. 2008; Liu et al. 2000). Chk1 phosphoryle ensuite les phosphatases CDC25A et C (Y. Sanchez et al. 1997; Jin et al. 2003) réprimant l'activité des CDK 1 et 2 en phase S et G2 (Jin et al. 2003). Cette répression inhibe la transition de la phase G2 à la mitose et bloque la phase S. Les blocages en phase S se nomment aussi points de contrôle intra-S. Ils sont principalement dus à des blocages de fourches de réplication dus aux dommages activant ATR/ATRIP-Chk1 (Iyer et al. 2017).

La voie ATM est, elle aussi, importante permettant notamment l'activation de Chk2 en détectant notamment les cassures double brin (Blackford et al. 2017). Ce dernier agit également sur CDC25A et C et médie un blocage en S et un blocage de transition entre phase G2 et mitose (J. Yang 2003).

ATM, ATR, CHK2 ainsi que la DNA-PKcs permettent aussi une activation de P53 (via une activation de Mdm2) entraînant un blocage de la cellule en phase G1 via une activation de CDKN1A/p21 (J. Yang 2003; Murray et al. 2020). Les PIKK permettent aussi une signalisation des dommages sur les nucléosomes via notamment une phosphorylation de l'histone H2AX (on parle de  $\gamma$ H2AX). La DNA-PKcs n'est pas connue pour interagir avec Chk1 et 2 mais participe néanmoins à cette signalisation sur l'histone H2AX (Blackford et al. 2017).

Dans le cas du cisplatine, il est documenté dans différents types cellulaires que ce dernier peut bloquer les cellules en phase G2/M et en phase S en activant Chk1 et Chk2. L'activation de ces kinases serait médiée essentiellement par ATR induisant aussi une activation de P53 (Pabla et al. 2008).

### IV.3.3 Régulation de l'expression des gènes

Il est à distinguer deux types de régulation génique en réponse aux dommages de l'ADN. D'une part, pour certains agents génotoxiques (UV-C, camptothécine, cisplatine), les dommages induits dans les gènes induisent des pauses de l'ARN polymérase II et inhibent l'élongation de la transcription (Muñoz et al. 2009; Cullinane et al. 1999). D'autre part, la

cascade de signalisation DDR active des facteurs régulateurs de la transcription, notamment le facteur de transcription p53 (Lanz et al. 2019). Dans le cas du cisplatine, il a été montré une forte activation de p53 par phosphorylation liée à l'activation d'ATR et médiant un rôle central dans l'induction de l'apoptose. P53 va activer une série de gènes impliqués dans la réparation, le cycle cellulaire et l'apoptose (Ghosh 2019).

#### IV.3.3.1 Régulation co- et post-transcriptionnelle des ARN

En ce qui concerne les régulations géniques apportées par les agents génotoxiques, ces derniers peuvent réguler l'épissage alternatif, le clivage/ polyadénylation, la stabilité et la traduction des ARNm en modulant l'activité (ou l'expression) de facteurs impliqués dans ces processus ou dans leur régulation (protéines liant l'ARN ou miARN par exemple) (Dutertre, Lambert, et al. 2014; Dutertre et al. 2021). Il est notamment documenté que les UV peuvent affecter l'épissage alternatif (Muñoz et al. 2009). La camptothécine (inhibiteur de topoisomérase I induisant des cassures d'ADN) a été montrée comme perturbant l'épissage alternatif de gènes clés comme *MDM2* via une inhibition de l'interaction entre les facteurs EWS et YB-1 (qui sont associés à l'ARN pol II). Cet effet se produit également avec la doxorubicine (inhibiteur de topoisomérase II) et le cisplatine (Dutertre et al. 2010). Plus tard, dans le cadre d'un traitement à la doxorubicine, une régulation d'exons terminaux alternatifs d'un groupe de gènes a pu être notée. Cette régulation passe en partie par une inhibition d'une protéine liant l'ARN et régulant l'épissage, HuR (Dutertre, Chakrama, et al. 2014)

Dans le cas du cisplatine, il a notamment été montré dans le carcinome du sein une altération de 717 événements d'épissage dont 245 touchant des exons cassettes. Les gènes dans lesquels cette altération prend place sont des gènes de cycle cellulaire, de maturation des ARNm et d'épissage des ARN pré-messagers. Cette altération a été montrée comme dépendant de P110 $\beta$ , une PI3K mais aussi de SRSF4 un régulateur de l'épissage. Lorsque ce facteur est déplété, une partie des altérations de l'épissage de même que la mortalité induite par le cisplatine sont perdues (Gabriel et al. 2015). Publié récemment dans un autre article sur le cancer du sein, une équipe a confirmé l'impact du cisplatine à l'échelle du génome sur l'épissage des ARN pré-messagers. Elle s'est notamment concentrée sur un gène, *COASY*, impliqué dans le métabolisme mitochondrial. Elle a notamment observé qu'en réponse au cisplatine, une isoforme de ce gène excluant les exons 4 et 5 était générée. Cette régulation de l'isoforme passerait par l'inactivation de RBM39, un effecteur de l'épissage via une interaction avec c-Jun en réponse au cisplatine. Expérimentalement, les régulations apportées par

l'inhibition de RBM39 recoupe celles induites par le cisplatine, sans toutefois toutes les expliquer. Le cisplatine induirait donc par d'autres mécanismes son rôle sur l'épissage. L'isoforme COASY délétée des exons 4 et 5 augmenterait la sensibilité au cisplatine en entravant la fonction mitochondriale (Lemaitre et al. 2022). Dans ces deux études, il semblerait que ce soit l'action du cisplatine sur des facteurs impliqués dans la maturation des ARN pré-messagers qui soit responsable de son effet sur l'épissage des gènes. Entre autres, une traduction différentielle des ARNm semble aussi impliquée dans la résistance au cisplatine. Dans des cellules NSCLC, il a notamment été observé que la traduction de l'ARNm du gène *USP1*, impliqué dans les voies de réparation, était inhibée en réponse au cisplatine dans les cellules sensibles, mais pas dans les cellules résistantes au cisplatine. Cette absence d'inhibition de traduction a été montrée comme médiant la résistance en resensibilisant les cellules par inactivation ou inhibition de la protéine USP1 (Sourisseau et al. 2016)

#### IV.4 Cisplatine et cancer

Le cisplatine est principalement utilisé dans 3 types de cancer : les cancers du poumon, les cancers de l'ovaire et les cancers du testicule.

Dans le cas des cancers du poumon, le cisplatine est utilisé pour les deux types de cancer les plus communs, le cancer du poumon à petites cellules (SCLC : *Small cell lung cancer*) et le cancer du poumon non à petites cellules (NSCLC : *non small cell lung cancer*), qui est le plus fréquent. Traitement le plus efficace pour un SCLC, le cisplatine est utilisé systématiquement après diagnostic (Abrams et al. 2003). Etant plus agressif en se développant rapidement et générant facilement des métastases, le diagnostic d'un SCLC est posé en général alors qu'il y a déjà plusieurs tumeurs disséminées dans le patient (Ghosh 2019). Dans le cas du NSCLC, le cisplatine est utilisé sur les stades précoces de la maladie après une ablation chirurgicale de la tumeur supposée unique (Ghosh 2019).

Pour les cancers épithéliaux de l'ovaire, ces derniers sont en général traités par ablation chirurgicale de la tumeur suivie d'un traitement au cisplatine dans le but de tuer les potentielles cellules restantes de l'opération (Dasari et al. 2014). Pour éviter les rechutes dues à la résistance, des thérapies combinant le cisplatine avec d'autres agents chimiques peuvent être utilisées (Ghosh 2019).

Dans le cas des cancers du testicule, le cisplatine est aussi utilisé pour les deux principaux types de cancers, le cancer du testicule séminome et non séminome (Ghosh 2019). Le cisplatine est le traitement le plus efficace pour les patients avec cancer du testicule séminome avec 85% de guérison (Motzer 1993). Pour les cancers du testicule non séminome, le cisplatine est utilisé en combinaison avec d'autres drogues en fonction du sous-type tissulaire avec au moins 90% de patients guéris dans le cas de tératome par exemple (Dearnaley et al. 1991).

Le cisplatine peut aussi être utilisé de façon moins systématique dans d'autres types de cancers tels que les cancers du cerveau chez l'enfant, les cancers gastriques, du côlon, du sein, de la tête et du cou, de la vessie et certains lymphomes et leucémie (Wiltshaw 1979; Khan et al. 1982; Koizumi et al. 2008; Previati et al. 2006; Ajani et al. 2008; Decatris et al. 2004; Tsimberidou et al. 2009; Dhar et al. 2011). Il n'est cependant pas le traitement de prédilection, car n'étant pas pour ces cancers le traitement le plus optimal existant.

Bien que le cisplatine soit une drogue très utilisée en chimiothérapie, ce dernier est aussi connu pour sa grande toxicité.

#### IV.4.1 Toxicité

Bon nombre d'effets secondaires indésirables apparaissent lors du traitement au cisplatine. Une corrélation a pu être établie entre la sévérité des effets secondaires indésirables observée et la dose de cisplatine utilisée (Astolfi et al. 2013), montrant l'importance d'éviter le surdosage. Cependant, un équilibre doit être trouvé entre efficacité du traitement et tolérance à la toxicité du cisplatine par le patient. La palette des effets indésirables du traitement cisplatine est très large allant de symptômes de nausées et de vomissements (Kosmas et al. 2001) à une importante toxicité sur les organes notamment vis-à-vis des reins (Tang et al. 2023).

Lors d'un traitement au cisplatine, plus de 90% du cisplatine est éliminé du corps par excrétion rénale (Tsang et al. 2009). Une accumulation de cisplatine dans les reins peut avoir lieu. Cette accumulation cause des dommages sévères aux reins et provoque des maladies chroniques du rein. À l'échelle des tissus et des cellules tubulaires du rein, cette accumulation provoque des dommages à l'ADN, des pathologies mitochondriales, du stress oxydatif et du stress du réticulum endoplasmique (Tang et al. 2023). Cela provoque par la suite de l'autophagie, de l'apoptose, de la nécrose et inflammation programmée, de l'arrêt de cycle

cellulaire et de la sénescence. Cette toxicité rénale est observée lors d'un traitement cisplatine unique pour 28 à 36% des patients (Ghosh 2019).

D'autres organes peuvent être touchés par des lésions notamment le foie, l'appareil gastrointestinal et l'oreille interne. Dans le cas du foie et de l'oreille interne, cette toxicité est principalement médiée par la génération de ROS (Ghosh 2019). Pour l'oreille interne, cet excès de ROS dans les cellules de la cochlée peut provoquer des pertes auditives permanentes, bilatérales dans les hautes fréquences (McKeage 1995; Sheth et al. 2017). Pour l'appareil gastrointestinal, le cisplatine peut provoquer des diarrhées, perte de goût et de nombreuses inflammations notamment au niveau du pancréas, des muqueuses. Une toxicité touchant un nombre de patients plus faible a aussi été observée notamment au niveau du système cardiovasculaire et nerveux ainsi qu'au niveau des yeux (Ghosh 2019).

Pour éviter ces effets de toxicité, différentes stratégies sont mises en œuvre. Pour contrer les effets dus aux ROS générées par le cisplatine, des antioxydants sont prescrits aux patients. Entre autres, il a été noté qu'une hydratation importante des patients limitait les phénomènes de néphrotoxicité (Tsang et al. 2009; Skinner 1995). Des équipes tentent aussi de synthétiser des analogues basés sur du platine moins toxiques mais tout aussi efficaces pour substituer le cisplatine, une partie de la toxicité du cisplatine reposant sur les doses utilisées. D'autres équipes essayent de combiner le cisplatine avec d'autres agents chimiques afin de diminuer la quantité de cisplatine utilisée tout en gardant la même efficacité thérapeutique (Ghosh 2019).

La toxicité du cisplatine n'est cependant pas le seul problème apparaissant lors des chimiothérapies. En effet, comme mentionné ci-dessus malgré la grande efficacité des premiers traitements cisplatine d'un patient, permettant même une guérison pour certains, il est souvent observé l'apparition de cellules résistantes au cisplatine. Ces cellules moins sensibles aux doses thérapeutiques provoquent la rechute des patients traités et sont très souvent le signe d'un mauvais pronostic de survie sur le long terme (Ghosh 2019).

#### IV.4.2 Résistance

De façon générale, certains types de cancers sont plus résistants à des agents génotoxiques que d'autres. Pour le cisplatine, il a été notamment observé que les cancers des testicules, des ovaires, SCLC, de la tête et du cou y sont sensibles alors que les cancers du côlon et NSCLC y sont plus résistants (Fuertes et al. 2003; Muggia et al. 1993).



La résistance au cisplatine survient aux différentes étapes d'action du cisplatine (Figure 20). Une résistance au cisplatine peut notamment être acquise en limitant la concentration de cisplatine dans la cellule. Pour ce faire, il peut y avoir une diminution de l'import actif du cisplatine par répression des canaux de transport CTR1, le plus souvent associée à un mauvais pronostic. À l'inverse, la résistance peut aussi être due à une augmentation du nombre et de l'activité des pompes responsables de l'export du cisplatine telles que les transporteurs du Cuivre ATP7A et B ou encore le transporteur de cassette liant l'ATP MRP2 et MXR (Song et al. 2022). Globalement, la concentration de cisplatine dans la cellule est corrélée au pronostic à long terme du patient (S.-H. Chen et al. 2019).

Une fois entrés dans la cellule, certains agents génotoxiques peuvent être rendus inactifs grâce à divers processus. Dans le cadre du cisplatine, cela passe par de fortes réactions chimiques avec des molécules chargées négativement ciblées par le cisplatine comme le glutathion réduit et les protéines metallothionein. Dans le cas de la liaison glutathion réduit-cisplatine, cette liaison est notamment catalysée par la glutathion-S-transférase (GST). En liant le cisplatine, ces molécules neutralisent l'effet du cisplatine sur l'ADN, mais aussi son effet sur l'induction de stress oxydatif (Song et al. 2022).

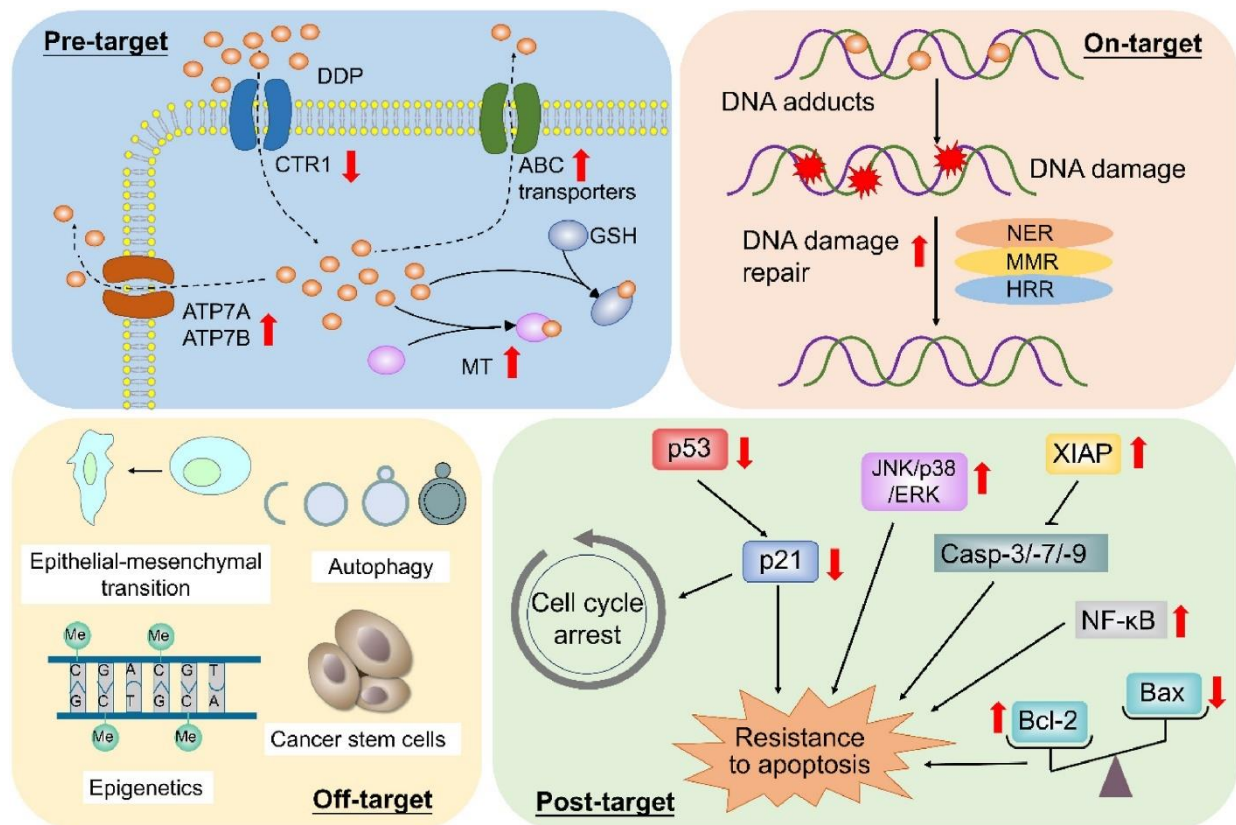


Figure 20 : Mécanismes de résistances au cisplatine (de Song et al. 2022)

Des mécanismes de résistance au cisplatine peuvent intervenir avant l'action du cisplatine sur l'ADN ou des ROS qu'il génère (*Pre-target*), lors des dommages à l'ADN induit par le cisplatine ou les ROS qu'il génère (*On-target*) ou lors du déclenchement de la cascade de signalisation des dommages pouvant aboutir à l'induction de mort cellulaire ou à des arrêts de cycle (*Post-target*). Eventuellement, certains états physiologiques particuliers participent indirectement à la résistance au cisplatine par des changements de processus cellulaires ou de voies de signalisations majeurs. Des marques épigénétiques peuvent aussi participer à ce processus (*Off-target*).

La résistance au cisplatine passe aussi par la capacité de la cellule à réparer rapidement les dommages à l'ADN occasionnés par le cisplatine ou les ROS générés par ce dernier. Comme mentionné précédemment, la grande majorité des dommages cisplatine est réparée par le NER. On observe de ce fait une surexpression de ERCC1, protéine clé du NER, dans les cellules résistantes (S.-H. Chen et al. 2019), augmentant l'activité réparatrice de la voie NER. Il est également connu que le cisplatine est plus efficace sur des cellules déficientes en polymérase TLS (Y. Zhao et al. 2012; Hanaoka et al. 2016), en HR et dans les voies de réparation (Fanconi et TC-NER couplées aux polymérase TLS) des cassures double brin induites notamment par les pontages interbrins du cisplatine, telle que BRCA1 et 2 (S.-H. Chen et al. 2019; Hanaoka et al. 2016; Enoiu et al. 2012). D'autres voies de réparation, telles que le MMR, ont été documentées comme favorisant une résistance au cisplatine. Cette résistance par le MMR passe par des mutations ou répressions notamment vis-à-vis des facteurs en amont de la voie telles que MSH2 et MLH1 (Galluzzi et al. 2012). Pour certains, l'hypothèse admise serait que le MMR tenterait de réparer les pontages cisplatine, mais en y insérant beaucoup d'erreurs. Ces erreurs seraient alors reconnues et déclencheraient une cascade de signalisation pro-apoptotique arrêtant le cycle et tuant les cellules (Galluzzi et al. 2012). Cependant, d'autres études tendent à montrer que le rôle de sensibilisation apportée par les facteurs de MMR ne serait pas lié à leurs rôles sur la réparation en elle-même (Basu et al. 2010). Au niveau des polymérase TLS, l'expression de la polymérase H a par exemple était montrée comme associée à la survie de patients atteints de cancer du poumon. La surexpression de REV3 sous-unité catalytique de la polymérase  $\zeta$  induirait aussi une résistance au cisplatine (Galluzzi et al. 2012). Par ailleurs des cellules acquièrent une résistance face au cisplatine en contrant la génération de ROS. Elles favorisent l'expression de facteurs de transcription telles que Nrf2 médiant le contrôle transcriptionnel de gènes impliqués dans l'homéostasie cellulaire de l'oxydo-réduction (S.-H. Chen et al. 2019).

Comme mentionné précédemment, si la cellule cancéreuse n'arrive pas à réparer à temps certains pontages, des mécanismes de signalisation apoptotique se déclenchent. Dans le cas de cellules résistantes, ces dernières modulent notamment ces cascades de signalisation

responsables de l'induction de l'apoptose ou d'autres formes de mort. Globalement, dans le cadre de la résistance au cisplatine, les cellules résistantes vont favoriser l'expression d'oncogènes comme *c-myc*, *c-abl*, *c-fos* et *H-ras* et réprimer par inhibition, mutation ou diminution de l'expression les gènes suppresseurs de tumeur tels que *TP53* (Ghosh 2019). Une activation de la voie NF- $\kappa$ B est aussi responsable d'une augmentation de la prolifération et une diminution de l'apoptose. Plusieurs facteurs de la voie *MAPK* ont pu aussi être reliés aux phénomènes de résistance sans pour autant conclure sur le rôle pro ou anti-apoptotique (Galluzzi et al. 2012). Des membres de la famille de *BCL2* ainsi que de nombreuses caspases ou ligands de caspases ont pu être montrés comme régulés dans le cadre de la résistance au cisplatine (Galluzzi et al. 2012). Il a aussi été rapporté que des cellules tiraient leur résistance à l'apoptose induite par le cisplatine de certains facteurs de la voie NER. Une diminution d'XPC pourrait empêcher le recrutement d'ATM, bloquant la cascade de signalisation (Colton et al. 2006).

La résistance des cellules cancéreuses aux traitements génotoxiques est déterminée par la mutation ou la régulation d'expression de nombreux gènes impliqués dans les processus mentionnés ci-dessus. Outre les régulations transcriptionnelles, un grand nombre de miARN et circARN agissant sur la stabilité ou la traduction ont été impliqués dans la résistance au cisplatine (F. Li et al. 2023; S. Wang et al. 2020).

D'autres caractéristiques telles que le microenvironnement cellulaire de la tumeur, l'immunogénicité et la topologie de la tumeur (comprenant l'irrigation sanguine et l'accessibilité aux cellules immunitaires) en elle-même ont été montrés comme impliqués dans la résistance (S.-H. Chen et al. 2019). Cependant, ne travaillant que sur des modèles de lignées cellulaires immortalisés ou d'origine tumorale, nous ferons l'impasse sur de telles régulations dans le cadre de mon projet.

Certains mécanismes de résistance passent par des changements majeurs dans les processus cellulaires et voies de signalisations qui perturbent indirectement l'action du cisplatine ou qui bloquent l'induction de mort cellulaire. C'est notamment le cas de l'autophagie, de la transition épithelio-mésenchymale, de la réponse aux chocs thermiques et à certains cancers acquérant des propriétés proches des cellules souches (*stemness*, en anglais) (Song et al. 2022). Des modifications épigénétiques telles que les méthylations M6A ou encore certaines marques d'histones sont aussi connues pour altérer l'expression de gènes impliqués dans la résistance au cisplatine (F. Li et al. 2023). La résistance aux génotoxiques

peut aussi impliquer des régulations de l'épissage alternatif. Dans des modèles de cancer du sein résistant à la doxorubicine, plus d'une quarantaine de facteurs d'épissage et des milliers d'événements d'épissage alternatif ont pu être montrés comme régulés. Deux facteurs d'épissage, ZRANB2 et SYF2, ont été montrés comme contribuant à cette résistance en modulant l'épissage alternatif de gène (*ECT2*) participant à la résistance (Tanaka et al. 2020). Dans le carcinome épidermoïde buccal, il a été montré que le gène *CD44* était épissé différemment dans les cellules résistantes au cisplatine (Qiao et al. 2023). Cette différence d'épissage permet par la suite une activation des cellules dendritiques médiant cette résistance.

## IV.5 Liens entre IPA et DDR

### IV.5.1 Régulation des gènes de DDR au niveau de l'IPA

Au cours des différentes études définissant le rôle de CDK12 et U1 snRNA à l'aide d'inhibiteurs comme répresseurs de l'IPA, les auteurs ont noté un enrichissement en gènes impliqués dans la DDR.

#### IV.5.1.1 Régulation par CDK12

Plusieurs moyens ont été développés pour inhiber ou réprimer l'expression de CDK12, qui agit en complexe avec la cycline K (Dixon-Clarke et al. 2015). Des papiers utilisent notamment des approches avec des siARN/shARN contre CDK12 (Blazek et al. 2011; Tien et al. 2017; Krajewska et al. 2019) ou la cycline K (Tien et al. 2017; Sun et al. 2022). Certains auteurs ont développé à l'aide de système CRISPR-cas9 des cellules présentant une mutation homozygote de CDK12 sensible à des analogues d'ATP l'inhibant lorsque délivrés à la cellule (Chirackal Manavalan et al. 2019). Cela permet d'être le plus spécifique possible à CDK12 tout en gardant les avantages d'utilisation des inhibiteurs. Plusieurs types d'analogues sont utilisés dans la littérature, mais le fonctionnement est similaire (Chirackal Manavalan et al. 2019; Tellier et al. 2020; Fan et al. 2020). Par ailleurs, des études ont réalisé des inactivations par CRISPR des gènes *CDK12* et/ou *CDK13*, qui est très homologue à *CDK12* (Dubbury et al. 2018; Quereda et al. 2019). Une série d'inhibiteurs de CDK12 a été développée dont le plus utilisé est THZ531 (T. Zhang et al. 2016; Krajewska et al. 2019; Fan et al. 2020), qui inhibe aussi CDK13. D'autres auteurs utilisent le dinaciclib mais celui-ci inhibe aussi CDK 1, 2, 5 et 9 (Johnson et al. 2016). Une étude a développé un inhibiteur dual de CDK12 et CDK13 appelé SR-4835 (Quereda

et al. 2019). Du fait de la spécificité variable de ces inhibiteurs, une approche complémentaire pour réprimer CDK12 est souvent réalisée dans ces études.

L'un des premiers papiers montrant l'impact de la répression de CDK12 sur les gènes de DDR utilisait des siARN dans les HEK et les HeLa. À la suite de la déplétion de CDK12 par siARN, une diminution de l'expression de BRCA1, FANCI, FANCD2 et ATR a été observée (Blazek et al. 2011). Dans une autre étude, à la suite d'un traitement avec THZ531, les auteurs montrent une diminution d'expression des gènes *FANCF*, *ERCC4* et *BRCA1* (Paculová et al. 2017; T. Zhang et al. 2016). Lors d'un traitement à la dinaciclib confirmé ensuite par shARN contre CDK12, les auteurs observent une diminution du recrutement de BRCA1 et RAD51 aux cassures doubles brins (Johnson et al. 2016). Plus tard, une équipe a montré que l'inhibition de CDK12, cette fois-ci dans le neuroblastome, réprimait BRCA1, FAN1, RAD51, BLM, PALB2 et EXO1 tous impliqués dans le DDR et particulièrement le HR (Krajewska et al. 2019). Il est à noter que toutes les études sont en accord pour montrer la répression de BRCA1 à la suite de la perte fonctionnelle de CDK12. Une grande partie d'entre elles mentionnent un phénotype dit de « *BRCAness* », c'est-à-dire des effondrements de fourches répliquatives dus à une HR défectueuse (Dubbury et al. 2018; H.-W. Wang 2019; Quereda et al. 2019; Sun et al. 2022). En revanche, chaque étude trouve à chaque fois d'autres gènes différents réprimés dans la voie HR ou dans la voie Fanconi. Cette différence pourrait s'expliquer soit par la spécificité des différentes méthodes utilisées pour réprimer CDK12, soit par une potentielle régulation tissu spécifique du mode d'action de CDK12. Une étude a montré que la perte fonctionnelle de CDK12 était différente d'une perte totale de HR car présentant des marques de large duplication en tandem témoignant d'une activité résiduelle de la HR (Popova et al. 2016).

De façon cohérente avec l'impact de l'inhibition de CDK12 sur divers gènes de DDR (notamment HR), les auteurs ont aussi montré que ces inhibitions pouvaient sensibiliser des cellules cancéreuses aux génotoxiques. Blazek et al. ont trouvé que la déplétion par siARN de CDK12 sensibilisait les cellules à l'étoposide, la mitomycine C et la camptothécine. Montré tout d'abord dans le cancer du sein triple-négatif, l'inhibition de CDK12 permet aussi de resensibiliser les cellules aux PARPi (Johnson et al. 2016). Plus tard, dans des cellules de cancer du sein triple-négatif, une sensibilisation aux PARPi et aux agents génotoxiques (cisplatine, doxorubicine et irinotecan) par CDK12i a pu être validée (Quereda et al. 2019). Cette sensibilisation passe par une meilleure induction de dommages de l'ADN, de l'apoptose et une diminution de la croissance tumorale chez des souris xénotreffées. Une autre étude sur le

cancer de la prostate résistant à la castration, observe qu'un traitement inhibant CDK12 ou la cycline K permettrait aussi de sensibiliser ces cellules aux PARPi (Sun et al. 2022). Ces sensibilisations passent notamment par l'obtention d'une mortalité similaire face à l'agent génotoxique, mais pour des doses de cet agent beaucoup plus modérées. Par ailleurs, il a été montré que les cellules cancéreuses présentant des mutations rendant inactif CDK12 étaient plus sensibles au cisplatine (Joshi et al. 2014).

En plus d'une répression de voie de DDR telle que la HR, un effet de la répression de CDK12 sur la mortalité endogène a pu être noté. En effet, déjà observé par Blazek et al dans les HEK et les Hela en siARN, CDK12i a été remarqué comme provoquant par lui-même la mort des cellules dans des lignées Jurkat (T. Zhang et al. 2016). Cet effet a pu être confirmé dans des cellules de neuroblastome où CDK12i a été montré comme cytotoxique (Krajewska et al. 2019). Une autre étude montre par inhibition endogène de CDK12 et/ou de CDK13 une certaine redondance de ces kinases dans le maintien de la viabilité et de la prolifération cellulaire (Fan et al. 2020). Ces résultats sont à mettre en parallèle avec le fait que CDK12 se retrouve muté et sa mutation associée à un mauvais pronostic dans certains cancers. C'est notamment le cas pour le cancer du sein (Paculová et al. 2017), le carcinome des ovaires (Ekumi et al. 2015) et le cancer de la prostate (Sun et al. 2022). Ces mutations répriment CDK12 et affectent la HR provoquant de la *BRCAness*. Ces effondrements et impossibilités de réparer de façon fidèle une cassure double brin induisent de l'instabilité génétique favorisant la tumorigénèse (Dubbury et al. 2018). Cela permettrait aux cellules cancéreuses d'augmenter leur taux de mutations et de réarrangements chromosomiques (Paculová et al. 2017) explorant ainsi plus de pistes d'échappement thérapeutique. Cependant, cette instabilité est à double tranchant et permet d'être plus sensible aux génotoxiques comme montré ci-dessus (Dubbury et al. 2018). La mutation de CDK12 ou son inhibition a donc un double rôle dans l'oncogénèse et la réponse aux traitements (Paculová et al. 2017).

En tout premier lieu, il a d'abord été observé que l'inhibition de CDK12 diminuait l'expression de gènes de DDR en diminuant l'élongation de la transcription et la processivité. Il a été montré que CDK12 est présent au niveau des gènes actifs et des *enhancers* (T. Zhang et al. 2016) et que THZ531 inhibe l'élongation de la transcription. Une corrélation a pu être observée entre traitement par THZ531 et une inhibition des facteurs de transcription (T. Zhang et al. 2016). Une autre étude sur des cellules Jurkat et HEK en TT-seq et mNET-seq, des séquençages basés sur l'ARN pol II en cours de transcription, mesurant les effets directs de

l'inhibition de CDK12 (par THZ531 et des analogues d'ATP, 1-NM-PP1), a notamment montré que l'inhibition de CDK12 diminue globalement l'élongation de la transcription de l'ARN pol II et réduit la phosphorylation de la sérine 2, mais aussi de la sérine 5 de la queue CTD. Les chercheurs expliquent la perte d'élongation avec une perte totale d'association entre l'ARN pol II et LEO1, CDC73 et partielle avec PAF1 et SPT6, des facteurs d'élongation, en interactome (Tellier et al. 2020). Cependant, une étude basée sur des RNA-seq et ChIP-seq a montré que l'inhibition de CDK12 ne diminuait pas le niveau global de phosphorylation de l'ARN pol II en sérine 2 bien qu'observant une diminution des ARN pol II porteuses de la phosphorylation en sérine 2 dans les gènes où une perte de processivité de l'ARN pol II a été montrée (Chirackal Manavalan et al. 2019). Cette non diminution de la phosphorylation de la sérine 2 observée globalement pourrait cependant être attribuée à un problème de sensibilité du ChIP-seq, l'anticorps utilisé pourrait être saturé sans traitement avec l'inhibiteur. Au niveau transcriptionnel, il a été montré que l'inhibition de CDK12 provoquait des variations d'épissage (Sun et al. 2022; Krajewska et al. 2019) et que CDK12 était de façon générale impliqué dans l'épissage (Paculová et al. 2017). Il a été aussi montré que CDK12 régule l'épissage alternatif du dernier exon notamment dans *ATM* et *DNAJB6*, deux gènes de DDR (Tien et al. 2017). L'état de phosphorylation de la sérine 2 de la queue CTD de l'ARN pol II a par ailleurs pu être relié à la polyadénylation (Hsin et al. 2012). CDK12 a été montré comme médiant cette phosphorylation régulant ainsi le PCPA (Davidson et al. 2014). Globalement, la perte de processivité et l'inhibition d'élongation notées précédemment ont pu être reliées à une augmentation des événements d'IPA et d'utilisation des polyA proximaux (Krajewska et al. 2019; Fan et al. 2020). L'inhibition de CDK12 augmente globalement l'IPA permettant ainsi la diminution de la production d'ARNm pleine longueur, et cela, préférentiellement dans les gènes de DDR (Dubbury et al. 2018) confirmée plus tard par d'autres études (Quereda et al. 2019; Magnuson et al. 2022). CDK12 aurait aussi des effets sur la dégradation des ARN en plus de ses effets sur la terminaison prémature de la transcription et l'épissage (Magnuson et al. 2022). Certaines études ont voulu regarder les cibles de phosphorylation de CDK12. L'une d'elle trouve plus d'une cinquantaine de facteurs régulés dans les 30 min après ajout de CDK12i. Ils sont principalement impliqués dans la transcription et la maturation des ARN. Une autre partie, étonnamment, serait engagée dans des processus d'export nucléaire, de différenciation cellulaire et de mitose (Bartkowiak et al. 2019). L'autre étude montre, elle aussi, une phosphorylation par CDK12 de facteurs de maturation de l'ARN dont U1 snRNP (Krajewska et al. 2019).

Par ailleurs, la répression de la cycline K permettrait aussi d'induire l'IPA au niveau de l'intron 3 du gène codant le récepteur aux androgènes dans le cancer de la prostate résistant à la castration sans que les auteurs aient pu déterminer la fonction de cette isoforme IPA (Sun et al. 2022).

La spécificité de CDK12 pour réguler les gènes de DDR serait due aux caractéristiques de ces gènes. La première étude basée sur des siARN contre CDK12 avait montré que la répression d'expression des gènes de HR était notamment due au fait que ces gènes sont longs avec beaucoup d'introns (Blazek et al. 2011). Cette tendance a pu être confirmée par la suite, trouvant au passage que les gènes régulés au niveau de leurs polyA proximaux par CDK12 contenaient plus de sites polyA (Tien et al. 2017). Une autre étude a montré par la suite que CDK12 inhiberait, de façon dépendante de la longueur, les gènes d'une taille supérieure à 45 kb dont une bonne proportion sont effectivement impliqués dans le DDR. Ce défaut d'élongation serait dû aux caractéristiques propres des gènes possédant plus d'introns, plus longs avec un contenu AT riche présentant donc plus de sites IPA et moins de sites liant U1 snRNA. La longueur du premier intron semblerait jouer aussi un rôle important. De par son influence sur l'élongation de la transcription et la polyadénylation intronique, l'inhibition de CDK12 ciblerait donc préférentiellement les gènes longs comme les gènes de DDR (Krajewska et al. 2019). Toutefois, une autre étude montre que même si les gènes de HR sont enrichis en sites polyA dans les introns, ils sont quand même plus sensibles à la perte de CDK12 que d'autres gènes possédant le même nombre de sites polyA (Dubbury et al. 2018).

En plus de son action sur le DDR, il a été montré que l'inhibition de CDK12 provoquait une inhibition du cycle cellulaire. En effet, par RNA-seq et CHIP-seq en utilisant THZ531 ou des analogues de l'ATP (3-MB-PP1) sur des lignées HCT116 CDK12 sensibles à l'analogue, des chercheurs ont bien observé que les gènes de DDR étaient enrichis dans cette diminution de l'expression, mais qu'en deuxième position figurait le groupe des gènes de réplication de l'ADN. Ils observent qu'à la suite de cette diminution de l'expression de ces gènes, les cellules se bloquent progressivement en phase G1 ne pouvant continuer en phase S (Chirackal Manavalan et al. 2019).

#### IV.5.1.2 Régulation par U1 snRNA

Il a été montré dans des expériences utilisant l'inhibiteur de U1 snRNA, U1-AMO, que certains groupes de gènes semblaient plus sensibles à l'induction d'IPA. En effet, au cours



d'une étude décrivant les fonctions de U1 snRNA notamment vis-à-vis de son rôle d'inducteur d'IPA, les gènes impliqués dans la réplication de l'ADN et les points de contrôle du cycle cellulaire ont été décrits comme subissant préférentiellement des événements d'IPA (Oh et al. 2017). Une autre étude a observé que lorsque U1 snRNP était surexprimée dans des cellules HCT116, une irradiation aux UV-C induisait moins de fragmentations d'ADN (phénotype associé à la mort par apoptose) et moins d'inductions d'IPA dans des gènes testés (Devany et al. 2016), mais le lien potentiel entre les deux effets n'a pas été montré.

On pourrait ici supputer d'un possible impact de U1 snRNP sur la DDR favorisant ainsi une possible sensibilisation aux génotoxiques comme dans le cas des inhibiteurs de CDK12.

#### IV.5.2 Régulation des isoformes IPA par les agents génotoxiques

Les inducteurs d'IPA influencent l'expression des gènes de DDR, mais à l'inverse, il a aussi été montré que des traitements génotoxiques régulent l'expression des isoformes IPA.

Une première étude à l'aide de puces exons a observé une régulation à grande échelle des isoformes ALE en réponse à la camptothécine et à la doxorubicine (inhibiteurs de topoisomérase I et II, respectivement) dans des cellules de cancer du sein (Dutertre, Chakrama, et al. 2014). Les topoisomérases sont une famille de 6 protéines chez l'Homme permettant de désenchevêtrer l'ADN par cassure simple brin-ligation lorsque ce dernier se retrouve trop enroulé dans le cadre de la réplication, transcription, recombinaison et compaction chromatinienne créant des torsions (Ashour et al. 2015). La doxorubicine et la camptothécine répriment les exons ALE proximaux (utilisant un site IPA) par rapport aux exons ALE distaux (dernier exon du gène) dans 80% et environ 50% des cas respectivement. La régulation des exons ALE par la doxorubicine, qui a été confirmée et étendue par des analyses des isoformes IPA par 3'-seq, est enrichie dans les gènes de DDR et de cycle cellulaire et est médiée en partie par une protéine de liaison aux ARN, HuR/ELAVL1 (Dutertre, Chakrama, et al. 2014; Chakraborty et al. 2022).

Deux études, par RNA-seq et 3'-seq dans des lignées cellulaires humaines, ont montré qu'un traitement aux UV-C augmente l'expression des isoformes IPA par rapport aux isoformes pleine-longueur dans de nombreux gènes (Devany et al. 2016; Williamson et al. 2017). Selon les études, cette régulation était médiée par une inhibition de l'élongation et de la processivité de la transcription dans les gènes longs (ce qui diminue la production des isoformes pleine-longueur) ou par une diminution de U1 snRNA, qui est un répresseur de l'IPA (voir ci-dessus).

La diminution d'U1 snRNA par les UV avait été précédemment décrite (Morra et al. 1986). Il est à noter que chez la levure *S. cerevisiae*, en réponse au 4-nitroquinoline 1-oxide (4NQO) qui mime l'exposition aux UV, plus de 2000 gènes ont montré un changement d'utilisation des sites polyA, le plus souvent une diminution des sites polyA intragéniques (à noter qu'il n'y a pas d'intron dans la quasi-totalité des gènes chez *S. cerevisiae*) par rapport au polyA canonique (Graber et al. 2013). Les causes de ce changement ont été attribuées à une inhibition d'expression protéique de la machinerie de clivage/polyadénylation ainsi qu'une différence dans les séquences régulatrices en Cis du site polyA contrôlant ainsi probablement leur activité.

Récemment, mon équipe a observé qu'un autre agent génotoxique, le cisplatine, régule les isoformes IPA à grande échelle dans les cellules humaines, comme nous le verrons dans la partie suivante.

## V. Données préliminaires et Objectifs

Avant mon arrivée dans mon équipe d'accueil, celle-ci avait pu mettre en évidence, par des analyses de 3'-seq dans des lignées cellulaires de cancer du poumon NSCLC (H358 et A549), qu'en réponse au cisplatine les cellules favorisaient l'expression des isoformes IPA par rapport aux transcrits pleine-longueur (ratio IPA :LE) dans un grand nombre de gènes (voir l'article 1 ci-dessous). De plus, le statut traductionnel des isoformes IPA étant globalement peu connu, mon équipe avait étudié l'engagement des isoformes IPA dans les polysomes, en séquençant par 3'-seq les ARN de fractions polysomes lourds (enrichies en ARNm activement traduits) et cytosol total (pour comparaison). Cette analyse a révélé un groupe d'isoformes IPA qui étaient très peu abondantes dans les polysomes lourds par rapport au cytosol total (et par rapport à l'isoforme LE du gène correspondant). Ces isoformes IPA étaient enrichies dans la partie 5' des gènes (5'IPA). Comme discuté plus haut, ce type d'isoforme IPA était très peu caractérisée et considérée soit comme dégradée dans le noyau (Ogami et al. 2017; Chiu et al. 2018) soit comme étant des ARN longs non-codants (Singh et al. 2018) mais un seul d'entre eux avait une fonction montrée (Williamson et al. 2017).

Au vu de ces données, le premier objectif de ma thèse a été de déterminer si les isoformes 5'IPA induites par le cisplatine ont un rôle dans la réponse cellulaire à ce génotoxique, rôle qui pourrait être dû soit à une fonction non-codante soit au codage de microprotéines.

Par ailleurs, au vu des liens entre IPA et DDR, en particulier au vu des données suggérant que U1 snRNA régule l'IPA (PCPA) dans des gènes de DDR et de cycle cellulaire (Oh et al. 2017) et impacte l'apoptose induite par les radiations UV-C (Devany et al. 2016), mon deuxième objectif a été de déterminer si le blocage de U1 snRNA avec U1-AMO pourrait permettre de sensibiliser les cellules NSCLC au cisplatine, via une induction de l'IPA.

Globalement, ces deux objectifs permettraient de mieux comprendre le rôle et l'impact de l'IPA dans la réponse des cellules cancéreuses au cisplatine, et, pour la première partie, de mieux connaître les fonctions potentielles des isoformes IPA produites dans la partie 5' des gènes.

## VI. Articles

### VI.1 Les isoformes de polyadénylation intronique dans la partie 5' des gènes constituent une source de microprotéines et sont impliquées dans la réponse cellulaire au cisplatine

#### VI.1.1 Introduction

Avant mon arrivée au laboratoire, mon équipe avait démarré une analyse à grande échelle de la régulation et de la traduction des isoformes IPA en réponse au cisplatine dans des lignées de cancer du poumon NSCLC (Figures 1 et 2). Précédemment d'autres génotoxiques comme les UV-C ou la doxorubicine avait été montrés comme régulant l'IPA, mais la traduction des isoformes IPA était peu caractérisée globalement. Je me suis notamment intéressé aux isoformes 5'IPA, qui sont générées dans la partie 5' des gènes et qui sont souvent considérées comme non fonctionnelles et avec un faible potentiel codant. J'ai contribué à l'étude de la fonction de ces isoformes 5'IPA au regard de la réponse au cisplatine, ainsi que leur traduction en microprotéines, classe émergente de petites protéines (moins de 100 acides aminés)(Figures 3 à 7). Les analyses bioinformatiques des séquençages ont été réalisées par des bioinformaticiens mais j'ai pu participer au premier essai de combinaison des bases de données de Ribo-seq et de Spectrométrie de Masse. Une ingénieure m'a aidé à l'obtention des résultats présentés en Figure 7. Cette étude a été déposée dans bioRxiv et est en évaluation à *review commons* (une plateforme d'EMBO press) depuis le 5 décembre 2023.

**Intronic polyadenylation isoforms in the 5' part of genes constitute a source of microproteins and are involved in cell response to cisplatin**

Alexandre DEVAUX<sup>1-3§</sup>, Iris TANAKA<sup>1-3§</sup>, Mandy CADIX<sup>1-4</sup>, Amélie HENEMAN-MASUREL<sup>1-3</sup>, Sophie MICHALLET<sup>5</sup>, Quentin FOUILLEUL<sup>1-3</sup>, Alina CHAKRABORTY<sup>1-3</sup>, Céline M. LABBE<sup>1-3</sup>, Nicolas FONTRONDA<sup>6</sup>, Jean-Baptiste CLAUDE<sup>6</sup>, Marc DELOGER<sup>4</sup>, Pierre GESTRAUD<sup>4</sup>, Ludovic TESSIER<sup>1-3</sup>, Hussein MORTADA<sup>6</sup>, Sonia LAMEIRAS<sup>7</sup>, Virginie RAYNAL<sup>7,8</sup>, Sylvain BAULANDE<sup>7</sup>, Nicolas SERVANT<sup>4</sup>, Didier AUBOEUF<sup>6</sup>, Béatrice EYMIN<sup>5</sup>, Stéphan VAGNER<sup>1-3</sup>, Martin DUTERTRE<sup>1-3\*</sup>

<sup>1</sup>Institut Curie, Université PSL, CNRS UMR3348, INSERM U1278, 91400 Orsay, France

<sup>2</sup>Université Paris-Saclay, CNRS UMR3348, INSERM U1278, 91400 Orsay, France

<sup>3</sup>Equipe Labellisée Ligue Nationale Contre le Cancer, INSERM U1278, 91400 Orsay, France

<sup>4</sup>INSERM U900, Mines Paris Tech, Institut Curie, 75000 Paris, France

<sup>5</sup>INSERM U1209, CNRS UMR5309, Université Grenoble Alpes, Institute for Advanced Biosciences, 38700 La Tronche, France

<sup>6</sup>CNRS UMR 5239, Ecole Normale Supérieure de Lyon, 69007 Lyon, France

<sup>7</sup>ICGex NGS platform, Institut Curie, 75000 Paris, France

<sup>8</sup>INSERM U830, Institut Curie, 75000 Paris, France

<sup>§</sup>These authors equally contributed to this work and should be considered as co-first authors.

\*Correspondence: martin.dutertre@curie.fr

## ABSTRACT

Transcript isoforms generated by intronic polyadenylation (IPA) are widely regulated in various biological processes and often encode protein isoforms. Microproteins are small proteins translated from small open reading frames (sORFs) in noncoding RNAs and mRNAs, but their production by IPA isoforms is unknown. Using 3'-seq and long-read RNA-seq analyses in lung cancer cells, we show that cisplatin, a DNA-crosslinking anticancer agent, upregulates IPA isoforms relative to full-length mRNAs in long genes. A subset of cisplatin-regulated IPA isoforms are poorly associated with heavy polysomes and terminate upstream of the annotated translation initiation codon of genes. Such IPA isoforms in the *PHF20* and *PRKAR1B* genes are associated with light polysomes, contain Ribo-Seq-supported sORFs in an alternative last exon within the annotated 5'UTR part of genes, and are translated into microproteins. For *PRKAR1B*, the microprotein was detected by Western blot and immunofluorescence after transfection of a tagged isoform; and siRNA depletion of the endogenous IPA isoform, CRISPR deletion of the IPA site, or CRISPR mutation of the sORF initiation codon led to increased cell survival to cisplatin. Based on Ribo-Seq and mass-spectrometry data sets, we identified 156 genes producing both a canonical protein-coding mRNA and a microprotein-coding 5'UTR-located IPA isoform (coined miP-5'UTR-IPA isoform) regulated by cisplatin. Finally, the regulation of (miP-5'UTR-)IPA *versus* full-length isoforms by cisplatin involved an inhibition of transcription processivity in a FANCD2 and senataxin-dependent manner. Altogether, these findings reveal the novel paradigm of miP-5'UTR-IPA genes and their role in cancer cell response to a genotoxic agent.

## KEYWORDS

Intronic polyadenylation – Alternative last exon – DNA-damaging agent – Small ORF – Micropeptide

## HIGHLIGHTS

- Cisplatin increases intronic-polyadenylation *versus* full-length transcript isoforms in long genes through a FANCD2 and senataxin-dependent decrease of transcription processivity
- A subset of cisplatin-regulated intronic-polyadenylation isoforms terminate in the annotated 5'UTR part of genes and encode microproteins, thus we coined them miP-5'UTR-IPA isoforms
- The miP-5'UTR-IPA isoform of *PRKAR1B* impacts cisplatin sensitivity and its effect is mediated by its small ORF
- We identify 156 genes producing both a canonical protein-coding mRNA and a microprotein-coding miP-5'UTR-IPA transcript

## INTRODUCTION

Intronic polyadenylation (IPA), sometimes referred to as splicing-dependent, upstream region- or coding region- alternative polyadenylation (APA), corresponds to the use of a polyadenylation site (polyA site) located upstream of the last exon (LE) of a gene and allows the production of transcripts with alternative last exons (ALEs) (1–4). Genome-wide analyses using 3'-seq (RNA-seq focused on the 3' end of polyadenylated transcripts) and other approaches have revealed regulation of the IPA:LE isoform ratio in many human genes across tissues and in various biological processes, such as cell differentiation, proliferation, and cell responses to DNA-damaging agents and other stress inducers, with different stressors inducing different global patterns of IPA:LE isoform regulation (5–13). The mechanisms of IPA regulation have been intensively studied, with many regulatory factors acting at the levels of splicing, polyadenylation, and transcription elongation and processivity (1–4).

In contrast, relatively little is known about the translation outcome of IPA isoforms. On the one hand, dozens of IPA isoforms were shown to encode protein isoforms (3) and a recent genome-wide analysis showed that the majority of cytosolic IPA isoforms are associated with polysomes, suggesting they are translated (11). On the other hand, some studies identified two IPA isoforms with noncoding RNA functions (8, 14), IPA isoforms degraded by the nuclear exosome (15, 16), and abundant IPA isoforms with a low predicted coding potential that were therefore classified as long noncoding RNAs (lncRNAs), but their function and translation status were not studied (17). Microproteins, also called micropeptides or sORF-encoded peptides, are an emerging class of proteins of less than 100 amino acids that are encoded by small open reading frames (sORFs) found in lncRNAs, canonical mRNAs (either upstream, downstream, overlapping or inside the annotated canonical ORF), and circular RNAs (18, 19). An increasing number of human microproteins are involved in biological processes (18, 20, 21). IPA isoforms constitute a potential but unexplored source of microprotein production.

Cisplatin is a DNA-crosslinking anticancer agent that is widely used in non-small cell lung cancer (NSCLC) treatment (22). It widely regulates alternative splicing (23, 24) but little is known about its effects on IPA isoforms. In this study, by investigating on IPA isoform regulation by cisplatin in NSCLC cells, we identified cisplatin-regulated IPA isoforms that terminate within the annotated 5'UTR part of genes, encode microproteins, and impact cell response to cisplatin. These findings reveal the existence of genes (which we call miP-5'UTR-IPA genes) producing both a canonical protein-coding mRNA and a microprotein-coding IPA transcript.

## MATERIAL AND METHODS

### Plasmid construction

*PHF20* sORF#1, *PRKAR1B* sORF#1 and *PRKAR1B* sORF#2, together with the ALE sequence upstream of the sORF and with a carboxy-terminal Flag tag, were obtained by RT-PCR (using primers described in Table S1) and cloned into the pTRIP $\Delta$ U3-MND-IRES-GFP plasmid (25) between the MND promoter and an internal ribosome entry site driving the expression of green fluorescent protein (GFP).

### Cell culture, transfection, and treatment

Cells were cultured in RPMI-1640-Glutamax (H358) or DMEM (A549, HeLa, HEK-293T) medium (GibcoBRL, Life Technologies, Cergy Pontoise, France) supplemented with 10% (v/v) heat-inactivated fetal calf serum (GibcoBRL), in 5% CO<sub>2</sub> at 37°C. All cell lines and CRISPR clones were authenticated by STR analysis. Cisplatin was obtained from Selleckchem (Euromedex, Souffelweyersheim, France) and resuspended in DMSO. Unless otherwise stated, cisplatin was used at 25 and 100 µM in A549 and H358 cells, respectively. Cell treatment with vehicle (DMSO) was always done in parallel to cisplatin treatment. Reverse transfection of siRNAs (Eurogentec, Belgium; Table S2) was performed using Lipofectamine RNAiMAX (ThermoFischer Scientific, France) and following the manufacturer's instructions. For plasmid transfection, cells were seeded in 6-well plates containing Marienfeld Superior cover glasses; the next day, cells were transfected for 24 hours with 5 µg of plasmid, 7.5 µL of Lipofectamine 3000 and 10 µL of P3000 (ThermoFischer Scientific, France) in the case of HeLa cells, or with 4 µg of plasmid and 9 µL of Lipofectamine 2000 (ThermoFischer Scientific, France) in the case of HEK-293T cells.

### **WST1 and FACS analysis**

For cell growth and viability analysis, cells were seeded in 96-well plates. Following cisplatin treatment, cell viability was assayed in triplicate wells using WST1 (Sigma Aldrich, France) according to the manufacturer's instructions. For FACS analysis of DNA content, cells were fixed with 70% cold ethanol for 30 min on ice, treated with RNase A (20 µg/ml) for 20 min and stained with propidium iodide (10 µg/ml). Flow cytometric analysis of 10000 cells was performed on a FACScan flow cytometer (BD Biosciences) and data were recovered using the CellQuest software (BD Biosciences).

### **Protein extraction and Western blot**

For immunoblotting, cells washed three times in PBS were lysed in RIPA buffer (150mM NaCl, 50mM Tris HCl pH 8, 0.1% SDS, 1% Nonidet P40, 0.5% Na deoxycholate, 0.1mM PMSF, 2.5µg/ml pepstatin, 10µg/ml aprotinin, 5µg/ml leupeptin, 0.2mM Na<sub>3</sub>VO<sub>4</sub>) for 30 min on ice and pelleted. Protein concentration was determined using the Pierce BCA Protein Assay kit (Biorad). Proteins were then separated in 10-12% SDS-PAGE gels (or Tris-Tricine gel 16% gels for microproteins) and electroblotted onto PVDF membranes. Membranes were incubated overnight at +4°C with antibodies against PRIM2 (PA5-48859; ThermoFisher Scientific), PRKAR1B (PA5-55392, ThermoFischer Scientific), actin (Sigma-Aldrich, Lyon, France), GAPDH (G8795, Merck) or Flag (F1804, Sigma Aldrich), then for one hour with a horse-radish peroxidase-conjugated goat anti-mouse or anti-rabbit antibody (Jackson ImmunoResearch Laboratories, West Grove, PA, USA). After washing, blots were revealed using the ECL chemiluminescence method (Amersham, Les Ulis, France), according to the manufacturer's protocol.

### **Immunofluorescence analysis**

Transfected cells grown on Marienfeld Superior cover glasses were washed twice with ice-cold PBS, fixed using 4% paraformaldehyde in PBS during 20 min, washed three times with PBS, permeabilized for 10 min with PBS-0.1% Triton-X, washed twice with PBS, blocked for 5 min with PBS-5% BSA, incubated with Anti-Flag M2 antibody (F1804, Sigma Aldrich) at RT for 1 hour, washed twice with PBS, blocked again, incubated with F(ab')<sub>2</sub>-rabbit anti-mouse IgG (H+L) cross-adsorbed secondary antibody, Alexa Fluor 594 (A-21205, Invitrogen) in PBS-5% BSA for 1 hour, washed twice with PBS, incubated at RT with PBS containing 4',6-diamidino-2-phenylindole (DAPI) 0.1 µg/mL, and washed twice with PBS. Cover glasses were mounted on slides in PBS, glycerol 15%, 1,4-diazabicyclo-(2.2.2) octane (DABCO, Sigma)



100 mg/ml. For microscopy, acquisition was done with an exposure time of 50 ms for Trans-DIC, 20 ms for DAPI 405, 200 ms for FITC and 600 ms for Tx2. Z stacking was used for imaging DAPI, FITC and Tx2 with a step of 0,3µm, ±3 µm around the acquisition point.

### **Preparation of cytosol and polysome fractions**

Polysome profiling was performed as described previously (26). Briefly, cells treated or not with CisPt were incubated at 37 °C with 100 µg/mL cycloheximide in fresh medium for 5 min. Cells were then washed, scraped into ice-cold PBS supplemented with 100 µg/mL cycloheximide, centrifuged at 3000 r.p.m. for 5 min. The cell pellets were resuspended into 400 µL of LSB buffer (20 mM Tris, pH 7.4, 100 mM NaCl, 3 mM MgCl<sub>2</sub>, 0.5 M sucrose, 1 mM DTT, 100 U/mL RNasin and 100 µg/mL cycloheximide). After homogenization, 400 µL LSB buffer supplemented with 0.2 % Triton X-100 and 0.25 M sucrose was added. Samples were stayed on ice for 30 min and centrifuged at 12,000 g for 15 min at 4 °C to pellet nuclei. The supernatant (cytosolic extract) was adjusted to 5 M NaCl and 1 M MgCl<sub>2</sub>. The lysates were then loaded onto a 5-50% sucrose density gradient and centrifuged in an SW41 Ti rotor (Beckman) at 36,000 rpm for 2 h at 4 °C. Fractions were monitored and collected using a gradient fractionation system (Isco). RNA was extracted from the four heaviest polysomal fractions (pooled) for 3'-seq analysis or from each fraction for RT-qPCR analysis.

### **RNA extraction**

RNA from whole cells or fractionated cell lysates were extracted with TRIzol Reagent (TRIzol-LS for polysomes; ThermoFisher Scientific) according to the manufacturer's instructions, and 1 µl of GlycoBlue (ThermoFisher Scientific) was added for RNA precipitation. RNA from whole cells was treated with DNase I (TURBO DNA-free, ThermoFisher Scientific). RNA samples were quantified using a Nanodrop 2000 spectrophotometer (ThermoFisher Scientific). For sequencing, RNA samples were analyzed using an RNA 2100 Bioanalyzer (Agilent).

### **RT-(q)PCR**

Reverse transcription was performed on RNA using SuperScript III Reverse Transcriptase (ThermoFisher Scientific) and random primers. PCR was performed using GoTaq Flexi DNA Polymerase (Promega), and PCR products were migrated on agarose gels. Quantitative PCR (qPCR) was performed using Power SYBR Green PCR Master Mix (ThermoFisher Scientific) on a CFX96 Real-Time PCR Detection System (BioRad).

### **3'-seq experiments and bioinformatic analysis**

5 µg of RNA were used for poly(A)<sup>+</sup> RNA purification with Dynabeads mRNA DIRECT Micro kit (ThermoFisher Scientific). Poly(A)<sup>+</sup> RNA was fragmented at 70°C for 5 min in RNA Fragmentation Reagents (ThermoFisher Scientific). After ethanol precipitation, RNA was reverse transcribed using anchored oligo-dT and SMARTScribe RT enzyme (Clontech) allowing to add adapter sequences to the 5' and 3' ends of RNA fragments. Libraries were amplified by 16 cycles of PCR with GoTaq Flexi DNA Polymerase (Promega), purified using Agencourt AMPure XP (Beckman Coulter) magnetic beads, and quantified with Quant-iT Picogreen dsDNA kit (ThermoFisher Scientific). A size selection step was done using SPRIselect (Beckman Coulter) magnetic beads to obtain fragments of 150–300 bp. Purified libraries were controlled by capillary electrophoresis (LabChip - Perkin) and quantitated by qPCR (KAPA

Library Quantification Kits Illumina Platforms - Roche). Pooled libraries (12 pM) with 30% of phiX were subjected to single-end, 50 bp sequencing using the HiSeq 2500 machine (Illumina). Read 1 was read with primer HP6 (Illumina) with 3 dark cycles (first 3 bases of read 1 were not read). Index i7 (6 pb barcode) was read with primer HP8 (Illumina).

Raw reads were trimmed in their 5' and 3' ends to remove uninformative nucleotides due to primer sequences, nucleotides added by SMART Scribe RT enzyme and polyA tail of mRNAs. Trimmed reads of 25 bp or more were aligned on the human reference genome (hg19) using Bowtie2 (version 2.2.5) (27). Only reads with a mapping quality score (MAPQ) of 20 or more were retained (Samtools version 1.1) for downstream analysis. Reads were then clustered along the genome using Bedtools (version 2.17.0) (28), allowing a maximum distance of 170 bp (max library fragment length (300bp) – min fragment length (60bp) – read length (50bp) – oligodT length (25bp)) and a minimum number of 5 reads per peak. Peaks with a stretch of 6 consecutive As (or 8 As out of 9 nucleotides) within 150 bp downstream were filtered out, as they are likely due to internal priming of oligo-dT. Overlapping peaks from all analyzed samples were merged to define a common set of genomic windows corresponding to polyA sites. To annotate peak location within genes, gene coordinates were obtained on the basis of overlapping Refseq transcripts with the same gene symbol. Peaks overlapping any intronic region of a gene were classified as intronic polyA (IPA) peaks. Peaks overlapping the last exon of a gene were classified as LE peaks. Differential analyses between two conditions were done using three independent biological replicates per condition. To compare the regulation of each IPA to the regulation of the gene's last exon (taken as the sum of the peaks in this exon), we used DESeq2 (version 1.4.5) (29) and the following statistical model:

$$Y_{ij} = \mu + L_i + C_j + (LC)_{ij} + E_{ij}$$

where  $Y_{ij}$  is the normalized counts of peak  $i$  in biological condition  $j$ ,  $\mu$  is the mean,  $L_i$  is the peak localization (IPA or LE),  $C_j$  is the biological condition,  $(LC)_{ij}$  is the interaction between peak localization and biological condition, and  $E_{ij}$  is the residual. P-values and adjusted P-values (Benjamini-Hochberg) were calculated. For H358 cells, data using a false discovery rate (FDR) of 10% are shown. For A549 cells, data with  $p < 0.05$  are shown, to avoid under-estimations of cross-comparisons between lists. The complete bioinformatics pipeline (3'-SMART package) described above can be freely downloaded at GitHub (<https://github.com/InstitutCurie/3-SMART>) and can be run through a configuration file and a simple command line. Annotated polyadenylation sites were retrieved from the PolyA\_DB 3 and PolyASite 2.0 databases (30, 31).

### **Long-read RNA-seq (Iso-seq) and bioinformatic analysis**

A starting amount of 1  $\mu$ g was used for long read library preparation. The first step of cDNA synthesis was done using the SMARTer PCR cDNA Synthesis Kit (Takara). Then, cDNA was amplified by 14 PCR cycles and divided in two fractions to be purified with different AMPure PB beads (Pacific Biosciences) ratio. In order to enrich for larger transcripts, a ratio of 0.4x was applied to a fraction (corresponding to 5 out of 8 of the initial volume), while the other fraction was purified with 1x ratio. For each sample, both fractions were quantified with Qubit dsDNA high-sensitivity kit (ThermoFisher Scientific), qualified by capillary electrophoresis (Bioanalyzer High Sensitivity DNA kit - Agilent), pooled equimolarly, and subjected to the SMRTbell Template Prep Kit 1.0 (Pacific Biosciences). The first DNA damage and end repair steps were applied according to the manufacturers' recommendations. The following ligation was

extended to 16 hours to improve efficiency. The final exonuclease treatment was applied to digest unwanted molecules. The SMRTbells were quantified with Qubit dsDNA high-sensitivity kit (ThermoFisher Scientific) and qualified by capillary electrophoresis as above. Complexes were prepared following the manufacturer's recommendations using the Sequel Binding Kit 3.0 and the sequencing Primer v3 (Pacific Biosciences). Each sample was sequenced on 4 SMRTcells 1M, on the Sequel I system (Pacific Biosciences) using a loading molarity of 5 pM. For each SMRTcell, 10 hours of movie, 2 hours of immobilization, and 4 hours of pre-extension were set up. For bioinformatic analysis, subreads coming from PacBio SEQUEL SMRTcells were processed via the SMRTLink's IsoSeq v3.0 pipeline with default parameters (<https://github.com/PacificBiosciences/IsoSeq/blob/master/isoseq-clustering.md>). This pipeline is composed of the following 6 steps: Circular Consensus Sequence (CCS) calling, primer removal/demultiplexing, refining, merging, clustering and polishing. Resulting CCS reads and HQ transcripts were mapped with minimap2 (v2.17-r941) against hg19 genome with the following parameters: -ax splice -uf --secondary=no -C5 -O6,24 -B4. Then, CCS reads were classified into 'last exon (LE)' and 'intronic polyadenylation (IPA)' isoforms, depending on whether the 3'-end of the read was located in the last exon of a gene or within a gene but upstream of the last exon.

### **Total-RNA-seq and analysis of transcription processivity**

For total-RNA-seq, total RNA from whole H358 cells treated or not with cisplatin (2 biological replicates of each condition) was subjected to DNase I treatment. 500 ng of good quality RNA (RIN > 9) were used for Illumina compatible library preparation using the TruSeq Stranded total RNA protocol allowing to take into account strand information. A first step of ribosomal RNA depletion was performed using the RiboZero Gold kit (Illumina). After fragmentation, cDNA synthesis was performed and resulting fragments were used for dA-tailing followed by ligation of TruSeq indexed adapters. PCR amplification was finally achieved to generate the final barcoded cDNA libraries. Libraries were equimolarly pooled and subjected to qPCR quantification using the KAPA library quantification kit (Roche). Sequencing was carried out on the NovaSeq 6000 instrument from Illumina based on a 2\*100 cycle mode (paired-end reads, 100 bases) using a S1 flow cell in order to obtain around 100 million clusters (200 million raw paired-end reads) per sample. Fastq files were generated from raw sequencing data using bcl2fastq where demultiplexing was performed according to barcodes. For each gene, the number of reads in the last intron was divided by the number of reads in the first intron. Genes with at least 20% reduction of this ratio in cisplatin *versus* vehicle conditions ( $p < 0.05$ ) were considered to have their transcription processivity inhibited. A similar method was used to analyze intronic probes in exon-arrays.

### **Other bioinformatic and statistical analyses**

Intersection between lists of 3'-seq peaks were done using Bedtools (version 2.17.0) (28) allowing a gap of 300 bp (Fig. 2C). Functional gene annotation analyses were done using the DAVID software (32, 33), using the human genome as a reference. Genomic coordinates of sORFs were obtained from the sORFs.org (using a FLOSS classification of 'good') and OpenProt databases (34, 35) and were crossed with the genomic coordinates of cisplatin-regulated 5'UTR-IPA (from gene start to IPA site). For each experimental analysis, at least three independent experiments were performed. In bar charts, error bars represent the standard error of the mean (SEM) that is the standard deviation divided by the square root of sample number. Student's paired t-tests were considered significant if  $p < 0.05$ .

## CRISPR-Cas9 edited clones

Single guide RNAs (sgRNAs) and a donor block oligonucleotide (for the mATG design) were designed using Integrated DNA technologies [IDT] design tool and those with the best on and off-target scores were selected (Table S3). PAM sequences were mutated on the donor block oligo to avoid recutting. For cell electroporation, 3.33 µg of Cas9 HiFi (IDT) in 0.5 µL of buffer R (Neon transfection Kit™, ThermoFisher) was mixed in a final volume of 1 µL with 22 picomol of sgRNA and incubated 10 to 20 min at room temperature. 10<sup>5</sup> A549 cells were resuspended in 9 µL of Buffer R and added. 21.9 picomol of electroporation enhancer (IDT) and 22 picomol of donor block oligo (in the case of mATG) were added to the mix for a final volume of 11.5 µL. Electroporation was performed on a Neon transfectoator (ThermoFisher) at 1400 V with 4 pulses of 15 ms using Neon transfection kit 10 µL. Cells were seeded into 96 well plates. For mATG, media contained HDR enhancer V2 (IDT) at 1.34 µM and were changed 1 day after transfection. For cell cloning, 100 µL of cell suspension at 0.8 cell per 100 µL of medium (half fresh medium, half cell-conditioned medium filtered at 22 µm) were seeded in 96 well plates. For genomic DNA extraction, 25 µL of a trypsinized confluent cell plate was mixed with 25 µL of lysis buffer (20 mM Tris HCl pH 8.3, 100 mM KCl, 5 mM MgCl<sub>2</sub>, 1% NP-40, 1% Tween-20, 1:50 proteinase K). After 10 min at 65°C and 10 min at 95°C, high fidelity PCR was performed using Phusion enzyme in 20 µL final volume. Then, PCR products underwent either T7E1 enzymatic assay, restriction enzyme assay, or DNA cleaning for sequencing. For T7E1 enzymatic assay, 5 µL of the PCR product was dehybridized for 10 min at 95°C, then rehybridized by decreasing temperature to 85°C at 2°C/sec and to 25°C at 0.3°C/sec, then 1 U of T7 endonuclease 1 (New England Biolabs) and NEB Buffer 2 were added, samples were incubated for 1 hour at 37°C and analysed by electrophoresis on 1.2% agarose gel.

## RESULTS

### Cisplatin upregulates the IPA:LE isoform ratio in many genes

To analyze the impact of cisplatin on IPA isoform regulation, total RNA from the H358 human NSCLC cell line treated for 24 hours with either cisplatin or vehicle were analyzed by 3'-seq, that relies on the sequencing of regions preceding the 3'-terminal polyA tail of transcripts (36). Because 3'-seq analysis can be flawed by internal priming of the oligo(dT) at genomic stretches of A's, peaks ending near a genomic stretch of A's were discarded. We also verified that most of the selected peaks contained a potential polyadenylation signal (AAUAAA-like motif; Fig. S1A). For each peak overlapping an annotated intron, we analyzed the effect of cisplatin on its expression relative to the last exon (LE) of the cognate gene. With a false discovery rate (FDR) of 10%, 2963 intronic peaks in 1987 genes were regulated by cisplatin at 24 hours of treatment (Fig. 1A, 'all'; Table S4). The vast majority of them (89%) were detected at a level of at least 5% of the LE of the gene (Fig. 1A, 'abundant'). Because internal priming artefacts can be difficult to filter out bioinformatically, we also crossed our list of regulated intronic peaks with lists of already published polyA sites identified in normal tissues (30, 31). Among the 2,963 regulated intronic peaks, 1197 (40%) matched an annotated polyA site (Fig. 1A, 'annotated'). Also, we found a similar percentage of annotated polyA sites in our subsequent 3'-seq analyses on cytosolic fractions (see below), which are unlikely to be subject to internal priming in introns. These data suggest that many of the

cisplatin-regulated intronic peaks that we identified may correspond to genuine polyA sites, that were not annotated in normal tissues. We therefore decided to keep all regulated peaks for subsequent analyses.

Strikingly, about 95% of IPA regulation events by cisplatin corresponded to an upregulation of the IPA:LE isoform ratio (Fig. 1A). This was in sharp contrast with the previously described effect of doxorubicin, which repressed the IPA:LE ratio in the majority of cases (7, 11), but was reminiscent of the effect of ultraviolet C (UV-C) irradiation (8, 9). Out of 84 IPA isoforms that were found to be upregulated by UV-C by RNA-seq in a previous study (8), 34 (40%) were also upregulated by cisplatin (Fig. S1B).

Likewise, 3'-seq analysis on another human NSCLC cell line (A549 cells) treated for 16 hours with cisplatin or with vehicle showed large-scale upregulation of IPA:LE isoform ratio (Fig. 1B and Table S7). 52% of the upregulation events in A549 cells were also found in H358 cells (Fig. S1C). Thus, large-scale up-regulation of IPA:LE isoform ratio is a robust effect of cisplatin in two different NSCLC cell lines. Analysis of gene functional annotation indicated that genes with cisplatin upregulation of the IPA:LE ratio were enriched in functions related to the DNA damage response, cell cycle and cell death (Fig. 1C), which are known to be impacted by cisplatin. Indeed, cisplatin induced cell cycle arrest and cell death in a time and dose-dependent manner (Fig. S1D).

An example of regulated gene is *PRIM2* (DNA Primase Subunit 2), which encodes a regulatory subunit of DNA primase, a DNA replication enzyme. In *PRIM2* intron 7, our 3'-seq data identified an IPA site, which is annotated in PolyA\_DB\_3 (chr6:57299993:+) and the expression of which was increased by cisplatin relative to the LE of the gene (exon 14 ; Fig. 1D). Although no RefSeq transcript matched this IPA site, long-read RNA-seq data that we generated in cisplatin-treated H358 cells showed the existence of a spliced mRNA containing exon 7 and an ALE within intron 7, matching the regulated IPA site (Fig. 1D). This IPA isoform was confirmed by RT-PCR analysis using primers located in exon 7 and the ALE (Fig. 1E-F). RT-qPCR analysis in both H358 and A549 cells validated the upregulation by cisplatin of the IPA:LE ratio of *PRIM2*, which was mainly due to a decrease of the LE isoform, while IPA isoform levels were maintained (Fig. 1G). This regulation by cisplatin was dose-dependent (Fig. S1E), led to decreased *PRIM2* protein levels (Fig. S1F), and preceded cell death induction (Fig. S1D-E). The upregulation of IPA:LE isoform ratio by cisplatin was validated for additional genes by RT-PCR (Fig. S1G). In some genes (*ZFC3H1*, *HERC4*), the IPA isoform was readily increased by cisplatin, but in most cases, the regulation was merely due to a decrease of the LE isoform (Fig. S1G), as noted above for *PRIM2* (Fig. 1G). Finally, the IPA:LE ratio was not upregulated by cisplatin in the *HOMEZ* gene, used as a control (Fig. S1H).

To further assess the impact of cisplatin on IPA versus LE isoform regulation, we used single-molecule long-read RNA-seq (Iso-Seq) in H358 cells treated or not with cisplatin. We obtained more than one million transcripts in each condition (Fig. S1I). For each detected transcript, we determined whether it terminated within the LE of a gene (LE transcript) or upstream of the LE (IPA transcript). Cisplatin globally increased the proportion of IPA compared to LE transcripts by 20% (going from 12,7% to 15,3%; Fig. 1H); this represents a substantial increase when considering that the IPA:LE isoform ratio was upregulated in only 1875 genes in our 3'-seq data. Furthermore, there were 2.53 times less LE transcripts that were detected in cisplatin-treated cells only, when compared to vehicle-treated cells only; meanwhile, there were 1.22 times more IPA transcripts (Fig. 1I). Consistently, the fraction of genes detected at the IPA but not LE level was increased by 2.08-fold by cisplatin treatment, going from 5.9%

to 12.3%; meanwhile, the fraction of genes detected at the LE but not IPA level was slightly decreased, going from 39% to 34% (1.15-fold decrease; Fig. 1J). Thus, both our 3'-seq and long-read RNA-seq data indicate that cisplatin globally favors the expression of IPA isoforms when compared to LE isoforms.

### **A subset of IPA isoforms are depleted in heavy polysomes and terminate in the annotated 5'UTR part of genes**

To gain insights into the potential function of IPA isoforms, and because little is known about IPA isoform translation on a genome-wide scale, we carried out 3'-seq analyses on the heavy polysomes fraction (HP, corresponding to actively translated mRNAs) of cisplatin- and vehicle-treated A549 cells. As in the case of cytosolic RNA that was analyzed in parallel (Fig. 1B), in the HP fraction cisplatin mostly upregulated the IPA:LE isoform ratio (Fig. 2A and Table S8). These data suggested that at least a subset of cisplatin-upregulated IPA isoforms are translated.

To compare the translation efficiency of IPA and LE isoforms, we then analyzed the relative abundance of IPA and matched LE isoforms in HPs *versus* cytosol. Performing this analysis on all IPA isoforms (not just those regulated by cisplatin), we observed that a subset of them exhibited a differential translation efficiency (more often lower) relative to the corresponding LE isoform (Fig. 2B and S2A; Tables S9 and S10). Thanks to these analyses, we identified a subset of cisplatin-upregulated IPA isoforms that were efficiently translated: they were upregulated by cisplatin (relative to matched LE isoforms) in both cytosol and HPs, and did not have a translation efficiency lower than the corresponding LE isoform (N=91; Fig. 2C and Table S11). These IPA isoforms likely encode protein isoforms with a truncated or distinct carboxy-terminal domain, compared to the protein encoded by the full-length mRNA (LE isoform), as previously shown for various genes (3).

These analyses also identified a subset of cisplatin-upregulated IPA isoforms that were inefficiently recruited to HPs: they were upregulated by cisplatin (relative to matched LE isoforms) in the cytosol but not in HPs, and had a lower HP:cytosol ratio than the corresponding LE isoform (N=210; Fig. 2C and Table S11). One of these 210 IPA isoforms is produced by the *ASCC3* (Activating Signal Cointegrator 1 Complex Subunit 3) gene and was previously reported to have a noncoding RNA function (8). In agreement with our 3'-seq data, RT-qPCR analyses showed that the *ASCC3* IPA isoform was upregulated (relative to the LE isoform) by cisplatin in the cytosol (Fig. 2D, left) and had a much lower HP:cytosol ratio than the LE isoform (Fig. 2E, left), which is consistent with its noncoding function.

When compared to the IPA isoforms that were efficiently translated (recruited to HPs), the inefficiently translated ones were on average 2.9 times shorter (Fig. 2F). In addition, they were enriched 2.8 times in the 5' part of genes (Fig. 2G and S2B) and 7.3 times in the annotated 5'UTR part of genes (Fig. 2H). The latter configuration means that the IPA isoform terminates upstream of the annotated translation initiation site of the gene, which could explain the defect in recruitment to HPs. We thus coined such isoforms 5'UTR-IPA isoforms. (Of note, the *ASCC3* IPA isoform is not a 5'UTR-IPA isoform because it is produced from the coding region of the gene.)

For four genes with 5'UTR-IPA isoforms, namely *PHF20* (PHD Finger Protein 20), *PRKAR1B* (Protein Kinase cAMP-Dependent Type I Regulatory Subunit Beta), *EIF4G3* (Eukaryotic Translation Initiation Factor 4 Gamma 3) and *LRRFIP2* (LRR Binding FLII Interacting Protein 2), we validated by RT-qPCR

that the 5'UTR-IPA isoform was upregulated (relative to the LE isoform) by cisplatin in the cytosol and had a much lower HP:cytosol ratio than the LE isoform (Fig. 2D-E). In other words, for these genes, the IPA:LE isoform ratio was upregulated by cisplatin in the cytosol, and was much lower in HPs than in the cytosol, in agreement with our 3'-seq data. Altogether, our 3'-seq data identify a set of IPA isoforms, especially 5'UTR-IPA isoforms, that are upregulated (relative to the LE isoform) by cisplatin in the cytosol, but are depleted in HPs when compared to cytosol.

### **The 5'UTR-IPA isoforms of the *PHF20* and *PRKAR1B* genes impact cisplatin sensitivity**

To determine whether the *PHF20* and *PRKAR1B* 5'UTR-IPA isoforms may play a functional role in cisplatin response, we designed two independent siRNAs that depleted each IPA isoform without affecting the corresponding LE isoform (Fig. 3A). In A549 cells transfected with a control siRNA, cisplatin decreased cell growth by about 60% (data not shown). siRNA-mediated depletion of both IPA isoforms had no robust effect on cell growth in the absence of cisplatin (Fig. 3B), but reproducibly increased cell survival to cisplatin, that is, cell growth in the presence of cisplatin normalized to cell growth without drug (Fig. 3C). These data indicate that the *PHF20* and *PRKAR1B* 5'UTR-IPA isoforms, whose expression is maintained following cisplatin treatment (Fig. 2D), inhibit cell survival to cisplatin. Because cisplatin decreases the expression levels of the LE isoforms of these genes (Fig. 2D), we also determined the effects of depleting these isoforms on cell growth. For both genes, two independent siRNAs that depleted the LE but not the IPA isoform led to decreased cell growth in both the absence and presence of cisplatin (Fig. 3D-E). Thus, for both genes, our data suggest that in response to cisplatin, cell growth is inhibited both by the 5'UTR-IPA isoform (whose expression is maintained) and by the decreased expression of the canonical LE isoform (Fig. 3F).

### **The 5'UTR-IPA isoforms of the *PHF20* and *PRKAR1B* genes are translated into microproteins**

Because the cytosolic IPA isoforms of *PHF20* and *PRKAR1B* are depleted in HPs (Fig. 2E) and are generated from the annotated 5'UTR part of genes, we reasoned that their depletion phenotype (Fig. 3) could be due either to a noncoding function or to the production of microproteins encoded by sORFs (see introduction) (18, 19). To determine whether 5'UTR-IPA isoforms may contain sORFs, we first determined their exact exon content. The *PHF20* and *PRKAR1B* 5'UTR-IPA isoforms are not annotated. Our long-read RNA-seq data indicated that the *PHF20* 5'UTR-IPA isoform starts in the first exon of the gene, like the full-length *PHF20* mRNA (Fig. 4A). In the 5'UTR-IPA isoform, exon 1 is spliced to a non-annotated exon that is located within the first intron of the full-length transcript and that is supported by both our long-read RNA-seq and total-RNA-seq data (that is, RNA-seq on total RNA depleted of ribosomal RNA). This exon ends with a polyA site, that is also supported by our 3'-seq data (Fig. 4A). The *PRKAR1B* gene has two alternative first exons, both of which can be used to produce the full-length *PRKAR1B* mRNAs. In this gene, our long-read RNA-seq data revealed two 5'UTR-IPA isoforms that contain either one of the two alternative first exons, which are then spliced to a non-annotated exon (Fig. 4B). This novel exon is also supported by our total-RNA-seq data and ends with a polyA site supported by our 3'-seq data. Thus, in both genes, the 5'UTR-IPA isoforms contain an ALE exon located within the annotated first intron of the gene.

We then searched for sORFs supported by Ribo-seq datasets and matching the *PHF20* and *PRKAR1B* 5'UTR-IPA isoforms. In the OpenProt database (34), one such sORF (hereafter referred to as *PHF20* sORF#1) was found for *PHF20* and two were found for *PRKAR1B* (sORF#1 and sORF#2) (Fig. 4A-B 'Ribo-Seq', Fig. S3A-B, and Table S12). For both genes, these sORFs are entirely contained within the ALE exon that defines the 5'UTR-IPA isoform. For *PHF20*, additional potential Ribo-seq-supported sORFs were found in the sORFs.org database (35) (Table S13). These data suggest that the *PHF20* and *PRKAR1B* 5'UTR-IPA isoforms are translated.

To further test whether the *PHF20* and *PRKAR1B* 5'UTR-IPA isoforms are translated in cisplatin-treated A549 cells, we analyzed by RT-qPCR the sucrose density gradient fractions that were obtained from the cytosol of these cells and that correspond to sub-, light- and heavy-polysomal fractions. For both genes, the 5'UTR-IPA isoforms were associated with lower-density fractions when compared to matched LE isoforms (Fig. 4C). The 5'UTR-IPA isoforms were almost absent from the heaviest polysomal fractions, while they were predominant in sub-polysomal fractions. Importantly, about 15% of each 5'UTR-IPA isoform was associated with a single full ribosome (corresponding to a recognized initiation codon), and about 45% with light polysomes (mainly 2 or 3 ribosomes; Fig. 4C). Similar results were obtained in vehicle-treated cells (Fig. S3D). As controls, a lncRNA (*SCARNA18*) was scarcely found in light polysomes (Fig. 4C, bottom), while the IPA isoform of *PATZ1* was abundant in HPs in agreement with our 3'-seq data (Fig. S3C and Table S11). These data indicate that the *PHF20* and *PRKAR1B* 5'UTR-IPA isoforms are enriched in light polysomes. Altogether, these polysome profiling data and the above-mentioned Ribo-seq data indicate that the *PHF20* and *PRKAR1B* 5'UTR-IPA isoforms are translated. Their association with light, rather than heavy polysomes, is consistent with the small size of their ORFs (coding 35 to 38 amino acids), given that a ribosome covers about 30 nucleotides.

Then, we investigated whether microproteins that are predicted to be produced by sORFs in the *PHF20* and *PRKAR1B* 5'UTR-IPA isoforms (Fig. 4D) could indeed be detected. In mass spectrometry datasets available in OpenProt (34), we found peptide detection evidence for the microprotein corresponding to sORF#1 of *PHF20* (Fig. 4A, 'Mass Spec') but no such evidence for the sORFs of *PRKAR1B*. To further test microprotein production, we cloned *PHF20* sORF#1, *PRKAR1B* sORF#1 and *PRKAR1B* sORF#2 in expression plasmids, with a carboxy-terminal Flag tag, the ALE sequence upstream of the sORF (the sequence downstream of the sORF was omitted), and a separately encoded GFP. Following transfection into HeLa cells, immunofluorescence with an anti-Flag antibody detected no signal in the case of a GFP-only plasmid (without Flag) but a strong signal in the case of *PRKAR1B* sORF#2 (Fig. 4E). In the absence of cisplatin, this signal was exclusively nuclear and mainly located in a few large nuclear areas (Fig. 4E, bottom, 'vehicle'). However, following cisplatin treatment, the signal became diffuse throughout the nucleus and cytoplasm (Fig. 4E, bottom, 'cisplatin'). We could not detect signal for *PRKAR1B* sORF#1 and *PHF20* sORF#1 (data not shown). The microprotein encoded by *PRKAR1B* sORF#2 was also detected by Western blot analysis on transfected HEK-293T cells (Fig. 4F). Thus, multiple approaches (*i.e.*, polysome profiling, Ribo-Seq, mass spectrometry, immunofluorescence, and/or Western blot) indicate that the 5'UTR-IPA isoforms of *PRKAR1B* and *PHF20* are translated and encode microproteins. Altogether, our data identify 5'UTR-IPA isoforms of *PHF20* and *PRKAR1B*, that are translated (Fig. 4) and inhibit cell survival to cisplatin (Fig. 3).



### **The cisplatin survival phenotype of the *PRKAR1B* 5'UTR-IPA isoform is attributable to its sORF#2**

To determine, whether the phenotype of the *PRKAR1B* 5'UTR-IPA isoform is due to its sORF#2, we mutated by CRISPR the initiation codon of this sORF in the endogenous *PRKAR1B* gene in A549 cells and we obtained 7 such homozygously mutated clones (Fig. 5A, mATG). We also generated 4 homozygous clones with CRISPR deletion of the IPA site ( $\Delta$ IPA). As control cells, we used 8 unsuccessful CRISPR clones and parental A549 cells. We verified by RT-qPCR that the expression of the *PRKAR1B* 5'UTR-IPA isoform was compromised in  $\Delta$ IPA (but not mATG) clones, when compared to control cells (Fig. 5B). In addition, Western blot analysis showed that the expression of the canonical *PRKAR1B* protein was not altered in  $\Delta$ IPA and mATG clones (Fig. 5C). Finally, we looked at cell growth in the three groups of clones. In the absence of cisplatin, cell growth was slightly increased in  $\Delta$ IPA but not mATG clones, when compared to control cells (Fig. 5D). In conditions where cell survival to cisplatin was about 60% in the control group, both the  $\Delta$ IPA and mATG groups of clones survived significantly more to the drug, when compared to the control group (Fig. 5E). This increased survival to cisplatin, which was also observed with siRNAs targeting the *PRKAR1B* 5'UTR-IPA isoform (Fig. 3C), was found in the majority of both  $\Delta$ IPA and mATG clones (Fig. 5E; a minority did not survive more than controls, possibly due to off-target mutations or compensatory mechanisms). These data indicate that both the depletion of the *PRKAR1B* 5'UTR-IPA isoform (by either siRNA or CRISPR) and the CRISPR mutation of its sORF#2 ATG lead to increased cell survival to cisplatin. Thus, the phenotype of the *PRKAR1B* 5'UTR-IPA isoform is due at least in part to its sORF#2.

### **The concept of miP-5'UTR-IPA genes and their widespread regulation by cisplatin**

Based on our data on the *PHF20* and *PRKAR1B* genes, we propose to define the novel paradigm of miP-5'UTR-IPA genes, where an IPA isoform is generated in the annotated 5'UTR part of a canonical protein-coding gene and encodes a microprotein (Fig. 6A). To further identify miP-5'UTR-IPA genes regulated by cisplatin, we first identified 290 genes with 5'UTR-IPA isoforms that were upregulated (relative to the last exon of genes) in response to cisplatin in either H358 or A549 cells (Table S14). We found that 156 of them had, between the gene 5'-end and the IPA site, an sORF supported by Ribo-Seq (135 cases) and/or mass spectrometry datasets (50 cases; Fig. 6B and Table S15). These data identify a set of 156 candidate genes with cisplatin-regulated miP-5'UTR-IPA. Looking at the annotated function of these 156 genes, we found them to be enriched in specific functions, most notably signal transduction and transcription regulation (Fig. 6C), as exemplified by *PRKAR1B* and *PHF20*, respectively.

### **Cisplatin effect on the IPA:LE ratio is mediated by decreased transcription processivity and depends on *FANCD2* and *senataxin***

Finally, we investigated on the molecular mechanisms underlying IPA:LE isoform regulation in response to cisplatin. Based on our long-read RNA-seq data, the average length of detected mRNAs was 1.54 times smaller in cisplatin-treated cells than in vehicle-treated cells (1.3 versus 2.0 kb; Fig. 7A, left panel). The cisplatin-induced decrease in transcript length was even more pronounced (1.86-fold) when introns were taken into account (pre-mRNA length of 5.0 versus 9.3 kb; Fig. 7A, right panel). These data suggest that cisplatin favors the expression of shorter primary transcripts. In our 3'-seq data, genes with

IPA:LE isoform ratio upregulation by cisplatin were about 3 times longer than genes with non-regulated IPA, indicating that the upregulation of IPA:LE isoform ratio by cisplatin is enriched in long genes (Fig. 7B). Altogether, our 3'-seq and long-read RNA-seq data indicate that cisplatin treatment favors the expression of IPA *versus* full-length (LE) isoforms in long genes. (Cisplatin also seemed to favor the expression of shorter genes, because pre-mRNA length of LE transcripts in cisplatin-treated cells was similar to that of IPA transcripts in vehicle-treated cells [Fig. 7A, right panel]; and the length of cisplatin-upregulated IPA transcripts was similar to the length of LE transcripts of non-regulated genes [Fig. 7B]).

Because the cisplatin-induced upregulation of the IPA:LE isoform ratio was enriched in long genes and was accompanied by a decrease of LE isoform levels, both globally (Fig. 1I-J) and in many tested genes (Fig. 1G, 2D, and S1G), we reasoned that this cisplatin effect may be due to an inhibition of transcription processivity toward the 3'-end of long genes. Also consistent with this hypothesis, cisplatin was reported to stall RNA polymerase and inhibit transcription elongation in several genes or gene constructs (37–39). To test this hypothesis, we carried out total-RNA-seq (that is, RNA-seq on total RNA depleted of ribosomal RNA) on H358 cells treated or not with cisplatin, and we used the ratio of intronic reads at the 3' *versus* 5' part of genes (3':5' of gene) as a proxy for transcription processivity. Cisplatin inhibited the relative expression of the 3' *versus* 5' part of genes in 1,465 (22%) out of 6,653 analyzed genes (>20% decrease and  $p < 0.05$ ) and this effect was enriched in long genes (Fig. 7C and Table S5), suggesting a decrease of transcription processivity. Genes exhibiting IPA:LE isoform ratio upregulation by cisplatin had more often their transcription processivity inhibited by the drug, when compared to genes with non-regulated IPAs (81% *versus* 51%; Fig. 7D, middle). Conversely, IPA:LE isoform ratio upregulation was more frequent in genes exhibiting processivity inhibition (50% *versus* 20%; Fig. 7D, right). Thus, the upregulation of IPA:LE isoform ratio and the downregulation of transcription processivity in response to cisplatin were correlated. By RT-qPCR using primers in introns at the 3' *versus* 5' part of genes, we verified that cisplatin treatment inhibited transcription processivity in 10 out of 13 tested genes with IPA:LE isoform ratio upregulation (Fig. S4A-B and see below). More detailed analyses on the *PRIM2* gene indicated that in response to cisplatin, pre-mRNA levels strongly decreased between the regulated IPA and the last intron (Fig. S4C); the dose-dependent effects on processivity and isoform ratio were correlated (Fig. 1H and S4D); and the processivity effect was seen at as early as 1 hour of treatment (Fig. S4E). Thus, the cisplatin-induced decrease of processivity likely explains the selective decrease of the LE but not IPA isoform levels and the increase in the IPA:LE isoform ratio.

We then investigated on factors that could explain the cisplatin-induced decrease of transcription processivity and increase of IPA:LE isoform ratio. Cisplatin-DNA crosslinks in transcribed genes trigger RNA polymerase II stalling and transcription-coupled nucleotide excision repair (TC-NER) mediated by CSB and XPF (XPF is also involved in other repair pathways) (40). In addition, interstrand DNA crosslinks induced by cisplatin are repaired by the Fanconi anemia pathway involving FANCD2 (40). Depletion of either CSB or XPF with siRNAs did not prevent cisplatin effects on processivity and IPA:LE isoform ratio in the *PRKAR1B* and *PHF20* genes and on *PRIM2* processivity (Fig. 7E-F and S4F), thus arguing against a role of TC-NER in these cisplatin effects. In contrast, siRNA depletion of FANCD2 with two independent siRNAs partially prevented cisplatin effects on processivity and IPA:LE ratio in the *PRKAR1B* and *PHF20* genes and on *PRIM2* processivity (Fig. 7E-F and S4F). These effects of FANCD2 are unlikely due to its canonical role in the Fanconi repair pathway, that occurs in S phase (40). FANCD2

was also shown to interact with R-loops (41–44), which are structures made of RNA:DNA hybrids (and a displaced DNA strand) and formed during transcription; and to colocalize with senataxin (SETX) (45), an R-loop helicase that promotes the use of some polyA sites (46). As in the case of FANCD2, siRNA depletion of SETX partially prevented cisplatin effects on processivity and IPA:LE ratio in the *PRKAR1B* and *PHF20* genes and on *PRIM2* processivity (Fig. 7G-H and S4G). Thus, our data suggest that the cisplatin-induced increase in the ratio of (miP-5'UTR-)IPA to LE isoforms in long genes is mediated, at least in part, by a decrease of transcription processivity, that is dependent on FANCD2 and SETX (Fig. 7I).

## DISCUSSION

IPA isoforms are widely regulated in various biological processes and are usually thought to encode protein isoforms (they are sometimes referred to as coding region-APA isoforms (4, 47)). Our findings that many cisplatin-regulated IPA isoforms are generated in the annotated 5'UTR part of genes and are translated into microproteins provide new insights into the translation outcome and function of IPA isoforms and reveal the novel paradigm of miP-5'UTR-IPA genes. In addition, our findings that cisplatin upregulates the IPA:LE isoform ratio in many genes provides new insights into the effects of this anticancer drug and suggest a role for miP-5'UTR-IPA isoform regulation in cancer cell response to cisplatin.

The almost systematic up- (as opposed to down-) regulation pattern of the IPA:LE isoform ratio in whole cells in response to cisplatin is distinct from the effects previously described for other genotoxic anticancer drugs, with doxorubicin downregulating the IPA:LE isoform ratio in most cases and camptothecin inducing both patterns in approximately equal proportions (7, 11). Our data strongly suggest a role for transcription processivity inhibition in IPA:LE isoform ratio upregulation by cisplatin. Previous analyses on gene constructs showed that cisplatin-DNA adducts induce RNA polymerase stalling and inhibit transcription elongation; several mechanisms are involved, including interstrand covalent bonds (crosslinks), RNA polymerase modifications, and reduced nucleosome mobility (37–39, 48, 49). Our analyses of intronic RNA levels by total-RNA-seq in thousands of endogenous genes and by RT-qPCR in several genes (Fig. 7 and S4A-E) suggest that cisplatin inhibits transcription processivity in long genes, which can directly explain the decrease of full-length mRNA (LE isoform) levels relative to IPA isoform levels. In future studies, analyses of nascent RNA could be used to more directly assess transcription dynamics, including elongation rate. Because cleavage/ polyadenylation is extensively coupled to transcription, reduced elongation rate near the IPA site could favor its cotranscriptional recognition, which in turn could favor intronic transcription termination. Upregulation of the IPA:LE isoform ratio through elongation inhibition in long genes was proposed for UV-C (8); consistently, many regulation events induced by UV-C were found in our cisplatin dataset (Fig. S1B; the limited dataset available for UV-C precludes the converse analysis).

Mechanistically, our data indicate that processivity inhibition and IPA:LE upregulation by cisplatin are dependent on FANCD2 and SETX (Fig. 7E-I and S4F-G). Both of these proteins are known to colocalize with each other (45) and to interact with R-loops (41–44), and the effects of SETX on cleavage/ polyadenylation are R-loop dependent (46). Consistent with this, we found that depletion of RNase H1

(which selectively degrades RNA hybridized to DNA) prevented processivity inhibition by cisplatin in the *PRIM2* gene (Fig. S4H-I), but effects on other genes were more elusive (data not shown). Thus, the potential role of R-loops in IPA:LE regulation by cisplatin, FANCD2 and SETX remains to be investigated. Additional factors may contribute to IPA:LE upregulation by cisplatin, because IPA is regulated by many factors involved in cleavage/ polyadenylation, splicing, transcription elongation and termination, and epigenetic marks (1–4). For example, depletion or inhibition of CDK12 (a protein kinase that phosphorylates RNA polymerase II) inhibits transcription processivity and upregulates the IPA:LE isoform ratio preferentially in long genes (13, 50). Finally, IPA:LE isoform ratio upregulation in several genes upon UV-C irradiation was shown to be mediated by downregulation of the U1 snRNA, which as part of the U1 snRNP is a widespread repressor of IPA (9). Although we found a significant overlap between cisplatin and UV-C regulated IPA events (Fig. S1B), U1 snRNA levels did not seem to be decreased by cisplatin treatment (data not shown).

Our findings also provide new insights into the translation outcome and function of IPA isoforms, especially those produced from the 5' part of genes (5'IPA isoforms) and more specifically from the annotated 5'UTR part of genes (5'UTR-IPA isoforms). First, while two studies identified 5'IPA isoforms that were degraded in the nucleus, thereby preventing their export and translation (15, 16), our data identify a set of IPA isoforms that are found in the cytosol but are lowly associated with HPs (Fig. 2C) and this set is enriched in 5'IPA, more precisely 5'UTR-IPA (Fig. 2F-H and S2B). Some of these 5'IPA isoforms may have a noncoding function, as shown for the IPA isoform of *ASCC3* (8), which is one of the cisplatin-upregulated 5'IPA (but not 5'UTR-IPA) isoforms that we found to be inefficiently associated with HPs (Fig. 2D-E). Along these lines, a 5'UTR-IPA isoform (called SPUD) with a noncoding function was recently described in the *CDKN1A* gene (14).

Second, while a study identified immune cell-enriched 5'IPA isoforms that have a low coding potential and were therefore classified as lncRNAs (17), their translation status and potential function were not determined, and to date, IPA isoforms were not shown to encode microproteins. Here, we provide evidence that many 5'UTR-IPA transcripts, including those in *PHF20* and *PRKAR1B*, contain sORFs that are translated into microproteins, as indicated by our analyses of Ribo-seq and mass-spectrometry datasets and by our polysome profiling, immunofluorescence and Western blot data (Fig. 4, 6B and S3). Thus, our findings reveal 5'UTR-IPA isoforms as a novel source of microproteins. In the case of *PHF20* sORF#1, the fact that the microprotein is supported by mass spectrometry data but not by our immunofluorescence analysis might be due to diffusion out of permeabilized cells, secretion or instability. Alternatively, the additional sORFs (detected by sORFs.org; Table S13) upstream or 5'-overlapping the sORF#1 of *PHF20* might inhibit its translation in some biological conditions (they might also encode their own microproteins but we have not tagged them).

Third, our data reveal the novel paradigm of miP-5'UTR-IPA genes, that produce -from the same promoter but through an IPA switch- a canonical protein-coding full-length mRNA and a microprotein-coding miP-5'UTR-IPA isoform (Fig. 6A). This paradigm is distinct from the case of sORFs located within mRNAs along with the canonical ORF, as in the case of upstream ORFs (21). Indeed, in miP-5'UTR-IPA genes, the sORF and the canonical ORF lie in distinct transcript isoforms. Such a situation has only been shown for some microproteins encoded by circular RNAs, which are alternative transcripts generated by backsplicing (51).

Our data support the functional relevance of miP-5'UTR-IPA genes. Our data indicate that the miP-5'UTR-IPA isoforms of the *PHF20* and *PRKAR1B* genes are functional, because their depletion decreases cisplatin toxicity (Fig. 3). CRISPR editing of the initiation codon of sORF#2 in the *PRKAR1B* miP-5'UTR-IPA isoform shows that the role of this isoform in cisplatin survival is mediated at least in part by sORF#2 (Fig. 5). The cisplatin-induced relocalization of the microprotein encoded by this sORF (Fig. 4E) also argues for its functionality. In addition, for both the *PHF20* and *PRKAR1B* genes, our data suggest that the miP-5'UTR-IPA and full-length mRNA isoforms have opposite effects on cell growth, and that both miP-5'UTR-IPA isoform expression and full-length mRNA down-regulation may contribute to the anti-proliferative effects of cisplatin (Fig. 3F). Interestingly, the PHF20 canonical protein is involved in the DNA damage response, more specifically in the regulation of TP53 (Tumor Protein P53) protein stability in response to DNA damage (52–54). The potential relevance of the downregulation of the PRKAR1B canonical protein, which is a regulatory subunit of protein kinase A, is less clear, but cyclic AMP signalling has been involved in lung cancer cell sensitivity to DNA damage (55).

Besides miP-5'UTR-IPA, our study identifies other IPA isoforms of potential functional interest. In particular, our finding that cisplatin upregulates a 5'IPA isoform in the coding region of *ZFC3H1* (Zinc Finger C3H1-Type Containing; Fig. S1G), a gene encoding a protein involved in the nuclear degradation of 5'IPA transcripts (15), raises the possibility of a positive feed-back mechanism, whereby *ZFC3H1* transcript truncation in response to cisplatin might enhance the cellular accumulation of 5'IPA isoforms. Likewise, we found that cisplatin upregulates a 5'IPA isoform in the coding region of *INTS6* (Integrator Complex Subunit 6; Fig. S1G), a gene encoding a component of the integrator complex that controls the 3'-end maturation and transcription termination of U snRNAs and the early elongation of specific coding genes; *INTS6* was also involved in the DNA damage response (56–58). Regarding the *PRIM2* IPA isoform identified in this study, it is produced from the coding region of the gene (Fig. 1D) and could be detected in polysomes, but we did not detect a truncated PRIM2 protein by Western blot (data not shown), which might be due to a lack of stability. Nevertheless, our siRNA analyses indicated that this *PRIM2* IPA isoform promotes cell survival to cisplatin (Fig. S5); and the strong decrease of full-length PRIM2 mRNA and protein levels in response to cisplatin (Fig. 1G and S1F) is expected to affect DNA replication, where this protein plays a key role as a subunit of DNA primase. Further studies are required to determine the role of the cisplatin-induced large-scale regulation of IPA:LE isoforms (which is enriched in genes related to the DNA damage response, cell cycle and cell death; Fig. 1C) in cancer cell responses to cisplatin.

In conclusion, our data indicate that upregulation of the IPA:LE isoform ratio is a widespread mechanism of transcriptome regulation by cisplatin, that links different layers of gene expression regulation by cisplatin (*i.e.*, it is impacted by transcription processivity and it impacts the translome) and is involved in cancer cell responses to the widely used anticancer drug, cisplatin. While little was known about the translation status of IPA isoforms, our findings identify distinct subsets of IPA isoforms that are differentially recruited to polysomes. In addition to 3'UTR-APA that affects 3'UTR length, coding region-APA that typically generates protein isoforms (1–4), and two instances of IPA isoforms with noncoding functions (8, 14), our data also reveal 5'UTR-IPA isoforms that encode microproteins and are functional. The novel paradigm of miP-5'UTR-IPA, that we reveal in this study, implicate in cancer cell response to cisplatin, and find enriched in genes involved in specific functions (*e.g.*, signaling and transcription; Fig.

6C), opens wide avenues of research. Future studies should identify more extensively miP-5'UTR-IPA isoforms (here we focused on cisplatin-regulated ones) and determine their function in physiology and disease.

## **DATA AVAILABILITY**

The datasets generated in this study have been deposited in the UCSC genome browser (private pre-publication link: [https://genome.ucsc.edu/s/clabbe/visuCDDPpaper\\_Dutertre2021](https://genome.ucsc.edu/s/clabbe/visuCDDPpaper_Dutertre2021)) and in the Gene Expression Omnibus repository (GEO) under accession numbers GSE119978 (token: azazmciqfbatzyd), GSE180854 (token: mnevgiouzfetjz) and GSE180918 (token: ihmlucmubhyzil). The complete bioinformatics pipeline for 3'-seq analysis of IPA (3'-SMART package) can be freely downloaded at GitHub (<https://github.com/InstitutCurie/3-SMART>). Any additional information required to reanalyze the data reported in this paper is available upon request. The plasmids are available upon request.

**Supplementary Data are available online.**

## **FUNDING**

This work was supported by grants of the Fondation pour la Recherche Médicale (DBI20141231314), Institut National du Cancer (2015-141), and Institut Curie (ICGex) to MDu, by a grant of Ligue Nationale Contre le Cancer (équipe labellisée) to SV, by fellowships from Ministère de l'Enseignement Supérieur et de la Recherche to IT and AD, and by a fellowship from Fondation ARC to IT. The ICGex NGS platform of the Institut Curie is supported by grants ANR-10-EQPX-03 (Equipex) and ANR-10-INBS-09-08 (France Génomique Consortium) from the Agence Nationale de la Recherche ("Investissements d'Avenir" program), by the Canceropôle Ile-de-France and by the SiRIC-Curie program - SiRIC Grant INCa-DGOS-4654.

## **ACKNOWLEDGEMENTS**

We thank Patricia Uguen, Christine Tran Quang, and Albertas Navickas for helpful technical advice.

## **CONFLICT OF INTEREST**

SV is a scientific cofounder of Ribonexus. The other authors declare no competing interests.

## **AUTHORS' CONTRIBUTIONS**

Experiments were performed by AD, IT, QF, AHM, AC and LT under the supervision of MDu, and by SM under the supervision of BE. Illumina and PacBio sequencing were performed by VR and SL, respectively, under the supervision of SB. Bioinformatics and biostatistics analyses were performed by MC, PG, MDe and CML under the supervision of MDu and NS, and by HM, JBC and NF under the

supervision of DA. MDu wrote the paper with help of BE, SV, IT, and AD. All the authors read and approved the final manuscript.

## REFERENCES

1. Ogorodnikov,A., Kargapolova,Y. and Danckwardt,S. (2016) Processing and transcriptome expansion at the mRNA 3' end in health and disease: finding the right end. *Pflugers Archiv European Journal of Physiology*, **468**, 993–1012.
2. Gruber,A.J. and Zavolan,M. (2019) Alternative cleavage and polyadenylation in health and disease. *Nature Reviews Genetics*, 10.1038/s41576-019-0145-z.
3. Tian,B. and Manley,J.L. (2016) Alternative polyadenylation of mRNA precursors. *Nature Reviews Molecular Cell Biology*, **18**, 18–30.
4. Neve,J., Patel,R., Wang,Z., Louey,A. and Furger,A.M. (2017) Cleavage and polyadenylation: Ending the message expands gene regulation. *RNA Biol*, **14**, 865–890.
5. Sandberg,R., Neilson,J.R., Sarma,A., Sharp,P. a and Burge,C.B. (2008) Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites. *Science (New York, N.Y.)*, 10.1126/science.1155390.
6. Elkon,R., Drost,J., van Haaften,G., Jenal,M., Schrier,M., Vrieling,J.A.O. and Agami,R. (2012) E2F mediates enhanced alternative polyadenylation in proliferation. *Genome Biology*, **13**.
7. Dutertre,M., Chakrama,F.Z., Combe,E., Desmet,F.O., Mortada,H., Espinoza,M.P., Gratadou,L. and Auboeuf,D. (2014) A recently evolved class of alternative 3'-terminal exons involved in cell cycle regulation by topoisomerase inhibitors. *Nature Communications*, 10.1038/ncomms4395.
8. Williamson,L., Saponaro,M., Boeing,S., East,P., Mitter,R., Kantidakis,T., Kelly,G.P., Lobley,A., Walker,J., Spencer-Dene,B., *et al.* (2017) UV Irradiation Induces a Non-coding RNA that Functionally Opposes the Protein Encoded by the Same Gene. *Cell*, 10.1016/j.cell.2017.01.019.
9. Devany,E., Park,J.Y., Murphy,M.R., Zakusilo,G., Baquero,J., Zhang,X., Hoque,M., Tian,B. and Kleiman,F.E. (2016) Intronic cleavage and polyadenylation regulates gene expression during DNA damage response through U1 snRNA. *Cell Discovery*, 10.1038/celldisc.2016.13.
10. Hollerer,I., Curk,T., Haase,B., Benes,V., Hauer,C., Neu-Yilik,G., Bhuvanagiri,M., Hentze,M.W. and Kulozik,A.E. (2016) The differential expression of alternatively polyadenylated transcripts is a common stress-induced response mechanism that modulates mammalian mRNA expression in a quantitative and qualitative fashion. *RNA*, **22**, 1441–1453.
11. Chakraborty,A., Cadix,M., Relier,S., Taricco,N., Alaeitabar,T., Devaux,A., Labbé,C.M., Martineau,S., Heneman-Masurel,A., Gestraud,P., *et al.* (2022) Compartment-specific and ELAVL1-coordinated regulation of intronic polyadenylation isoforms by doxorubicin. *Genome Res*, **32**, 1271–1284.
12. Cugusi,S., Mitter,R., Kelly,G.P., Walker,J., Han,Z., Pisano,P., Wierer,M., Stewart,A. and Svejstrup,J.Q. (2022) Heat shock induces premature transcript termination and reconfigures the human transcriptome. *Mol Cell*, **82**, 1573-1588.e10.
13. Dutertre,M., Sfaxi,R. and Vagner,S. (2021) Reciprocal Links between Pre-messenger RNA 3'-End Processing and Genome Stability. *Trends Biochem Sci*, **46**, 579–594.

14. Murphy, M.R., Ramadei, A., Doymaz, A., Varriano, S., Natelson, D.M., Yu, A., Aktas, S., Mazzeo, M., Mazzeo, M., Zakusilo, G., *et al.* (2023) Long non-coding RNA generated from CDKN1A gene by alternative polyadenylation regulates p21 expression during DNA damage response. *Nucleic Acids Res*, 10.1093/nar/gkad899.
15. Ogami, K., Richard, P., Chen, Y., Hoque, M., Li, W., Moresco, J.J., Yates, J.R., Tian, B. and Manley, J.L. (2017) An Mtr4/ZFC3H1 complex facilitates turnover of unstable nuclear RNAs to prevent their cytoplasmic transport and global translational repression. *Genes and Development*, **31**, 1257–1271.
16. Chiu, A.C., Suzuki, H.I., Wu, X., Mahat, D.B., Kriz, A.J. and Sharp, P.A. (2018) Transcriptional Pause Sites Delineate Stable Nucleosome-Associated Premature Polyadenylation Suppressed by U1 snRNP. *Molecular Cell*, **69**, 648-663.e7.
17. Singh, I., Lee, S.H., Sperling, A.S., Samur, M.K., Tai, Y.T., Fulciniti, M., Munshi, N.C., Mayr, C. and Leslie, C.S. (2018) Widespread intronic polyadenylation diversifies immune cell transcriptomes. *Nature Communications*, **9**.
18. Orr, M.W., Mao, Y., Storz, G. and Qian, S.-B. (2020) Alternative ORFs and small ORFs: shedding light on the dark proteome. *Nucleic Acids Res*, **48**, 1029–1042.
19. Mudge, J.M., Ruiz-Orera, J., Prensner, J.R., Brunet, M.A., Calvet, F., Jungreis, I., Gonzalez, J.M., Magrane, M., Martinez, T.F., Schulz, J.F., *et al.* (2022) Standardized annotation of translated open reading frames. *Nat Biotechnol*, **40**, 994–999.
20. Sandmann, C.-L., Schulz, J.F., Ruiz-Orera, J., Kirchner, M., Ziehm, M., Adami, E., Marczenke, M., Christ, A., Liebe, N., Greiner, J., *et al.* (2023) Evolutionary origins and interactomes of human, young microproteins and small peptides translated from short open reading frames. *Mol Cell*, **83**, 994-1011.e18.
21. Chen, J., Brunner, A.-D., Cogan, J.Z., Nuñez, J.K., Fields, A.P., Adamson, B., Itzhak, D.N., Li, J.Y., Mann, M., Leonetti, M.D., *et al.* (2020) Pervasive functional translation of noncanonical human open reading frames. *Science*, **367**, 1140–1146.
22. Cepeda, V., Fuertes, M.A., Castilla, J., Alonso, C., Quevedo, C. and Perez, J.M. (2008) Biochemical Mechanisms of Cisplatin Cytotoxicity. *Anti-Cancer Agents in Medicinal Chemistry*, **7**, 3–18.
23. Gabriel, M., Delforge, Y., Deward, A., Habraken, Y., Hennuy, B., Piette, J., Klinck, R., Chabot, B., Colige, A. and Lambert, C. (2015) Role of the splicing factor SRSF4 in cisplatin-induced modifications of pre-mRNA splicing and apoptosis. *BMC Cancer*, **15**, 227.
24. Lemaitre, F., Chakrama, F., O’Grady, T., Peulen, O., Rademaker, G., Deward, A., Chabot, B., Piette, J., Colige, A., Lambert, C., *et al.* (2022) The transcription factor c-Jun inhibits RBM39 to reprogram pre-mRNA splicing during genotoxic stress. *Nucleic Acids Res*, **50**, 12768–12789.
25. Kusy, S., Gerby, B., Goardon, N., Gault, N., Ferri, F., Gérard, D., Armstrong, F., Ballerini, P., Cayuela, J.-M., Baruchel, A., *et al.* (2010) NKX3.1 is a direct TAL1 target gene that mediates proliferation of TAL1-expressing human T cell acute lymphoblastic leukemia. *J Exp Med*, **207**, 2141–2156.
26. Boussemart, L., Malka-Mahieu, H., Girault, I., Allard, D., Hemmingsson, O., Tomasic, G., Thomas, M., Basmadjian, C., Ribeiro, N., Thuaud, F., *et al.* (2014) eIF4F is a nexus of resistance to anti-BRAF and anti-MEK cancer therapies. *Nature*, **513**, 105–109.
27. Langmead, B. and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods*, **9**, 357–359.



28. Quinlan,A.R. and Hall,I.M. (2010) BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.
29. Love,M.I., Huber,W. and Anders,S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, **15**.
30. Herrmann,C.J., Schmidt,R., Kanitz,A., Artimo,P., Gruber,A.J. and Zavolan,M. (2020) PolyASite 2.0: A consolidated atlas of polyadenylation sites from 3' end sequencing. *Nucleic Acids Research*, **48**, D174–D179.
31. Wang,R., Nambiar,R., Zheng,D. and Tian,B. (2018) PolyA-DB 3 catalogs cleavage and polyadenylation sites identified by deep sequencing in multiple genomes. *Nucleic Acids Research*, **46**, D315–D319.
32. Huang,D.W., Sherman,B.T. and Lempicki,R.A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, **4**, 44–57.
33. Huang,D.W., Sherman,B.T. and Lempicki,R.A. (2009) Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research*, **37**, 1–13.
34. Brunet,M.A., Brunelle,M., Lucier,J.F., Delcourt,V., Levesque,M., Grenier,F., Samandi,S., Leblanc,S., Aguilar,J.D., Dufour,P., *et al.* (2019) OpenProt: A more comprehensive guide to explore eukaryotic coding potential and proteomes. *Nucleic Acids Research*, **47**, D403–D410.
35. Olexiuk,V., Van Criekinge,W. and Menschaert,G. (2018) An update on sORFs.org: A repository of small ORFs identified by ribosome profiling. *Nucleic Acids Research*, **46**, D497–D502.
36. Lianoglou,S., Garg,V., Yang,J.L., Leslie,C.S. and Mayr,C. (2013) Ubiquitously transcribed genes use alternative polyadenylation to achieve tissue-specific expression. *Genes and Development*, **27**, 2380–2396.
37. Tremeau-Bravard,A., Riedl,T., Egly,J.M. and Dahmus,M.E. (2004) Fate of RNA Polymerase II Stalled at a Cisplatin Lesion. *Journal of Biological Chemistry*, **279**, 7751–7759.
38. Jung,Y. and Lippard,S.J. (2003) Multiple States of Stalled T7 RNA Polymerase at DNA Lesions Generated by Platinum Anticancer Agents. *Journal of Biological Chemistry*, **278**, 52084–52092.
39. Zhu,G., Song,L. and Lippard,S.J. (2013) Visualizing inhibition of nucleosome mobility and transcription by cisplatin-DNA interstrand crosslinks in live mammalian cells. *Cancer Research*, **73**, 4451–4460.
40. Rocha,C.R.R., Silva,M.M., Quinet,A., Cabral-Neto,J.B. and Menck,C.F.M. (2018) DNA repair pathways and cisplatin resistance: an intimate relationship. *Clinics (Sao Paulo)*, **73**, e478s.
41. Okamoto,Y., Abe,M., Itaya,A., Tomida,J., Ishiai,M., Takaori-Kondo,A., Taoka,M., Isobe,T. and Takata,M. (2019) FANCD2 protects genome stability by recruiting RNA processing enzymes to resolve R-loops during mild replication stress. *FEBS J*, **286**, 139–150.
42. Okamoto,Y., Iwasaki,W.M., Kugou,K., Takahashi,K.K., Oda,A., Sato,K., Kobayashi,W., Kawai,H., Sakasai,R., Takaori-Kondo,A., *et al.* (2018) Replication stress induces accumulation of FANCD2 at central region of large fragile genes. *Nucleic Acids Res*, **46**, 2932–2944.
43. García-Rubio,M.L., Pérez-Calero,C., Barroso,S.I., Tumini,E., Herrera-Moyano,E., Rosado,I.V. and Aguilera,A. (2015) The Fanconi Anemia Pathway Protects Genome Integrity from R-loops. *PLoS Genet*, **11**, e1005674.

44. Liang,Z., Liang,F., Teng,Y., Chen,X., Liu,J., Longerich,S., Rao,T., Green,A.M., Collins,N.B., Xiong,Y., *et al.* (2019) Binding of FANCI-FANCD2 Complex to RNA and R-Loops Stimulates Robust FANCD2 Monoubiquitination. *Cell Rep*, **26**, 564-572.e5.
45. Yüce,Ö. and West,S.C. (2013) Senataxin, defective in the neurodegenerative disorder ataxia with oculomotor apraxia 2, lies at the interface of transcription and the DNA damage response. *Mol Cell Biol*, **33**, 406–417.
46. Skourti-Stathaki,K., Proudfoot,N.J. and Gromak,N. (2011) Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Mol Cell*, **42**, 794–805.
47. Di Giammartino,D.C., Nishida,K. and Manley,J.L. (2011) Mechanisms and Consequences of Alternative Polyadenylation. *Molecular Cell*, **43**, 853–866.
48. Jung,Y. and Lippard,S.J. (2006) RNA polymerase II blockage by cisplatin-damaged DNA: Stability and polyubiquitylation of stalled polymerase. *Journal of Biological Chemistry*, **281**, 1361–1370.
49. Todd,R.C. and Lippard,S.J. (2010) Consequences of cisplatin binding on nucleosome structure and dynamics. *Chemistry and Biology*, **17**, 1334–1343.
50. Chou,J., Quigley,D.A., Robinson,T.M., Feng,F.Y. and Ashworth,A. (2020) Transcription-Associated Cyclin-Dependent Kinases as Targets and Biomarkers for Cancer Therapy. *Cancer Discov*, **10**, 351–370.
51. van Heesch,S., Witte,F., Schneider-Lunitz,V., Schulz,J.F., Adami,E., Faber,A.B., Kirchner,M., Maatz,H., Blachut,S., Sandmann,C.-L., *et al.* (2019) The Translational Landscape of the Human Heart. *Cell*, **178**, 242-260.e29.
52. Cui,G., Park,S., Badeaux,A.I., Kim,D., Lee,J., Thompson,J.R., Yan,F., Kaneko,S., Yuan,Z., Botuyan,M.V., *et al.* (2012) PHF20 is an effector protein of p53 double lysine methylation that stabilizes and activates p53. *Nat Struct Mol Biol*, **19**, 916–924.
53. Klein,B.J., Wang,X., Cui,G., Yuan,C., Botuyan,M.V., Lin,K., Lu,Y., Wang,X., Zhao,Y., Bruns,C.J., *et al.* (2016) PHF20 Readers Link Methylation of Histone H3K4 and p53 with H4K16 Acetylation. *Cell Rep*, **17**, 1158–1170.
54. Li,Y., Park,J., Piao,L., Kong,G., Kim,Y., Park,K.A., Zhang,T., Hong,J., Hur,G.M., Seok,J.H., *et al.* (2013) PKB-mediated PHF20 phosphorylation on Ser291 is required for p53 function in DNA damage. *Cell Signal*, **25**, 74–84.
55. Desai,S., Kumar,A., Laskar,S. and Pandey,B.N. (2014) Differential roles of ATF-2 in survival and DNA repair contributing to radioresistance induced by autocrine soluble factors in A549 lung cancer cells. *Cell Signal*, **26**, 2424–2435.
56. Zhang,F., Ma,T. and Yu,X. (2013) A core hSSB1-INTS complex participates in the DNA damage response. *Journal of Cell Science*, **126**, 4850–4855.
57. Stadelmayer,B., Micas,G., Gamot,A., Martin,P., Malirat,N., Koval,S., Raffel,R., Sobhian,B., Severac,D., Rialle,S., *et al.* (2014) Integrator complex regulates NELF-mediated RNA polymerase II pause/release and processivity at coding genes. *Nature Communications*, **5**.
58. Gardini,A., Baillat,D., Cesaroni,M., Hu,D., Marinis,J.M., Wagner,E.J., Lazar,M.A., Shilatifard,A. and Shiekhattar,R. (2014) Integrator regulates transcriptional initiation and pause release following activation. *Molecular Cell*, **56**, 128–139.

## FIGURE LEGENDS

**Figure 1: Cisplatin upregulates the IPA:LE isoform ratio in many genes.** **A**, 3'-seq analysis of IPA:LE isoform ratio regulation by cisplatin in whole H358 cells treated with cisplatin or vehicle for 24 hours. **B**, Similar 3'-seq analysis in A549 cells (cytosol) treated with cisplatin or vehicle for 16 hours. **C**, Gene Ontology terms enriched in genes with IPA:LE regulation by cisplatin in H358 cells. Analysis with the DAVID software; the fold enrichment is indicated; only a subset of enriched functions are shown. **D**, Visualization of 3'-seq and long-read RNA-seq data for the *PRIM2* gene in the UCSC genome browser. The sequence around the regulated IPA site is shown and the polyadenylation signal is framed. **E**, Schematics of the *PRIM2* gene and PCR primers. **F**, RT-PCR detection of the *PRIM2* IPA isoform. SM, size markers. **G**, RT-qPCR analysis of *PRIM2* isoforms in H358 and A549 cells treated for 24 or 16 hours, respectively, with either cisplatin (CisPt) or vehicle (Veh.). **H-J**, Analysis of IPA and LE transcripts by long-read RNA-seq (Iso-Seq, PacBio) in H358 cells treated with cisplatin or vehicle for 24 hours. All detected transcripts were classified into IPA and LE transcripts. **H**, Percentage of IPA transcripts among detected transcripts. **I-J**, Number of genes detected at the level of IPA or LE in the presence or absence of cisplatin.

**Figure 2: A subset of IPA isoforms are depleted in heavy polysomes and terminate in the annotated 5'UTR part of genes.** **A**, 3'-seq analysis of IPA:LE isoform ratio regulation by cisplatin in heavy polysomes (HP) of A549 cells treated with cisplatin or vehicle for 16 hours. **B**, 3'-seq analysis of IPA:LE isoform ratio regulation in HP *versus* cytosol from cisplatin-treated A549 cells. **C**, Subsets of cisplatin-upregulated IPA isoforms are efficiently or inefficiently recruited to HP. Comparison between lists of IPA isoforms that are regulated relative to matched LE isoforms in the indicated conditions. **D-E**, RT-qPCR analysis of IPA and LE isoforms of the indicated genes, in the cytosol of A549 cells treated with cisplatin or vehicle (**D**) and in HP and cytosol of cisplatin-treated A549 cells (**E**). **F-H**, Comparison between the efficiently and inefficiently HP-recruited subsets of cisplatin-upregulated IPA isoforms identified in panel C. **F**, Distance from transcription start site (TSS) to IPA site. Median and SD. **G-H**, Percent of IPA isoforms that are located in the 5' part of the gene (**G**) or upstream of the annotated translation initiation site (indicated as ATG) of the gene (**H**).

**Figure 3: The *PHF20* and *PRKAR1B* 5'UTR-IPA isoforms impact cisplatin sensitivity.** Effects of transfected siRNAs targeting either the IPA or LE isoform of *PHF20* or *PRKAR1B* in A549 cells. Two independent siRNAs were used for each isoform. Data in the presence of a negative control siRNA (siCTL) were set to 100% or 1. **A** and **D**, RT-qPCR analysis of IPA and LE isoforms abundance in vehicle-treated cells. **B**, **C** and **E**, WST1 assay analysis of A549 cell viability in the presence or absence of cisplatin for 24 hours. In C, cell viability with cisplatin was normalized to cell viability without cisplatin. **F**, Model showing the impact of the IPA and LE isoforms of *PHF20* and *PRKAR1B* on cell growth in the absence and presence of cisplatin.

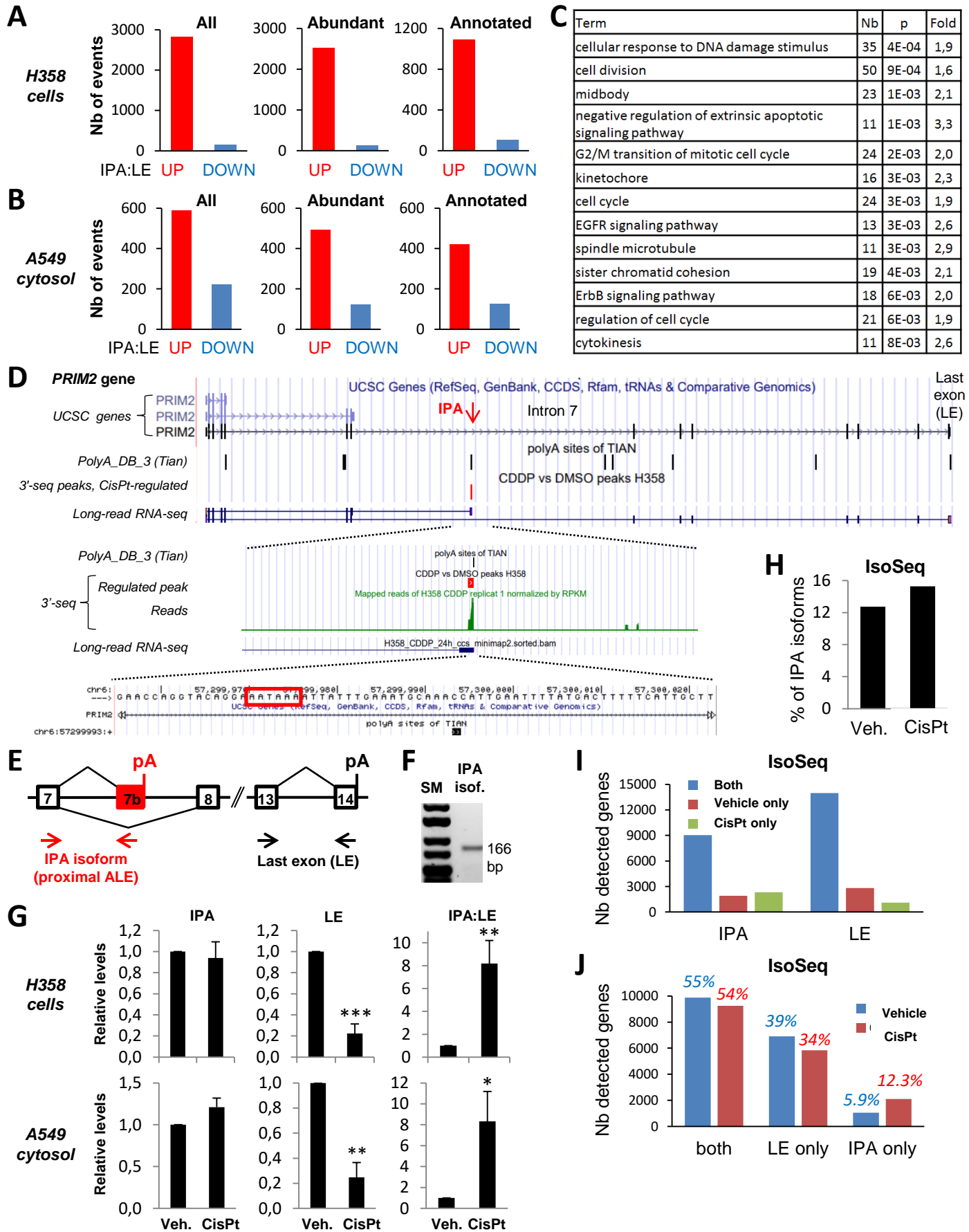
**Figure 4: The *PHF20* and *PRKAR1B* 5'UTR-IPA isoforms are translated into microproteins.** **A-B**, Visualization of our 3'-seq, total-RNA-seq and long-read RNA-seq data, as well as Ribo-seq and mass spectrometry data from OpenProt, for the 5' part of the *PHF20* and *PRKAR1B* genes in the UCSC genome browser. The regulated IPA site and the annotated translation initiation site (ATG) are shown. **C**, RT-qPCR analysis of the *PHF20* and *PRKAR1B* IPA and LE isoforms and of the *SCARNA18* lncRNA in sucrose gradient fractions prepared from the cytosol of cisplatin-treated A549 cells. The number of ribosomes in the different fractions is indicated. **D**, Sequence of the predicted microproteins encoded by sORFs in the *PHF20* and *PRKAR1B* 5'UTR-IPA isoforms. The OpenProt accession number of the sORFs is given. **E-F**, Analysis of cells transfected with expression plasmids encoding GFP and containing or not the *PRKAR1B* 5'UTR-IPA isoform with a Flag tag in frame with sORF#2. **E**, Anti-Flag immunofluorescence, GFP fluorescence, and DAPI staining (nuclei) in HeLa cells grown in the presence or absence of 10  $\mu$ M Cisplatin for 16 hours. The bars represent 10  $\mu$ m. **F**, Western blot analysis of HEK-293T cells.

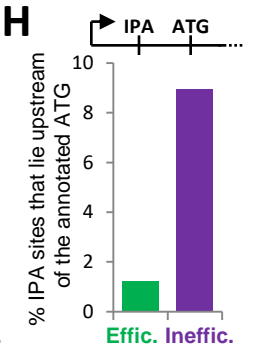
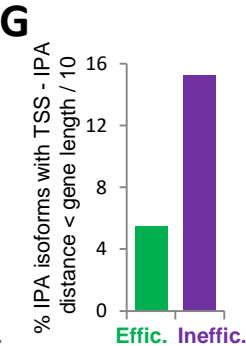
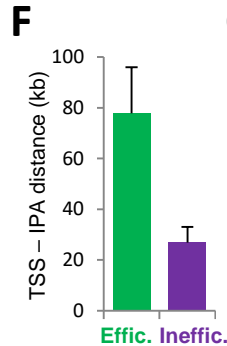
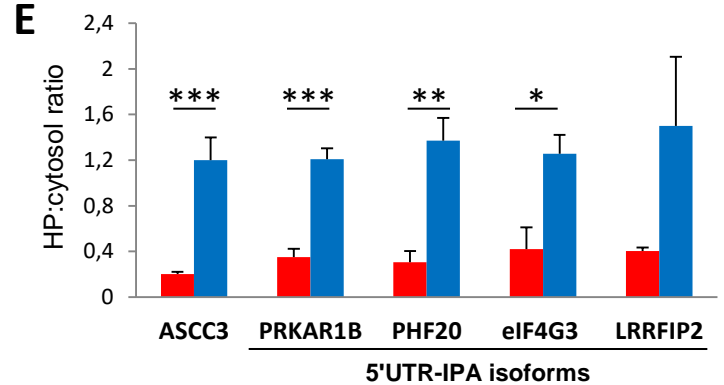
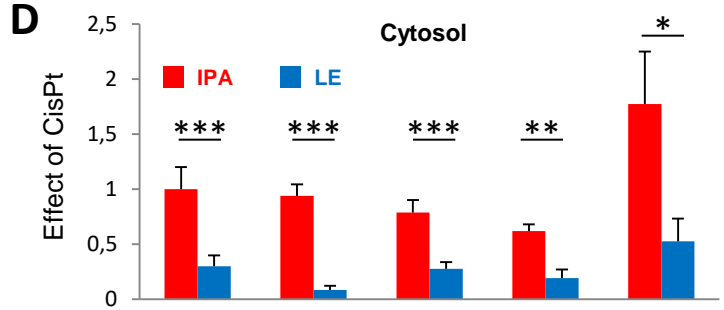
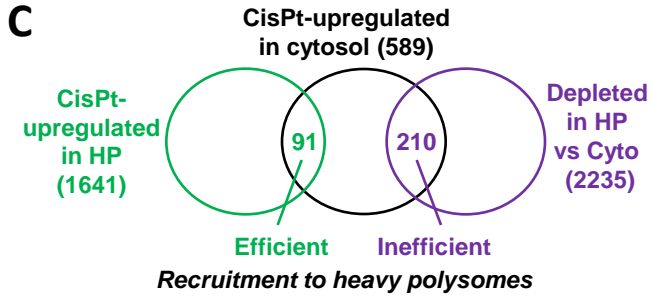
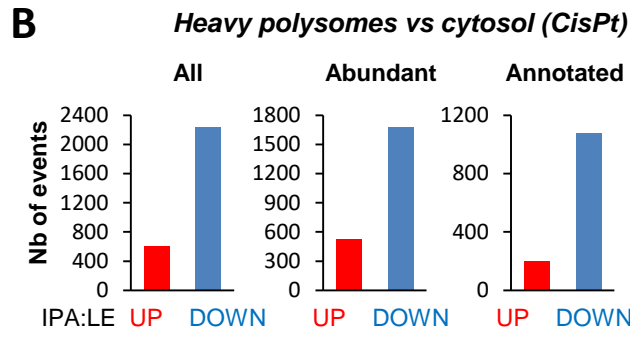
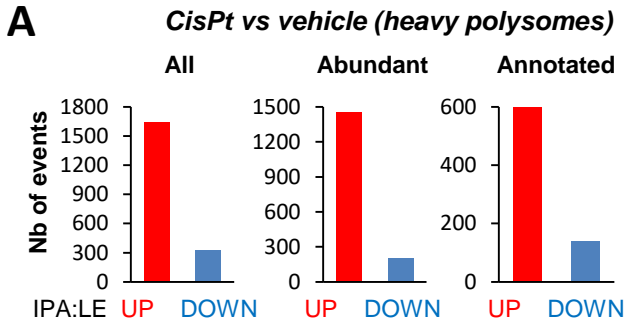
**Figure 5: The cisplatin survival phenotype of the *PRKAR1B* 5'UTR-IPA isoform is attributable to its small ORF#2.** **A**, Schematics of CRISPR-edited and control (Ctrl) cell lines. **B-E**, Analyses of the Ctrl,  $\Delta$ IPA, and mATG cell lines. The graphs use arbitrary units. Statistical tests use Ctrl cells for comparison. **B**, RT-qPCR analysis of expression levels of the *PRKAR1B* IPA isoform, normalized to TBP mRNA levels. **C**, Western blot quantification of the PRKAR1B canonical protein, normalized to GAPDH protein levels. **D-E**, WST1 analysis of cell viability in the presence or absence of cisplatin for 24 hours. **D**, Vehicle-treated cells. **E**, Cell viability with cisplatin was normalized to cell viability without cisplatin.

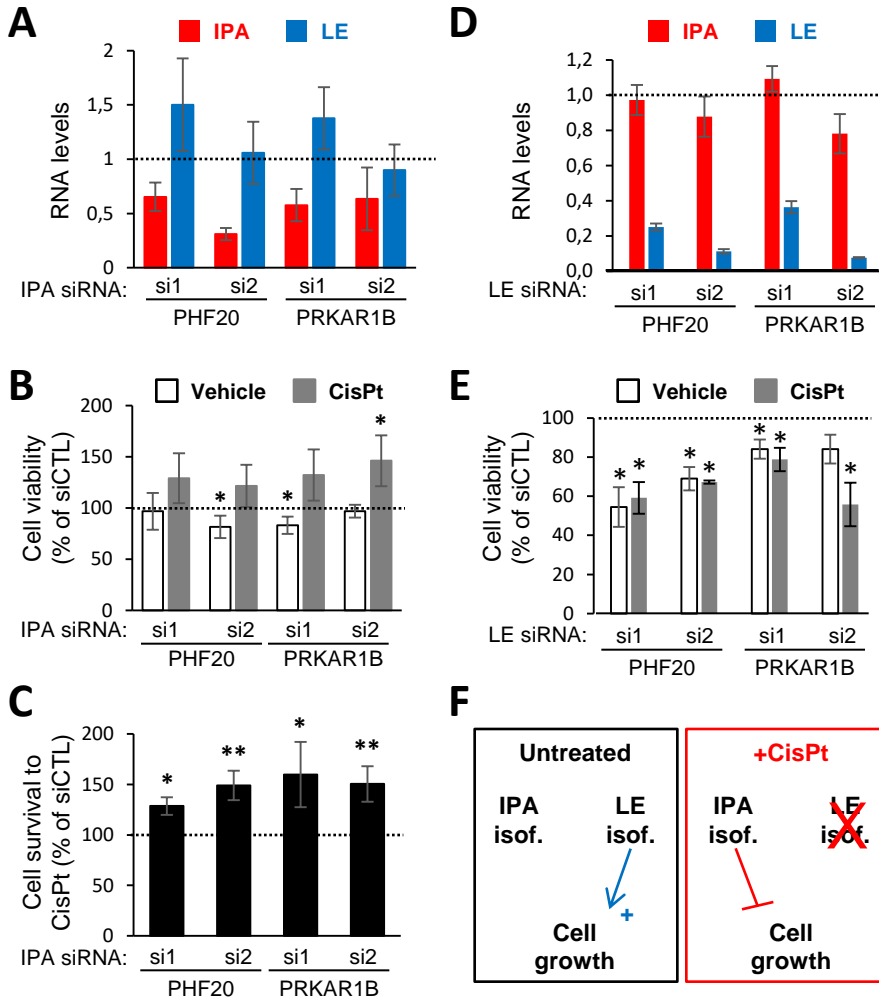
**Figure 6: Genome-wide identification of putative miP-5'UTR-IPA isoforms upregulated by cisplatin relative to LE isoform.** **A**, Model of the novel genetic paradigm of miP-5'UTR-IPA. **B**, Subsets of cisplatin-upregulated 5'UTR-IPA isoforms (from A549 and H358 cells) that contain sORFs supported by Ribo-Seq and/or mass spectrometry datasets. **C**, Gene Ontology terms enriched in the set of 156 miP-5'UTR-IPA candidate genes identified in B. Analysis with the DAVID software. Shown are functions that are enriched by at least 2-fold, with an adjusted p value below 0.05, and with at least 10 genes. Nb, number of genes. Fold, fold enrichment.

**Figure 7: Cisplatin effect on the IPA:LE ratio is mediated by decreased transcription processivity and depends on FANCD2 and senataxin.** **A**, Analysis of IPA and LE transcripts by long-read RNA-seq (Iso-Seq, from Fig. 1H) in H358 cells treated with cisplatin or vehicle for 24 hours. Median length of IPA, LE and overall transcripts detected in the presence or absence of cisplatin, either without or with introns taken into account. **B**, Distance (showing median and SD) from transcription start site (TSS) to polyA site of IPA and matched LE isoforms. Comparison between IPA isoforms that were either non-regulated or upregulated by cisplatin in 3'-seq analyses (from Fig. 1A). **C**, Analysis of transcription processivity by total-RNA-seq in H358 cells treated with cisplatin or vehicle for 24 hours. For each gene, processivity was assessed as the ratio of read number in the last versus first intron (3':5' ratio). Median length and

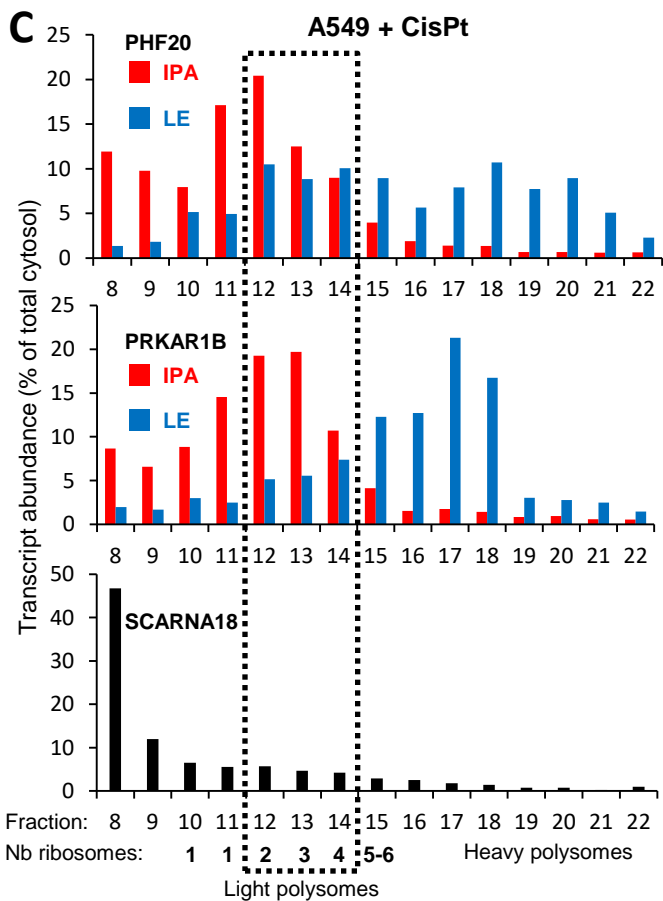
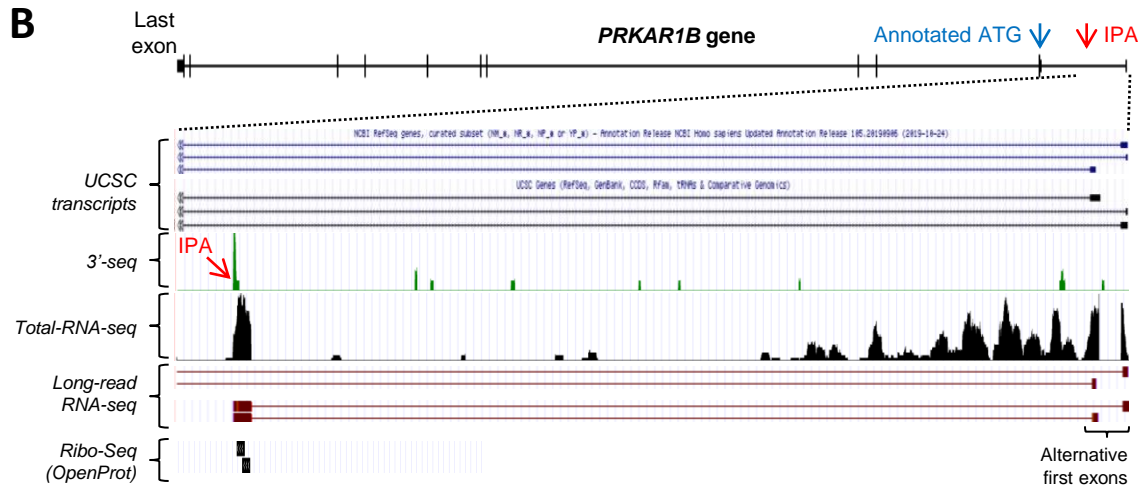
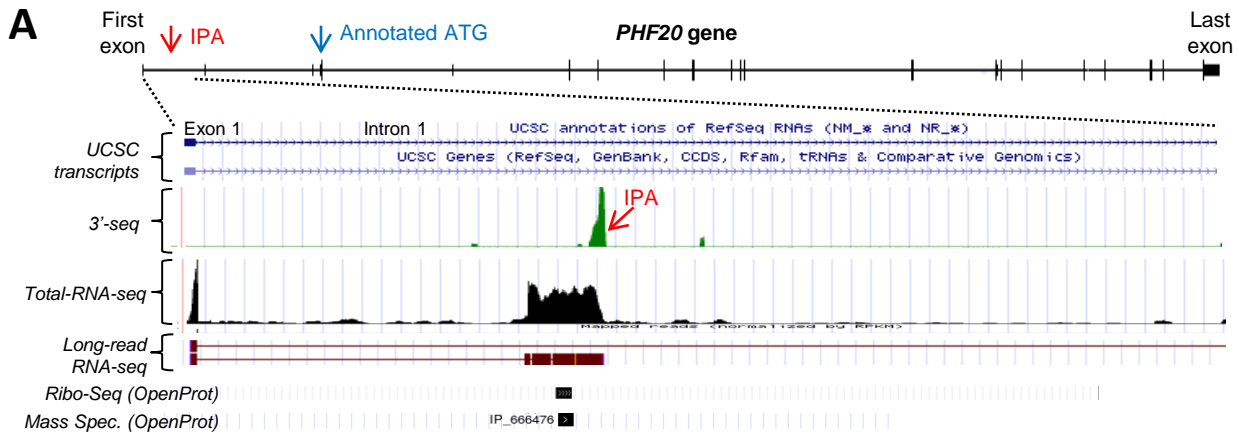
number of genes, whose processivity was either up-, down- or non-regulated by cisplatin. **D**, Enrichment of IPA:LE isoform ratio upregulation in genes with processivity inhibition in response to cisplatin; and *vice versa*, enrichment of processivity inhibition in genes with IPA:LE isoform ratio upregulation. The regulation of IPA:LE isoform ratio and of processivity come from 3'-seq and total-RNA-seq analyses, respectively. **E-H**, A549 cells were transfected with siRNAs targeting the indicated factors or a control siRNA (siCTL) and were treated with cisplatin or vehicle for 8 hours. In G-H, *FANCD2* was depleted using a pool of the two siRNAs used in E-F. **E** and **G**, RT-qPCR analysis of the siRNA-targeted factors. **F** and **H**, RT-qPCR analysis of the fold-effects of cisplatin on the 5':3' ratio (that is, the inverse of processivity) and on the IPA:LE isoform ratio for the *PRKAR1B* and *PHF20* genes. **I**, Model of the molecular mechanism, by which cisplatin upregulates the ratio of (miP-5'UTR-)IPA to LE isoforms in long genes.









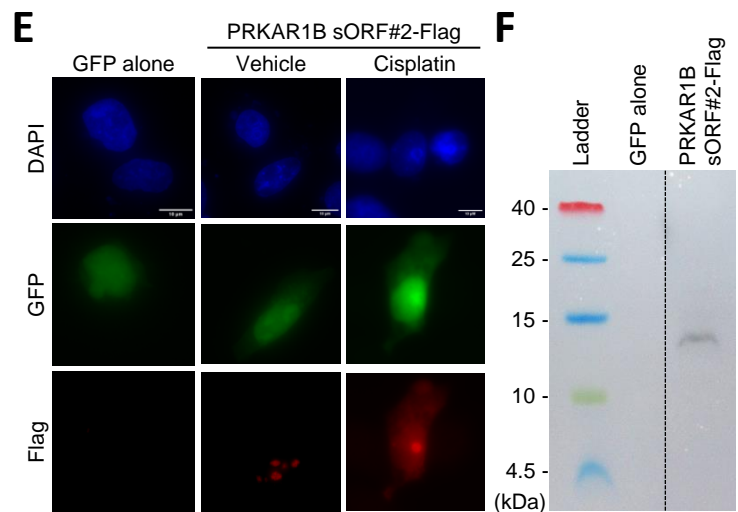


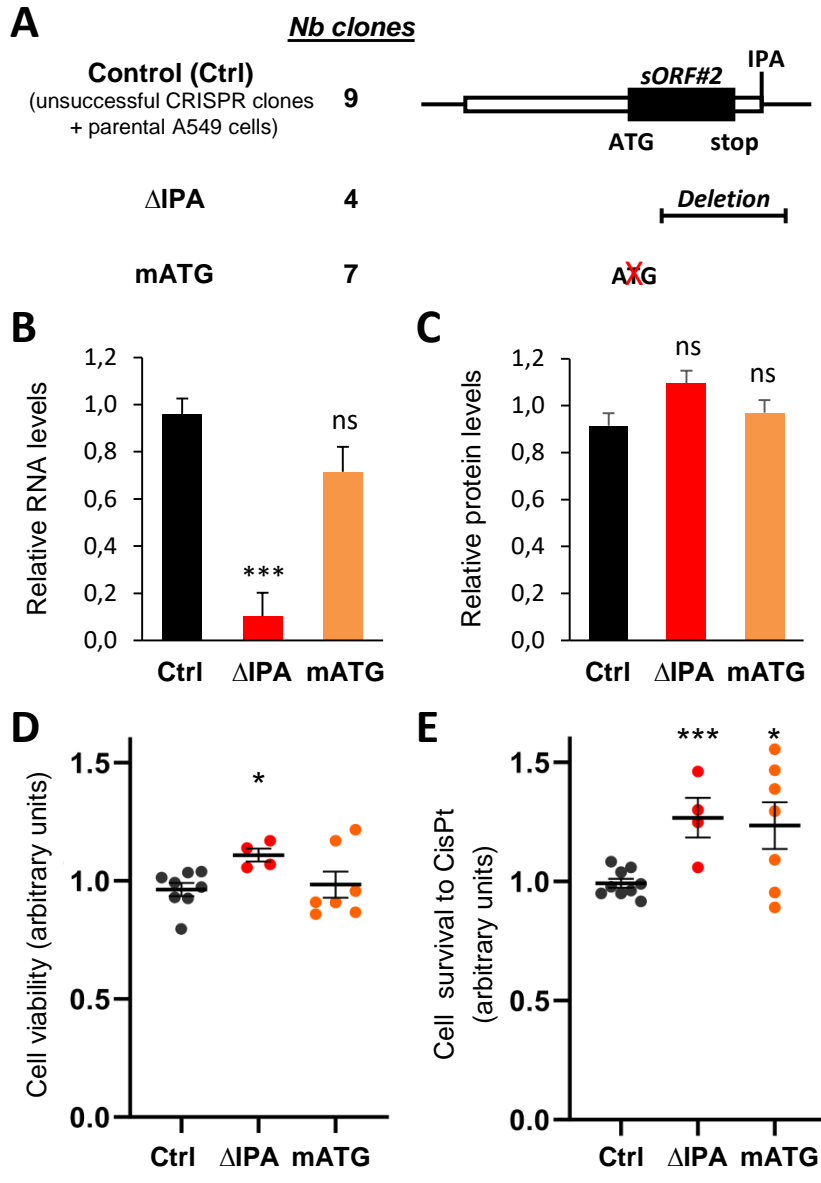
**D**

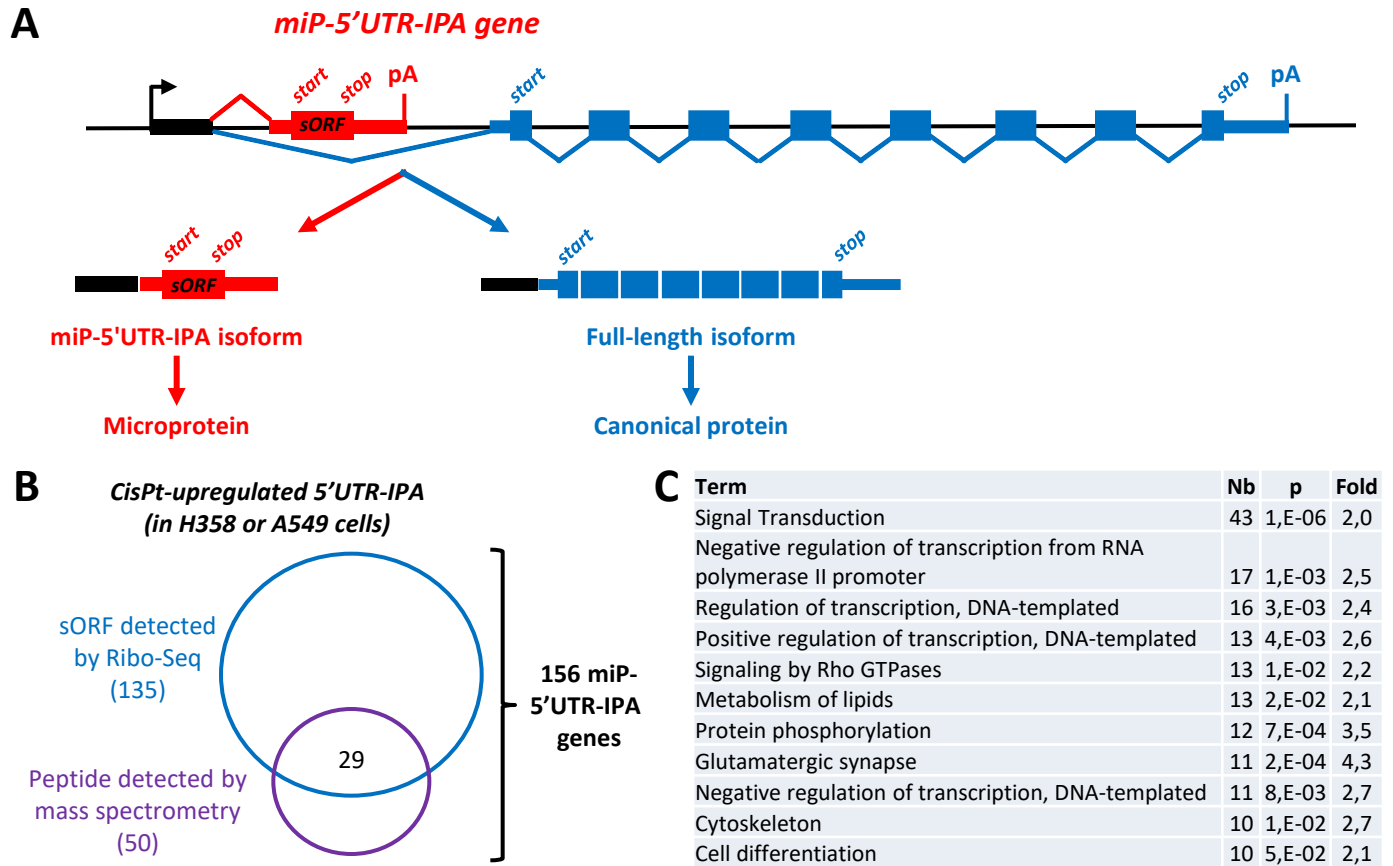
**PHF20 sORF#1 (IP\_666476):**  
MFAKHLLLKDAEGFNDTPRNLRLMGPMEKCSLSVGT

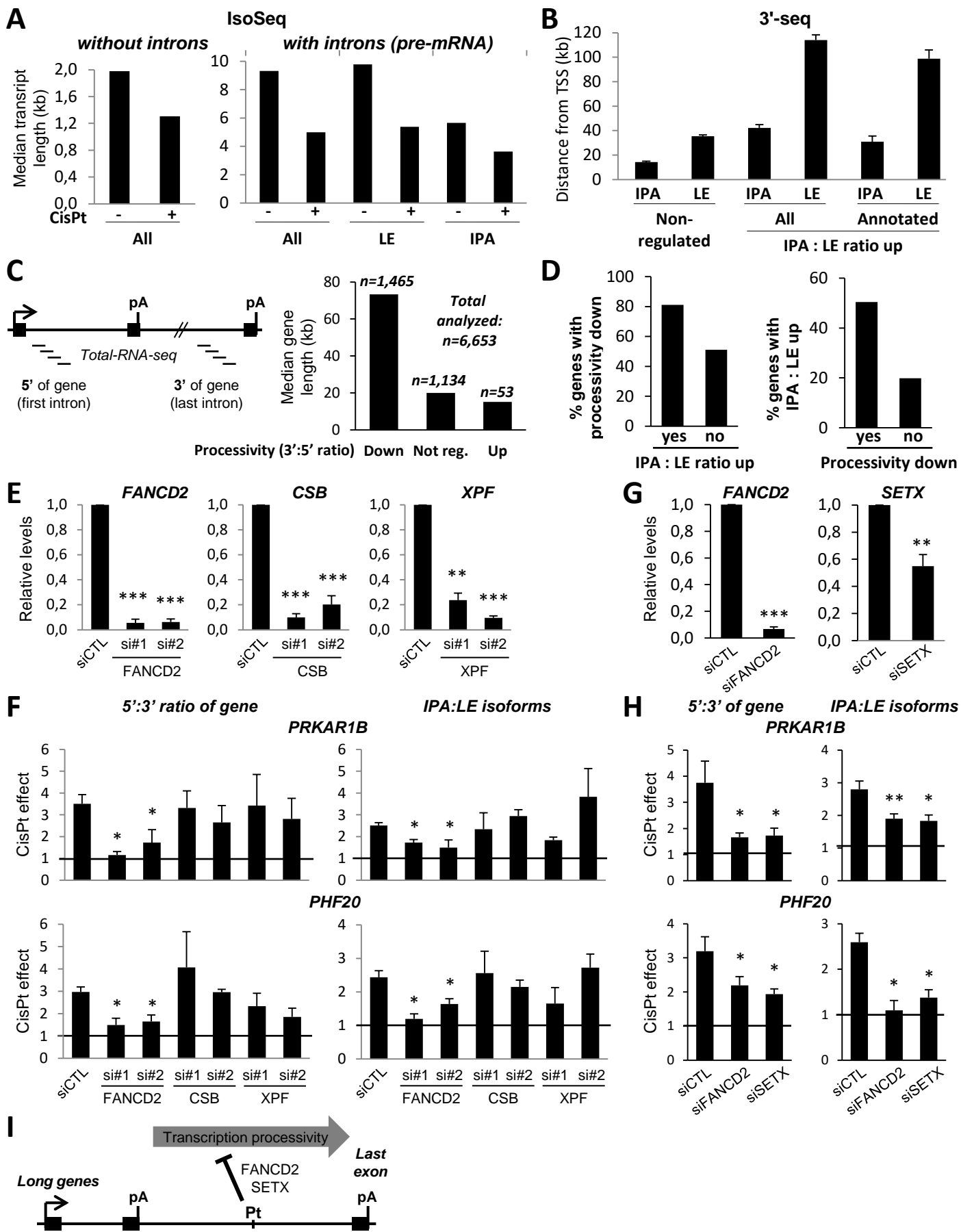
**PRKAR1B sORF#1 (IP\_590577):**  
MQLAGEGSGTAPRKPTYRSHLKPRAHAASPVAAA

**PRKAR1B sORF#2 (IP\_590576):**  
MLQPPQWPEQLNKASLFPFCTPENSCVGLFAMQSE





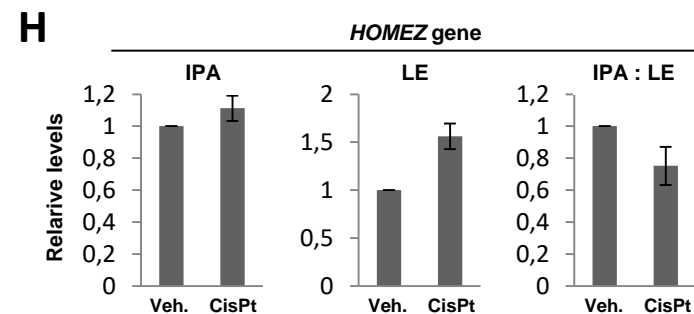
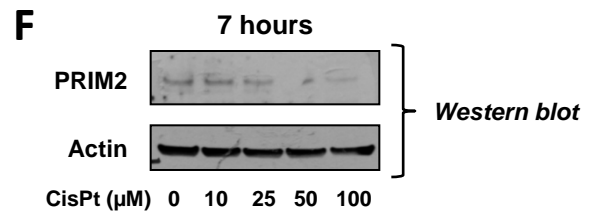
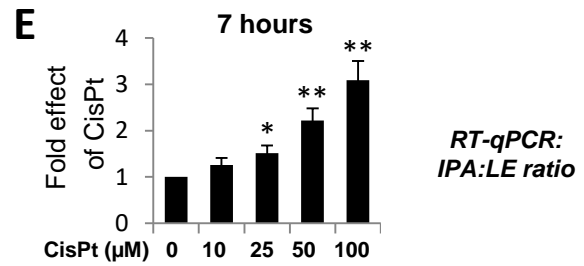
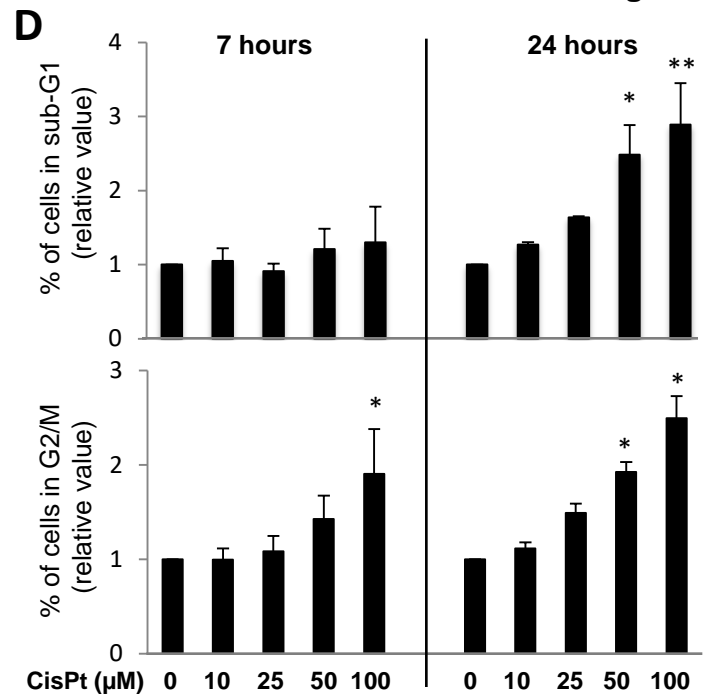
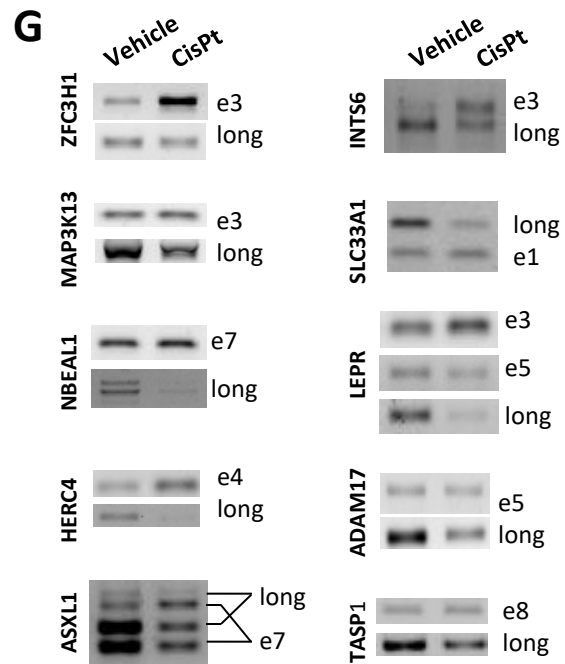
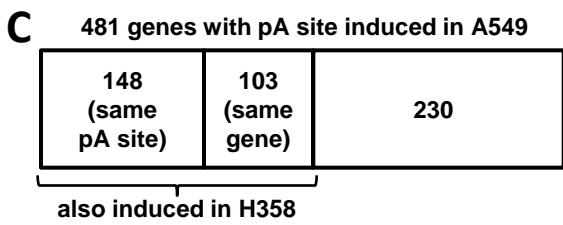
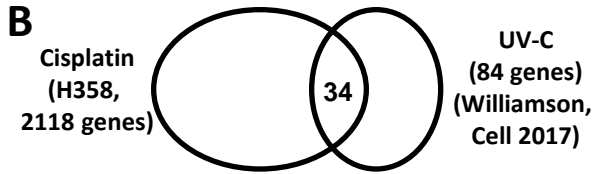
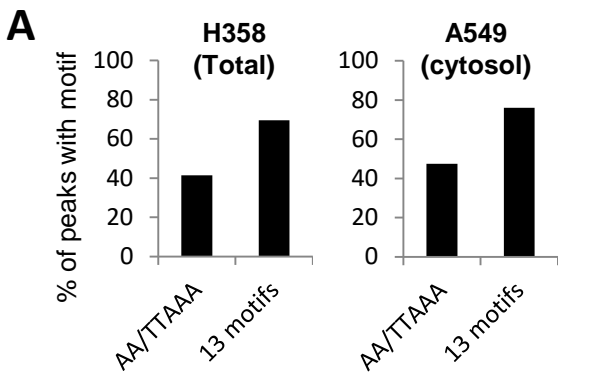




# **SUPPLEMENTARY DATA**

**SUPPLEMENTAL FIGURES S1-S5**

**SUPPLEMENTAL TABLES S1-S15**



**I** IsoSeq (PacBio)

		Nb of reads (millions)
IPA	Vehicle	0,17
	CisPt	0,17
LE	Vehicle	1,20
	CisPt	0,95

**Figure S1: Cisplatin upregulates the IPA:LE isoform ratio in many genes.**

**A**, Percentage of peaks from our 3'-seq analyses on H358 and A549 cells that have a sequence motif matching known polyadenylation signals, either focusing on the two main motifs (AATAAA and ATTTAAA) or including other identified variants.

**B**, Venn diagram comparing IPA:LE regulation events by cisplatin (this study) and UV-C (Williamson et al., Cell 2017).

**C**, Comparison of IPA:LE isoform ratio regulation by cisplatin in H358 and A549 cells.

**D-F**, Analyses of H358 cells treated with cisplatin at different doses and for different time.

**D**, FACS analysis using propidium iodide, showing cells in G2/M phase of the cell cycle (bottom panels) and in sub-G1 (dead cells, top panels).

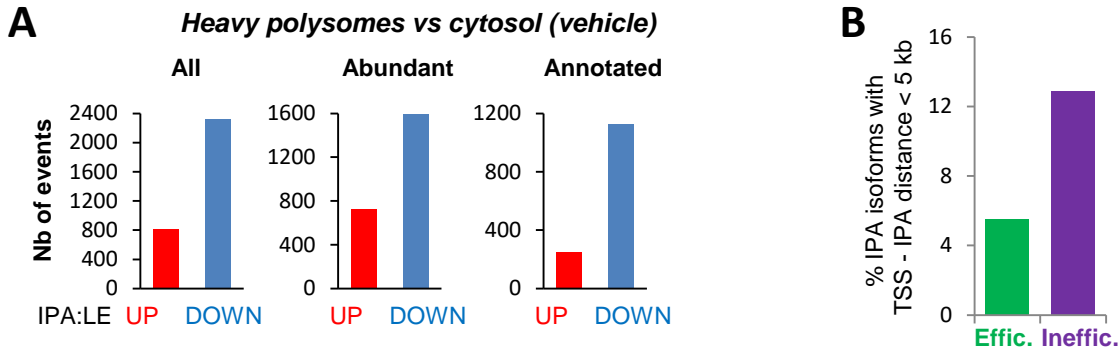
**E**, RT-qPCR analysis of *PRIM2* IPA:LE isoform ratio.

**F**, Western blot analysis of *PRIM2* protein levels.

**G**, RT-PCR validation of IPA:LE isoform ratio regulation by cisplatin in various genes. Long isoforms and the position of proximal ALEs are indicated.

**H**, RT-qPCR analysis of the IPA and LE isoforms of *HOMEZ* in H358 cells grown in the presence of cisplatin (CisPt) or vehicle (Veh.) for 24 hours. Transcript levels without cisplatin were set to 1.

**I**, Long-read RNA-seq data in H358 cells treated with cisplatin or vehicle. Each read is a transcript. Number of reads corresponding to IPA and LE transcripts.



**Figure S2: A subset of IPA isoforms are depleted in heavy polysomes and terminate in the 5' part of genes.**

**A**, 3'-seq analysis of IPA:LE isoform ratio regulation in heavy polysomes *versus* cytosol from vehicle-treated A549 cells.

**B**, Comparison between the efficiently and inefficiently translated subsets of cisplatin-induced IPA isoforms identified in Fig. 2C. Percent of IPA isoforms that are located within 5 kb of the gene's transcription start site (TSS).

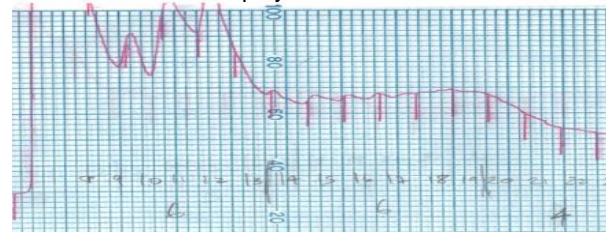
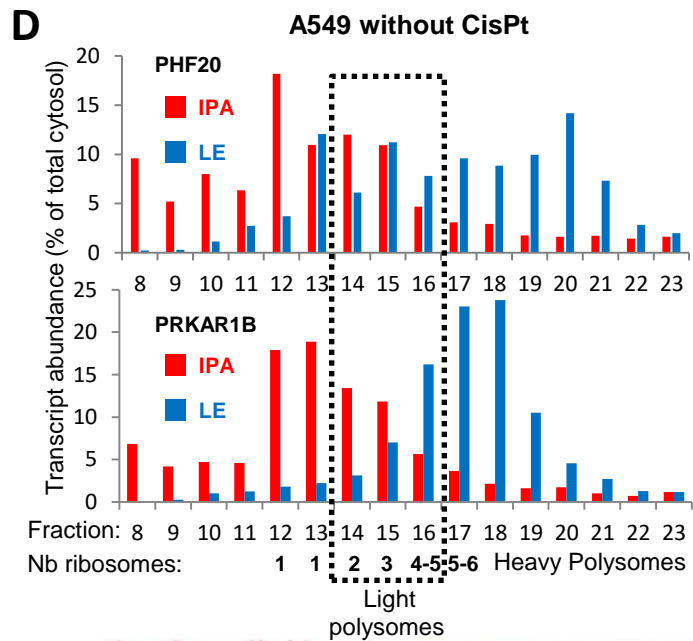
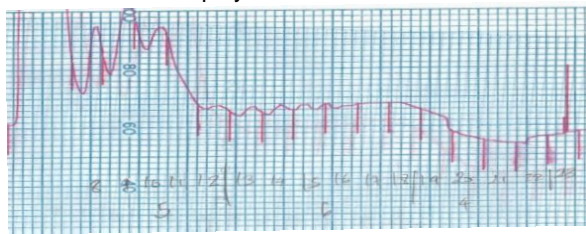
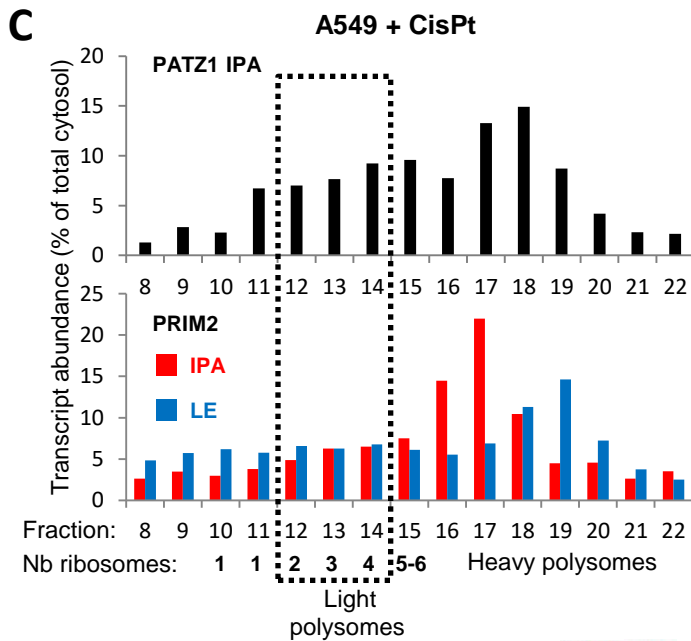


**A** PHF20

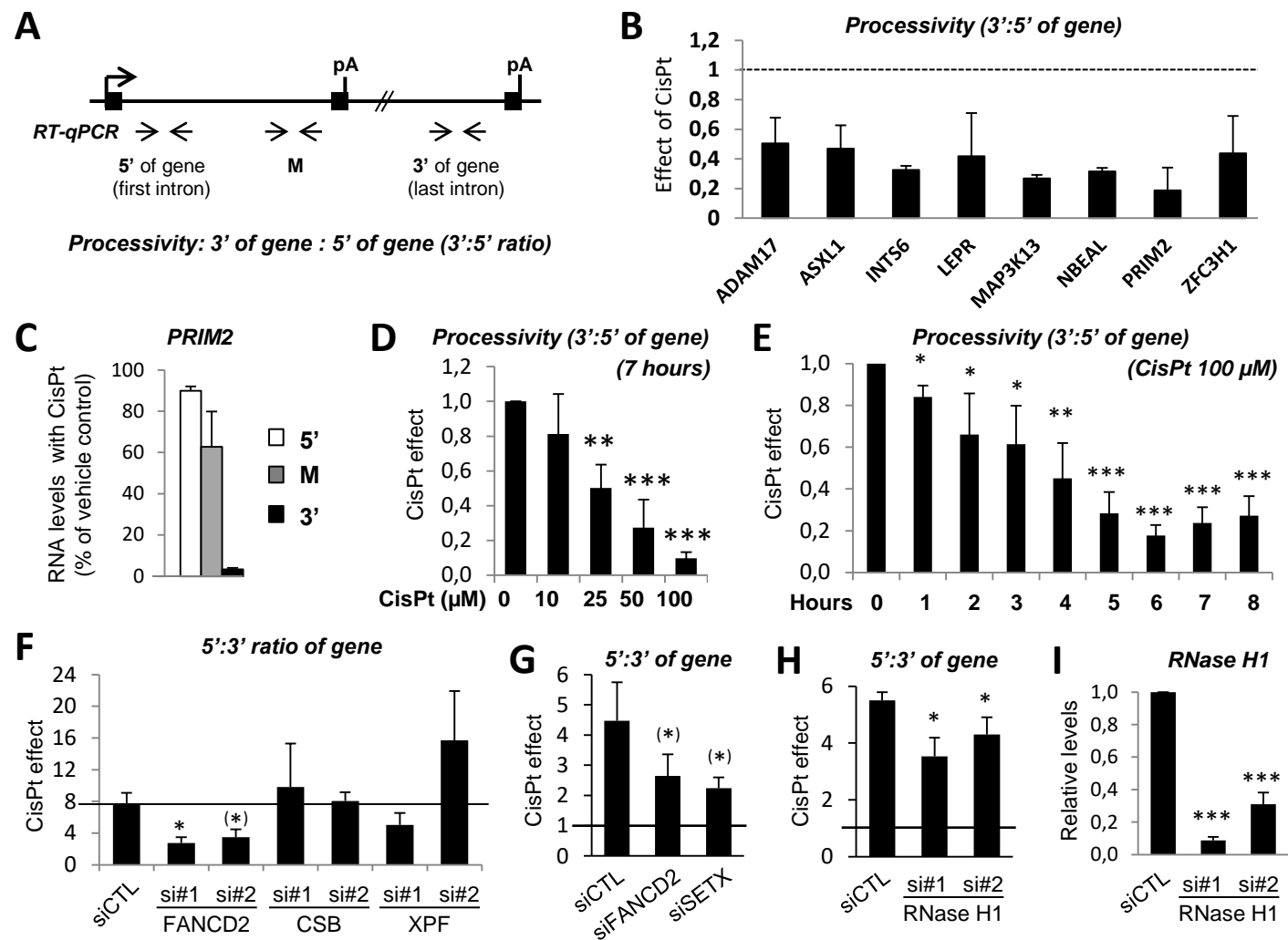
GGGGGTCACGTGGTGCAGAGCACGCAGAGTCTTCTGTGCGGTTGGTCTCTGGAGCGGGCGCAGCCGGAGTGGG  
 GCCGTGAGATCTTATTCAGGGCTTCTCAGCCTGAACCCCAACCAGATTAATTTCTCCACAGCTGGAATTCCAA  
 GTGCTTGGGGATAATGATACCTCTGACCTTTCTTCTTTTGGGAAGTACTTGAGTGTGCAGCTGCATGAGGCCT  
 CAGCAGGAGAGAGATTTTAGGTCCAAGAAGCTATACCAGTAGGACAAGGCAGGAAAATACTACACTTTCAGGA  
 TCAAGCCCCTCTGACTCTCATTTGAAACTGG**ATGTTTGCTAAGCACCTGCTTCTTAAGGATGCCGAGGGATT**  
**TAATGATACTCCAGAAACCTGGAGAGATTAATGGGGCCTATGGAGAAGTGTCTGAACTCAGTGTGGGA**  
**CTTGA**AATAAAATTAACCATTGTCATGTTTTCAGAACAATAAGCTGTTTTATATTTTCATGTGCATGAAAGCCCTAG  
 AACTAAGTTGTATTTCAGAAATGAAATAGATCCCACAGTTAGATGATGTGGCCATTAGGAAGTACCAATT  
 TATAAAATCACTGGAGGTCTGTCTGAGCAGTACCTAATAAAATATAGTATACTGAAAGTGAACAGA

**B** PRKAR1B

CTCAGGGAGCATACGGACGTTGCGCAGTGA**ATGCAGCTGGCTGGGGAAGGCTCAGGAACCGCACCCAGGA**  
**AGCCACCTACAGATCCACCTCAAACCACGACATGCTGCAGCCTCCAGTGGCCAGAGCAGCTTAACA**  
**AAGCATCTTTGTTTCCTTTCTGCACCCAGAAATAGCTGTGTTGGTCTCTTCGCCATGCAGAGTGAATAAA**  
 GAATAAGACAAACACC



**Figure S3: The 5'UTR-IPA isoforms of the *PHF20* and *PRKAR1B* genes contain small ORFs and are translated in light polysomes. A-B**, Sequence of the 5'UTR-IPA isoforms of *PHF20* and *PRKAR1B*. For *PHF20*, the exon 1 is in italics; for *PRKAR1B*, only the part corresponding to the ALE is shown (because there are two alternative first exons). The sORFs detected by OpenProt are indicated in bold, either in red (*PHF20* sORF#1 and *PRKAR1B* sORF#2) or underlined (*PRKAR1B* sORF#1). **C-D**, RT-qPCR analysis of IPA and LE isoforms in sucrose gradient fractions prepared from cytosol of A549 cells. Analysis of the *PATZ1* and *PRIM2* genes in cisplatin-treated cells (**C**) and of the *PHF20* and *PRKAR1B* genes in vehicle-treated cells (**D**). The number of ribosomes in the different fractions is indicated. Typical polysome profiles from these experiments are shown at the bottom.



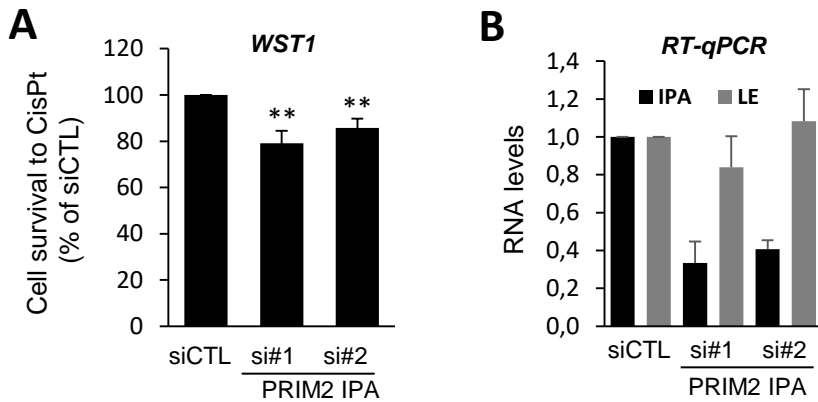
**Figure S4: Repression of transcription processivity by cisplatin in the *PRIM2* gene.**

**A**, Design of transcription processivity analysis by RT-qPCR. For each gene, processivity was assessed as the ratio of RNA levels in the last *versus* first intron (3':5' ratio of gene).

**B-E**, RT-qPCR analyses in H358 cells treated with vehicle or cisplatin. Unless otherwise stated, cisplatin treatment was done for 7 hours **B**, Processivity analysis by RT-qPCR in the indicated genes. **C**, Analysis of *PRIM2* pre-mRNA levels by RT-qPCR using the indicated primers pairs, as defined in panel A. **D-E**, Dose-response (**D**) and time-course (**E**) analyses of processivity repression by cisplatin in the *PRIM2* gene.

**F-G**, RT-qPCR analysis of the fold-effects of cisplatin on the 5':3' ratio (that is, the inverse of processivity) in the *PRIM2* gene in A549 cells transfected with the indicated siRNAs and treated with cisplatin or vehicle for 8 hours. Same cDNA samples as in Fig. 7E-H. (\*) p value between 0.05 and 0.08.

**H-I**, A549 cells were transfected with siRNAs targeting RNase H1 or a control siRNA (siCTL) and treated with cisplatin or vehicle for 8 hours. **H**, RT-qPCR analysis of the fold-effects of cisplatin on the 5':3' ratio in the *PRIM2* gene. **I**, RT-qPCR analysis of RNase H1 levels.



**Figure S5: Effect of *PRIM2* IPA isoform depletion on cell survival to cisplatin.**

**A-B**, Effects of siRNAs targeting the *PRIM2* IPA isoform on H358 cell viability as measured by WST1 assay in the presence of cisplatin normalized to cell viability with vehicle (treatment for 48 hours, panel A) and on IPA and LE isoforms abundance as measured by RT-qPCR (panel B). Two independent siRNAs were used. In panel A, the data obtained for a negative control siRNA (siCTL, about 50% cell survival to cisplatin) were set to 100%.

## SUPPLEMENTAL TABLES

Table S1: PCR primers used in this study.

Table S2: siRNAs used in this study.

Table S3: CRISPR guide RNAs and donor block used in this study.

Table S4: 3'-seq dataset of IPA isoform regulation by cisplatin in H358 cells.

Table S5: Dataset of transcription processivity analysis.

Table S6: 3'-seq dataset of IPA isoform regulation by oxaliplatin in H358 cells.

Table S7: 3'-seq dataset of IPA isoform regulation by cisplatin in A549 cytosol.

Table S8: 3'-seq dataset of IPA isoform regulation by cisplatin in A549 polysomes.

Table S9: 3'-seq dataset of IPA isoform regulation in polysomes *versus* cytosol of cisplatin-treated A549 cells.

Table S10: 3'-seq dataset of IPA isoform regulation in polysomes *versus* cytosol of vehicle-treated A549 cells.

Table S11: Cisplatin-upregulated IPA isoforms that are either efficiently (a) or inefficiently (b) recruited to heavy polysomes.

Table S12: Small ORFs identified by OpenProt in PHF20 and PRKAR1B 5'UTR-IPA isoforms.

Table S13: Small ORFs identified by sORFs.org in the PHF20 5'UTR-IPA isoform.

Table S14: Genes with 5'UTR-IPA isoforms upregulated by cisplatin.

Table S15: Genes with miP-5'UTR-IPA isoforms upregulated by cisplatin.

## VI.1.2 Conclusion

Dans cette étude, nous avons donc observé une nouvelle régulation du cisplatine sur la ratio IPA:LE des isoformes notamment due à son effet sur la processivité transcriptionnelle. Des isoformes 5'UTR-IPA ont notamment étaient montrées comme induites par le cisplatine se retrouvant exportées dans le cytoplasme mais sont mal recrutées dans les fractions de polysomes lourds. Deux de ces isoformes dans les gènes *PHF20* et *PRKAR1B* présentent une fonction dans la survie des cellules cancéreuses traitées au cisplatine. Après analyse sur fraction de polysome, ces isoformes se sont révélées comme engagées dans des fractions de polysomes légers. L'isoforme IPA de *PRKAR1B* s'est révélée comme codant une microprotéine. La délétion de l'IPA et la mutation de l'ATG de la microprotéine de *PRKAR1B* en CRISPR révèlent que l'isoforme IPA et plus précisément la traduction de la sORF sur celle-ci médierait le phénotype de survie au cisplatine. Cette étude montre ici une nouvelle fonction des isoformes ARN issues d'IPA dans la 5' UTR comme source de microprotéines.

## VI.2 Le blocage de U1 snRNA sensibilise les cellules au cisplatine en augmentant les dommages à l'ADN

### VI.2.1 Introduction

U1snRNP joue un rôle majeur dans la répression de l'IPA. Dans une étude, il était indiqué que cet effet était enrichi dans les gènes de DDR (Oh et al. 2017). Une autre étude a observé que l'induction de la mort cellulaire par les UV-C peut être prévenue par surexpression de U1snRNA (Devany et al. 2016). Par ailleurs, plusieurs études ont montré que l'inhibition ou la déplétion de CDK12 augmente l'IPA dans des gènes de DDR (notamment HR) et augmente la sensibilité de cellules cancéreuses à des génotoxiques (voir introduction). Durant ma thèse, j'ai étudié l'impact du blocage de U1 snRNA sur la sensibilité au cisplatine dans des lignées cellulaires de cancer du poumon NSCLC en lien avec la régulation des gènes de DDR. Cet article sera soumis pour publication après acceptation du premier article ci-dessus. J'ai entièrement contribué à l'élaboration de ce papier hormis les analyses bioinformatiques du séquençage réalisées par une bioinformaticienne de mon équipe.

## U1 snRNA blockade sensitizes lung cancer cells to cisplatin by increasing DNA damage

Alexandre DEVAUX<sup>1-3</sup>, Céline M. LABBE<sup>1-3</sup>, Stéphan VAGNER<sup>1-3</sup>, Martin DUTERTRE<sup>1-3\*</sup>

<sup>1</sup> Institut Curie, Université PSL, CNRS UMR3348, INSERM U1278, 91400 Orsay, France

<sup>2</sup> Université Paris-Saclay, CNRS UMR3348, INSERM U1278, 91400 Orsay, France

<sup>3</sup> Equipe Labellisée Ligue Nationale Contre le Cancer

\* Correspondence: martin.dutertre@curie.fr

### ABSTRACT

Genotoxic agents, such as cisplatin that is a DNA-crosslinking anticancer drug, regulate the ratio of intronic polyadenylation (IPA) to last-exon (LE) transcripts in many genes. Conversely, multiple genes involved in the DNA damage response (DDR), especially homologous recombination, are regulated at this level. The U1 small nuclear RNA (snRNA) widely represses IPA, but little is known about its effects on the DDR and cell sensitivity to genotoxic agents. Here, we show that U1 snRNA blockade using an antisense oligonucleotide (U1-AMO) decreases cancer cell survival to cisplatin, by increasing cisplatin-induced DNA damage. 3'-seq analysis indicates that U1 snRNA blockade represses full-length mRNA expression of genes involved in the nucleotide-excision repair and Fanconi anemia pathways, that are involved in the repair of cisplatin-DNA crosslinks. This 3'-seq analysis also reveals that moderate doses of U1-AMO and cisplatin upregulate the IPA:LE isoform ratio in distinct sets of genes, and that U1-AMO prevents cisplatin effects on the IPA:LE ratio in a large subset of genes. Altogether, these data extend the recently emerged crosstalk between IPA and the DDR, and suggest that U1 snRNA targeting may be used to sensitize cancer cells to genotoxic agents.

**KEYWORDS:** Intronic polyadenylation – U1 snRNA – DNA-damaging agent – Platinum compound

### Highlights

- U1 snRNA blockade sensitizes cells to cisplatin
- U1 snRNA blockade increases cisplatin-induced DNA damage
- U1 snRNA blockade represses genes of the NER and Fanconi anemia repair pathways
- U1 snRNA blockade partially prevents cisplatin effect on the IPA:LE isoform ratio

## INTRODUCTION

The U1 small nuclear RNA (snRNA), as part of the U1 small nuclear ribonucleoprotein (snRNP), is a core component of the spliceosome. U1 snRNA hybridization with pre-messenger RNA (pre-mRNA) 5' splice sites mediates their recognition, that is required for splicing. The U1 snRNP also has splicing-independent activities, including stimulation of transcription initiation (1, 2) and elongation (3). In addition, U1 snRNA binding to pre-mRNA can impact the use of a nearby cleavage and polyadenylation site (polyA site) (4, 5). Cell transfection with an antisense morpholino oligonucleotide (U1-AMO) targeting the 5' splice site-binding region of the U1 snRNA increases the use of intronic polyadenylation (IPA) sites (6–10). IPA sites are sometimes referred to as premature cleavage and polyadenylation (PCPA) sites, although a subset of them are used to generate transcript isoforms with alternative last exons. As a result of PCPA, U1-AMO decreases the synthesis of full-length mRNAs in thousands of genes, as shown by RNA-seq analysis of nascent RNA (10).

Genotoxic agents -including both radiations and drugs- can be used to kill cancer cells through the induction of DNA lesions. The so-called DNA damage response (DDR) comprises the recognition of various types of DNA lesions, their repair by multiple pathways, the activation of cell-cycle checkpoints, and other processes, leading eventually to either cell survival or cell death (11, 12). Among genotoxic anticancer agents is cisplatin, which is widely used in several cancer types such as non-small lung cancer. Cisplatin induces both intra- and inter-strand DNA crosslinks, that are repaired mainly by the nucleotide-excision repair (NER) and Fanconi anemia (FA) pathways, respectively (13).

A reciprocal crosstalk between IPA and the DDR has emerged in recent years (14). On the one hand, the IPA to last-exon (LE) isoform ratio is widely regulated by genotoxic agents (15–19), with mainly upregulation in the case of ultraviolet-C irradiation (16, 17) and cisplatin treatment (18). For both of these agents, IPA:LE regulation events are enriched in long genes and are mediated at least in part by a repression of transcription processivity (16, 18). In addition, following ultraviolet-C irradiation, U1 snRNA levels are decreased and its overexpression prevents several IPA:LE regulation events and apoptosis induction (17).

On the other hand, genes that are regulated at the level of IPA are enriched in DDR genes. Indeed, depletion or inhibition of cyclin-dependent kinase 12 (CDK12) increases the IPA:LE isoform ratio and decreases transcription processivity in multiple genes involved in the DDR –especially in homologous recombination– and sensitizes cells to genotoxic agents, presumably due to the IPA-mediated regulation of DDR genes (20–25). However, CDK12 interacts with and phosphorylates many proteins in addition to the RNA polymerase II large subunit (23, 25, 26) and, through the phosphorylation of a translation factor, has been involved in translational regulation of a DDR protein (CHK1) and of multiple mitosis proteins controlling mitotic chromosome stability (27).

In contrast with CDK12, the effects of the U1 snRNA on DNA damage and cell sensitivity to genotoxic and/or anticancer agents have been little studied and were therefore the subject of the present study.



## RESULTS

### **U1 snRNA blockade sensitizes lung cancer cells to growth inhibition by cisplatin**

Because U1 snRNA widely regulates IPA, which is enriched in DDR genes, we tested the ability of U1-AMO to sensitize cancer cell lines to a genotoxic agent. For this, we used two cell lines (A549 and H358) derived from non-small cell lung cancer, which is commonly treated with cisplatin in the clinic. Cells were transfected with either CTRL-AMO or U1-AMO, grown in culture, then treated or not with cisplatin, and finally assayed for cell viability by WST1 assay. In the absence of cisplatin, U1-AMO decreased the growth of both cell lines, when compared to a control oligonucleotide (CTRL-AMO), with a nice dose-dependent effect in H358 cells (Fig. 1A, vehicle). This is consistent with previous studies in other cell lines.

To look at cell survival to cisplatin, we used cisplatin doses that had mild effects on cell viability, and for each dose of AMO we calculated the percent of cell viability with cisplatin/AMO combination *versus* AMO alone. In both cell lines, when compared to CTRL-AMO, prior transfection with U1-AMO decreased the percent of cells surviving to a subsequent cisplatin treatment, and there was a nice dose-dependent effect of U1-AMO in H358 cells (Fig. 1B, 'cisplatin survival'). Conversely, the percent of cells that did not survive to cisplatin was increased by U1-AMO, when compared to CTRL-AMO (Fig. 1B, 'cisplatin mortality'). Thus, in both cell lines, U1-AMO decreased cell survival to cisplatin.

Finally, to test whether U1-AMO may also sensitize cells that acquired resistance to cisplatin, we used the A549-R cell line that was previously obtained by selection of A549 cells with increasing doses of cisplatin (28). As in parental A549 cells, U1-AMO decreased the growth of A549-R cells more strongly in the presence of cisplatin, decreased cisplatin survival, and increased cisplatin mortality (Fig. 1C-D).

### **U1 snRNA blockade prevents the ability of cisplatin to inhibit cell cycle progression**

Because genotoxic agents block cell cycle progression, and because little is known on U1-AMO effects on the cell cycle, we analyzed cell distribution in cell cycle phases following treatment with U1-AMO and cisplatin either alone or in combination, as above. For this, we carried out FACS analysis using propidium iodide and bromodeoxyuridine (BrdU) to measure DNA content and synthesis, respectively (Fig. 2A-B). In both cell lines, the G0-G1 to G2-M ratio, which reflects the ability of cells to progress from G2-M to G1 phase, was moderately decreased by U1-AMO alone (which may explain in part its growth-inhibitory effects) and was strongly decreased by cisplatin alone, but U1-AMO at least partially prevented this cisplatin effect (Fig. 2C-D, left panels). In addition, cisplatin increased the S to G0-G1 ratio in both cell lines, suggesting a replication blockade, but U1-AMO prevented these cisplatin effects at least partially (Fig. 2C-D, right panels). Thus, the ability of U1 snRNA blockade to enhance the growth-inhibitory effect of cisplatin (Fig. 1B-C) is not due to an enhancement of the cell cycle effects of the drug. Instead, U1 snRNA blockade prevents (at least partially) the ability of cisplatin to inhibit cell cycle progression.

### **U1 snRNA blockade increases cisplatin-induced DNA damage**

Then, we tested whether cell sensitization to cisplatin by U1-AMO might be due to decreased repair of cisplatin-induced DNA damage. For this, we measured DNA damage by immunofluorescence of the  $\gamma$ H2AX histone mark. In both A549 (Fig. 3A-B) and H358 cells (Fig. 3C-D),  $\gamma$ H2AX staining (quantitated as either mean fluorescence intensity or foci number) was increased by cisplatin and was further enhanced by prior treatment with U1-AMO (in a dose-dependent manner), which had little effect by itself on  $\gamma$ H2AX staining. These data indicate that U1-AMO decreases the repair of cisplatin-induced DNA damage.

### **U1 snRNA blockade represses genes of the NER and FA repair pathways**

To determine the mechanism, by which U1-AMO decreases the repair of cisplatin-induced DNA damage, we analyzed the impact of U1-AMO on gene expression by using 3'-seq (RNA-seq focused on the 3'-end of polyadenylated transcripts), because U1-AMO was shown to increase PCPA and to decrease transcription processivity in long genes (10). Treatment of A549 cells with U1-AMO at 500 and 1500 nM led to a decrease of last-exon expression levels by at least 2-fold in 250 and 3118 genes, respectively, with 22 and 535 genes repressed by at least 4-fold (Fig. 4A). The 535 repressed genes were enriched in genes involved in cell cycle (most notably, mitotic spindle organization and G1/S and G2/M transitions of mitotic cell cycle) and DNA repair (Fig. 4B). The most enriched DNA-repair pathways were the NER (especially, transcription-coupled [TC]-NER) and interstrand cross-link (especially FA) repair pathways (Fig. 4B). Remarkably, these are the two main repair pathways of cisplatin-DNA crosslinks. We validated by RT-qPCR that U1-AMO decreased in a dose-dependent manner the expression of the last exon of genes involved in the FA (e.g., *ATR*, *FANCC*, *FAN1*, *NEIL3*), TC-NER (e.g., *ERCC3*, *CDK7*) or both pathways (e.g., *ERCC4*) in A549 cells (Fig. 4C, left panels) and independently of cisplatin treatment (data not shown). Finally, most of these genes were also repressed by U1-AMO in H358 cells (Fig. 4C, right panels). Altogether, these data indicate that U1 snRNA blockade decreases the expression of the full-length mRNAs of many genes involved in the FA and TC-NER pathways.

### **U1 snRNA blockade partially prevents cisplatin effect on the IPA:LE isoform ratio**

Because both U1-AMO (10) and cisplatin (18) were shown to regulate IPA isoforms in many genes, we also used our 3'-seq analyses in A549 cells to compare their effects, alone and in combination, on the IPA:LE isoform ratio on a genome-wide scale. As expected, both agents alone regulated IPA:LE events in many genes, with mainly upregulation events, when compared to control cells that received CTRL-AMO and solvent (Fig. 5A-B, lines 1, 2 and 6). U1-AMO upregulated IPA:LE events in a dose-dependent manner, with 2088 events in 1475 genes at 500 nM, and 29677 events in 5776 genes at 1500 nM (Fig. 5A-B, lines 2 and 3). Meanwhile, treatment with cisplatin alone at 5  $\mu$ M upregulated 2881 IPA:LE events in 1396 genes, when compared to control cells (Fig. 5A-B, line 1). Comparison of the genes with IPA:LE events upregulated by U1-AMO (500 nM) or cisplatin alone revealed an overlap

of only one third (Fig. 5C), suggesting that they have different specificity. Consistently with this observation, cisplatin treatment did not appear to regulate U1 snRNA levels (Fig. 5D), which is in contrast with its downregulation upon ultraviolet-C irradiation (17).

The combined treatment with U1-AMO at 500 nM and cisplatin upregulated about the same number of IPA:LE events (2949 events in 1560 genes) as each agent alone (Fig. 5A-B, lines 1-3). In addition, 64% of the genes with IPA:LE events upregulated by the combined agents were also regulated by at least one agent alone (Fig. 5C). Thus, the effects of U1-AMO and cisplatin on IPA:LE isoforms are less than additive on a global scale. (Of note, we cannot exclude that U1-AMO and cisplatin may cooperate to regulate some IPA:LE events, such as those that seem to be upregulated only by the combined agents; Fig. 5C). Finally, there were 2.4 times less (1224 *versus* 2881) IPA:LE events upregulated by cisplatin relative to vehicle if cells were preably transfected with 500 nM U1-AMO instead of CTRL-AMO (Fig. 5A, comparing lines 1 and 4). Conversely, there were 2.1 times more downregulated events (252 *versus* 121). These data suggest that U1-AMO prevents a large subset of cisplatin effects on IPA:LE events.

## DISCUSSION

A reciprocal crosstalk between IPA and the DDR has emerged: the IPA:LE isoform ratio is widely regulated in cell response to genotoxic agents (15–19); conversely, CDK12 inhibition increases the IPA:LE isoform ratio in homologous recombination genes and –presumably thereby– sensitizes cells to genotoxic anticancer agents (20–25). The U1 snRNA widely represses IPA (6–10) and was involved in cell response to ultraviolet-C irradiation (17), but little is known about its effects on DNA damage and on cell sensitivity to genotoxic and/or anticancer agents. This study shows that U1 snRNA blockade can sensitize lung cancer cells to cisplatin, a DNA-crosslinking anticancer agent, by increasing IPA and decreasing full-length transcripts in the NER and FA repair genes, and increasing in turn drug-induced DNA damage. We also find that U1 snRNA blockade partially prevents cisplatin effect on the IPA:LE isoform ratio on a genome-wide scale.

Our data show that U1 snRNA blockade can sensitize lung cancer cells to cisplatin (Fig. 1). This finding is consistent with the previous observation that U1 snRNA overexpression prevents apoptosis induction by ultraviolet-C irradiation, but the underlying mechanism was unknown (17). Our data show that U1 snRNA blockade increases cisplatin-induced DNA damage (Fig. 3), presumably due to the repression of multiple genes involved in the NER and FA pathways (Fig. 4), that play key roles in the repair of DNA crosslinks induced by this drug (13). Previous studies showed that CDK12 inhibition leads to the repression of genes involved in the homologous recombination pathway of DNA repair, that is particularly important for the repair of double-strand breaks and at stalled replication forks (20–25). In the future, it would be interesting to compare the effects of U1 snRNA and CDK12 targeting on the various pathways of DNA repair and DDR signaling, and on cancer cell sensitivity to various DNA-damaging agents.

Our data also suggest differences in the way cisplatin and ultraviolet-C treatment regulate the IPA:LE isoform ratio, with respect to the role of the U1 snRNA in this regulation. Indeed, a previous study showed that ultraviolet-C irradiation leads to a decrease of U1 snRNA levels and that its overexpression prevents IPA:LE isoform up-regulation in tested genes (17). In contrast, in the case of cisplatin treatment, U1 snRNA levels do not seem to be regulated (Fig. 5D) and U1 snRNA blockade partially prevents cisplatin effect on the IPA:LE isoform ratio on a genome-wide scale (Fig. 5A). Nevertheless, for both agents, the up-regulation of the IPA:LE isoform ratio is due at least in part to a decrease of transcription processivity, which decreases LE isoforms (16, 18).

Altogether, this study increases our understanding of IPA isoform regulation by genotoxic agents and of DDR gene regulation at the IPA level, and identifies the U1 snRNA as a relevant target for cancer cell sensitization to genotoxic chemotherapy.

## **MATERIAL AND METHODS**

### **Cell culture, electroporation, and treatment**

Cells were cultured in RPMI-1640-Glutamax (H358) or DMEM (A549, A549-R0) medium (GibcoBRL, Life Technologies, Cergy Pontoise, France) supplemented with 10% (v/v) heat-inactivated fetal calf serum (GibcoBRL), in 5% CO<sub>2</sub> at 37°C. Cells were electroporated using Cell Line Nucleofector™ Kit T (VCA-1002, Lonza). Following manufacturer's recommendations, 1 million cells were mixed with 100 µL of buffer T kit with Supplement 1. The mix was put into the electroporation cuve provided by Lonza, with the indicated amounts of the U1-AMO (GGTATCTCCCCTGCCAGGTAAGTAT) and/or Standard Control morpholino oligonucleotides (Ctrl-AMO, Gene Tools, Philomath, Oregon, USA). The cuve was inserted into Nucleofector™ II and the A-023 (H358) or X-001 (A549/A549-R0) program. For H358 or A549/A549-R0 respectively, 300 000 or 100 000 cells/mL were plated into 96-well plates, at 300 000 or 200 000 cells/mL for 12 well plates, and at 250 000 cells/mL for 6-well plates. For H358 or A549/A549-R0 respectively, 24 or 8 hrs after electroporation, cells were treated with either cisplatin (Selleckchem, Euromedex, Souffelweyersheim, France) or vehicle (DMSO) for 48 or 24 hrs.

### **WST1 assay and FACS analysis**

For cell growth and viability analysis, cells were seeded in 96-well plates. Following cisplatin treatment, cell viability was assayed in triplicate wells using WST1 (Sigma Aldrich, France) according to the manufacturer's instructions. For FACS analysis, cells were treated 2 hrs with BrdU 20 µM in a 6-well plate at the end of the indicated treatment. Cells were harvested then fixed with 70% cold ethanol and put at -20°C for storage. H358 and A549 cells were processed using FITC-BrdU Flow Kits according to the manufacturer's instructions (BD Biosciences, San Jose, CA, USA cat.51–2354AK). Flow cytometric analysis of 10000 cells was performed on a FACScan flow cytometer (BD Biosciences) and data were recovered using the CellQuest software (BD Biosciences).

### **Immunofluorescence analysis**

Transfected cells grown on Marienfeld Superior cover glasses were washed twice with ice cold PBS, fixed using 4% paraformaldehyde in PBS during 20 min, washed three times with PBS, permeabilized for 10 min with PBS-0.1% Triton-X, washed twice with PBS, blocked for 5 min with PBS-5% BSA, incubated with Anti- $\gamma$ H2AX (phospho S139) rabbit antibody (ab26350, Abcam, 1:2000) at RT for 1 hour, washed twice with PBS, blocked again, incubated with F(ab')<sub>2</sub>-rabbit anti-mouse IgG (H+L) cross-adsorbed secondary antibody, Alexa Fluor 488 (A-21206, ThermoFischer, 1:1000) in PBS-5% BSA for 1 hour, washed twice with PBS, incubated at RT with PBS containing 4',6-diamidino-2-phenylindole (DAPI) 0.1  $\mu$ g/mL, and washed twice with PBS. Cover glasses were mounted on slides in PBS, glycerol 15%, 1,4-diazabicyclo-(2.2.2) octane (DABCO, Sigma) 100 mg/ml. For microscopy, acquisition was done with an exposure time of 50 ms for Trans-DIC, 20 ms for DAPI 405 and 50 ms for FITC. Z stacking was used for imaging DAPI and FITC with 31 panels for a range of  $\pm 5$   $\mu$ m around the acquisition point.

### **RNA extraction**

RNA from whole cells or fractionated cell lysates were extracted with TRIzol Reagent (TRIzol-LS for polysomes; ThermoFischer Scientific) according to the manufacturer's instructions, and 1  $\mu$ l of GlycoBlue (ThermoFischer Scientific) was added for RNA precipitation. RNA from whole cells was treated with DNase I (TURBO DNA-free, ThermoFischer Scientific). RNA samples were quantified using a Nanodrop 2000 spectrophotometer (ThermoFischer Scientific). For sequencing, RNA samples were analyzed using an RNA 2100 Bioanalyzer (Agilent).

### **RT-(q)PCR**

Reverse transcription was performed on RNA using SuperScript III Reverse Transcriptase (ThermoFisher Scientific) and oligo-dT primers (except for U1 snRNA, for which random primers were used). Quantitative PCR (qPCR) was performed using Power SYBR Green PCR Master Mix (ThermoFisher Scientific) on a CFX96 Real-Time PCR Detection System (BioRad). Primer sequences are given in Suppl. Table S1.

### **3'-seq experiments and bioinformatic analysis**

3'-seq libraries were prepared with QuantSeq 3' mRNA-Seq Library Prep Kit REV for Illumina (Lexogen) using 250 ng (for polysomal RNA) or 500 ng (for all other samples) of DNase I treated RNA (n=3 for each condition) following manufacturer's instructions. Purified libraries were quantified with Quant-iT Picogreen dsDNA kit (Thermo Fisher Scientific) and run on Experion automated electrophoresis system (BIO-RAD). Pooled libraries were quantitated by qPCR (KAPA Library Quantification Kits Illumina Platforms, Roche), diluted to 12 pM, and subjected to single-end, 50 bp sequencing using the HiSeq 2500 machine (Illumina).

For each sample, raw reads were trimmed to remove uninformative nucleotides due to primer sequences. Trimmed reads of 25 bp or more were aligned on the human reference genome (hg19) using Bowtie 2 (version 2.2.5) (29). Only reads with a mapping quality score (MAPQ) of 20 or more were retained (SAMtools version 1.1) for downstream analysis (30). Reads were then clustered along the genome using BEDTools (version 2.17.0) (31), allowing a maximum distance of 50 bp and a minimum number of 5 reads per peak. Peaks with a stretch of 6 consecutive As (or 8 As out of 9 nucleotides) within 50 bp downstream were filtered out, as they are likely due to internal priming of Oligo(dT). Overlapping peaks from all samples of the compared conditions were merged to define a common set of genomic windows corresponding to poly(A) sites. To annotate peak location within genes, gene coordinates were obtained on the basis of overlapping RefSeq transcripts with the same gene symbol. Peaks overlapping any intronic region of a gene were classified as intronic poly(A) (IPA) peaks. Peaks overlapping the last exon of a gene were classified as LE peaks. Differential analyses between two conditions were done using three independent biological replicates per condition. To compare the regulation of each IPA to the regulation of the gene's last exon (taken as the sum of the peaks in this exon), we used DESeq2 (version 1.4.5) (32) and the following statistical model:

$$Y_{ij} = \mu + L_i + C_j + (LC)_{ij} + E_{ij}$$

where  $Y_{ij}$  is the normalized counts of peak  $i$  in biological condition  $j$ ,  $\mu$  is the mean,  $L_i$  is the peak localization (IPA or LE),  $C_j$  is the biological condition,  $(LC)_{ij}$  is the interaction between peak localization and biological condition, and  $E_{ij}$  is the residual. P-values and adjusted P-values (Benjamini-Hochberg) were calculated. Data with  $p < 0.05$  are shown. The complete bioinformatics pipeline (3'-SMART package) described above can be freely downloaded at GitHub (<https://github.com/InstitutCurie/3-SMART>) and can be run through a configuration file and a simple command line. Annotated polyadenylation sites were retrieved from the PolyA\_DB 3 and PolyASite 2.0 databases (33, 34). Functional gene annotation analyses were done using the DAVID software (35, 36), using the human genome as a reference.

### **Statistical analyses**

For each experimental analysis, at least three independent experiments were performed. In all bar charts, error bars represent the standard error of the mean (SEM) that is the standard deviation divided by the square root of sample number. A Student's paired t-test was used. Tests were considered significant if  $p < 0.05$ .

### **DATA AVAILABILITY**

The datasets generated in this study will be deposited in the UCSC genome browser (private pre-publication link) and in the Gene Expression Omnibus repository (GEO). The complete bioinformatics pipeline for 3'-seq analysis of IPA (3'-SMART package) can be freely downloaded at GitHub (<https://github.com/InstitutCurie/3-SMART>).

## FUNDING

This work was supported by grants of the Institut National du Cancer (2015-141), and Institut Curie (ICGex) to MD, by a grant of Ligue Nationale Contre le Cancer (équipe labellisée) to SV, by fellowships from Ministère de l'Enseignement Supérieur et de la Recherche, Université Paris Saclay and Institut Curie to AD. The ICGex NGS platform of the Institut Curie is supported by grants ANR-10-EQPX-03 (Equipex) and ANR-10-INBS-09-08 (France Génomique Consortium) from the Agence Nationale de la Recherche ("Investissements d'Avenir" program), by the Canceropôle Ile-de-France and by the SiRIC-Curie program - SiRIC Grant INCa-DGOS-4654.

## CONFLICT OF INTEREST

SV is a scientific cofounder of Ribonexus. The other authors declare no competing interests.

## AUTHORS' CONTRIBUTIONS

Experiments were performed by AD under the supervision of MD. Bioinformatics analyses were performed by CML under the supervision of MD. MD wrote the paper.

## REFERENCES

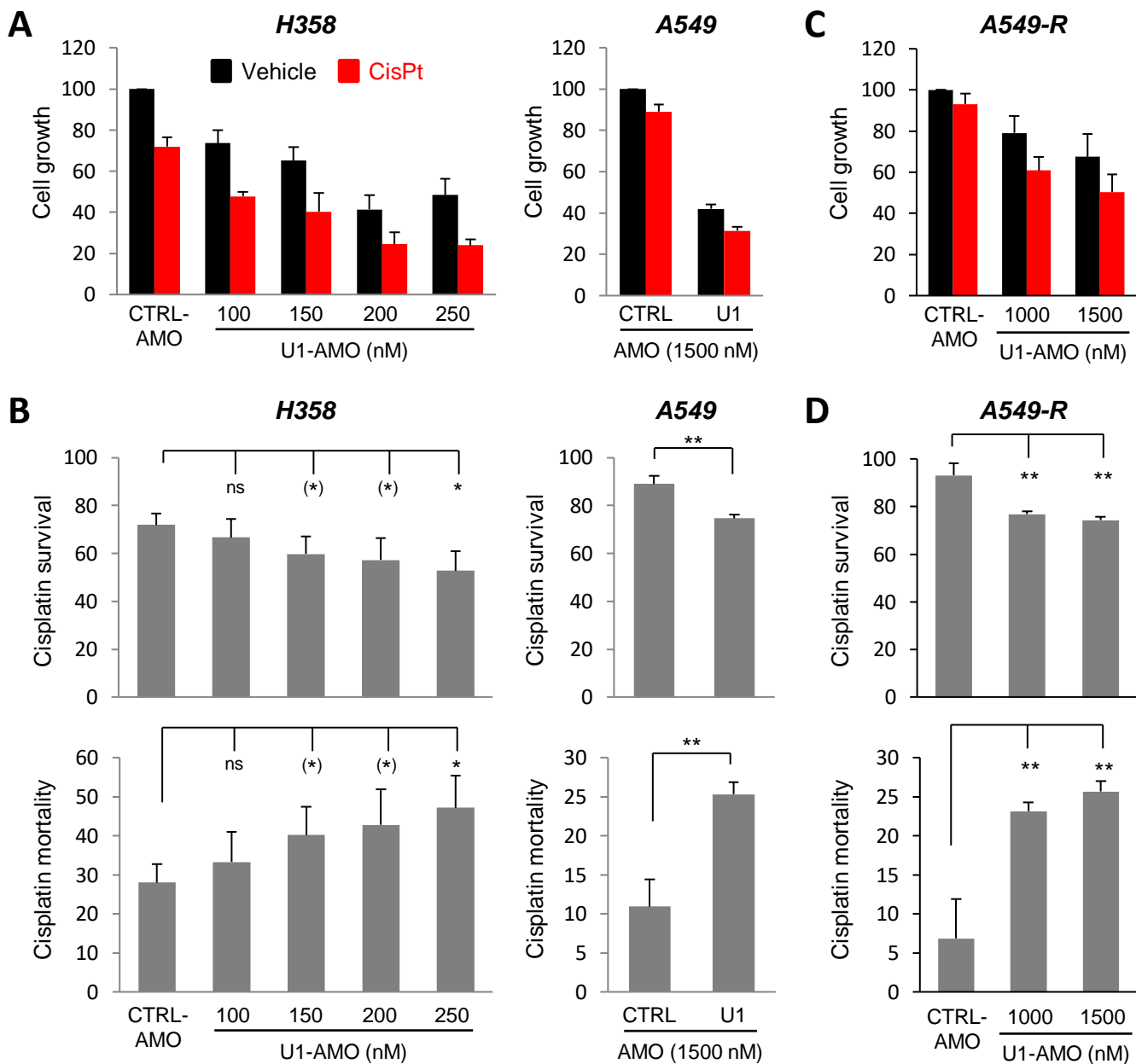
1. Kwek, K.Y., Murphy, S., Furger, A., Thomas, B., O'Gorman, W., Kimura, H., Proudfoot, N.J. and Akoulitchev, A. (2002) U1 snRNA associates with TFIIH and regulates transcriptional initiation. *Nat Struct Biol*, **9**, 800–805.
2. Damgaard, C.K., Kahns, S., Lykke-Andersen, S., Nielsen, A.L., Jensen, T.H. and Kjems, J. (2008) A 5' splice site enhances the recruitment of basal transcription initiation factors in vivo. *Mol Cell*, **29**, 271–278.
3. Mimoso, C.A. and Adelman, K. (2023) U1 snRNP increases RNA Pol II elongation rate to enable synthesis of long genes. *Mol Cell*, **83**, 1264-1279.e10.
4. Gunderson, S.I., Polycarpou-Schwarz, M. and Mattaj, I.W. (1998) U1 snRNP inhibits pre-mRNA polyadenylation through a direct interaction between U1 70K and poly(A) polymerase. *Mol Cell*, **1**, 255–264.
5. Vagner, S., Rügsegger, U., Gunderson, S.I., Keller, W. and Mattaj, I.W. (2000) Position-dependent inhibition of the cleavage step of pre-mRNA 3'-end processing by U1 snRNP. *RNA*, **6**, 178–188.
6. Kaida, D., Berg, M.G., Younis, I., Kasim, M., Singh, L.N., Wan, L. and Dreyfuss, G. (2010) U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature*, **468**, 664–668.

7. 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.
8. Vorlová, S., Rocco, G., Lefave, C.V., Jodelka, F.M., Hess, K., Hastings, M.L., Henke, E. and Cartegni, L. (2011) Induction of antagonistic soluble decoy receptor tyrosine kinases by intronic polyA activation. *Mol Cell*, **43**, 927–939.
9. Andersen, P.K., Lykke-Andersen, S. and Jensen, T.H. (2012) Promoter-proximal polyadenylation sites reduce transcription activity. *Genes Dev*, **26**, 2169–2179.
10. Oh, J.-M., Di, C., Venters, C.C., Guo, J., Arai, C., So, B.R., Pinto, A.M., Zhang, Z., Wan, L., Younis, I., *et al.* (2017) U1 snRNP telescripting regulates a size-function-stratified human genome. *Nat Struct Mol Biol*, **24**, 993–999.
11. Ciccia, A. and Elledge, S.J. (2010) The DNA damage response: making it safe to play with knives. *Mol Cell*, **40**, 179–204.
12. Dutertre, M., Lambert, S., Carreira, A., Amor-Guéret, M. and Vagner, S. (2014) DNA damage: RNA-binding proteins protect from near and far. *Trends Biochem Sci*, **39**, 141–149.
13. Rocha, C.R.R., Silva, M.M., Quinet, A., Cabral-Neto, J.B. and Menck, C.F.M. (2018) DNA repair pathways and cisplatin resistance: an intimate relationship. *Clinics (Sao Paulo)*, **73**, e478s.
14. Dutertre, M., Sfaxi, R. and Vagner, S. (2021) Reciprocal Links between Pre-messenger RNA 3'-End Processing and Genome Stability. *Trends Biochem Sci*, **46**, 579–594.
15. Dutertre, M., Chakrama, F.Z., Combe, E., Desmet, F.-O., Mortada, H., Polay Espinoza, M., Gratadou, L. and Auboeuf, D. (2014) A recently evolved class of alternative 3'-terminal exons involved in cell cycle regulation by topoisomerase inhibitors. *Nat Commun*, **5**, 3395.
16. Williamson, L., Saponaro, M., Boeing, S., East, P., Mitter, R., Kantidakis, T., Kelly, G.P., Lobley, A., Walker, J., Spencer-Dene, B., *et al.* (2017) UV Irradiation Induces a Non-coding RNA that Functionally Opposes the Protein Encoded by the Same Gene. *Cell*, **168**, 843-855.e13.
17. Devany, E., Park, J.Y., Murphy, M.R., Zakusilo, G., Baquero, J., Zhang, X., Hoque, M., Tian, B. and Kleiman, F.E. (2016) Intronic cleavage and polyadenylation regulates gene expression during DNA damage response through U1 snRNA. *Cell Discov*, **2**, 16013.
18. Devaux, A., Tanaka, I., Cadix, M., Heneman-Masurel, A., Michallet, S., Fouilleul, Q., Chakraborty, A., Labbe, C.M., Fontrodona, N., Claude, J.-B., *et al.* (2023) Intronic polyadenylation isoforms in the 5' part of genes constitute a source of microproteins and are involved in cell response to cisplatin. 10.1101/2023.12.05.569446.
19. Chakraborty, A., Cadix, M., Relier, S., Taricco, N., Alaeitabar, T., Devaux, A., Labbé, C.M., Martineau, S., Heneman-Masurel, A., Gestraud, P., *et al.* (2022) Compartment-specific and ELAVL1-coordinated regulation of intronic polyadenylation isoforms by doxorubicin. *Genome Res*, **32**, 1271–1284.



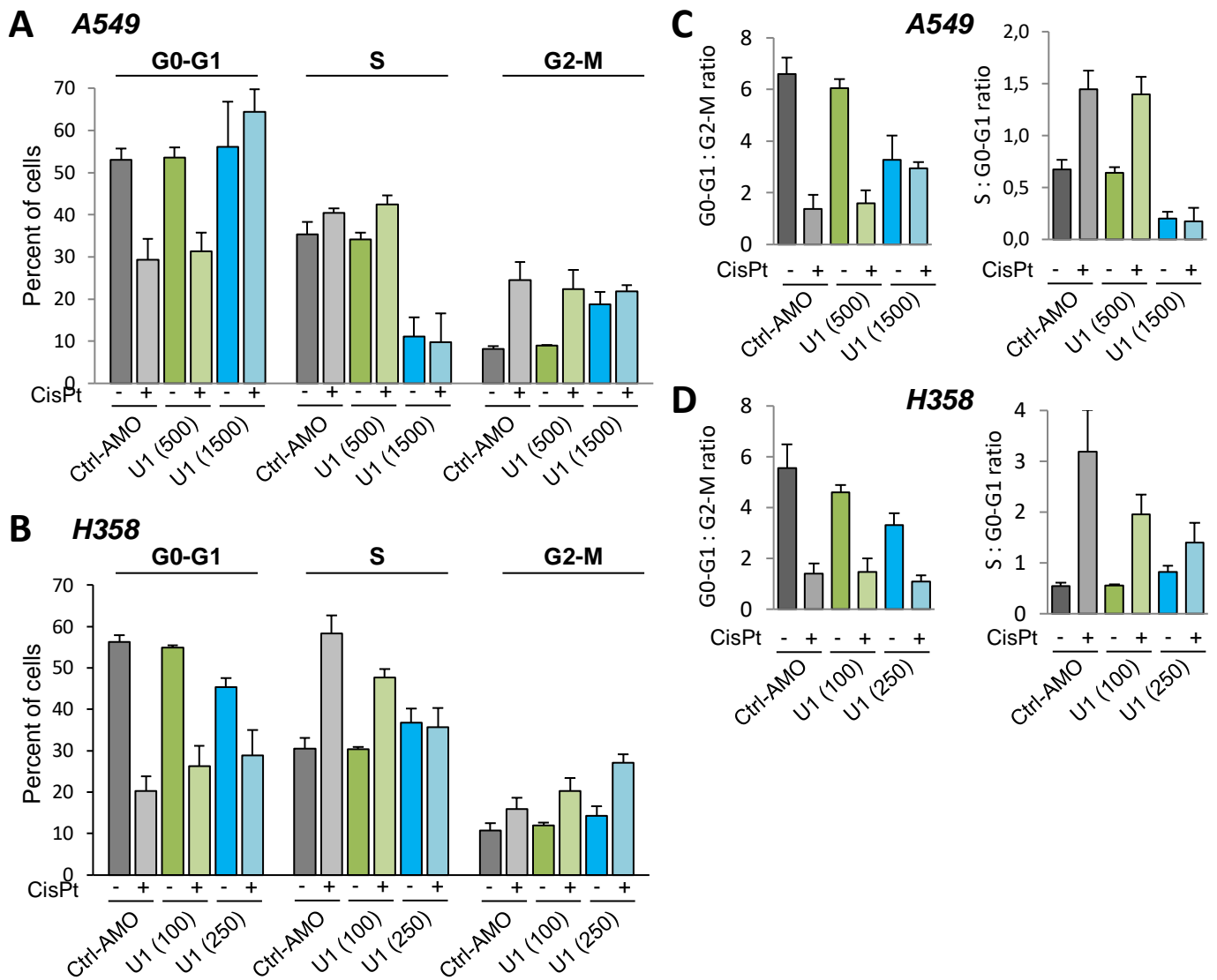
20. Chirackal Manavalan,A.P., Pilarova,K., Kluge,M., Bartholomeeusen,K., Rajecy,M., Oppelt,J., Khirsariya,P., Paruch,K., Krejci,L., Friedel,C.C., *et al.* (2019) CDK12 controls G1/S progression by regulating RNAPII processivity at core DNA replication genes. *EMBO Rep*, **20**, e47592.
21. Dobbury,S.J., Boutz,P.L. and Sharp,P.A. (2018) CDK12 regulates DNA repair genes by suppressing intronic polyadenylation. *Nature*, **564**, 141–145.
22. Fan,Z., Devlin,J.R., Hogg,S.J., Doyle,M.A., Harrison,P.F., Todorovski,I., Cluse,L.A., Knight,D.A., Sandow,J.J., Gregory,G., *et al.* (2020) CDK13 cooperates with CDK12 to control global RNA polymerase II processivity. *Sci Adv*, **6**, eaaz5041.
23. Krajewska,M., Dries,R., Grasseti,A.V., Dust,S., Gao,Y., Huang,H., Sharma,B., Day,D.S., Kwiatkowski,N., Pomaville,M., *et al.* (2019) CDK12 loss in cancer cells affects DNA damage response genes through premature cleavage and polyadenylation. *Nat Commun*, **10**, 1757.
24. Quereda,V., Bayle,S., Vena,F., Frydman,S.M., Monastyrskiy,A., Roush,W.R. and Duckett,D.R. (2019) Therapeutic Targeting of CDK12/CDK13 in Triple-Negative Breast Cancer. *Cancer Cell*, **36**, 545-558.e7.
25. Tien,J.F., Mazloomian,A., Cheng,S.-W.G., Hughes,C.S., Chow,C.C.T., Canapi,L.T., Oloumi,A., Trigo-Gonzalez,G., Bashashati,A., Xu,J., *et al.* (2017) CDK12 regulates alternative last exon mRNA splicing and promotes breast cancer cell invasion. *Nucleic Acids Res*, **45**, 6698–6716.
26. Bartkowiak,B., Yan,C.M., Soderblom,E.J. and Greenleaf,A.L. (2019) CDK12 Activity-Dependent Phosphorylation Events in Human Cells. *Biomolecules*, **9**, 634.
27. Choi,S.H., Martinez,T.F., Kim,S., Donaldson,C., Shokhirev,M.N., Saghatelian,A. and Jones,K.A. (2019) CDK12 phosphorylates 4E-BP1 to enable mTORC1-dependent translation and mitotic genome stability. *Genes Dev*, **33**, 418–435.
28. Michels,J., Vitale,I., Galluzzi,L., Adam,J., Olaussen,K.A., Kepp,O., Senovilla,L., Talhaoui,I., Guegan,J., Enot,D.P., *et al.* (2013) Cisplatin resistance associated with PARP hyperactivation. *Cancer Res*, **73**, 2271–2280.
29. Langmead,B. and Salzberg,S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods*, **9**, 357–359.
30. Li,H., Handsaker,B., Wysoker,A., Fennell,T., Ruan,J., Homer,N., Marth,G., Abecasis,G., Durbin,R., and 1000 Genome Project Data Processing Subgroup (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.
31. Quinlan,A.R. and Hall,I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.
32. Love,M.I., Huber,W. and Anders,S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*, **15**, 550.

33. Herrmann,C.J., Schmidt,R., Kanitz,A., Artimo,P., Gruber,A.J. and Zavolan,M. (2020) PolyASite 2.0: a consolidated atlas of polyadenylation sites from 3' end sequencing. *Nucleic Acids Res*, **48**, D174–D179.
34. Wang,R., Nambiar,R., Zheng,D. and Tian,B. (2018) PolyA\_DB 3 catalogs cleavage and polyadenylation sites identified by deep sequencing in multiple genomes. *Nucleic Acids Res*, **46**, D315–D319.
35. Huang,D.W., Sherman,B.T. and Lempicki,R.A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*, **4**, 44–57.
36. Huang,D.W., Sherman,B.T. and Lempicki,R.A. (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res*, **37**, 1–13.



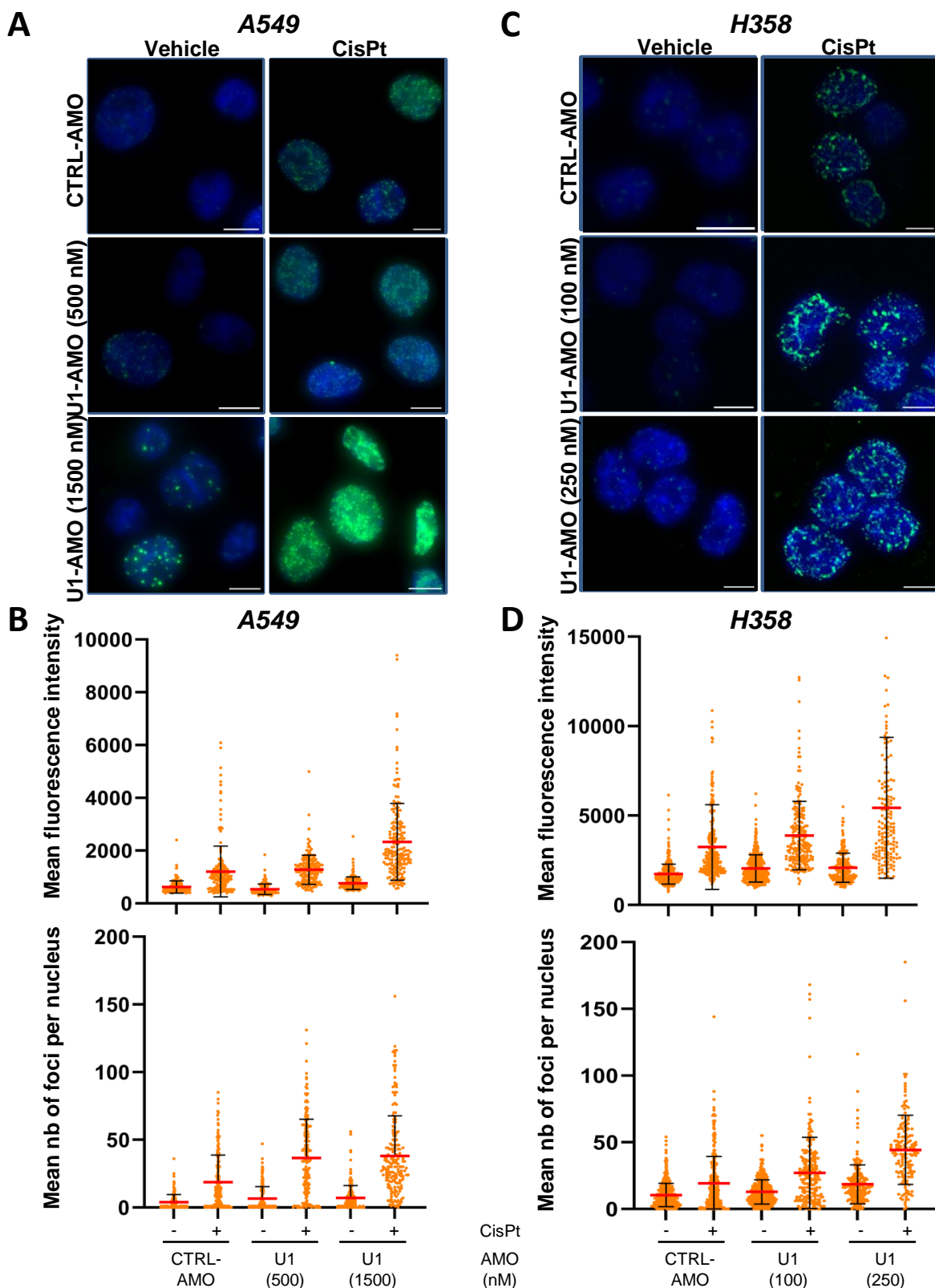
**Fig. 1: U1 snRNA blockade sensitizes cells to growth inhibition by cisplatin**

**A-B**, H358, A549 and A549-R cells were electroporated with CTRL- and/or U1-AMO and subsequently treated with either cisplatin (CisPt at 10, 5 or 15  $\mu$ M, respectively) or vehicle. In each panel, the total amount of AMO (CTRL + U1) is the same in every condition. Cell growth was measured by WST1 assay. **B**, For each AMO condition, cisplatin survival corresponds to the percent of cell viability with cisplatin/AMO combination *versus* AMO alone. Cisplatin mortality corresponds to the percent of cells that did not survive to cisplatin.



**Fig. 2: U1 snRNA blockade prevents the ability of cisplatin to inhibit cell cycle progression.**

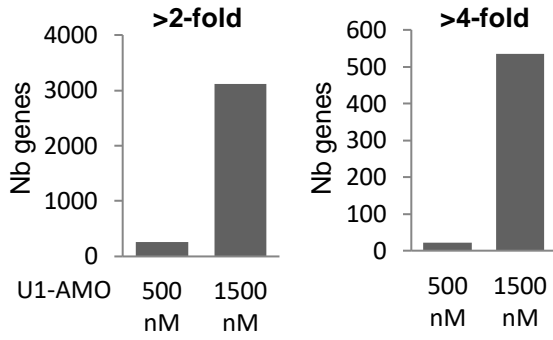
H358 and A549 cells were electroporated with CTRL- and/or U1-AMO and subsequently treated with either cisplatin (at 10 or 5  $\mu$ M, respectively) or vehicle. In each panel, the total amount of AMO (CTRL + U1) is the same in every condition. Cell cycle distribution was analyzed by FACS with propidium iodide and BrdU.



**Fig. 3: U1 snRNA blockade increases cisplatin-induced DNA damage.**

H358 and A549 cells were electroporated with CTRL- and/or U1-AMO and subsequently treated with either cisplatin (at 10 or 5  $\mu$ M, respectively) or vehicle. In each panel, the total amount of AMO (CTRL + U1) is the same in every condition. Immunofluorescence analysis of  $\gamma$ H2AX. **A**, Typical images of  $\gamma$ H2AX (green) and DAPI (blue) staining. Bars represent 10  $\mu$ m. **B**, Quantification of  $\gamma$ H2AX staining.

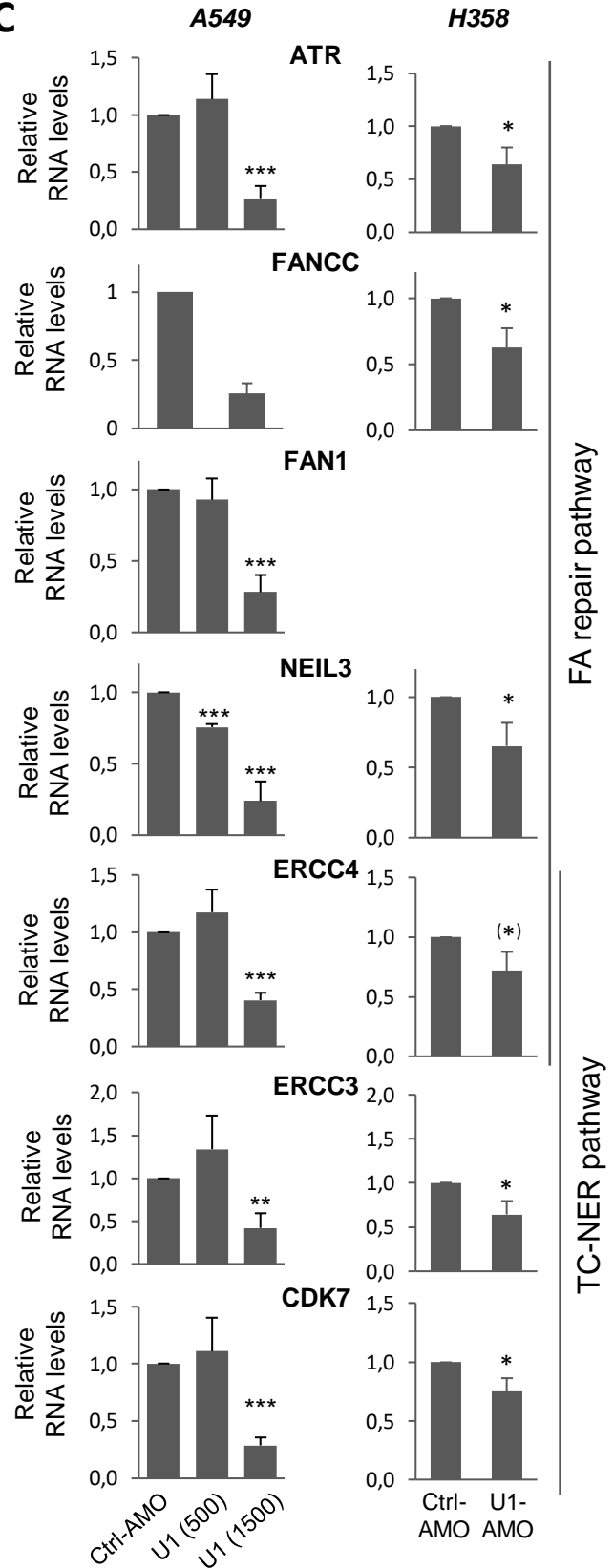
### A LE down-regulation by U1-AMO (3'-seq)



### B Enrichment in Gene ontology terms

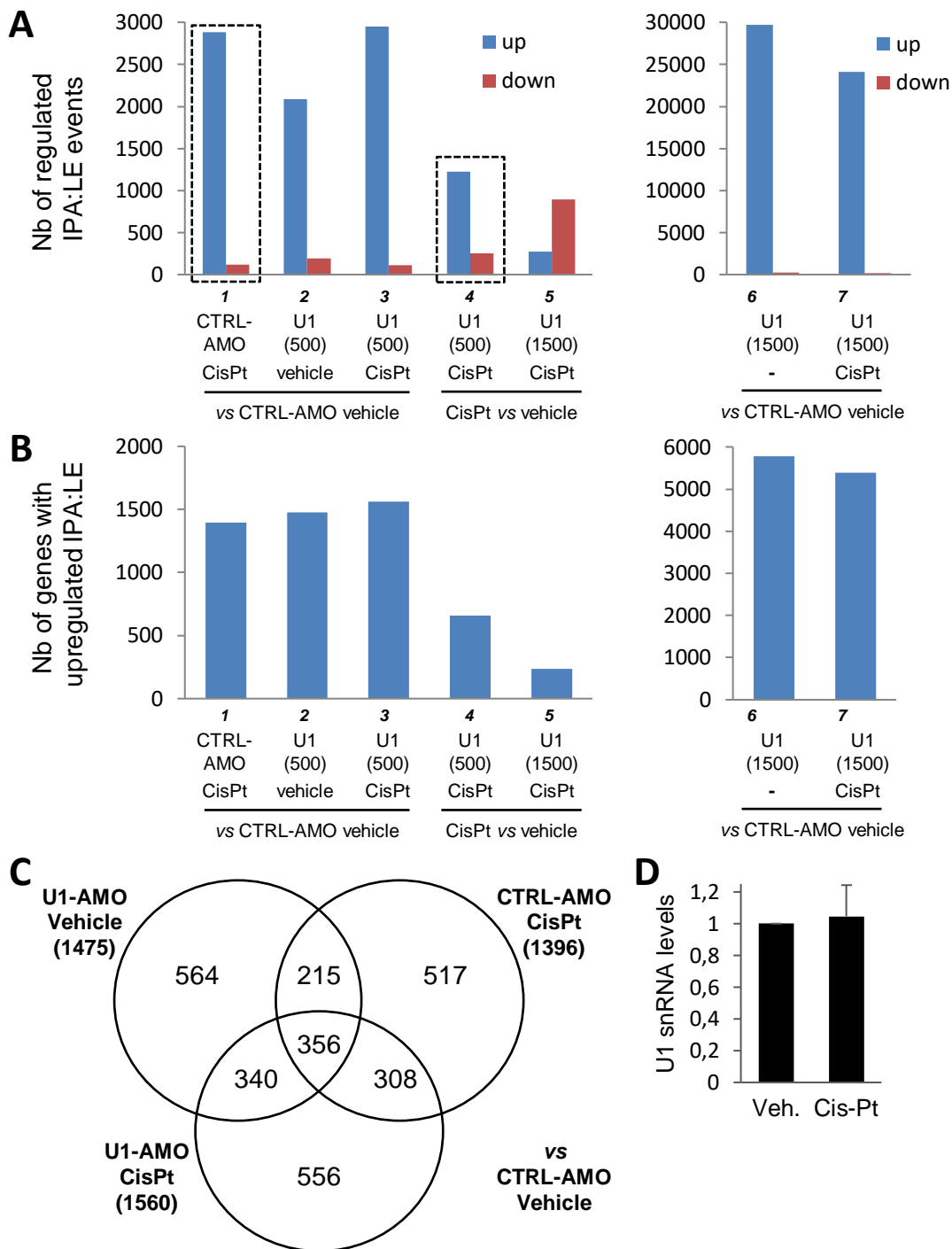
Term	Nb	Enrich°	P val	FDR
Cell cycle	50	2,5	7,E-09	6,E-07
Mitotic spindle organization	13	4,2	6,E-05	3,E-02
G1/S transition of mitotic cell cycle	9	5,0	4,E-04	9,E-02
G2/M transition of mitotic cell cycle	9	2,6	2,E-02	7,E-01
DNA damage	32	2,6	2,E-06	9,E-05
DNA repair	26	2,5	4,E-05	7,E-04
Transcription-coupled nucleotide-excision repair	11	6,1	1,E-05	8,E-03
Nucleotide excision repair	7	5,3	2,E-03	1,E-01
Interstrand cross-link repair	6	4,5	1,E-02	6,E-01
Fanconi anemia pathway	5	3,3	6,E-02	4,E-01

### C



### Fig. 4: U1 snRNA blockade represses genes of the NER and FA repair pathways.

A549 cells were electroporated with CTRL- and/or U1-AMO (with a total of 1500 nM AMO in every condition) and subsequently treated with either cisplatin at 5  $\mu$ M or vehicle. **A**, Number of genes whose expression level is down-regulated by U1-AMO when compared to the same dose of CTRL-AMO in 3'-seq analysis. **B**, Enrichment analysis of gene ontology terms in the list of genes that are down-regulated by at least 4-fold by U1-AMO at 1500 nM. **C**, RT-qPCR analysis using primers located in the last exon of the indicated genes in A549 and H358 cells electroporated with CTRL- and/or U1-AMO, as indicated.



**Fig. 5: U1 snRNA blockade partially prevents cisplatin effect on the IPA:LE isoform ratio.**

A549 cells were electroporated with CTRL- and/or U1-AMO (with a total of 1500 nM AMO in every condition) and subsequently treated with either cisplatin at 5  $\mu$ M or vehicle. 3'-seq analysis of IPA:LE isoform ratio regulation between the indicated conditions. U1, U1-AMO. **A**, Number of events (IPA sites) with up- or down-regulation of the IPA:LE isoform ratio. **B**, Number of genes with upregulated IPA:LE ratio. **C**, Venn diagram of genes with upregulated IPA:LE ratio in different conditions when compared to control cells. **D**, RT-qPCR analysis of U1 snRNA levels in H358 cells treated with cisplatin or vehicle.

## VI.2.2 Conclusion

Dans cette étude, U1-AMO (inhibiteur de U1 snRNA) a été montré dans trois lignées NSCLC différentes comme accentuant la diminution de viabilité induite par le cisplatine. Cette sensibilisation au cisplatine par U1-AMO passe notamment par une augmentation des foyers  $\gamma$ H2AX correspondant aux dommages à l'ADN. Cette sensibilisation est aussi associée avec une répression des gènes de DDR appartenant aux voies Fanconi et NER, principales voies de réparation des pontages d'ADN induits par le cisplatine. Ces résultats suggèrent la possibilité de sensibiliser les cellules cancéreuses à des génotoxiques par un ciblage de U1 snRNA.



## VII. Discussion et perspectives

### VII.1 Mécanismes de régulation du ratio d'isoformes IPA:LE en réponse au cisplatine et U1-AMO

Comme discuté dans le premier article, nos données indiquent que la diminution de processivité transcriptionnelle et l'augmentation du ratio d'isoformes IPA:LE par le cisplatine sont dépendantes des protéines FANCD2 et SETX, qui sont connues pour interagir fonctionnellement avec les R-loops (Said et al. 2022), mais ces effets ne semblent pas être affectés par la RNase H1 pourtant connue pour dégrader ARN au sein des hybrides ARN:ADN des R-loops (Crossley et al. 2019). Afin d'éclairer ce constat, nous pourrions tester par déplétion d'autres facteurs de la voie R-loop intervenant à différentes étapes de la cinétique de résolution afin de déterminer si ce n'est pas une étape précise qui modulerait la réponse du cisplatine au ratio IPA:LE.

Une expérience de *Proximity Ligation Assay* (PLA), regardant la proximité entre deux facteurs, avait été mise en place avec un couple d'anticorps dirigés contre FANCD2 et les hybrides ADN:ARN (S9.6) (données non montrées). Il était en théorie possible de regarder si FANCD2 se trouve à proximité de R-loops, résultat nouveau par rapport à la littérature mais aussi de montrer si le cisplatine induit des R-loops. Un traitement à la RNase H1 comme contrôle négatif était aussi nécessaire afin de montrer que les foyers observés étaient bien issus de R-

s. FANCD2/S9.6 pourraient être présent à des foyers de dommages et de fourches bloquées. Ce contrôle négatif a révélé que nous ne regardions pas que des R-loops à l'aide de ce couple. A court terme, nous envisageons de reprendre cette expérience mais avec des facteurs documentés comme plus spécifiques aux R-loops comme un couple RNase H1/S9.6 ou SETX/BRCA1 (Hatchi et al. 2015) en utilisant des inducteurs de R-loop connus en parallèle comme DRB, la camptothécine ou la pyridostatine (Castillo-Guzman et al. 2021; Chappidi et al. 2020).

Actuellement non montré rigoureusement, certaines revues s'attendent à ce que le cisplatine induise des R-loops. En effet, des génotoxiques créant des dommages similaires comme la mitomycine C, agent générant des pontages interbrins à l'ADN, ont été montré comme induisant des R-loops (Wells et al. 2019). De plus, une étude a récemment montré que

les ROS, générés par des agents génotoxiques comme le cisplatine, peuvent aussi induire des R-loops (Andrs et al. 2023). Nous savons que le cisplatine diminue la processivité de l'ARN pol II (voir papier 1) et supposons que cela est due aux adduits générés bloquant l'ARN pol II en transcription. Ce blocage permettrait la formation d'une R-loop aboutissant potentiellement à de l'IPA. Ce modèle complète celui proposé par Skourti-Stathaki, Proudfoot, et Gromak. Dans ce dernier, il fait mention chez l'Homme que la formation puis la résolution de R-loops uniquement en fin de gène contribue à la terminaison de la transcription. La R-loop se forme et se résout après le signal polyA dans une région GC riche permettant un événement de clivage/polyadénylation puis la dégradation de la partie 3' de l'ARN dépendant de Xrn2, montré dans ce cas pour un groupe de gènes dont la  *$\beta$ -actin*.

L'expérience la plus rigoureuse serait de réaliser un CHIP-seq en utilisant des anticorps dirigés contre l'ARN pol II et contre un facteur de R-loop (se nommant DRIP-seq lorsque dirigé contre S9.6) avec ou sans cisplatine. Dans le but de montrer le lien potentiel entre R-loop et induction d'IPA. Un dernier anticorps témoin de l'activité de clivage/polyadénylation tel que Xrn2 pourrait être utilisé. Le CHIP-seq à la différence de PLA permettrait de montrer réellement pour des locus donnés l'induction de R-loops à l'emplacement de l'arrêt de l'ARN pol II par le cisplatine déclenchant potentiellement un événement de polyadénylation.

Dans notre second article, nous avons observé que certains IPA induits par U1-AMO ou le cisplatine étaient communément régulés tandis que d'autres sont spécifiques à chacun sous tendant des mécanismes de régulation communs mais aussi exclusifs. Il serait intéressant d'aller regarder les séquences et motifs de ces différents IPA afin de déterminer les critères ou les potentiels facteurs régulant ces IPA de façon U1-AMO ou cisplatine dépendant. Étonnamment, nous avons observé de potentiels effets antagonistes entre cisplatine et U1-AMO pour l'induction du ratio IPA:LE dans de nombreux gènes (figure 5). En effet, à confirmer toutefois en RT-qPCR, en co-traitement, le cisplatine et U1-AMO, préviennent une partie de l'effet respectif de l'autre drogue. Or, l'abondance de U1snRNA ne semble pas régulée par le cisplatine (figure 5D). Il est possible qu'U1snRNA interagisse avec la machinerie régulant l'effet du cisplatine sur les IPA et inversement. Nous savons par ailleurs qu'U1-AMO réprime certains gènes de la voie Fanconi dont une partie est impliquée dans le traitement des R-loops (Bhatia et al. 2017). Une autre possibilité serait que l'inhibition d'élongation transcriptionnelle par les adduits cisplatine sur les gènes donne plus de temps aux U1snRNA non inhibés par U1-AMO d'accomplir leur rôle de répresseur d'IPA. Pour tester cela, nous pourrions voir si la régulation du ratio IPA :LE par U1-AMO change en cotraitement avec d'autres inhibiteurs d'élongation de

la transcription comme DRB induisant des pauses de polymérasas (Ip et al. 2011) ou encore une répression des facteurs SCAF4 ou 8 (voir partie II).

## VII.2 Traductibilité de nos isoformes miP-5'UTR IPA

Les isoformes PHF20 IPA et PRKAR1B IPA bien que déplétées dans les polysomes lourds sont abondantes dans les fractions plus légères de 2 à 3 ribosomes pour PHF20 et de 3 à 4 ribosomes dans le cas de PRKAR1B. Un des résultats étonnants de notre étude est que nous détectons en Western et IF la microprotéine codée par PRKAR1B sORF2 mais pas celle de PHF20 qui pourtant a été détectée en spectrométrie de masse d'après OpenProt. Une des explications potentielles pour PHF20 IPA serait que cet engagement en fraction 2 à 3 pourrait s'apparenter à du *ribosome pausing*. En effet, dans la littérature, il est admis que les séquences sur l'ARNm, leur structure, la disponibilité des ARNt et le peptide en cours de synthèse peuvent induire une pause de ribosome en cours d'élongation. Cette pause permet le repliement correct de la protéine, des interactions protéiques mais peut aussi induire une dégradation de la protéine en cours de synthèse ainsi que son ARNm (Collart et al. 2020). Une autre explication potentielle est liée au fait que lors de l'élaboration de l'insert de PHF20 IPA, sa partie la plus en 5' n'a pas pu être incluse (en raison de son contenu trop GC riche il était très difficile d'obtenir un insert par PCR). Des séquences régulatrices pourraient s'y trouver.

Il est aussi possible que PHF20 IPA soit traduite mais dans des conditions particulières, par exemple un type cellulaire. Nous avons collecté nos données de Ribo-seq et Spectrométrie de Masse en combinant OpenProt et sORFs.org. Ces données proviennent de tissus et lignées humaines différents des lignées utilisées dans cette étude. Il est connu que les microprotéines peuvent être assez tissu-spécifiques (voir partie II). La détection de PHF20 en spectrométrie de masse a par exemple été réalisée dans des tissus supposés neuronaux. Il serait alors intéressant d'aller regarder par transfection si PHF20 sORF qui n'a pas pu être détectée dans nos lignées Hela, HEK, et NSCLC puisse être détecté dans des cellules de type neuronal.

Dans une série d'expérience en cours nous nous sommes aperçus qu'en enlevant une partie de la séquence de l'isoforme IPA en amont du codon d'initiation de la sORF2 de PRKAR1B, l'expression de la microprotéine pouvait être perdue. Afin d'identifier dans cette isoforme 5'UTR-IPA des séquences régulatrices de la traduction de la sORF, une série d'expérience délétant plus finement des parties de cette région pourrait être envisagée. La sORF1 de PRKAR1B pourrait (malgré qu'elle semble non traduite dans nos conditions) aider au

recrutement de ribosomes pour PRKAR1B sORF 2. Il y aurait ici comme une uORF chevauchante (uoORF), aidant la traduction de PRKAR1B sORF2. Ces résultats sont à mettre en parallèle avec les études utilisant des vecteurs d'expressions comme stratégie pour tester la traduction des sORF. Ces études pourraient sous-estimer en partie leur potentiel codant par omission de leurs séquences environnantes en amont.

### VII.3 Fonctions et destin de la microprotéine PRKAR1B sORF 2

Nos données d'IF indiquent que PRKAR1B sORF2 code une microprotéine qui, balisée en C-ter, se retrouve enrichie dans de grands compartiments nucléaires, correspondant probablement aux nucléoles. Il serait intéressant de le vérifier, en réalisant notamment des IF pour la nucléoline et la fibrillarine. Par ailleurs, à la vue de sa séquence peptidique, PRKAR1B sORF2 est prédite intrinsèquement désordonnée, l'écartant des microprotéines membranaires discutées dans l'introduction. Cependant aucun court motif d'interaction linéaire (SLIM) n'a été identifié comme la localisant au noyau. Cette localisation pourrait être un artefact de la surexpression (Gibson et al. 2013) et nous attendons de voir la localisation de la microprotéine endogène dans les lignées CRISPR-tag que nous essayons de générer actuellement. Mais nous pensons que cette localisation subnucléaire résulterait de l'interaction de la microprotéine avec des facteurs. Une spectrométrie de masse des interactants de la microprotéine (BioID, Apex ou ViroTrap ; Cardon et al. 2019) pourrait être entreprise afin de trouver des candidats expliquant sa localisation et permettant aussi d'expliquer sa fonction. PRKAR1B sORF2 pourrait en effet être impliquée dans le fonctionnement des nucléoles et moduler des interactions entre facteurs. Impliquée dans la survie au cisplatine, la microprotéine pourrait aussi agir sur des facteurs nucléaires comme la machinerie de réparation ou des facteurs de signalisation et de transcription.

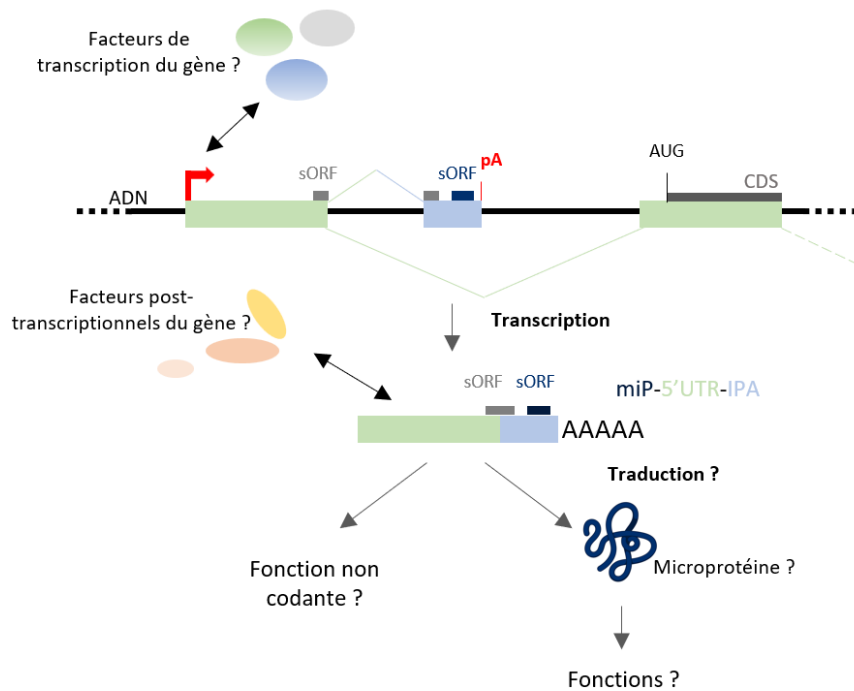
En balisant en C-ter PRKAR1B sORF2, une deuxième bande est observée par WB si on réalise d'abord une immunoprécipitation. Cette deuxième bande pourrait être la preuve de deux microprotéines partageant un même codon STOP. Or nous nous sommes rendus compte qu'en amont et en phase de l'ATG de PRKAR1B sORF2 était présent un CTG, codon alternatif d'initiation de la traduction. Pour tester cette hypothèse, nous balisons actuellement les deux codons d'initiation possibles en N-ter. Nous modifions aussi dans notre construction avec *tag* en C-ter, chacun de ces codons d'initiation en codons non-initiateurs. Nous nous attendons à voir un changement dans les bandes observées en WB. Il n'est pas non plus exclu que cette

bande supplémentaire correspond en réalité à la modification post-traductionnelle de notre microprotéine. Plusieurs séquences favorisant différentes modifications sont, en effet, retrouvées sur notre microprotéine comme des séquences de fucosylation ou de phosphorylation. Nous pourrions faire des expériences de traitement de notre microprotéine par diverses enzymes (phosphatases, etc...) afin de voir si ces doubles bandes changent.

Il serait aussi intéressant de développer un anticorps dirigé contre PRKAR1B sORF2 pour regarder sa distribution dans les tissus humains. Nous pourrions regarder dans des collections de tissus sains mais aussi tumoraux afin d'observer la régulation de l'expression et de la localisation de la microprotéine.

#### VII.4 Concept des miP-5'UTR-IPA

Le concept que les isoformes 5'UTR-IPA sont une source de microprotéines est nouveau. En effet, aucune étude ne documente des isoformes IPA comme sources de sORF codant des microprotéines. Une estimation par notre équipe fait monter ce nombre aux alentours de 156 isoformes miP-5'UTR-IPA potentielles induites par le cisplatine (Figure 6 du papier 1). Les isoformes 5'UTR-IPA possèdent deux parties distinctes, une première constituée du début de la 5'UTR de l'isoforme longue et une seconde issue de l'exon terminal alternatif avec le site IPA. La première partie pourrait apporter une partie de la régulation transcriptionnelle et post-transcriptionnelle de l'isoforme longue à la sORF. Elle pourrait contribuer à sa traduction, permettre sa stabilisation ou une localisation particulière. La deuxième partie, elle, est complètement nouvelle et, au moins dans le cas de PRKAR1B et PHF20, contient la sORF. Cela signifierait que cette partie pourrait porter une fonction indépendante du gène « hôte » (Figure 21). *In fine*, cela pourrait faire converger les fonctions de la microprotéine et de la protéine. Curieusement pour PRKAR1B sORF2, nous avons observé que le codon STOP était incorporé dans le signal polyA (AATAA) de l'isoforme. Cela signifie que PRKAR1B 5'UTR-IPA ne possède quasiment pas de 3'UTR.



### Figure 21 : Modèle des miP-5'UTR-IPA

Des isoformes ARN issues d'IPA ne comportent qu'une partie de la 5'UTR du gène ainsi qu'un exon terminal alternatif (ALE), on les nomme 5'UTR-IPA. Elles pourraient être régulées par des facteurs de transcription du gène. Certaines de ces 5'UTR-IPA sont porteuses de sORF nichées totalement ou partiellement dans l'ALE (sORF bleu ou grise respectivement) et codant des microprotéines, on les qualifie ainsi de miP-5'UTR-IPA. Ces dernières pourraient aussi être régulées post-transcriptionnellement. La microprotéine peut être vectrice de fonctions mais l'isoforme ARN pourrait aussi avoir une fonction non codante.

## VII.5 Fonctions des isoformes 5'UTR-IPA

Surprenamment, pour nos deux isoformes 5'UTR-IPA de PRKAR1B et PHF20 ainsi que pour deux autres actuellement à l'étude dans mon équipe, leur déplétion augmente la survie cellulaire (en présence ou en absence de cisplatine, selon les cas). Le fait que nous trouvions seulement des isoformes IPA allant dans ce sens peut relever pour le moment du hasard statistique mais l'on peut se demander s'il n'y a pas derrière une tendance globale liant les isoformes 5'UTR-IPA. On pourrait faire l'hypothèse que ces isoformes ARN, par des fonctions codantes ou non, sont des signaux de mort ou de stress pour la cellule. En réponse à un stress, ces isoformes agiraient en tant que vecteur de stress supplémentaire pour la cellule. La cellule doit alors reprendre à temps ses régulations sur ses 5'UTR-IPA, si elle n'y arrive pas une réponse apoptotique se met en place dans le cas contraire elle survit. Cette hypothèse est à mettre en parallèle avec l'observation de Ogami et al. que la répression de l'exosome nucléaire conduit à la fois à une induction d'isoformes 5'IPA et à une diminution de la traduction globale dans la cellule. Pour tester cela, il serait possible de réaliser un criblage sgARN dirigé contre les 156 isoformes miP-5'UTR-IPA que nous avons identifiées (Figure 6 du papier 1) et de regarder

leur phénotype de croissance cellulaire (avec ou sans agent stressant). Ce crible pourrait être dirigé contre l'expression de l'ARN (délétion de l'isoforme) ou l'expression de la microprotéine (mutation du codon d'initiation). Nous pourrions aussi étudier l'interactome de ces isoformes ARN au niveau de leurs partenaires potentiels protéiques ou ribonucléiques. Ces interactants pourraient nous indiquer les potentielles voies de signalisation en jeu. Nous voudrions également étudier les interactants protéiques de leurs microprotéines potentielles par BioID (comme décrit ci-dessus pour PRKAR1B sORF2) ou *pull-down*-MS et éventuellement les interactants ribonucléiques (interaction ARN-proteine) afin d'en apprendre plus sur leurs mécanismes d'action.

Il serait intéressant d'étendre les expériences de viabilité cellulaire sur PRKAR1B dans d'autres lignées cellulaires et d'autres tissus. L'approche CRISPR permet de faire abstraction des effets *off-targets* potentiels des siARN ciblant la 5'UTR-IPA. Il serait donc intéressant de réaliser ces expériences par délétion en CRISPR. De même, il serait intéressant de ne plus se limiter à la réponse au cisplatine mais d'utiliser d'autres agents génotoxiques. On peut en effet se demander si le phénotype trouvé est spécifique au cisplatine ou correspond à une réponse plus générale de la cellule au stress génotoxique.

Dans cette étude, nous nous sommes concentrés sur les isoformes IPA courtes finissant avant l'ATG annoté et pas ou peu détectées dans les polysomes lourds pour finir par découvrir des microprotéines. Cependant, nos données révèlent aussi qu'une poignée d'isoformes IPA courtes finissant avant l'ATG sont bien engagées dans les polysomes lourds. On pourrait s'attendre à ce que certaines de ces isoformes possèdent le même mode d'action non-codant que le transcrite SPUD. En effet, ces isoformes en interagissant avec les ribosomes actifs pourraient entrer en compétition avec le transcrite canonique de leur gène pour la liaison à des régulateurs traductionnels (Murphy et al. 2023). D'autres pourraient coder pour des microprotéines très activement traduites qui pourraient faire l'objet d'études.

## VII.6 Sensibilisation par U1-AMO au cisplatine : Perspectives et limites

Nos données montrent que l'inhibition de U1snRNA par U1-AMO dans des lignées cellulaires NSCLC sensibilise au cisplatine et induit une répression forte et au moins jusqu'à 48 heures, de nombreux gènes impliqués dans des voies de DDR. Il est à noter que les effets publiés de U1-AMO sur des gènes de DDR étaient basés sur des analyses d'ARN naissant, 4 à 8 heures après transfection et n'avaient pas été validés (Oh et al. 2017). Dans notre étude, les

gènes de DDR affectés et validés en RT-qPCR sont des gènes clés comme *ATR*, *XPF (ERCC4)*, *XPB (ERCC3)*, *FANCC* et *FAN1*. Les voies de DDR impactées ici semblent être à première vue celle de la signalisation cellulaire avec la répression d'ATR. Les autres gènes de DDR se rapportent ici aux voies FANCONI, NER et TC-NER. D'autres voies annexes de réparations de liaisons interbrins combinant le TC-NER aux polymérases TLS semblent aussi touchées (partie IV ; données 3'-seq). Ces voies peuvent traiter les dommages induits par le cisplatine (partie IV). L'inhibition d'ATR pourrait cependant prévenir une partie de l'effet sensibilisant de U1-AMO en inhibant l'induction d'apoptose en réponse au cisplatine médiée justement par ATR (partie IV). La répression de ces gènes est toutefois associée à une augmentation des foyers  $\gamma$ H2AX (phosphorylation médiée par ATR ou ATM) en IF en cotraitement cisplatine-U1-AMO par rapport aux traitements cisplatine seul dans nos deux lignées. Cela suggère que la signalisation ATR reste suffisamment efficace, dans nos conditions de temps, pour médier sa fonction de signalisation. Il serait intéressant de regarder l'expression et la phosphorylation de la protéine ATR afin de vérifier cela.

Une des premières incertitudes lors de la mise en place des traitements U1-AMO était ses propres effets cytotoxiques. En effet, peu d'études s'étaient penchées sur ses effets à long terme (seulement à maximum 8h) ni sur ses propres effets physiologiques, s'intéressant plus aux mécanismes d'action. Nos observations montrent qu'U1-AMO inhibe la croissance cellulaire dans les trois lignées NSCLC étudiées (A549, H358 et A549-R0; figure 1 du papier 2) mais qu'il n'inhibe la progression du cycle cellulaire que dans les cellules A549 (pas dans les H358 ; figure 2). En effet, nous observons une diminution de cellules en phase S et une augmentation de cellules en G2/M, ce qui pourrait être dû à des blocages en G1 et en G2. Cela reste à vérifier par Western Blot sur des protéines de points de contrôle. Ces effets sur le cycle ne semblent pas être dépendant de dommages à l'ADN car nous n'observons pas d'augmentation des foyers  $\gamma$ H2AX par U1-AMO en absence de cisplatine (figure 3). Le blocage en G1 d'A549 pourrait être dépendant de la voie p53-p21 car p53 est sauvage dans A549 et muté dans H358. Nos données 3'-seq révèlent, en outre, une potentielle répression de l'expression des gènes de cycle cellulaire. En lien avec cela, dans notre FACS avec BrdU, il a pu être observé une diminution notable de l'incorporation de BrdU par les cellules traitées à U1-AMO de façon dose dépendante (données non montrées). Cette diminution de l'incorporation pourrait être validée par une autre approche mesurant l'incorporation d'ADN sur nos extraits cellulaires *ex vivo*. U1-AMO a donc ici de très importants effets sur le cycle cellulaire et la réplication qui pourraient aussi être intéressants dans le contexte de thérapie anti-cancéreuse.



Dans nos données FACS très peu de cellules en subG1 sont observées dans toutes nos conditions malgré une mortalité cellulaire accrue entre un génotoxique et un inducteur d'IPA dans nos tests de viabilité. Cela pourrait s'expliquer dans un biais du protocole que j'ai utilisé, dans lequel je ne récoltais pas le milieu (et donc la plupart des cellules mourantes). Un autre protocole pourrait être pris, afin de vérifier cela mais le mieux serait de regarder par FACS ou WB les cellules exprimant des facteurs de mort de divers types (apoptose, nécrose, etc...).

Les régulations du cycle apportées par le cisplatine sur nos cellules sont donc en majorité prévenues par U1-AMO à l'exception de l'augmentation des cellules en phase G2/M pour H358. Cette inhibition de progression du cycle cellulaire, de la réplication et de la croissance par U1-AMO doivent diminuer la coopération de U1-AMO avec le cisplatine qui reste néanmoins mesurable mais qui pourrait peut-être optimisée. Cela laisse entrevoir les limitations d'utilisation de cette drogue pour sensibiliser aux génotoxiques. La dose employée pour impacter le DDR affecte aussi de nombreux autres gènes. Visant les gènes du cycle cellulaire et de la réplication comme mentionné ci-dessus, ces effets pourraient prévenir une partie de la sensibilisation au cisplatine. Cibler des facteurs plus en aval avec des effets plus spécifiques permettrait de continuer à cibler ces gènes du DDR tout en écartant ces effets indésirables. Un ciblage de facteurs tel que des facteurs du spliceosome seraient possibles.

Il serait cependant intéressant d'étudier l'effet d'U1-AMO sur des tissus ou des cellules sains. En effet, si le blocage se trouve éventuellement plus fort dans les cellules saines, il permettrait de mieux protéger les cellules saines face au cisplatine. Plusieurs drogues actuellement à l'étude en synergie avec des agents génotoxiques possèdent un effet bloqueur de cycle affectant plus les tissus sains que les tumeurs dérégulées au niveau de leur cycle cellulaire. Cela permet ainsi la protection des tissus sains face à l'apoptose et réduit la toxicité des traitements (Sesink 2023). De plus, il serait intéressant de regarder pour d'autres génotoxiques si une sensibilisation par U1-AMO est aussi observée comme cela est par exemple le cas pour CDK12i avec divers génotoxiques. En effet, U1-AMO réprime des gènes de plusieurs voies de réparation et on pourrait s'attendre à ce que d'autres génotoxiques provoquant d'autres types de dommages à l'ADN puissent obtenir un gain de mortalité en combinaison avec U1-AMO. En outre, on pourrait se demander si une combinaison U1-AMO et CDK12i ne serait pas envisageable afin d'augmenter l'induction d'IPA dans les gènes de DDR. Par ces deux approches focalisées sur l'IPA, même à dose faible, on pourrait anticiper de forts effets de répression sur la voie Fanconi ciblée par les deux drogues. De plus, CDK12i cible les

gènes de HR avec *ATM* tandis que U1-AMO cible plus le NER avec *ATR* dévoilant une potentielle complémentarité des drogues. La combinaison à dose modérée diminuerait aussi les effets *off-targets* de U1-AMO.

## VIII. Bibliographie :

- Abrams, Tinya J., Leslie B. Lee, Lesley J. Murray, Nancy K. Pryer, et Julie M. Cherrington. 2003. « SU11248 Inhibits KIT and Platelet-Derived Growth Factor Receptor Beta in Preclinical Models of Human Small Cell Lung Cancer ». *Molecular Cancer Therapeutics* 2 (5): 471-78.
- Ajani, Jaffer A., Kathryn A. Winter, Leonard L. Gunderson, John Pedersen, Al B. Benson, Charles R. Thomas, Robert J. Mayer, Michael G. Haddock, Tyvin A. Rich, et Christopher Willett. 2008. « Fluorouracil, Mitomycin, and Radiotherapy vs Fluorouracil, Cisplatin, and Radiotherapy for Carcinoma of the Anal Canal: A Randomized Controlled Trial ». *JAMA* 299 (16): 1914-21. <https://doi.org/10.1001/jama.299.16.1914>.
- Alfonso-Gonzalez, Carlos, Ivano Legnini, Sarah Holec, Laura Arrigoni, Hasan Can Ozbulut, Fernando Mateos, David Koppstein, Agnieszka Rybak-Wolf, Ulrike Bönisch, Nikolaus Rajewsky, et Valérie Hilgers. 2023. « Sites of transcription initiation drive mRNA isoform selection ». *Cell* 186 (11): 2438-2455.e22. <https://doi.org/10.1016/j.cell.2023.04.012>.
- Alt, Aaron, Katja Lammens, Claudia Chicchini, Alfred Lammens, J. Carsten Pieck, David Kuch, Karl-Peter Hopfner, et Thomas Carell. 2007. « Bypass of DNA Lesions Generated during Anticancer Treatment with Cisplatin by DNA Polymerase Eta ». *Science (New York, N.Y.)* 318 (5852): 967-70. <https://doi.org/10.1126/science.1148242>.
- Alt, F. W., A. L. Bothwell, M. Knapp, E. Siden, E. Mather, M. Koshland, et D. Baltimore. 1980. « Synthesis of Secreted and Membrane-Bound Immunoglobulin Mu Heavy Chains Is Directed by mRNAs That Differ at Their 3' Ends ». *Cell* 20 (2): 293-301. [https://doi.org/10.1016/0092-8674\(80\)90615-7](https://doi.org/10.1016/0092-8674(80)90615-7).
- Amante, Samuele M, Bertille Montibus, Michael Cowley, Nikolaos Barkas, Jessica Setiadi, Heba Saadeh, Joanna Giemza, Stephania Contreras-Castillo, Karin Fleischanderl, Reiner Schulz, et Rebecca J Oakey. 2020. « Transcription of intragenic CpG islands influences spatiotemporal host gene pre-mRNA processing ». *Nucleic Acids Research* 48 (15): 8349-59. <https://doi.org/10.1093/nar/gkaa556>.
- Amara, Susan G., Ronald M. Evans, et Michael G. Rosenfeld. 1984. « Calcitonin/Calcitonin Gene-Related Peptide Transcription Unit: Tissue-Specific Expression Involves Selective Use of Alternative Polyadenylation Sites ». *Molecular and Cellular Biology* 4 (10): 2151-60. <https://doi.org/10.1128/mcb.4.10.2151-2160.1984>.
- Anczuków, Olga, et Adrian R. Krainer. 2016. « Splicing-Factor Alterations in Cancers ». *RNA (New York, N.Y.)* 22 (9): 1285-1301. <https://doi.org/10.1261/rna.057919.116>.
- Anderson, Douglas M., Catherine A. Makarewich, Kelly M. Anderson, John M. Shelton, Svetlana Bezprozvannaya, Rhonda Bassel-Duby, et Eric N. Olson. 2016. « Widespread Control of Calcium Signaling by a Family of SERCA-Inhibiting Micropeptides ». *Science Signaling* 9 (457): ra119-ra119. <https://doi.org/10.1126/scisignal.aaj1460>.
- Andrs, Martin, Henriette Stoy, Barbora Boleslavska, Nagaraja Chappidi, Radhakrishnan Kanagaraj, Zuzana Nascakova, Shruti Menon, Satyajeet Rao, Anna Oravetzova, Jana Dobrovolna, Kalpana Surendranath, Massimo Lopes, et Pavel Janscak. 2023.

- « Excessive Reactive Oxygen Species Induce Transcription-Dependent Replication Stress ». *Nature Communications* 14 (1): 1791. <https://doi.org/10.1038/s41467-023-37341-y>.
- Arana, Mercedes E., Mineaki Seki, Richard D. Wood, Igor B. Rogozin, et Thomas A. Kunkel. 2008. « Low-fidelity DNA synthesis by human DNA polymerase theta ». *Nucleic Acids Research* 36 (11): 3847-56. <https://doi.org/10.1093/nar/gkn310>.
- Araújo, Sofia J., Franck Tirode, Frederic Coin, Helmut Pospiech, Juhani E. Syväoja, Manuel Stucki, Ulrich Hübscher, Jean-Marc Egly, et Richard D. Wood. 2000. « Nucleotide excision repair of DNA with recombinant human proteins: definition of the minimal set of factors, active forms of TFIIH, and modulation by CAK ». *Genes & Development* 14 (3): 349-59.
- Arnoult, Nausica, Adriana Correia, Jiao Ma, Anna Merlo, Sara Garcia-Gomez, Marija Maric, Marco Tognetti, Christopher W. Benner, Simon J. Boulton, Alan Saghatelian, et Jan Karlseder. 2017. « Regulation of DNA Repair Pathway Choice in S and G2 Phases by the NHEJ Inhibitor CYREN ». *Nature* 549 (7673): 548-52. <https://doi.org/10.1038/nature24023>.
- Ashour, Mohamed E., Reham Atteya, et Sherif F. El-Khamisy. 2015. « Topoisomerase-Mediated Chromosomal Break Repair: An Emerging Player in Many Games ». *Nature Reviews Cancer* 15 (3): 137-51. <https://doi.org/10.1038/nrc3892>.
- Astolfi, Laura, Sara Ghiselli, Valeria Guaran, Milvia Chicca, Edi Simoni, Elena Olivetto, Giorgio Lelli, et Alessandro Martini. 2013. « Correlation of Adverse Effects of Cisplatin Administration in Patients Affected by Solid Tumours: A Retrospective Evaluation ». *Oncology Reports* 29 (4): 1285-92. <https://doi.org/10.3892/or.2013.2279>.
- Bai, Rui, Ruixue Wan, Lin Wang, Kui Xu, Qiangfeng Zhang, Jianlin Lei, et Yigong Shi. 2021. « Structure of the Activated Human Minor Spliceosome ». *Science* 371 (6535): eabg0879. <https://doi.org/10.1126/science.abg0879>.
- Bartkowiak, Bartłomiej, Christopher M. Yan, Erik J. Soderblom, et Arno L. Greenleaf. 2019. « CDK12 Activity-Dependent Phosphorylation Events in Human Cells ». *Biomolecules* 9 (10): 634. <https://doi.org/10.3390/biom9100634>.
- Basu, Alakananda, et Soumya Krishnamurthy. 2010. « Cellular Responses to Cisplatin-Induced DNA Damage ». *Journal of Nucleic Acids* 2010 (août): e201367. <https://doi.org/10.4061/2010/201367>.
- Benech, P., Y. Mory, M. Revel, et J. Chebath. 1985. « Structure of two forms of the interferon-induced (2'-5') oligo A synthetase of human cells based on cDNAs and gene sequences. » *The EMBO Journal* 4 (9): 2249-56. <https://doi.org/10.1002/j.1460-2075.1985.tb03922.x>.
- Berg, Michael G., Larry N. Singh, Ihab Younis, Qiang Liu, Anna Maria Pinto, Daisuke Kaida, Zhenxi Zhang, Sungchan Cho, Scott Sherrill-Mix, Lili Wan, et Gideon Dreyfuss. 2012. « U1 snRNP Determines mRNA Length and Regulates Isoform Expression ». *Cell* 150 (1): 53-64. <https://doi.org/10.1016/j.cell.2012.05.029>.

- Bernardes de Jesus, Bruno M., Magnar Bjørås, Frédéric Coin, et Jean Marc Egly. 2008. « Dissection of the Molecular Defects Caused by Pathogenic Mutations in the DNA Repair Factor XPC ». *Molecular and Cellular Biology* 28 (23): 7225-35. <https://doi.org/10.1128/MCB.00781-08>.
- Bhati, Kaushal Kumar, Anko Blaakmeer, Esther Botterweg Paredes, Ulla Dolde, Tenai Eguen, Shin-Young Hong, Vandasue Rodrigues, Daniel Straub, Bin Sun, et Stephan Wenkel. 2018. « Approaches to Identify and Characterize microProteins and Their Potential Uses in Biotechnology ». *Cellular and Molecular Life Sciences* 75 (14): 2529-36. <https://doi.org/10.1007/s00018-018-2818-8>.
- Bhatia, Vaibhav, Emilia Herrera-Moyano, Andrés Aguilera, et Belén Gómez-González. 2017. « The Role of Replication-Associated Repair Factors on R-Loops ». *Genes* 8 (7): 171. <https://doi.org/10.3390/genes8070171>.
- Bi, Baoyuan, Nataliya Rybalchenko, Efim I. Golub, et Charles M. Radding. 2004. « Human and yeast Rad52 proteins promote DNA strand exchange ». *Proceedings of the National Academy of Sciences of the United States of America* 101 (26): 9568-72. <https://doi.org/10.1073/pnas.0403205101>.
- Blackford, Andrew N., et Stephen P. Jackson. 2017. « ATM, ATR, and DNA-PK: The Trinity at the Heart of the DNA Damage Response ». *Molecular Cell* 66 (6): 801-17. <https://doi.org/10.1016/j.molcel.2017.05.015>.
- Blazek, Dalibor, Jiri Kohoutek, Koen Bartholomeeusen, Eric Johansen, Petra Hulinkova, Zeping Luo, Peter Cimermancic, Jernej Ule, et B. Matija Peterlin. 2011. « The Cyclin K/Cdk12 Complex Maintains Genomic Stability via Regulation of Expression of DNA Damage Response Genes ». *Genes & Development* 25 (20): 2158-72. <https://doi.org/10.1101/gad.16962311>.
- Bozorgmehr, Joseph Hannon. 2023. « The Non-de Novo Origins of “Functional” Micropeptides in the Human Lineage: A Response to Vakirlis et al (2022) in Cell Reports ». <https://doi.org/10.13140/RG.2.2.11681.68960/3>.
- Brozovic, Anamaria, Andreja Ambriović-Ristov, et Maja Osmak. 2010. « The Relationship between Cisplatin-Induced Reactive Oxygen Species, Glutathione, and BCL-2 and Resistance to Cisplatin ». *Critical Reviews in Toxicology* 40 (4): 347-59. <https://doi.org/10.3109/10408441003601836>.
- Brunet, Marie A., Sebastien Leblanc, et Xavier Roucou. 2020. « Reconsidering proteomic diversity with functional investigation of small ORFs and alternative ORFs ». *Experimental Cell Research* 393 (1): 112057. <https://doi.org/10.1016/j.yexcr.2020.112057>.
- Burkovics, Peter, Marek Sebesta, Alexandra Sisakova, Nicolas Plault, Valeria Szukacsov, Thomas Robert, Lajos Pinter, Victoria Marini, Peter Kolesar, Lajos Haracska, Serge Gangloff, et Lumir Krejci. 2013. « Srs2 Mediates PCNA-SUMO-Dependent Inhibition of DNA Repair Synthesis ». *The EMBO Journal* 32 (5): 742-55. <https://doi.org/10.1038/emboj.2013.9>.

- Calvo, Olga, et James L. Manley. 2003. « Strange Bedfellows: Polyadenylation Factors at the Promoter ». *Genes & Development* 17 (11): 1321-27. <https://doi.org/10.1101/gad.1093603>.
- Cao, Xiongwen, et Sarah A. Slavoff. 2020. « Non-AUG start codons: Expanding and regulating the small and alternative ORFeome ». *Experimental Cell Research* 391 (1): 111973. <https://doi.org/10.1016/j.yexcr.2020.111973>.
- Cardon, Tristan, Michel Salzet, Julien Franck, et Isabelle Fournier. 2019. « Nuclei of HeLa cells interactomes unravel a network of ghost proteins involved in proteins translation ». *Biochimica et Biophysica Acta (BBA) - General Subjects* 1863 (10): 1458-70. <https://doi.org/10.1016/j.bbagen.2019.05.009>.
- Carter, Robert, et Guy Drouin. 2009. « Structural Differentiation of the Three Eukaryotic RNA Polymerases ». *Genomics* 94 (6): 388-96. <https://doi.org/10.1016/j.ygeno.2009.08.011>.
- Castillo-Guzman, Daisy, et Frédéric Chédin. 2021. « Defining R-loop classes and their contributions to genome instability ». *DNA Repair* 106 (octobre): 103182. <https://doi.org/10.1016/j.dnarep.2021.103182>.
- Chakraborty, Alina, Mandy Cadix, Sébastien Relier, Nicolò Taricco, Tina Alaeitabar, Alexandre Devaux, Céline M. Labbé, Sylvain Martineau, Amélie Heneman-Masurel, Pierre Gestraud, Alberto Inga, Nicolas Servant, Stéphan Vagner, et Martin Dutertre. 2022. « Compartment-Specific and ELAVL1-Coordinated Regulation of Intronic Polyadenylation Isoforms by Doxorubicin ». *Genome Research* 32 (7): 1271-84. <https://doi.org/10.1101/gr.276192.121>.
- Chalkley, Gillian E., et C. Peter Verrijzer. 1999. « DNA binding site selection by RNA polymerase II TAFs: a TAFII250–TAFII150 complex recognizes the Initiator ». *The EMBO Journal* 18 (17): 4835-45. <https://doi.org/10.1093/emboj/18.17.4835>.
- Chan, D. W., et S. P. Lees-Miller. 1996. « The DNA-Dependent Protein Kinase Is Inactivated by Autophosphorylation of the Catalytic Subunit ». *The Journal of Biological Chemistry* 271 (15): 8936-41. <https://doi.org/10.1074/jbc.271.15.8936>.
- Chappidi, Nagaraja, Zuzana Nascakova, Barbora Boleslavská, Ralph Zellweger, Esin Isik, Martin Andrs, Shruti Menon, Jana Dobrovolna, Chiara Balbo Pogliano, Joao Matos, Antonio Porro, Massimo Lopes, et Pavel Janscak. 2020. « Fork Cleavage-Religation Cycle and Active Transcription Mediate Replication Restart after Fork Stalling at Co-transcriptional R-Loops ». *Molecular Cell* 77 (3): 528-541.e8. <https://doi.org/10.1016/j.molcel.2019.10.026>.
- Charlesworth, Amanda, Hedda A. Meijer, et Cornelia H. de Moor. 2013. « Specificity Factors in Cytoplasmic Polyadenylation: Factors in Cytoplasmic Polyadenylation ». *Wiley Interdisciplinary Reviews: RNA* 4 (4): 437-61. <https://doi.org/10.1002/wrna.1171>.
- Chen, Helen H. W., Wen-Chung Chen, Zhang-Dong Liang, Wen-Bin Tsai, Yan Long, Isamu Aiba, Siqing Fu, Russell Broadus, Jinsong Liu, Lynn G. Feun, Niramol Savaraj, et Macus Tien Kuo. 2015. « Targeting Drug Transport Mechanisms for Improving Platinum-Based

- Cancer Chemotherapy ». *Expert Opinion on Therapeutic Targets* 19 (10): 1307-17. <https://doi.org/10.1517/14728222.2015.1043269>.
- Chen, Jin, Andreas-David Brunner, J. Zachery Cogan, James K. Nuñez, Alexander P. Fields, Britt Adamson, Daniel N. Itzhak, Jason Y. Li, Matthias Mann, Manuel D. Leonetti, et Jonathan S. Weissman. 2020. « Pervasive Functional Translation of Noncanonical Human Open Reading Frames ». *Science*, mars. <https://doi.org/10.1126/science.aay0262>.
- Chen, Shang-Hung, et Jang-Yang Chang. 2019. « New Insights into Mechanisms of Cisplatin Resistance: From Tumor Cell to Microenvironment ». *International Journal of Molecular Sciences* 20 (17): 4136. <https://doi.org/10.3390/ijms20174136>.
- Chen, Xin, Sarah M. Lloyd, Junghun Kweon, Giovanni M. Gamalong, et Xiaomin Bao. 2021. « Epidermal Progenitors Suppress GRHL3-Mediated Differentiation through Intronic Polyadenylation Promoted by CPSF-HNRNPA3 Collaboration ». *Nature Communications* 12 (1): 448. <https://doi.org/10.1038/s41467-020-20674-3>.
- Cherry, Sara, et Kristen W. Lynch. 2020. « Alternative Splicing and Cancer: Insights, Opportunities, and Challenges from an Expanding View of the Transcriptome ». *Genes & Development* 34 (15-16): 1005-16. <https://doi.org/10.1101/gad.338962.120>.
- Chirackal Manavalan, Anil Paul, Kveta Pilarova, Michael Kluge, Koen Bartholomeeusen, Michal Rajecky, Jan Oppelt, Prashant Khirsariya, Kamil Paruch, Lumir Krejci, Caroline C Friedel, et Dalibor Blazek. 2019. « CDK12 controls G1/S progression by regulating RNAPII processivity at core DNA replication genes ». *EMBO reports* 20 (9): e47592. <https://doi.org/10.15252/embr.201847592>.
- Chiu, Anthony C., Hiroshi I. Suzuki, Xuebing Wu, Dig B. Mahat, Andrea J. Kriz, et Phillip A. Sharp. 2018. « Transcriptional Pause Sites Delineate Stable Nucleosome-Associated Premature Polyadenylation Suppressed by U1 snRNP ». *Molecular Cell* 69 (4): 648-663.e7. <https://doi.org/10.1016/j.molcel.2018.01.006>.
- Chong, Chloe, Markus Müller, HuiSong Pak, Dermot Harnett, Florian Huber, Delphine Grun, Marion Leleu, Aymeric Auger, Marion Arnaud, Brian J. Stevenson, Justine Michaux, Ilija Bilic, Antje Hirsekorn, Lorenzo Calviello, Laia Simó-Riudalbas, Evarist Planet, Jan Lubiński, Marta Bryśkiewicz, Maciej Wiznerowicz, Ioannis Xenarios, Lin Zhang, Didier Trono, Alexandre Harari, Uwe Ohler, George Coukos, et Michal Bassani-Sternberg. 2020. « Integrated Proteogenomic Deep Sequencing and Analytics Accurately Identify Non-Canonical Peptides in Tumor Immunopeptidomes ». *Nature Communications* 11 (1): 1293. <https://doi.org/10.1038/s41467-020-14968-9>.
- Chu, Qian, Thomas F. Martinez, Sammy Weiser Novak, Cynthia J. Donaldson, Dan Tan, Joan M. Vaughan, Tina Chang, Jolene K. Diedrich, Leo Andrade, Andrew Kim, Tong Zhang, Uri Manor, et Alan Saghatelian. 2019. « Regulation of the ER stress response by a mitochondrial microprotein ». *Nature Communications* 10 (octobre): 4883. <https://doi.org/10.1038/s41467-019-12816-z>.
- Cleaver, J. E., J. Jen, W. C. Charles, et D. L. Mitchell. 1991. « Cyclobutane Dimers and (6-4) Photoproducts in Human Cells Are Mended with the Same Patch Sizes ». *Photochemistry and Photobiology* 54 (3): 393-402. <https://doi.org/10.1111/j.1751-1097.1991.tb02033.x>.

- Coin, Frédéric, Valentyn Oksenysh, Vincent Mocquet, Stefanie Groh, Christine Blattner, et Jean Marc Egly. 2008. « Nucleotide Excision Repair Driven by the Dissociation of CAK from TFIIH ». *Molecular Cell* 31 (1): 9-20. <https://doi.org/10.1016/j.molcel.2008.04.024>.
- Collart, Martine A, et Benjamin Weiss. 2020. « Ribosome pausing, a dangerous necessity for co-translational events ». *Nucleic Acids Research* 48 (3): 1043-55. <https://doi.org/10.1093/nar/gkz763>.
- Colton, Stephanie L., Xiaoxin S. Xu, Y. Alan Wang, et Gan Wang. 2006. « The Involvement of Ataxia-telangiectasia Mutated Protein Activation in Nucleotide Excision Repair-facilitated Cell Survival with Cisplatin Treatment\* ». *Journal of Biological Chemistry* 281 (37): 27117-25. <https://doi.org/10.1074/jbc.M602826200>.
- Critchlow, S. E., R. P. Bowater, et S. P. Jackson. 1997. « Mammalian DNA Double-Strand Break Repair Protein XRCC4 Interacts with DNA Ligase IV ». *Current Biology: CB* 7 (8): 588-98. [https://doi.org/10.1016/s0960-9822\(06\)00258-2](https://doi.org/10.1016/s0960-9822(06)00258-2).
- Crossley, Madzia P., Michael Bocek, et Karlene A. Cimprich. 2019. « R-Loops as Cellular Regulators and Genomic Threats ». *Molecular Cell* 73 (3): 398-411. <https://doi.org/10.1016/j.molcel.2019.01.024>.
- Cugusi, Simona, Richard Mitter, Gavin P. Kelly, Jane Walker, Zhong Han, Paola Pisano, Michael Wierer, Aengus Stewart, et Jesper Q. Svejstrup. 2022. « Heat Shock Induces Premature Transcript Termination and Reconfigures the Human Transcriptome ». *Molecular Cell* 82 (8): 1573-1588.e10. <https://doi.org/10.1016/j.molcel.2022.01.007>.
- Cullinane, Carleen, Sharlyn J. Mazur, John M. Essigmann, Don R. Phillips, et Vilhelm A. Bohr. 1999. « Inhibition of RNA Polymerase II Transcription in Human Cell Extracts by Cisplatin DNA Damage ». *Biochemistry* 38 (19): 6204-12. <https://doi.org/10.1021/bi982685+>.
- Dasari, Shaloam, et Paul Bernard Tchounwou. 2014. « Cisplatin in cancer therapy: molecular mechanisms of action ». *European journal of pharmacology* 740 (octobre): 364-78. <https://doi.org/10.1016/j.ejphar.2014.07.025>.
- Davidson, Lee, Lisa Muniz, et Steven West. 2014. « 3' End Formation of Pre-mRNA and Phosphorylation of Ser2 on the RNA Polymerase II CTD Are Reciprocally Coupled in Human Cells ». *Genes & Development* 28 (4): 342-56. <https://doi.org/10.1101/gad.231274.113>.
- Davies, M. S., S. J. Berners-Price, et T. W. Hambley. 2000. « Slowing of Cisplatin Aquation in the Presence of DNA but Not in the Presence of Phosphate: Improved Understanding of Sequence Selectivity and the Roles of Monoaquated and Diaquated Species in the Binding of Cisplatin to DNA ». *Inorganic Chemistry* 39 (25): 5603-13. <https://doi.org/10.1021/ic000847w>.
- Dearnaley, D. P., A. Horwich, R. A'Hern, J. Nicholls, G. Jay, W. F. Hendry, et M. J. Peckham. 1991. « Combination Chemotherapy with Bleomycin, Etoposide and Cisplatin (BEP) for Metastatic Testicular Teratoma: Long-Term Follow-Up ». *European Journal of Cancer (Oxford, England: 1990)* 27 (6): 684-91. [https://doi.org/10.1016/0277-5379\(91\)90166-b](https://doi.org/10.1016/0277-5379(91)90166-b).



- Decatris, M. P., S. Sundar, et K. J. O'Byrne. 2004. « Platinum-Based Chemotherapy in Metastatic Breast Cancer: Current Status ». *Cancer Treatment Reviews* 30 (1): 53-81. [https://doi.org/10.1016/S0305-7372\(03\)00139-7](https://doi.org/10.1016/S0305-7372(03)00139-7).
- Deriano, Ludovic, et David B. Roth. 2013. « Modernizing the Nonhomologous End-Joining Repertoire: Alternative and Classical NHEJ Share the Stage ». *Annual Review of Genetics* 47 (1): 433-55. <https://doi.org/10.1146/annurev-genet-110711-155540>.
- Devany, Emral, Ji Yeon Park, Michael R. Murphy, George Zakusilo, Jorge Baquero, Xiaokan Zhang, Mainul Hoque, Bin Tian, et Frida E. Kleiman. 2016. « Intronic Cleavage and Polyadenylation Regulates Gene Expression during DNA Damage Response through U1 snRNA ». *Cell Discovery* 2 (1): 1-14. <https://doi.org/10.1038/celldisc.2016.13>.
- Dhar, Shanta, Nagesh Kolishetti, Stephen J. Lippard, et Omid C. Farokhzad. 2011. « Targeted Delivery of a Cisplatin Prodrug for Safer and More Effective Prostate Cancer Therapy in Vivo ». *Proceedings of the National Academy of Sciences of the United States of America* 108 (5): 1850-55. <https://doi.org/10.1073/pnas.1011379108>.
- Dharmalingam, Prakash, Rajasekaran Mahalingam, Hari Krishna Yalamanchili, Tingting Weng, Harry Karmouty-Quintana, Ashrith Guha, et Rajarajan A. Thandavarayan. 2021. « Emerging Roles of Alternative Cleavage and Polyadenylation (APA) in Human Disease ». *Journal of Cellular Physiology* n/a (n/a). <https://doi.org/10.1002/jcp.30549>.
- Di Giammartino, Dafne Campigli, Kensei Nishida, et James L. Manley. 2011. « Mechanisms and Consequences of Alternative Polyadenylation ». *Molecular Cell* 43 (6): 853-66. <https://doi.org/10.1016/j.molcel.2011.08.017>.
- Ding, Qi, Yeturu V. R. Reddy, Wei Wang, Timothy Woods, Pauline Douglas, Dale A. Ramsden, Susan P. Lees-Miller, et Katheryn Meek. 2003. « Autophosphorylation of the Catalytic Subunit of the DNA-Dependent Protein Kinase Is Required for Efficient End Processing during DNA Double-Strand Break Repair ». *Molecular and Cellular Biology* 23 (16): 5836-48. <https://doi.org/10.1128/MCB.23.16.5836-5848.2003>.
- Dixon-Clarke, Sarah E., Jonathan M. Elkins, S.-W. Grace Cheng, Gregg B. Morin, et Alex N. Bullock. 2015. « Structures of the CDK12/CycK complex with AMP-PNP reveal a flexible C-terminal kinase extension important for ATP binding ». *Scientific Reports* 5 (novembre): 17122. <https://doi.org/10.1038/srep17122>.
- Drapkin, Ronny, Joyce T. Reardon, Athar Ansari, Juch-Chin Huang, Leigh Zawel, KyuJeong Ahn, Aziz Sancar, et Danny Reinberg. 1994. « Dual Role of TFIIH in DNA Excision Repair and in Transcription by RNA Polymerase II ». *Nature* 368 (6473): 769-72. <https://doi.org/10.1038/368769a0>.
- Dubbury, Sara J., Paul L. Boutz, et Phillip A. Sharp. 2018. « CDK12 Regulates DNA Repair Genes by Suppressing Intronic Polyadenylation ». *Nature* 564 (7734): 141-45. <https://doi.org/10.1038/s41586-018-0758-y>.
- Dujardin, Gwendal, Élisabeth Daguene, Delphine G. Bernard, Marion Flodrops, Stéphanie Durand, Aurélie Chauveau, Flaria El Khoury, Catherine Le Jossic-Corcus, et Laurent Corcos. 2016. « L'épissage des ARN pré-messagers : quand le splicéosome perd pied ». *médecine/sciences* 32 (12): 1103-10. <https://doi.org/10.1051/medsci/20163212014>.

- Durante, M., J. S. Bedford, D. J. Chen, S. Conrad, M. N. Cornforth, A. T. Natarajan, D. C. van Gent, et G. Obe. 2013. « From DNA Damage to Chromosome Aberrations: Joining the Break ». *Mutation Research* 756 (1-2): 5-13. <https://doi.org/10.1016/j.mrgentox.2013.05.014>.
- Dutertre, Martin, Fatima Zahra Chakrama, Emmanuel Combe, François-Olivier Desmet, Hussein Mortada, Micaela Polay Espinoza, Lise Gratadou, et Didier Auboeuf. 2014. « A Recently Evolved Class of Alternative 3'-Terminal Exons Involved in Cell Cycle Regulation by Topoisomerase Inhibitors ». *Nature Communications* 5 (1): 3395. <https://doi.org/10.1038/ncomms4395>.
- Dutertre, Martin, Sarah Lambert, Aura Carreira, Mounira Amor-Gu ret, et St phan Vagner. 2014. « DNA Damage: RNA-Binding Proteins Protect from near and Far ». *Trends in Biochemical Sciences* 39 (3): 141-49. <https://doi.org/10.1016/j.tibs.2014.01.003>.
- Dutertre, Martin, Gabriel Sanchez, Marie-C cile De Cian, J r me Barbier, Etienne Dardenne, Lise Gratadou, Gwendal Dujardin, Catherine Le Jossic-Corcos, Laurent Corcos, et Didier Auboeuf. 2010. « Cotranscriptional Exon Skipping in the Genotoxic Stress Response ». *Nature Structural & Molecular Biology* 17 (11): 1358-66. <https://doi.org/10.1038/nsmb.1912>.
- Dutertre, Martin, Rym Sfaxi, et St phan Vagner. 2021. « Reciprocal Links between Pre-Messenger RNA 3'-End Processing and Genome Stability ». *Trends in Biochemical Sciences* 46 (7): 579-94. <https://doi.org/10.1016/j.tibs.2021.01.009>.
- Dye, M. J., et N. J. Proudfoot. 2001. « Multiple Transcript Cleavage Precedes Polymerase Release in Termination by RNA Polymerase II ». *Cell* 105 (5): 669-81. [https://doi.org/10.1016/s0092-8674\(01\)00372-5](https://doi.org/10.1016/s0092-8674(01)00372-5).
- Ebbesen, Karoline K., Thomas B. Hansen, et J rgen Kjems. 2017. « Insights into circular RNA biology ». *RNA Biology* 14 (8): 1035-45. <https://doi.org/10.1080/15476286.2016.1271524>.
- Eckmann, Christian R., Christiane Rammelt, et Elmar Wahle. 2011. « Control of Poly(A) Tail Length ». *WIREs RNA* 2 (3): 348-61. <https://doi.org/10.1002/wrna.56>.
- Edwards-Gilbert, Gretchen, Kristen L. Veraldi, et Christine Milcarek. 1997. « Alternative poly(A) site selection in complex transcription units: means to an end? ». *Nucleic Acids Research* 25 (13): 2547-61. <https://doi.org/10.1093/nar/25.13.2547>.
- Ekumi, Kingsley M., Hana Paculova, Tina Lenasi, Vendula Pospichalova, Christian A. B sken, Jana Rybarikova, Vitezslav Bryja, Matthias Geyer, Dalibor Blazek, et Matjaz Barboric. 2015. « Ovarian Carcinoma CDK12 Mutations Misregulate Expression of DNA Repair Genes via Deficient Formation and Function of the Cdk12/CycK Complex ». *Nucleic Acids Research* 43 (5): 2575-89. <https://doi.org/10.1093/nar/gkv101>.
- Engreitz, Jesse M., Klara Sirokman, Patrick McDonel, Alexander A. Shishkin, Christine Surka, Pamela Russell, Sharon R. Grossman, Amy Y. Chow, Mitchell Guttman, et Eric S. Lander. 2014. « RNA-RNA Interactions Enable Specific Targeting of Noncoding RNAs to Nascent Pre-mRNAs and Chromatin Sites ». *Cell* 159 (1): 188-99. <https://doi.org/10.1016/j.cell.2014.08.018>.

- Enoiu, Milica, Josef Jiricny, et Orlando D. Schärer. 2012. « Repair of cisplatin-induced DNA interstrand crosslinks by a replication-independent pathway involving transcription-coupled repair and translesion synthesis ». *Nucleic Acids Research* 40 (18): 8953-64. <https://doi.org/10.1093/nar/gks670>.
- Fan, Zheng, Jennifer R. Devlin, Simon J. Hogg, Maria A. Doyle, Paul F. Harrison, Izabela Todorovski, Leonie A. Cluse, Deborah A. Knight, Jarrod J. Sandow, Gareth Gregory, Andrew Fox, Traude H. Beilharz, Nicholas Kwiatkowski, Nichollas E. Scott, Ana Tufegdžic Vidakovic, Gavin P. Kelly, Jesper Q. Svejstrup, Matthias Geyer, Nathanael S. Gray, Stephin J. Vervoort, et Ricky W. Johnstone. 2020. « CDK13 cooperates with CDK12 to control global RNA polymerase II processivity ». *Science Advances* 6 (18): eaz5041. <https://doi.org/10.1126/sciadv.aaz5041>.
- Fichtinger-Schepman, A. M., J. L. van der Veer, J. H. den Hartog, P. H. Lohman, et J. Reedijk. 1985. « Adducts of the Antitumor Drug Cis-Diamminedichloroplatinum(II) with DNA: Formation, Identification, and Quantitation ». *Biochemistry* 24 (3): 707-13. <https://doi.org/10.1021/bi00324a025>.
- Fishel, R., et T. Wilson. 1997. « MutS Homologs in Mammalian Cells ». *Current Opinion in Genetics & Development* 7 (1): 105-13. [https://doi.org/10.1016/s0959-437x\(97\)80117-7](https://doi.org/10.1016/s0959-437x(97)80117-7).
- Florea, Ana-Maria, et Dietrich Büsselberg. 2011. « Cisplatin as an Anti-Tumor Drug: Cellular Mechanisms of Activity, Drug Resistance and Induced Side Effects ». *Cancers* 3 (1): 1351-71. <https://doi.org/10.3390/cancers3011351>.
- Fousteri, Maria, et Leon HF Mullenders. 2008. « Transcription-Coupled Nucleotide Excision Repair in Mammalian Cells: Molecular Mechanisms and Biological Effects ». *Cell Research* 18 (1): 73-84. <https://doi.org/10.1038/cr.2008.6>.
- Fuertes, Miguel A., Carlos Alonso, et José M. Pérez. 2003. « Biochemical Modulation of Cisplatin Mechanisms of Action: Enhancement of Antitumor Activity and Circumvention of Drug Resistance ». *Chemical Reviews* 103 (3): 645-62. <https://doi.org/10.1021/cr020010d>.
- Gabriel, Maude, Yves Delforge, Adeline Deward, Yvette Habraken, Benoit Hennuy, Jacques Piette, Roscoe Klinck, Benoit Chabot, Alain Colige, et Charles Lambert. 2015. « Role of the Splicing Factor SRSF4 in Cisplatin-Induced Modifications of Pre-mRNA Splicing and Apoptosis ». *BMC Cancer* 15 (1): 227. <https://doi.org/10.1186/s12885-015-1259-0>.
- Galluzzi, L, L Senovilla, I Vitale, J Michels, I Martins, O Kepp, M Castedo, et G Kroemer. 2012. « Molecular Mechanisms of Cisplatin Resistance ». *Oncogene* 31 (15): 1869-83. <https://doi.org/10.1038/onc.2011.384>.
- Germann, Susanne M., Vibe H. Oestergaard, Caroline Haas, Pauline Salis, Akira Motegi, et Michael Lisby. 2011. « Dpb11/TopBP1 plays distinct roles in DNA replication, checkpoint response and homologous recombination ». *DNA Repair* 10 (2): 210-24. <https://doi.org/10.1016/j.dnarep.2010.11.001>.
- Ghosh, Sumit. 2019. « Cisplatin: The first metal based anticancer drug ». *Bioorganic Chemistry* 88 (juillet): 102925. <https://doi.org/10.1016/j.bioorg.2019.102925>.

- Gibson, Toby J., Markus Seiler, et Reiner A. Veitia. 2013. « The Transience of Transient Overexpression ». *Nature Methods* 10 (8): 715-21. <https://doi.org/10.1038/nmeth.2534>.
- Giuliani, Virginia, Meredith A. Miller, Chiu-Yi Liu, Stella R. Hartono, Caleb A. Class, Christopher A. Bristow, Erika Suzuki, Lionel A. Sanz, Guang Gao, Jason P. Gay, Ningping Feng, Johnathon L. Rose, Hideo Tomihara, Joseph R. Daniele, Michael D. Peoples, Jennifer P. Bardenhagen, Mary K. Geck Do, Qing E. Chang, Bhavatarini Vangamudi, Christopher Vellano, Haoqiang Ying, Angela K. Deem, Kim-Anh Do, Giannicola Genovese, Joseph R. Marszalek, Jeffrey J. Kovacs, Michael Kim, Jason B. Fleming, Ernesto Guccione, Andrea Viale, Anirban Maitra, M. Emilia Di Francesco, Timothy A. Yap, Philip Jones, Giulio Draetta, Alessandro Carugo, Frederic Chedin, et Timothy P. Heffernan. 2021. « PRMT1-Dependent Regulation of RNA Metabolism and DNA Damage Response Sustains Pancreatic Ductal Adenocarcinoma ». *Nature Communications* 12 (1): 4626. <https://doi.org/10.1038/s41467-021-24798-y>.
- Goellner, Eva M., Catherine E. Smith, Christopher S. Campbell, Hans Hombauer, Arshad Desai, Christopher D. Putnam, et Richard D. Kolodner. 2014. « PCNA and Msh2-Msh6 Activate an Mlh1-Pms1 Endonuclease Pathway Required for Exo1-Independent Mismatch Repair ». *Molecular Cell* 55 (2): 291-304. <https://doi.org/10.1016/j.molcel.2014.04.034>.
- Gottlieb, T. M., et S. P. Jackson. 1993. « The DNA-Dependent Protein Kinase: Requirement for DNA Ends and Association with Ku Antigen ». *Cell* 72 (1): 131-42. [https://doi.org/10.1016/0092-8674\(93\)90057-w](https://doi.org/10.1016/0092-8674(93)90057-w).
- Graber, Joel H., Fathima I. Nazeer, Pei-chun Yeh, Jason N. Kuehner, Sneha Borikar, Derick Hoskinson, et Claire L. Moore. 2013. « DNA Damage Induces Targeted, Genome-Wide Variation of Poly(A) Sites in Budding Yeast ». *Genome Research* 23 (10): 1690-1703. <https://doi.org/10.1101/gr.144964.112>.
- Grange, Pierre de la. 2016. « Altérations de l'épissage et maladies rares ». *médecine/sciences* 32 (12): 1111-19. <https://doi.org/10.1051/medsci/20163212015>.
- Grawunder, U., M. Wilm, X. Wu, P. Kulesza, T. E. Wilson, M. Mann, et M. R. Lieber. 1997. « Activity of DNA Ligase IV Stimulated by Complex Formation with XRCC4 Protein in Mammalian Cells ». *Nature* 388 (6641): 492-95. <https://doi.org/10.1038/41358>.
- Gregersen, Lea H., Richard Mitter, Alejandro P. Ugalde, Takayuki Nojima, Nicholas J. Proudfoot, Reuven Agami, Aengus Stewart, et Jesper Q. Svejstrup. 2019. « SCAF4 and SCAF8, mRNA Anti-Terminator Proteins ». *Cell* 177 (7): 1797-1813.e18. <https://doi.org/10.1016/j.cell.2019.04.038>.
- Grosso, Ana R, Ana P Leite, Sílvia Carvalho, Mafalda R Matos, Filipa B Martins, Alexandra C Vítor, Joana MP Desterro, Maria Carmo-Fonseca, et Sérgio F de Almeida. 2015. « Pervasive transcription read-through promotes aberrant expression of oncogenes and RNA chimeras in renal carcinoma ». Édité par Chi Van Dang. *eLife* 4 (novembre): e09214. <https://doi.org/10.7554/eLife.09214>.

- Grubb, A, et H Löfberg. 1982. « Human gamma-trace, a basic microprotein: amino acid sequence and presence in the adenohypophysis. » *Proceedings of the National Academy of Sciences* 79 (9): 3024-27. <https://doi.org/10.1073/pnas.79.9.3024>.
- Gruber, Andreas J., et Mihaela Zavolan. 2019. « Alternative Cleavage and Polyadenylation in Health and Disease ». *Nature Reviews Genetics* 20 (10): 599-614. <https://doi.org/10.1038/s41576-019-0145-z>.
- Gunderson, S. I., M. Polycarpou-Schwarz, et I. W. Mattaj. 1998. « U1 snRNP Inhibits Pre-mRNA Polyadenylation through a Direct Interaction between U1 70K and Poly(A) Polymerase ». *Molecular Cell* 1 (2): 255-64. [https://doi.org/10.1016/s1097-2765\(00\)80026-x](https://doi.org/10.1016/s1097-2765(00)80026-x).
- Hanaoka, Fumio, et Kaoru Sugasawa, éd. 2016. *DNA Replication, Recombination, and Repair: Molecular Mechanisms and Pathology*. Tokyo: Springer Japan. <https://doi.org/10.1007/978-4-431-55873-6>.
- Haracska, Lajos, Ildiko Unk, Robert E. Johnson, Erik Johansson, Peter M.J. Burgers, Satya Prakash, et Louise Prakash. 2001. « Roles of yeast DNA polymerases  $\delta$  and  $\zeta$  and of Rev1 in the bypass of abasic sites ». *Genes & Development* 15 (8): 945-54. <https://doi.org/10.1101/gad.882301>.
- Harrington, Jill M., et Richard D. Kolodner. 2007. « Saccharomyces Cerevisiae Msh2-Msh3 Acts in Repair of Base-Base Mismatches ». *Molecular and Cellular Biology* 27 (18): 6546-54. <https://doi.org/10.1128/MCB.00855-07>.
- Hartwig, Andrea, Michael Arand, Bernd Epe, Sabine Guth, Gunnar Jahnke, Alfonso Lampen, Hans-Jörg Martus, Bernhard Monien, Ivonne M. C. M. Rietjens, Simone Schmitz-Spanke, Gerlinde Schriever-Schwemmer, Pablo Steinberg, et Gerhard Eisenbrand. 2020. « Mode of Action-Based Risk Assessment of Genotoxic Carcinogens ». *Archives of Toxicology* 94 (6): 1787-1877. <https://doi.org/10.1007/s00204-020-02733-2>.
- Hassel, Keira R., Omar Brito-Estrada, et Catherine A. Makarewich. 2023. « Microproteins: Overlooked Regulators of Physiology and Disease ». *iScience* 26 (6): 106781. <https://doi.org/10.1016/j.isci.2023.106781>.
- Hatchi, Elodie, Konstantina Skourti-Stathaki, Steffen Ventsch, Luca Pinello, Angela Yen, Kinga Kamieniarz-Gdula, Stoil Dimitrov, Shailja Pathania, Kristine M. McKinney, Matthew L. Eaton, Manolis Kellis, Sarah J. Hill, Giovanni Parmigiani, Nicholas J. Proudfoot, et David M. Livingston. 2015. « BRCA1 Recruitment to Transcriptional Pause Sites Is Required for R-Loop-Driven DNA Damage Repair ». *Molecular Cell* 57 (4): 636-47. <https://doi.org/10.1016/j.molcel.2015.01.011>.
- Heesch, Sebastiaan van, Franziska Witte, Valentin Schneider-Lunitz, Jana F. Schulz, Eleonora Adami, Allison B. Faber, Marieluise Kirchner, Henrike Maatz, Susanne Blachut, Clara-Louisa Sandmann, Masatoshi Kanda, Catherine L. Worth, Sebastian Schafer, Lorenzo Calviello, Rhys Merriott, Giannino Patone, Oliver Hummel, Emanuel Wyler, Benedikt Obermayer, Michael B. Mücke, Eric L. Lindberg, Franziska Trnka, Sebastian Memczak, Marcel Schilling, Leanne E. Felkin, Paul J. R. Barton, Nicholas M. Quafe, Konstantinos Vanezis, Sebastian Diecke, Masaya Mukai, Nancy Mah, Su-Jun Oh, Andreas Kurtz, Christoph Schramm, Dorothee Schwinge, Marcial Sebode, Magdalena Harakalova,

- Folkert W. Asselbergs, Aryan Vink, Roel A. de Weger, Sivakumar Viswanathan, Anissa A. Widjaja, Anna Gärtner-Rommel, Hendrik Milting, Cris Dos Remedios, Christoph Knosalla, Philipp Mertins, Markus Landthaler, Martin Vingron, Wolfgang A. Linke, Jonathan G. Seidman, Christine E. Seidman, Nikolaus Rajewsky, Uwe Ohler, Stuart A. Cook, et Norbert Hubner. 2019. « The Translational Landscape of the Human Heart ». *Cell* 178 (1): 242-260.e29. <https://doi.org/10.1016/j.cell.2019.05.010>.
- Hoang, Margaret L., Frederick J. Tan, David C. Lai, Sue E. Celniker, Roger A. Hoskins, Maitreya J. Dunham, Yixian Zheng, et Douglas Koshland. 2010. « Competitive Repair by Naturally Dispersed Repetitive DNA during Non-Allelic Homologous Recombination ». *PLoS Genetics* 6 (12). <https://doi.org/10.1371/journal.pgen.1001228>.
- Hollerer, Ina, Tomaz Curk, Bettina Haase, Vladimir Benes, Christian Hauer, Gabriele Neu-Yilik, Madhuri Bhuvanagiri, Matthias W. Hentze, et Andreas E. Kulozik. 2016. « The Differential Expression of Alternatively Polyadenylated Transcripts Is a Common Stress-Induced Response Mechanism That Modulates Mammalian mRNA Expression in a Quantitative and Qualitative Fashion ». *RNA* 22 (9): 1441-53. <https://doi.org/10.1261/rna.055657.115>.
- Hon, Chung-Chau, Jordan A. Ramilowski, Jayson Harshbarger, Nicolas Bertin, Owen J. L. Rackham, Julian Gough, Elena Denisenko, Sebastian Schmeier, Thomas M. Poulsen, Jessica Severin, Marina Lizio, Hideya Kawaji, Takeya Kasukawa, Masayoshi Itoh, A. Maxwell Burroughs, Shohei Noma, Sarah Djebali, Tanvir Alam, Yulia A. Medvedeva, Alison C. Testa, Leonard Lipovich, Chi-Wai Yip, Imad Abugessaisa, Mickaël Mendez, Akira Hasegawa, Dave Tang, Timo Lassmann, Peter Heutink, Magda Babina, Christine A. Wells, Soichi Kojima, Yukio Nakamura, Harukazu Suzuki, Carsten O. Daub, Michiel J. L. de Hoon, Erik Arner, Yoshihide Hayashizaki, Piero Carninci, et Alistair R. R. Forrest. 2017. « An Atlas of Human Long Non-Coding RNAs with Accurate 5' Ends ». *Nature* 543 (7644): 199-204. <https://doi.org/10.1038/nature21374>.
- Hoque, Mainul, Zhe Ji, Dinghai Zheng, Wenting Luo, Wencheng Li, Bei You, Ji Yeon Park, Ghassan Yehia, et Bin Tian. 2013. « Analysis of Alternative Cleavage and Polyadenylation by 3' Region Extraction and Deep Sequencing ». *Nature Methods* 10 (2): 133-39. <https://doi.org/10.1038/nmeth.2288>.
- Hsin, Jing-Ping, et James L. Manley. 2012. « The RNA Polymerase II CTD Coordinates Transcription and RNA Processing ». *Genes & Development* 26 (19): 2119-37. <https://doi.org/10.1101/gad.200303.112>.
- Hsu, Polly Yingshan, Lorenzo Calviello, Hsin-Yen Larry Wu, Fay-Wei Li, Carl J. Rothfels, Uwe Ohler, et Philip N. Benfey. 2016. « Super-resolution ribosome profiling reveals unannotated translation events in Arabidopsis ». *Proceedings of the National Academy of Sciences* 113 (45): E7126-35. <https://doi.org/10.1073/pnas.1614788113>.
- Hung, Putzer J., Britney Johnson, Bo-Ruei Chen, Andrea K. Byrum, Andrea L. Bredemeyer, William T. Yewdell, Tanya E. Johnson, Brian J. Lee, Shruthi Deivasigamani, Issa Hindi, Parmeshwar Amatya, Michael L. Gross, Tanya T. Paull, David J. Pisapia, Jayanta Chaudhuri, John J. H. Petrini, Nima Mosammamarast, Gaya K. Amarasinghe, Shan Zha, Jessica K. Tyler, et Barry P. Sleckman. 2018. « MRI Is a DNA Damage Response Adaptor during Classical Non-Homologous End Joining ». *Molecular Cell* 71 (2): 332-342.e8. <https://doi.org/10.1016/j.molcel.2018.06.018>.

- Ingolia, Nicholas T., Sina Ghaemmaghami, John R. S. Newman, et Jonathan S. Weissman. 2009. « Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling ». *Science*, avril. <https://doi.org/10.1126/science.1168978>.
- Ip, Joanna Y., Dominic Schmidt, Qun Pan, Arun K. Ramani, Andrew G. Fraser, Duncan T. Odom, et Benjamin J. Blencowe. 2011. « Global Impact of RNA Polymerase II Elongation Inhibition on Alternative Splicing Regulation ». *Genome Research* 21 (3): 390-401. <https://doi.org/10.1101/gr.111070.110>.
- Ira, Grzegorz, Anna Malkova, Giordano Liberi, Marco Foiani, et James E. Haber. 2003. « Srs2 and Sgs1–Top3 Suppress Crossovers during Double-Strand Break Repair in Yeast ». *Cell* 115 (4): 401-11.
- Ismailov, Z. B., E. S. Belykh, A. A. Chernykh, A. M. Udoratina, D. V. Kazakov, A. V. Rybak, S. N. Kerimova, et I. O. Velegzhaninov. 2023. « Systematic review of comparative transcriptomic studies of cellular resistance to genotoxic stress ». *Mutation Research/Reviews in Mutation Research* 792 (juillet): 108467. <https://doi.org/10.1016/j.mrrev.2023.108467>.
- Iyer, Divya, et Nicholas Rhind. 2017. « The Intra-S Checkpoint Responses to DNA Damage ». *Genes* 8 (2): 74. <https://doi.org/10.3390/genes8020074>.
- Jacob, Aishwarya G., et Christopher W. J. Smith. 2017. « Intron Retention as a Component of Regulated Gene Expression Programs ». *Human Genetics* 136 (9): 1043-57. <https://doi.org/10.1007/s00439-017-1791-x>.
- Jin, Jianping, Takahiro Shirogane, Lai Xu, Grzegorz Nalepa, Jun Qin, Stephen J. Elledge, et J. Wade Harper. 2003. « SCF $\beta$ -TRCP links Chk1 signaling to degradation of the Cdc25A protein phosphatase ». *Genes & Development* 17 (24): 3062-74. <https://doi.org/10.1101/gad.1157503>.
- Johnson, Shawn F., Cristina Cruz, Ann Katrin Greifenberg, Sofia Dust, Daniel G. Stover, David Chi, Benjamin Primack, Shiliang Cao, Andrea J. Bernhardt, Rhiannon Coulson, Jean-Bernard Lazaro, Bose Kochupurakkal, Heather Sun, Christine Unitt, Lisa A. Moreau, Kristopher A. Sarosiek, Maurizio Scaltriti, Dejan Juric, José Baselga, Andrea L. Richardson, Scott J. Rodig, Alan D. D'Andrea, Judith Balmaña, Neil Johnson, Matthias Geyer, Violeta Serra, Elgene Lim, et Geoffrey I. Shapiro. 2016. « CDK12 Inhibition Reverses De Novo and Acquired PARP Inhibitor Resistance in BRCA Wild-Type and Mutated Models of Triple-Negative Breast Cancer ». *Cell Reports* 17 (9): 2367-81. <https://doi.org/10.1016/j.celrep.2016.10.077>.
- Joshi, Poorval M., Shari L. Sutor, Catherine J. Huntoon, et Larry M. Karnitz. 2014. « Ovarian Cancer-Associated Mutations Disable Catalytic Activity of CDK12, a Kinase That Promotes Homologous Recombination Repair and Resistance to Cisplatin and Poly(ADP-Ribose) Polymerase Inhibitors ». *The Journal of Biological Chemistry* 289 (13): 9247-53. <https://doi.org/10.1074/jbc.M114.551143>.
- Kaida, Daisuke, Michael G. Berg, Ihab Younis, Mumtaz Kasim, Larry N. Singh, Lili Wan, et Gideon Dreyfuss. 2010. « U1 snRNP Protects Pre-mRNAs from Premature Cleavage and Polyadenylation ». *Nature* 468 (7324): 664-68. <https://doi.org/10.1038/nature09479>.

- Kamieniarz-Gdula, Kinga, et Nick J. Proudfoot. 2019. « Transcriptional Control by Premature Termination: A Forgotten Mechanism ». *Trends in Genetics* 35 (8): 553-64. <https://doi.org/10.1016/j.tig.2019.05.005>.
- Kanwal, Rajnee, Mitali Pandey, Natarajan Bhaskaran, Gregory T. MacLennan, Pingfu Fu, Lee E. Ponsky, et Sanjay Gupta. 2014. « Protection against Oxidative DNA Damage and Stress in Human Prostate by Glutathione S-Transferase P1 ». *Molecular Carcinogenesis* 53 (1): 8-18. <https://doi.org/10.1002/mc.21939>.
- Kastenhuber, Edward R., et Scott W. Lowe. 2017. « Putting P53 in Context ». *Cell* 170 (6): 1062-78. <https://doi.org/10.1016/j.cell.2017.08.028>.
- Kazlauskas, Darius, Mart Krupovic, Julien Guglielmini, Patrick Forterre, et Āeslovas Venclovas. 2020. « Diversity and evolution of B-family DNA polymerases ». *Nucleic Acids Research* 48 (18): 10142-56. <https://doi.org/10.1093/nar/gkaa760>.
- Ke, Shengdong, Endalkachew A. Alemu, Claudia Mertens, Emily Conn Gantman, John J. Fak, Aldo Mele, Bhagwattie Haripal, Ilana Zucker-Scharff, Michael J. Moore, Christopher Y. Park, Cathrine Broberg Vågbø, Anna Kusńnierczyk, Arne Klungland, James E. Darnell, et Robert B. Darnell. 2015. « A Majority of m6A Residues Are in the Last Exons, Allowing the Potential for 3' UTR Regulation ». *Genes & Development* 29 (19): 2037-53. <https://doi.org/10.1101/gad.269415.115>.
- Kearse, Michael G., et Jeremy E. Wilusz. 2017. « Non-AUG Translation: A New Start for Protein Synthesis in Eukaryotes ». *Genes & Development* 31 (17): 1717-31. <https://doi.org/10.1101/gad.305250.117>.
- Kelemen, Olga, Paolo Convertini, Zhaiyi Zhang, Yuan Wen, Manli Shen, Marina Falaleeva, et Stefan Stamm. 2013. « Function of alternative splicing ». *Gene* 514 (1): 1-30. <https://doi.org/10.1016/j.gene.2012.07.083>.
- Kesner, Jordan S., Ziheng Chen, Peiguo Shi, Alexis O. Aparicio, Michael R. Murphy, Yang Guo, Aditi Trehan, Jessica E. Lipponen, Yocelyn Recinos, Natura Myeku, et Xuebing Wu. 2023. « Noncoding Translation Mitigation ». *Nature* 617 (7960): 395-402. <https://doi.org/10.1038/s41586-023-05946-4>.
- Khan, A. B., B. J. D'Souza, M. D. Wharam, L. A. Champion, L. F. Sinks, S. Y. Woo, D. C. McCullough, et B. G. Leventhal. 1982. « Cisplatin Therapy in Recurrent Childhood Brain Tumors ». *Cancer Treatment Reports* 66 (12): 2013-20.
- Kim, Hyungjin, et Alan D. D'Andrea. 2012. « Regulation of DNA Cross-Link Repair by the Fanconi Anemia/BRCA Pathway ». *Genes & Development* 26 (13): 1393-1408. <https://doi.org/10.1101/gad.195248.112>.
- Knipscheer, Puck, Markus Räschele, Agata Smogorzewska, Milica Enoiu, The Vinh Ho, Orlando D. Schärer, Stephen J. Elledge, et Johannes C. Walter. 2009. « The Fanconi Anemia Pathway Promotes Replication-Dependent DNA Interstrand Cross-Link Repair ». *Science (New York, N.Y.)* 326 (5960): 1698-1701. <https://doi.org/10.1126/science.1182372>.
- Koizumi, Wasaburo, Hiroyuki Narahara, Takuo Hara, Akinori Takagane, Toshikazu Akiya, Masakazu Takagi, Kosei Miyashita, Takashi Nishizaki, Osamu Kobayashi, Wataru



- Takiyama, Yasushi Toh, Takashi Nagaie, Seiichi Takagi, Yoshitaka Yamamura, Kimihiko Yanaoka, Hiroyuki Orita, et Masahiro Takeuchi. 2008. « S-1 plus Cisplatin versus S-1 Alone for First-Line Treatment of Advanced Gastric Cancer (SPIRITS Trial): A Phase III Trial ». *The Lancet. Oncology* 9 (3): 215-21. [https://doi.org/10.1016/S1470-2045\(08\)70035-4](https://doi.org/10.1016/S1470-2045(08)70035-4).
- Kosmas, C., N. B. Tsavaris, N. A. Malamos, M. Vadiaka, et C. Koufos. 2001. « Phase II Study of Paclitaxel, Ifosfamide, and Cisplatin as Second-Line Treatment in Relapsed Small-Cell Lung Cancer ». *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology* 19 (1): 119-26. <https://doi.org/10.1200/JCO.2001.19.1.119>.
- Kottemann, Molly C., et Agata Smogorzewska. 2013. « Fanconi anemia and the repair of Watson and Crick crosslinks ». *Nature* 493 (7432): 356-63. <https://doi.org/10.1038/nature11863>.
- Krajewska, Malgorzata, Ruben Dries, Andrew V. Grasseti, Sofia Dust, Yang Gao, Hao Huang, Bandana Sharma, Daniel S. Day, Nicholas Kwiatkowski, Monica Pomaville, Oliver Dodd, Edmond Chipumuro, Tinghu Zhang, Arno L. Greenleaf, Guo-Cheng Yuan, Nathanael S. Gray, Richard A. Young, Matthias Geyer, Scott A. Gerber, et Rani E. George. 2019. « CDK12 Loss in Cancer Cells Affects DNA Damage Response Genes through Premature Cleavage and Polyadenylation ». *Nature Communications* 10 (1): 1757. <https://doi.org/10.1038/s41467-019-09703-y>.
- Krejci, Lumir, Veronika Altmannova, Mario Spirek, et Xiaolan Zhao. 2012. « Homologous recombination and its regulation ». *Nucleic Acids Research* 40 (13): 5795-5818. <https://doi.org/10.1093/nar/gks270>.
- Kwak, Hojoong, et John T. Lis. 2013. « Control of Transcriptional Elongation ». *Annual Review of Genetics* 47 (1): 483-508. <https://doi.org/10.1146/annurev-genet-110711-155440>.
- Kwon, Buki, Mervin M. Fansler, Neil D. Patel, Jihye Lee, Weirui Ma, et Christine Mayr. 2022. « Enhancers Regulate 3' End Processing Activity to Control Expression of Alternative 3'UTR Isoforms ». *Nature Communications* 13 (1): 2709. <https://doi.org/10.1038/s41467-022-30525-y>.
- Lanz, Michael Charles, Diego Dibitetto, et Marcus Bustamante Smolka. 2019. « DNA damage kinase signaling: checkpoint and repair at 30 years ». *The EMBO Journal* 38 (18): e101801. <https://doi.org/10.15252/embj.2019101801>.
- Lau, Patrick J., et Richard D. Kolodner. 2003. « Transfer of the MSH2.MSH6 Complex from Proliferating Cell Nuclear Antigen to Mismatched Bases in DNA ». *The Journal of Biological Chemistry* 278 (1): 14-17. <https://doi.org/10.1074/jbc.C200627200>.
- Lee, Changhan, Jennifer Zeng, Brian G. Drew, Tamer Sallam, Alejandro Martin-Montalvo, Junxiang Wan, Su-Jeong Kim, Hemal Mehta, Andrea L. Hevener, Rafael de Cabo, et Pinchas Cohen. 2015. « The Mitochondrial-derived peptide MOTS-c promotes metabolic homeostasis and reduces obesity and insulin resistance ». *Cell metabolism* 21 (3): 443-54. <https://doi.org/10.1016/j.cmet.2015.02.009>.
- Lee, Cheryl Q. E., Baptiste Kerouanton, Sonia Chothani, Shan Zhang, Ying Chen, Chinmay Kumar Mantri, Daniella Helena Hock, Radiance Lim, Rhea Nadkarni, Vinh Thang Huynh, Daryl

- Lim, Wei Leong Chew, Franklin L. Zhong, David Arthur Stroud, Sebastian Schafer, Vinay Tergaonkar, Ashley L. St John, Owen J. L. Rackham, et Lena Ho. 2021. « Coding and Non-Coding Roles of MOCCI (C15ORF48) Coordinate to Regulate Host Inflammation and Immunity ». *Nature Communications* 12 (1): 2130. <https://doi.org/10.1038/s41467-021-22397-5>.
- Lee, Shih-Han, Irtisha Singh, Sarah Tisdale, Omar Abdel-Wahab, Christina S. Leslie, et Christine Mayr. 2018. « Widespread Intronic Polyadenylation Inactivates Tumour Suppressor Genes in Leukaemia ». *Nature* 561 (7721): 127-31. <https://doi.org/10.1038/s41586-018-0465-8>.
- Legnini, Ivano, Gaia Di Timoteo, Francesca Rossi, Mariangela Morlando, Francesca Briganti, Olga Sthandier, Alessandro Fatica, Tiziana Santini, Adrian Andronache, Mark Wade, Pietro Laneve, Nikolaus Rajewsky, et Irene Bozzoni. 2017. « Circ-ZNF609 Is a Circular RNA That Can Be Translated and Functions in Myogenesis ». *Molecular Cell* 66 (1): 22-37.e9. <https://doi.org/10.1016/j.molcel.2017.02.017>.
- Lemaitre, Florence, Fatima Chakrama, Tina O'Grady, Olivier Peulen, Gilles Rademaker, Adeline Deward, Benoit Chabot, Jacques Piette, Alain Colige, Charles Lambert, Franck Dequiedt, et Yvette Habraken. 2022. « The Transcription Factor C-Jun Inhibits RBM39 to Reprogram Pre-mRNA Splicing during Genotoxic Stress ». *Nucleic Acids Research* 50 (22): 12768-89. <https://doi.org/10.1093/nar/gkac1130>.
- Lengsfeld, Bettina M., Alison J. Rattray, Venugopal Bhaskara, Rodolfo Ghirlando, et Tanya T. Paull. 2007. « Sae2 is an endonuclease that processes hairpin DNA cooperatively with the Mre11/Rad50/Xrs2 complex ». *Molecular cell* 28 (4): 638-51. <https://doi.org/10.1016/j.molcel.2007.11.001>.
- Leong, Alyssa Zi-Xin, Pey Yee Lee, M. Aiman Mohtar, Saiful Effendi Syafruddin, Yuh-Fen Pung, et Teck Yew Low. 2022. « Short Open Reading Frames (sORFs) and Microproteins: An Update on Their Identification and Validation Measures ». *Journal of Biomedical Science* 29 (1): 19. <https://doi.org/10.1186/s12929-022-00802-5>.
- Li, Fei, Zaosong Zheng, Wei Chen, Dongqing Li, Henghui Zhang, Yuanchao Zhu, Qixin Mo, Xinlei Zhao, Qin Fan, Fan Deng, Conghui Han, et Wanlong Tan. 2023. « Regulation of cisplatin resistance in bladder cancer by epigenetic mechanisms ». *Drug Resistance Updates* 68 (mai): 100938. <https://doi.org/10.1016/j.drug.2023.100938>.
- Li, Weimin, Wencheng Li, Rakesh S. Laishram, Mainul Hoque, Zhe Ji, Bin Tian, et Richard A. Anderson. 2017. « Distinct regulation of alternative polyadenylation and gene expression by nuclear poly(A) polymerases ». *Nucleic Acids Research* 45 (15): 8930-42. <https://doi.org/10.1093/nar/gkx560>.
- Li, Wencheng, Bei You, Mainul Hoque, Dinghai Zheng, Wenting Luo, Zhe Ji, Ji Yeon Park, Samuel I. Gunderson, Auinash Kalsotra, James L. Manley, et Bin Tian. 2015. « Systematic Profiling of Poly(A)+ Transcripts Modulated by Core 3' End Processing and Splicing Factors Reveals Regulatory Rules of Alternative Cleavage and Polyadenylation ». *PLoS Genetics* 11 (4): e1005166. <https://doi.org/10.1371/journal.pgen.1005166>.

- Lieber, Michael R. 2010. « The Mechanism of Double-Strand DNA Break Repair by the Nonhomologous DNA End Joining Pathway ». *Annual review of biochemistry* 79: 181-211. <https://doi.org/10.1146/annurev.biochem.052308.093131>.
- Liu, Qinghua, Saritha Guntuku, Xian-Shu Cui, Shuhei Matsuoka, David Cortez, Katsuyuki Tamai, Guangbin Luo, Sandra Carattini-Rivera, Francisco DeMayo, Allan Bradley, Larry A. Donehower, et Stephen J. Elledge. 2000. « Chk1 is an essential kinase that is regulated by Atr and required for the G2/M DNA damage checkpoint ». *Genes & Development* 14 (12): 1448-59.
- Llorente, Bertrand, Catherine E. Smith, et Lorraine S. Symington. 2008. « Break-Induced Replication: What Is It and What Is It For? » *Cell Cycle (Georgetown, Tex.)* 7 (7): 859-64. <https://doi.org/10.4161/cc.7.7.5613>.
- Long, David T., Markus Räschle, Vladimir Joukov, et Johannes C. Walter. 2011. « Mechanism of RAD51-Dependent DNA Interstrand Cross-Link Repair ». *Science (New York, N.Y.)* 333 (6038): 84-87. <https://doi.org/10.1126/science.1204258>.
- Luco, Reini F., Qun Pan, Kaoru Tominaga, Benjamin J. Blencowe, Olivia M. Pereira-Smith, et Tom Misteli. 2010. « Regulation of Alternative Splicing by Histone Modifications ». *Science (New York, N.Y.)* 327 (5968): 996-1000. <https://doi.org/10.1126/science.1184208>.
- Luo, Wenting, Zhe Ji, Zhenhua Pan, Bei You, Mainul Hoque, Wencheng Li, Samuel I. Gunderson, et Bin Tian. 2013. « The Conserved Intronic Cleavage and Polyadenylation Site of CstF-77 Gene Imparts Control of 3' End Processing Activity through Feedback Autoregulation and by U1 snRNP ». *PLOS Genetics* 9 (7): e1003613. <https://doi.org/10.1371/journal.pgen.1003613>.
- Luo, Xiaotong, Yuantai Huang, Huiqin Li, Yihai Luo, Zhixiang Zuo, Jian Ren, et Yubin Xie. 2022. « SPENCER: a comprehensive database for small peptides encoded by noncoding RNAs in cancer patients ». *Nucleic Acids Research* 50 (D1): D1373-81. <https://doi.org/10.1093/nar/gkab822>.
- Ma, Xuelian, Shumin Cheng, Ruofan Ding, Zhaozhao Zhao, XuDong Zou, Shouhong Guang, Qixuan Wang, Huan Jing, Chen Yu, Ting Ni, et Lei Li. 2023. « ipaQTL-atlas: an atlas of intronic polyadenylation quantitative trait loci across human tissues ». *Nucleic Acids Research* 51 (D1): D1046-52. <https://doi.org/10.1093/nar/gkac736>.
- Ma, Yingli, Yang Yue, Yanbin Ma, Qing Zhang, Qingtong Zhou, Yunpeng Song, Yuqing Shen, Xun Li, Xiaochuan Ma, Chao Li, Michael A. Hanson, Gye Won Han, E. Allen Sickmier, Gayathri Swaminath, Suwen Zhao, Raymond C. Stevens, Liaoyuan A. Hu, Wenge Zhong, Mingqiang Zhang, et Fei Xu. 2017. « Structural Basis for Apelin Control of the Human Apelin Receptor ». *Structure* 25 (6): 858-866.e4. <https://doi.org/10.1016/j.str.2017.04.008>.
- MacDonald, Clinton C., et K. Wyatt McMahon. 2010. « Tissue-Specific Mechanisms of Alternative Polyadenylation: Testis, Brain, and beyond: Tissue-Specific Mechanisms of Alternative Polyadenylation ». *Wiley Interdisciplinary Reviews: RNA* 1 (3): 494-501. <https://doi.org/10.1002/wrna.29>.

- Maffeo, Christopher, et Aleksei Aksimentiev. 2009. « Structure, Dynamics, and Ion Conductance of the Phospholamban Pentamer ». *Biophysical Journal* 96 (12): 4853-65. <https://doi.org/10.1016/j.bpj.2009.03.053>.
- Magnuson, Brian, Karan Bedi, Ishwarya Venkata Narayanan, Bartłomiej Bartkowiak, Hailey Blinkiewicz, Michelle T. Paulsen, Arno Greenleaf, et Mats Ljungman. 2022. « CDK12 Regulates Co-Transcriptional Splicing and RNA Turnover in Human Cells ». *iScience* 25 (9): 105030. <https://doi.org/10.1016/j.isci.2022.105030>.
- Makarewich, Catherine A. 2020. « The hidden world of membrane microproteins ». *Experimental Cell Research* 388 (2): 111853. <https://doi.org/10.1016/j.yexcr.2020.111853>.
- Makarewich, Catherine A., et Eric N. Olson. 2017. « Mining for Micropeptides ». *Trends in Cell Biology* 27 (9): 685-96. <https://doi.org/10.1016/j.tcb.2017.04.006>.
- Malinge, Jean-Marc, Marie-Josèphe Giraud-Panis, et Marc Leng. 1999. « Interstrand cross-links of cisplatin induce striking distortions in DNA ». *Journal of Inorganic Biochemistry* 77 (1): 23-29. [https://doi.org/10.1016/S0162-0134\(99\)00148-8](https://doi.org/10.1016/S0162-0134(99)00148-8).
- Malkova, Anna, et Grzegorz Ira. 2013. « Break-Induced Replication: Functions and Molecular Mechanism ». *Current Opinion in Genetics & Development* 23 (3): 271-79. <https://doi.org/10.1016/j.gde.2013.05.007>.
- Maloisel, Laurent, Francis Fabre, et Serge Gangloff. 2008. « DNA Polymerase  $\delta$  Is Preferentially Recruited during Homologous Recombination To Promote Heteroduplex DNA Extension ». *Molecular and Cellular Biology* 28 (4): 1373-82. <https://doi.org/10.1128/MCB.01651-07>.
- Marasco, Luciano E., et Alberto R. Kornblihtt. 2023. « The Physiology of Alternative Splicing ». *Nature Reviews Molecular Cell Biology* 24 (4): 242-54. <https://doi.org/10.1038/s41580-022-00545-z>.
- Marsischky, Gerald T., et Richard D. Kolodner. 1999. « Biochemical Characterization of the Interaction between the *Saccharomyces Cerevisiae* MSH2-MSH6 Complex and Mispaiored Bases in DNA \* ». *Journal of Biological Chemistry* 274 (38): 26668-82. <https://doi.org/10.1074/jbc.274.38.26668>.
- Matsumoto, Akinobu, Alessandra Pasut, Masaki Matsumoto, Riu Yamashita, Jacqueline Fung, Emanuele Monteleone, Alan Saghatelian, Keiichi I. Nakayama, John G. Clohessy, et Pier Paolo Pandolfi. 2017. « mTORC1 and Muscle Regeneration Are Regulated by the LINC00961-Encoded SPAR Polypeptide ». *Nature* 541 (7636): 228-32. <https://doi.org/10.1038/nature21034>.
- Mayr, Christine. 2019. « What Are 3' UTRs Doing? » *Cold Spring Harbor Perspectives in Biology* 11 (10): a034728. <https://doi.org/10.1101/cshperspect.a034728>.
- Mayr, Christine, et David P. Bartel. 2009. « Widespread Shortening of 3'UTRs by Alternative Cleavage and Polyadenylation Activates Oncogenes in Cancer Cells ». *Cell* 138 (4): 673-84. <https://doi.org/10.1016/j.cell.2009.06.016>.

- Mazurek, Anthony, Christopher N. Johnson, Markus W. Germann, et Richard Fishel. 2009. « Sequence context effect for hMSH2-hMSH6 mismatch-dependent activation ». *Proceedings of the National Academy of Sciences of the United States of America* 106 (11): 4177-82. <https://doi.org/10.1073/pnas.0808572106>.
- McCulloch, Scott D., Robert J. Kokoska, Parie Garg, Peter M. Burgers, et Thomas A. Kunkel. 2009. « The efficiency and fidelity of 8-oxo-guanine bypass by DNA polymerases  $\delta$  and  $\eta$  ». *Nucleic Acids Research* 37 (9): 2830-40. <https://doi.org/10.1093/nar/gkp103>.
- McKeage, M. J. 1995. « Comparative Adverse Effect Profiles of Platinum Drugs ». *Drug Safety* 13 (4): 228-44. <https://doi.org/10.2165/00002018-199513040-00003>.
- McMahill, Melissa S., Caroline W. Sham, et Douglas K. Bishop. 2007. « Synthesis-Dependent Strand Annealing in Meiosis ». *PLOS Biology* 5 (11): e299. <https://doi.org/10.1371/journal.pbio.0050299>.
- Meng, Kun, Shaohua Lu, Yu-Ying Li, Li-Ling Hu, Jing Zhang, Yun Cao, Yang Wang, Chris Zhiyi Zhang, et Qing-Yu He. 2023. « LINC00493-encoded microprotein SMIM26 exerts anti-metastatic activity in renal cell carcinoma ». *EMBO reports* 24 (6): e56282. <https://doi.org/10.15252/embr.202256282>.
- Miettinen, Teemu P., et Mikael Björklund. 2015. « Modified Ribosome Profiling Reveals High Abundance of Ribosome Protected mRNA Fragments Derived from 3' Untranslated Regions ». *Nucleic Acids Research* 43 (2): 1019-34. <https://doi.org/10.1093/nar/gku1310>.
- Millevoi, Stefania, Clarisse Loulergue, Sabine Dettwiler, Sarah Zeïneb Karaa, Walter Keller, Michael Antoniou, et Stéphan Vagner. 2006. « An interaction between U2AF 65 and CF Im links the splicing and 3' end processing machineries ». *The EMBO Journal* 25 (20): 4854-64. <https://doi.org/10.1038/sj.emboj.7601331>.
- Mimitou, Eleni P., et Lorraine S. Symington. 2008. « Sae2, Exo1 and Sgs1 Collaborate in DNA Double-Strand Break Processing ». *Nature* 455 (7214): 770-74. <https://doi.org/10.1038/nature07312>.
- . 2010. « Ku Prevents Exo1 and Sgs1-Dependent Resection of DNA Ends in the Absence of a Functional MRX Complex or Sae2 ». *The EMBO Journal* 29 (19): 3358-69. <https://doi.org/10.1038/emboj.2010.193>.
- Mimoso, Claudia A., et Karen Adelman. 2023. « U1 snRNP Increases RNA Pol II Elongation Rate to Enable Synthesis of Long Genes ». *Molecular Cell* 83 (8): 1264-1279.e10. <https://doi.org/10.1016/j.molcel.2023.03.002>.
- Min, Kyung-Won, Sylvia Davila, Richard W. Zealy, Lawson T. Lloyd, In Young Lee, Rumi Lee, Kyung Hye Roh, Ahjin Jung, Jacek Jemielity, Eui-Ju Choi, Jeong Ho Chang, et Je-Hyun Yoon. 2017. « eIF4E phosphorylation by MST1 reduces translation of a subset of mRNAs, but increases lncRNA translation ». *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* 1860 (7): 761-72. <https://doi.org/10.1016/j.bbagr.2017.05.002>.
- Miura, Tohru, Yoshimasa Yamana, Takehiko Usui, Hiroaki I. Ogawa, Masa-Toshi Yamamoto, et Kohji Kusano. 2012. « Homologous Recombination via Synthesis-Dependent Strand

- Annealing in Yeast Requires the Irc20 and Srs2 DNA Helicases ». *Genetics* 191 (1): 65-78. <https://doi.org/10.1534/genetics.112.139105>.
- Moehle, Erica A, Hannes Braberg, Nevan J Krogan, et Christine Guthrie. 2014. « Adventures in time and space ». *RNA Biology* 11 (4): 313-19. <https://doi.org/10.4161/rna.28646>.
- Moggs, Jonathan G., David E. Szymkowski, Masami Yamada, Peter Karran, et Richard D. Wood. 1997. « Differential Human Nucleotide Excision Repair of Paired and Mismatched Cisplatin-DNA Adducts ». *Nucleic Acids Research* 25 (3): 480-90. <https://doi.org/10.1093/nar/25.3.480>.
- Mohsen, Jessica J., Alina A. Martel, et Sarah A. Slavoff. 2023. « Microproteins—Discovery, Structure, and Function ». *PROTEOMICS* 23 (23-24): 2100211. <https://doi.org/10.1002/pmic.202100211>.
- Montigny, Audrey, Patrizia Tavormina, Carine Duboe, H el ene San Cl emente, Marielle Aguilar, Philippe Valenti, Dominique Laouressergues, Jean-Philippe Combier, et Serge Plaza. 2021. « Drosophila primary microRNA-8 encodes a microRNA-encoded peptide acting in parallel of miR-8 ». *Genome Biology* 22 (1): 118. <https://doi.org/10.1186/s13059-021-02345-8>.
- Morra, D. S., S. H. Lawler, B. P. Eliceiri, et G. L. Eliceiri. 1986. « Inhibition of Small Nuclear RNA Synthesis by Ultraviolet Radiation ». *The Journal of Biological Chemistry* 261 (7): 3142-46.
- Motzer, R. J. 1993. « Optimal Treatment for Advanced Seminoma? » *Cancer* 72 (1): 3-4. [https://doi.org/10.1002/1097-0142\(19930701\)72:1<3::aid-cnrc2820720103>3.0.co;2-6](https://doi.org/10.1002/1097-0142(19930701)72:1<3::aid-cnrc2820720103>3.0.co;2-6).
- Mouilleron, H el ene, Vivian Delcourt, et Xavier Roucou. 2016. « Death of a dogma: eukaryotic mRNAs can code for more than one protein ». *Nucleic Acids Research* 44 (1): 14-23. <https://doi.org/10.1093/nar/gkv1218>.
- Mudge, Jonathan M., Jorge Ruiz-Orera, John R. Prensner, Marie A. Brunet, Ferriol Calvet, Irwin Jungreis, Jose Manuel Gonzalez, Michele Magrane, Thomas F. Martinez, Jana Felicitas Schulz, Yucheng T. Yang, M. Mar Alb a, Julie L. Aspden, Pavel V. Baranov, Ariel A. Bazzini, Elspeth Bruford, Maria Jesus Martin, Lorenzo Calviello, Anne-Ruxandra Carvunis, Jin Chen, Juan Pablo Couso, Eric W. Deutsch, Paul Flicek, Adam Frankish, Mark Gerstein, Norbert Hubner, Nicholas T. Ingolia, Manolis Kellis, Gerben Menschaert, Robert L. Moritz, Uwe Ohler, Xavier Roucou, Alan Saghatelian, Jonathan S. Weissman, et Sebastiaan van Heesch. 2022. « Standardized Annotation of Translated Open Reading Frames ». *Nature Biotechnology* 40 (7): 994-99. <https://doi.org/10.1038/s41587-022-01369-0>.
- Muggia, Franco M., et Gerrit Los. 1993. « Platinum Resistance: Laboratory Findings and Clinical Implications ». *STEM CELLS* 11 (3): 182-93. <https://doi.org/10.1002/stem.5530110304>.
- Mu oz, Manuel J., M. Soledad P erez Santangelo, Maria P. Paronetto, Manuel de la Mata, Federico Pelisch, St ephane Boireau, Kira Glover-Cutter, Claudia Ben-Dov, Mat as Blaustein, Juan J. Lozano, Gregory Bird, David Bentley, Edouard Bertrand, et Alberto R. Kornblihtt. 2009. « DNA Damage Regulates Alternative Splicing through Inhibition of

- RNA Polymerase II Elongation ». *Cell* 137 (4): 708-20. <https://doi.org/10.1016/j.cell.2009.03.010>.
- Murphy, Michael R., Anthony Ramadei, Ahmet Doymaz, Sophia Varriano, Devorah Natelson, Amy Yu, Sera Aktas, Marie Mazzeo, Michael Mazzeo, George Zakusilo, et Frida E. Kleiman. 2023. « Long Non-Coding RNA Generated from CDKN1A Gene by Alternative Polyadenylation Regulates P21 Expression during DNA Damage Response ». bioRxiv. <https://doi.org/10.1101/2023.01.10.523318>.
- Murray, David, et Razmik Mirzayans. 2020. « Cellular Responses to Platinum-Based Anticancer Drugs and UVC: Role of P53 and Implications for Cancer Therapy ». *International Journal of Molecular Sciences* 21 (16): 5766. <https://doi.org/10.3390/ijms21165766>.
- Naftelberg, Shiran, Ignacio E. Schor, Gil Ast, et Alberto R. Kornblihtt. 2015. « Regulation of Alternative Splicing Through Coupling with Transcription and Chromatin Structure ». *Annual Review of Biochemistry* 84 (1): 165-98. <https://doi.org/10.1146/annurev-biochem-060614-034242>.
- Nanavaty, Vishal, Elizabeth W. Abrash, Changjin Hong, Sunho Park, Emily E. Fink, Zhuangyue Li, Thomas J. Sweet, Jeffrey M. Bhasin, Srinidhi Singuri, Byron H. Lee, Tae Hyun Hwang, et Angela H. Ting. 2020. « DNA Methylation Regulates Alternative Polyadenylation via CTCF and the Cohesin Complex ». *Molecular Cell* 78 (4): 752-764.e6. <https://doi.org/10.1016/j.molcel.2020.03.024>.
- Nelson, J. R., C. W. Lawrence, et D. C. Hinkle. 1996. « Deoxycytidyl Transferase Activity of Yeast REV1 Protein ». *Nature* 382 (6593): 729-31. <https://doi.org/10.1038/382729a0>.
- Neville, Matthew D.C., Robin Kohze, Chaitanya Erady, Narendra Meena, Matthew Hayden, David N. Cooper, Matthew Mort, et Sudhakaran Prabakaran. 2021. « A Platform for Curated Products from Novel Open Reading Frames Prompts Reinterpretation of Disease Variants ». *Genome Research* 31 (2): 327-36. <https://doi.org/10.1101/gr.263202.120>.
- Nishi, Ryotaro, Wataru Sakai, Daisuke Tone, Fumio Hanaoka, et Kaoru Sugasawa. 2013. « Structure-Function Analysis of the EF-Hand Protein Centrin-2 for Its Intracellular Localization and Nucleotide Excision Repair ». *Nucleic Acids Research* 41 (14): 6917-29. <https://doi.org/10.1093/nar/gkt434>.
- Noe Gonzalez, Melvin, Daniel Blears, et Jesper Q. Svejstrup. 2021. « Causes and Consequences of RNA Polymerase II Stalling during Transcript Elongation ». *Nature Reviews Molecular Cell Biology* 22 (1): 3-21. <https://doi.org/10.1038/s41580-020-00308-8>.
- Nojima, Takayuki, Martin Dienstbier, Shona Murphy, Nicholas J. Proudfoot, et Michael J. Dye. 2013. « Definition of RNA Polymerase II CoTC Terminator Elements in the Human Genome ». *Cell Reports* 3 (4): 1080-92. <https://doi.org/10.1016/j.celrep.2013.03.012>.
- Noll, David M., Tracey McGregor Mason, et Paul S. Miller. 2006. « Formation and Repair of Interstrand Cross-Links in DNA ». *Chemical Reviews* 106 (2): 277-301. <https://doi.org/10.1021/cr040478b>.

- Nourse, Jamie, Stefano Spada, et Sven Danckwardt. 2020. « Emerging Roles of RNA 3'-End Cleavage and Polyadenylation in Pathogenesis, Diagnosis and Therapy of Human Disorders ». *Biomolecules* 10 (6): 915. <https://doi.org/10.3390/biom10060915>.
- Ogami, Koichi, Patricia Richard, Yaqiong Chen, Mainul Hoque, Wencheng Li, James J. Moresco, John R. Yates, Bin Tian, et James L. Manley. 2017. « An Mtr4/ZFC3H1 Complex Facilitates Turnover of Unstable Nuclear RNAs to Prevent Their Cytoplasmic Transport and Global Translational Repression ». *Genes & Development* 31 (12): 1257-71. <https://doi.org/10.1101/gad.302604.117>.
- Ogi, Tomoo, et Alan R. Lehmann. 2006. « The Y-Family DNA Polymerase Kappa (Pol Kappa) Functions in Mammalian Nucleotide-Excision Repair ». *Nature Cell Biology* 8 (6): 640-42. <https://doi.org/10.1038/ncb1417>.
- Oh, Jung-Min, Chao Di, Christopher C. Venters, Jiannan Guo, Chie Arai, Byung Ran So, Anna Maria Pinto, Zhenxi Zhang, Lili Wan, Ihab Younis, et Gideon Dreyfuss. 2017. « U1 snRNP Telescripting Regulates a Size-Function-Stratified Human Genome ». *Nature Structural & Molecular Biology* 24 (11): 993-99. <https://doi.org/10.1038/nsmb.3473>.
- Ohashi, E., K. Bebenek, T. Matsuda, W. J. Feaver, V. L. Gerlach, E. C. Friedberg, H. Ohmori, et T. A. Kunkel. 2000. « Fidelity and Processivity of DNA Synthesis by DNA Polymerase Kappa, the Product of the Human DINB1 Gene ». *The Journal of Biological Chemistry* 275 (50): 39678-84. <https://doi.org/10.1074/jbc.M005309200>.
- Oksenyshyn, Valentyn, Bruno Bernardes de Jesus, Alexander Zhovmer, Jean-Marc Egly, et Frédéric Coin. 2009. « Molecular insights into the recruitment of TFIIH to sites of DNA damage ». *The EMBO Journal* 28 (19): 2971-80. <https://doi.org/10.1038/emboj.2009.230>.
- Oliver, José L., et Antonio Marín. 1996. « A Relationship between GC Content and Coding-Sequence Length ». *Journal of Molecular Evolution* 43 (3): 216-23. <https://doi.org/10.1007/BF02338829>.
- Ormancey, Mélanie, Patrice Thuleau, Jean-Philippe Combier, et Serge Plaza. 2023. « The Essentials on microRNA-Encoded Peptides from Plants to Animals ». *Biomolecules* 13 (2): 206. <https://doi.org/10.3390/biom13020206>.
- Orr, Mona Wu, Yuanhui Mao, Gisela Storz, et Shu-Bing Qian. 2020. « Alternative ORFs and small ORFs: shedding light on the dark proteome ». *Nucleic Acids Research* 48 (3): 1029-42. <https://doi.org/10.1093/nar/gkz734>.
- Oxenoid, Kirill, et James J. Chou. 2005. « The structure of phospholamban pentamer reveals a channel-like architecture in membranes ». *Proceedings of the National Academy of Sciences* 102 (31): 10870-75. <https://doi.org/10.1073/pnas.0504920102>.
- Pabla, Navjotsingh, Shuang Huang, Qing-Sheng Mi, Rene Daniel, et Zheng Dong. 2008. « ATR-Chk2 Signaling in p53 Activation and DNA Damage Response during Cisplatin-induced Apoptosis\* ». *Journal of Biological Chemistry* 283 (10): 6572-83. <https://doi.org/10.1074/jbc.M707568200>.



- Pack, Lindsey R, Leighton H Daigh, et Tobias Meyer. 2019. « Putting the Brakes on the Cell Cycle: Mechanisms of Cellular Growth Arrest ». *Current Opinion in Cell Biology* 60 (octobre): 106-13. <https://doi.org/10.1016/j.ceb.2019.05.005>.
- Paculová, Hana, et Jiří Kohoutek. 2017. « The Emerging Roles of CDK12 in Tumorigenesis ». *Cell Division* 12 (1): 7. <https://doi.org/10.1186/s13008-017-0033-x>.
- Pan, Qun, Ofer Shai, Leo J. Lee, Brendan J. Frey, et Benjamin J. Blencowe. 2008. « Deep Surveying of Alternative Splicing Complexity in the Human Transcriptome by High-Throughput Sequencing ». *Nature Genetics* 40 (12): 1413-15. <https://doi.org/10.1038/ng.259>.
- Pâques, F., et J. E. Haber. 1999. « Multiple Pathways of Recombination Induced by Double-Strand Breaks in *Saccharomyces Cerevisiae* ». *Microbiology and Molecular Biology Reviews: MMBR* 63 (2): 349-404. <https://doi.org/10.1128/MMBR.63.2.349-404.1999>.
- Pavlov, Y. I., D. Nguyen, et T. A. Kunkel. 2001. « Mutator Effects of Overproducing DNA Polymerase Eta (Rad30) and Its Catalytically Inactive Variant in Yeast ». *Mutation Research* 478 (1-2): 129-39. [https://doi.org/10.1016/s0027-5107\(01\)00131-2](https://doi.org/10.1016/s0027-5107(01)00131-2).
- Peck, Sarah A., Katlyn D. Hughes, Jose F. Victorino, et Amber L. Mosley. 2019. « Writing a Wrong: Coupled RNA Polymerase II Transcription and RNA Quality Control ». *WIREs RNA* 10 (4): e1529. <https://doi.org/10.1002/wrna.1529>.
- Pestova, Tatyana V., et Christopher U. T. Hellen. 1999. « Ribosome Recruitment and Scanning: What's New? ». *Trends in Biochemical Sciences* 24 (3): 85-87. [https://doi.org/10.1016/S0968-0004\(99\)01356-0](https://doi.org/10.1016/S0968-0004(99)01356-0).
- Popova, Tatiana, Elodie Manié, Valentina Boeva, Aude Battistella, Oumou Goundiam, Nicholas K. Smith, Christopher R. Mueller, Virginie Raynal, Odette Mariani, Xavier Sastre-Garau, et Marc-Henri Stern. 2016. « Ovarian Cancers Harboring Inactivating Mutations in CDK12 Display a Distinct Genomic Instability Pattern Characterized by Large Tandem Duplications ». *Cancer Research* 76 (7): 1882-91. <https://doi.org/10.1158/0008-5472.CAN-15-2128>.
- Powers, Kyle T, Jenn-Yeu Alvin Szeto, et Christiane Schaffitzel. 2020. « New insights into no-go, non-stop and nonsense-mediated mRNA decay complexes ». *Current Opinion in Structural Biology, Catalysis and Regulation • Protein Nucleic Acid Interaction*, 65 (décembre): 110-18. <https://doi.org/10.1016/j.sbi.2020.06.011>.
- Prakash, Rohit, Dominik Satory, Eloïse Dray, Almas Papusha, Jürgen Scheller, Wilfried Kramer, Lumir Krejci, Hannah Klein, James E. Haber, Patrick Sung, et Grzegorz Ira. 2009. « Yeast Mph1 helicase dissociates Rad51-made D-loops: implications for crossover control in mitotic recombination ». *Genes & Development* 23 (1): 67-79. <https://doi.org/10.1101/gad.1737809>.
- Prensner, John R., Oana M. Enache, Victor Luria, Karsten Krug, Karl R. Clauser, Joshua M. Dempster, Amir Karger, Li Wang, Karolina Stumbraite, Vickie M. Wang, Ginevra Botta, Nicholas J. Lyons, Amy Goodale, Zohra Kalani, Briana Fritchman, Adam Brown, Douglas Alan, Thomas Green, Xiaoping Yang, Jacob D. Jaffe, Jennifer A. Roth, Federica Piccioni, Marc W. Kirschner, Zhe Ji, David E. Root, et Todd R. Golub. 2021. « Noncanonical Open

- Reading Frames Encode Functional Proteins Essential for Cancer Cell Survival ». *Nature Biotechnology* 39 (6): 697-704. <https://doi.org/10.1038/s41587-020-00806-2>.
- Previati, Maurizio, Irene Lanzoni, Elisa Corbacella, Sara Magosso, Valeria Guaran, Alessandro Martini, et Silvano Capitani. 2006. « Cisplatin-Induced Apoptosis in Human Promyelocytic Leukemia Cells ». *International Journal of Molecular Medicine* 18 (3): 511-16.
- Proudfoot, Nick J. 2016. « Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut ». *Science* 352 (6291): aad9926. <https://doi.org/10.1126/science.aad9926>.
- Putnam, Christopher D., Tikvah K. Hayes, et Richard D. Kolodner. 2009. « Specific Pathways Prevent Duplication-Mediated Genome Rearrangements ». *Nature* 460 (7258): 984-89. <https://doi.org/10.1038/nature08217>.
- Qiao, Xue, Li Zhu, Rongbo Song, Chao Shang, et Yan Guo. 2023. « CD44 occurring alternative splicing promotes cisplatin resistance and evokes tumor immune response in oral squamous cell carcinoma cells ». *Translational Oncology* 31 (mai): 101644. <https://doi.org/10.1016/j.tranon.2023.101644>.
- Quereda, Victor, Simon Bayle, Francesca Vena, Sylvia M. Frydman, Andrii Monastyrskyi, William R. Roush, et Derek R. Duckett. 2019. « Therapeutic Targeting of CDK12/CDK13 in Triple-Negative Breast Cancer ». *Cancer Cell* 36 (5): 545-558.e7. <https://doi.org/10.1016/j.ccell.2019.09.004>.
- Rabik, Cara A., et M. Eileen Dolan. 2007. « Molecular Mechanisms of Resistance and Toxicity Associated with Platinating Agents ». *Cancer Treatment Reviews* 33 (1): 9-23. <https://doi.org/10.1016/j.ctrv.2006.09.006>.
- Ramanathan, Anand, G. Brett Robb, et Siu-Hong Chan. 2016. « mRNA capping: biological functions and applications ». *Nucleic Acids Research* 44 (16): 7511-26. <https://doi.org/10.1093/nar/gkw551>.
- Räschle, Markus, Puck Knipscheer, Milica Enoiu, Todor Angelov, Jingchuan Sun, Jack D. Griffith, Tom E. Ellenberger, Orlando D. Schärer, et Johannes C. Walter. 2008. « Mechanism of Replication-Coupled DNA Interstrand Crosslink Repair ». *Cell* 134 (6): 969-80. <https://doi.org/10.1016/j.cell.2008.08.030>.
- Ray, Partho Sarothi, Richa Grover, et Saumitra Das. 2006. « Two internal ribosome entry sites mediate the translation of p53 isoforms ». *EMBO reports* 7 (4): 404-10. <https://doi.org/10.1038/sj.embor.7400623>.
- Ren, Xi, Bin Zhang, Jia Li, Thamizhanban Manoharan, Beijia Liu, Yangyang Song, Shuye Tian, Kar-Tong Tan, Lingwen Ding, Ying Li, Omer An, Ming Li, Chan-Shuo Wu, Yang Liu, Boon Heng Dennis Teo, Sze Jing Tang, Jinhua Lu, Yuhui Hu, Wei Chen, Leilei Chen, Gloryn Chia, et Henry Yang. 2022. « Pervasive Intronic Polyadenylation Serves as a Potential Source of Cancer Neoantigens ». Preprint. In Review. <https://doi.org/10.21203/rs.3.rs-1537870/v1>.

- Renkawitz, Jörg, Claudio A. Lademann, et Stefan Jentsch. 2014. « Mechanisms and Principles of Homology Search during Recombination ». *Nature Reviews. Molecular Cell Biology* 15 (6): 369-83. <https://doi.org/10.1038/nrm3805>.
- Riedl, Thilo, Fumio Hanaoka, et Jean-Marc Egly. 2003. « The Comings and Goings of Nucleotide Excision Repair Factors on Damaged DNA ». *The EMBO Journal* 22 (19): 5293-5303. <https://doi.org/10.1093/emboj/cdg489>.
- Robertson, A. B., A. Klungland, T. Rognes, et I. Leiros. 2009. « DNA Repair in Mammalian Cells: Base Excision Repair: The Long and Short of It ». *Cellular and Molecular Life Sciences* 66 (6): 981-93. <https://doi.org/10.1007/s00018-009-8736-z>.
- Rogers, J., P. Early, C. Carter, K. Calame, M. Bond, L. Hood, et R. Wall. 1980. « Two mRNAs with Different 3' Ends Encode Membrane-Bound and Secreted Forms of Immunoglobulin  $\mu$  Chain ». *Cell* 20 (2): 303-12. [https://doi.org/10.1016/0092-8674\(80\)90616-9](https://doi.org/10.1016/0092-8674(80)90616-9).
- Said, Maha, Viviana Barra, Elisa Balzano, Ibtissam Talhaoui, Franca Pelliccia, Simona Giunta, et Valeria Naim. 2022. « FANCD2 Promotes Mitotic Rescue from Transcription-Mediated Replication Stress in SETX-Deficient Cancer Cells ». *Communications Biology* 5 (1): 1-13. <https://doi.org/10.1038/s42003-022-04360-2>.
- Sakharkar, Meena Kishore, Vincent T. K. Chow, et Pandjassarame Kanguene. 2004. « Distributions of Exons and Introns in the Human Genome ». *In Silico Biology* 4 (4): 387-93.
- Sanchez, Gabriel, Danielle Bittencourt, Karine Laud, Jérôme Barbier, Olivier Delattre, Didier Auboeuf, et Martin Dutertre. 2008. « Alteration of cyclin D1 transcript elongation by a mutated transcription factor up-regulates the oncogenic D1b splice isoform in cancer ». *Proceedings of the National Academy of Sciences* 105 (16): 6004-9. <https://doi.org/10.1073/pnas.0710748105>.
- Sanchez, Yolanda, Calvin Wong, Richard S. Thoma, Ron Richman, Zhiqi Wu, Helen Piwnicka-Worms, et Stephen J. Elledge. 1997. « Conservation of the Chk1 Checkpoint Pathway in Mammals: Linkage of DNA Damage to Cdk Regulation Through Cdc25 ». *Science* 277 (5331): 1497-1501. <https://doi.org/10.1126/science.277.5331.1497>.
- Sandmann, Clara-L., Jana F. Schulz, Jorge Ruiz-Orera, Marieluise Kirchner, Matthias Ziehm, Eleonora Adami, Maike Marczenke, Annabel Christ, Nina Liebe, Johannes Greiner, Aaron Schoenenberger, Michael B. Muecke, Ning Liang, Robert L. Moritz, Zhi Sun, Eric W. Deutsch, Michael Gotthardt, Jonathan M. Mudge, John R. Prensner, Thomas E. Willnow, Philipp Mertins, Sebastiaan Van Heesch, et Norbert Hubner. 2023. « Evolutionary Origins and Interactomes of Human, Young Microproteins and Small Peptides Translated from Short Open Reading Frames ». *Molecular Cell* 83 (6): 994-1011.e18. <https://doi.org/10.1016/j.molcel.2023.01.023>.
- Sattler, Ulrike, Philippe Frit, Bernard Salles, et Patrick Calsou. 2003. « Long-patch DNA repair synthesis during base excision repair in mammalian cells ». *EMBO Reports* 4 (4): 363-67. <https://doi.org/10.1038/sj.embor.embor796>.

- Scharer, O. D. 2013. « Nucleotide Excision Repair in Eukaryotes ». *Cold Spring Harbor Perspectives in Biology* 5 (10): a012609-a012609. <https://doi.org/10.1101/cshperspect.a012609>.
- Schlesinger, Dörte, et Simon J. Elsässer. 2022. « Revisiting sORFs: Overcoming Challenges to Identify and Characterize Functional Microproteins ». *The FEBS Journal* 289 (1): 53-74. <https://doi.org/10.1111/febs.15769>.
- Sesink, Anouk. 2023. « The molecular and cellular consequences of AsiDNA™ combined with radiotherapy on healthy tissue ». These de doctorat, université Paris-Saclay. <https://www.theses.fr/2023UPASL101>.
- Shatkin, Aaron J., et James L. Manley. 2000. « The Ends of the Affair: Capping and Polyadenylation ». *Nature Structural Biology* 7 (10): 838-42. <https://doi.org/10.1038/79583>.
- Sheth, Sandeep, Debashree Mukherjea, Leonard P. Rybak, et Vickram Ramkumar. 2017. « Mechanisms of Cisplatin-Induced Ototoxicity and Otoprotection ». *Frontiers in Cellular Neuroscience* 11: 338. <https://doi.org/10.3389/fncel.2017.00338>.
- Shigechi, Tomoko, Junya Tomida, Koichi Sato, Masahiko Kobayashi, John K. Eykelenboom, Fabio Pessina, Yanbin Zhang, Emi Uchida, Masamichi Ishiai, Noel F. Lowndes, Kenichi Yamamoto, Hitoshi Kurumizaka, Yoshihiko Maehara, et Minoru Takata. 2012. « ATR-ATRIP Kinase Complex Triggers Activation of the Fanconi Anemia DNA Repair Pathway ». *Cancer Research* 72 (5): 1149-56. <https://doi.org/10.1158/0008-5472.CAN-11-2904>.
- Shim, Eun Yong, Woo-Hyun Chung, Matthew L Nicolette, Yu Zhang, Melody Davis, Zhu Zhu, Tanya T Paull, Grzegorz Ira, et Sang Eun Lee. 2010. « Saccharomyces cerevisiae Mre11/Rad50/Xrs2 and Ku proteins regulate association of Exo1 and Dna2 with DNA breaks ». *The EMBO Journal* 29 (19): 3370-80. <https://doi.org/10.1038/emboj.2010.219>.
- Shimada, Midori, Hiroyuki Niida, Doaa H. Zineldeen, Hideaki Tagami, Masafumi Tanaka, Hiroyuki Saito, et Makoto Nakanishi. 2008. « Chk1 Is a Histone H3 Threonine 11 Kinase That Regulates DNA Damage-Induced Transcriptional Repression ». *Cell* 132 (2): 221-32. <https://doi.org/10.1016/j.cell.2007.12.013>.
- Shin, Jihae, Qingbao Ding, Luyang Wang, Yange Cui, Erdene Baljinnyam, Aysegul Guvenek, et Bin Tian. 2021. « CRISPRpas: Programmable Regulation of Alternative Polyadenylation by dCas9 ». *Nucleic Acids Research*, juillet. <https://doi.org/10.1093/nar/gkab519>.
- Shivji, Mahmud K. K., Vladimir N. Podust, Ulrich Huebscher, et Richard D. Wood. 1995. « Nucleotide Excision Repair DNA Synthesis by DNA Polymerase .epsilon. in the Presence of PCNA, RFC, and RPA ». *Biochemistry* 34 (15): 5011-17. <https://doi.org/10.1021/bi00015a012>.
- Singh, Irtisha, Shih-Han Lee, Adam S. Sperling, Mehmet K. Samur, Yu-Tzu Tai, Mariateresa Fulciniti, Nikhil C. Munshi, Christine Mayr, et Christina S. Leslie. 2018. « Widespread Intronic Polyadenylation Diversifies Immune Cell Transcriptomes ». *Nature Communications* 9 (1): 1716. <https://doi.org/10.1038/s41467-018-04112-z>.

- Skinner, R. 1995. « Strategies to Prevent Nephrotoxicity of Anticancer Drugs ». *Current Opinion in Oncology* 7 (4): 310-15. <https://doi.org/10.1097/00001622-199507000-00003>.
- Skourti-Stathaki, Konstantina, Nicholas J. Proudfoot, et Natalia Gromak. 2011. « Human Senataxin Resolves RNA/DNA Hybrids Formed at Transcriptional Pause Sites to Promote Xrn2-Dependent Termination ». *Molecular Cell* 42 (6): 794-805. <https://doi.org/10.1016/j.molcel.2011.04.026>.
- Slavoff, Sarah A., Jinho Heo, Bogdan A. Budnik, Leslyn A. Hanakahi, et Alan Saghatelian. 2014. « A Human Short Open Reading Frame (sORF)-encoded Polypeptide That Stimulates DNA End Joining ». *The Journal of Biological Chemistry* 289 (16): 10950-57. <https://doi.org/10.1074/jbc.C113.533968>.
- Slavoff, Sarah A., Andrew J. Mitchell, Adam G. Schwaib, Moran N. Cabili, Jiao Ma, Joshua Z. Levin, Amir D. Karger, Bogdan A. Budnik, John L. Rinn, et Alan Saghatelian. 2013. « Peptidomic Discovery of Short Open Reading Frame-Encoded Peptides in Human Cells ». *Nature Chemical Biology* 9 (1): 59-64. <https://doi.org/10.1038/nchembio.1120>.
- Snowden, Timothy, Samir Acharya, Charles Butz, Mark Berardini, et Richard Fishel. 2004. « hMSH4-hMSH5 Recognizes Holliday Junctions and Forms a Meiosis-Specific Sliding Clamp That Embraces Homologous Chromosomes ». *Molecular Cell* 15 (3): 437-51. <https://doi.org/10.1016/j.molcel.2004.06.040>.
- So, Byung Ran, Chao Di, Zhiqiang Cai, Christopher C. Venters, Jiannan Guo, Jung-Min Oh, Chie Arai, et Gideon Dreyfuss. 2019. « A Complex of U1 snRNP with Cleavage and Polyadenylation Factors Controls Telescripting, Regulating mRNA Transcription in Human Cells ». *Molecular Cell* 76 (4): 590-599.e4. <https://doi.org/10.1016/j.molcel.2019.08.007>.
- Sommer, Jan, Christoph Garbers, Janina Wolf, Ahmad Trad, Jens M. Moll, Markus Sack, Rainer Fischer, Joachim Gröttinger, Georg H. Waetzig, Doreen M. Floss, et Jürgen Scheller. 2014. « Alternative Intronic Polyadenylation Generates the Interleukin-6 Trans-Signaling Inhibitor Sgp130-E10 \* ». *Journal of Biological Chemistry* 289 (32): 22140-50. <https://doi.org/10.1074/jbc.M114.560938>.
- Song, Mengdi, Mingxiao Cui, et Kehai Liu. 2022. « Therapeutic strategies to overcome cisplatin resistance in ovarian cancer ». *European Journal of Medicinal Chemistry* 232 (mars): 114205. <https://doi.org/10.1016/j.ejmech.2022.114205>.
- Sormanni, Pietro, Francesco A. Aprile, et Michele Vendruscolo. 2015. « Rational Design of Antibodies Targeting Specific Epitopes within Intrinsically Disordered Proteins ». *Proceedings of the National Academy of Sciences* 112 (32): 9902-7. <https://doi.org/10.1073/pnas.1422401112>.
- Sourisseau, Tony, Carole Helissey, Céline Lefebvre, Florence Ponsonailles, Hélène Malka-Mahieu, Ken A Olausson, Fabrice André, Stephan Vagner, et Jean-Charles Soria. 2016. « Translational regulation of the mRNA encoding the ubiquitin peptidase USP1 involved in the DNA damage response as a determinant of Cisplatin resistance ». *Cell Cycle* 15 (2): 295-302. <https://doi.org/10.1080/15384101.2015.1120918>.

- Staresincic, Lidija, Adebanye F Fagbemi, Jacqueline H Enzlin, Audrey M Gourdin, Nils Wijgers, Isabelle Dunand-Sauthier, Giuseppina Giglia-Mari, Stuart G Clarkson, Wim Vermeulen, et Orlando D Schärer. 2009. « Coordination of dual incision and repair synthesis in human nucleotide excision repair ». *The EMBO Journal* 28 (8): 1111-20. <https://doi.org/10.1038/emboj.2009.49>.
- Staudt, Annica-Carolin, et Stephan Wenkel. 2011. « Regulation of protein function by 'microProteins' ». *EMBO reports* 12 (1): 35-42. <https://doi.org/10.1038/embor.2010.196>.
- Sugasawa, Kaoru, Tomoko Okamoto, Yuichiro Shimizu, Chikahide Masutani, Shigenori Iwai, et Fumio Hanaoka. 2001. « A multistep damage recognition mechanism for global genomic nucleotide excision repair ». *Genes & Development* 15 (5): 507-21. <https://doi.org/10.1101/gad.866301>.
- Sun, Rui, Ting Wei, Donglin Ding, Jianong Zhang, Sujun Chen, Housheng Hansen He, Ligu Wang, et Haojie Huang. 2022. « CYCLIN K down-regulation induces androgen receptor gene intronic polyadenylation, variant expression and PARP inhibitor vulnerability in castration-resistant prostate cancer ». *Proceedings of the National Academy of Sciences* 119 (39): e2205509119. <https://doi.org/10.1073/pnas.2205509119>.
- Takata, Kei-ichi, Tatsuhiko Shimizu, Shigenori Iwai, et Richard D. Wood. 2006. « Human DNA Polymerase N (POLN) Is a Low Fidelity Enzyme Capable of Error-Free Bypass of 5S-Thymine Glycol ». *The Journal of Biological Chemistry* 281 (33): 23445-55. <https://doi.org/10.1074/jbc.M604317200>.
- Tanaka, Iris, Alina Chakraborty, Olivier Saulnier, Clara Benoit-Pilven, Sophie Vacher, Dalila Labiod, Eric W. F. Lam, Ivan Bièche, Olivier Delattre, Frédéric Pouzoulet, Didier Auboeuf, Stéphane Vagner, et Martin Dutertre. 2020. « ZRANB2 and SYF2-Mediated Splicing Programs Converging on ECT2 Are Involved in Breast Cancer Cell Resistance to Doxorubicin ». *Nucleic Acids Research* 48 (5): 2676-93. <https://doi.org/10.1093/nar/gkz1213>.
- Tang, Chengyuan, Man J. Livingston, Robert Safirstein, et Zheng Dong. 2023. « Cisplatin Nephrotoxicity: New Insights and Therapeutic Implications ». *Nature Reviews Nephrology* 19 (1): 53-72. <https://doi.org/10.1038/s41581-022-00631-7>.
- Tapias, Angels, Jerome Auriol, Diane Forget, Jacqueline H. Enzlin, Orlando D Schärer, Frederic Coin, Benoit Coulombe, et Jean-Marc Egly. 2004. « Ordered Conformational Changes in Damaged DNA Induced by Nucleotide Excision Repair Factors ». *The Journal of biological chemistry* 279 (18): 19074-83. <https://doi.org/10.1074/jbc.M312611200>.
- Tatemoto, Kazuhiko, Masaki Hosoya, Yugo Habata, Ryo Fujii, Tadao Kakegawa, Min-Xu Zou, Yuji Kawamata, Shoji Fukusumi, Shuji Hinuma, Chieko Kitada, Tsutomu Kurokawa, Haruo Onda, et Masahiko Fujino. 1998. « Isolation and Characterization of a Novel Endogenous Peptide Ligand for the Human APJ Receptor ». *Biochemical and Biophysical Research Communications* 251 (2): 471-76. <https://doi.org/10.1006/bbrc.1998.9489>.
- Tellier, Michael, Justyna Zaborowska, Livia Caizzi, Eusra Mohammad, Taras Velychko, Björn Schwalb, Ivan Ferrer-Vicens, Daniel Blears, Takayuki Nojima, Patrick Cramer, et Shona

- Murphy. 2020. « CDK12 globally stimulates RNA polymerase II transcription elongation and carboxyl-terminal domain phosphorylation ». *Nucleic Acids Research* 48 (14): 7712-27. <https://doi.org/10.1093/nar/gkaa514>.
- Terrone, Sophie, Jessica Valat, Nicolas Fontrodona, Guillaume Giraud, Jean-Baptiste Claude, Emmanuel Combe, Audrey Lapendry, H el ene Polv eche, Lamya Ben Ameer, Arnaud Duvermy, Laurent Modolo, Pascal Bernard, Franck Mortreux, Didier Auboeuf, et Cyril F Bourgeois. 2022. « RNA Helicase-Dependent Gene Looping Impacts Messenger RNA Processing ». *Nucleic Acids Research* 50 (16): 9226-46. <https://doi.org/10.1093/nar/gkac717>.
- Tian, Bin, et Joel H. Graber. 2012. « Signals for Pre-mRNA Cleavage and Polyadenylation: Polyadenylation Signals ». *Wiley Interdisciplinary Reviews: RNA* 3 (3): 385-96. <https://doi.org/10.1002/wrna.116>.
- Tian, Bin, et James L. Manley. 2017. « Alternative Polyadenylation of mRNA Precursors ». *Nature Reviews Molecular Cell Biology* 18 (1): 18-30. <https://doi.org/10.1038/nrm.2016.116>.
- Tian, Bin, Zhenhua Pan, et Ju Youn Lee. 2007. « Widespread mRNA Polyadenylation Events in Introns Indicate Dynamic Interplay between Polyadenylation and Splicing ». *Genome Research* 17 (2): 156-65. <https://doi.org/10.1101/gr.5532707>.
- Tien, Jerry F., Alborz Mazloomian, S.-W. Grace Cheng, Christopher S. Hughes, Christalle C.T. Chow, Leanna T. Canapi, Arusha Oloumi, Genny Trigo-Gonzalez, Ali Bashashati, James Xu, Vicky C.-D. Chang, Sohrab P. Shah, Samuel Aparicio, et Gregg B. Morin. 2017. « CDK12 regulates alternative last exon mRNA splicing and promotes breast cancer cell invasion ». *Nucleic Acids Research* 45 (11): 6698-6716. <https://doi.org/10.1093/nar/gkx187>.
- Todd, Ryan C., et Stephen J. Lippard. 2010. « Structure of duplex DNA containing the cisplatin 1,2- $\{Pt(NH_3)_2\}_2$ -d(GpG) cross-link at 1.77  resolution ». *Journal of Inorganic Biochemistry* 104 (9): 902-8. <https://doi.org/10.1016/j.jinorgbio.2010.04.005>.
- Trotta, Edoardo. 2016. « Selective Forces and Mutational Biases Drive Stop Codon Usage in the Human Genome: A Comparison with Sense Codon Usage ». *BMC Genomics* 17 (1): 366. <https://doi.org/10.1186/s12864-016-2692-4>.
- Tsang, Roger Y., Turki Al-Fayea, et Heather-Jane Au. 2009. « Cisplatin Overdose: Toxicities and Management ». *Drug Safety* 32 (12): 1109-22. <https://doi.org/10.2165/11316640-000000000-00000>.
- Tsimberidou, Apostolia-Maria, Fadi Braiteh, David J. Stewart, et Razelle Kurzrock. 2009. « Ultimate Fate of Oncology Drugs Approved by the Us Food and Drug Administration without a Randomized Trial ». *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology* 27 (36): 6243-50. <https://doi.org/10.1200/JCO.2009.23.6018>.
- Vagner, S, U R uegsegger, S I Gunderson, W Keller, et I W Mattaj. 2000. « Position-dependent inhibition of the cleavage step of pre-mRNA 3'-end processing by U1 snRNP. » *RNA* 6 (2): 178-88.

- Vaisman, Alexandra, Maria Varchenko, Asad Umar, Thomas A. Kunkel, John I. Risinger, J. Carl Barrett, Thomas C. Hamilton, et Stephen G. Chaney. 1998. « The Role of hMLH1, hMSH3, and hMSH6 Defects in Cisplatin and Oxaliplatin Resistance: Correlation with Replicative Bypass of Platinum-DNA Adducts<sup>1</sup> ». *Cancer Research* 58 (16): 3579-85.
- Vakirlis, Nikolaos, Zoe Vance, Kate M. Duggan, et Aoife McLysaght. 2022. « De Novo Birth of Functional Microproteins in the Human Lineage ». *Cell Reports* 41 (12): 111808. <https://doi.org/10.1016/j.celrep.2022.111808>.
- Vannini, Alessandro. 2013. « A structural perspective on RNA polymerase I and RNA polymerase III transcription machineries ». *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms, Transcription by Odd Pols, 1829* (3): 258-64. <https://doi.org/10.1016/j.bbagr.2012.09.009>.
- Vorlová, Sandra, Gina Rocco, Clare V. LeFave, Francine M. Jodelka, Ken Hess, Michelle L. Hastings, Erik Henke, et Luca Cartegni. 2011. « Induction of Antagonistic Soluble Decoy Receptor Tyrosine Kinases by Intronic PolyA Activation ». *Molecular Cell* 43 (6): 927-39. <https://doi.org/10.1016/j.molcel.2011.08.009>.
- Wakasugi, Mitsuo, Aki Kawashima, Hiroshi Morioka, Stuart Linn, Aziz Sancar, Toshio Mori, Osamu Nikaido, et Tsukasa Matsunaga. 2002. « DDB Accumulates at DNA Damage Sites Immediately after UV Irradiation and Directly Stimulates Nucleotide Excision Repair ». *The Journal of Biological Chemistry* 277 (3): 1637-40. <https://doi.org/10.1074/jbc.C100610200>.
- Walker, J. R., R. A. Corpina, et J. Goldberg. 2001. « Structure of the Ku Heterodimer Bound to DNA and Its Implications for Double-Strand Break Repair ». *Nature* 412 (6847): 607-14. <https://doi.org/10.1038/35088000>.
- Wang, Eric T., Rickard Sandberg, Shujun Luo, Irina Khrebtukova, Lu Zhang, Christine Mayr, Stephen F. Kingsmore, Gary P. Schroth, et Christopher B. Burge. 2008. « Alternative Isoform Regulation in Human Tissue Transcriptomes ». *Nature* 456 (7221): 470-76. <https://doi.org/10.1038/nature07509>.
- Wang, Hang, Becky L. Sartini, Clarke F. Millette, et Daniel L. Kilpatrick. 2006. « A Developmental Switch in Transcription Factor Isoforms During Spermatogenesis Controlled by Alternative Messenger RNA 3'-End Formation<sup>1</sup> ». *Biology of Reproduction* 75 (3): 318-23. <https://doi.org/10.1095/biolreprod.106.052209>.
- Wang, Hong-Wei. 2019. « A Link between Intronic Polyadenylation and HR Maintenance Discovered ». *Biochemistry* 58 (14): 1835-36. <https://doi.org/10.1021/acs.biochem.9b00202>.
- Wang, Qi-En, Qianzheng Zhu, Gulzar Wani, Jianming Chen, et Altaf A. Wani. 2004. « UV Radiation-Induced XPC Translocation within Chromatin Is Mediated by Damaged-DNA Binding Protein, DDB2 ». *Carcinogenesis* 25 (6): 1033-43. <https://doi.org/10.1093/carcin/bgh085>.
- Wang, Ruijia, Ram Nambiar, Dinghai Zheng, et Bin Tian. 2018. « PolyA\_DB 3 Catalogs Cleavage and Polyadenylation Sites Identified by Deep Sequencing in Multiple Genomes ». *Nucleic Acids Research* 46 (D1): D315-19. <https://doi.org/10.1093/nar/gkx1000>.



- Wang, Ruijia, Dinghai Zheng, Lu Wei, Qingbao Ding, et Bin Tian. 2019. « Regulation of Intronic Polyadenylation by PCF11 Impacts mRNA Expression of Long Genes ». *Cell Reports* 26 (10): 2766-2778.e6. <https://doi.org/10.1016/j.celrep.2019.02.049>.
- Wang, Shanshan, Ming-Yue Li, Yi Liu, Alexander C Vlantis, Jason Yk Chan, Lingbin Xue, Bao-Guang Hu, Shucaai Yang, Mo-Xian Chen, Shaoming Zhou, Wei Guo, Xianhai Zeng, Shuqi Qiu, C Andrew Van Hasselt, Michael Cf Tong, et George G Chen. 2020. « The Role of microRNA in Cisplatin Resistance or Sensitivity ». *Expert Opinion on Therapeutic Targets* 24 (9): 885-97. <https://doi.org/10.1080/14728222.2020.1785431>.
- Wawrzynow, Alicja, Janet L. Theibert, Constance Murphy, Istvan Jona, Anthony Martonosi, et John H. Collins. 1992. « Sarcolipin, the “Proteolipid” of Skeletal Muscle Sarcoplasmic Reticulum, Is a Unique, Amphipathic, 31-Residue Peptide ». *Archives of Biochemistry and Biophysics* 298 (2): 620-23. [https://doi.org/10.1016/0003-9861\(92\)90457-8](https://doi.org/10.1016/0003-9861(92)90457-8).
- Wells, James P., Justin White, et Peter C. Stirling. 2019. « R Loops and Their Composite Cancer Connections ». *Trends in Cancer* 5 (10): 619-31. <https://doi.org/10.1016/j.trecan.2019.08.006>.
- West, Steven, Nicholas J. Proudfoot, et Michael J. Dye. 2008. « Molecular Dissection of Mammalian RNA Polymerase II Transcriptional Termination ». *Molecular Cell* 29 (5): 600-610. <https://doi.org/10.1016/j.molcel.2007.12.019>.
- Weterings, Eric, Nicole S. Verkaik, Hennie T. Brüggerwirth, Jan H. J. Hoeijmakers, et Dik C. van Gent. 2003. « The role of DNA dependent protein kinase in synapsis of DNA ends ». *Nucleic Acids Research* 31 (24): 7238-46. <https://doi.org/10.1093/nar/gkg889>.
- Wilkinson, Max E., Clément Charenton, et Kiyoshi Nagai. 2020. « RNA Splicing by the Spliceosome ». *Annual Review of Biochemistry* 89 (1): 359-88. <https://doi.org/10.1146/annurev-biochem-091719-064225>.
- Williamson, Laura, Marco Saponaro, Stefan Boeing, Philip East, Richard Mitter, Theodoros Kantidakis, Gavin P. Kelly, Anna Lobley, Jane Walker, Bradley Spencer-Dene, Michael Howell, Aengus Stewart, et Jesper Q. Svejstrup. 2017. « UV Irradiation Induces a Non-Coding RNA That Functionally Opposes the Protein Encoded by the Same Gene ». *Cell* 168 (5): 843-855.e13. <https://doi.org/10.1016/j.cell.2017.01.019>.
- Wilson, T., S. Guerrette, et R. Fishel. 1999. « Dissociation of Mismatch Recognition and ATPase Activity by hMSH2-hMSH3 ». *The Journal of Biological Chemistry* 274 (31): 21659-64. <https://doi.org/10.1074/jbc.274.31.21659>.
- Wiltshaw, Eve. 1979. « Cisplatin in the Treatment of Cancer ». *Platinum Metals Review* 23 (3): 90-98.
- Wright, Bradley W., Zixin Yi, Jonathan S. Weissman, et Jin Chen. 2022. « The Dark Proteome: Translation from Noncanonical Open Reading Frames ». *Trends in Cell Biology* 32 (3): 243-58. <https://doi.org/10.1016/j.tcb.2021.10.010>.
- Xiang, Yu, Youqiong Ye, Yanyan Lou, Yang Yang, Chunyan Cai, Zhao Zhang, Tingting Mills, Ning-Yuan Chen, Yoonjin Kim, Fatma Muge Ozguc, Lixia Diao, Harry Karmouty-Quintana, Yang Xia, Rodney E Kellems, Zheng Chen, Michael R Blackburn, Seung-Hee Yoo, Ann-Bin Shyu, Gordon B Mills, et Leng Han. 2018. « Comprehensive Characterization of

- Alternative Polyadenylation in Human Cancer ». *JNCI: Journal of the National Cancer Institute* 110 (4): 379-89. <https://doi.org/10.1093/jnci/djx223>.
- Xiao, Wen, Samir Adhikari, Ujwal Dahal, Yu-Sheng Chen, Ya-Juan Hao, Bao-Fa Sun, Hui-Ying Sun, Ang Li, Xiao-Li Ping, Wei-Yi Lai, Xing Wang, Hai-Li Ma, Chun-Min Huang, Ying Yang, Niu Huang, Gui-Bin Jiang, Hai-Lin Wang, Qi Zhou, Xiu-Jie Wang, Yong-Liang Zhao, et Yun-Gui Yang. 2016. « Nuclear m6A Reader YTHDC1 Regulates mRNA Splicing ». *Molecular Cell* 61 (4): 507-19. <https://doi.org/10.1016/j.molcel.2016.01.012>.
- Yang, J. 2003. « ATM, ATR and DNA-PK: Initiators of the Cellular Genotoxic Stress Responses ». *Carcinogenesis* 24 (10): 1571-80. <https://doi.org/10.1093/carcin/bgg137>.
- Yang, Wei. 2014. « An Overview of Y-Family DNA Polymerases and a Case Study of Human DNA Polymerase  $\eta$  ». *Biochemistry* 53 (17): 2793. <https://doi.org/10.1021/bi500019s>.
- Yates, Andrew, Wasiru Akanni, M. Ridwan Amode, Daniel Barrell, Konstantinos Billis, Denise Carvalho-Silva, Carla Cummins, Peter Clapham, Stephen Fitzgerald, Laurent Gil, Carlos García Girón, Leo Gordon, Thibaut Hourlier, Sarah E. Hunt, Sophie H. Janacek, Nathan Johnson, Thomas Juettemann, Stephen Keenan, Ilias Lavidas, Fergal J. Martin, Thomas Maurel, William McLaren, Daniel N. Murphy, Rishi Nag, Michael Nuhn, Anne Parker, Mateus Patricio, Miguel Pignatelli, Matthew Rahtz, Harpreet Singh Riat, Daniel Sheppard, Kieron Taylor, Anja Thormann, Alessandro Vullo, Steven P. Wilder, Amonida Zadissa, Ewan Birney, Jennifer Harrow, Matthieu Muffato, Emily Perry, Magali Ruffier, Giulietta Spudich, Stephen J. Trevanion, Fiona Cunningham, Bronwen L. Aken, Daniel R. Zerbino, et Paul Flicek. 2016. « Ensembl 2016 ». *Nucleic Acids Research* 44 (D1): D710-16. <https://doi.org/10.1093/nar/gkv1157>.
- Yeasmin, Fouzia, Tetsushi Yada, et Nobuyoshi Akimitsu. 2018. « Micropeptides Encoded in Transcripts Previously Identified as Long Noncoding RNAs: A New Chapter in Transcriptomics and Proteomics ». *Frontiers in Genetics* 9. <https://doi.org/10.3389/fgene.2018.00144>.
- Yoshimoto, Rei, Jagat K. Chhipi-Shrestha, Tilman Schneider-Poetsch, Masaaki Furuno, A. Maxwell Burroughs, Shohei Noma, Harukazu Suzuki, Yoshihide Hayashizaki, Akila Mayeda, Shinichi Nakagawa, Daisuke Kaida, Shintaro Iwasaki, et Minoru Yoshida. 2021. « Spliceostatin A Interaction with SF3B Limits U1 snRNP Availability and Causes Premature Cleavage and Polyadenylation ». *Cell Chemical Biology* 28 (9): 1356-1365.e4. <https://doi.org/10.1016/j.chembiol.2021.03.002>.
- Zaccara, Sara, Ryan J. Ries, et Samie R. Jaffrey. 2019. « Reading, Writing and Erasing mRNA Methylation ». *Nature Reviews Molecular Cell Biology* 20 (10): 608-24. <https://doi.org/10.1038/s41580-019-0168-5>.
- Zamft, Bradley, Lacramioara Bintu, Toyotaka Ishibashi, et Carlos Bustamante. 2012. « Nascent RNA structure modulates the transcriptional dynamics of RNA polymerases ». *Proceedings of the National Academy of Sciences* 109 (23): 8948-53. <https://doi.org/10.1073/pnas.1205063109>.
- Zeman, Michelle K., et Karlene A. Cimprich. 2014. « Causes and Consequences of Replication Stress ». *Nature Cell Biology* 16 (1): 2-9. <https://doi.org/10.1038/ncb2897>.

- Zhang, Maolei, Kun Zhao, Xiaoping Xu, Yibing Yang, Sheng Yan, Ping Wei, Hui Liu, Jianbo Xu, Feizhe Xiao, Huangkai Zhou, Xuesong Yang, Nunu Huang, Jinglei Liu, Kejun He, Keping Xie, Gong Zhang, Suyun Huang, et Nu Zhang. 2018. « A Peptide Encoded by Circular Form of LINC-PINT Suppresses Oncogenic Transcriptional Elongation in Glioblastoma ». *Nature Communications* 9 (1): 4475. <https://doi.org/10.1038/s41467-018-06862-2>.
- Zhang, Shan, Boris Reljić, Chao Liang, Baptiste Kerouanton, Joel Celio Francisco, Jih Hou Peh, Camille Mary, Narendra Suhas Jagannathan, Volodimir Olexiouk, Claire Tang, Gio Fidelito, Srikanth Nama, Ruey-Kuang Cheng, Caroline Lei Wee, Loo Chien Wang, Paula Duek Roggli, Prabha Sampath, Lydie Lane, Enrico Petretto, Radoslaw M. Sobota, Suresh Jesuthasan, Lisa Tucker-Kellogg, Bruno Reversade, Gerben Menschaert, Lei Sun, David A. Stroud, et Lena Ho. 2020. « Mitochondrial peptide BRAWNIN is essential for vertebrate respiratory complex III assembly ». *Nature Communications* 11 (mars): 1312. <https://doi.org/10.1038/s41467-020-14999-2>.
- Zhang, Tinghu, Nicholas Kwiatkowski, Calla M Olson, Sarah E Dixon-Clarke, Brian J Abraham, Ann K Greifenberg, Scott B Ficarro, Jonathan M Elkins, Yanke Liang, Nancy M Hannett, Theresa Manz, Mingfeng Hao, Bartlomiej Bartkowiak, Arno L Greenleaf, Jarrod A Marto, Matthias Geyer, Alex N Bullock, Richard A Young, et Nathanael S Gray. 2016. « Covalent targeting of remote cysteine residues to develop CDK12 and 13 inhibitors ». *Nature chemical biology* 12 (10): 876-84. <https://doi.org/10.1038/nchembio.2166>.
- Zhang, Y., F. Yuan, X. Wu, M. Wang, O. Rechkoblit, J. S. Taylor, N. E. Geacintov, et Z. Wang. 2000. « Error-Free and Error-Prone Lesion Bypass by Human DNA Polymerase Kappa in Vitro ». *Nucleic Acids Research* 28 (21): 4138-46. <https://doi.org/10.1093/nar/28.21.4138>.
- Zhang, Yanbin, Fenghua Yuan, Steven R. Presnell, Keli Tian, Yin Gao, Alan E. Tomkinson, Liya Gu, et Guo-Min Li. 2005. « Reconstitution of 5'-Directed Human Mismatch Repair in a Purified System ». *Cell* 122 (5): 693-705. <https://doi.org/10.1016/j.cell.2005.06.027>.
- Zhang, Yanbin, Fenghua Yuan, Xiaohua Wu, John-Stephen Taylor, et Zhigang Wang. 2001. « Response of human DNA polymerase  $\iota$  to DNA lesions ». *Nucleic Acids Research* 29 (4): 928-35. <https://doi.org/10.1093/nar/29.4.928>.
- Zhao, Ye, Christian Biertümpfel, Mark T. Gregory, Yue-Jin Hua, Fumio Hanaoka, et Wei Yang. 2012. « Structural basis of human DNA polymerase  $\eta$ -mediated chemoresistance to cisplatin ». *Proceedings of the National Academy of Sciences of the United States of America* 109 (19): 7269-74. <https://doi.org/10.1073/pnas.1202681109>.
- Zhao, Zhaozhao, Qiushi Xu, Ran Wei, Leihuan Huang, Weixu Wang, Gang Wei, et Ting Ni. 2021. « Comprehensive Characterization of Somatic Variants Associated with Intronic Polyadenylation in Human Cancers ». *Nucleic Acids Research*, septembre. <https://doi.org/10.1093/nar/gkab772>.
- Zhao, Zhaozhao, Qiushi Xu, Ran Wei, Weixu Wang, Dong Ding, Yu Yang, Jun Yao, Liye Zhang, Yue-Qing Hu, Gang Wei, et Ting Ni. 2021. « Cancer-Associated Dynamics and Potential Regulators of Intronic Polyadenylation Revealed by IPAfinder Using Standard RNA-Seq Data ». *Genome Research*, septembre, gr.271627.120. <https://doi.org/10.1101/gr.271627.120>.

- Zheng, Xiao, Lujun Chen, You Zhou, Qi Wang, Zhuojun Zheng, Bin Xu, Chen Wu, Qi Zhou, Wenwei Hu, Changping Wu, et Jingting Jiang. 2019. « A novel protein encoded by a circular RNA circPPP1R12A promotes tumor pathogenesis and metastasis of colon cancer via Hippo-YAP signaling ». *Molecular Cancer* 18 (1): 47. <https://doi.org/10.1186/s12943-019-1010-6>.
- Zhou, Zheng, Ji Xu, Ximing Bao, Jiali Shi, Bin Liu, Yanqing Chen, et Jiping Li. 2019. « Nuclear Nrf2 Activity in Laryngeal Carcinoma Is Regulated by SENP3 After Cisplatin-Induced Reactive Oxygen Species Stress ». *Journal of Cancer* 10 (15): 3427-34. <https://doi.org/10.7150/jca.30318>.
- Zhu, Zhu, Woo-Hyun Chung, Eun Yong Shim, Sang Eun Lee, et Grzegorz Ira. 2008. « Sgs1 Helicase and Two Nucleases Dna2 and Exo1 Resect DNA Double-Strand Break Ends ». *Cell* 134 (6): 981-94. <https://doi.org/10.1016/j.cell.2008.08.037>.
- Zietlow, Laura, Leigh Anne Smith, Mika Bessho, et Tadayoshi Bessho. 2009. « Evidence for the involvement of human DNA polymerase N in the repair of DNA interstrand cross-links ». *Biochemistry* 48 (49): 11817-24. <https://doi.org/10.1021/bi9015346>.
- Zingone, Adriana, Sanju Sinha, Michael Ante, Cu Nguyen, Dalia Daujotyte, Elise D. Bowman, Neelam Sinha, Khadijah A. Mitchell, Qingrong Chen, Chunhua Yan, Phillippe Loher, Daoud Meerzaman, Eytan Ruppim, et Bríd M. Ryan. 2021. « A Comprehensive Map of Alternative Polyadenylation in African American and European American Lung Cancer Patients ». *Nature Communications* 12 (1): 5605. <https://doi.org/10.1038/s41467-021-25763-5>.
- Zotter, Angelika, Martijn S. Luijsterburg, Daniël O. Warmerdam, Shehu Ibrahim, Alex Nigg, Wiggert A. van Cappellen, Jan H. J. Hoeijmakers, Roel van Driel, Wim Vermeulen, et Adriaan B. Houtsmuller. 2006. « Recruitment of the Nucleotide Excision Repair Endonuclease XPG to Sites of UV-Induced Dna Damage Depends on Functional TFIIH ». *Molecular and Cellular Biology* 26 (23): 8868-79. <https://doi.org/10.1128/MCB.00695-06>.